Stochastic behavior of tropical convection in observations and a multicloud model

KARSTEN PETERS, * CHRISTIAN JAKOB,

ARC Centre of Excellence for Climate System Science,

School of Mathematical Sciences, Monash University, Clayton, VIC 3800, Australia

LAURA DAVIES

School of Mathematical Sciences, Monash University, Clayton, VIC 3800, Australia

BOUALEM KHOUIDER

Department of Mathematics and Statistics,

University of Victoria, PO BOX 3045 STN CSC, B.C. Victoria, Canada V8W 3P4

ANDREW J. MAJDA

Centre for Atmosphere Ocean Science and Courant Institute of Mathematical Sciences,

New York University, 251 Mercer Street, Room 902, New York, NY 10012

* *Corresponding author address:* ARC Centre of Excellence for Climate System Science, School of Mathematical Sciences, Monash University, Clayton, VIC 3800, Australia.

E-mail: karsten.peters@monash.edu

ABSTRACT

The aim for a more accurate representation of tropical convection in global circulation models is a long-standing issue. Here, we investigate the relationships between large- and convective scales in observations and a Stochastic Multicloud Model (SMCM) to ultimately support the design of a novel convection parametrization with stochastic elements. Observations of tropical convection obtained at Darwin and Kwajalein are used here. We find that the variability of observed tropical convection generally decreases with increasing large-scale forcing, implying a transition from stochastic to more deterministic behaviour with increasing forcing. Convection shows to yield a more systematic relationship with measures related to large-scale convergence compared to measures related to energetics, e.g. CAPE. Using the observations, we adjust the parameters in the SMCM, force it with the time series of the observed large scale state and compare the simulated convective behaviour to that observed. We find that the SMCM-modelled cloud fields compare better with observations when using predictors related to convergence rather than energetics. Furthermore, the underlying framework of the SMCM is able to reproduce the observed functional dependencies of convective variability on the imposed large-scale state – an encouraging result on the road towards a novel convection parametrization approach. However, establishing sound cause-and-effect relationships between tropical convection and the large-scale environment remains problematic and warrants further research.

1. Introduction

Climate projections using general circulation models (GCMs) are the tool of choice when it comes to quantifying the anthropogenic influence on Earth's climate, ultimately answering the question to what degree humanity has an influence on global mean surface temperature. Although GCMs have undergone considerable development, mainly manifested in an ever-more increase in complexity, uncertainty in climate sensitivity has not been substantially reduced since its ad hoc introduction by Charney et al. (1979) and major atmospheric processes are still subject to considerable uncertainties. Of these, atmospheric convection and the clouds and feedbacks associated with it are most probably the most uncertain in the latest generation of GCMs (Randall et al. 2007). This is not only true for the multi-model ensemble of the CMIP3 (Coupled Model Intercomparison Project phase 3, Meehl et al. 2007), but model parameters associated with convection are often the most sensitive in perturbed parameter ensembles (Murphy et al. 2004; Klocke et al. 2011).

Uncertainties in the representation of convection in current generation GCMs not only lead to uncertainties in estimates of climate sensitivity, but also manifest themselves in an erroneous simulation of precipitation. Generally, GCMs are capable of capturing the overall amount of precipitation well, but the spatial distribution and variance often compare poorly to observations (*e.g.* Dai 2006; Pincus et al. 2008). Due to the limited spatial resolution of a GCM, atmospheric convection is of subgrid-scale nature and can thus not be explicitly resolved and must be parameterised. Since the emergence of the first convection parametrization techniques some four decades ago, the response of convective elements to a given large-scale atmospheric state has mostly been formulated as purely deterministic (see Arakawa (2004) for a review) which implicitly prevents a particular model integration from developing convective variability beyond that given by the atmospheric state at the grid-point level.

It is just in the last decade that a possible solution to this lack of variability in parameterised subgrid-scale processes has emerged. This solution is based on representing the variability in the response of unresolved processes to the large-scale environment in a dynamically-stochastic rather than in a purely deterministic manner (Palmer 2001), and has been shown to increase predictive skill of numerical weather prediction (*i.e.* Buizza et al. 1999).

