UvA-DARE (Digital Academic Repository)

Stochastic Convection Parameterization with Markov Chains in an Intermediate-Complexity GCM

Dorrestijn, J.; Crommelin, D.T.; Siebesma, A.P.; Jonker, H.J.J. ; Selten, F.

Stochastic Convection Parameterization with Markov Chains in an Intermediate-Complexity GCM

JESSE DORRESTIJN

CWI, Amsterdam, Netherlands

DAAN T. CROMMELIN

CWI, and Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Amsterdam, Netherlands

A. PIER SIEBESMA

KNMI, De Bilt, and Delft University of Technology, Delft, Netherlands

HARMEN J. J. JONKER

Delft University of Technology, Delft, Netherlands

FRANK SELTEN

KNMI, De Bilt, Netherlands

ABSTRACT

Conditional Markov chain (CMC) models have proven to be promising building blocks for stochastic convection parameterizations. In this paper, it is demonstrated how two different CMC models can be used as mass flux closures in convection parameterizations. More specifically, the CMC models provide a stochastic estimate of the convective area fraction that is directly proportional to the cloud-base mass flux. Since, in one of the models, the number of CMCs decreases with increasing resolution, this approach makes convection parameterizations scale aware and introduces stochastic fluctuations that increase with resolution in a realistic way. Both CMC models are implemented in a GCM of intermediate complexity. It is shown that with the CMC models, trained with observational data, it is possible to improve both the subgrid-scale variability and the autocorrelation function of the cloud-base mass flux as well as the distribution of the daily accumulated precipitation in the tropics. Hovmöller diagrams and wavenumber–frequency diagrams of the equatorial precipitation indicate that, in this specific GCM, convectively coupled equatorial waves are more sensitive to the mean cloud-base mass flux than to stochastic fluctuations. A smaller mean mass flux tends to increase the power of the simulated MJO and to diminish equatorial Kelvin waves.

1. Introduction

Deep convection is an atmospheric process of major importance in Earth's weather and climate system. Locally, it transports heat, moisture, and momentum vertically in the atmosphere (Arakawa 2004). Globally, it affects the large-scale circulation (Randall et al. 1989). Further, deep convection largely determines precipitation in the tropics. Of specific interest is its coupling to equatorial waves (e.g., equatorial Kelvin waves, Rossby waves, and the MJO) that largely determine the variability of precipitation (Kiladis et al. 2009; Wheeler and Kiladis 1999). Most GCMs do not resolve deep convection. Instead, this process is represented by parameterizations, assuming, for example, a cumulus ensemble that is in quasi equilibrium with the large-scale forcing (Arakawa and Schubert 1974).

Availability of larger computational resources allows GCMs to be run at finer resolutions. At horizontal grid resolutions below ~100 km, and especially in the Grey

Corresponding author address: Jesse Dorrestijn, CWI, P.O. Box 94079, Amsterdam 1090 GB, Netherlands.
E-mail: j.dorrestijn@cwi.nl

Zone (1–10 km), where convection becomes partially resolved, the quasi-equilibrium assumption breaks down. As a result, the assumption that there is a unique relation between the cumulus ensemble and the large-scale conditions is not reasonable anymore, because the ensemble in the GCM grid column is too small, and life cycles of individual cumulus clouds cause large fluctuations in the convective response and associated subgrid fluxes. Therefore, convection parameterizations should become scale aware (Arakawa et al. 2011) and stochastic ingredients are required in absence of quasi equilibrium at increasing resolutions (e.g., Palmer 2001; Plant and Craig 2008). Stochastic physics have been introduced in GCMs for various reasons: to more realistically represent the subgrid-scale variability (Lin and Neelin 2000), but also to enlarge the model spread in ensemble prediction systems (e.g., Buizza et al. 1999; Teixeira and Reynolds 2008; Bengtsson et al. 2013).

The stochastic subgrid-process parameterization approach used in this paper has been introduced by Crommelin and Vanden-Eijnden (2008). The main idea behind this approach is to represent subgrid processes of an atmosphere or ocean model by stochastic processes of which the properties are inferred from high-resolution data prior to implementation. More specifically, the processes are represented by finite-state Markov chains with transition probability matrices that are estimated from data and are conditioned on the resolved model variables. In Crommelin and Vanden-Eijnden (2008), the conditional Markov chains (CMCs) were shown to adequately represent subgrid-scale variables in the Lorenz '96 model (Lorenz 1996). Using the same CMC approach in a GCM to parameterize convection is a challenging task.

In a GCM, both the large-scale and the subgrid-scale state are not single scalars as is the case in the Lorenz '96 model but, instead, are formed by various vertical profiles of resolved and subgrid variables respectively. Another difficulty is the availability of high-resolution data of convection. As explained by Crommelin and Vanden-Eijnden (2008), Markov chains can be inferred from high-resolution convection resolving model data as well as observational data. Inferring CMCs from high-resolution model data has been explored by Dorrestijn et al. (2013b) and Dorrestijn et al. (2013a).

Inspired by the stochastic multicloud model of Khouider et al. (2010), Dorrestijn et al. (2015) constructed a stochastic multicloud model on a two-dimensional square lattice, using CMCs inferred from observational data. The model was inferred from an extensive dataset, consisting of a combination of high-resolution data of deep convection (Kumar et al. 2013) and large-scale reanalysis data improved with

observational data (Davies et al. 2013). The high-resolution (2.5×2.5 km^2) data originated from a rain radar located in the tropics (Darwin, Australia) and were available every 10 min for several months in a region of size ~$1.5° \times 1.5°$. Thresholds for the cloud-top height and the rain rate were used for classification into a finite number of convective or stratiform cloud types (Dorrestijn et al. 2013a; Khouider et al. 2010). Observations of cloud-type transitions were used to estimate the transition probabilities of the CMCs. When conditioned on the large-scale vertical velocity and choosing 100 CMCs, the cloud-type area fractions of the scheme were comparable to the observational fractions in the radar domain. By varying the number of CMCs, the multicloud model could be adapted to the size of a GCM column, thereby making the parameterization scale aware.

In Gottwald et al. (2016), a similar data-driven stochastic scheme has been developed. Observational datasets from Darwin and Kwajalein were used to construct parameterizations of the convective area fraction σ_c, also conditioned on the large-scale vertical velocity. The convective area fraction was obtained by sampling directly from the area fraction distribution that was estimated from the data before, conditioned on the large-scale state. Introducing time correlation was explored as well by using CMCs. The scheme was able to adequately reproduce observational time series of σ_c.

Testing the schemes in a dynamical environment, in which the CMCs are interacting with the resolved model variables in a GCM, is a necessary step in the development of the CMC-based schemes for the usage in state-of-the-art GCMs. Therefore, in the present paper, we show results of the implementation of the stochastic multicloud model of Dorrestijn et al. (2015), referred to as Dor15, and a scheme similar to the CMC scheme of Gottwald et al. (2016), referred to as Gott15, in a GCM of intermediate complexity; the climate model Simplified Parameterizations, Primitive Equation Dynamics (SPEEDY) (Molteni 2003; Kucharski et al. 2006, 2013).

The stochastic schemes produce σ_c, which serves as a closure for the cloud-base mass-flux M_b in the convection parameterization scheme. So, SPEEDY's traditional deterministic convection scheme, a simplified Tiedtke mass-flux scheme (Tiedtke 1989), is made stochastic by using σ_c as stochastic input for the determination of M_b. This is a crucial step in the coupling of the stochastic schemes to the convection scheme of SPEEDY. The coupling of a stochastic scheme to the convection scheme of a NWP model, via σ_c and M_b, has been successfully applied earlier by Bengtsson et al. (2013).
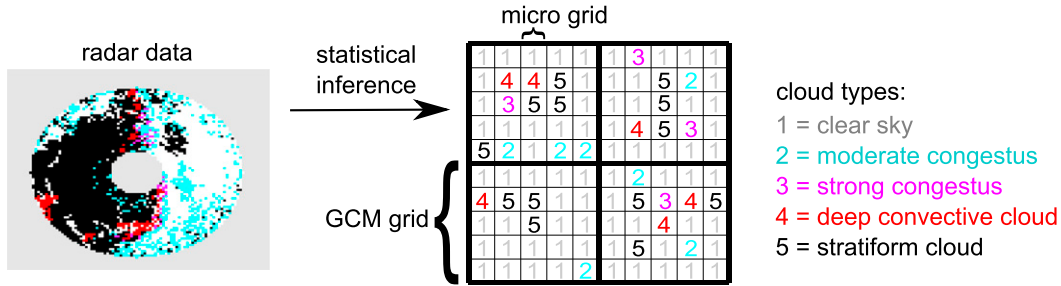
FIG. 1. Illustration of the stochastic multicloud model (the Dor15 scheme). The thick black lines indicate the GCM grid of which we see four columns from a top view. Inside the four columns, the thin black lines form the 2D microgrid of the multicloud model. Here, each GCM grid column contains $N = 25$ nodes, with a CMC on each node, switching between the five cloud types. A snapshot from the discretized radar data from Darwin is included to point out that the transition probabilities of the CMCs are estimated from observational data.

Our paper is organized as follows. In section 2, we describe the Dor15 scheme, followed by a description of the Gott15 scheme in section 3. Then, we explain how we implement the schemes in SPEEDY in section 4. We specify the observational datasets in section 5, and we present model results in section 6. A discussion follows in section 7.

