A stochastic framework for modeling the population dynamics of convective clouds

Samson Hagos[*], Zhe Feng[*], Robert S. Plant[@], Robert A. Houze Jr[#], and Heng Xiao[*]*

[*]*Pacific Northwest National Laboratory Richland WA99354*

[#]*University of Washington, Seattle, WA 98195*

[@]*Department of Meteorology, University of Reading, UK.*

Submitted to Journal of Advances in Modelling Earth Systems, October 15 2017

Corresponding Author Address

Samson Hagos

Pacific Northwest National Laboratory

902 Battelle Boulevard

Richland WA 99352

Email: samson.hagos@pnnl.gov

Abstract

A stochastic prognostic framework for modeling the population dynamics of convective clouds and representing them in climate models is proposed. The framework follows the non-equilibrium statistical mechanical approach to constructing a master equation for representing the evolution of the number of convective cells of a specific size and their associated cloud-base mass flux, given a large-scale forcing. In this framework, referred to as STOchastic framework for Modeling Population dynamics of convective clouds (STOMP), the evolution of convective cell size is predicted from three key characteristics of convective cells: (i) the probability of growth, (ii) the probability of decay, and (iii) the cloud-base mass flux. STOMP models are constructed and evaluated against CPOL radar observations at Darwin and convection permitting model (CPM) simulations.

Multiple models are constructed under various assumptions regarding these three key parameters and the realisms of these models are evaluated. It is shown that in a model where convective plumes prefer to aggregate spatially and the cloud-base mass flux is a non-linear function of convective cell area, then the mass flux manifests a recharge-discharge behavior under steady forcing. Such a model also produces observed behavior of convective cell populations and CPM simulated cloud-base mass flux variability under diurnally varying forcing. In addition to its use in developing understanding of convection processes and the controls on convective cell size distributions, this modeling framework is also designed to be capable of serving as a non-equilibrium closure formulations for spectral mass flux parameterizations.

1. Introduction

In traditional cumulus parameterizations, cumulus convection is assumed to be in statistical equilibrium with a slowly varying environment and to respond to any changes in forcing almost instantaneously and deterministically with little memory or internal variability of its own. Such an assumption implicitly requires model grid columns to be large compared to the mean distance between convective elements so that the columns contain a meaningful number of updrafts. However, it has been known since the Global Atmospheric Research Program's Atlantic Tropical Experiment [GATE, *Houze and Betts*, 1981] that large, long-lasting mesoscale convective systems (MCSs) make important contributions to heat, moisture and momentum budgets, and that scale-separation is not present in either time or space [*Moncrieff*, 2010]. Advances in computational resources have made operational global weather and experimental climate models with spatial resolution ≤ 10 km [*Hólm et al.*, 2016; *Satoh et al.*, 2014] possible, which makes such assumptions even more problematic, not least because stochastic effects become increasingly relevant [eg. *Plant and Craig*, 2008; *Jones and Randall*, 2011]. On the other hand, radar, aircraft and satellite observations, as well as cloud-resolving limited-area simulations are providing deeper understanding of processes within the cloud population and interactions with the environment at various scales [*Burleyson et al*., 2016; *Heinze et al*., 2017].

These challenges, advances and opportunities require rethinking of the community's approach, specifically for the issues of departures from quasi-equilibrium, internal cloud population dynamics and the associated stochasticity [*Randall, 2013; Holloway et al., 2014*]. In order to discuss these challenges and efforts at addressing them and put this work in context, we consider the original pair of energy equations of *Arakawa and Schubert* [1974] for an ensemble of convective updrafts, written here in discrete form.

$$\frac{dA_i}{dt} = -\sum_{j=1}^{N} \gamma_{ij} M_{Bj} + F_i \qquad (1)$$

$$\frac{dK_i}{dt} = A_i M_{Bi} - \frac{K_i}{\tau_d} \qquad (2)$$

Here the subscript i represents a convective cell (for example, with a given entrainment rate, or, as we shall later consider here, with a given cell size). As will be discussed in detail in the next section, a cell is defined as a contiguous area (a set of connected pixels) within which much of upward mass transport and convective precipitation takes place. F_i is the external forcing acting on cloud type i K_i is the convective kinetic energy, A_i is the vertical integral of in-plume buoyancy (also called the 'cloud work function') and M_{Bi} is the cloud-base mass flux. γ_{ij} represents the effect of a unit of mass flux associated with cloud type *j* on the potential energy for type i. Although negative values can arise [*Yano and Plant*, 2012a], the elements of γ are often assumed to be positive in accordance with the overall stabilizing effect of convective clouds: i.e convective damping via warming of the troposphere..

The most common and drastic simplifications to the above equations are to average over the ensemble of cloud types in order to produce a "bulk plume," and to apply the quasi-equilibrium assumption. For example, in Eq. 1, the quasi-equilibrium assumption means that the two terms on the right-hand side approximately balance, while the first of these terms is greatly simplified because the interaction matrix γ_{ij} reduces to a single quantity γ that multiplies the bulk cloud-base mass flux. These have been very common simplifications in convective parameterizations [e.g. *Fritsch and Chappell*, 1980; *Tiedtke*, 1989; *Gregory and Rowntree*, 1990]. Over recent years, stochastic fluctuations about an equilibrium solution have been proposed and included in some convective parameterizations, based on either a bulk plume formulation [*Palmer et al.*, 2009; *Sakradzija et al*., 2016] or allowing a spectrum of cloud types [*Plant and Craig*, 2008; *Wang and Zhang*, 2016].

Early efforts at removing the quasi-equilibrium assumption were made by *Randall and Pan* [1993] and *Pan and Randall* [1998] who explained how a diagnostic relationship between convective kinetic energy and cloud-base mass flux would be sufficient to close the pair of equations and allow them to be used prognostically. They postulated the form $K = \alpha M_B^2$ and showed that in General Circulation Model (GCM) tests the parameter α controls the relative frequency of shallow convection. Later *Yano and Plant* [2012b] argued for $K = \beta M_B$ as a more appropriate postulate and demonstrated that for that relationship under constant forcing, a nonlinear oscillation can occur between 'discharging' and 'recharging' states.

Another development from Eqs. 1 and 2 is to try to solve them for the population dynamics of clouds and obtain the spectral distribution of mass flux M_{Bi} for a set of types i. One advantage of a spectral approach to representing convective clouds is that microphysical processes, aerosol and radiative processes can be considered for individual cloud types rather than as averages over the population. Thus, size-dependent non-linear processes (entrainment/detrainment for example) can be treated directly. However, it should still be recognized that a steady-plume hypothesis is normally made in the representation of each type without any consideration of the individual cloud lifecycle [*Yano*, 2015]. Moreover, the advantages come with the challenge of understanding and modeling the cloud-cloud and cloud-environment interactions that shape the cloud spectrum. In the *Wagner and Graf [2010]* scheme, for example, the cloud types are assumed to compete in a manner similar to competitive Lotka-Volterra [*Volterra*, 1928] systems for population dynamics.

Their system is integrated so as to satisfy convective quasi-equilibrium conditions [*Plant and Yano*, 2011]. In the European Center Hamburg Atmospheric Model [ECHAM, *Roeckner et al*., 2003], the *Wagner and Graf* [2010] scheme improves the spatial and temporal variability of convective events in comparison to a bulk mass flux scheme.

Stochastic models of convective clouds using birth-death processes and interactions among them were introduced by *Khouider et al*.[2010, 2014], and recent developments of the approach can be found in *Gottwald et al*. [2016] and *Dorrestijn et al*. [2016]. These multicloud models consider three modes of convective heating (deep, congestus and stratiform) and are concerned with the interplay between these modes and their couplings to aspects of the large-scale flow, particularly moisture and large-scale vertical velocity [e.g. *Peters et al*., 2017]. *Bengtsson et al. [2013, 2016]* use a cellular automata model for convective area fraction as a way to introduce stochasticity and estimate uncertainty associated with lateral communication of convection fluctuations in a numerical weather prediction model. They show some improvement in short term forecast of accumulated precipitation.

Plant [2012] also proposed a stochastic cloud population model. It evolves according to probabilities of transitions using a master equation, and the focus was to make direct contacts with Eqs. 1 and 2 (for different assumed K-M_B relationships) in the limit of large system size. Although studied for a single cloud type in an idealized setting, the method allows prognostic treatments that are consistent with the energy equations to be combined with the stochastic nature of the assumed underlying processes. In this study, we also consider a probabilistic representation for the non-equilibrium dynamics of cloud populations. Of major interest here is the resulting distribution of the full spectrum of cloud sizes, and its development due to evolution of the imposed forcing. Those are not issues addressed by the previous work described above. In radiative convective equilibrium, distributions of cloud number and mass flux can be predicted from equilibrium statistical mechanics [*Craig and Cohen*, 2006] and these have proved robust in cloud-resolving simulations even for convection exhibiting some organization [*Cohen and Craig*, 2006] or with some departures from equilibrium [*Davoudi et al*., 2010]. However, our investigations here will include consideration of diurnally varying forcings that may be far from equilibrium. We attempt to construct possible representations for the evolution of the cloud size distribution over the day. The representations considered will be informed by analysis of radar observations and convection permitting model (CPM) simulations. The framework that is developed is designed to contribute to (i) the testing of hypotheses regarding the roles of specific physical processes that could influence the evolution of the size distribution of convective cells including direct cloud-cloud interactions and (ii) the development of a stochastic, prognostic parameterization that includes a realistic representation of cloud population dynamics.

In the next section, the observational data and model simulations used are described. In the subsequent section, a detailed description of the modeling framework and the behavior of multiple models constructed under various simplifying assumptions are examined.

2. Description of observational data and CPM simulations

In order to inform the development of the stochastic framework, we examine the cloud population dynamics from radar observations and CPM simulations. While the primary purpose of the study is not to extensively compare radar observations and the CPM simulations, as will be shown throughout the paper, their consistency provides us with confidence on the conclusions inferred to develop the stochastic framework. The radar observations used in this study are obtained from the C-band polarimetric (CPOL) scanning radar located at Darwin, Australia [*Kumar et al*., 2013a,b]. We use three wet seasons of CPOL data collected between November

2005-March 2006, October 2006-March 2007, and December 2009-April 2010. In total, approximately 11,760 hours of CPOL volumetric data are used to construct the cloud population statistics. The CPOL radar collects a 3-D volume of data within a 150 km radius (Fig. 1a) every 10 min. Each volume scan consists of a total of 16 sweeps at elevation angles ranging from 0.5º to 42º. The sweep data are then gridded to a Cartesian grid of (ΔX, ΔY, ΔZ) = (2.5, 2.5, 0.5) km. The vertical extent of the gridded data is from 0.5 to 20 km. Although the CPOL radar collects polarimetric observations that provide insights into microphysical processes, only the horizontal reflectivity is used for this study. For more details of the CPOL radar data processing, see *Kumar et al.*, [2013a].

To identify convective cells from the CPOL radar data, the *Steiner et al*. [1995] algorithm is applied to the radar reflectivity field at 2.5 km height. The Steiner algorithm mainly uses the horizontal texture (i.e., peakedness) of radar reflectivity to identify areas of intense radar echo return and designates them as convective. An individual radar pixel is classified as convective if 1) its reflectivity value is above 40 dBZ, or 2) it exceeds its area-averaged background reflectivity within an 11 km radius centered on the pixel. Surrounding pixels up to 5 km radius (based on background reflectivity value) can also be assigned as convective. All connected convective pixels are grouped and, a collection of at least five connected pixels is labeled as a convective cell. Thus, the smallest cells that are considered to be resolved by the gridded CPOL radar have an area of 31.5 km^2.