Specifically targeted towards improving the representation of convection, Lin and Neelin (2000, 2003) introduced random perturbations to convective available potential energy (CAPE) or the heating profile of the host convective scheme and found that even such a simple approach significantly enhanced precipitation variance towards that of observations. Randomly perturbing the trigger function of the Kain-Fritsch convection scheme also proved to yield an increase in predictive skill (Bright and Mullen 2002). Teixeira and Reynolds (2008) randomly sampled convective-parametrization relevant variables from a subgrid-scale distribution and found an increase in the spread of an ensemble prediction system and in particular a better representation of tropical convection. A similarly simple approach was taken by Tompkins and Berner (2008) who randomly sampled a subgrid-scale relative humidity distribution to perturb a convective parcel's initial humidity and/or the humidity of the entrained air during ascent. Although promising results were obtained for mid-latitudes, the methodology employed did not yield improvements in tropical convection. In all the studies mentioned above, the randomly sampled deviations were assumed proportional to the mean of the perturbed variable – an assumption shown to be valid when using cloud resolving model data as surrogate for observations (Shutts and Palmer 2007) .

Taking a step further from just modifying the input parameters for existing convective parametrization closures and cloud models, several recent studies focused on formulating more advanced stochastic schemes. Majda and Khouider (2002) introduced a stochastic parameterization of convective inhibition (CIN) based on the Ising model of statistical mechanics. It is further coarse grained to obtain a Markov birth-death process, which is two-way coupled to the large-scale dynamics and which can be integrated with very little computational overhead (Khouider et al. 2003). The stochastic CIN model is used in Khouider et al. (2003) and in Majda et al. (2008) to improve the wave variability and climate in an otherwise deficient mass-flux like parameterization in the context of a simple one and half layer toy GCM. Plant and Craig (2008) calculated a distribution of convective plumes and then randomly sampled this distribution to obtain a plume-ensemble which matches a required grid-box mean mass-flux given by a CAPE closure. Testing in a single-column model environment yielded high variability for small grid-boxes, approaching the deterministic limit with increasing grid-box size. Recently, this scheme was tested in a limited area model-ensemble over central Europe and results showed a promising increase in precipitation variance (Groenemeijer and Craig 2012). Although not concentrating on deep convection, the study of Dorrestijn et al. (2012) represents a notable approach to stochastic parametrization of shallow cumulus convection. They applied a Markov chain method to sample pairs of turbulent heat and moisture fluxes obtained form Large-Eddy Simulations (LES) and found a good agreement in the calculated ensemble spread compared to the LES data. Following the coarse graining ideas used in Khouider et al. (2003), Khouider et al. (2010) designed the Stochastic Multi-Cloud Model (SMCM) based on a birth-death process to represent tropical convection. The SMCM calculates the evolution of a cloud population consisting of three cloud types associated with tropical convection (congestus, deep convection, stratiform) constrained by the large-scale atmospheric state. The state of the cloud ensemble at any given time and large-scale forcing is represented by area fractions per cloud type on a subgrid-scale lattice. The SMCM was shown to reasonably simulate tropical convection and associated wave-features when coupled to a simple two-layer atmospheric model (Khouider et al. 2010; Frenkel et al. 2012, 2013).

As the vast majority of today's GCM convection schemes are mass flux schemes, the cloud area fractions simulated by the SMCM could prove valuable for introducing a stochastic component to such schemes. Then at least one part (area) of the cloud base mass flux would yield a stochastic component, leaving the other part (updraft velocity) to be assigned in another suitable fashion.

It is the aim of this study to provide an assessment of whether the underlying framework of the SMCM is suitable to reproduce observed convective behavior. In doing so, we analyse observed convective behavior and subsequently adjust the model parameters, which have so far been based on sensible empirical assumptions (Khouider et al. 2010), to match the observed mean response of convection to the large-scale state. We then use the resulting, adjusted model to test whether its underlying framework is suitable to reproduce the statistical mean behavior of observed convection, the positive outcome of which would render the SMCM a useful tool for convection parametrization.

