




## RESEARCH ARTICLE

10.1029/2022MS003391

†Deceased 12 March 2021

# A Shallow-Deep Unified Stochastic Mass Flux Cumulus Parameterization in the Single Column Community Climate Model

**B. Khouider<sup>1</sup>** , **B. B. Goswami<sup>2,3</sup>**, **R. Phani<sup>4</sup>**, and **A. J. Majda<sup>2,5,†</sup>**
<sup>1</sup>Department of Mathematics and Statistics, University of Victoria, Victoria, BC, Canada, <sup>2</sup>Center for Prototype Climate Models, New York University Abu Dhabi, Abu Dhabi, UAE, <sup>3</sup>Now at Institute of Science and Technology, Austria (ISTA), Klosterneuburg, Austria, <sup>4</sup>Indian Institute of Tropical Meteorology, Pune, India, <sup>5</sup>Department of Mathematics, Center for Atmosphere and Ocean Sciences, Courant Institute for Mathematical Sciences, New York University, New York, NY, USA
**Key Points:**

- The Zhang-McFarlane cumulus parameterization is modified by using randomly detraining plumes according to a stochastic multicloud model (SMCM)
- The SMCM tracks the evolution throughout the convection life cycle of the area fractions associated with multiple cloud types that characterize organized convection
- The SMCM framework allows for a natural coupling of shallow and deep convection and reproduces the tri-modal convection structure

**Correspondence to:**
 B. Khouider,  
[khouider@uvic.ca](mailto:khouider@uvic.ca)
**Citation:**
 Khouider, B., Goswami, B. B., Phani, R., & Majda, A. J. (2023). A shallow-deep unified stochastic mass flux cumulus parameterization in the single column Community Climate Model. *Journal of Advances in Modeling Earth Systems*, 15, e2022MS003391. <https://doi.org/10.1029/2022MS003391>

 Received 15 SEP 2022  
 Accepted 27 OCT 2023

**Abstract** Cumulus parameterization (CP) in state-of-the-art global climate models is based on the quasi-equilibrium assumption (QEA), which views convection as the action of an ensemble of cumulus clouds, in a state of equilibrium with respect to a slowly varying atmospheric state. This view is not compatible with the organization and dynamical interactions across multiple scales of cloud systems in the tropics and progress in this research area was slow over decades despite the widely recognized major shortcomings. Novel ideas on how to represent key physical processes of moist convection-large-scale interaction to overcome the QEA have surged recently. The stochastic multicloud model (SMCM) CP in particular mimics the dynamical interactions of multiple cloud types that characterize organized tropical convection. Here, the SMCM is used to modify the Zhang-McFarlane (ZM) CP by changing the way in which the bulk mass flux and bulk entrainment and detraining rates are calculated. This is done by introducing a stochastic ensemble of plumes characterized by randomly varying detraining level distributions based on the cloud area fraction of the SMCM. The SMCM is here extended to include shallow cumulus clouds resulting in a unified shallow-deep CP. The new stochastic multicloud plume CP is validated against the control ZM scheme in the context of the single column Community Climate Model of the National Center for Atmospheric Research using data from both tropical ocean and midlatitude land convection. Some key features of the SMCM CP such as its capability to represent the tri-modal nature of organized convection are emphasized.

**Plain Language Summary** Current climate models perform poorly in the way they represent convection and cloud systems in the tropics. A stochastic multicloud model (SMCM), based on an interacting particles lattice model that tracks the random variations of the cloud area fractions (CAF) of the main cloud types that characterize organized tropical convection, is used here to modify/stochasticize a state-of-the-art cumulus parametrization. The main idea is to use the CAF of various cloud types, as predicted by the SMCM to decide on the distribution of the vertical penetration of convecting air parcels that form the clouds. Accordingly the distribution of the associated cloud top heights changes randomly based on the large scale environment depending on whether the later favors shallow, mid-level, or deep convection. The new parametrization is tested in the NCAR Single Column Community Climate Model and compared with observations when available. The main features of the new parameterization are discussed here. Based on the results, the stochastic parameterization possess various features that are desirable in order to improve the representation of organized convection in 3D climate models. This is in agreement with the success of the SMCM when used in different climate models, although based on much simplistic approaches.

© 2023 The Authors. Journal of Advances in Modeling Earth Systems published by Wiley Periodicals LLC on behalf of American Geophysical Union. This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

## 1. Introduction

The state-of-the-art Earth System Models (ESMs) have limitations in their projections of changes in spatial and temporal patterns of precipitation and the frequency of weather-extremes in a warmer world (Dai, 2006; IPCC, 2021; Meehl et al., 2000). There are large uncertainties in the way ESMs deal with clouds and precipitation in the tropics (Bony & Dufresne, 2005; Bony et al., 2015; Brient & Bony, 2013; Ceppi et al., 2017). Climate models have significant biases in rainfall intensity such as overestimations and underestimations of the mean precipitation in many key tropical and polar regions (Flato et al., 2013; Sabeerali et al., 2013) leading to uncertainties in the greenhouse effect by water vapor and clouds. These uncertainties are in large part due to modeling

errors (Palmer, 2001). Clouds cover on average two-thirds of the planet. Depending on their optical depth and cloud top pressure, clouds affect Earth's radiative energy budget differently. While high clouds mainly absorb long waves and re-emit them back to the surface and contribute to global warming, low clouds, that are abundant over extended regions of the globe, have mostly a cooling effect by reflecting shortwaves back to space and contribute to Earth's albedo. Unfortunately recent studies suggest that these (favorable) low-level clouds that help keep our planet relatively cool, may become significantly reduced in a warmer climate (Flato et al., 2013) and may even disappear all together in the most extreme warming scenarios (Schneider et al., 2019). This is in fact a positive climate feedback phenomenon that would exacerbate global warming and may contribute to the uncertainty of future climate projections. Overall, the representation of moist convection and clouds is one of the most significant modeling errors in climate science.

The need for improved sub-grid models of moist convection, a.k.a, cumulus parameterizations (CP), in coarse resolution global climate models (GCMs) has been an active research area since the 60s (Arakawa, 2004; Stensrud, 2007). CPs in most state-of-the-art GCMs are based on the quasi-equilibrium assumption (QEA) that views convection as the action of an ensemble of cumulus clouds, in a state of equilibrium with respect to a slowly varying atmospheric large scale state (Arakawa & Schubert, 1974; Stensrud, 2007; G. J. Zhang & McFarlane, 1995). Despite significant improvements, GCMs (and ESMs) are still unable to represent adequately the multiscale spacial and temporal variability of precipitation and atmospheric dynamics in the tropics (Dai, 2006; Hung et al., 2013; Randall et al., 2003). This limitation has been associated with the incompatibility of the QEA with the observed organization and dynamical interactions (with the environment and with each other) of cloud systems in the tropics, over a large spectrum of scales (Khouider, 2019; Moncrieff & Klinker, 1997; Rio et al., 2019).

Convective systems range from the convective cells of 1–10 km and a few minutes, to mesoscale systems of 100–500 km and a few hours, to synoptic scale convectively coupled waves of 1,000–5,000 km and a few days, to intra-seasonal planetary scales disturbances such as the Madden-Julian oscillation (MJO), monsoons, and the inter-tropical convergence zones (Houze, 2004; Khouider et al., 2013; Kiladis et al., 2009; Murakami et al., 1986; Nakazawa, 1988; C. Zhang, 2005, 2013). The “CP deadlock” was first broken by the cloud resolving convective parametrization (CRCP) concept, which uses a 2d cloud resolving model in place of a CP (Grabowski, 2016; Randall et al., 2003). However, the insurmountable CRM's computational overhead made the CRCP method impractical. But it led the GCM community to look beyond the QEA theory as the community came to realize that the unresolved degrees of freedom associated with unresolved processes are very important (Palmer, 2001, 2022).

The last two decades have seen a surge of novel ideas to overcome the QEA which led to new breakthroughs in CP (Rio et al., 2019). While some models are designed to capture specific physical processes of organized convection, in a deterministic fashion (e.g., Khouider et al., 2011; Moncrieff et al., 2017; Pan & Randall, 1998; Tompkins, 2002), many probabilistic models that mimic the stochastic nature of convection have also been proposed and used (Bengtsson & Karnich, 2016; Buizza et al., 1999; D'Andrea et al., 2014; Dorrestijn et al., 2013; Hagos et al., 2018; Khouider et al., 2010, 2003; J. Lin & Neelin, 2000, 2003; Majda & Khouider, 2002; Plant & Craig, 2008; Suselj et al., 2019a, 2019b; Y. Wang et al., 2016; Y. Wang & Zhang, 2016). See Berner et al. (2016) for a comprehensive review. Khouider et al. (2010) in particular used a stochastic multicloud model (SMCM) CP, based on a Markov jump lattice system, to represent the subgrid variability associated with the multiple cloud types that characterize organized tropical convection (Johnson et al., 1999).

The last few years have seen an increase in data driven-machine learning strategies to improve CPs and GCMs (Brenowitz & Bretherton, 2018; Cardoso-Bihlo et al., 2019; Dorrestijn et al., 2015; Gentine et al., 2018; Gottwald et al., 2016; Schneider et al., 2017). The ongoing search for the most reliable CPs will allow further breakthroughs in ESMs and in our basic understanding of organized moist convection (Rio et al., 2019).

The implementation and testing of the SMCM in both idealized (Ajayamohan et al., 2016; Deng et al., 2015; Frenkel et al., 2012; K. Peters et al., 2013) and state-of-the-art, coupled atmosphere-ocean, GCMs (Goswami et al., 2017a, 2017b, 2017c; Khouider, 2019; K. Peters et al., 2017) demonstrated the feasibility of a realistic representation of the temporal and spacial variability of tropical precipitation and the associated atmospheric multi-scale dynamical modes, such as the MJO, convectively coupled waves, and monsoon systems (Khouider, 2019). Goswami et al. (2017c) in particular implemented the SMCM in the Climate Forecasting System version 2 of the U. S. National Centers for Environmental Predictions in lieu of the preexisting relaxed Arakawa-Schubert scheme

(Saha et al., 2014) while K. Peters et al. (2017) used the deep convective area fraction from the SMCM to improve the triggering of deep convection and the mass flux closure in the ECHAM 6.5 GCM.

Apart from K. Peters et al. (2017), the implementations of the SMCM so far relied on imposed vertical heating structure functions—one for each one of the congestus, deep, and stratiform cloud types—that together mimic the observed temperature and humidity tendency profiles (the equivalents of Q1 and Q2 of Yanai et al. (1973); see Section 4.2) and their dynamical evolution throughout the convection life-cycle (Johnson et al., 1999; Mapes et al., 2006). However, these vertical structure functions and other model parameters depend heavily on the observed current climatology and are mostly-only suited for the tropics (Casey et al., 2007; Goswami et al., 2017c; J.-L. Lin et al., 2004; Schumacher et al., 2007; Stachnik et al., 2013). Because of these structural limitations of the heating profiles on which these earlier versions of the SMCM CP rely, the resulting models are not suitable for climate change projections. The goal here is to instead use the SMCM as a building block to propose a stochastic plume model within the framework of a state-of-the-art mass flux CP. This also constitutes a more significant effort compared to what has been done in K. Peters et al. (2017) and it results in a unified shallow-deep, scale aware, and stochastic multi-cloud plume mass-flux CP. Here, the new parameterization is introduced and tested in the context of the single-column Community Climate Model (SCAM) (Gettelman et al., 2019; Neale et al., 2010; Randall et al., 1996; M. Zhang et al., 2016).

The paper is organized as follows. The new unified shallow-deep stochastic multi-cloud plume mass-flux CP is described in Section 2. Section 3 summarizes the observational test cases that are used to test and validate the new CP. The validation results are presented in Section 4 and Section 5 provides a summary discussion and conclusions.

## 2. A Unified Stochastic Multi-Cloud Plume-Based Mass-Flux Scheme

The new unified shallow-deep stochastic multi-cloud plume cumulus parameterization (SMPCP), presented here, is based on the Zhang-McFarlane (G. J. Zhang & McFarlane, 1995, hereafter: ZM95) cumulus parametrization. Loosely speaking the ZM scheme is modified in the way the plume ensemble is used to produce a bulk mass flux and the associated bulk entrainment and detrainment rates. Instead of assuming that convective plumes can detrain at all levels, beyond the level of minimum moist static energy (MSE), as done in ZM, we use a scheme where the plume ensemble is formed by plumes whose detrainment levels are determined in a probabilistic fashion based on the SMCM of Khouider et al. (2010), with the stochasticity introducing subgrid variability to break away from the QEA.

To unify the treatment of shallow and deep convection, the original SMCM is extended to four cloud types by including shallow cumulus cloud types in addition to the readily existing congestus, deep, and stratiform cloud types. The mathematical framework for the four cloud SMCM remains the same, for both the case without local interactions and the case with local interactions introduced in Khouider (2014) and Khouider et al. (2010), respectively. The four cloud types have a relatively long time memory, as they define the convective systems life cycle, and interact with each other across both time and space. In essence, this is where the QEA, which attempts to keep convective activity in equilibrium with respect to the slowly varying large scale dynamics, fails. Instead, the SMCM introduces variability at the subgrid scale in the CP in both time and space as it mimics this multiple scale convective organization.

In a nutshell a mass flux scheme for CP, whether it is based on a spectral plume model (e.g., Arakawa & Schubert, 1974, ZM95) or it uses a bulk mass flux model (Tiedke, 1984; Tiedtke, 1993), is a mathematical framework that approximates the updraft and downdraft mass fluxes associated with the vertical displacement of air parcels between the mixed layer and the free troposphere, and the associated dynamical and thermodynamical variables. According to QEA, a steady state model is assumed and conveniently used for the updraft and downdraft dynamics. In this context, the conservation of upward mass flux,  $M(z)$ , for example, takes the form

$$\frac{\partial M}{\partial z} = E(z) - D(z). \quad (1)$$

Here,  $E(z)$  and  $D(z)$  are, respectively, the entrainment and detrainment rates of environmental air into the cloud and of cloud air into the environment. There are two big challenging issues associated with Equation 1. One is associated with defining  $E(z)$  and  $D(z)$  and the other is related to setting a proper boundary condition at cloud

base so that the  $M$ -equation can be solved from bottom to top, and it need to be solved until the top of the tallest cumulus cloud, which also needs to be predicted. The later, that is, finding the cloud top, is perhaps the hardest problem because it depends on the many unknown parameters that control turbulent mixing of the rising plumes and their local buoyancy acceleration, to name a few. Nonetheless, this boils down to finding a model for the entrainment and detrainment rates  $E(z)$  and  $D(z)$ , respectively. This is in fact a free boundary problem as the cloud top height and the  $E(z)$ - $D(z)$  couplet are intimately related and closure assumptions are needed at many levels.

The solution to both issues needs to be both physically sound and computationally efficient. Here, we are mainly concerned with the issue of computing the bulk mass flux profile and the associated bulk entrainment and detrainment rates using a stochastic model; for the cloud base mass flux issue, below, we use a combined convective available potential energy (CAPE) and turbulent kinetic energy closures to treat simultaneously deep and shallow convection, respectively. An original idea on how to use the SMCM framework to relax the quasi-equilibrium closure, for the cloud base mass flux, is proposed in Khouider and Leclerc (2019) but this new closure is not employed here to focus on the effect of using the SMCM to change the entrainment and detrainment rates. Nonetheless, in a nutshell, the relaxed quasi-equilibrium closure using the SMCM framework amounts to proposing and using evolution equations (that can be either deterministic-mean field limit or stochastic) for the cloud area fractions (CAF) of multiple cloud types, to obtain closed prognostic equations for the cloud work function and cloud base mass flux of Pan and Randall (1998). The inclusion of both the stochastic closure based on the SMCM and the SMCCP, in the same parametrization scheme, will be the subject of future research.

### 2.1. The Stochastic Multi-Cloud Plume Model

We assume a steady state plume model as in ZM95. Let

$$m_j(z) = M_b^j e^{\lambda_j(z-z_b)}, \quad z_b \leq z \leq z_D^j \quad (2)$$

be the total mass flux associated with the sub-ensemble of plumes that all detrain at the same level  $z_D^j$  and have the same fractional entrainment rate  $\lambda_j$ ,  $j = 0, 1, 2, \dots$ . Here  $z_b$  is the cloud base height and  $M_b^j$  is the associated mass flux at cloud base, which, for the time being, we assume is given. Also,  $z$  is the height coordinate variable, with  $z = 0$  at the surface, and  $j$  is the index of the convecting plume sub-ensemble. As we will see later, the calculation of  $M_b^j$  follows the QEA-based CAPE closure of ZM95 for deep convection and the turbulent kinetic energy closure of Bretherton and Park (2008) for shallow convection, consistent with the control SCAM model to which the new parameterization is compared to (see Section 2.7).

We assume that there is a one-to-one correspondence between the fractional entrainment rates  $\lambda_j$  and the detrainment levels  $z_j$ :

$$z \rightarrow \lambda_z,$$

and that the two variables can be interchangeably used to index the plume sub-ensemble (ZM95; Arakawa & Schubert, 1974).