2. The Dor15 scheme

The stochastic multicloud model consists of a 2D square lattice with N nodes, with at each node a CMC, denoted Y_n ($1 \leq n \leq N$), that switches, every 10 min, between the following states: 1) clear sky, 2) moderate congestus, 3) strong congestus, 4) deep convective cloud, and 5) stratiform cloud. We refer to these states as cloud types. In Fig. 1, we illustrate how the multicloud square lattice, or "microgrid," can be embedded in a GCM grid. In the figure, we see four GCM columns with $N = 25$ CMCs for each column. The value $N = 25$ has only been chosen for the sake of illustration and results will be presented for $N = 100$ and $N = 500$.

The transition probabilities of the CMCs depend on the large-scale state of the atmosphere: they are conditioned on the vertical velocity averaged over the lower part of the troposphere defined by

$$\langle \omega \rangle := \frac{1}{p_0 - p*} \int_{p*}^{p_0} \overline{\omega}(p)\, dp,$$

in which $\overline{\omega}$ is the large-scale vertical velocity (hPa h^{-1}), p_0 is the pressure at the surface, and $p*$ is the pressure level at 340 hPa. We condition the CMCs on $\langle \omega \rangle$, because in Dorrestijn et al. (2015) this variable was shown to have the largest correlation with deep convection; see Davies et al. (2013) and Peters et al. (2013) for similar findings. Since the CMCs have 5 states, the transition probability matrices are of size 5×5 and since we bin all possible

values of $\langle \omega \rangle$ into 25 intervals, we obtain 25 matrices; for each interval there is a different 5×5 matrix.

In a GCM grid column, the N CMCs yield area fractions σ_m for each cloud type $1 \leq m \leq 5$, which are defined by

$$\sigma_m = \frac{1}{N} \sum_{n=1}^{N} 1[Y_n = m], \tag{1}$$

in which $1[\cdot]$ is the indicator function ($1[Y_n = m] = 1$ if $Y_n = m$ and $1[Y_n = m] = 0$ if $Y_n \neq m$). Previous studies based on observational data (Dorrestijn et al. 2015; Gottwald et al. 2016) show that the expectation value of σ_4 is an increasing function of $\langle \omega \rangle$, with a maximum of around 0.03 for $\langle \omega \rangle \approx 15$ hPa h^{-1}.

Ideally, one would like to choose N such that the size of the microgrid cells corresponds to the typical size of a convective updraft area L^2_{conv}. This implies that N should be the ratio between the GCM horizontal grid size area ΔX^2 and L^2_{conv}; that is, $N \approx \Delta X^2 / L^2_{conv}$. The parameter N is a scaling parameter enabling the Dor15 scheme to adapt to the GCM grid resolution and determines the magnitude of the stochastic fluctuations of the area fractions σ_m: the larger N, the smaller the deviations from the expectation values, to which the fractions converge if $N \to \infty$. This gives a deterministic version of the model. Previous offline studies (Dorrestijn et al. 2015) showed that for $N = 100$ the temporal fluctuations of the deep convective fractions resemble the observational fluctuations on an area of size 170×170 km^2, and therefore, $L^2_{conv} \approx 17^2$ km^2. The value $N = 100$ is ideal for usage in a GCM with grid size $\Delta X^2 = 170^2$ km^2. We test the multicloud model in SPEEDY for the relatively small value $N = 100$, referred to as Dor15–100, to be able to assess the impact of stochastic fluctuations. As an extra sensitivity test, we do an additional experiment with $N = 500$, referred to as Dor15–500, which is a more appropriate value for SPEEDY.

For the implementation in SPEEDY we use the sum of the strong congestus and deep convective area fractions to estimate the convective area fraction:

$$\sigma_c = \sigma_4 + \sigma_3, \qquad (2)$$

which is used in the closure for M_b in the convection scheme, explained in detail in section 4.

More information about multicloud models can be found in, for example, Ajayamohan et al. (2014), Deng et al. (2015), Dorrestijn et al. (2013a, 2015), Frenkel et al. (2013), Khouider and Majda (2006), Khouider et al. (2010), Majda et al. (2007), and Peters et al. (2013).

3. The Gott15 scheme

In the Gott15 scheme, the CMCs switch between σ_c values instead of cloud types, and only one CMC is used for each GCM column (by contrast, the multicloud model has N CMCs in each GCM column). Thus, the scheme is less complex than the multicloud model but it is not scale aware. The fluctuations of σ_c cannot be adapted to the GCM resolution. We will now describe in detail how we construct the Gott15 scheme.

Again, we use the discretized Darwin radar dataset. The deep convective area fractions σ_4 are added to the strong congestus area fractions σ_3 forming σ_c. We cluster the fractions with k means (MacQueen 1967; Gan et al. 2007) using $K = 10$ cluster centroids. This results in 10 possible σ_c values, which are the states of the CMCs. We use the observational σ_c to estimate transition probability matrices of size 10×10. As in Dor15, the CMCs are conditioned on the 25 intervals of $\langle \omega \rangle$, so we estimate 25 matrices; for each interval of $\langle \omega \rangle$ there is a different 10×10 matrix. The transition probabilities of the CMC correspond to a time step of 10 min, since observational fractions are available every 10 min, and to an area size of $\sim 1.5° \times 1.5°$, which is the size of the radar domain.

The Gott15 scheme is implemented in SPEEDY in the same way as the multicloud model: σ_c is used as a closure for M_b. We stress that the main difference between the Gott15 scheme and the Dor15 scheme is that the Gott15 scheme does not make use of a multicloud model; instead, its CMCs make transitions between σ_c values.

4. Implementation in SPEEDY

SPEEDY is a GCM of intermediate complexity: only the most important processes are incorporated in the model, they are represented in a simplified way, and the GCM's resolution is coarse (Kucharski et al. 2013). It is a hydrostatic spectral model that solves the primitive equations on the entire globe. The prognostic variables are vorticity, horizontal divergence, absolute temperature, surface pressure, and specific humidity. The time integration is performed by a leapfrog scheme and the time step in the standard version of SPEEDY is 40 min. In our version, the horizontal resolution is T30, referring to a triangular truncation at total wavenumber 30. The prognostic model fields are expanded into series of spherical harmonic functions of total wavenumber 30 and smaller. Along latitude circles these functions correspond to cosine and sine functions with maximum zonal wavenumber 30. This corresponds to a size of $\sim 3.75° \times 3.75°$ for each of the $96 \times 48 = 4608$ vertical columns. In the vertical, the model has eight pressure levels. SSTs are prescribed by using observational climatological fields, while land skin temperatures are prognosed using a soil model. SPEEDY has a seasonal cycle but no daily cycle. Simplified parameterizations are used to represent shortwave and longwave radiation, deep convection, clouds, surface heat and moisture fluxes, large-scale condensation, and vertical diffusion (representing, e.g., shallow convection). Precipitation is the sum of the large-scale and convective precipitation. The large-scale precipitation is derived from a large-scale condensation scheme and the convective precipitation is derived from the deep convection scheme.

The reason why we choose such a simplified GCM is that it provides a perfect playground to explore new stochastic concepts in convection parameterizations and the impact on the representation of intraseasonal variability caused by equatorial waves. In that respect this explorative study should be considered as a natural intermediate step from recent offline studies (Dorrestijn et al. 2015; Gottwald et al. 2016) toward an implementation into the state-of-the-art GCMs.

a. The relaxation closure (CTRL)

The deep convection scheme is a simplified Tiedtke mass flux scheme (Tiedtke 1989). Convection in a grid column is triggered if the atmosphere is conditionally unstable with respect to the lowest model level and if the relative humidity in the two lowest model levels exceeds a critical value (Molteni 2003). In the standard version of SPEEDY, the cloud-base mass flux M_b is estimated by a relaxation closure. This closure determines a value of M_b such that the convection scheme relaxes back to a prescribed relative humidity threshold in 6 h. The control experiments are done using this relaxation closure and are referred to as CTRL. In the vertical, the mass, heat, and moisture fluxes are modified by a prescribed entrainment profile, while detrainment is assumed only to occur at the highest level where the convection scheme is active by depositing the

convective updraft mass, heat, and moisture into the environment.

b. Implementation of the stochastic schemes

The stochastic schemes are implemented by replacing the standard relaxation closure for M_b, which can instead be estimated by using the definition

$$M_b = \rho w_c \sigma_c, \qquad (3)$$

with a typical prescribed value of the updraft momentum at cloud base, $\rho w_c = 1 \, \mathrm{kg \, m^{-2} \, s^{-1}}$ (Möbis and Stevens 2012). For the multicloud model, we will also test the influence of this particular choice by varying this updraft momentum. In one experiment, we set ρw_c at $0.5 \, \mathrm{kg \, m^{-2} \, s^{-1}}$ while using $N = 100$, referred to as Dor15–100w0.5, and later we choose other values of ρw_c.

When the multicloud model is used, we evolve $N = 100$ or $N = 500$ CMCs in every vertical column of SPEEDY, yielding cloud-type area fractions σ_m for each cloud type at every model time step. The convective area fraction σ_c is calculated with (2) and used in (3). Note that we also evolve the CMCs for columns without deep convection (in case the trigger function did not activate convection) to be sure that the Markov chains do not have to spin up when convection is activated. Since the transition probabilities of the CMCs correspond to a time step of 10 min, we set the time step of SPEEDY at 10 min for all runs.

In each vertical column, the input of the CMCs is the large-scale vertical velocity $\langle \omega \rangle$. The value $\langle \omega \rangle$ is assigned to one of the 25 interval numbers and the CMCs will all use the same transition probability matrix that corresponds to this interval number. Given its cloud type, each CMC will switch to another cloud type (or does not change cloud type), and after that, the new area fractions σ_m are calculated using (1). In the present study, we only use σ_4 and σ_3. In Dorrestijn et al. (2015), the possible usage of the other cloud-type area fractions is described.