For each convective cell, the averaged 10 dBZ echo-top height of the cell is determined as a proxy for the intensity of the convection (i.e., deeper echo-top heights indicate stronger updrafts lofting larger particles up in the troposphere). For better estimate of echo-top heights, only data in the range 20-140 km from CPOL are analyzed. The radar processing procedures described above have been used in previous studies over Darwin region [e.g., *Kumar et al.*, 2013] and over tropical Indian Ocean region [e.g., *Hagos et al.*, 2014a, b].

The CPM component of this study focuses on the 1 January 2006 to 28 February 2006 monsoon period, within the first CPOL season. The Weather Research and Forecasting (WRF) model [*Skamarock et al*., 2008] is used, with details of the model set-up provided in Table 1 and the simulation domain shown in Fig. 1b. The domain covers the region between 25°S-5°S and 120°E-150°E, with 2.5 km grid spacing and the simulation is run without a cumulus parameterization. Lateral and surface boundary conditions are obtained from ERA-Interim reanalysis [*Dee et al*., 2011] and are updated 6 hourly. Sea surface temperatures are prescribed and are also updated 6 hourly. The reflectivity from the model is calculated online from a particle size distribution using a radar simulator [*Smith et al*., 1984]. Evaluations of the model performance in representing the radar-observed aspects of the convection are discussed throughout this paper, along with the analysis of the results. In order to increase the sample size of simulated radar reflectivity from the two-month long model simulation, thirteen additional "virtual radar" sites are considered along the northern coast of Australia in addition to the Darwin CPOL site and the reflectivity fields from circular areas equivalent in size to the CPOL radar domain (i.e. 150 km radius) are extracted (Fig. 1b). The identification of convective cells within the domains of fourteen 'virtual radars' was done in the same way as for the observations.

For each of the convective cells identified in the simulation, the cloud-base mass flux was calculated. This was done in two steps. First, the cloud-base height was identified for every grid column c identified as part of a convective cell. The base was defined as the lowest level z_{bc} for which the cloud liquid water content $q_{cloud}(z)$ was both larger than a threshold value of 10^{-5} kg kg^{-1} and was below the level of peak q_{cloud}. Second, the cloud-base mass-flux per unit area for a convective cell was calculated as:

$$m_b = \frac{1}{N}\sum_{c=1}^{N} \rho\,(z_{bc}) w(z_{bc}) \quad (3)$$

where ρ is the density in kg m^{-3} and w the vertical velocity in m s^{-1} for the N individual grid columns comprising the cell. The cell mass flux is then

$$M_B = m_b a = m_b N \Delta a \quad (4)$$

with a the cell area and Δa the area of a grid column. The distinction between the cell mass flux per unit area, m_b, and the cell mass flux, M_B, is important for the discussion later.

3. Stochastic modeling framework

(a) General description

As discussed in the Introduction, this study aims to develop a modeling framework for representing the evolution of the size distribution of convective cells. The general framework presented in this section is common to the hierarchy of models we develop in this study. In the subsequent section, specific models are constructed and evaluated against observations and the CPM simulations.

We define a state of the cloud population in a given domain by the size distribution of the convective cells: i.e., a vector $\boldsymbol{n}$, with elements n_i denoting the number of cells of each possible size $a_i = i\Delta a$, where Δa is the area of a single grid point. Often in statistical mechanics population-dynamics problems like the one at hand can be formulated in the form of a master equation. We follow that approach below, although the dynamics will be evaluated numerically and we do not seek an analytic solution. In this context, the master equation for the evolution of n_i is given by:

$$\frac{dn_i}{dt} = \sum_{j \neq i} \left(W_{ji} n_j - W_{ij} n_i \right) \quad (5)$$

W_{ji} is a transition rate from size a_j to size a_i and W_{ij} is a transition rate from size a_i to size a_j. It is convenient to define a size bin of zero area, a_0, with $n_0 = 1$, so that Eq. 5 describes the evolution for all $i \geq 1$and where W_{0i} represents the formation of new clouds of size a_i and W_{i0} represents the removal of clouds of size a_i. For non-zero values of the indices, the first term on the right hand side represents the gain in the number of clouds of size a_i that have evolved from other sizes a_j while the second term represents the loss in the clouds of size a_i due to their evolution into clouds of other sizes a_j. For the origins, derivation and applications of the master equation in other fields see for example *Gardiner* [2004], van *Kampen* [2007], and *Liang and Qian* [2010]. In order to solve this set of coupled differential equations one has to know the transition rates under the given environmental conditions. Obviously W_{ij} and W_{ji} are not known for general conditions for all pairs of cell sizes but here we consider whether some simple assumptions may nonetheless be sufficient to produce *W* elements that give a reasonable description of the size distribution.

At any given time, we consider a number of convective pixels *p* within the domain of interest, so that the fraction of the domain *f* covered by convective pixels is:

$$f = \frac{p\Delta a}{A_{domain}} \quad (6)$$

where A_{domain} is the area of the grid box. The model is evolved by removing and adding pixels with rates that are determined respectively by the first and second terms on the right-hand side of the following equation:

$$\frac{dp}{dt} = \frac{1}{a_1 m_{b1}} \left(-\frac{\sum_i M_{Bi}}{\tau} + \bar{F} \right) \quad (7)$$

$a_1 = \Delta a$ is the area of a single pixel and $m_{b1}$is the cloud-base mass flux per unit area for such a pixel. The forcing $\bar{F}$ with dimensions of mass flux per unit time dictates the rate of formation of new pixels and is assumed to be provided as an input to the model according to the prevailing large-scale conditions. For application in a GCM, $\bar{F}$ could be provided by an existing equilibrium-based closure calculation. When divided by the denominator it becomes the number of pixels being added to the system per unit time. M_{Bi} represents the cloud-base mass flux associated with the convective cells of size a_i and the removal rate is assumed to be such as to produce a simple Newtonian damping of the mass flux with an associated convective relaxation timescale τ. The damping characterizes the dissipation of momentum and thermal contrasts as the convective air mixes with the environment and instability is removed. The key assumption in (7) is that the imbalance between cloud-base mass flux and the external forcing controls the amount of instability for further growth of existing convective cells or formation of new cells. However the equation does not specifically determine how this instability is distributed spatially and hence the size distribution of the cells (i.e., the connections among convective pixels or lack thereof). This process presumably involves internal variability as well as some degree of randomness. Furthermore note that Eq. 7 is perforce an approximation, since the number of pixels is an integer which is written above as a continuous variable. Whenever a pixel is added or removed in the model, it is further necessary to specify how that relates to the existing state vector $\boldsymbol{n}$ in order to complete the definition of the transition matrix elements W.

Equation 7 is inspired by Eq. 1 with some key similarities and differences. A destabilizing role of the forcing and a stabilizing role of the cloud-base mass flux are preserved in Eq. 7 but we assume the large-scale forcing to be manifest directly in terms of the resulting area fraction of convection rather than via an instability measure. In other words, the forcing for pixel number in Eq. 7 is assumed to be related to the instability forcing in Eq. 1 by a factor of the form $1/\gamma\tau$ that is treated as constant. The obvious advantage of framing the forcing in this way is that the area fraction is directly observable using radars. Moreover the large-scale forcing only determines the evolution of the total convective area fraction and does not specifically determine what cloud sizes/types will be produced. Rather, the size distribution is assumed to be controlled by internal cloud population dynamics that we aim to model below. Instead of using Eq. 2 and an ansatz for the relationship between cloud-base mass flux and kinetic energy, we make use of the CPM results (with support from the observations) to specify the relationship between cloud-base mass flux and cell size as will be discussed below.

To determine the relation of an added pixel to the existing ones, we define a probability of growth vector $\boldsymbol{G}$ such that $G_{i=0}$ represents the probability that the new pixel will be located in free space away from existing cells while $G_{i>0}$ represents the probability that the new pixel will be located adjacent to an existing cell of size a_i and so will constitute growth of that cell. The probability that the pixel will land on a cloud free space can be expressed as

$$G_0 = 1 - \sum_{i>0} G_i \quad (8)$$

If a pixel is added to the free space then the state vector is updated by

$$n_1(t+dt) = n_1(t) + 1 \quad (9)$$

whereas if a convective cell of size a_i gains a pixel according to this procedure then the state vector is updated by

$$n_{i+1}(t+dt) = n_{i+1}(t) + 1 \quad ; \quad n_i(t+dt) = n_i(t) - 1 \quad (10)$$

Similarly the probability of decay vector $\boldsymbol{D}$ is defined as D_i for $i>0$ as the distribution of the probability that cells of a given size will lose a pixel when a pixel is removed from the domain. If a pixel is removed from a cell of size $a_{i>1}$ then the corresponding state vector update is

$$n_i(t+dt) = n_i(t) - 1 \quad ; \quad n_{i-1}(t+dt) = n_{i-1}(t) + 1 \quad (11)$$

whereas the removal of a single-pixel cell corresponds to the update

$$n_1(t+dt) = n_1(t) - 1 \quad (12)$$

The final $\boldsymbol{n}(t+dt)$ size distribution is obtained when all the dp pixels are added/removed to the domain one at a time according to the procedure discussed above. The flow chart in Figure 2 summarizes the procedure. For a given time-step, which in this case is 10 minutes (motivated by the amount of time it takes for the CPOL radar to make a full circle), the given forcing determines the number of pixels to be added. These pixels are added one at a time as discussed above. The cells that gain these pixels are randomly drawn according to the probability of growth ***G,*** and ***n*** is updated. The number of pixels to be removed is determined by the cloud-base mass flux and they are removed by the process above. The cells that lose pixels are randomly drawn according to the probability of decay ***D,*** and ***n*** is once again updated. The final ***n*** is then used to calculate the new cloud-base mass flux, $\boldsymbol{G}$ and $\boldsymbol{D}$ for use in the next time step. In order to use Eq. 7, it remains to specify a relationship for the cloud-base mass flux M_{Bi}, as a function of the n_i cells in that area category a_i. Two different possibilities for such a relation will be considered in the models below.

Specific models constructed under this framework are defined by the assumed functional forms of the probability of growth vector $\boldsymbol{G}$, the probability of decay $\boldsymbol{D}$ and the cloud-base mass flux relationship. Hereafter we refer to these models as STOchastic Models for Population dynamics of convective clouds (STOMP). Below we present and discuss the specific models, their corresponding assumptions, and evaluate their degree of realism. The consideration of $\boldsymbol{G}$ and $\boldsymbol{D}$ leads to a tridiagonal transition matrix which does not take account of (for example) merging and splitting of pre-existing cells. In the future, we aim to explore further populating the transition matrix with observation-based and physically-sound elements to represent such processes.