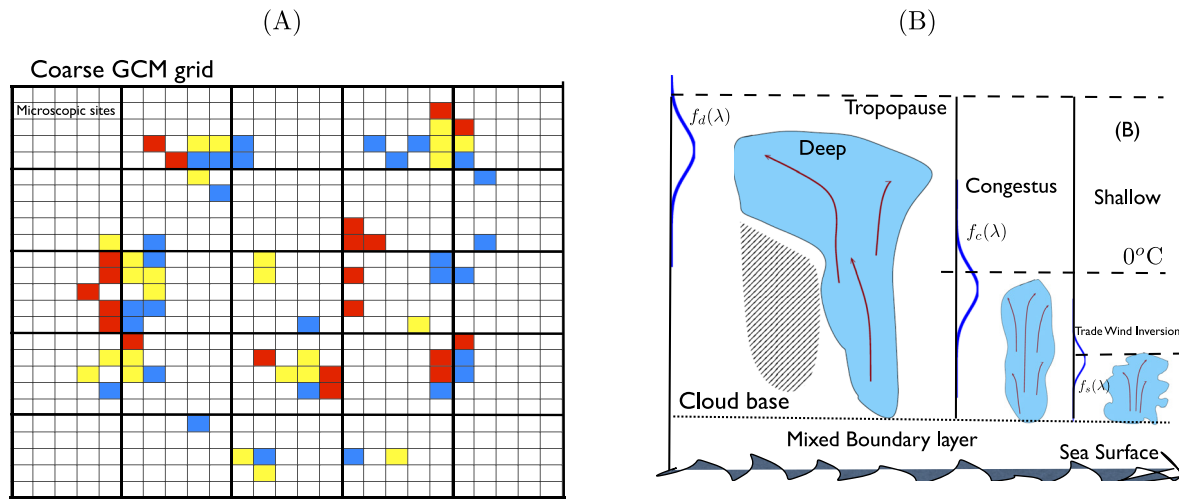
Consider the total mass flux,  $M(z)$ , of the grand ensemble of plumes that detrain at level  $z$  or higher or equivalently that have an entrainment rate  $\lambda_j \leq \lambda_z$ . For a fixed level  $z$ , we have

$$M(z) = \sum_{j, \lambda_j \leq \lambda_z} m_j(z) = \sum_{j, \lambda_j \leq \lambda_z} M_b^j e^{\lambda_j(z-z_b)}. \quad (3)$$

The summation is over all plumes  $j$  whose entrainment rate  $\lambda_j \leq \lambda_z$  or equivalently the detrainment level  $z_j \geq z$ . Let  $N_z$  be the number of detrainment levels  $z_j$  greater or equal to  $z$  or equivalently the number of entrainment rates  $\lambda_j$  such that  $\lambda_j \leq \lambda_z$ . We assume that  $N_z$  is a random variable with a probability distribution to be specified. For large enough  $N_z$ , we have approximately

$$M(z) \approx \frac{N_z}{\lambda_z} \int_0^{\lambda_z} M_b(\lambda) e^{\lambda(z-z_b)} d\lambda. \quad (4)$$

Mathematically speaking, Equation 4 is only loosely related to Equation 3 as explained above. One has to view Equation 3 as a mere physical motivation for Equation 4, which, from now on, becomes our main definition for the bulk updraft mass flux.



**Figure 1.** (a) The multicloud model lattice. The global climate models (GCM) grid box is divided into a certain number of microscopic site and each site is either clear sky or occupied a cloud of certain type, shallow, congestus, deep as illustrated by the different colors. (b) Sketch of the plume detrainment level distribution functions and the associated cloud types. The detrainment probability distribution functions are centered near and below the empirical detrainment levels (dashed lines) formed by the trade wind inversion, the 0°C or freezing/melting level, and the tropopause, which serve to cap shallow cumulus, cumulus congestus, and deep cumulonimbus clouds.

To determine the distribution of  $N_z$ , we use a hybrid method that combines a counting process in the vertical and a lattice model in the horizontal. The horizontal lattice model is based on the SMCM while the counting process is reminiscent to the equilibrium statistics model of Craig and Cohen (2006).

To start, we consider a square lattice overlaid over the averaging horizontal area of size  $A_L$  with a size  $N_L \times N_L$  lattice sites. Let  $z_k$  be a fixed vertical level on the GCM grid. We denote by  $P_{n,m,k}$  the number of plumes going through the lattice site  $n, m$ ,  $1 \leq n, m \leq N_L$ , that detrain at an arbitrary level  $z \geq z_k$ .

We divide the plumes into groups belonging to three cloud types or categories: shallow cumulus, cumulus congestus, and deep convection according to their detrainment levels. The shallow cumulus group or type includes all plumes that detrain near the level of free convection (LFC), that is, having a cloud-top height below or near say 2 km, and the cumulus congestus type comprises all plumes that detrain near the freezing level (FRZ) while the deep convection type regroups all plumes detraining far above FRZ, typically near the level of neutral buoyancy of non mixing parcels. We assume that each one of these cloud types has its own probability distribution function (PDF) of detrainment levels. The PDFs are accordingly centered near three specified levels and they are intentionally allowed to overlap, consistent with observations of cloud top echo heights (See e.g., Figure 1 of Zermelo-Daz et al. (2015)). Our approach is illustrated in Figure 1. Figure 1b, in particular, illustrates the three cloud types of interest and the corresponding detrainment level PDFs, each centered at a chosen level  $z_x$  or equivalently a fractional entrainment rate  $\lambda_x$ .

We introduce an order parameter  $\sigma_{n,m}$  that takes values from 0, 1, 2, 3, and 4, at each site or cell  $(n, m)$  of the horizontal lattice, according to whether the given site is respectively a clear sky or it is occupied by one of the cloud types: shallow cumulus, cumulus congestus, deep, or stratiform. See Figure 1a. This is an extension of the original SMCM (Khouider, 2014; Khouider et al., 2010), which had only congestus, deep, and stratiform cloud types.

Let  $P_{n,m,k}^\sigma$  be the number of plumes with a detrainment level  $z_D \geq z_k$ , originating over the lattice site  $(n, m)$ , conditional on the realization of  $\sigma_{n,m}$ . Clearly when  $\sigma_{n,m} = 0$ , we have  $P_{n,m,k}^\sigma = 0$ . Let  $d_\sigma$  be the detrainment level of a random plume of type  $\sigma$ ,  $\sigma = 1, 2, 3$ . Stratiform clouds are not considered part of the convecting plumes although they contribute in the SMCM dynamics, to account for the interaction of cumulus clouds with stratiform cloud decks which have been recognized, a long time ago, to be of paramount importance for organized convection (Majda & Shefter, 2001; Mapes, 2000; Pan & Randall, 1998). There is a possibility for taking advantage of the availability of this additional-stratiform order parameter value ( $\sigma = 4$ ) to systematically include stratiform anvils within the SMCCP framework. This however is postponed to the near future so we can focus on the modification and stochasticization of the vertical distribution of the convective mass flux and the associated entrainment and detrainment rates. Let  $\mathcal{P}_\sigma$  be the probability distribution of  $d_\sigma$  and  $f_\sigma$  be the associated density functions for  $\sigma = 1, 2, 3$ .



Let us assume that  $\sigma_{nm} = \sigma$  was realized, that is, that we know that clouds of type shallow, congestus, or deep do occur at lattice site  $(n, m)$ . Then, for  $\sigma = 1, 2, 3$ , the conditional probability of the plume number is given by,

$$\begin{aligned} \text{Prob}\{P_{n,m,k}^\sigma = j | \sigma_{n,m} = \sigma\} &\equiv \text{Prob}\{j \text{ plumes of type } \sigma \text{ detrain at } z_D \geq z_k, \text{ over the site } (n, m)\} \\ &= \text{Prob}\{j \text{ plumes of type } \sigma \text{ detrain at } z_D \geq z_k, \text{ over the site } (n, m) | d_\sigma \geq z_k\} \times \text{Prob}\{d_\sigma \geq z_k\}. \end{aligned} \quad (5)$$

Using statistical mechanics arguments, Craig and Cohen (2006), demonstrated, in a cloud resolving simulation, that the number of convecting plumes, in an ensemble that contribute to a given mass flux amount, is Poisson distributed. Inspired by this result, we assume that the number of plumes of type  $\sigma$ , for  $\sigma = 1, 2, 3$ , occurring over the site  $(n, m)$  and detraining at a level  $z \geq z_k$  given that  $d_\sigma \geq z_k$  is Poisson, with rate  $\Lambda_{\{\sigma,n,m,k\}}$ . For the sake of simplicity we assume that  $\Lambda$  depends only on the size of the lattice site and not on the specific location of the site. In particular, this dependency of the maximum size of the plume ensemble on the GCM grid size makes the new CP scale aware, even when the number of lattice sites is fixed, and at the same time determines the assumed size of the microscopic (cloud) cells. Moreover, given the conditional nature of this distribution, it is meaningful to assume that  $\Lambda$  is independent of the detraining level  $z_k$ , which is already accounted for in Equation 5 through  $P_\sigma$ . Accordingly, we set

$$\Lambda = \alpha \left( \frac{\Delta X}{LN_L} \right)^\beta, \quad (6)$$

where  $\Delta X$  is the horizontal extend of the averaging area of the ensemble (in practice the GCM grid size),  $L$  is a reference length scale and  $N_L$  is the number of lattice points while  $\alpha$  and  $\beta$  are scaling parameters. While the values of  $N_L, \alpha, \beta$  can be obtained from data (e.g., cloud resolving simulations), here, for the purpose of the single column validation, we assume that  $\Lambda$  takes the fixed value of  $\Lambda = 1$  (one plume per lattice site). Variations of  $\Lambda$  with the lattice size can be re-incorporated in the future to make the SMPCP resolution aware, especially in the context of full 3d simulations.

## 2.2. The Cloud Type Lattice Model

The order parameter  $\sigma^{n,m}$  defines a continuous time Markov process on the state space  $\Sigma = \{0, 1, 2, 3, 4\}^{N_L \times N_L}$  denoted here by  $\sigma_l$ . Each possible realization  $X \in \Sigma$  is called a configuration. See Figure 1a.

We introduce the area fractions

$$\bar{\sigma}_l = \frac{1}{N_L^2} \sum_{n,m} \mathbb{1}_{\{\sigma^{n,m}=l\}} = \frac{N_l}{N_L^2}, \quad l = 1, 2, 3, 4, \quad (7)$$

of shallow cumulus, cumulus congestus, deep convective clouds, and stratiform clouds, respectively. The corresponding numbers of sites that are occupied by shallow cumulus, cumulus congestus, and deep convecting (cumulonimbus) clouds are denoted by  $N_l = N_L^2 \bar{\sigma}_l, l = 1, \dots, 4$ , respectively. In Equation 7,  $\mathbb{1}_{\{\sigma=l\}}$  is the indicator function that takes the value 1 if  $\sigma = l$  and 0 otherwise.

As time progresses, transitions from one configuration to the next occur according to transition probabilities defined through an infinitesimal generator  $\mathcal{L}$ , a matrix of dimension the cardinality of  $\Sigma$ . However, for the sake of simplicity, we assume that the transition rates depend only on the environmental state (i.e., the GCM variables) and ignore local interactions between lattice sites. Thus, the  $\sigma_{n,m}$ 's, defined at the different sites, are independent identically distributed random variables and the infinitesimal generator is fully described through the transition rates at a single site. As a result, the area fractions in Equation 7 define a coarse-grained stochastic process that can be regarded as a multidimensional stochastic birth-death process with immigration, for the cloud type populations and its coarse grained transition rates are easily obtained from the microscopic transition rates, in closed form and without any approximation. This in particular allows a seamless simulations of the  $\bar{\sigma}_l$ 's as a stand alone coarse grained Markov process with an insignificant computational overhead (Khouider et al., 2010). Nonetheless, a similar coarse graining procedure could be applied to the case with local interactions albeit an extra assumption of strong mixing (Khouider, 2014).

We denote by  $R_{lk}$  the transition rate from state  $l$  to state  $k, 0 \leq l, k \leq 4$ . On physical grounds some transitions are forbidden on the infinitesimal time and the associated transition rates are thus set to zero. Namely, infinitesimal

transitions (or sudden jumps) from deep to congestus, from congestus to shallow, from deep to shallow, and from stratiform to any other cloudy state are not allowed. We have

$$R_{21} = R_{32} = R_{31} = R_{04} = R_{14} = R_{24} = R_{41} = R_{42} = R_{43} = 0. \quad (8)$$

For simplicity, the death rates, corresponding to transitions for a given cloud type to clear sky, that is, from state 1, 2, 3, or 4 to state 0 are assumed independent of the large scale state. We set

$$R_{l0} = \frac{1}{\tau_{l0}}, l = 1, 2, 3, 4. \quad (9)$$

The remaining–non death transition rates (namely for births and immigration, in the jargon of population dynamics) are set to depend on specific large-scale indicators or predictors that determine whether the underlying environment is favorable to the formation of clouds of certain types or to clear sky. Here, we make the tacit assumptions that shallow cumulus is favored in regions of high convective inhibition (CIN), strong trade wind inversion, or strong subsidence, congestus clouds are favored when there is high CAPE but the mid troposphere is dry while deep convection is favored when there is both high CAPE and the mid-troposphere is moist. Deep convective clouds are assumed to transition to stratiform clouds according to a prescribed transition rate independent of the large scale state (Majda & Shefter, 2001; Mapes, 2000).

Following Khouider et al. (2010), we assume that the transition rates take the form of Arrhenius-like functions and set

$$\begin{aligned} R_{01} &= \frac{1}{\tau_{01}} F(C_N, W_N) \\ R_{02} &= \frac{1}{\tau_{02}} \Gamma(C_L) * \Gamma(D) * [1 - F(C_N, W_N)] \\ R_{03} &= \frac{1}{\tau_{03}} \Gamma(C) * (1 - \Gamma(D)) * [1 - F(C_N, W_N)] \\ R_{12} &= \frac{1}{\tau_{12}} \Gamma(C_L) * \Gamma(D) * [1 - F(C_N, W_N)] \\ R_{13} &= \frac{1}{\tau_{13}} \Gamma(C) * (1 - \Gamma(D)) * [1 - F(C_N, W_N)] \\ R_{23} &= \frac{1}{\tau_{23}} \Gamma(C) * (1 - \Gamma(D)) * [1 - F(C_N, W_N)] \\ R_{34} &= \frac{1}{\tau_{34}}, \end{aligned} \quad (10)$$

where  $F(C_N, W_N) = 0.5[\Gamma(C_N) + \Gamma(W_N)]$  and  $\Gamma(X) = 1 - \exp(-X)$ , if  $X > 0$  while  $\Gamma(X) = 0$  if  $X \leq 0$ . Here,  $X$  is a generic variable spanning the large scale predictors CAPE,  $C = \text{CAPE}/\text{CAPE}_0$ , low level CAPE,  $C_L = \text{LCAPE}/\text{LCAPE}_0$ , dryness,  $D = (\mathcal{H} - \mathcal{H}_0)/\text{MTD}_0$ , the large scale subsidence at the top of the planetary boundary layer,  $W_N = -\min(0, W/W_0)$ , and CIN,  $C_N = -\text{CIN}/\text{CIN}_0$ . Further,  $\mathcal{H}$  is the relative humidity at 700 hPa with  $\mathcal{H}_0 = 5\%$  so that deep convection is inhibited when the mid-troposphere is dry and LCAPE stands for low-level CAPE and is defined as the part of the CAPE integral between LFC and the freezing level.  $\text{CAPE}_0 = 1,000 \text{ J/Kg}$ ,  $\text{LCAPE}_0 = 100 \text{ J/Kg}$ ,  $\text{MTD}_0 = 10\%$ ,  $W_0 = 1 \text{ m/s}$  and  $\text{CIN}_0 = 50 \text{ J/Kg}$  are reference scales used for normalization. We note that here CIN is considered as a negative definite quantity, so that when CIN is large,  $\Gamma(C_N)$  approaches unity. Also, CAPE and LCAPE can be replaced by other like quantities that can be used to measure the convective potential such as vertical velocity as done in K. Peters et al. (2017) or even the cloud work function of Arakawa and Schubert (1974).

Noteworthy, the transition rates in Equation 10 are not used as such but they directly yield the coarse grained birth, death, and cloud type switching rates to seamlessly evolve the cloud area fraction in Equation 7. For example, the probability to have the congestus population  $N_2$  augmented by one individual, during a short time lapse  $\Delta t$ , is given by

$$\text{Prob}\{N_2(t + \Delta t) = N_2(t) + 1\} = [N_1(t)R_{12} + (N_L^2 - N_1(t) - N_3(t) - N_4(t))R_{02}]\Delta t + o(\Delta t).$$

Here the transition rates  $R_{kj}$  are fixed. The interested reader is referred to Khouider (2019) and Khouider et al. (2010) for more details.

**Table 1**

*Transition Time Scale Parameters (in Hours) Used in the Five Different Experiments Performed for Each One of the Six Study Cases*

Description	Parameter	Value			
		SMCM	SMCM1	SMCM2	SMCM3
		(DYNAMO-radar)	(GATE-LES)	(Ad hoc)	(Guess)
Clear sky to shallow cumulus	$\tau_{01}$	1	1	9	2
Clear sky to congestus	$\tau_{02}$	5.6	32	32	2
Clear sky to deep	$\tau_{03}$	0.1	12	12	0.75
Shallow cumulus to congestus	$\tau_{12}$	3	1	12	2
Shallow cumulus to deep	$\tau_{13}$	0.14	1	1	1
Congestus to deep	$\tau_{23}$	0.31	0.25	0.25	1.5
Deep to stratiform	$\tau_{34}$	2	0.25	0.25	1
Shallow cumulus to clear sky	$\tau_{10}$	5	2	20	5
Congestus to clear sky	$\tau_{20}$	7	2	2	7
Deep to clear sky	$\tau_{30}$	14.3	9.5	9.5	10
Stratiform to clear sky	$\tau_{40}$	30	1	1	20

*Note.* See text for details.

The parameters  $\tau_{lk}$ ,  $l, k = 0, \dots, 4$  constitute the set of the transition time scales that define the transition rates  $R_{lk}$ ,  $l, k = 0, \dots, 4$  together with their explicit dependencies on the large scale dynamics and thermodynamics—the predictors. In practice,  $\tau_{kl}$ 's are either determined from data using statistical inference as done in Cardoso-Bihlo et al. (2019), De La Chevrotiere et al. (2014), and De La Chevrotière et al. (2015) or prescribed in an ad hoc fashion based on physical intuition as done in Khouider et al. (2010). De La Chevrotière et al. (2015) inferred the seven transition timescales for the original three cloud-type SMCM using large eddy simulation (LES) for a tropical Atlantic field experiment test case (Khairoutdinov et al., 2009) while Cardoso-Bihlo et al. (2019) did the same using radar cloud data from the Indian Ocean field campaign known as DYNAMO (Feng et al., 2014). The later-DYNAMO parameters are augmented with ad hocly chosen values for the transitions from and to the shallow cumulus state (4 in total:  $\tau_{01}$ ,  $\tau_{12}$ ,  $\tau_{13}$ ,  $\tau_{10}$ ) and used here as a benchmark test case. The sensitivity to this parameter choice is tested by comparing to results obtained with three variants of the parameter set, one is an extension of the parameters obtained by De La Chevrotière et al. (2015) based on LES data while the other two are based on intuition. The first of the last two is a slight modification of the DYNAMO parameters while the other is completely an educated guess. The actual values are reported in Table 1.