When the Gott15 scheme is used, we evolve only 1 CMC in every column of SPEEDY, which directly yields σ_c. The value of $\langle \omega \rangle$ in a model column determines which transition probability matrix is used by the CMC.

5. Observations

We will compare the model behavior of SPEEDY with observations. We will use two observational datasets. The first dataset is the Darwin radar dataset. We will compare M_b at time step level (10 min) of the two stochastic schemes and CTRL with M_b observed in Darwin. We emphasize that we do not have observations

of M_b; however, since we use σ_c of the stochastic schemes directly as M_b in (3), we will also use the observational σ_c as a proxy for the observational M_b by assuming again that ρw_c is equal to $1 \, \mathrm{kg \, m^{-2} \, s^{-1}}$.

The second observational dataset is the daily accumulated precipitation GPI dataset ($1° \times 1°$) (Huffman and Bolvin 2013). Since SPEEDY has a resolution of ~$3.75° \times 3.75°$, we average the observational precipitation values over blocks of this size.

6. Results

We run SPEEDY several times for 11 yr with different closures for M_b. To avoid spinup effects, data from the first year are excluded. We store variables (e.g., M_b and precipitation values) at every time step and for all vertical columns around the equator between 15°N and 15°S, which are eight vertical columns for each longitude.

a. M_b at time step level

To get a first impression of the convective behavior of SPEEDY with the several closures, we show M_b at time-step level for 2 weeks for a vertical column located at ~13°S, 130°E in Fig. 2. We choose this particular grid column because it is closest to Darwin, Australia, for which we can show the time series of M_b using σ_c as a proxy in Fig. 2a. The time series should be compared in a statistical sense. The goal is not to give an identical reproduction of the time series observed in Darwin; instead, we show "typical" time series of the several closures during the rainy season.

In Fig. 2b, we see that the mass flux of CTRL is nonzero for specific time intervals, only when the trigger function is active (an inactive trigger function is indicated by a red dot at the horizontal axis). If the trigger function allows for convection, the mass flux is always close to $0.03 \, \mathrm{kg \, m^{-2} \, s^{-1}}$; CTRL has small variability. Further, there are periods when the trigger function switches convection on and off too rapidly—for example, from day 2 until day 5. The too-intermittent behavior of CTRL is due to the trigger function.

In Fig. 2c, we clearly see the discrete character of Dor15–100: only values that are integer multiples of $1/100 = 0.01$ are attained, because $N = 100$ CMCs are used to calculate σ_c. If only one CMC is in a convective state (state 3 or 4), $\sigma_c = 1/100 = 0.01$, if two CMCs are in a convective state, then $\sigma_c = 2/100 = 0.02$, etc. The mass flux fluctuates between 0 and $0.07 \, \mathrm{kg \, m^{-2} \, s^{-1}}$, in this period of this particular realization, which suggests that the variability has improved compared to CTRL. Note that a zero mass flux can be the result of an inactive trigger function or a convective area fraction σ_c equal to
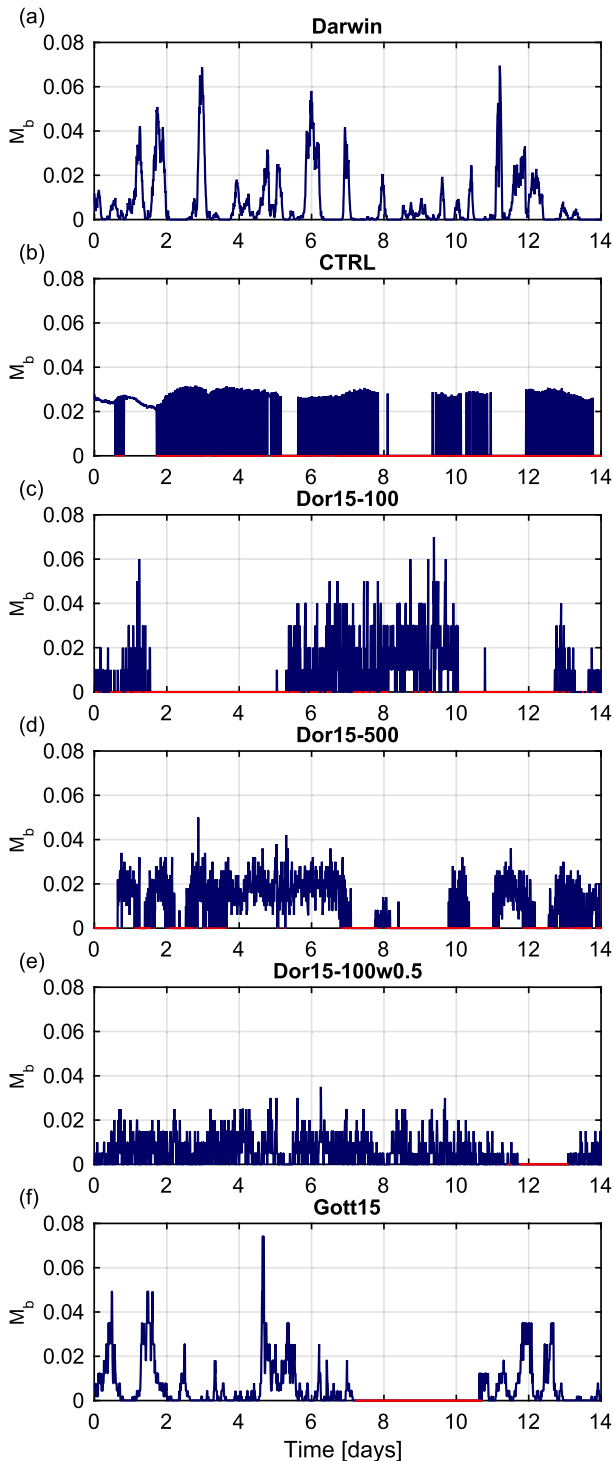
FIG. 2. Typical time series of M_b (kg m^{-2} s^{-1}) (a) observed in Darwin (σ_c observations used as a proxy for M_b, assuming $\rho w_c = 1$ kg m^{-2} s^{-1}) and produced by SPEEDY at 13°S, 130°E for (b) CTRL, (c) Dor15–100, (d) Dor15–500, (e) Dor15–100w0.5, and (f) Gott15. An inactive trigger function is indicated by a red dot at the horizontal axis.

zero. For example, from day 2 until day 5, the zero M_b is a result of an inactive trigger function. The character of the time series is too intermittent compared to the series observed in Darwin, which cannot be exclusively attributed to the trigger function.

In Fig. 2d, we see the mass flux produced by Dor15–500. Mass-flux values higher than 0.04 kg m^{-2} s^{-1} are rare. Deviations from the expectation values are expected to be smaller compared to the $N = 100$ experiment. By increasing N even more, the time series start to resemble the series of CTRL. However, note that for the deterministic limit $N \rightarrow \infty$, the closure still differs from the standard relaxation closure and, therefore, convergence of the stochastic closure to CTRL should not be expected.

In Fig. 2e, we see that Dor15–100w0.5 produces lower M_b values than Dor15–100 and that the values are multiples of 0.005. Lower mass fluxes imply that convective instabilities are less quickly removed, leading to prolonged periods of convective activity. As opposed to Dor15–100 the trigger function is active for almost the entire period: it is only inactive around day 12.

Finally, the Gott15 scheme (Fig. 2f) produces M_b time series that are similar to the series as observed in Darwin. The highest value of M_b lies between 0.07 and 0.08 kg m^{-2} s^{-1}. The general shape of the convective peaks looks quite realistic for this scheme. It is less intermittent than the multicloud and CTRL time series.

Clearly, compared to CTRL, the two stochastic schemes (Dor15 and Gott15) are better reproducing the fluctuations as observed in Darwin.

b. The distribution of M_b

In Fig. 3, the distributions of M_b are visualized by showing histograms of the relative frequency of occurrence of the nonzero M_b and the corresponding mean and standard deviation observed in Darwin (Fig. 3a) and for model data between 15°N and 15°S based on the different closures (Figs. 3b–f). The y axes are scaled logarithmically to make the tails of the distributions better visible. The Darwin histogram corresponds to a distribution that is approximately exponential with a maximum M_b of around 0.10 kg m^{-2} s^{-1}.

In Fig. 3b, we see that the mass flux of CTRL has a peak value at 0.03 kg m^{-2} s^{-1} and that the relative frequencies are rapidly decreasing to zero for larger mass fluxes. The maximum value lies below 0.05 kg m^{-2} s^{-1}. The mean mass flux of CTRL is larger than the mean mass flux observed in Darwin and the standard deviation is smaller. This is also the case if we evaluate the model data near Darwin instead of the entire tropical belt.