(b) A uniform probability model (STOMP-UP)

In the uniform probability (UP) model, we assume that new pixels can land anywhere in the domain independent of the spatial distribution of the existing pixels. In other words, the existing convective cells have no effect on where the new pixel is added. As we show and discuss below, such a model excludes important processes that are likely to be important for the cloud population dynamics, but it constitutes a useful base case for later developments. The growth vector in this model is thus defined only by the areas currently occupied by the corresponding cells: specifically, the probability that an existing cell of size a_i will grow by acquiring the new pixel is

$$G_i = \frac{n_i a_i}{A_{domain}} \qquad (13)$$

and the probability of formation of a new single-pixel cell G_0 is given by the probability that the pixel lands on the convection-free area, which is related to the convective area fraction f as

$$G_0 = 1 - f \quad (14)$$

Similarly the decay vector is defined so that all convective pixels in the domain have equal probability of being removed, such that

$$D_i = \frac{n_i a_i}{A_{domain}} \quad (15)$$

A relationship between convective cell size and cloud-base mass flux is also needed and the simplest possibility is to assume a linear relationship. This is consistent with the common assumption that cloud-base mass flux variations are dominated by the variation in the total area fraction and that variation in vertical velocity is secondary [e.g. *Robe and Emanuel*, 1996; *Kumar et al.*, 2015]. Therefore the cloud-base mass flux per unit area, m_b in Eq. 4, is set to be constant, $m_{bi} = m_b = 0.78$ kg m^{-2} s^{-1}, a mean value obtained by averaging the cloud-base mass flux per area obtained from all of the convective cells in the CPM simulation, regardless of their size.

Before discussing the behavior of this model, the nature and magnitude of the forcing deserves a brief discussion. No particular assumption is made about the origin of the forcing other than it maintains a certain amount of average cloud-base mass flux in long-term sense while maintaining temporal behavior of interest. In this particular study it either follows the solar cycle or it is constant in time. It is imposed on the system in a form of a rate of change of cloud-base mass flux (Eq. 7). Its long-term mean is given by a domain-average cloud-base mass flux obtained from the CPM simulation of 0.01 kg m^{-2} s^{-1} divided by the prescribed adjustment time τ. This form of forcing is meant to make the coupling of stochastic model to a broad range of traditional cumulus parameterizations rather straightforward. Given the rate of change of deterministic mass flux from a traditional closure, this model would produce the stochastic cloud-base mass flux without any reference to how the deterministic mass flux is calculated in the first place.

STOMP-UP is run for 10 years with a diurnally-varying forcing that mimics the solar cycle and its behavior is examined for two adjustment times of $\tau = 1$ hr and $\tau = 4$ hr. Such values for the adjustment timescale are consistent with values found in the literature for weak-temperature gradient studies of the interactions of convection and the large scale [e.g. *Daleu et al* 2015] and are representative of the time taken for gravity wave signals to propagate across the domain and adjust the large-scale atmospheric state. Figure 3 shows the mean diurnal cycle of the prescribed forcing (dashed line) and the response of the domain-mean cloud-base mass flux for the two adjustment times. As one might expect the lag between the forcing and the cloud-base mass flux is quite sensitive to the adjustment time: for a smaller adjustment time the mass flux is closer to the phase of the forcing, and the model would reduce to quasi-equilibrium for $\tau \rightarrow 0$. With $\tau = 4$ hr the mass flux lags behind the forcing by about three hours in agreement with the CPM simulated diurnal cycle of the cloud-base mass flux.

Note that since the cloud-base mass flux is a linear function of cell area (i.e. the mass flux per area is independent of cell size by design), the total cloud-base mass flux in this case depends only on the total convective area fraction, and not on the cell size distribution. Nonetheless it is instructive to compare the cell size distribution from the stochastic model (STOMP-UP) with those obtained from radar observations and the CPM simulation. That comparison is shown in Figure 4 as a function of the total convective area fraction. Since the numbers of convective cells in the various size bins cover a broad range of scales the frequency of cells is shown on a log-scale. It is immediately apparent that the uniform probability model greatly underestimates the frequency of large cells: for example, cells larger than 100 km^2 are practically absent. Clearly chance alone cannot explain the existence of large convective cells found in both the CPOL observations and the CPM simulation. Rather some physical mechanism must exist that favors the formation of convective pixels in the neighborhood of existing cells and hence allows growth of large cells. In other words the empty spaces among convective cells must be less favorable for the formation of new convection than what a uniform probability suggests, and the STOMP-UP model likely underestimates the probability of existing cells growing as their lifecycle develops ($G_{i>0}$) and overestimates probability of new cell formation ($G_{i=0}$).

It is well known that formation of new convective cells is not random. *López* [1973, 1976, 1977] and *Houze and Cheng* [1977] showed that the smaller convective cell sizes (below mesoscale dimension) over tropical oceans follow a lognormal rather than a normal distribution. *López* [1976] demonstrates mathematically how the lognormal distribution is the frequency distribution of a variable that is subject to the law of proportionate effects, i.e., a variable whose change in value at any step of a process is a random proportion of the previous value of the variable. This interpretation is discussed in the book of *Aitcheson and Brown* [1957], who traced the interpretation back to much earlier statistical work. If the change in a value of a variable *x* is a random proportion of its current value, then after *n* steps the logarithm of *x* is normally distributed. Thus, it is evident that the growth mechanism of cells is important for determining their population statistics. The growth of a cell is not a completely random amount but likely depends on the current size of the cell.

While not accurate, the uniform probability stochastic model is informative to the extent that it identifies the limitations of a purely random process for convective cell formation and growth. In the next subsection, we take a closer look at the CPOL observations and CPM simulation to obtain a deeper insight into aspects of the physics missing in the simple stochastic model and develop a more complex version that aims to address these issues.

(c) An aggregation probability model (STOMP-AP)

As discussed above, an obvious limitation of the STOMP-UP model is that the uniform probability assumption leads to a large number of isolated convective cells. These cells do not grow by chance because they cover only a small fraction of the domain. In reality however, small cells grow quite readily and certainly more strongly than their size suggests (Figure 4). Thus a physical mechanism for growth has to be incorporated, allowing convective pixels to aggregate into fewer, larger cells. Another important issue to consider is the lifecycle of convective cells. In STOMP-UP, it is assumed that the convective cells grow by acquiring the pixels assigned to them randomly with probability proportional to the fraction of the domain they cover. If that is the case, the mean size of convective cells in a scene at any time is proportional to the number of cells in the scene. Figure 5 shows the diurnal cycles of the number of convective cells and mean cell sizes from STOMP-UP model compared with those from the CPM simulation and the CPOL observations. In addition to the expected differences in the magnitude of size and number of cells, there is a phase difference in the diurnal cycle. For STOMP-UP, the evolution of the number of convective cells and mean cell size are in phase while the larger cells appear several hours after the peak number of cells for the CPM or for the CPOL radar observed cells.

One potential growth mechanism arises through humidification by detrainment from the clouds. As *Cohen and Craig* [2004] and *Craig and Mack* [2013] note, the subsidence effect of a convective cell is more or less uniformly distributed in the surrounding space through the rapid action of gravity waves, while the moistening effect is a much slower process because moisture has to be carried away from the cell by much slower advection processes. This could make environments near existing convection relatively humid and so potentially more favorable for the development of new convection. In an idealized modeling study, *Craig and Mack* [2013] demonstrated that the incorporation of such a distinction between the warming and moistening effects of convection can lead to the formation of larger dry and moist areas through a process which they refer to as coarsening.

A simple way to represent a localized moistening process (or lifecycle processes in the development of cells, other indeed any other processes which favor the local growth of convection) in our framework is to modify the probability of growth vector $\boldsymbol{G}$. Specifically we introduce a single parameter δ to describe the relative probability of growth of existing cells to the formation of new cells. Eqs. 13 and 14 are modified to

$$G_{i>0} = \frac{\delta n_i a_i}{A_{domain}} \quad (16)$$

and

$$G_0 = 1 - \delta f \quad (17)$$

respectively. Physically δ can be interpreted as determining how likely a new convection pixel is to be formed in the vicinity of an existing convective cell in comparison to a clear environment.

One could also modify the representation of the warming and stabilizing effect of convection, but as *Craig and Mack* [2013] argued, the warming effect of convection is likely to act relatively uniformly across the whole domain. Lacking a strong motivation to do otherwise, we leave the probability of decay vector $\boldsymbol{D}$ unchanged.

We consider the effect of the δ parameter on the diurnal cycle of convective cell count and mean cell size. Figure 6 shows the diurnal cycle of these quantities for $\delta = 1$ (as in STOMP-UP), 15 and 30. As intended, with increasing δ the number of small isolated cells decreases and so the mean cell size increases. Importantly, the mean cell size peaks several hours after the cell number for the case of $\delta = 30$ rather than peaking at around the same time as in $\delta = 1$. A larger δ parameter results in qualitatively better agreement with the observations and CPM simulation. This can be interpreted as that the probability of forming a convective pixel in the vicinity of an existing cell is around 30 times more likely than forming a new isolated pixel. We could conceivably develop a more sophisticated representation of $\boldsymbol{G}$ with dependencies on environmental conditions, or on the sizes and number of existing cells, according to the dominant local enhancement process that is assumed. The constant δ parameter introduced here is simply a demonstration of the framework.

In developing the reference STOMP-UP model discussed in the last section and in modifying $\boldsymbol{G}$ as just discussed above, the cloud-base mass-flux per area was assumed to be constant and so the way in which convective pixels are spatially distributed has no bearing on the total cloud-base mass flux. Thus, the total cloud-base mass flux in the domain is proportional to the number of pixels irrespective of whether the pixels exist as individual cells or are connected into a large convective cell. This assumption can be tested using the CPM simulation results and the CPOL observations. Figure 7a shows the CPM simulation results for the mean cloud-base mass flux per unit area plotted as functions of cell area and convective area fraction. It is apparent that for given the area fraction, the cloud-base mass flux per unit area increases with the cell area. This implies that even for the same total area fraction (the same number of convective pixels), the scenes with larger cells will have a larger domain-average cloud-base mass flux. The dependence may be interpreted in terms of the entrainment and detrainment of mass into and out of convective cells [e.g., *de Rooy et al*., 2013]. Smaller convective plumes have a larger perimeter to area ratio rendering them relatively more exposed to the drier and less buoyant environment. In comparison larger convective cells are more likely to have individual updrafts enclosed within the interior of the cell and shielded from direct interactions with environmental air. Convectively induced cold pools are reported to facilitate such cloud-cloud and cloud-environment interactions [*Feng et al*., 2015]. As a consequence of the interactions, larger cells are more likely to grow deep, and this may be observed from the corresponding cell-average 10 dBZ echo-top heights (Fig 7c). Unfortunately cell-level observation of cloud-base mass flux is not directly available from the radar observations and so we consider the cell-average 10 dBZ echo-top height from the CPOL radar as a proxy. Remarkably the relationship between cell size and echo-top height from the CPM simulation is in good agreement with the observation in describing how the observed cell-average echo-top height increases with cell size, consistent with the behavior of organized convection associated with the Madden-Julian Oscillation over tropical oceans [*Hagos et al.*, 2014]. This point provides us with some confidence that the CPM simulation results are fit for the purpose of deriving a relationship between cloud-base mass flux and convective cell area.

Figure 8 shows the CPM relationship between convective cell size a_i and the cloud-base mass flux per unit area m_{bi}. The cloud-base mass flux increases by around a third up to a cell area of about 500 km^2 and more gradually for larger cell areas. A reasonable and simple approximation is provided by two linear relationships of the form

$$m_{bi} = \lambda + \mu \left(\frac{a_i - a_1}{a_1} \right) \quad (18)$$

where λ and μ are the fitted parameters corresponding to the red lines in Fig. 8. Substituting Eq. 18 into Eq. 4 results in a nonlinear relationship between the cell cloud-base mass flux M_{Bi} and cell area:

$$M_{bi} = \left[\lambda + \mu\left(\frac{a_i - a_1}{a_1}\right)\right] n_i a_i \quad (19)$$

Specifically, for $a_i \leq 500$ km^2, we use $\lambda = 0.3$ kg m^{-2} s^{-1}, $\mu = 0.023$ kg m^{-2} s^{-1}, and for $a_i > 500$ km^2, we use λ =0.54 kg m^{-2} s^{-1}, $\mu = 0.0027$ kg m^{-2} s^{-1} , respectively.