It is worth noting here that the Bayesian inference framework, used to learn the transition time scales  $\tau_{k,l}$  (Cardoso-Bihlo et al., 2019; De La Chevrotière et al., 2015) could easily be extended to include, as deemed necessary, other parameters such as the dryness and the subsidence reference scales  $MTD_0$  and  $W_0$ , respectively, to have a somewhat more complete learning process.

### 2.3. The Stochastic Bulk Mass Flux Model

With the relative cloud-type plume numbers, or the cloud-type area fractions, defined through the stochastic-lattice plume model above, we are ready to derive the stochastic bulk mass flux equations. First, we note that a sufficient condition to achieve a Poisson process for the distribution of plumes in Equation 5 is to assume that the individual probabilities

$$p_{k,\sigma}(j) = \text{Prob}\{j \text{ plumes of type } \sigma \text{ detraining at } z_D, \text{ for } z_k \leq z_D \leq z_{k+1} | z_k \leq d_\sigma \leq z_{k+1}\}$$

are Poisson distributed with the same rate  $\Lambda$ , independently of  $\sigma$ , as readily justified by the Craig and Cohen (2006) model. We obtain from Equation 4, after taking expectations over the Poisson distribution,

$$M(z) = \frac{\Lambda M_b}{N_L^2} \int_0^{\lambda_z} e^{\lambda(z-z_b)} [N_1 f_1(\lambda) + N_2 f_2(\lambda) + N_3 f_3(\lambda)] d\lambda, \quad (11)$$



where  $f_l$ ,  $l = 1, 2, 3$  are the probability densities illustrated in Figure 1b that are associated with the plume distributions  $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3$ , which in turn are assumed to make up  $\mathcal{P}_\sigma$ . Here,  $N_1, N_2$ , and  $N_3$  are the numbers of lattice sites that are occupied by shallow, congestus, and deep cumulus, respectively, while the Poisson parameter  $\Lambda$  is given in Equation 6, although as already noted the value  $\Lambda = 1$  is used for the single column runs herein.

Note that Equation 11 can be alternatively written in terms of the CAF as

$$M(z) = \Lambda M_b \int_0^{\lambda_z} e^{\lambda(z-z_b)} [\bar{\sigma}_1 f_1(\lambda) + \bar{\sigma}_2 f_2(\lambda) + \bar{\sigma}_3 f_3(\lambda)] d\lambda, \quad (12)$$

where  $\bar{\sigma}_l = N_l/N_L^2$  for  $l = 1, 2, 3$  as defined in Equation 7.

#### 2.4. Entrainment and Detrainment Calculation

For convenience, we use the short-hand notation

$$f(\lambda) = \bar{\sigma}_1 f_1(\lambda) + \bar{\sigma}_2 f_2(\lambda) + \bar{\sigma}_3 f_3(\lambda) \quad (13)$$

and rewrite Equation 12 simply as

$$M(z) = \Lambda \int_0^{\lambda_z} m_\lambda(z) f(\lambda) d\lambda. \quad (14)$$

The precise formulation of  $\lambda_z$  as a function of  $z$ , which is used here, will be given below in Section 2.5. For now, we note that this function is a strictly decreasing function of  $z$ ,  $z_b \leq z \leq z_\infty$ , where  $z_\infty$  is the detrainment level of the deepest penetrating plume, and rewrite the bulk mass flux equation in Equation 1 as

$$\frac{1}{M(z)} \frac{\partial M(z)}{\partial z} = \epsilon(z) - \delta(z), \quad (15)$$

where  $\epsilon(z) = E(z)/M(z)$  and  $\delta(z) = D(z)/M(z)$  are respectively the bulk entrainment and bulk detrainment rates per unit mass flux. After differentiating Equation 14, we obtain

$$\begin{aligned} \epsilon(z) - \delta(z) &= \frac{\Lambda}{M(z)} m_{\lambda_z}(z) f(\lambda_z) \frac{d\lambda_z}{dz} + \frac{\Lambda}{M(z)} \int_0^{\lambda_z} \frac{\partial m_\lambda(z)}{\partial z} f(\lambda) d\lambda \\ &= \frac{\Lambda M_b}{M(z)} \int_0^{\lambda_z} \lambda e^{\lambda(z-z_b)} f(\lambda) d\lambda + \frac{\Lambda M_b}{M(z)} e^{\lambda_z(z-z_b)} f(\lambda_z) \frac{d\lambda_z}{dz}. \end{aligned}$$

Consequently, we set

$$\epsilon(z) = \frac{\Lambda M_b}{M(z)} \int_0^{\lambda_z} \lambda e^{\lambda(z-z_b)} f(\lambda) d\lambda = \frac{\int_0^{\lambda_z} \lambda e^{\lambda(z-z_b)} f(\lambda) d\lambda}{\int_0^{\lambda_z} e^{\lambda(z-z_b)} f(\lambda) d\lambda} \quad (16)$$

and

$$\delta(z) = -\frac{\Lambda}{M(z)} M_b e^{\lambda_z(z-z_b)} f(\lambda_z) \frac{d\lambda_z}{dz} = -\frac{e^{\lambda_z(z-z_b)} f(\lambda_z)}{\int_0^{\lambda_z} e^{\lambda(z-z_b)} f(\lambda) d\lambda} \frac{d\lambda_z}{dz}. \quad (17)$$

Since  $\lambda_z$  is a decreasing function of  $z$ ,  $\delta(z)$  is non-negative and peaks precisely at (detrainment) levels corresponding to the centers of individual cloud-type PDFs that form up  $f(\lambda_z)$  in Equation 13, consistent with Figure 1 of Zermelo-Daz et al. (2015).

To facilitate the approximation of the integrals in Equations 14, 16, and 17, we assume the simple Gaussian shapes for the distributions  $f_1, f_2, f_3$ :

$$f_j(\lambda) = \frac{1}{\sqrt{2\pi\alpha_j}} \exp\left\{-\frac{(\lambda - \lambda_j)^2}{2\alpha_j^2}\right\}, j = 1, 2, 3, \quad (18)$$

where  $\lambda_j$  and  $\alpha_j$ ,  $j = 1, 2, 3$  are respectively the means and standard deviations of the plume entrainment rates corresponding to shallow cumulus, cumulus congestus, and deep cumulonimbus cloud types, respectively.

It remains to approximate the integrals

$$I_1^j = \frac{1}{\alpha_j \sqrt{2\pi}} \int_0^{\lambda_z} \exp \left\{ \lambda(z - z_b) - \frac{(\lambda - \lambda_j)^2}{2\alpha_j^2} \right\} d\lambda \quad (19)$$

and

$$I_2^j = \frac{1}{\alpha_j \sqrt{2\pi}} \int_0^{\lambda_z} \lambda \exp \left\{ \lambda(z - z_b) - \frac{(\lambda - \lambda_j)^2}{2\alpha_j^2} \right\} d\lambda, \quad (20)$$

for  $j = 1, 2, 3$ , to obtain  $\epsilon(z)$  and  $\delta(z)$ . These integrals are further simplified and written in closed form in terms of the error function (erf) and coded as such.

### 2.5. Entrainment Function

To define the entrainment function  $\lambda_z = \lambda(z)$  we follow ZM95 and define  $\lambda_z$  based on MSE. We consider the steady state equation of the MSE of a convecting plume, with a fractional entrainment rate  $\lambda$ , rising from a cloud base  $z_b$  to its detrainment level  $z_D$ :

$$\frac{dh_u}{dz} = \lambda(h - h_u), \quad z_b \leq z \leq z_D, \quad (21)$$

where  $h_u$  is the MSE of the rising plume and  $h$  is the MSE of the environment. The idea is to use this equation to compute  $\lambda$  as a function of the detrainment level  $z_D$  under the two physical assumptions:

1.  $h_u(z_b) = h(z_s) \equiv h_0$ , where  $z_s$  is the surface level, that is, the plume rising from the surface conserves its MSE and it is unsaturated below  $z_b$ .
2.  $h_u(z_D) = h^*(z_D) \equiv h^*$ , the environmental saturation MSE. In other words, at its detrainment level, the rising plume is in thermodynamic equilibrium with the environment at saturation.

A quick integration of the ordinary differential equation in Equation 21 yields

$$h_0 - h^* = \lambda \int_{z_b}^{z_D} (h_0 - h(z')) e^{\lambda(z' - z_D)} dz'. \quad (22)$$

A one-to-one function for  $\lambda$  in terms of  $z_D$  is then obtained by inverting this equation. An easy and cheap way to do this is to simply approximate this integral based on the mean value theorem. However, the mid-point approximation breaks down when  $h^*(z)$ 's concavity varies significantly with height. Thus, instead, we compute the integral in Equation 22 with more accuracy (using a composite Newton-Cotes rule) and then invert the resulting equation for  $\lambda$  using the Newton-Raphson method. Let

$$F(\lambda) = \lambda \int_{z_b}^z (h_0 - h(z')) e^{\lambda(z' - z)} dz' - (h_0 - h^*(z))$$

and

$$F'(\lambda) = \int_{z_b}^z (h_0 - h(z')) e^{\lambda(z' - z)} [1 + \lambda(z' - z)] dz'$$

its derivative. Given an initial guess  $\lambda_0$ , we define

$$\lambda_{i+1} = \lambda_i - \frac{F(\lambda_i)}{F'(\lambda_i)}, \quad i = 0, 1, 2, \dots \quad (23)$$

This yields approximate solution  $\lambda_*$  of  $F(\lambda) = 0$ , when  $i$  is sufficiently large. An initial guess  $\lambda_0 = 10^{-5} \text{ m}^{-1}$  and a maximum of four iterations are used. It has to be noted that the authors haven't investigated whether this iterative

process converges for all initial guesses  $\lambda_0$  and given background moist static energies but the computations seem to be stable and led to reasonable results, for all the case studies considered.

## 2.6. Fixing the Detrainment Level Distributions Parameters

To complete the definition of the detrainment level distributions in Equation 18, which can be loosely thought of as the cloud top distributions for each cloud type, we need to specify the means and standard deviation parameters  $\lambda_j, \alpha_j$  for each  $f_j(\lambda), j = 1, 2, 3$ . Consider the model levels that are closest to the lifting condensation level, the LFC, the freezing level, and the level of neutral buoyancy, denoted here for convenience as by  $z_{lcl}, z_{lfc}, z_{frz}, z_{lnb}$ , respectively, that are associated with the given environmental profile. Let  $\lambda_{lcl}, \lambda_{lfc}, \lambda_{frz}, \lambda_{lnb}$  be the fractional entrainment rates of plumes that detrain at each one of these levels, respectively, based on an appropriate inversion of Equation 22. Let  $0 < \delta_j < 1$  and  $\gamma_j > 0, j = 1, 2, 3$ , be a set of tunable parameters.

We set

$$\begin{aligned}\lambda_1 &= \delta_1 \lambda_{lfc} \\ \lambda_2 &= \delta_2 \lambda_{frz} \\ \lambda_3 &= \delta_3 \lambda_{frz} + (1 - \delta_3) \lambda_{lnb}\end{aligned}\tag{24}$$

and

$$\begin{aligned}\alpha_1 &= \gamma_1 \lambda_1 \\ \alpha_2 &= \gamma_2 (\lambda_{lfc} - \lambda_{frz}) \\ \alpha_3 &= \gamma (\lambda_{frz} - \lambda_{lnb}).\end{aligned}\tag{25}$$

Preliminary numerical experiments revealed that the values  $\delta_1 = 0.6, \delta_2 = 1.25, \delta_3 = 0.2, \gamma_1 = 0.35, \gamma_2 = 0.15$  and  $\gamma_3 = 1.0$  are a good choice for a wide range of test cases. They are fixed as such and used for the results presented here. Further tuning is probably required for 3d simulations.

## 2.7. Implementation

The model used is CAM5, the atmospheric component of CESM1.2.2 which can be downloaded from the UCAR website: [https://svn-ccsm-models.cgd.ucar.edu/cesm1/release\\_tags/cesm1\\_2\\_2](https://svn-ccsm-models.cgd.ucar.edu/cesm1/release_tags/cesm1_2_2), which we run in single column mode.

Generally speaking, a single column climate model (SCM) is an idealized version of a GCM where the horizontal variations are ignored, leading to an isolation of the atmospheric vertical motion. As such SCM's are ideal testbeds for convection parameterizations while keeping the computational cost to the minimum (Gettelman et al., 2019; M. Zhang et al., 2016). In a way, an SCM simulates the atmospheric processes using a single/isolated GCM grid column thus the name SCM.

We recall here that our main strategy for achieving a stochastic mass flux CP is to modify the ZM scheme so that the updraft mass flux and the associated entrainment and detrainment rates are calculated according to the probabilistic formulations in Equations 14, 16, and 17, respectively. The stochasticity arises from the SMCM CAFs that act as weights in the definition of the PDF  $f(\lambda)$  in Equation 13. In addition to the realism of having a dynamically evolving detrainment level distribution, the SMCM introduces variability at subgrid level into the cumulus scheme and we obtain a naturally unified shallow-deep convective parametrization.

While the MSE profile,  $h(z)$ , is used to compute the entrainment function  $\lambda(z)$  using Newton-Raphson in Equation 23, the SMCM is run in parallel to produce the CAF  $\bar{\sigma}_j, j = 1, \dots, 4$ . To this end, the large scale predictors, CAPE, LCAPE, middle tropospheric dryness (D), vertical velocity at the top of the boundary layer (W), CIN, and the strength of the inversion are fed to the SMCM routine to estimate the transition rates in Equation 10 at each GCM time step (and at each column of the GCM, if run in full 3D mode; it is not the case here!).

The bulk mass flux and the bulk fractional entrainment rates are estimated from adequate approximations of the integrals in Equations 14 and 16, respectively. A composite Simpson rule and semi-exact evaluations based on known values of the erf function  $\left(\text{erf}(y) = \frac{2}{\sqrt{\pi}} \int_0^y e^{-x^2} dx\right)$  have been tested and the results were found to be

only mildly sensitive to the choice of either method. To guarantee conservation of mass with machine precision, instead of approximating the derivative  $\frac{d\lambda}{dz}$  in Equation 17 in a similar fashion, we use the mass conservation equation in Equation 15 to estimate  $\delta(z)$  from the already-made estimations of  $M(z)$  and  $\epsilon(z)$ :

$$\delta(z_k) := \epsilon(z_k) + \frac{M_{k+1} - M_k}{z_{k+1} - z_k}, \quad (26)$$

where  $k = 1, 2, \dots$  are the model levels. We note that this is possible because the mass flux  $M$  is known a priori at all levels.

When implementing the SMPCP in SCAM, we altered the way convection is triggered. This was possible only because the SMCM is a shallow-deep unified scheme. In the original-ZM CP the trigger is based on a threshold CAPE value. In other words, the ZM convection routine is called only when CAPE is larger than a certain threshold value. Since in the SMPCP the mass flux is calculated based on the evolving SMCM's CAF, it will be automatically zero if the CAF are all zero. As such the CAPE trigger is not needed and the stochastic convective cloud fraction aggregate  $\bar{\sigma}_1 + \bar{\sigma}_2 + \bar{\sigma}_3$  becomes implicitly the new convection trigger function. This also allows the ZM-SMCM code to be called and used even when the SMCM predicts only shallow convection (or only congestus) and no deep cloud fraction.

Meanwhile, the cloud base mass flux is split into its shallow and deep components,

$$M_b = \begin{cases} \frac{\bar{\sigma}_1 M_{bs} + (\bar{\sigma}_2 + \bar{\sigma}_3) M_{bc}}{\bar{\sigma}_1 + \bar{\sigma}_2 + \bar{\sigma}_3}, & \text{if } \bar{\sigma}_1 + \bar{\sigma}_2 + \bar{\sigma}_3 > 0. \\ 0, & \text{otherwise,} \end{cases} \quad (27)$$

In Equation 27  $M_{bc}$  represents the deep convective cloud base mass flux while  $M_{bs}$  represents its shallow cumulus counterpart.  $M_{bc}$  is computed based on the original CAPE closure and  $M_{bs}$  is based on the boundary layer turbulence closure of Bretherton and Park (2008).

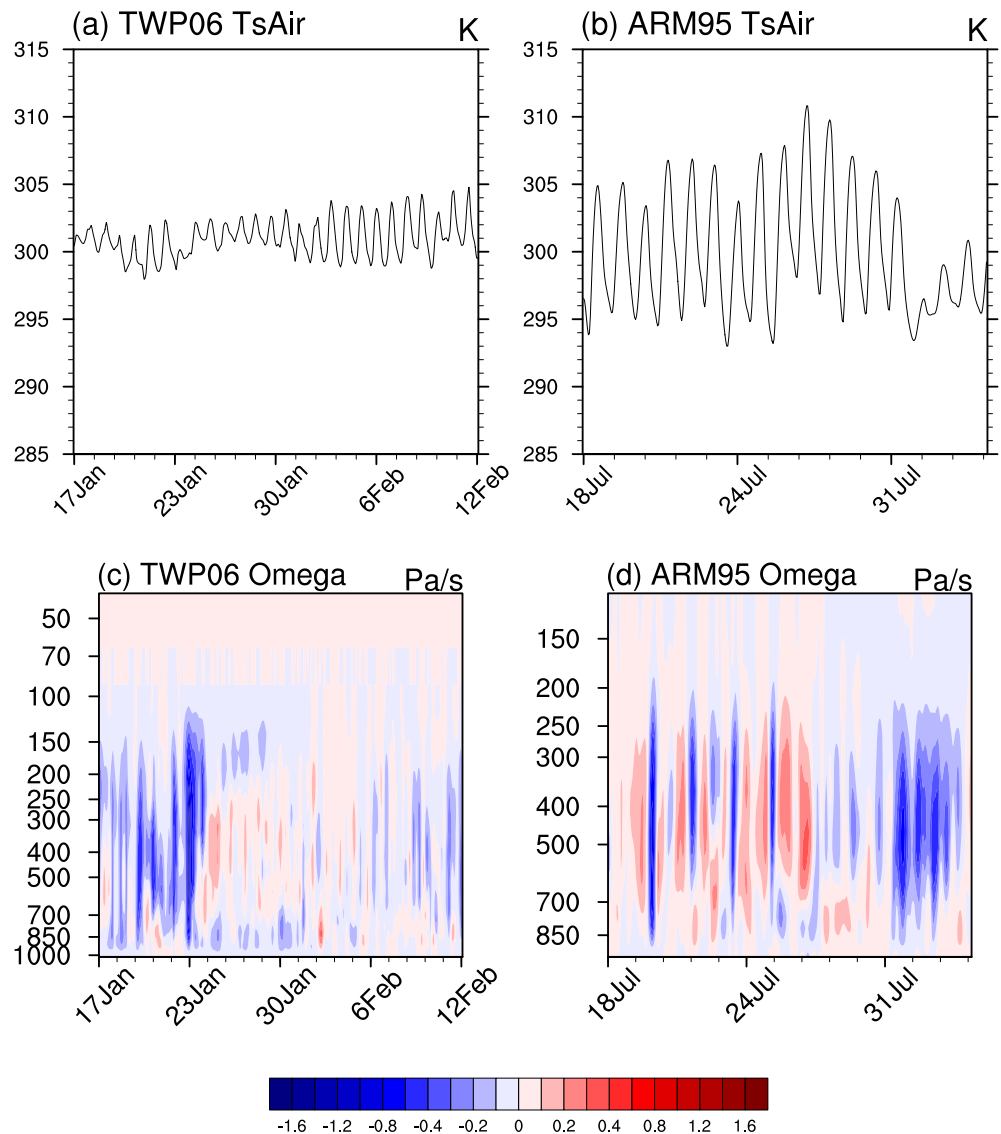
This essentially implies that, convection strength in SMPCP is seamless. It also means that, SMPCP is not only a shallow-deep unified scheme, but this framework also unifies the triggers and closures of shallow and deep convection. We recall that since SMPCP is a unified scheme, the shallow convection parameterization scheme in CAM is turned off when SMPCP is used.