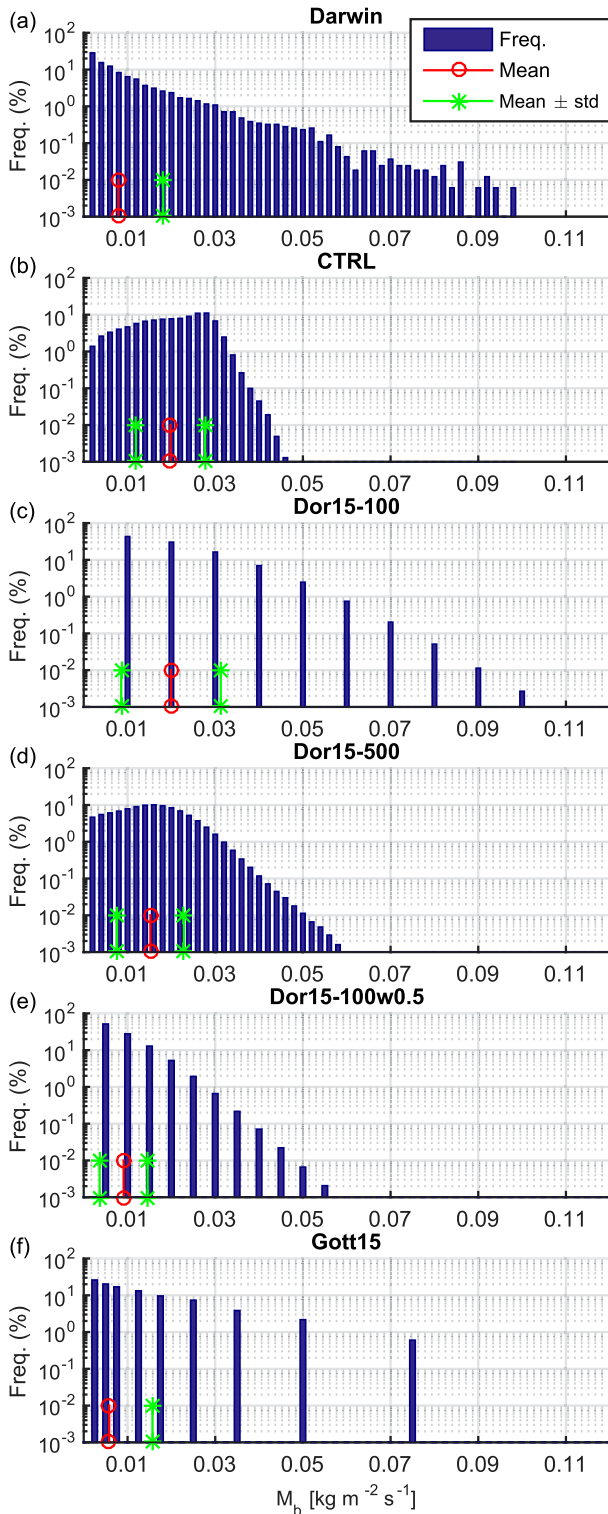The mass flux of Dor15–100 (Fig. 3c) can attain values up to 0.10 kg m^{-2} s^{-1}. The discrete character of the

FIG. 3. Histograms showing the relative frequency of occurrence of the nonzero M_b, the mean M_b, and the standard deviation for (a) the Darwin observations (using σ_c as a proxy) and for SPEEDY (1 yr of model data between 15°N and 15°S) using (b) CTRL, (c) Dor15–100, (d) Dor15–500, (e) Dor15–100w0.5, and (f) Gott15.

scheme is visible, with only integer multiples of $0.01 \, \text{kg m}^{-2} \, \text{s}^{-1}$. The mean flux is close to the mean flux of CTRL and its standard deviation is slightly larger. Dor15–500 (Fig. 3d) displays a histogram that resembles the histogram of CTRL, except that a higher maximum mass flux is possible. The histogram looks smoother than the histogram of Dor15–100, since integer multiples of $0.002 \, \text{kg m}^{-2} \, \text{s}^{-1}$ can be attained. The mean mass flux is lower than the mean mass flux of Dor15–100 and it has a smaller standard deviation. Dor15–100w0.5 produces lower M_b than Dor15–100 and the histogram suggests that M_b is approximately exponentially distributed.

Gott15 (Fig. 3f) attains 10 different mass flux values, which are exactly the values of the 10 cluster centroids. Its maximum mass flux is around $0.07 \, \text{kg m}^{-2} \, \text{s}^{-1}$; higher maximum values can be obtained, for example, by using a larger number of cluster centroids. This last option would, however, need reconstruction of the Gott15 scheme through a revised estimation of the transition matrices. The relative frequency of the bins of Gott15 seems to decrease approximately exponentially. The mean and standard deviation are close to the observational values.

We conclude that, compared to CTRL, the stochastic schemes (Dor15 and Gott15) produce mass flux distributions that are more similar to the Darwin distribution. However, the discrete character of the stochastic schemes is not very realistic.

c. Autocorrelation functions

Deep convection is correlated in time and probabilities of the occurrence and strength of convection depend strongly on earlier time instances. This is one of the reasons why we choose to parameterize convection with Markov chain models: to be able to capture this correlation. How well the several closures reproduce observational correlations can be assessed by calculating autocorrelation functions (ACFs) (Dorrestijn et al. 2015).

In Fig. 4a, we plot ACFs of M_b averaged over 15°N–15°S for 1 yr of model data with the Gott15 scheme, the multicloud model ($N = 100$ and $N = 500$), and CTRL and compare them to the ACF of M_b observed in Darwin. Compared to the observations, the ACFs of all models except Gott15 decrease too rapidly initially as a result of the intermittent character and too slow thereafter. In contrast, the ACF of Gott15 is close to the observational ACF and the discrepancies can be partly contributed to the absence of a daily cycle in SPEEDY. The absence of a daily cycle in SPEEDY contributes to a slower decay of the ACFs and the absence of a peak after 1 day.
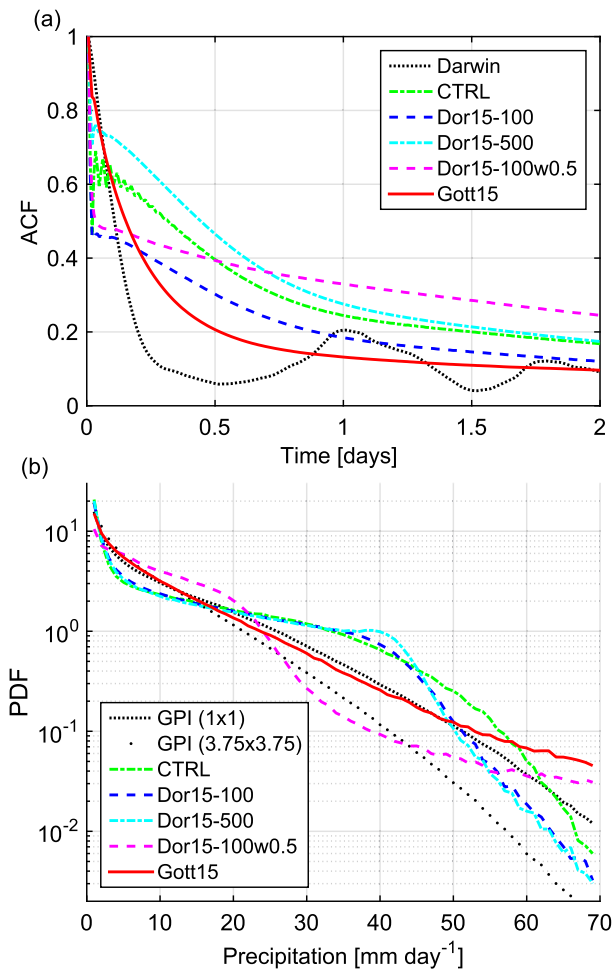
FIG. 4. (a) The autocorrelation function of M_b for Darwin observations and for SPEEDY (15°N–15°S) with CTRL, Dor15–100, Dor15–500, Dor15–100w0.5, and the Gott15 scheme. (b) The PDFs of the nonzero daily accumulated precipitation for GPI and for SPEEDY (15°N–15°S) with the same closures.

d. Precipitation

The daily accumulated precipitation is an important output of GCMs. We will assess the different mass flux closures by comparing the model's precipitation output with observations. In Fig. 4b, we show the PDFs of the nonzero daily accumulated precipitation for 10 yr of data between 15°N and 15°S. Note the logarithmic scale of the y axis. We see that the PDF produced while using Gott15 is very close to the PDF of the GPI observations (1° × 1°) for precipitation values less than 50 mm day^{-1} and that higher values are too frequent. Its PDF is not so close to the GPI observations that are averaged over blocks of size 3.75° × 3.75°; only for precipitation values below 20 mm day^{-1} there is a good fit. Gott15 has been trained with data corresponding to an area of ~1.5° × 1.5°, which may explain why it is closer to GPI 1° × 1° than to GPI 3.75° × 3.75°.

The PDFs of Dor15–100, Dor15–500, and CTRL are similar, but not close to the observational PDFs. Above 45 mm day^{-1}, the PDFs decrease with the correct slope compared to GPI (3.75° × 3.75°). The PDF of Dor15–100w0.5 differs from the PDF of Dor15–100, but it is still not close to the observational PDFs. For values higher than 50 mm day^{-1}, the PDF is close to the PDF of Gott15. In section 6f, we will further examine the impact of ρw_c.

In Fig. 5, we show 10-yr-averaged equatorial precipitation. The general patterns produced by SPEEDY (Figs. 5b–f) are somewhat similar to the GPI observations (Fig. 5a): a narrow ITCZ in the northeastern Pacific Ocean and a wide one over the Maritime Continent. However, there are some major errors: for example, the precipitation in CTRL, Dor15–100 and Dor15–500 in the northeastern Pacific Ocean, is twice as high as in GPI. Also SPEEDY's spatial patterns in the Indian Ocean differ significantly from the patterns in GPI.

Dor15–100 (Fig. 5c) and Dor15–500 (Fig. 5d) do hardly change the precipitation patterns of CTRL (Fig. 5b). So, the schemes, based on different closures, produce similar 10-yr-average precipitation. This can be explained by realizing that precipitation scales with mass flux at cloud base. Inspection of Figs. 3b–d shows that the different closures give similar mean mass flux values of 0.02 kg m^{-2} s^{-1}. Dor15–100w0.5 and Gott15 produce significantly lower mean mass flux values, which explains the reduction of the intensity of the precipitation patterns (Figs. 5e–f). These schemes do not improve the patterns in general. Only the ITCZ in the northeastern Pacific Ocean seems to improve. Precipitation in the warm pool (140°E) is still too intense and too localized compared to the observations. We conclude that the intensity of M_b, rather than the variability of M_b, seems to have a major impact on mean precipitation in SPEEDY.

e. Equatorial waves

At the equator, the Coriolis force vanishes, and it increases north and south of the equator. This results in dynamics that are typical for the tropics. The governing equations of atmosphere and ocean admit solutions that describe waves traveling along the equator. It is possible to discern atmospheric waves in satellite observations of precipitation because of their tendency to couple to deep convection.