In the remainder of this section we present the model behavior with modified ***G*** under the linear and non-linear relationship between cloud-base mass flux and cell size. For brevity this version of the model will be referred to as STOMP-AP (Aggregation Probability) to highlight the fact that the probability of cell growth favors aggregation.

(i) Response to constant forcing

The behaviors of the linear and non-linear versions of the STOMP-AP model in comparison to those of STOMP-UP are examined. Eight one-year long simulations are performed. The simulations differ by whether they are linear (Eq. 4) or non-linear (Eq. 18, the relationship between cloud-base mass flux and cell area, as discussed in section 3b), the stochastic model used (STOMP-UP with δ = 1 and STOMP-AP with δ = 30 in Eq. 16 and Eq. 17) and the adjustment time (τ = 1 hr or 4 hr). A steady forcing equivalent to adding 8 and 2 pixels of area $a_1 = 31.5\, km^2$ every 15 minutes for adjustment timescales of one and four hours respectively is prescribed. In both cases the forcing results in the equilibrium cloud-base mass flux per area of 0.0097 kg m^{-2} s^{-1} and 0.0078 kg m^{-2} s^{-1}, respectively, which are comparable to the long-term mean obtained from the CPM simulation.

Time series for the area-averaged cloud-base mass flux in the simulations are shown in Figure 9. As expected, all of the linear simulations produce a steady equilibrium solution, and because the cloud-base mass flux per area is independent of cell size in these runs, any stochasticity of the cell sizes has no impact on this diagnostic. For the non-linear solutions, however, the cloud-base mass flux per unit area depends on cell size, and hence the stochasticity in the instantaneous distribution of cell sizes manifest in modifying the averaged mass flux. Using the STOMP-UP formulation, the cell size variability is small (recall Fig. 4a) and so the stochasticity remains weak and the solution for averaged cloud-base mass flux remains close to the corresponding linear simulations (Figs. 9a,b). As we investigate in more detail below, the STOMP-AP formulation produces cells covering a broader range of sizes. With $\delta = 10.0$ the amplitude of the cloud-base mass flux fluctuation increases dramatically (Figs. 9c,d).

In order to understand what is happening in this case, suppose that the system starts in a quiescent state with a few small cells. By virtue of their small number and size, these cells are unlikely to grow and instead new cells will be formed. The cloud-base mass flux increases rather gradually with the increase in the number of cells. Later, as some cells grow larger, the non-linear effects of aggregation $\boldsymbol{G}$ on one hand and the non-linear dependence on mass flux on the other result in a rapid increase of cloud-base mass flux. This can produce cloud-base mass flux that overshoots the equilibrium. The damping term $\boldsymbol{D}$ in the pixel evolution equation (Eq. 15) then becomes more important than the forcing and leads back towards a quiescent period. Such an evolution is reminiscent of the recharge-discharge cycle response to steady forcing found by *Yano and Plant* [2012], albeit with a different origin for the nonlinear growth phase. Here the nonlinearity arises because larger cells account for more than their share (by area) of the cloud-base mass flux in the system and because those larger cells are allowed to develop preferentially over small isolated cells. The adjustment timescale influences the frequency of this oscillation. Larger adjustment time-scale leads to the appearance of lower frequency of oscillation and episodes of potentially large cloud-base mass flux because some convective cells would have more time to grow.

(ii) Response to diurnal forcing

We now consider the response of the STOMP-AP model to a diurnally-varying forcing. The diurnal variation of the forcing is identical to that shown in Fig. 3, with its amplitude chosen to produce a mean cloud-base mass flux that is comparable to that obtained from the CPM simulation about 0.01 kg m^{-2} s^{-1}. For this case we set δ = 30 and τ = 4 hr.

Figure 10 shows the diurnal cycle of the cloud-base mass flux. The linear and non-linear models produce similar lags of the mass flux peak compared to the forcing, in agreement with that in the CPM simulation. However, it is noteworthy that the non-linear model produces an additional lag of around an hour relative to its linear counterpart. In the linear case, the area-averaged cloud-base mass flux depends only on the total area fraction and hence the lag is determined by the adjustment time, as previously discussed. Recall that by design the STOMP-AP model has a lag between the peak number of cells and the peak in the mean cell size (Fig. 6) as also found in observations and the CPM simulation (Fig. 5). For the non-linear model a large number of small cells provides less cloud-base mass flux than a small number of large cells, and hence the delay of STOMP-AP in producing large cell sizes also induces a delay in the mass flux peak.

To illustrate these points, an example diurnal cycle of cell number and size evolution in STOMP-AP for an arbitrary day is shown in Figure 11. In conjunction with Fig. 6, it suggests that the diurnal cycle of cloud populations can be considered in three stages.

1. With the onset of the forcing at 06 AM, small convective cells start to appear and their number increases throughout the morning, peaking around noon.
2. From early afternoon larger cells start to appear. The mean size of the cells peaks around 03 PM, by which time the number of cells is reduced because of the preferential growth of the larger cells (or, equivalently, due to the relatively unfavorable environment for new, isolated, convective pixels).
3. Late in the afternoon and through the night, as the forcing declines, the convective cells decay, with reductions to both mean number and size.

Having demonstrated some interesting and encouraging behavior from the STOMP-AP model we can now revisit our objective of using it to predict the size distribution of convective cells, and the dependence of that distribution on the total area fraction. Figure 4 shows that the STOMP-UP model greatly overestimates the number of small cells and underestimates the number of larger cells. Such limitations motivated the development of STOMP-AP, which can account for aggregation of cells and the non-linearity of mass flux dependence on cell size. Comparison of Figure 12 and Fig. 4a shows that both the linear and non-linear forms of STOMP-AP greatly improve the size distribution over STOMP-UP. There is slight difference between the two in that the non-linear model generally produces larger cells, in better agreement with observations and the CPM simulation. On the other hand, the non-linear model somewhat underestimates the frequency of smaller cells and overestimates the frequency of larger cells at the higher values of the total area fraction.

4. Conclusion

This article proposes a new prognostic framework for understanding the population dynamics of convective clouds and representing them in climate models. The approach used follows the non-equilibrium statistical mechanical approach to modelling population dynamics through a master equation. The aim is to represent the evolution of the number of convective cells of a specific size and their associated cloud-base mass flux, given a large-scale forcing for the convective area.

In this framework, referred to as STOchastic framework for Modeling Population dynamics of convective clouds (STOMP), the evolution of convective cell size is predicted from three key characteristics, which may depend on the convective cell size a_i. These characteristics are (i) the probability of growth (G_i), (ii) the probability of decay (D_i), and (iii) the cloud-base mass flux M_{Bi}. STOMP models are constructed and evaluated against CPOL radar observations at Darwin, Australia and CPM simulations. In the first model, the evolution of convective cell sizes is treated through the random addition and removal of convective pixels with a uniform probability (STOMP-UP) across the domain. Thus, a new pixel is sited irrespective of whether the location is currently convective or environmental. The cloud-base mass flux of a cell is assumed to be a linear function of cell size. It was shown that STOMP-UP underestimates the frequency of large convective cells (Fig. 4) and that it has diurnal cycles of the mean numbers of cells and mean cell sizes in phase, while for observations and the CPM the latter lags by about three hours (Fig. 5).

To overcome those deficiencies we developed the STOMP-Aggregation Probability model (STOMP-AP), in which the probability of growth is modified such that a simple aggregation parameter δ allows growth of existing cells to be favored over the formation of new ones. The aggregation parameter was chosen to reproduce the observed lag between the diurnal cycles of the mean numbers of cells and the mean cell sizes (Fig. 6). We also used CPM simulation results to develop the model further, demonstrating that cloud-base mass flux is a non-linear function of cell size (Figs. 7 and 8), and incorporated the fitted relationship within STOMP-AP. Under steady forcing, the model with aggregation and with non-linear dependence of mass flux on convective cell size can result in a solution with a stochastic oscillation: this is between a 'recharge' period when small convective cells increase in number but mass flux and mean cell size are relatively low, and a 'discharge' period when large cells appear and their damping due to their associated mass flux overwhelms the forcing, thereby reducing the number of convective cells (Fig. 9d). Under a diurnally-varying forcing, the non-linearity increases the lag between peak forcing and the mass flux peak because much of the mass flux is carried by the larger cells which form later in the afternoon (Fig. 10). Finally it was shown that the treatment of aggregation and (to a lesser extent) the non-linearity leads to much-improved cell size statistics for a given total convective area fraction (Fig. 12) compared to the linear model.

Besides its use in developing understanding of convection processes and the controls on convective size distributions, this framework is also designed to be capable of providing alternative, non-equilibrium, closure formulations for spectral mass flux parameterizations. Given the appropriate forcing from the host climate model (which could be estimated from the pre-existing closure method in many GCMs), the framework can be used to evolve the cloud-base mass flux according to the assumed cloud population dynamics. In addition it provides a spectrum of convective cell sizes, which may be used to close a spectral parameterization [e.g. *Zhang and McFarlane*, 1995; *Plant and Craig*, 2008; *Wagner and Graf*, 2010] for which cloud processes can be treated more directly at the cloud scale, with possible benefits for radiative processes, aerosols and microphysical processes [e.g. *Song et al*., 2012] involved in the formation of stratiform rain and MCSs. The cell size distribution may also be useful for the treatment of scale-awareness in grey zone parameterizations, through including only a suitable part of the convective cell size spectrum for the calculation of unresolved mass flux. Future work will involve incorporation of this framework into a mass flux cumulus scheme and examination of its impact on model climatology and variability.

Acknowledgements

This research is based on work supported by the U.S. Department of Energy Office of Science Biological and Environmental Research as part of the Atmospheric Systems Research Program. Computing resources for the model simulations are provided by the National Energy Research Scientific Computing Center (NERSC). Pacific Northwest National Laboratory is operated by Battelle for the U.S. Department of Energy under Contract DE-AC05-76RLO1830. The convective cloud population data derived from the simulation and C-POL radar observational data are available at https://portal.nersc.gov/project/cpmmjo/STOMP/.

Table 1. Convection permitting model simulation configuration

Parameter or initial condition	Configuration
Horizontal grid spacing	2.5 km
Cumulus	None
Longwave radiation	The Rapid Radiative Transfer Model [*Mlawer et al*., 1997]
Shortwave radiation	The Rapid Radiative Transfer Model [*Morcrette et al*., 2008]

Microphysics	Thompson [*Thompson et al*., 2008]
Boundary layer	Yonsei State University scheme [*Hong et al*., 2006]
Surface, initial and boundary condition data	ERA-Interim, updated every 6 hours
Number of vertical levels	30
Model top	50 hPa

References

Arakawa, A. and W. H. Schubert, 1974: Interaction of a cumulus cloud ensemble with the large-scale environment, Part I. *J. Atmos. Sci.,* **31,** 674-701.