### 3. Data and Observation Test Cases

The comparative results from two case studies that cover both tropical ocean and midlatitude-land convection, namely, the Tropical Warm Pool-International Cloud Experiment (TWP-ICE) which took place in 2006 (denoted here TWP06 for short) and the Atmospheric Radiation Measurement Program Intensive Observation Period of 1995 (ARM95), are reported here. Details on these field campaigns and the inherent weather events can be found for example, in Gettelman et al. (2019) and M. Zhang et al. (2016) and references therein. Nonetheless, a brief summary of their main weather features are given below for the sake of convenience. We emphasize the fact that when combined, these two field experiments contain a rich variety of convection regimes covering land, ocean, and coastal areas, which help to highlight some of the key features of the new parametrization when compared to the default ZM scheme as implemented in CAM and to the available observations. Although not shown here, many other test cases including GATE, TOGA-COARE, ARM97, have showed similar performance for both models, which proves the robustness and representativity of the reported results.

The observed surface fluxes and vertical velocity are used to force the models while the corresponding temperature and relative humidity profiles, and precipitation time series are used as benchmarks to validate the models. The (air) surface temperature and vertical velocity profile time series for the 2 test cases are shown in Figure 2 while the corresponding precipitation records are reported in Figure 3. As we can readily see from these figures, while the time span of the test cases are comparable, roughly 2.5 and 3 weeks, the variabilities of the illustrated forcing fields vary strongly from one case to another, as discussed below.

The forcing fields from both test cases have a temporal resolution of 20 min while the vertical velocity is given at 18 and 40 pressure levels in ARM95 and TWP06, respectively.

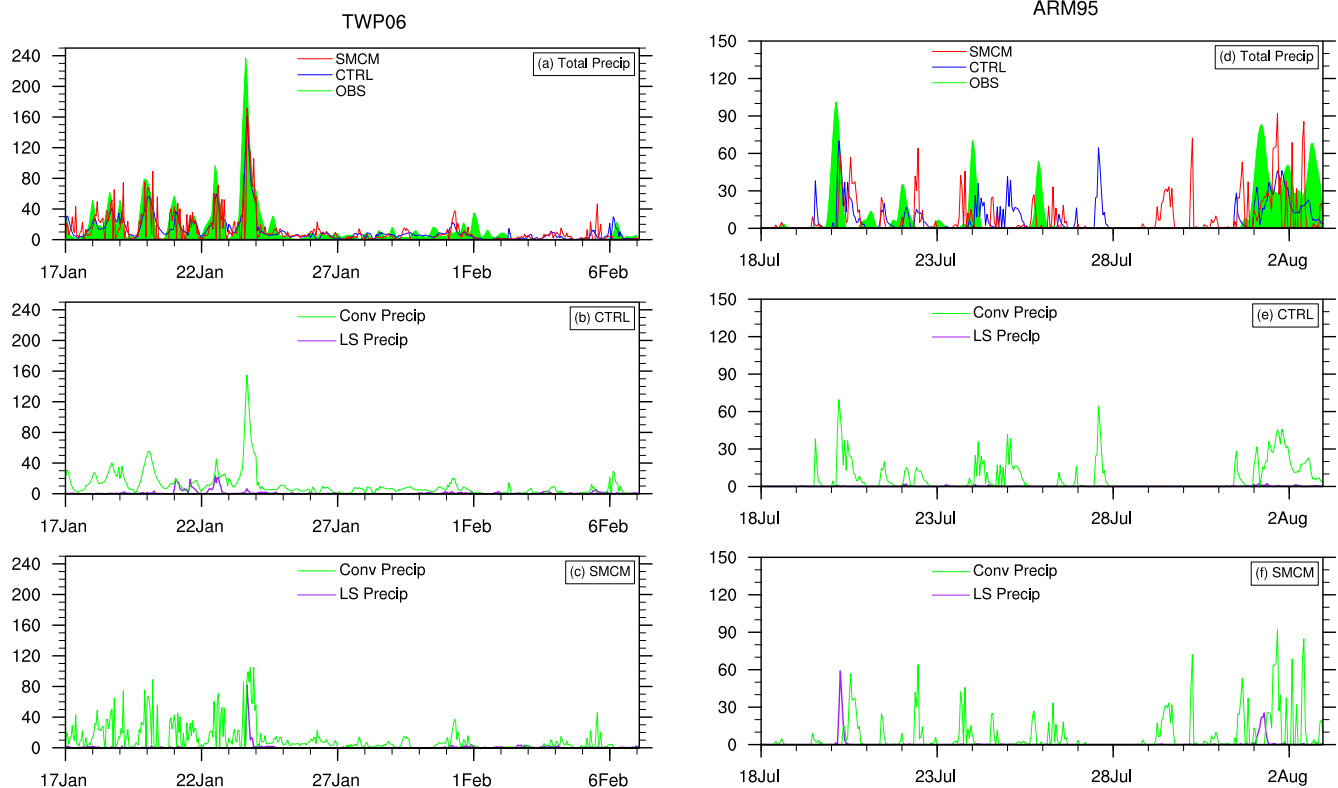


**Figure 2.** Observation test cases used to validate the new convective parameterization. The two main fields used to force the model simulations are plotted, namely, the (air) surface temperature (a) and (b) and the (pressure) vertical velocity  $\omega$  time-height color maps, in  $\text{Pa s}^{-1}$ , (c) and (d). TWP06 (a, c); ARM95 (b, d). See text for details.

The Tropical Warm Pool International Cloud Experiment also known as TWP-ICE, was undertaken in the Darwin, Northern Australia, from 21 January to 13 February 2006 with a view to improve the parameterization of tropical convection and clouds in numerical weather prediction and climate models (May et al., 2008). It was expected during this campaign to understand convection and cloud characteristics which are representative of conditions typical for wide areas of the tropics. A large amplitude MJO passed through the region and, as it can be surmised from Figures 2a and 2c, different phases of the Australian monsoon were observed, including active (21–25 January), suppressed (26 January to 2 February), clear skies (3–5 February) and break (6–13 February) conditions. It should be noted that the break conditions are different from the clear sky conditions in that, during the break phase, isolated convective events were observed. The isolated events tended to be localized in coastal regions and they exhibited a strong diurnal cycle. For more details, we refer the reader to the corresponding precipitation time series plot in the result section below and to Figure 4 of May et al. (2008).

ARM95, on the other hand, is an intensive observation period (IOP) conducted over the southern Great Plains site of the ARM program, supported by the United States Department of Energy, covering an 18-day period in summer of 1995, starting from 0000 UTC 18 July and ending at 2300 UTC 4 August (M. Zhang et al., 2016).





**Figure 3.** Precipitation time series (in mm per hour) throughout the intensive observation period period for TWP06 (a, b, c) and ARM95 (d, e, f). The two top panels (a, d) show the observed (green shading) and simulated times series using the CTRL (blue) and stochastic multicloud model (SMCM) (red) models. The middle and bottom panels (b, e) and (c, f) are for the convective (green) versus the large scale (magenta) precipitation for the CTRL and SMCM models, respectively.

This campaign was led by the SCM Working Group, a subgroup under ARM motivated to improve the parameterization of physical processes in the single column setup of a GCM. In this IOP, three distinct weather regimes of mid-latitude land-convection were observed. In the first 10 days, variable cloudiness and precipitation were observed as associated with an upper level trough over North America. Clear sky conditions associated with a high pressure system prevailed in the following 3 days. Increasing cloudiness, thunderstorms, and intense precipitation, probably associated with mesoscale convective systems, were observed in the last few days of the campaign as can be seen in Figure 4a of Xu and Randall (2000). The forcing field time series in Figures 2b and 2d are consistent with this behavior, especially from the reduced diurnal cycle in the surface temperature and from the strong and extended upward vertical velocity during those last few days of increased cloudiness. This is also reflected in the precipitation time series reported in the result section where a corresponding extended rainy period was recorded.

## 4. Numerical Tests and Validation

### 4.1. Model Set-Up

The SMPCP is implemented in SCAM and tested for the two IOP test cases discussed in Section 3, against the control SCAM which uses the original Zhang-McFarlane scheme (ZM95). Details on the benefits of single column models for climate model development and on precisely how SCAM was carved out of CAM can be found in Gettelman et al. (2019) and references therein. Gettelman et al. (2019) particularly emphasizes that in its latest version (CAM6) SCAM can be run in three different modes: The large-scale temperature profile can be either (a) nudged to the corresponding observed temperature profile, (b) nudged to the corresponding temperature profile obtained from the full 3D CAM simulation, or (c) the model is run freely without any nudging and the full radiative transfer is used to force the temperature field instead. The idea behind (a) and (b) is that the nudging helps alleviate some of the most apparent shortcomings of the single column set up, namely, the missing effects of

horizontal advection and convergence and momentum transfer, and as such the third option result in temperature profiles that deviate significantly from their observed counterparts.

Unfortunately, Option 1 is not available in CAM5, which is the version in which we have presently implemented the SMPCP and Option 3 is computationally too expensive. Thus, to avoid testing the models in unphysical regimes, in which the model climatologies have significantly diverged over time from reality, we divided each one the two IOP's (that are roughly 20 days long) into three periods of 1 week or less. In other words, for each test case, the simulations were run for the total duration of the IOP but the models were reinitialized at the end of each sub period lasting at most 1 week. The TWP06 simulations are thus initialized/re-initialized on 2 August 1996, 24 January 2006 and 1 February 2006 while the ARM95 simulations were initialized/reinitialized on 18 July 1995, 25 July 1995, and 2 August 1995. This allowed to reduce the deviation (in both models) from the observed mean temperature from as much as  $8^\circ$ , if the free runs were instead made for the total duration of the test cases without re-initializations, down to  $2^\circ$ . These resulting temperature and humidity biases are not shown here for the sake of streamlining.

SCAM is run for the duration of the test case under consideration with a 10 min time step and 30 vertical levels, based on a hybrid  $\sigma - p$  coordinate system (Neale et al., 2010). The lowest level is at  $\sim 992.55$  hPa while the highest is at  $\sim 3.64$  hPa and there are 5 full levels below 900 hPa. The forcing fields are updated every two time steps and the vertical velocity is interpolated linearly to fit into the model's vertical resolution.

To test the sensitivity of the results to the transition time scales ( $\tau_{kl}$ ,  $k, l = 0, 1, 2, 3, 4$ ) in Equation 10, we consider the suite of experiments for the SMPCP listed in Table 1 that use 4 different sets of these parameters. As already mentioned, the four parameter sets in Table 1 are obtained based on a combination of rigorous inference from observational and LES data and educated guesses. Because of the importance of these parameters in terms of the SMCM sensitivity, we briefly describe here these parameter sets.

We recall that the SMCM used here is extended from the original model first introduced in Khouider et al. (2010) to include the shallow cumulus cloud type as the fourth cloudy state of the Markov process. Thus, all the transition timescales involving this new cloud type are novel to this study.

Noteworthy, despite the sensitivity of the results to the transition scales as discussed below, the results shown here are highly encouraging, in the sense that the most desired features of the new parameterization remained robust. Further research for finding the most optimal set of transition time scale parameters for the SMCM, using statistical inference from observed data, is still ongoing and this may potentially lead to a highly competitive-improved version of the SMPCP.

Since the SMCM0 set of transition timescales on the first column in Table 1 is the closest to reality, we use it as our default parameter set and below, we refer to the SCAM model using the SMPCP scheme based on this parameter set as simply SMCM while the control SCAM model is referred to as CTRL, for the sake of brevity. For all the other parameter sets, the corresponding SMPCP experiments are referred to as SMCM1, SMCM2, and SMCM3, respectively.

#### 4.2. Main Features of the SMPCP

The precipitation timeseries for the TWP06 and ARM95 test cases are reported in Figure 3. The top panels compare the observed (green) hourly rainfall with its simulated counterpart using the CTRL (blue) and the SMCM (red) single column models. We recall that TWP06 is for convection over a tropical ocean and coastal areas while ARM95 is for mid-latitude continental convection. Overall, the two models seem to be in agreement in terms of capturing the observed main rainy episodes despite some significant departures in amplitude and timing, especially for the ARM95 test case. There are also significant departures between the models and the observation toward the end of the TWP06 experiment.

While none of the models seem to clearly outperform the other, there are some desirable features that seem to be present in the SMCM and not in the CTRL. For instance, the SMCM shows overall more variability and appears to capture higher rainfall events, that are comparable to the observation, especially during the suppressed period of the TWP06 IOP, roughly between 24 January and 12 February, corresponding to the period with weak vertical velocity in Figure 2f, where convection does not seem to be very strong. Also, putting together the vertical velocity plot in Figure 2d and the ARM95 rainfall time series in Figure 3d, one may speculate that the precipitation peaks around that period, that are predicted by the two models but not present in the observations (28 July for

CTL and 30 July for SMCM), are perhaps a direct result of the few weak lifting events that are seen to occur during this period. The discrepancy between the models and the observation may be due to the presence (in the real world) of physical processes such as dry intrusions or microphysical details (turbulent eddies and variability in aerosols), which are meant to inhibit precipitation but are not captured by the single column framework due to the missing horizontal transport.

The simulated total precipitation is divided into its convective and large scale components akin to the parts of rainfall accounted for by the convection parameterization (i.e., sub-grid precipitation) and by rainfall generated from condensation occurring at the grid scale. The bottom four panels in Figure 3 contain the plots of the convective and large scale precipitation time series for TWP06 and ARM95 as simulated by the CTRL and the SMCM models. While both models depict relatively much less large scale rain, SMCM seems to predict higher amounts when it does. This is reminiscent to the fact that the SMCM does not always promote deep convection but instead allows for the environment to progressively moisten. Also, based on many other experiments not shown here, apart from isolated instances such as the one event around 20 July in ARM95, the SMCM's large scale precipitation seems to occur in connection with major (strongly peaked and/or long-lasting) convective rain episodes, either within or following right after convection events. This may be an indication that the large scale precipitation in the SMCM is likely linked to the presence of environmental moisture from cumulus cloud detrainment, and as such it could be rightfully associated to stratiform rain (as it is often referred by the GCM community) following deep convection, as it occurs in organized convective systems (Khouider & Majda, 2006; Majda & Shefter, 2001; Mapes, 2000). On the one hand, organized stratiform precipitation is often associated with propagating mesoscale convective systems that can occur within monsoon intra-seasonal oscillations and MJO convective envelopes (Ajayamohan et al., 2016; Choudhury & Krishnan, 2011; Deng et al., 2016; Kumar et al., 2017; Moncrieff, 2010; Moncrieff et al., 2017; Parker & Johnson, 2004; Schumacher & Houze, 2003) and on the other hand it has been argued in several studies that the increase in the simulated large scale precipitation, at the expense of convective precipitation, improves the fidelity of the underlying model in terms of simulating tropical weather and climate variability (Abhik et al., 2017; Chattopadhyay et al., 2009; Choudhury & Krishnan, 2011; Ganai et al., 2019; Kumar et al., 2017; Tao et al., 2010; Tokay et al., 1999) as top-heavy heating seems to be beneficial (Lappen & Schumacher, 2014) for the same dynamical reasons of exciting the second baroclinic mode for example, (Khouider & Majda, 2006; Majda & Shefter, 2001; Moncrieff et al., 2017). Given the aforementioned success stories behind the implementation of the SMCM in GCMs, this feature is good news.

The observed and simulated perturbation temperature (with respect to the corresponding time-mean vertical profile associated with the IOP, not shown) and the relative humidity (with respect to saturation over liquid water) for both the observations and the two model simulations are reported in Figure 4. In agreement with the observation made by Gettelman et al. (2019), in these single column free runs, the temperature and accordingly the relative humidity profiles deviate rather quickly from the corresponding observed profiles, resulting in mean biases of up two degrees (results not shown). As pointed out in that article, such biases are mainly due to the lack of horizontal coupling and other physical processes that are not represented by the SCM and as such it is not something expected to be corrected by the convective parametrization. Perhaps akin to the lack of lateral coupling, the ARM95 observed temperature perturbations in Figure 4g, show two successive streaks of positive and negative anomalies that are perhaps associated with a packet of vertically propagating gravity waves that are totally absent in the two model simulations.