A distinction can be made between waves that are mainly symmetrical with respect to the equator—for example, equatorial Kelvin waves traveling eastward with 15 m s^{-1} (or ~360° month^{-1}), equatorial Rossby waves (ER) traveling westward, westward inertio-gravity (WIG) waves, eastward inertio-gravity waves, and the MJO traveling eastward with 5 m s^{-1} (or ~120° month^{-1}) and

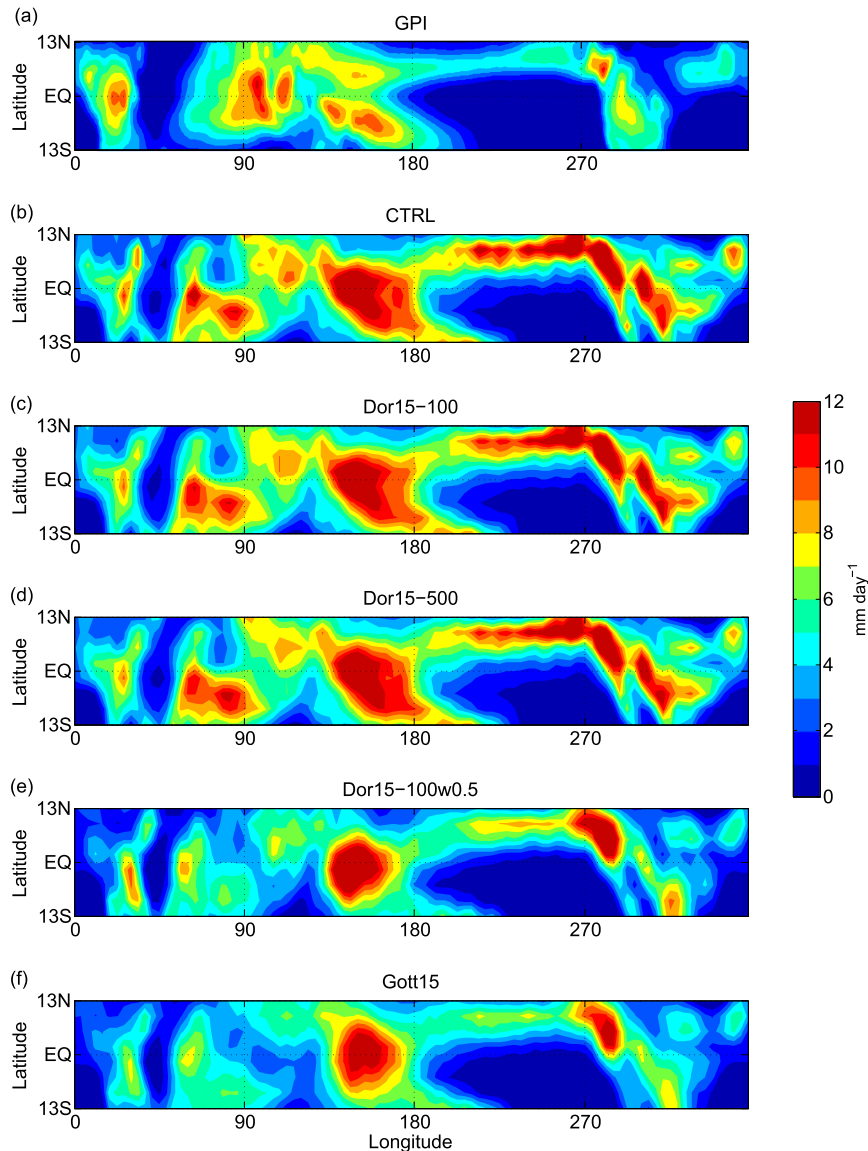FIG. 5. Mean equatorial precipitation (10-yr averaged) for (a) the GPI observations (3.75° × 3.75°) and SPEEDY (3.75° × 3.75°) with (b) CTRL, (c) Dor15–100, (d) Dor15–500, (e) Dor15–100w0.5, and (f) Gott15.

waves with an antisymmetric structure with respect to the equator (e.g., mixed Rossby–gravity). For a comprehensive treatise on equatorial waves, we refer to Wheeler and Kiladis (1999). State-of-the-art GCMs should be able to reproduce these waves. Producing realistic equatorial waves (especially the MJO) is one of the major challenges for weather and climate modelers (Kiladis et al. 2009; Biello and Majda 2005).

Exactly as in Zhang (2005), we show in Fig. 6 longitude–time plots, also known as Hovmöller diagrams, of the equatorial daily precipitation averaged over 15°N–15°S for 1 yr of GPI observations and for the

SPEEDY experiments. Hovmöller plots are useful to get a first insight in the model's ability to produce equatorial waves.

The eastward-moving Kelvin waves are clearly visible for CTRL, Dor15–100, and Dor15–500 (Figs. 6b–d). In the observations (Fig. 6a), these Kelvin waves are visible but not as prominent. The Hovmöller diagrams of the multicloud model and CTRL are in general very similar. The multicloud model seems to produce slightly larger coherent structures of heavy rainfall, which are visible as tiny red blobs—for example, in January at 90° and 150°E in Fig. 6d. The MJO events in the GPI observations—for

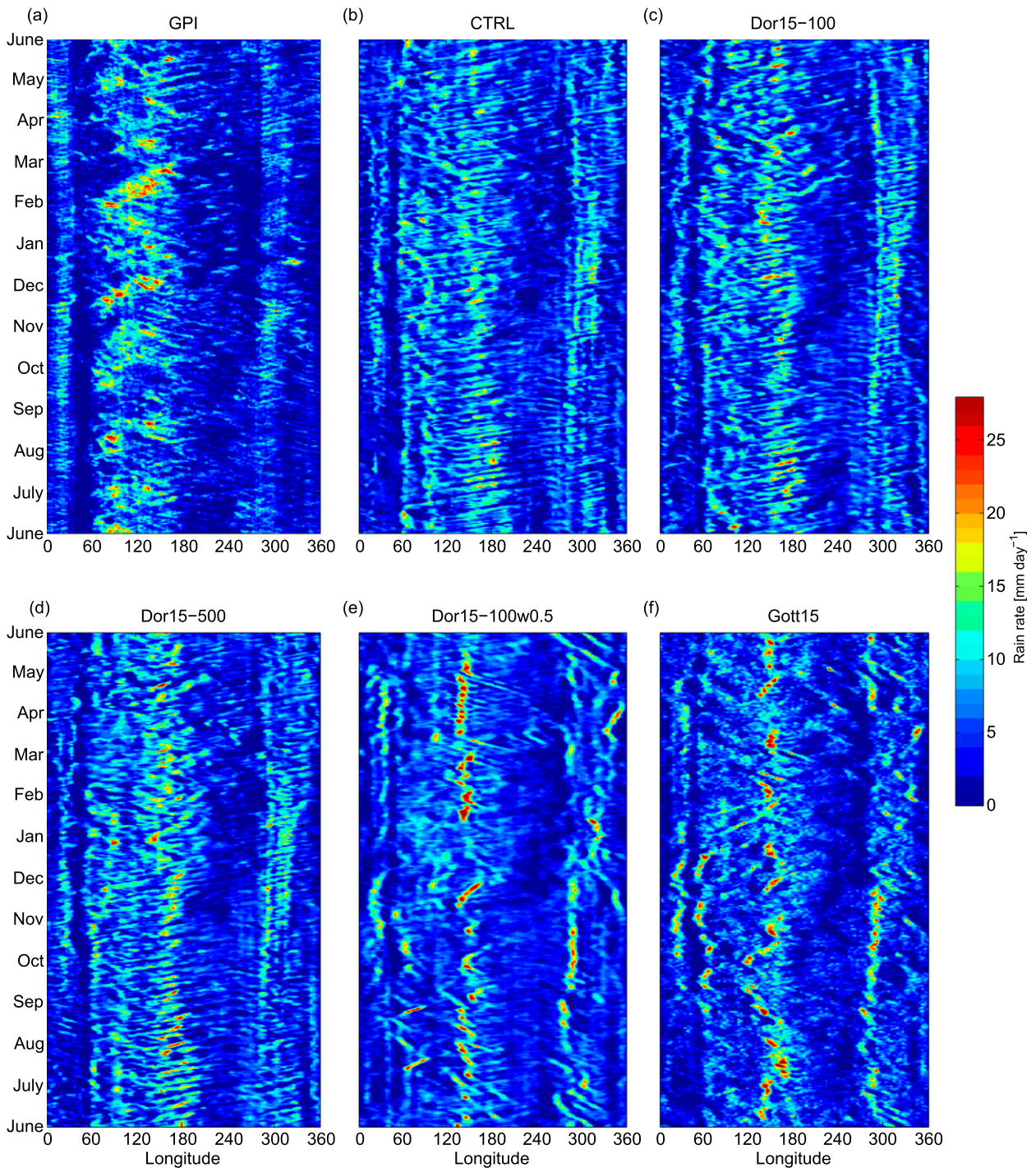FIG. 6. Hovmöller diagrams (Zhang 2005) of the daily precipitation (mm day^{-1}) averaged over 15°N–15°S for (a) the GPI observations (1° × 1°) from June 2000 to May 2001 and a typical year of SPEEDY with (b) CTRL, (c) Dor15–100, (d) Dor15–500, (e) Dor15–100w0.5, and (f) Gott15. Note that the diagrams should be compared in terms of general patterns; e.g., equatorial Kelvin waves are better visible in (b) than in (e).

example, in February (60°E–180°)—are prominent and are missing in CTRL, Dor15–100, and Dor15–500. In the Hovmöller diagram of the Gott15 scheme (Fig. 6f), large convective events are present (e.g., the red blobs

between 120°E and 180°), considerably more than in CTRL. We even see that MJO-like waves are present between 60°E and 180° in January. These MJO-like waves are, however, not as strong as in GPI, which

indicates that the representation of spatial organization of convection is still inadequate.