Burleyson, C. D., Z. Feng, S. Hagos, J. Fast, L. Machado and S. T. Martin, 2016: Spatial Variability of the Background Diurnal Cycle of Deep Convection around the GoAmazon2014/5 Field Campaign Sites. *J. Appl. Meteor. Climatol.*, **55**, 1579–1598.

Bengtsson, L., and H. Körnich, 2016: Impact of a stochastic parameterization of cumulus convection, using cellular automata, in a mesoscale ensemble prediction system. *Q. J. R. Meteorol. Soc.*, **142**, 1150–1159.

Bengtsson, M. Steinheimer, P. Bechtold, and J.-F. Geleyn, 2013: A stochastic parametrization for deep convection using cellular automata. *Q. J. R. Meteorol. Soc*., **139**, 1533–1543.

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

Cohen, B. G. and G. C. Craig, 2006: Fluctuations in an equilibrium convective ensemble. Part II: Numerical experiments. *J. Atmos. Sci*., **63**, 2005-2015.

Craig, G. C. and B. G. Cohen, 2006: Fluctuations in an equilibrium convective ensemble. Part I: Theoretical formulation. *J. Atmos. Sci*., **63**, 1996-2004.

Craig, G. C. and J. M. Mack, 2013: A coarsening model for self-organization of tropical convection, *J. Geophys. Res. Atmos*., **118**, 8761–8769.

Daleu, C. L., Plant, R. S., Woolnough, S. J.,Sessions, S., Herman, M. J., Sobel, A., Wang, S., Kim, D., Cheng, A., Bellon, G., Peyrille, P., Ferry, F., Siebesma, A. P. and van Ulft, B., 2015:

Intercomparison of methods of coupling between convection and large-scale circulation. Part I: Comparison over uniform surface conditions. *J. Adv. Model. Earth Syst.*, **7**, doi:10.1002/2015MS000468.

Davoudi, J., N. A. McFarlane and T. Birner, 2010: Fluctuation of mass flux in a cloud resolving simulation with interactive radiation, *J. Atmos. Sci*., **67**, 400-418.

de Rooy, W. C., P. Bechtold, K. Fröhlich, C. Hohenegger, H. Jonker, D. Mironov, A. P. Siebesma, J. Teixeira and J.-I. Yano, 2013: Entrainment and detrainment in cumulus convection: an overview. *Q. J. R. Meteorol. Soc*., **139**: 1–19.

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut and F. Vitart, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc*., **137**: 553–597.

Dorrestijn, J., D. T. Crommelin, A. P. Siebesma, H. J. Jonker and C. Jakob, 2015: Stochastic Parameterization of Convective Area Fractions with a Multicloud Model Inferred from Observational Data. *J. Atmos. Sci*., **72**, 854–869.

Feng, Z., S. Hagos, A. K. Rowe, C. D. Burleyson, M. N. Martini, and S. P. de Szoeke (2015), Mechanisms of convective cloud organization by cold pools over tropical warm ocean during the AMIE/DYNAMO field campaign, *J. Adv. Model. Earth Syst.*, 7(2), 357-381, doi: 10.1002/2014MS000384.

Fritsch, J. M. and C. F. Chappell, 1980: Numerical prediction of convectively driven mesoscale pressure systems. Part I: Convective parameterization. *J. Atmos. Sci.*, **37**, 1722–1733.

Gardiner, C. W., 2004: Handbook of stochastic methods: for physics, chemistry and the natural sciences. Series in Synergetics Vol. 13. 3rd edition. Springer, Berlin.

Gottwald, G. A., K. Peters and L. Davies, 2016: A data-driven method for the stochastic parametrisation of subgrid-scale tropical convective area fraction. *Q. J. R. Meteorol. Soc.*, **142**, 349–359.

Gregory, D. and P. R. Rowntree, 1990: A Mass Flux Convection Scheme with Representation of Cloud Ensemble Characteristics and Stability-Dependent Closure. *Mon. Wea. Rev*., **118**, 1483–1506.

Hagos, S., Z. Feng, K. Landu, and C. N. Long, 2014:, Advection, moistening, and shallow-to-deep convection transitions during the initiation and propagation of Madden-Julian Oscillation, *J. Adv. Model. Earth Syst*., **06**, doi:10.1002/2014MS000335.

Hagos, S., Z. Feng, C. D. Burleyson, K.-S. S. Lim, C. N. Long, D. Wu, and G. Thompson, 2014: Evaluation of convection permitting model simulations of cloud populations associated with the Madden-Julian Oscillation using data collected during the AMIE/DYNAMO field campaign, *J. Geophys. Res. Atmos*., **119**, 12,052–12,068.

Heinze, R., A. Dipankar, C. C. Henken, C. Moseley, O. Sourdeval, S. Trömel, X. Xie, P. Adamidis, F. Ament, H. Baars, C. Barthlott, A. Behrendt, U. Blahak, S. Bley, S. Brdar, M. Brueck, S. Crewell, H. Deneke, P. Di Girolamo, R. Evaristo, J. Fischer, C. Frank, P. Friederichs, T. Göcke, K. Gorges, L. Hande, M. Hanke, A. Hansen, H.-C.. Hege, C. Hoose, T. Jahns, N. Kalthoff, D. Klocke, S. Kneifel, P. Knippertz, A. Kuhn, T. van Laar, A.

Macke, V. Maurer, B. Mayer, C. I. Meyer, S. K. Muppa, R. A. J. Neggers, E. Orlandi, F. Pantillon, B. Pospichal, N. Röber, L. Scheck, A. Seifert, P. Seifert, F. Senf, P. Siligam, C. Simmer, S. Steinke, B. Stevens, K. Wapler, M. Weniger, V. Wulfmeyer, G. Zängl, D. Zhang and J. Quaas, 2017: Large-eddy simulations over Germany using ICON: A comprehensive evaluation. *Q. J. R. Meteorol. Soc*., **143**, 69–100.

Holloway, C. E., J. C. Petch, R. J. Beare, P Bechtold, G. C. Craig, S. H. Derbyshire, L. J. Donner, P. R. Field, S. L. Gray, J. H. Marsham, D. J. Parker, R. S. Plant, N. M. Roberts, D. M. Schultz, A. J. Stirling and S. J. Woolnough, 2014: Understanding and representing atmospheric convection across scales: Recommendations from the meeting held at Dartington Hall, Devon, UK, 28–30 January 2013. *Atmos. Sci. Lett*., **15**, 348–353.

Hólm, E., R. Forbes, S. Lang, L. Magnusson and S. Malardel, 2016: New model cycle brings higher resolution. *ECMWF Newsletter* No. **147**, 14–19.

Hong, S.-Y., Y. Noh and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev*., **134**, 2318-2341.

Houze, R. A., Jr., and A. K. Betts, 1981: Convection in GATE. *Rev. Geophys. Space Phys*., **19**, 541-576.

Houze, R. A., Jr. and C.-P. Cheng, 1977: Radar characteristics of tropical convection observed during GATE: Mean properties and trends over the summer season. Monthly Weather Review, 105, 964-980

Jones, T. R., and D. A. Randall, 2011: Quantifying the limits of convective parameterizations, *J. Geophys. Res.*, **116**, D08210, doi:10.1029/2010JD014913.

Khouider B., J. Biello and A. Majda, 2010: A stochastic multicloud model for tropical convection. *Commun. Math. Sci.*, **8**, 187–216.

Khouider, B., 2014: A coarse grained stochastic multi-type particle interacting model for tropical convection: Nearest neighbor interactions. *Commun. Math. Sci.*, **12**, 1379–1407.

Kumar, V. V., C. Jakob, A. Protat, P. T. May and L. Davies, 2013: The four cumulus cloud modes and their progression during rainfall events: A C-band polarimetric radar perspective, *J. Geophys. Res. Atmos.*, **118**, 8375–8389.

Kumar, V. V., A. Protat, P. T. May, C. Jakob, G. Penide, S. Kumar and L. Davies, 2013: On the Effects of Large-Scale Environment and Surface Types on Convective Cloud Characteristics over Darwin, Australia, *Mon. Wea. Rev.*, **141**, 1358-1374.

Kumar, V. V., C. Jakob, A. Protat, C. R. Williams and P. T. May, 2015: Mass-flux characteristics of tropical cumulus clouds from wind profiler observations at Darwin, Australia, *J. Atmos. Sci.*, **72**, 1837–1855.

Liang J. and H. Qian, 2010: Computational Cellular Dynamics Based on the Chemical Master Equation: A Challenge for Understanding Complexity. *J. Comput. Sci. Technol.*, **25**, 154-168.

Lopez, R. E., 1973: Cumulus convection and larger scale circulations, II. Cumulus and mesoscale interactions. Mon. Wea. Rev., 101, 856-870.

López, R. E., 1976: Radar characteristics of the cloud populations of tropical disturbances in the northwest Atlantic. Mon. Wea. Rev., 104, 268-283.

López, R. E., 1977: The lognormal distribution and cumulus cloud populations. Monthly Weather Review, 105, 865-872.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, *J. Geophys. Res*., 102(D14), 16663–16682.

Moncrieff, M. W., 2010: The Multiscale Organization of Moist Convection and the Intersection of Weather and Climate. In*: Climate Dynamics:* ***Why*** *Does Climate Vary? Geophys. Monogr*., Vol. **189**, Amer. Geophys. Union, 3–26, doi:10.1029/2008GM000838.

Morcrette, J., H. W. Barker, J. N. Cole, M. J. Iacono and R. Pincus, 2008: Impact of a New Radiation Package, McRad, in the ECMWF Integrated Forecasting System. *Mon. Wea. Rev*., **136**, 4773–4798.

Palmer T.N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. ECMWF Tech. Memo. **598**, 1–42.

Pan, D. M. and D. A. Randall, 1998: A cumulus parametrization with a prognostic closure. *Quart. J. Roy. Meteor. Soc*., **124**, 949–981.

Peters, K., T. Crueger, C. Jakob and B. Möbis, 2017: Improved MJO-simulation in ECHAM6.3 by coupling a Stochastic Multicloud Model to the convection scheme, *J. Adv. Model. Earth Syst*., **9**, 193–219.

Plant R. S. and G. C. Craig, 2008: A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci*., **65**, 87-105.

Plant, R.S. and J.-I. Yano, 2011: Comments on "An Ensemble Cumulus Convection Parameterization with Explicit Cloud Treatment". *J. Atmos. Sci*., **68**, 1541–1544.

Plant R. S., 2012: A new modelling framework for statistical cumulus dynamics, *Philos. Trans. R. Soc. A*, **370**, 1041–1060.

Randall, D. A., 2013: Beyond deadlock, *Geophys. Res. Lett*., **40**, 5970–5976.

Randall, D. A. and D. M. Pan, 1993: Implementation of the Arakawa–Schubert cumulus parameterization with a prognostic closure. In: The Representation of Cumulus Convection in Numerical Models, edited by Emanuel, K. A. and D. J. Raymond, Meteorological Monographs Vol 24, Chap 11, 137-144, American Meteorological Society.

Robe, F. R. and K. Emanuel, 1996: Dependence of tropical convection on radiative forcing, *J. Atmos. Sci*., **53**, 3265–3275.

Roeckner, E., and Coauthors, 2003: The atmospheric general circulation model ECHAM5. Part I:Model description.MPI Rep. 349, 140 pp.

Sakradzija, M., A. Seifert and A. Dipankar, 2016: A stochastic scale-aware parameterization of shallow cumulus convection across the convective gray zone, *J. Adv. Model. Earth Syst*., **8**, 786–812.