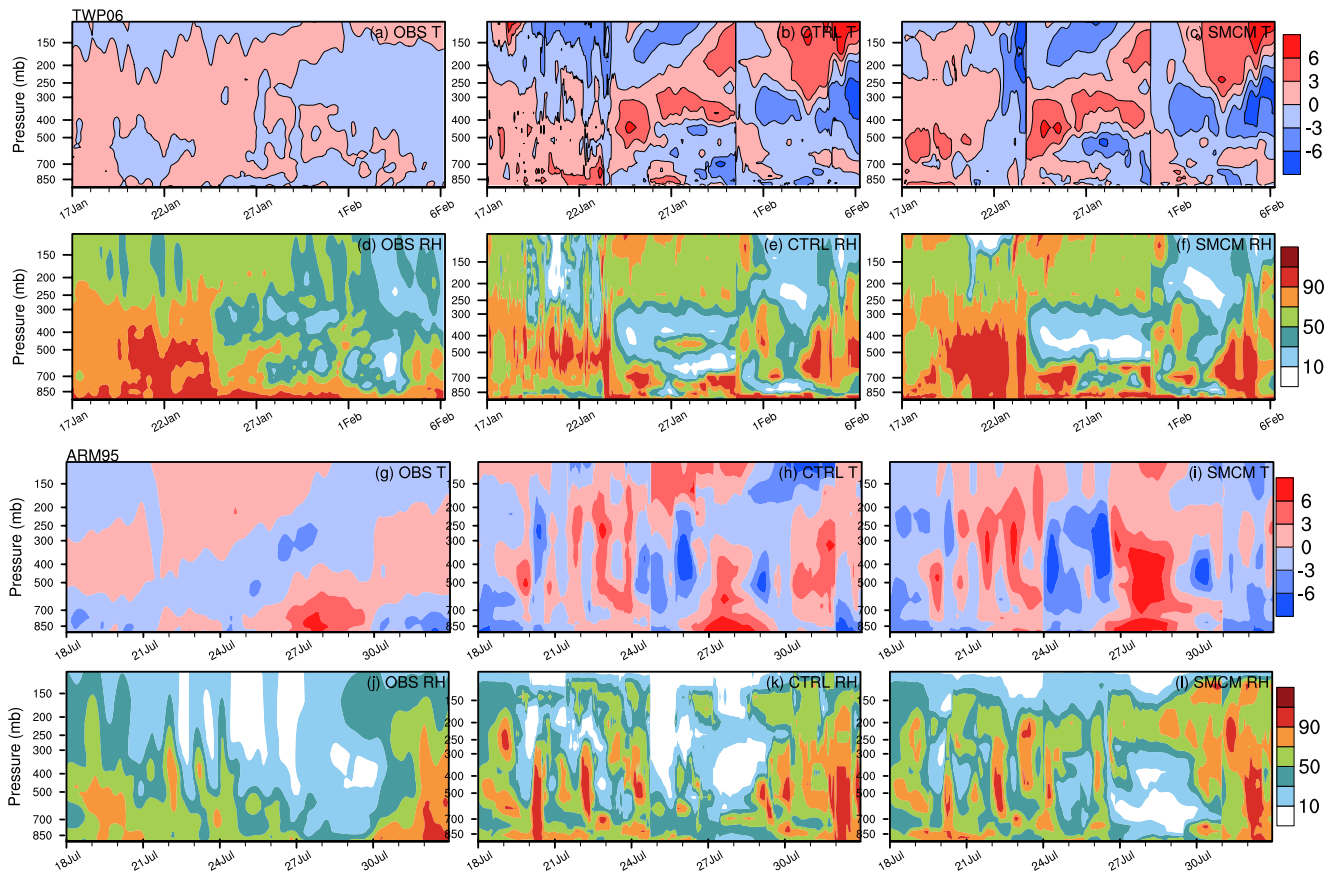
Nonetheless, the SMCM seems to better capture the structure and strength of the significant moisture anomaly during the first week of TWP06 experiment, shown in Figures 4d–4f. The SMCM shows a near surface as well as an upper tropospheric positive moisture anomaly in agreement with the observation, which may be associated with the unification of shallow and deep convection and the resulting intermittency of deep convection and a more progressive moistening of the atmospheric column prior to deep convection.

The observed and simulated profiles of the apparent heat source and moisture sink, known as  $Q_1$  and  $Q_2$ , respectively (Yanai et al., 1973), for the TWP06 and ARM95 cases are reported in Figure 5. To be precise, we recall that

$$Q_1 = c_p \frac{\partial T}{\partial t} - c_p \left( \bar{\omega} \frac{RT}{c_p p} - \bar{\omega} \frac{\partial T}{\partial p} \right), Q_2 = -L \frac{\partial q}{\partial t} - L \bar{\omega} \frac{\partial q}{\partial p},$$

where  $c_p$  is the specific heat at constant pressure,  $T$  is (large-scale) temperature,  $\bar{\omega}$  is the (imposed) vertical (pressure) velocity,  $R$  is the gas constant,  $L$  is the latent heat of condensation, and  $q$  is the (large-scale) specific humidity.

Unlike the temperature and humidity profiles in Figure 4, the simulated  $Q_1$  and  $Q_2$  are overall in good agreement with the observations, consistent with the agreements shown in the precipitation timeseries in Figure 3. This is

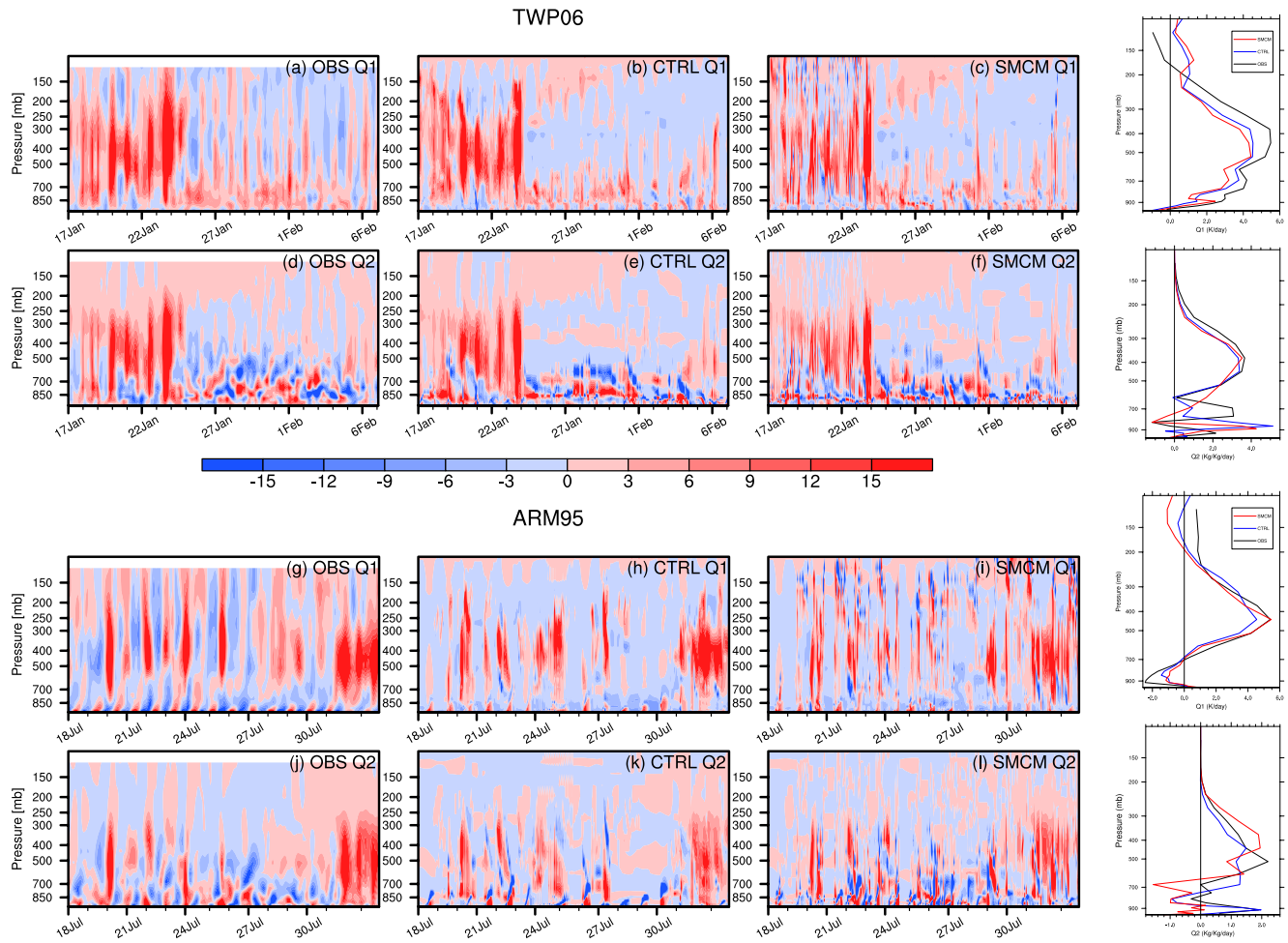


**Figure 4.** Time-height filled contour plots for the perturbations, from the corresponding mean profile, of temperature (a–c, g–i) and relative humidity (d–f, j–l). The observations are shown in panels (a, d, g, j) while the CTRL and stochastic multicloud model model simulations are in panels (b, e, h, k) and panels (c, f, i, l), respectively. The top two rows are for the TWP06 test case while the bottom two are for ARM95. The temperature units are in Kelvin and relative humidity is in percent.

another indication that the biases in the state variables are more likely due to the lack of horizontal coupling rather than to deficiencies in the CP. Nonetheless, there are a few important details in which the SMCM seems to better match the observations. Overall the CTRL model seems to be unable to capture some of the observed deeply penetrating heating and cooling events unlike the SMCM which seems to do so. This is the case, for instance, for the strong heating episode in the TWP06 case near Jan 23 and the observed succession of discrete events between 23 and 30 January. Also, toward the end of the TWP06 IOP, the heating anomalies in the CTRL seem to be capped below roughly 200 mb while they penetrate higher in both the observations and the SMCM. The same is true for the ARM95 test case, especially in terms of the succession of heating events before 26 July. The deeply penetrating heating and cooling events in the SMCM and observation time-height plots seem to cancel each other out in the time-mean profiles, as shown on the most-right panels of Figure 5; These panels show more agreement between the two models and the observations, and more so between the two models themselves. This may be justified by the near balance, on average, between cooling due adiabatic lifting and latent heating according to the weak temperature gradient approximation, which is heavily controlled by the imposed vertical velocity, but the episodic penetrative convective events have many more significant consequences to be ignored such as the impact of convection on cloud cover and the transport and distribution of aerosols and other important chemical tracers.

We now turn into the details of the sub-grid parameterization and present in Figure 6 the time-height plots of the bulk updraft mass flux,  $M_u$ , the bulk entrainment rates,  $E_u$ , and the bulk detrainment rate,  $D_u$ , for the TWP06 and ARM95 IOP's as they are simulated by both the SMCM and CTRL models. Although, there aren't any observations (of these variables) to compare to, one could infer from such plots, at least qualitatively, to what degree of realism the small scale process of convection is being emulated. As expected both models produce disturbances in those three quantities that are consistent with the occurrences of the major precipitation and heating-and-drying events seen in Figures 3 and 5. However, there are some fundamental differences inherent to how those disturbances are presented.



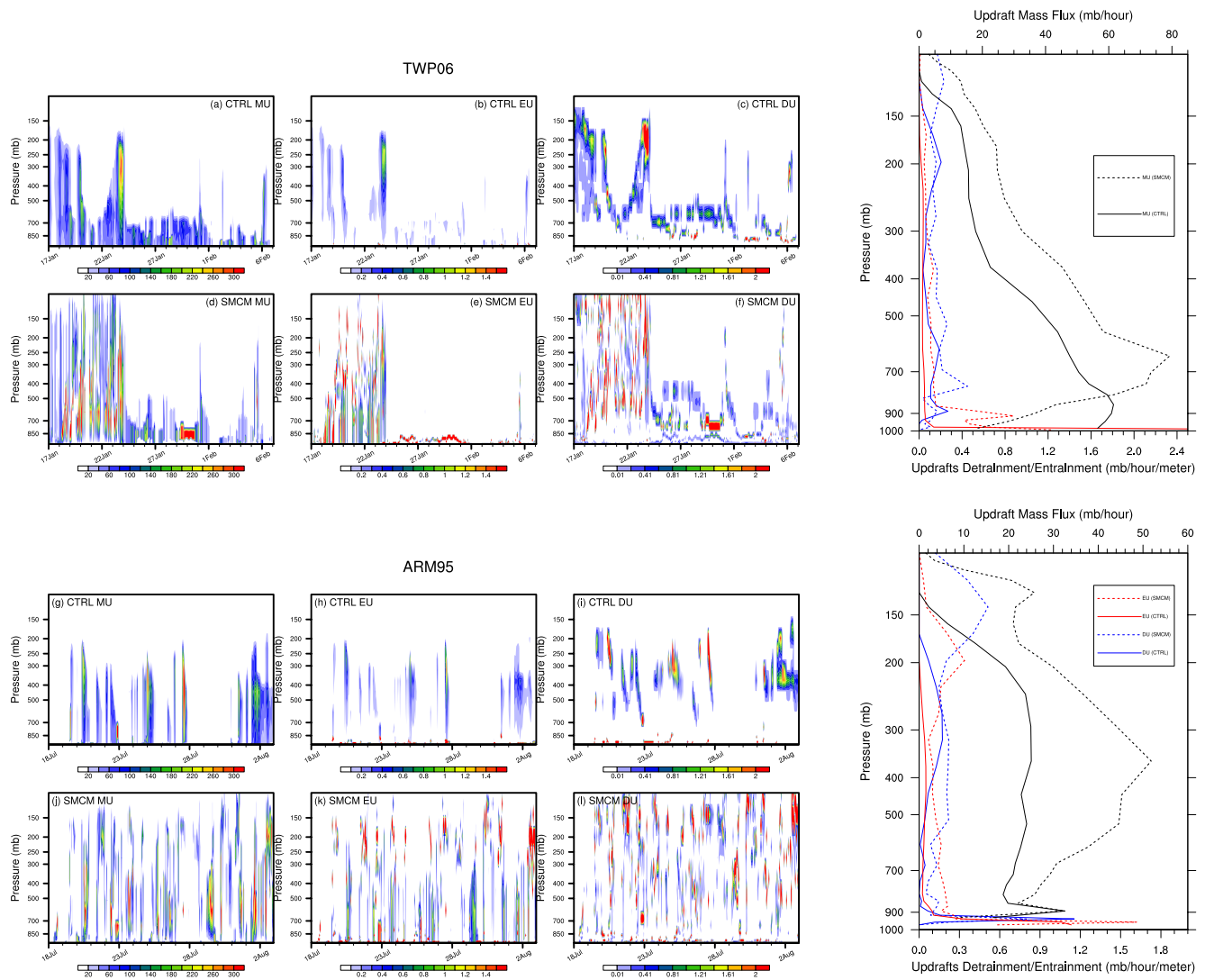


**Figure 5.** Time-height filled contour plots for Q1 (a–c, g–i) and Q2 (d–f, j–l), both in units of Kelvin per day. The observations and the CTRL and stochastic multicloud model (SMCM) simulations are in panels (a, d, g, h) and panels (b, e, h, k) and (c, f, i, l), respectively. The fourth and last column depicts the time-mean vertical profiles of Q1 (d, l) and Q2 (h, p) for the observations (gray) and the CTRL (blue) and SMCM (red) simulations. The top two panels are for TWP06 while the bottom two are for ARM95.

While the CTRL presents somewhat smooth and long lasting  $M_u$ ,  $E_u$ ,  $D_u$  events, the SMCM shows high frequency and more intermittent events. This likely results from the fact that the original ZM scheme used by the CTRL model is slaved to the large-scale dynamics, as per the quasi-equilibrium constraint, while the SMCP has significant degrees of freedom provided by the stochastic dynamics, which allows it to produce or rather emulate the subgrid variability of deep convection. Indeed the SMCM time series of  $M_u$  for example, are more in line with high-resolution simulations of deep convection, which show that single convective events are more abrupt and last only one to 2 hr and not days (e.g., Kuang & Bretherton, 2006; J. M. Peters et al., n.d.). Although one can argue that the goal of a CP is to represent the bulk effect of convection, the difference in event statistic and scaling is important for a realistic emulation of the interactions between convection and large scale dynamics, which are inherently nonlinear, happen rather discontinuously—on much finer scales (Arakawa, 2004), and are believed to be central for a better representation of organized convective systems at synoptic and intra-seasonal scales (Majda, 2007a, 2007b; Moncrieff & Klinker, 1997; Khouider, 2019; Randall et al., 2003).