The Hovmöller diagram of Dor15–100w0.5, with $\rho w_c = 0.5 \, \text{kg m}^{-2} \, \text{s}^{-1}$, differs from the Hovmöller diagram of CTRL: the Kelvin waves are less prominent and structures of heavy rainfall can be seen (mainly between 60°E and 180°) that are similar to the structures of Gott15. Also the MJO-like waves are present (60°E–180°, July–August) but are even weaker than for Gott15.

To further examine the model's ability to produce equatorial waves and investigate intraseasonal variability, we calculate Wheeler–Kiladis diagrams (Wheeler and Kiladis 1999) of the equatorial precipitation. We focus on the symmetric part of the precipitation, since we are mostly interested in equatorial Kelvin waves and the MJO—the waves with the largest contributions to intraseasonal variance in precipitation. We calculate zonal–wavenumber–frequency diagrams of the symmetric part of the equatorial precipitation (15°N–15°S) divided by the background spectrum, for which we apply smoothing with a 1–2–1 filter.

In Fig. 7, we plot the diagrams for the GPI observations (Lin et al. 2006) and the SPEEDY experiments. Note, first of all, that all the SPEEDY model results differ significantly from the GPI diagram. This is, besides the differences in the power of the waves, caused by the different background spectra by which the spectra are divided. In the observations (Fig. 7a), we clearly see the MJO peak (around zonal wavenumbers 1–5 with a period between 32 and 96 days) and the Kelvin waves for positive wavenumbers. Further, we see the ER and the WIG less prominently. The diagrams of CTRL, Dor15–100, and Dor15–500 (Figs. 7b–d) are very similar to each other and show too-prominent Kelvin waves while the MJO is essentially missing in these diagrams. These are typical model misrepresentations that occur in many state-of-the-art GCMs (Lin et al. 2006). Our multicloud scheme is not able to improve the MJO. Successful MJO-like simulation with similar stochastic multicloud models is possible as demonstrated by Deng et al. (2015).

In the diagram of the Gott15 (Fig. 7f), we see an MJO peak and the Kelvin waves are less prominent as in CTRL. The tropical depressions (TD type) are too prominent. For Dor15–100w0.5 (Fig. 7e), the Kelvin waves slightly diminish in comparison to Dor15–100, the TD type are even more prominent than in Gott15, and the MJO peak is missing.

f. The updraft momentum at cloud base

In the implementation of the stochastic schemes in SPEEDY, M_b was calculated by multiplying σ_c by $\rho w_c = 1 \, \text{kg m}^{-2} \, \text{s}^{-1}$ in (3). We find that changing ρw_c has a major impact on the model behavior. If we lower ρw_c, then the equatorial Kelvin waves get less prominent and the MJO strength increases. Also the time-averaged equatorial precipitation changes (Fig. 5e) as compared to CTRL (Fig. 5b). To examine the influence of ρw_c, we do additional runs with Dor15–100 with ρw_c values over the range $0 \leq \rho w_c \leq 1.5 \, \text{kg m}^{-2} \, \text{s}^{-1}$ and calculate the power of the equatorial Kelvin waves and MJO as a function of ρw_c. We define the power of the equatorial Kelvin waves and the MJO as the average powers of the corresponding wave regions in the Wheeler–Kiladis diagram as defined in Fig. 6 of Wheeler and Kiladis (1999).

Figure 8a displays the result of 12 independent 11-yr runs of SPEEDY using the Dor15–100 scheme with different values of ρw_c. We see that ρw_c has indeed an impact on the power of the equatorial Kelvin waves and the MJO. The equatorial Kelvin waves tend to have less power for smaller ρw_c values. The GPI observational power is 0.08, so the figure suggests that the equatorial Kelvin wave power is only correct when $\rho w_c \approx 0.45 \, \text{kg m}^{-2} \, \text{s}^{-1}$. The MJO power tends to increase for smaller updraft momentum values but never reaches the MJO observational power of 0.14. Note that for $\rho w_c = 0$, the convection scheme is essentially switched off, and all precipitation is formed by large-scale precipitation. The relative contributions of the large-scale and convective precipitation as a function of ρw_c is plotted in Fig. 8b. The general idea we get is that equatorial Kelvin waves are more prominent for schemes with a larger mean M_b and consequently a larger contribution of convective precipitation, and MJO-like features are more prominent for schemes with a smaller mean M_b and consequently a larger contribution of large-scale precipitation. The ratio between convective and large-scale precipitation seems to play a role in the type and the scale of organization of tropical convection in the model. Deng et al. (2016) similarly find that the strength of stratiform heating affects the formation of MJO-like or equatorial Kelvin wave structures in an aquaplanet GCM.

With this novel method of calculating the power of the MJO and equatorial Kelvin waves, it is possible to express the model's ability to simulate these waves in a single scalar. This enables modelers to directly tune parameters for optimal simulation of these waves. Note, however, that even if the powers are exactly equal to the observational powers, it is not yet sufficient to conclude that the model simulates the waves perfectly. Other requirements have to be fulfilled as well (Zhang 2005). The power only gives an impression. For example, CTRL gives a too-high equatorial Kelvin wave power, 0.12, and a too-low MJO power, 0.02, which is consistent with the patterns found in the Hovmöller diagrams (Figs. 6a,b).
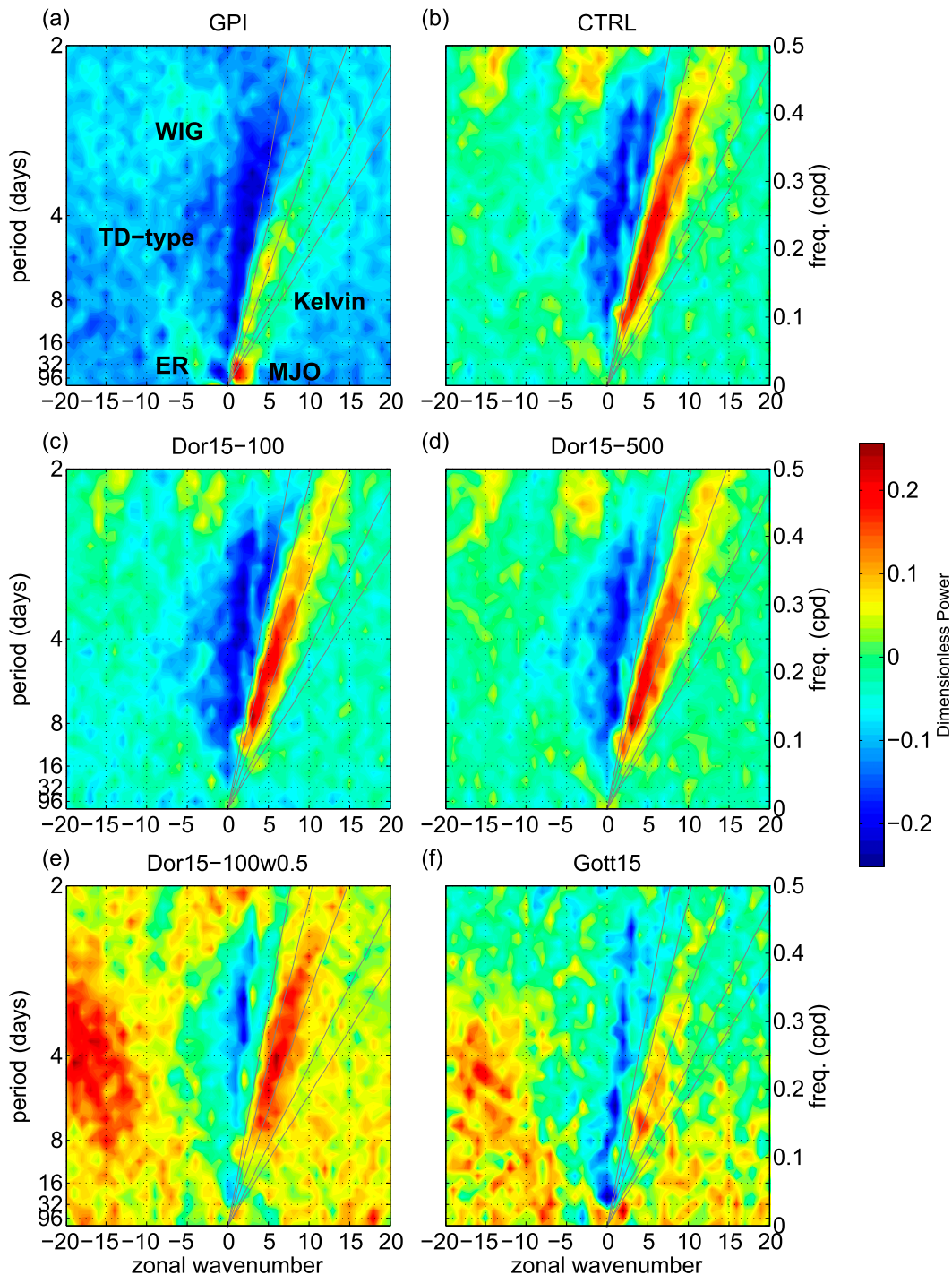
FIG. 7. Zonal wavenumber–frequency diagrams (Wheeler and Kiladis 1999; Lin et al. 2006) of the symmetric part of the equatorial precipitation (15°N–15°S) divided by the background spectrum for (a) GPI observations, (b) CTRL, (c) Dor15–100, (d) Dor15–500, (e) Dor15–100w0.5, and (f) Gott15.