Satoh, M., H. Tomita, H. Yashiro, H. Miura, C. Kodama, T. Seiki, A. T. Noda, Y. Yamada, D. Goto, M. Sawada, T. Miyoshi, Y. Niwa, M. Hara, T. Ohno, S.-I. Iga, T. Arakawa, T. Inoue and H. Kubokawa, 2014: The Non-hydrostatic Icosahedral Atmospheric Model: Description and development, *Progress in Earth and Planetary Science*, **1**, doi:10.1186/s40645-014-0018-1.

Skamarock, W. C., J. B. Klemp and J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang, W. Wang and J. G. Powers, 2008: A Description of the Advanced Research WRF Version 3, Tech. Note, NCAR/TN-475+STR. National Center for Atmospheric Research, Boulder.

Smith, P. L. (1984), Equivalent radar reflectivity factors for snow and ice particles, J. Clim. Appl. Meteorol., 23, 1258–1260.

Song, X., G. J. Zhang and J. F. Li, 2012: Evaluation of Microphysics Parameterization for Convective Clouds in the NCAR Community Atmosphere Model CAM5. *J. Climate*, **25**, 8568–8590.

Steiner, M., R. A. Houze and S.E. Yuter, 1995: Climatological Characterization of Three-Dimensional Storm Structure from Operational Radar and Rain Gauge Data. *J. Appl. Meteor*., **34**, 1978–2007.

Thompson, G., P. R. Field, R. M. Rasmussen and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev*., **136**, 5095–5115.

Tiedtke, M., 1989: A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models. *Mon. Wea. Rev*., **117**, 1779–1800.

van Kampen, N., 2007: Stochastic processes in physics and chemistry. 3rd edition. North Holland.

Volterra, V., 1928: Variations and fluctuations of the number of individuals in animal species living together. *ICES Journal of Marine Science*, **3**, 3–51.

Wagner, T. M. and H.-F. Graf, 2010: An ensemble cumulus convection parameterization with explicit cloud treatment. *J. Atmos. Sci*., **67**, 3854–3869.

Wang, Y. and G. J. Zhang, 2016: Global climate impacts of stochastic deep convection parameterization in the NCAR CAM5, *J. Adv. Model. Earth Syst*., **8**, 1641–1656.

Yano, J.-I. and R. S. Plant, 2012a: Interactions between shallow and deep convection under a finite departure from convective quasi-equilibrium. *J. Atmos. Sci*., **69,** 3463-3470.

Yano, J.-I. and R. S. Plant, 2012b: Finite departures from convective quasi-equilibrium: Periodic cycle and discharge–recharge mechanism. *Quart. J. Roy. Meteor. Soc*., **138**, 626–637.

Yano, J.-I., 2015: Formulation of the mass-flux convective parameterization. In: Parameterization of Atmospheric Convection, edited by Plant, R. S. and J.-I. Yano, Chap 7, 195-225, Imperial College Press.

Zhang, G. J. and N. A. McFarlane, 1995: Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Center General Circulation Model, *Atmos. Ocean*, **33**, 407–446.

Figure Captions

FIG. 1. (a) Example radar reflectivity snapshot at 2.5 km height showing the C-Pol radar site at Darwin, Australia. The black dot indicates the site and the red circle marks the approximate 150 km range of the radar. (b) Example simulated reflectivity snapshot at 2.5 km height showing the WRF model domain. The red circle marks the CPOL area and the black circles mark the ”virtual radar” areas, from which the reflectivities and convective cell mass fluxes simulated by the model are extracted for analysis.

FIG. 2. A flow-chart of the stochastic framework for modeling the population dynamics of convective clouds. In this framework a model is defined by how the probability of growth (**G**), the probability of decay (**D**) vectors and the relationship of mass flux with the convective cell size are specified.The green and red arrows represent a calculation at the current time-step and input from the previous time-step respectively.

FIG. 3. (a) Diurnal cycle of cloud base mass flux from the two STOMP-UP simulations (color), from the CPM simulation and the prescribed diurnal forcing (black dashed line). The forcing displayed is normalized by the daily mean and is therefore dimensionless.

FIG. 4. The logarithm of size distribution of convective cell size as a function of the total area fraction for (a) the STOMP uniform probability model, (b) CPM simulation and (c) C-Pol obser vation.

FIG. 5. The mean diurnal cycle of the number of convective cells (blue) and mean cell size (red) for (a) the STOMP uniform probability model, (b) the CPM simulation and (c) C-Pol radar.

FIG. 6. Same as Fig. 5, for STOMP Uniform Probability (δ = 1.0, solid lines) and STOMP Aggregation Probability model with δ = 15.0 (dashed lines) and δ = 30.0 (dotted lines).

FIG. 7. (a) cloud base mass flux from the CPM simulation, (b) 10dBZ echotop height as a function of total convective area fraction and cell size for the CPM simulation and (c) same for the C-Pol radar.

FIG. 8. The relationship between cloud-base mass flux per unit area and convective cell size. The red regression lines are used to parameterize the relationship.

FIG. 9. 30 day timeseries of the area average convective mass flux from (a,b) STOMP-UP and (c,d) STOMP-AP using the linear (red lines) or non-linear (blue lines) relationship between cloud-base mass flux and cell area. The adjustment time (τ) used is 1 hr for (a) and (c) and 4 hrs for (b) and (d).

FIG. 10. 100 day average of the diurnal cycle of cloud-base mass flux from STOMP-AP simula tions using the linear (red) or non-linear (blue) relationship between cloud-base mass flux and cell area. Also shown are results from the CPM simulation (green). The forcing (displayed in dashed line) is normalized by the daily mean and is therefore dimensionless.

FIG. 11. An example of the evolution of convective cell population simulated by STOMP Aggre gation Probability model under diurnally varying forcing (dashed solid line, Fig 10).

FIG. 12. The logarithm of the size distribution of convective cells as a function of area fraction for (a) STOMP-AP with linear dependence of cell mass flux on cell area, (b) STOMP-AP with non- linear dependence of cell mass flux on cell area, uniform probability model, (c) CPM simulation, and, (d) C-Pol observations.

Figure 1.

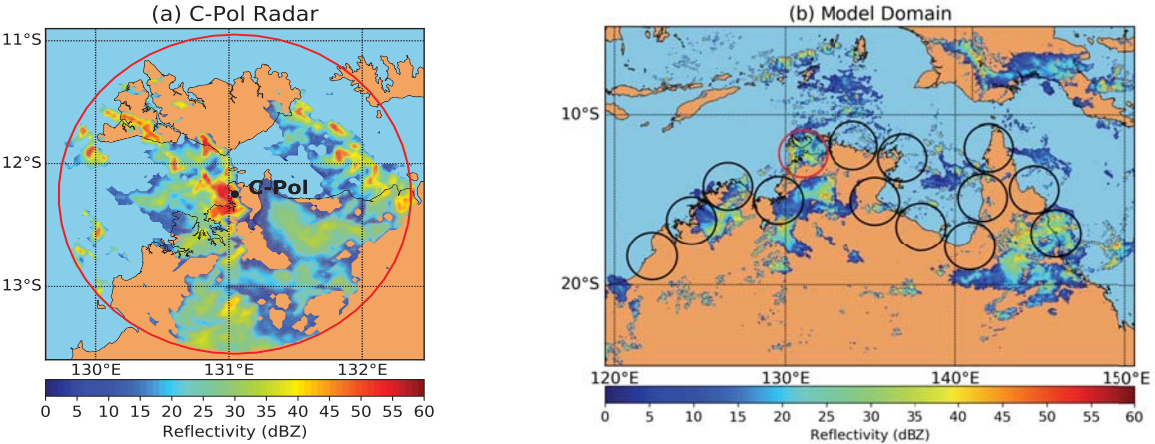

(a) C-Pol Radar
11°S
12°S
13°S
C-Pol
130°E
131°E
132°E
0 5 10 15 20 25 30 35 40 45 50 55 60
Reflectivity (dBZ)
(b) Model Domain
10°S
20°S
120°E
130°E
140°E
150°E
0 5 10 15 20 25 30 35 40 45 50 55 60
Reflectivity (dBZ)

Figure 2.

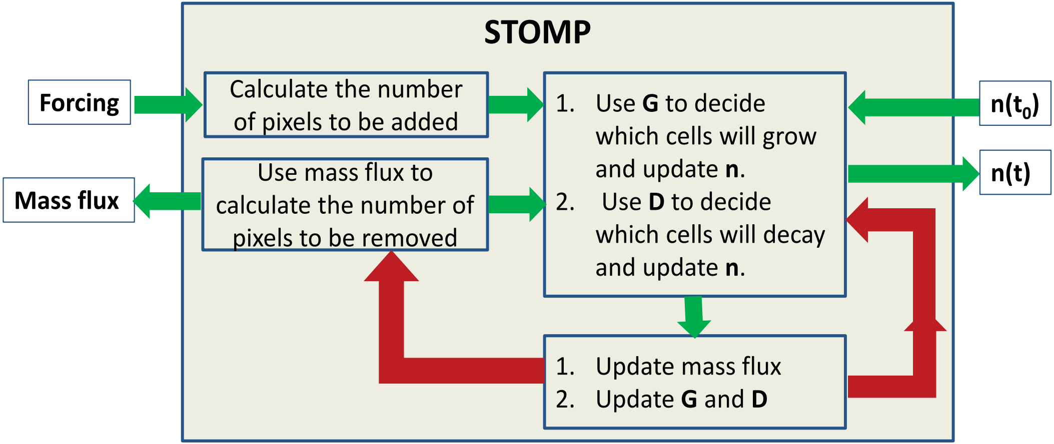
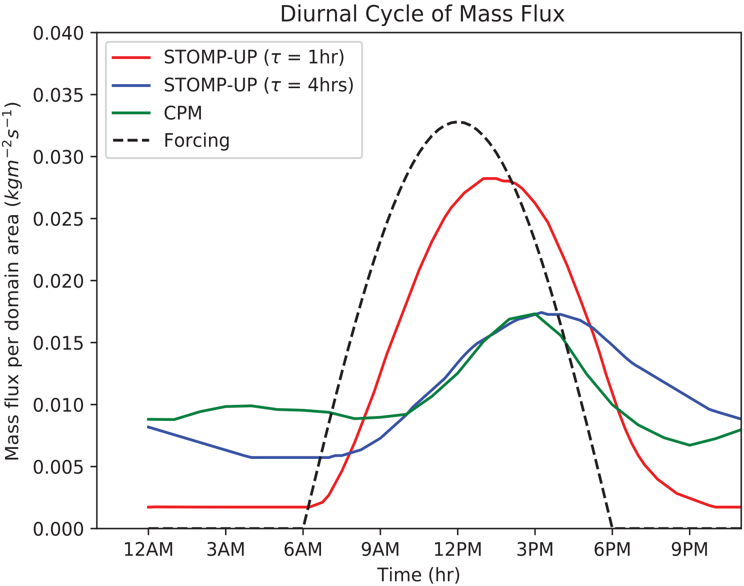
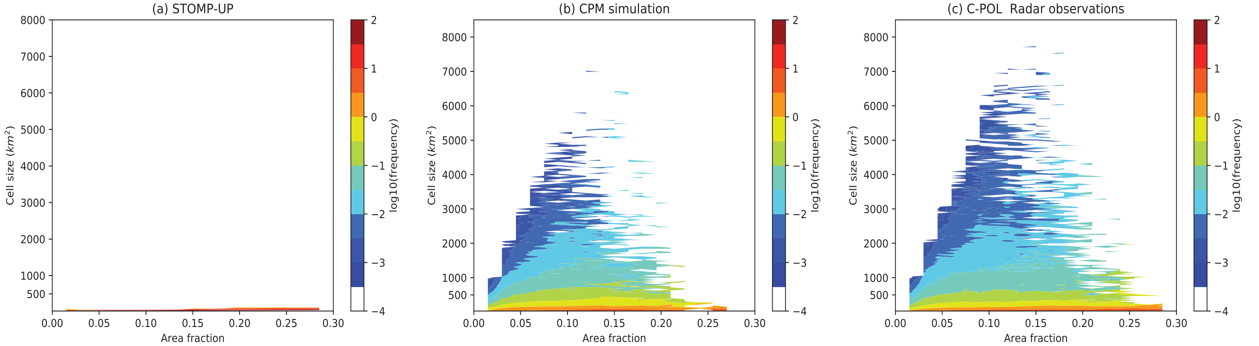

STOMP
Forcing
Calculate the number of pixels to be added
Mass flux
Use mass flux to calculate the number of pixels to be removed
1. Use **G** to decide which cells will grow and update **n**.
2. Use **D** to decide which cells will decay and update **n**.
n(t_0)
n(t)
1. Update mass flux
2. Update **G** and **D**

Figure 3.