Other fundamental differences between the CTRL and SMCM simulations of  $M_u$ ,  $E_u$  and  $D_u$  in Figure 6, include an overall deeper penetration in the upper troposphere of  $M_u$  disturbances in the SMCM, consistent with the deeper  $Q_1$  and  $Q_2$  instances that are captured by the SMCM model and not by the CTRL, very distinct shallow/congestus episodes of convection such as during the suppressed period of the TWP06 IOP between 23 January and 2 February, a more vertically spread out entrainment and more importantly the tri-modality exhibited by the detrainment, which occurs at the three main stable layers, near the inversion level, the freezing level, and the tropopause,



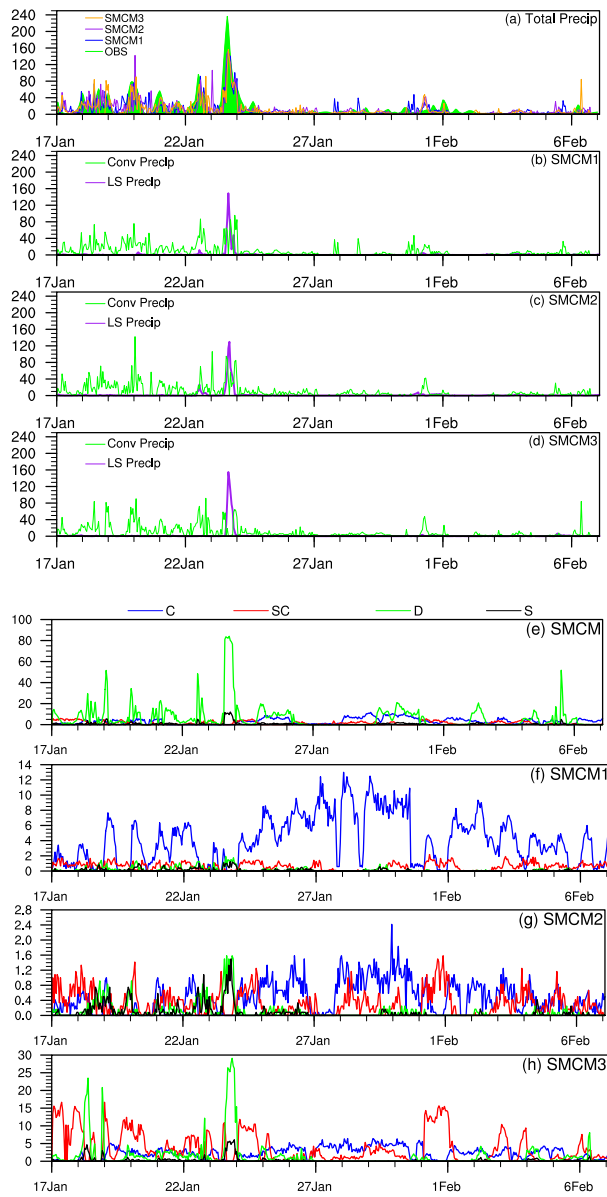


**Figure 6.** Time-height color-filled contours updraft mass flux per unit mass ( $M_u$ , in units of millibar per hour) and associated bulk entrainment and ( $E_u$ , in mb/hour/meter) detrainment ( $D_u$ , in mb/hour/meter) as simulated by the CTRL (a–c, g–i) and the stochastic multicloud model (d–f, j–l) models for the TWP06 test case (a–f) and the ARM95 (g–l) test cases. The panels of the last (right) column depict the corresponding time-mean vertical profiles.

consistent with observations (Johnson et al., 1999; Zermelo-Daz et al., 2015) and the model design as illustrated Figure 1. This has also been observed in other test cases (results not shown). In the CTRL, mass flux detrainment seems to occur only at the top of the cloud “aggregates” that are too bulky to emulate the subgrid variability due to intermittency of individual plume events. The detrainment events at lower level, which are a direct result of the addition of congestus cloud types, are important for mid-tropospheric moistening (Derbyshire et al., 2004; Hohenegger & Stevens, 2013; Johnson et al., 2015; Kuang & Bretherton, 2006; Waite & Khouider, 2010). Both the deepening and the tri-modality of convection (updraft mass flux) in SMCM are evident in the mean vertical profiles depicted on the two most right panels in Figure 6. The deepening of convection in the SMCM may indeed be correcting a major deficiency in CAM5 which tends to produce lower cloud tops compared to CloudSat observations as it has recently been reported in M. Wang and Zhang (2018).

### 4.3. Sensitivity to the Stochastic Transition Timescales

As pointed out above, the SMCM’s transition timescales are highly uncertain and that efforts to learn them from data are underway (Cardoso-Bihlo et al., 2019; De La Chevrotière et al., 2015). It is thus important to know how sensitive to these parameters, the SMCCP could be? In Figure 7, we plot the TWP06 precipitation time series



**Figure 7.** The TWP06 precipitation time series simulated by the single-column Community Climate Model model using the stochastic multi-cloud plume cumulus parameterization parametrization based on the different sets of transition time scales listed in Table 1. The total precipitation (a) from observations (green) and from the SCAM-SMPCP simulation using the SMCM1 (blue), SMCM2 (magenta), and SMCM3 (orange) timescales (See Figure 3 for SMCM0 and CTRL). The contributions from convective (green) and the large-scale (magenta) precipitation for stochastic multcloud model (SMCM) 1 (b), SMCM2 (c), and SMCM3 (d). The SMCM cloud area fraction time series obtained for the TWP06 test case with transition timescales: SMCM0 (e), SMCM1 (f), SMCM2 (g), SMCM3 (h). Each panel depicts the congestus (blue), shallow cumulus (red), deep convective (green), and stratiform (black) cloud type area fractions.

obtained with the SMPCP and the associated stochastic CAF of shallow, congestus, deep and stratiform cloud types as predicted by the SMCM when the different sets of transition time scales in Table 1 are used.

As we can see from Figure 7, the SMCM cloud area fraction time series change drastically from one parameter set to another. This is expected since those parameters affect directly the transition probabilities of the Markov chain which in turn directly affect the probability distribution of the lattice-stochastic process, that is, the filling fractions. However, the simulated precipitations remain fairly close to one another and to the observations. The amounts of convective versus large-scale precipitation are also not very sensitive to the probability transition parameters, as for example, the same strong peak in large scale precipitation near 23 January that appears to characterize the SMCM runs as opposed to the CTRL run. This suggests that the stability of the simulated precipitation by the SMPCP to changes in the transition timescales is not something solely inherent to stability of the SCAM model. Moreover, although not shown here, the other features such as the deeper and tri-modal nature of convection have also persisted.

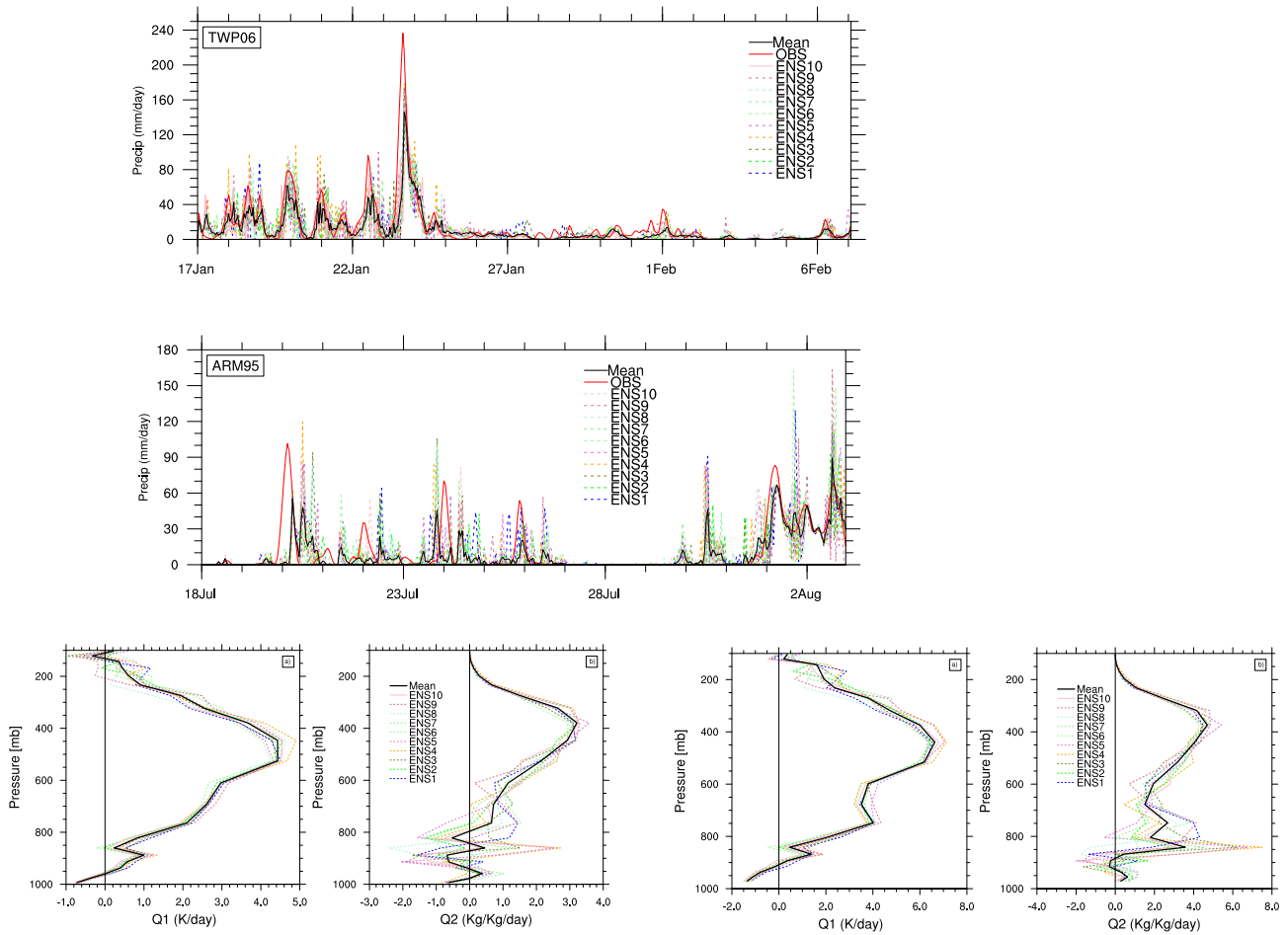
#### 4.4. Effect of Stochasticity: Similarity of Ensemble Members and Ensemble Spread

To assess the effect of the randomness on the SMCM simulations, we repeat the SMCM runs for both the TWP06 and the ARM95 test cases 10 times, each time with a randomly chosen seed used in the code's random number generator. This results in two 10-member ensembles of simulations for the two experiments. The corresponding precipitation time series and the Q1 and Q2 time-mean vertical profiles are shown in Figure 8. The observation mean profiles of Q1 and Q2 are not repeated here for the sake of clarity. The reader may refer to Figure 4 for a comparative view of the observed versus the simulated Q1 and Q2 profiles.

As can be seen from the figure panels, there is spread in both the precipitation time series and the Q1 and Q2 mean profiles. The spread is similar in the two experiments. Moreover, despite the large spread shown in the precipitation time series, for example, the individual ensemble members show very similar behaviors. They all show precipitation peaks of comparable strength albeit they do not always happen at the exact same time and are not at all identical strength-wise. More importantly, when the model is successful in capturing the observed precipitation, all the members seem to do so, as in the rainy period between 17 and 22 January in the TWP06 experiment, for example, and when the model fails to capture the exact timing of a major precipitation event, such as the ARM95 major event of 20 July, or when it shows a “fictitious” rainfall event, that is not recorded in the observation, such as the simulated event between 30 and 31 July, it fails similarly in all its ensemble members.

On the one hand, the last point in the above discussion is consistent with the idea that such systematic failures are perhaps associated with the limitations of the SCAM framework that does not represent physical processes associated with horizontal dynamics such as the advection of moisture and temperature such as occasional dry intrusion. On the other hand, the whole discussion, nonetheless, demonstrates that although there is inherently some

significant spread associated with the SMCM's randomness, the large scale effects are somewhat stable and that in this sense each single member can be deemed representative of the ensemble. This highly desirable feature of the SMCM is likely rooted from the fact that the stochastic dynamics are based on a Markov process with



**Figure 8.** Stochastic multicloud model ensemble simulations. Each dashed line on the six panels represents one ensemble member while the thick black line is for the mean ensemble. (I) and (II) the precipitation time series for the TWP06 and ARM95 experiments, respectively, with the observation plotted in red. (III) and (IV) are for the  $Q_1$  (a) and  $Q_2$  (b) time mean profiles.

obeys detailed balanced with respect to a slowly varying equilibrium measure that depends only on the grid-scale predictors. Also, the fact that, in the long run, the Gillespie algorithm, utilized here, provides an unbiased sampling of the equilibrium measure makes the different ensemble members independent and identically distributed, under identical large scale conditions.

Nevertheless, when affordable, the ensemble runs can be beneficial and the spread could provide a probabilistic solution that can take into account some of the uncertainties due to natural climate variability and due to model error. This is illustrated in Figures 8 (III) and 8 (IV), which both show a relatively significant spread in both the  $Q_1$  and the  $Q_2$  mean profiles, despite the underlying smoothing associated with the time averaging. Moreover, it is interesting to see that the largest discrepancies between the various ensemble members in the time-mean profiles occurs in areas where the model errors are the largest or the profiles themselves are more oscillatory, such as below 600 mb for the  $Q_2$  profiles and above 350 mb for the  $Q_1$  profiles.

### 5. Conclusion

An extended version of the SMCM of Khouider et al. (2010) has been proposed here and used to design a stochastic multicloud plume CP (SMCPCP) which is in essence a stochastic version of the ZM scheme (ZM95). The extended SMCM incorporates shallow cumulus clouds in addition to congestus, deep, and stratiform cloud types that characterize the original SMCM (Khouider et al., 2010). In addition to the generalization of the SMCM, this extension allows the treatment of shallow convection within the framework of the ZM scheme resulting in a unified shallow-deep CP (Park, 2014a, 2014b). We obtain a spectral mass-flux stochastic CP based on an

ensemble of randomly evolving plumes, determined from sampling predetermined distributions of detrainment levels or equivalently entrainment rates, conditional on the CAF from the SMCM. Empirically known stability layers of tropical convection (Johnson et al., 2015), namely the trade wind inversion, the freezing level, and the level of neutral buoyancy are used as a design principle to set the parameters of the detrainment level distributions of shallow cumulus, cumulus congestus, and cumulonimbus cloud types, respectively.

The SMCM predicts the area fractions of the four cloud types above and the first three are in turn used to determine the portions of the plume ensemble that belong to shallow cumulus, cumulus congestus, and deep cumulonimbus clouds. The dynamically evolving distributions are used to determine the bulk mass flux profile as the expected value of the plume ensemble based on (conditional) the SMCM CAF realizations. The bulk entrainment and detrainment rates are then obtained from the variation with height of the stochastic bulk mass flux.

The new stochastic parameterization is implemented here in the context of the single-column Community Atmospheric Model (Neale et al., 2010, CAM5) and validated based on test cases from two IOP that are made available through the CAM repository (Gettelman et al., 2019) involving a mixture of tropical convection over the ocean, monsoonal, and midlatitude land convection, namely the TWP-ICE and ARM95 experiments. The new parameterization is validated against the deterministic ZM scheme, which is the default CP in CAM5, in terms of its capacity to reproduce the observed precipitation timeseries, the temperature and relative humidity perturbations, and the corresponding latent heating and cooling and drying and moistening profiles. A large uncertainty in the SMCM remains the choice of the time scales associated with the transition probabilities of the Markov process between the various cloud types and clear sky conditions. Accordingly, here we considered four different sets of transition time scales to apprehend the sensitivity of the model to these parameters.

As can be seen in Figures 3 and 7, all the model experiments including the control runs (CTRL), using the default ZM scheme, capture the overall trends of the precipitation timeseries. Noteworthy, the stochastic runs exhibit more variability. However, an important key difference between the CTRL and stochastic experiments lies in the relative amounts of stratiform precipitation due to grid-scale condensation versus convective precipitation, which is produced by the CP, predicted for each test case. Two such occurrences are particularly noticeable in the ARM95 test case where in both cases all the SMPCP runs exhibit significant amounts of large-scale precipitation while the CTRL exhibits mainly only convective precipitation. This was also the case in other test cases though the results are not shown here. A better representation of grid scale precipitation in climate models is believed to be paramount for improved simulation of convectively coupled waves, the MJO and monsoon dynamics (Abhik et al., 2017; Ajayamohan et al., 2016; Choudhury & Krishnan, 2011; Deng et al., 2016; Ganai et al., 2019; Lappen & Schumacher, 2014; Kumar et al., 2017; Mapes, 2000; Moncrieff et al., 2017).

We note however that our main goal here is not per se to demonstrate—just yet—the superiority of the SMPCP over the original Zhang-Macfarlane scheme in the general setting, because the performance of the parameterization in the context of a full 3D/realistic climate simulation will largely differ from the single column setting. Thus, the results presented here should be regarded as mainly a sanity check used to demonstrate that the numerical code is working and the scheme's performance is acceptable. Nonetheless, the plots of subgrid quantities such as updraft mass flux and entrainment and detrainment rates, despite the lack of observational references, suggest that the SMCM provides a better representation of convective processes in many respects. For instance the updraft events are more intermittent and exhibit life cycles that are measured in hours and not days (unlike the CTRL), consistent with LES results (Kuang & Bretherton, 2006; J. M. Peters et al., n.d.). Updrafts are generally deeper in the SMCM than in the CTRL leading to occasionally deeper heating and drying events for which the  $Q_1$  and to some extent the  $Q_2$  bursts are deeper, a feature shared by the SMCM simulations and the corresponding observation as shown in Figure 5. The deepening of convection by the SMCM may be correcting a serious shortcoming in CAM5, which tends to predict lower cloud top heights as compared to CloudSat data (M. Wang & Zhang, 2018).

More importantly, the SMCM simulations systematically exhibit the tri-modal structure of tropical convection, visible in the bulk detrainment profiles (Figure 6 and other test cases not shown here) consistent with its design principle (Figure 2b). While the CTRL only rarely showcases three detrainment levels, the tri-modality of cumulus convection is widely accepted and believed to be a universal feature of organized convection as it is associated with the three main stability layers (the trade wind inversion, the freezing level, and the neutral buoyancy near the tropopause) that are responsible for the prevalence of the three main cloud types (shallow cumulus, cumulus congestus, and cumulonimbus) associated with organized convective systems (Johnson et al., 2015; Kiladis

et al., 2009; Mapes, 1993; Mapes & Houze, 1993; Mapes et al., 2006; Wu & Moncrieff, 1996; Zermelo-Daz et al., 2015).

Given the preceding success of the SMCM in the representation of atmospheric tropical dynamics and rainfall variability at synoptic and intra-seasonal/planetary scales (Dorrestijn et al., 2016; Goswami et al., 2017a, 2017b, 2017c; Khouider, 2019; K. Peters et al., 2017), the confirmation of the SMPCP in a fully coupled CAM and CESM is expected. Such research work is currently undertaken by the authors and collaborators and the results will be reported elsewhere in the near future.

As it can be expected the simulated CAF timeseries in Figure 7 are very sensitive to the choice of the transition timescales. Nonetheless, this did not prevent the SMPCP to simulate the overall trends of the precipitation timeseries and reproduce almost identical large scale and similar small scale features, such as deep penetration and tri-modality of convection independently on the choice of transition parameters. Indeed, despite this extreme sensitivity, bursts of deep-convective CAF associated with all the major convective precipitation events are visible in all the test cases and in all the model experiments. Nonetheless, the uncertainty of the transition timescale and other important parameters remains the main challenge of the SMCM and more research is needed in order to comprehensively learn these parameters from data. Arguably, this also can be viewed as a major advantage of the SMCM since the transition timescales and other parameters—such as the reference scales of the large scale predictors CAPE0, CIN0, DTM0, and  $W_0$ , as needed, can be systematically learned from data. Also, unlike a brute-force machine learning strategy that for example, tries to learn the whole temperature and moisture tendencies from observation data, the SMCM uses physics to reduce the learning task to a well defined and significantly smaller set of parameters.