7. Discussion

We have implemented two different stochastic parameterizations for the convective area fraction σ_c in the convection scheme of the intermediate complexity GCM SPEEDY and evaluated the impact in the tropics.

In both stochastic parameterizations σ_c is estimated with CMCs of which the transition probabilities are
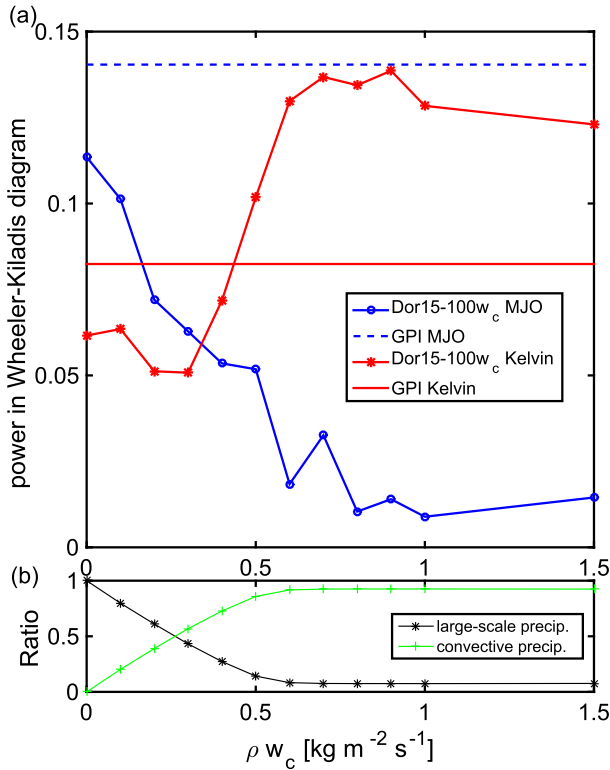
FIG. 8. (a) The average power of the equatorial Kelvin waves (line with asterisks) and the MJO (line with circles) in the Wheeler–Kiladis diagram of SPEEDY, using Dor15–100, as a function of ρw_c. Compare with the GPI Kelvin (red solid line) and MJO (blue dashed line) average power. (b) The relative contributions of the large-scale and convective precipitation as a function of ρw_c.

conditioned on the large-scale vertical velocity $\langle \omega \rangle$, as this is the large-scale variable that displays the largest correlation with the occurrence of deep convection (Dorrestijn et al. 2015). Note that a closure based on $\langle \omega \rangle$ effectively resembles a moist convergence closure, but because of the stochastic aspects our closures are not so hardwired as the more traditional deterministic moist convergence closures (e.g., Kuo 1965; Tiedtke 1989). Although it is difficult to disentangle convergence and convection in terms of causality, there is no reason not to use the large-scale vertical velocity to condition the transition probabilities of the CMCs.

On a local grid point level, both stochastic schemes produce mass flux time series that are more realistic than the series produced by the standard CTRL version (Fig. 2). This is also reflected in a broader and more realistic frequency of occurrence distribution of the cloud-base mass flux (Fig. 3). Gott15 and to a lesser extend Dor15 also improve the daily accumulated tropical precipitation compared to CTRL (Fig. 4b).

Substantial improvement of the temporal autocorrelation function for M_b is only observed for Gott15 (Fig. 4a).

Wheeler–Kiladis diagrams show that the equatorial Kelvin waves are too prominent in SPEEDY for CTRL and that the MJO is missing entirely. Gott15 significantly improves the representation of both the MJO and the equatorial Kelvin waves. Dor15 is only able to improve on this issue by strongly reducing ρw_c. By increasing the relaxation time scale of the relaxation closure in CTRL, similar changes are to be expected (Frierson 2007). For Dor15, it seems that changing the mean M_b has a larger impact on the representation of the equatorial waves than changing the magnitude of stochastic fluctuations of M_b on a time step level.

How many of the model errors are due to the convection schemes and how much due to the large-scale forcings of SPEEDY? The range of $\langle \omega \rangle$ values produced by SPEEDY compares well with the range observed around Darwin. The time series of M_b in Figs. 2c and 2f compare well with time series produced by the same schemes using observed $\langle \omega \rangle$ values (Dorrestijn et al. 2015; Gottwald et al. 2016). In addition, the large range in different mass flux behavior displayed in Figs. 2b–f suggests that most of the discrepancies between the Darwin time series and the model time series are due to the convection parameterizations and not the large-scale forcings of SPEEDY. The too-intermittent character of, for example, the Dor15–100 scheme is due to the scheme itself and not to SPEEDY.

An advantage of the Dor15 scheme over the Gott15 scheme is that it can be adapted to the scale of the GCM grid column, which makes it more universal in usage. We have, however, seen that the results for the Gott15 scheme are better than for the more involved Dor15 scheme. The main difference between the two methods is that the Gott15 scheme has been trained with the macroscopic data (i.e., averaged σ_c over the entire radar domain), while the Dor15 scheme has been trained on a finer scale: individual radar pixels. The Gott15 scheme works with only one CMC that directly yields σ_c corresponding to the size of the radar domain, while the Dor15 scheme works with N CMCs for which each CMC corresponds to the size of a convective updraft and σ_c is calculated later with (1) and (2). The main reason why Gott15 performs better is that it implicitly inferred spatial interactions between neighboring radar pixels, which are not captured by the independently evolving CMCs of Dor15. This could also be the reason that the Gott15 scheme is less intermittent than the multicloud schemes (Fig. 2) and has a more realistic ACF (Fig. 4a).

Including local interactions between neighboring cells in the Dor15 model could improve its performance but

TABLE 1. Computational costs (seconds per model day) of SPEEDY with the several closures compared to CTRL. The third column shows the number of random numbers that has to be generated each model day. The last column shows the number of random numbers that has to be generated for each model column (4608 columns) each time step (10 min). Calculations are performed on a PC with 7.7-GB memory and a 2.5-GHz processor.

Scheme	Computational costs (seconds per model day)	No. random numbers per day	No. random numbers per column per time step
CTRL	3.4	0	0
Dor15–100	5.0	6.6×10^7	100
Dor15–500	11.5	3.3×10^8	500
Gott15	3.4	6.6×10^5	1

lies beyond the scope of this paper. Including spatial interactions makes the model more complicated, because for every configuration of the neighboring cells a different transition probability matrix is needed. For successful inclusion of spatial interaction we refer to Bengtsson et al. (2013), in which a cellular automata approach (deterministic and stochastic) is applied to make convection interact spatially between different grid boxes of a NWP model, leading to a more realistic representation of convective organization. Further, in Dorrestijn et al. (2013a) locally interacting CMCs have been inferred from LES data and in Khouider (2014) the multicloud model of Khouider et al. (2010) is extended by including spatial dependencies.

The Dor15 multicloud model is inspired by the multicloud model of Khouider et al. (2010). The models are similar because in both models CMCs are positioned on a microgrid and randomly switch cloud type with probabilities that depend on the large-scale forcing. The main difference between the models is that the transition probabilities of the Dor15 scheme are estimated from data while the transition probabilities used in Khouider et al. (2010) are derived by choosing typical time scales of formation of clouds, conversion between cloud types, and decay of clouds, which are based on physical intuition. Furthermore, in the multicloud model of Khouider et al. (2010), probabilities are conditioned on CAPE and middle-troposphere dryness instead of large-scale vertical velocity for Dor15. In Khouider et al. (2010), a stochastic coarse-grained birth–death system is derived for the multicloud model, such that each GCM column only uses one CMC, which makes the method very effective. Further, the model of Khouider et al. (2010) is scale aware because the number of lattice sites in the microgrid can be adapted to the GCM gridbox size. We conclude that the beneficial properties of both methods could be combined to obtain an even better model. Especially the inclusion of spatial dependencies as in the extension in Khouider (2014) is promising.

In some recent studies (Ajayamohan et al. 2014; Ragone et al. 2015), new convection parameterizations

have been implemented in aquaplanet GCMs. SPEEDY can also run in aquaplanet mode, but for comparison to satellite observations, we have chosen to include land in the experiments.

A final remark on computational costs of the new stochastic schemes. The multicloud scheme, for which N CMCs have to be evolved for each grid column (including the generation of random numbers) increases the computational costs of the convective scheme substantially, while the computational burden of Gott15 is marginal. In Table 1, we list these computational costs. In GCMs with a large number of grid columns, using a large number of CMCs ($N > 100$) for each column could become computationally problematic. Khouider et al. (2010) showed that the usage of birth–death-like processes, with the same characteristics, is a solution to this problem.

REFERENCES

Ajayamohan, R. S., B. Khouider, and A. Majda, 2014: Simulation of monsoon intraseasonal oscillations in a coarse-resolution aquaplanet GCM. *Geophys. Res. Lett.*, **41**, 5662–5669, doi:10.1002/2014GL060662.

Arakawa, A., 2004: The cumulus parameterization problem: Past, present, and future. *J. Climate*, **17**, 2493–2525, doi:10.1175/1520-0442(2004)017<2493:RATCPP>2.0.CO;2.