Diurnal Cycle of Mass Flux
STOMP-UP (τ = 1hr)
STOMP-UP (τ = 4hrs)
CPM
Forcing
Mass flux per domain area ($kgm^{-2}s^{-1}$)
Time (hr)
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0.035
0.040
12AM
3AM
6AM
9AM
12PM
3PM
6PM
9PM

Figure 4.



(a) STOMP-UP
(b) CPM simulation
(c) C-POL Radar observations
Cell size (km^2)
Area fraction
log10(frequency)

Figure 5.

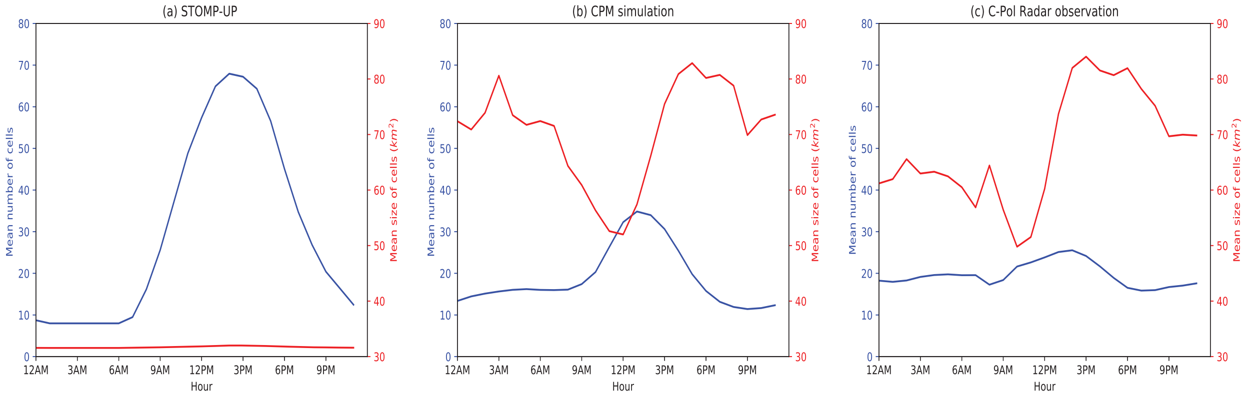
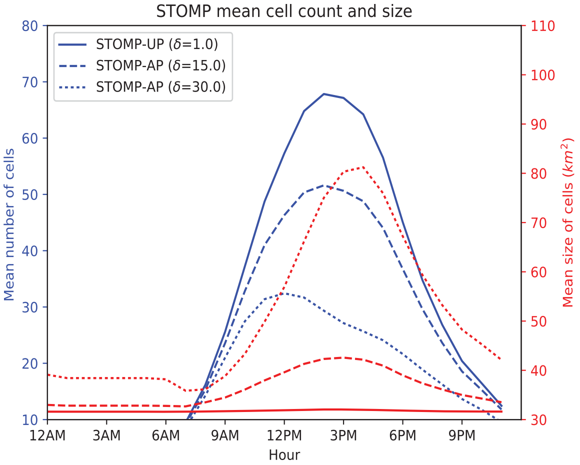

(a) STOMP-UP
(b) CPM simulation
(c) C-Pol Radar observation
Mean number of cells
Mean size of cells (km^2)
Hour
12AM
3AM
6AM
9AM
12PM
3PM
6PM
9PM
0
10
20
30
40
50
60
70
80
90

Figure 6.



STOMP mean cell count and size
STOMP-UP (δ=1.0)
STOMP-AP (δ=15.0)
STOMP-AP (δ=30.0)
Mean number of cells
Mean size of cells (km^2)
Hour
12AM
3AM
6AM
9AM
12PM
3PM
6PM
9PM
10
20
30
40
50
60
70
80
30
40
50
60
70
80
90
100
110

Figure 7.

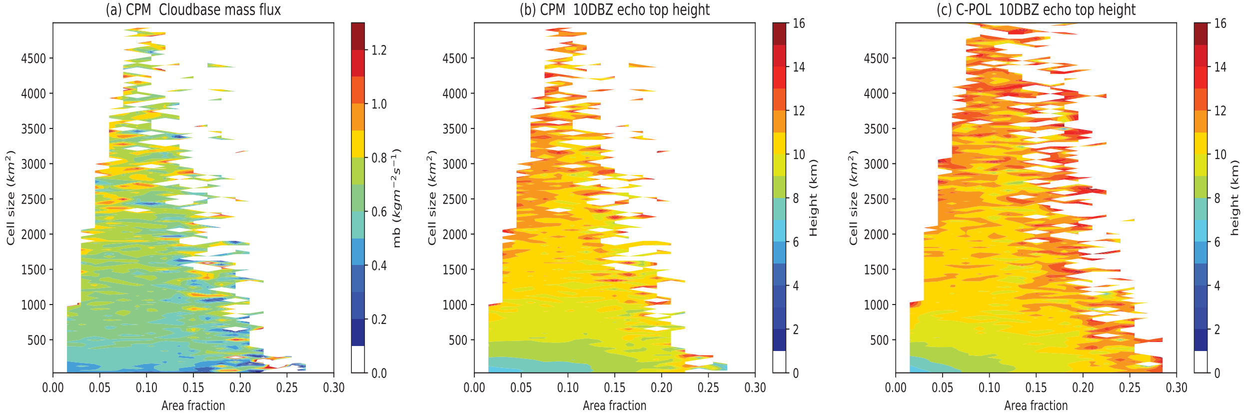

(a) CPM Cloudbase mass flux
(b) CPM 10DBZ echo top height
(c) C-POL 10DBZ echo top height
Cell size (km^2)
Area fraction
mb ($kgm^{-2}s^{-1}$)
Height (km)
height (km)
0.00 0.05 0.10 0.15 0.20 0.25 0.30
500 1000 1500 2000 2500 3000 3500 4000 4500
0.0 0.2 0.4 0.6 0.8 1.0 1.2
0 2 4 6 8 10 12 14 16

Figure 8.

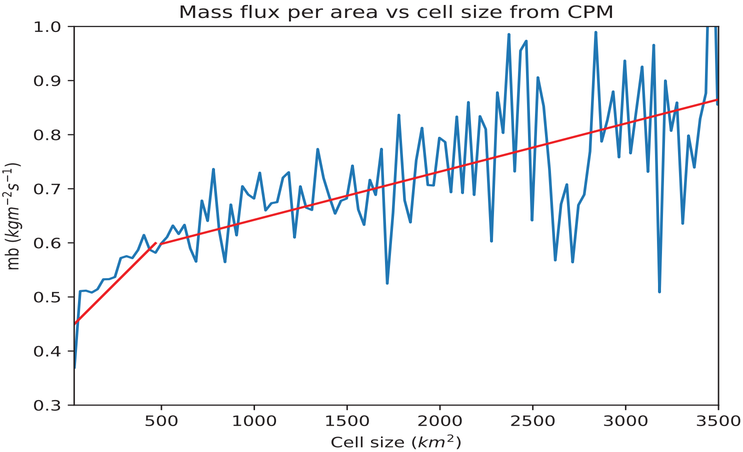

Mass flux per area vs cell size from CPM
mb ($kgm^{-2}s^{-1}$)
Cell size (km^2)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
500
1000
1500
2000
2500
3000
3500

Figure 9.

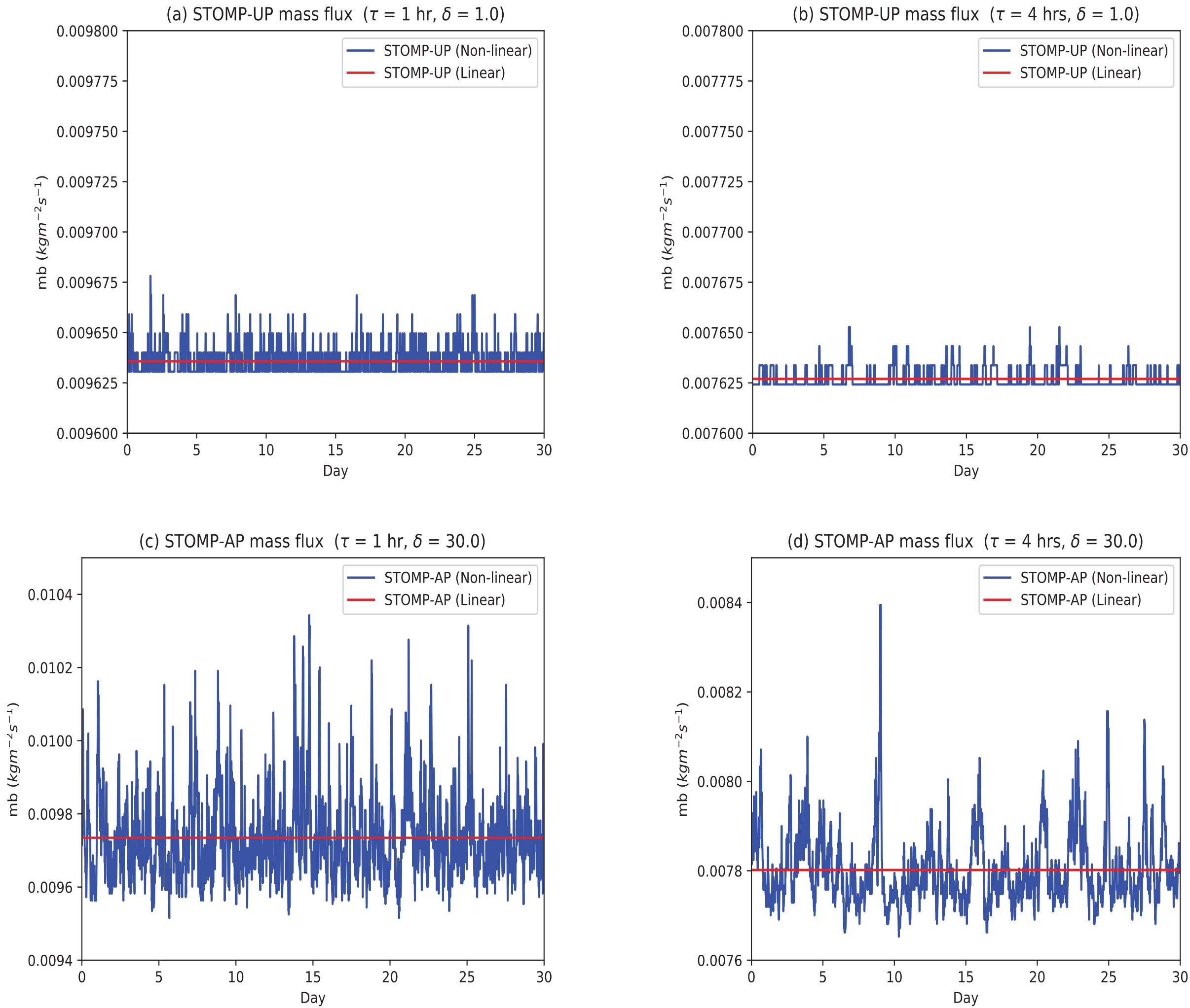
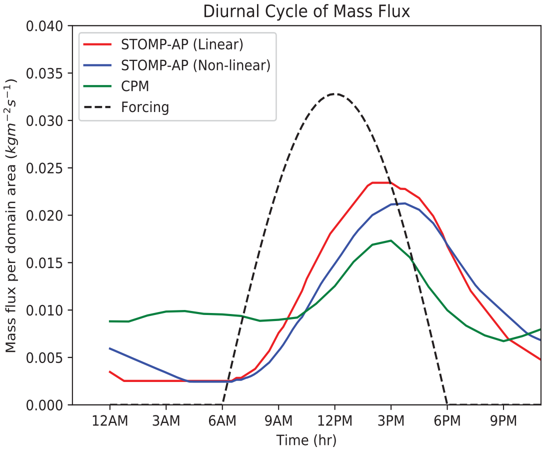