Cardoso-Bihlo et al. (2019) was able to use a Bayesian inference technique for an MJO case study but the sheer computational complexity did not allow a comprehensive study based on the whole observational data set of 5 months. Instead, three sets of transition timescales are obtained for three short data spans of 1–3 days corresponding to a suppressed, initiation, and active MJO periods. The refinement of the Bayesian inference methodology based on a variational inference to replace the Markov Chain Monte Carlo sampling which is currently used combined with an optimization of the parallelization technique may speed up the code and allow the inference of more comprehensive and perhaps lead to universal transition timescales for the SMCM.

Finally, as discussed in Section 4.4, given that the SMCM is based on a Markovian process with a slowly evolving equilibrium distribution, as the large-scale predictors vary in both time and space, the effect of random sampling does not affect the physics of the model. The set of 10 ensemble members chosen randomly based on the random choice of the seed of the random number generator, showed comparable results in terms of the overall performance of the SMPCP. Nonetheless, the stochasticity induces somewhat significant spread especially in areas where the model error is large. This is in essence consistent with the idea that stochastic parameterizations could help alleviate the model error problem in a probabilistic weather forecasting and climate predictions by adding spread into the predictive ensembles in areas where the model errors are the greatest (Berner et al., 2016; Palmer, 2001). The SMCM is a good candidate to accomplish this goal in a way that does not compromise the desired physical features sought by the parameterization.

## Data Availability Statement

The data used in this study are available through the CESM data repository (Hack et al., 2014; Phillips et al., 2020) and the software (the single column CAM model using the SMPCP) can be downloaded from Khouider (2023).

## References

- Abhik, S., Krishna, R. P. M., Mahakur, M., Ganai, M., Mukhopadhyay, P., & Dudhia, J. (2017). Revised cloud processes to improve the mean and intraseasonal variability of indian summer monsoon in climate forecast system: Part 1. *Journal of Advances in Modeling Earth Systems*, 9(2), 1002–1029. <https://doi.org/10.1002/2016MS000819>
- Ajayamohan, R. S., Khouider, B., Majda, A. J., & Deng, Q. (2016). Role of stratiform heating on the organization of convection over the monsoon trough. *Climate Dynamics*, 47(12), 1–20. <https://doi.org/10.1007/s00382-016-3033-7>
- Arakawa, A. (2004). The cumulus parameterization problem: Past, present, and future. *Journal of Climate*, 17(13), 2493–2525. [https://doi.org/10.1175/1520-0442\(2004\)017<2493:RATCPP>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<2493:RATCPP>2.0.CO;2)
- Arakawa, A., & Schubert, W. H. (1974). Interaction of a cumulus cloud ensemble with large-scale environment, part I. *Journal of the Atmospheric Sciences*, 31(3), 674–701. [https://doi.org/10.1175/1520-0469\(1974\)031<0674:ioacce>2.0.co;2](https://doi.org/10.1175/1520-0469(1974)031<0674:ioacce>2.0.co;2)

## Acknowledgments

The research of B.K. is supported in part by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (RGPIN-04246-2020). This research was conducted during the visits of P.M. Krishna to the Center for Prototype Climate Models at NYU Abu Dhabi and University of Victoria from November 2018 to June 2019 and July 2019 and October 2019, respectively. The authors are very grateful to the three anonymous reviewers who provided very thoughtful and constructive comments during the review process that helped greatly improve and shape the final version of the manuscript.



- Bengtsson, L., & Karnich, H. (2016). Impact of a stochastic parametrization of cumulus convection, using cellular automata, in a mesoscale ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, *142*(695), 1150–1159. <https://doi.org/10.1002/qj.2720>
- Berner, J., Achatz, U., Batta, L., Bengtsson, L., CaMara, A. D. L., Christensen, H. M., et al. (2016). Stochastic parameterization: Towards a new view of weather and climate models. *Bulletin of the American Meteorological Society*, *0*(0), Null. <https://doi.org/10.1175/BAMS-D-15-00268.1>
- Bony, S., & Dufresne, J. (2005). Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophysical Research Letters*, *32*(20), 20806. <https://doi.org/10.1029/2005gl023851>
- Bony, S., Stevens, B., Frierson, D. M. W., Jakob, C., Kageyama, M., Pincus, R., et al. (2015). Clouds, circulation and climate sensitivity. *Nature Geoscience*, *8*(4), 261–268. <https://doi.org/10.1038/ngeo2398>
- Brenowitz, N. D., & Bretherton, C. S. (2018). Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters*, *45*(12), 6289–6298. <https://doi.org/10.1029/2018gl078510>
- Bretherton, C. S., & Park, S. (2008). A new bulk shallow-cumulus model and implications for penetrative entrainment feedback on updraft buoyancy. *Journal of the Atmospheric Sciences*, *65*(7), 2174–2193. <https://doi.org/10.1175/2007JAS2242.1>
- Brient, F., & Bony, S. (2013). Interpretation of the positive low-cloud feedback predicted by a climate model under global warming. *Climate Dynamics*, *40*(9–10), 2415–2431. <https://doi.org/10.1007/s00382-011-1279-7>
- Buizza, R., Miller, M., & Palmer, T. N. (1999). Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, *125*(560), 2887–2908. <https://doi.org/10.1002/qj.49712556006>
- Cardoso-Bihlo, E., Khouider, B., Schumacher, C., & Chevrotière, M. D. L. (2019). Using radar data to calibrate a stochastic parametrization of organized convection. *Journal of Advances in Modeling Earth Systems*, *11*(6), 1655–1684. <https://doi.org/10.1029/2018ms001537>
- Casey, S. P. F., Dessler, A. E., & Schumacher, C. (2007). Frequency of tropical precipitating clouds as observed by the Tropical Rainfall Measuring Mission Precipitation Radar and ICESat/Geoscience Laser Altimeter System. *Journal of Geophysical Research*, *112*(D11), 14215+. <https://doi.org/10.1029/2007jd008468>
- Ceppi, P., Brient, F., Zelinka, M. D., & Hartmann, D. (2017). Cloud feedback mechanisms and their representation in global climate models. *Wiley Interdisciplinary Reviews-Climate Change*, *8*(4), 21. <https://doi.org/10.1002/wcc.465>
- Chattopadhyay, R., Goswami, B. N., Sahai, A. K., & Fraedrich, K. (2009). Role of stratiform rainfall in modifying the northward propagation of monsoon intraseasonal oscillation. *Journal of Geophysical Research*, *114*(D19), n/a. <https://doi.org/10.1029/2009JD011869>
- Choudhury, A. D., & Krishnan, R. (2011). Dynamical response of the south Asian monsoon trough to latent heating from stratiform and convective precipitation. *Journal of the Atmospheric Sciences*, *68*(6), 1347–1363. <https://doi.org/10.1175/2011JAS3705.1>
- Craig, G., & Cohen, B. (2006). Fluctuations in an equilibrium convective ensemble. Part I: Theoretical formulation. *Journal of the Atmospheric Sciences*, *63*(8), 1996–2004. <https://doi.org/10.1175/jas3709.1>
- Dai, A. (2006). Precipitation characteristics in 18 coupled climate models. *Journal of Climate*, *19*, 4605–4630. <https://doi.org/10.1175/jcli3884.1>
- D'Andrea, F., Gentine, P., Betts, A. K., & Lintner, B. R. (2014). Triggering deep convection with a probabilistic plume model. *Journal of the Atmospheric Sciences*, *71*(11), 3881–3901. <https://doi.org/10.1175/JAS-D-13-0340.1>
- De La Chevrotiere, M., Khouider, B., & Majda, A. (2014). Calibration of the stochastic multicloud model using Bayesian inference. *SIAM Journal on Scientific Computing*, *36*(3), B538–B560. <https://doi.org/10.1137/13094267x>
- De La Chevrotière, M., Khouider, B., & Majda, A. (2015). Stochasticity of convection in Giga-LES data. *Climate Dynamics*, *47*(5), 1845–1861. <https://doi.org/10.1007/s00382-015-2936-z>
- Deng, Q., Khouider, B., & Majda, A. J. (2015). The MJO in a coarse-resolution GCM with a stochastic multicloud parameterization. *Journal of the Atmospheric Sciences*, *72*(1), 55–74. <https://doi.org/10.1175/JAS-D-14-0120.1>
- Deng, Q., Khouider, B., Majda, A. J., & Ajayamohan, R. S. (2016). Effect of stratiform heating on the planetary-scale organization of tropical convection. *Journal of the Atmospheric Sciences*, *73*(1), 371–392. <https://doi.org/10.1175/JAS-D-15-0178.1>
- Derbyshire, S. H., Beau, I., Bechtold, P., Grandpeix, J.-Y., Piriou, J.-M., Redelsperger, J.-L., & Soares, P. M. M. (2004). Sensitivity of moist convection to environmental humidity. *Quarterly Journal of the Royal Meteorological Society*, *130*(604), 3055–3080. <https://doi.org/10.1256/qj.03.130>
- Dorrestijn, J., Crommelin, D., Biello, J., & Böing, S. (2013). A data-driven multi-cloud model for stochastic parametrization of deep convection. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *371*(1991), 20120374. <https://doi.org/10.1098/rsta.2012.0374>
- Dorrestijn, J., Crommelin, D. T., Siebesma, A. P., Jonker, H. J., & Jakob, C. (2015). Stochastic parameterization of convective area fractions with a multicloud model inferred from observational data. *Journal of the Atmospheric Sciences*, *72*(2), 854–869. <https://doi.org/10.1175/jas-d-14-0110.1>
- Dorrestijn, J., Crommelin, D. T., Siebesma, A. P., Jonker, H. J., & Selten, F. (2016). Stochastic convection parameterization with Markov chains in an intermediate-complexity GCM. *Journal of the Atmospheric Sciences*, *73*(3), 1367–1382. <https://doi.org/10.1175/jas-d-15-0244.1>
- Feng, Z., McFarlane, S. A., Schumacher, C., Ellis, S., & Bharadwaj, N. (2014). Constructing a merged cloud-precipitation radar dataset for tropical convective clouds during the DYNAMO/AMIE experiment at Addu Atoll. *Journal of Atmospheric and Oceanic Technology*, *31*(5), 1021–1042. <https://doi.org/10.1175/jtech-d-13-00132.1>
- Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., et al. (2013). *Evaluation of climate models*. Cambridge University Press.
- Frenkel, Y., Majda, A. J., & Khouider, B. (2012). Using the stochastic multicloud model to improve tropical convective parameterization: A paradigm example. *Journal of the Atmospheric Sciences*, *69*(3), 1080–1105. <https://doi.org/10.1175/jas-d-11-0148.1>
- Ganai, M., Krishna, R. P. M., Snehlata, T., Mukhopadhyay, P., Mahakur, M., & Han, J.-Y. (2019). The impact of modified fractional cloud condensate to precipitation conversion parameter in revised simplified Arakawa-Schubert convection parameterization scheme on the simulation of Indian summer monsoon and its forecast application on an extreme rainfall event over Mumbai. *Journal of Geophysical Research: Atmospheres*, *124*(10), 5379–5399. <https://doi.org/10.1029/2019JD030278>
- Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., & Yacalis, G. (2018). Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, *45*(11), 5742–5751. <https://doi.org/10.1029/2018gl078202>
- Gottelman, A., Truesdale, J. E., Bacmeister, J. T., Caldwell, P. M., Neale, R. B., Bogenschutz, P. A., & Simpson, I. R. (2019). The Single Column Atmosphere Model Version 6 (SCAM6): Not a scam but a tool for model evaluation and development. *Journal of Advances in Modeling Earth Systems*, *11*(5), 1381–1401. <https://doi.org/10.1029/2018MS001578>
- Goswami, B. B., Khouider, B., Krishna, R. P. M., Mukhopadhyay, P., & Majda, A. J. (2017a). Implementation and calibration of a stochastic convective parameterization in the NCEP Climate Forecast System. *Journal of Advances in Modeling Earth Systems*, *9*(3), 1721–1739. <https://doi.org/10.1002/2017ms001014>
- Goswami, B. B., Khouider, B., Krishna, R. P. M., Mukhopadhyay, P., & Majda, A. J. (2017b). Improved tropical modes of variability in the NCEP Climate Forecast System (version 2) via a stochastic multicloud model. *Journal of the Atmospheric Sciences*, *74*(10), 3339–3366. <https://doi.org/10.1175/JAS-D-17-0113.1>

- Goswami, B. B., Khouider, B., Krishna, R. P. M., Mukhopadhyay, P., & Majda, A. J. (2017c). Improving synoptic and intraseasonal variability in CFSv2 via stochastic representation of organized convection. *Geophysical Research Letters*, *44*(2), 1104–1113. <https://doi.org/10.1002/2016GL071542>
- Gottwald, G. A., Peters, K., & Davies, L. (2016). A data-driven method for the stochastic parametrisation of subgrid-scale tropical convective area fraction. *Quarterly Journal of the Royal Meteorological Society*, *142*(694), 349–359. <https://doi.org/10.1002/qj.2655>
- Grabowski, W. W. (2016). Towards global large eddy simulation: Superparameterization revisited. *Journal of the Meteorological Society of Japan*, *94*(4), 327–344. <https://doi.org/10.2151/jmsj.2016-017>
- Hack, J. J., Truesdale, J. E., Pedretti, J. A., & Petch, J. C. (2014). “SCAM” user’s guide [Dataset]. Retrieved from <https://svn-ccsm-inputdata.cgd.ucar.edu/trunk/inputdata/atm/cam/scam/iop/>
- Hagos, S., Feng, Z., Plant, R. S., Houze, R. A., Jr., & Xiao, H. (2018). A stochastic framework for modeling the population dynamics of convective clouds. *Journal of Advances in Modeling Earth Systems*, *10*(2), 448–465. <https://doi.org/10.1002/2017MS001214>
- Hohenegger, C., & Stevens, B. (2013). Preconditioning deep convection with cumulus congestus. *Journal of the Atmospheric Sciences*, *70*(2), 448–464. <https://doi.org/10.1175/jas-d-12-089.1>
- Houze, R. A., Jr. (2004). Mesoscale convective systems. *Reviews of Geophysics*, *42*(4), G4003+. <https://doi.org/10.1029/2004RG000150>
- Hung, M.-P., Lin, J.-L., Wang, W., Kim, D., Shinoda, T., & Weaver, S. J. (2013). MJO and convectively coupled equatorial waves simulated by CMIP5 climate models. *Journal of Climate*, *26*(17), 6185–6214. <https://doi.org/10.1175/JCLI-D-12-00541.1>
- IPCC. (2021). *Climate change 2021: The physical science basis. Contribution of Working Group I to the sixth assessment report of the Intergovernmental Panel on Climate Change* (Note: The landing pages for the digital object identifier (DOI) numbers shown below are not yet active. Those pages will be hosted by Cambridge University Press and will be accessible once the report is printed and available in open access on its website.). Cambridge University Press. Retrieved from <https://www.ipcc.ch/report/ar6/wg1/>
- Johnson, R. H., Ciesielski, P. E., Ruppert, J. H., & Katsumata, M. (2015). Sounding-based thermodynamic budgets for dynamo. *Journal of the Atmospheric Sciences*, *72*(2), 598–622. <https://doi.org/10.1175/JAS-D-14-0202.1>
- Johnson, R. H., Rickenbach, T. M., Rutledge, S. A., Ciesielski, P. E., & Schubert, W. H. (1999). Trimodal characteristics of tropical convection. *Journal of Climate*, *12*(8), 2397–2418. [https://doi.org/10.1175/1520-0442\(1999\)012<2397:tcotc>2.0.co;2](https://doi.org/10.1175/1520-0442(1999)012<2397:tcotc>2.0.co;2)
- Khairoutdinov, M. F., Krueger, S. K., Moeng, C.-H., Bogenschutz, P. A., & Randall, D. A. (2009). Large-eddy simulation of maritime deep tropical convection. *Journal of Advances in Modeling Earth Systems*, *1*(15), 13. <https://doi.org/10.3894/JAMES.2009.1.15>
- Khouider, B. (2014). A coarse grained stochastic multi-type particle interacting model for tropical convection: Nearest neighbour interactions. *Communications in Mathematical Sciences*, *12*(8), 1379–1407. <https://doi.org/10.4310/cms.2014.v12.n8.a1>
- Khouider, B. (2019). *Models for tropical climate dynamics: Waves, clouds, and precipitation*. Springer.
- Khouider, B. (2023). SMCM SINGLE COLUMN CAM [Software]. Borealis. <https://doi.org/10.5683/SP3/JBK0JD>
- Khouider, B., Biello, J., & Majda, A. J. (2010). A stochastic multicloud model for tropical convection. *Communications in Mathematical Sciences*, *8*(1), 187–216. <https://doi.org/10.4310/cms.2010.v8.n1.a10>
- Khouider, B., & Leclerc, E. (2019). Toward a stochastic relaxation for the QE theory of cumulus parameterization. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2474–2502. <https://doi.org/10.1029/2019MS001627>
- Khouider, B., Majda, A., & Katsoulakis, M. (2003). Coarse grained stochastic models for tropical convection and climate. *Proceedings of the National Academy of Sciences*, *100*(21), 11941–11946. <https://doi.org/10.1073/pnas.1634951100>
- Khouider, B., & Majda, A. J. (2006). A simple multicloud parameterization for convectively coupled tropical waves. Part I: Linear analysis. *Journal of the Atmospheric Sciences*, *63*(4), 1308–1323. <https://doi.org/10.1175/jas3677.1>
- Khouider, B., Majda, A. J., & Stechmann, S. N. (2013). Climate science in the tropics: Waves, vortices and PDEs. *Nonlinearity*, *26*(1), R1–R68. <https://doi.org/10.1088/0951-7715/26/1/r1>
- Khouider, B., St-Cyr, A., Majda, A. J., & Tribbia, J. (2011). The MJO and convectively coupled waves in a coarse resolution GCM with a simple multicloud parameterization. *Journal of the Atmospheric Sciences*, *68*(2), 240–264. <https://doi.org/10.1175/2010JAS3443.1>
- Kiladis, G. N., Wheeler, M. C., Haertel, P. T., Straub, K. H., & Roundy, P. E. (2009). Convectively coupled equatorial waves. *Reviews of Geophysics*, *47*(2), RG2003. <https://doi.org/10.1029/2008RG000266>
- Kuang, Z., & Bretherton, C. (2006). A mass-flux scheme view of a high-resolution simulation of a transition from shallow to deep cumulus convection. *Journal of the Atmospheric Sciences*, *63*(7), 1895–1909. <https://doi.org/10.1175/jas3723.1>
- Kumar, S., Arora, A., Chattopadhyay, R., Hazra, A., Rao, S. A., & Goswami, B. N. (2017). Seminal role of stratiform clouds in large-scale aggregation of tropical rain in boreal summer monsoon intraseasonal oscillations. *Climate Dynamics*, *48*(3), 999–1015. <https://doi.org/10.1007/s00382-016-3124-5>
- Lappen, C.-L., & Schumacher, C. (2014). The role of tilted heating in the evolution of the MJO. *Journal of Geophysical Research: Atmospheres*, *119*(6), 2966–2989. <https://doi.org/10.1002/2013JD020638>
- Lin, J., & Neelin, J. D. (2000). Influence of a stochastic moist convective parameterization on tropical climate variability. *Geophysical Research Letters*, *27*(22), 3691–3694. <https://doi.org/10.1029/2000gl011964>
- Lin, J., & Neelin, J. D. (2003). Toward stochastic deep convective parameterization in general circulation models. *Geophysical Research Letters*, *30*(4), 1162. <https://doi.org/10.1029/2002gl016203>
- Lin, J.-L., Mapes, B., Zhang, M., & Newman, M. (2004). Stratiform precipitation, vertical heating profiles, and the Madden-Julian oscillation. *Journal of the Atmospheric Sciences*, *61*(3), 296–309. [https://doi.org/10.1175/1520-0469\(2004\)061<0296:spvhpa>2.0.co;2](https://doi.org/10.1175/1520-0469(2004)061<0296:spvhpa>2.0.co;2)
- Majda, A. J. (2007a). Multiscale models with moisture and systematic strategies for superparameterization. *Journal of the Atmospheric Sciences*, *64*(7), 2726–2734. <https://doi.org/10.1175/jas3976.1>
- Majda, A. J. (2007b). New multi-scale models and self-similarity in tropical convection. *Journal of the Atmospheric Sciences*, *64*(4), 1393–1404. <https://doi.org/10.1175/jas3880.1>
- Majda, A. J., & Khouider, B. (2002). Stochastic and mesoscopic models for tropical convection. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(3), 1123–1128. <https://doi.org/10.1073/pnas.032663199>
- Majda, A. J., & Shefter, M. (2001). Models for stratiform instability and convectively coupled waves. *Journal of the Atmospheric Sciences*, *58*(12), 1567–1584. [https://doi.org/10.1175/1520-0469\(2001\)058<1567:mfsiac>2.0.co;2](https://doi.org/10.1175/1520-0469(2001)058<1567:mfsiac>2.0.co;2)
- Mapes, B., Tulich, S., Lin, J., & Zuidema, P. (2006). The mesoscale convection life cycle: Building block or prototype for large-scale tropical waves? *Dynamics of Atmospheres and Oceans*, *42*(1–4), 3–29. <https://doi.org/10.1016/j.dynatmoce.2006.03.003>
- Mapes, B. E. (1993). Gregarious tropical convection. *Journal of the Atmospheric Sciences*, *50*(13), 2026–2037. [https://doi.org/10.1175/1520-0469\(1993\)050<2026:gtc>2.0.co;2](https://doi.org/10.1175/1520-0469(1993)050<2026:gtc>2.0.co;2)
- Mapes, B. E. (2000). Convective inhibition, subgridscale triggering energy, and “stratiform instability” in a toy tropical wave model. *Journal of the Atmospheric Sciences*, *57*(10), 1515–1535. [https://doi.org/10.1175/1520-0469\(2000\)057<1515:sisste>2.0.co;2](https://doi.org/10.1175/1520-0469(2000)057<1515:sisste>2.0.co;2)