——, and W. Schubert, 1974: Interaction of a cumulus cloud ensemble with the large-scale environment, Part I. *J. Atmos. Sci.*, **31**, 674–701, doi:10.1175/1520-0469(1974)031<0674:IOACCE>2.0.CO;2.

——, J.-H. Jung, and C.-M. Wu, 2011: Toward unification of the multiscale modeling of the atmosphere. *Atmos. Chem. Phys.*, **11**, 3731–3742, doi:10.5194/acp-11-3731-2011.

Bengtsson, L., M. Steinheimer, P. Bechtold, and J.-F. Geleyn, 2013: A stochastic parametrization for deep convection using cellular automata. *Quart. J. Roy. Meteor. Soc.*, **139**, 1533–1543, doi:10.1002/qj.2108.

Biello, J., and A. Majda, 2005: A new multiscale model for the Madden–Julian oscillation. *J. Atmos. Sci.*, **62**, 1694–1721, doi:10.1175/JAS3455.1.

Buizza, R., M. Milleer, and T. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908, doi:10.1002/qj.49712556006.

Crommelin, D., and E. Vanden-Eijnden, 2008: Subgrid-scale parameterization with conditional Markov chains. *J. Atmos. Sci.*, **65**, 2661–2675, doi:10.1175/2008JAS2566.1.

Davies, L., C. Jakob, P. May, V. V. Kumar, and S. Xie, 2013: Relationships between the large-scale atmosphere and the small-scale convective state for Darwin, Australia. *J. Geophys. Res. Atmos.*, **118**, 11 534–11 545, doi:10.1002/jgrd.50645.

Deng, Q., B. Khouider, and A. Majda, 2015: The MJO in a coarse-resolution GCM with a stochastic multicloud parameterization. *J. Atmos. Sci.*, **72**, 55–74, doi:10.1175/JAS-D-14-0120.1.

——, ——, and ——, 2016: Effect of stratiform heating on the planetary-scale organization of tropical convection. *J. Atmos. Sci.*, **73**, 371–392, doi:10.1175/JAS-D-15-0178.1.

Dorrestijn, J., D. Crommelin, J. Biello, and S. Böing, 2013a: A data-driven multi-cloud model for stochastic parametrization of deep convection. *Philos. Trans. Roy. Soc. London*, **A371**, 20120374, doi:10.1098/rsta.2012.0374.

——, ——, A. P. Siebesma, and H. Jonker, 2013b: Stochastic parameterization of shallow cumulus convection estimated from high-resolution model data. *Theor. Comput. Fluid Dyn.*, **27**, 133–148, doi:10.1007/s00162-012-0281-y.

——, ——, ——, ——, and C. Jakob, 2015: Stochastic parameterization of convective area fractions with a multicloud model inferred from observational data. *J. Atmos. Sci.*, **72**, 854–869, doi:10.1175/JAS-D-14-0110.1.

Frenkel, Y., A. Majda, and B. Khouider, 2013: Stochastic and deterministic multicloud parameterizations for tropical convection. *Climate Dyn.*, **41**, 1527–1551, doi:10.1007/s00382-013-1678-z.

Frierson, D. M. W., 2007: Convectively coupled kelvin waves in an idealized moist general circulation model. *J. Atmos. Sci.*, **64**, 2076–2090, doi:10.1175/JAS3945.1.

Gan, G., C. Ma, and J. Wu, 2007: *Data Clustering: Theory, Algorithms, and Applications.* 3rd ed. ASA-SIAM Series on Statistics and Applied Probability, SIAM, 466 pp., doi:10.1137/1.9780898718348.

Gottwald, G. A., K. Peters, and L. Davies, 2016: A data-driven method for the stochastic parametrisation of subgrid-scale tropical convective area fraction. *Quart. J. Roy. Meteor. Soc.*, doi:10.1002/qj.2655, in press.

Huffman, G. J., and D. T. Bolvin, 2013: Version 1.2 GPCP one-degree daily precipitation data set documentation. WDC-A, NCDC, GPCP 1DD, 27 pp. [Available online at ftp://meso.gsfc.nasa.gov/pub/1dd-v1.2/1DD_v1.2_doc.pdf.]

Khouider, B., 2014: A coarse grained stochastic multi-type particle interacting model for tropical convection: Nearest neighbour interactions. *Commun. Math. Sci.*, **12**, 1379–1407, doi:10.4310/CMS.2014.v12.n8.a1.

——, and A. Majda, 2006: A simple multicloud parameterization for convectively coupled tropical waves. Part I: Linear analysis. *J. Atmos. Sci.*, **63**, 1308–1323, doi:10.1175/JAS3677.1.

——, J. Biello, and A. Majda, 2010: A stochastic multicloud model for tropical convection. *Commun. Math. Sci.*, **8**, 187–216, doi:10.4310/CMS.2010.v8.n1.a10.

Kiladis, G., M. Wheeler, P. T. Haertel, K. H. Straub, and P. E. Roundy, 2009: Convectively coupled equatorial waves. *Rev. Geophys.*, **47**, RG2003, doi:10.1029/2008RG000266.

Kucharski, F., F. Molteni, and A. Bracco, 2006: Decadal interactions between the western tropical Pacific and the North Atlantic Oscillation. *Climate Dyn.*, **26**, 79–91, doi:10.1007/s00382-005-0085-5.

——, ——, M. King, R. Farneti, I.-S. Kang, and L. Feudale, 2013: On the need of intermediate complexity general circulation models: A "SPEEDY" example. *Bull. Amer. Meteor. Soc.*, **94**, 25–30, doi:10.1175/BAMS-D-11-00238.1.

Kumar, V. V., C. Jakob, A. Protat, P. T. May, and L. Davies, 2013: The four cumulus cloud modes and their progression during rainfall events: A C-band polarimetric radar perspective. *J. Geophys. Res. Atmos.*, **118**, 8375–8389, doi:10.1002/jgrd.50640.

Kuo, H. L., 1965: On formation and intensification of tropical cyclones through latent heat release by cumulus convection. *J. Atmos. Sci.*, **22**, 40–63, doi:10.1175/1520-0469(1965)022<0040:OFAIOT>2.0.CO;2.

Lin, J.-B., and J. Neelin, 2000: Influence of a stochastic moist convective parameterization on tropical climate variability. *Geophys. Res. Lett.*, **27**, 3691–3694, doi:10.1029/2000GL011964.

Lin, J.-L., and Coauthors, 2006: Tropical intraseasonal variability in 14 IPCC AR4 climate models. Part I: Convective signals. *J. Climate*, **19**, 2665–2690, doi:10.1175/JCLI3735.1.

Lorenz, E. N., 1996: Predictability—A problem partly solved. *Proc. Seminar on Predictability*, Reading, United Kingdom, ECMWF, 1–18.

MacQueen, J., 1967: Some methods for classification and analysis of multivariate observations. *Proc. Fifth Berkeley Symp. Mathematical Statistics and Probability*, Berkeley, CA, Statistical Laboratory of the University of California, 281–297.

Majda, A., S. Stechmann, and B. Khouider, 2007: Madden-Julian Oscillation analog and intraseasonal variability in a multi-cloud model above the equator. *Proc. Natl. Acad. Sci. USA*, **104**, 9919–9924, doi:10.1073/pnas.0703572104.

Möbis, B., and B. Stevens, 2012: Factors controlling the position of the Intertropical Convergence Zone on an aquaplanet. *J. Adv. Model. Earth Syst.*, **4**, M00A04, doi:10.1029/2012MS000199.

Molteni, F., 2003: Atmospheric simulations using a GCM with simplified physical parametrizations. I. Model climatology and variability in multi-decadal experiments. *Climate Dyn.*, **20**, 175–191, doi:10.1007/s00382-002-0268-2.

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. *Quart. J. Roy. Meteor. Soc.*, **127**, 279–304, doi:10.1002/qj.49712757202.

Peters, K., C. Jakob, L. Davies, B. Khouider, and A. Majda, 2013: Stochastic behavior of tropical convection in observations and a multicloud model. *J. Atmos. Sci.*, **70**, 3556–3575, doi:10.1175/JAS-D-13-031.1.

Plant, R., and G. Craig, 2008: A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.*, **65**, 87–105, doi:10.1175/2007JAS2263.1.

Ragone, F., K. Fraedrich, H. Borth, and F. Lunkeit, 2015: Coupling a minimal stochastic lattice gas model of a cloud system to an atmospheric general circulation model. *Quart. J. Roy. Meteor. Soc.*, **141**, 37–51, doi:10.1002/qj.2331.

Randall, D. A., Harshvardhan, D. A. Dazlich, and T. G. Corsetti, 1989: Interactions among radiation, convection, and large-scale

dynamics in a general circulation model. *J. Atmos. Sci.*, **46**, 1943–1970, doi:10.1175/1520-0469(1989)046<1943: IARCAL>2.0.CO;2.

Teixeira, J., and C. Reynolds, 2008: Stochastic nature of physical parameterizations in ensemble prediction: A stochastic convection approach. *Mon. Wea. Rev.*, **136**, 483–496, doi:10.1175/ 2007MWR1870.1.

Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea.*

Rev., **117**, 1779–1800, doi:10.1175/1520-0493(1989)117<1779: ACMFSF>2.0.CO;2.

Wheeler, M., and G. Kiladis, 1999: Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber–frequency domain. *J. Atmos. Sci.*, **56**, 374–399, doi:10.1175/1520-0469(1999)056<0374: CCEWAO>2.0.CO;2.

Zhang, C., 2005: Madden-Julian Oscillation. *Rev. Geophys.*, **43**, RG2003, doi:10.1029/2004RG000158.