(a) STOMP-UP mass flux (τ = 1 hr, δ = 1.0)
STOMP-UP (Non-linear)
STOMP-UP (Linear)
mb (kgm⁻²s⁻¹)
Day
0.009800
0.009775
0.009750
0.009725
0.009700
0.009675
0.009650
0.009625
0.009600
0
5
10
15
20
25
30
(b) STOMP-UP mass flux (τ = 4 hrs, δ = 1.0)
STOMP-UP (Non-linear)
STOMP-UP (Linear)
mb (kgm⁻²s⁻¹)
Day
0.007800
0.007775
0.007750
0.007725
0.007700
0.007675
0.007650
0.007625
0.007600
(c) STOMP-AP mass flux (τ = 1 hr, δ = 30.0)
STOMP-AP (Non-linear)
STOMP-AP (Linear)
mb (kgm⁻²s⁻¹)
Day
0.0104
0.0102
0.0100
0.0098
0.0096
0.0094
(d) STOMP-AP mass flux (τ = 4 hrs, δ = 30.0)
STOMP-AP (Non-linear)
STOMP-AP (Linear)
mb (kgm⁻²s⁻¹)
Day
0.0084
0.0082
0.0080
0.0078
0.0076

Figure 10.



Diurnal Cycle of Mass Flux
STOMP-AP (Linear)
STOMP-AP (Non-linear)
CPM
Forcing
Mass flux per domain area ($kgm^{-2}s^{-1}$)
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
12AM
3AM
6AM
9AM
12PM
3PM
6PM
9PM
Time (hr)

Figure 11.

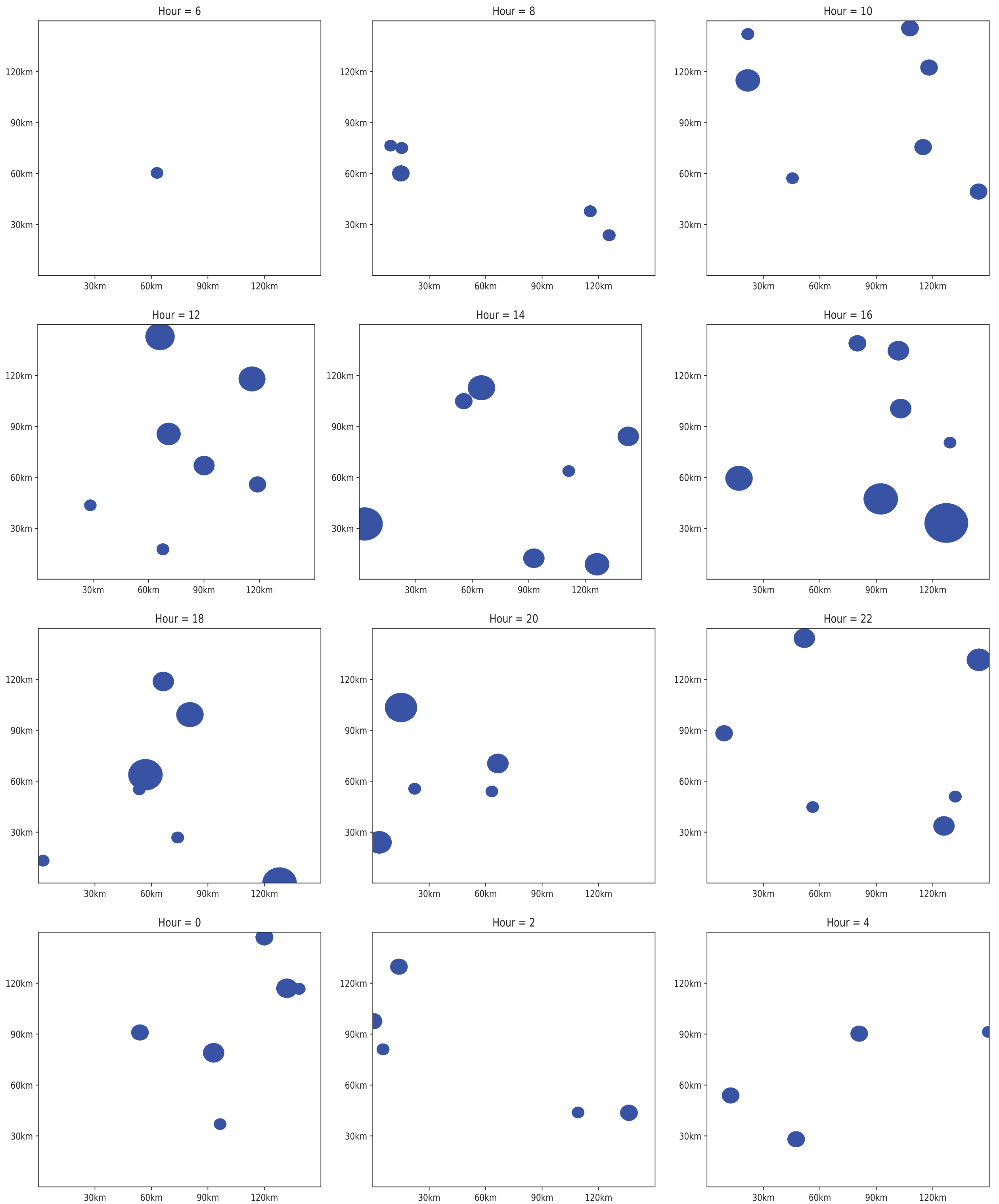
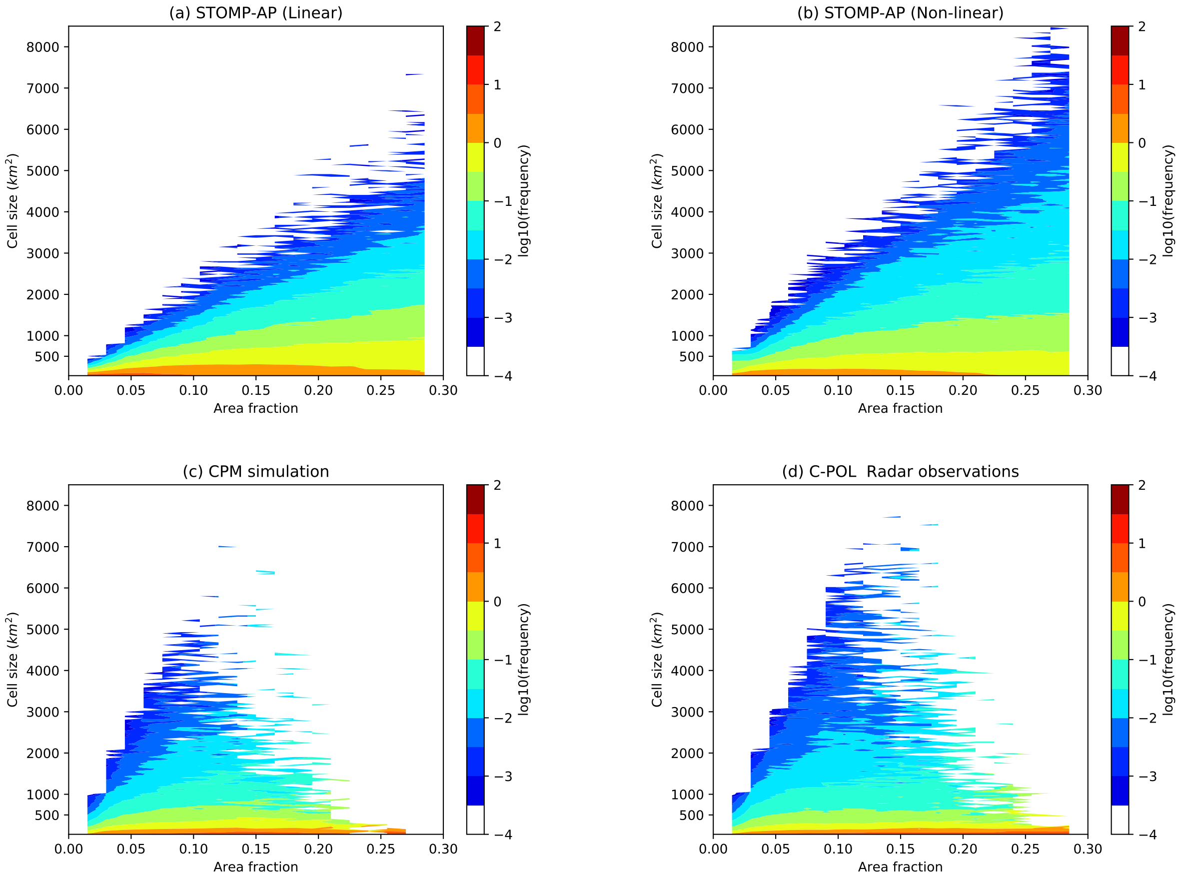

Hour = 6
Hour = 8
Hour = 10
Hour = 12
Hour = 14
Hour = 16
Hour = 18
Hour = 20
Hour = 22
Hour = 0
Hour = 2
Hour = 4
30km
60km
90km
120km

Figure 12.


(a) STOMP-AP (Linear)
(b) STOMP-AP (Non-linear)
(c) CPM simulation
(d) C-POL Radar observations
Cell size (km^2)
Area fraction
log10(frequency)
8000
7000
6000
5000
4000
3000
2000
1000
500
0.00
0.05
0.10
0.15
0.20
0.25
0.30
2
1
0
−1
−2
−3
−4