- Mapes, B. E., & Houze, R. A., Jr. (1993). Cloud clusters and superclusters over the oceanic warm pool. *Monthly Weather Review*, *121*(5), 1398–1416. [https://doi.org/10.1175/1520-0493\(1993\)121<1398:ccasot>2.0.co;2](https://doi.org/10.1175/1520-0493(1993)121<1398:ccasot>2.0.co;2)
- May, P. T., Mather, J. H., Vaughan, G., Jakob, C., McFarquhar, G. M., Bower, K. N., & Mace, G. G. (2008). The tropical warm pool international cloud experiment. *Bulletin of the American Meteorological Society*, *89*(5), 629–646. <https://doi.org/10.1175/BAMS-89-5-629>
- Meehl, G. A., Zwiers, F., Evans, J., Knutson, T., Mearns, L., & Whetton, P. (2000). Trend in extreme weather and climate events: Issues related to modelling extremes in projections of future climate change. *Bulletin America Meteorology Social*, *81*(3), 427–436. [https://doi.org/10.1175/1520-0477\(2000\)081<0427:tiewac>2.3.co;2](https://doi.org/10.1175/1520-0477(2000)081<0427:tiewac>2.3.co;2)
- Moncrieff, M. W. (2010). The multiscale organization of moist convection and the intersection of weather and climate. Why does climate vary? *Geophysical Monograph Series*, *189*, 3–26. <https://doi.org/10.1029/2008GM000838>
- Moncrieff, M. W., & Klinker, E. (1997). Organized convective systems in the tropical western Pacific as a process in general circulation models: A TOGA COARE case-study. *Quarterly Journal of the Royal Meteorological Society*, *123*(540), 805–827. <https://doi.org/10.1002/qj.49712354002>
- Moncrieff, M. W., Liu, C., & Bogenschutz, P. (2017). Simulation, modeling, and dynamically based parameterization of organized tropical convection for global climate models. *Journal of the Atmospheric Sciences*, *74*(5), 1363–1380. <https://doi.org/10.1175/JAS-D-16-0166.1>
- Murakami, T., Chen, L. X., & Xie, A. (1986). Relationship among seasonal cycles, low frequency oscillations and transient disturbances as revealed from outgoing long wave radiation. *Monthly Weather Review*, *114*(8), 1456–1465. [https://doi.org/10.1175/1520-0493\(1986\)114<1456:rasclf>2.0.co;2](https://doi.org/10.1175/1520-0493(1986)114<1456:rasclf>2.0.co;2)
- Nakazawa, T. (1988). Tropical super clusters within intraseasonal variations over the western Pacific. *Journal of the Meteorological Society of Japan*, *66*(6), 823–839. [https://doi.org/10.2151/jmsj1965.66.6\\_823](https://doi.org/10.2151/jmsj1965.66.6_823)
- Neale, R. B., Chen, C. C., Gettelman, A., Lauritzen, P. H., Park, S., Williamson, D. L., et al. (2010). *Description of the NCAR community atmosphere model (CAM5.0)* (Tech. Rep. NCAR/TN-486+STR ed.; Tech. Rep.). National Center for Atmospheric Research.
- Palmer, T. N. (2001). A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quarterly Journal of the Royal Meteorological Society*, *127*(572), 279–304. <https://doi.org/10.1002/qj.49712757202>
- Palmer, T. N. (2022). Stochastic weather and climate models. *Nature Reviews: Physics*, *1*(7), 463–471. <https://doi.org/10.1038/s42254-019-0062-2>
- Pan, D.-M., & Randall, D. D. A. (1998). A cumulus parameterization with a prognostic closure. *Quarterly Journal of the Royal Meteorological Society*, *124*(547), 949–981. <https://doi.org/10.1002/qj.49712454714>
- Park, S. (2014a). A unified convection scheme (UNICON). Part I: Formulation. *Journal of the Atmospheric Sciences*, *71*(11), 3902–3930. <https://doi.org/10.1175/JAS-D-13-0233.1>
- Park, S. (2014b). A unified convection scheme (UNICON). PART II: Simulation. *Journal of the Atmospheric Sciences*, *71*(11), 3931–3973. <https://doi.org/10.1175/JAS-D-13-0234.1>
- Parker, M. D., & Johnson, R. H. (2004). Structures and dynamics of quasi-2D mesoscale convective systems. *Journal of the Atmospheric Sciences*, *61*(5), 545–567. [https://doi.org/10.1175/1520-0469\(2004\)061<0545:sadoqm>2.0.co;2](https://doi.org/10.1175/1520-0469(2004)061<0545:sadoqm>2.0.co;2)
- Peters, J. M., Morrison, H., Zhang, G. J., & Powell, S. W. (n.d.). Improving the physical basis for updraft dynamics in deep convection parameterizations. *Journal of Advances in Modeling Earth Systems*, *13*(2), e2020MS002282. <https://doi.org/10.1029/2020MS002282>
- Peters, K., Cruieger, T., Jakob, C., & Möbis, B. (2017). Improved MJO-simulation in ECHAM6.3 by coupling a stochastic multicloud model to the convection scheme. *Journal of Advances in Modeling Earth Systems*, *9*(1), 193–219. <https://doi.org/10.1002/2016MS000809>
- Peters, K., Jakob, C., Davies, L., Khouider, B., & Majda, A. (2013). Stochastic behavior of tropical convection in observations and a multicloud model. *Journal of the Atmospheric Sciences*, *70*(11), 3556–3575. <https://doi.org/10.1175/JAS-D-13-031.1>
- Phillips, A. S., Deser, C., Fasullo, J., Schneider, D. P., & Simpson, I. R. (2020). Assessing climate variability and change in model large ensembles: A user's guide to the “climate variability diagnostics package for large ensembles” version 1.0 [Dataset]. <https://doi.org/10.5065/h7c7-f961>
- Plant, R. S., & Craig, G. C. (2008). A stochastic parameterization for deep convection based on equilibrium statistics. *Journal of the Atmospheric Sciences*, *65*(1), 87–105. <https://doi.org/10.1175/2007jas2263.1>
- Randall, D., Khairoutdinov, M., Arakawa, A., & Grabowski, W. (2003). Breaking the cloud parameterization deadlock. *Bulletin of the American Meteorological Society*, *84*(11), 1547–1564. <https://doi.org/10.1175/bams-84-11-1547>
- Randall, D. A., Xu, K.-M., Somerville, R. J. C., & Iacobellis, S. (1996). Single-column models and cloud ensemble models as links between observations and climate models. *Journal of Climate*, *9*(8), 1683–1697. [https://doi.org/10.1175/1520-0442\(1996\)009<1683:scmace>2.0.co;2](https://doi.org/10.1175/1520-0442(1996)009<1683:scmace>2.0.co;2)
- Rio, C., Del Genio, A., & Hourdin, F. (2019). Ongoing breakthroughs in convective parameterization. *Current Climate Change Reports*, *5*(2), 95–111. <https://doi.org/10.1007/s40641-019-00127-w>
- Sabeerali, C. T., Dandi, A. R., Dhakate, A., Salunke, K., Mahapatra, S., & Rao, S. A. (2013). Simulation of boreal summer intraseasonal oscillations in the latest CMIP5 coupled GCMs. *Journal of Geophysical Research: Atmospheres*, *118*(10), 4401–4420. <https://doi.org/10.1002/jgrd.50403>
- Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., et al. (2014). The NCEP Climate Forecast System Version 2. *Journal of Climate*, *27*(6), 2185–2208. <https://doi.org/10.1175/JCLI-D-12-00823.1>
- Schneider, T., Kaul, C. M., & Pressel, K. G. (2019). Possible climate transitions from breakup of stratocumulus decks under greenhouse warming. *Nature Geoscience*, *12*(3), 163–167. <https://doi.org/10.1038/s41561-019-0310-1>
- Schneider, T., Lan, S., Stuart, A., & Teixeira, J. (2017). Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophysical Research Letters*, *44*(12), 396–412. <https://doi.org/10.1002/2017gl076101>
- Schumacher, C., & Houze, R. (2003). Stratiform rain in the tropics as seen by the TRMM precipitation radar. *Journal of Climate*, *16*(11), 1739–1756. [https://doi.org/10.1175/1520-0442\(2003\)016<1739:sritta>2.0.co;2](https://doi.org/10.1175/1520-0442(2003)016<1739:sritta>2.0.co;2)
- Schumacher, C., Zhang, M. H., & Ciesielski, P. E. (2007). Heating structures of the TRMM field campaigns. *Journal of the Atmospheric Sciences*, *64*(7), 2593–2610. <https://doi.org/10.1175/JAS3938.1>
- Stachnik, J. P., Schumacher, C., & Ciesielski, P. E. (2013). Total heating characteristics of the ISCCP tropical and subtropical cloud regimes. *Journal of Climate*, *26*(18), 7097–7116. <https://doi.org/10.1175/JCLI-D-12-00673.1>
- Stensrud, D. J. (2007). *Parameterization schemes: Keys to understanding numerical weather prediction models*. Cambridge University Press.
- Suselj, K., Kurowski, M. J., & Teixeira, J. (2019a). On the factors controlling development of shallow convection in eddy-diffusivity/mass-flux models. *Journal of the Atmospheric Sciences*, *76*(2), 433–456. <https://doi.org/10.1175/JAS-D-18-0121.1>
- Suselj, K., Kurowski, M. J., & Teixeira, J. (2019b). A unified eddy-diffusivity/mass-flux approach for modeling atmospheric convection. *Journal of the Atmospheric Sciences*, *76*(8), 2505–2537. <https://doi.org/10.1175/JAS-D-18-0239.1>
- Tao, W.-K., Lang, S., Zeng, X., Shige, S., & Takayabu, Y. (2010). Relating convective and stratiform rain to latent heating. *Journal of Climate*, *23*(7), 1874–1893. <https://doi.org/10.1175/2009JCLI3278.1>

- Tiedke, M. (1984). The sensitivity of the time mean large-scale flow to cumulus convection in the ECMWF model. In *Workshop on convection in large-scale numerical models, ECMWF* (pp. 297–317).
- Tiedtke, M. (1993). Representation of clouds in large-scale models. *Monthly Weather Review*, *121*(11), 3040–3061. [https://doi.org/10.1175/1520-0493\(1993\)121<3040:rocils>2.0.co;2](https://doi.org/10.1175/1520-0493(1993)121<3040:rocils>2.0.co;2)
- Tokay, A., Short, D. A., Williams, C. R., Ecklund, W. L., & Gage, K. S. (1999). Tropical rainfall associated with convective and stratiform clouds: Intercomparison of disdrometer and profiler measurements. *Journal of Applied Meteorology*, *38*(3), 302–320. [https://doi.org/10.1175/1520-0450\(1999\)038<0302:trawca>2.0.co;2](https://doi.org/10.1175/1520-0450(1999)038<0302:trawca>2.0.co;2)
- Tompkins, A. M. (2002). A prognostic parameterization for the subgridscale variability of water vapor and clouds in large-scale models and its use to diagnose cloud cover. *Journal of the Atmospheric Sciences*, *59*(12), 1917–1942. [https://doi.org/10.1175/1520-0469\(2002\)059<1917:appfts>2.0.co;2](https://doi.org/10.1175/1520-0469(2002)059<1917:appfts>2.0.co;2)
- Waite, M., & Khouider, B. (2010). The deepening of tropical convection by congestus preconditioning. *Journal of the Atmospheric Sciences*, *67*(8), 2601–2615. <https://doi.org/10.1175/2010jas3357.1>
- Wang, M., & Zhang, G. J. (2018). Improving the simulation of tropical convective cloud-top heights in CAM5 with CloudSat observations. *Journal of Climate*, *31*(13), 5189–5204. <https://doi.org/10.1175/JCLI-D-18-0027.1>
- Wang, Y., & Zhang, G. J. (2016). Global climate impacts of stochastic deep convection parameterization in the NCAR CAM5. *Journal of Advances in Modeling Earth Systems*, *8*(4), 1641–1656. <https://doi.org/10.1002/2016MS000756>
- Wang, Y., Zhang, G. J., & Craig, G. C. (2016). Stochastic convective parameterization improving the simulation of tropical precipitation variability in the NCAR CAM5. *Geophysical Research Letters*, *43*(12), 6612–6619. <https://doi.org/10.1002/2016GL069818>
- Wu, X., & Moncrieff, M. W. (1996). Collective effects of organized convection and their approximation in general circulation models. *Journal of the Atmospheric Sciences*, *53*(10), 1477–1495. [https://doi.org/10.1175/1520-0469\(1996\)053<1477:ceocca>2.0.co;2](https://doi.org/10.1175/1520-0469(1996)053<1477:ceocca>2.0.co;2)
- Xu, K.-M., & Randall, D. A. (2000). Explicit simulation of midlatitude cumulus ensembles: Comparison with ARM data. *Journal of the Atmospheric Sciences*, *57*(17), 2839–2858. [https://doi.org/10.1175/1520-0469\(2000\)057<2839:ESOMCE;2.0.CO;2](https://doi.org/10.1175/1520-0469(2000)057<2839:ESOMCE;2.0.CO;2)
- Yanai, M., Esbensen, S., & Chu, J. H. (1973). Determination of bulk properties of tropical cloud clusters from large-scale heat and moisture budgets. *Journal of the Atmospheric Sciences*, *30*(4), 611–627. [https://doi.org/10.1175/1520-0469\(1973\)030<0611:dobpot>2.0.co;2](https://doi.org/10.1175/1520-0469(1973)030<0611:dobpot>2.0.co;2)
- Zermelo-Daz, D. M., Zhang, C., Kollias, P., & Kalesse, H. (2015). The role of shallow cloud moistening in MJO and non-MJO convective events over the ARM Manus site. *Journal of the Atmospheric Sciences*, *72*(12), 4797–4820. <https://doi.org/10.1175/JAS-D-14-0322.1>
- Zhang, C. (2005). Madden–Julian Oscillation. *Reviews of Geophysics*, *43*(2), G2003+. <https://doi.org/10.1029/2004RG000158>
- Zhang, C. (2013). Madden-Julian Oscillation: Bridging weather and climate. *Bulletin of the American Meteorological Society*, *94*(12), 1849–1870. <https://doi.org/10.1175/BAMS-D-12-00026.1>
- Zhang, G. J., & McFarlane, N. A. (1995). Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Center general circulation model. *Atmosphere-Ocean*, *33*(3), 407–446. <https://doi.org/10.1080/07055900.1995.9649539>
- Zhang, M., Somerville, R. C. J., & Xie, S. (2016). The SCM concept and creation of ARM forcing datasets. *Meteorological Monographs*, *57*, 241–2412. <https://doi.org/10.1175/AMSMONOGRAPHIS-D-15-0040.1>