The observational dataset we use in this study is described in Jakob et al. (2011) and represents a long-term, large scale dataset for three consecutive wet seasons over Darwin, Australia, complemented by an identically derived, but shorter dataset representative for Kwajalein. The Darwin-dataset has been shown to contain valuable information for characterising relationships between atmospheric convection and the large-scale state, with one of the most notable findings being that the relationships between convection and CAPE or vertical velocity show to be entirely stochastic or quasi-deterministic, respectively (Jakob et al. 2011).

We introduce the basics of the SMCM, the observational dataset as well as the observation derived forcing for the SMCM in Section 2 and present the statistical relationships of observed convection to large-scale variables in Section 3. We then adjust the parameters of the SMCM, force it with the observed large-scale state and analyse the statistics of the modeled convection as well as the stochasticity of the model solution in Section 4. Section 5 gives a summary, conclusions and short outlook.

2. Prerequisites: the model and the observations

In this study, we utilise the recently introduced stochastic multicloud model (SMCM, Khouider et al. 2010) in conjunction with a large scale observational dataset representative of a tropical location. In a nutshell, we investigate the degree to which the mathematical framework of the SMCM is suitable to reproduce the behavior of observed tropical convection – a necessary step towards a possible future usage in GCMs. In the following, we shortly introduce the SMCM (Sec. a) and the observational dataset (Sec. b)

a. The SMCM: a short introduction

Given the temporal evolution of a large scale atmospheric state representative of a tropical location, the SMCM simulates the evolution of an ensemble of three cloud types associated with tropical convection on a lattice containing $n \times n$ sites. The considered cloud types are congestus and deep convective as well as stratiform clouds (shallow convection is not considered) and the large scale atmospheric state is given by two variables: one representing a proxy for convective activity and the other representing a proxy for mid-tropospheric dryness (cf. Sec. c). In the SMCM, the evolution of the cloud ensemble is represented by a coarse grained birth-death process. This process is evolved in time by means of an acceptance-rejection Markov chain Monte Carlo method based on Gillespie's exact algorithm (Gillespie (1975), see Khouider et al. (2010) for details on the implementation). Each individual lattice site can take either one of four states: clear sky, congestus cloud, deep convective cloud, or stratiform cloud. The total size of this lattice, say 20×20 sites, is assumed as being representative of a GCM grid-box, but there is no explicit spatial scale associated with neither the individual lattice sites nor with the total lattice. There is also no spatial coherence between individual lattice sites, *i.e.* the temporal evolution at one site is completely independent of that of its neighbors. However, local interactions between lattice sites can be easily incorporated, provided the strength and nature of these interactions are understood.

The evolution of this birth-death process is determined by a set of equations which define transition rates from one of the four states (see above) to another. Individual transition rates can, but need not, be dependent on the given large scale state and their formulation is mainly inspired by physical intuition and based on specific rules, *e.g.* a deep convective cloud is not allowed to form from a stratiform cloud (see Khouider et al. 2010, for detail). The individual transition rates are associated with timescales assumed of being representative for a specific transition. These transition timescales have been chosen in an ad-hoc, but physically meaningful manner and represent the only parameters that can be used to tune the SMCM in its current formulation. Khouider et al. (2010) presented two sets of transition timescales, both of which should be considered as rough estimates. Recently, Frenkel et al. (2012) found a third set of transition rates more useful. In this study, we use observations to take a closer look at these previously made choices of transition timescales.

So far, the SMCM has not been used in combination with observations, but was coupled to a simple two-layer atmospheric model capable of capturing the main characteristics of tropical convection and associated wave features (Khouider and Majda 2006, 2008b,a; Khouider et al. 2010). There, simple formulations of precipitation formation and the associated heating profiles accounted for the feedback to the dynamics. Recently, Frenkel et al. (2012) used the SMCM to explore its capabilities in the context of improving GCM convection parametrizations by using the above mentioned two-layer model to flows about an equatorial ring. They found that using the SMCM increases the variability of tropical convection compared to a deterministic convection parametrization and that the SMCM is able to produce a realistic Walker cell circulation when forced with a longitudinal SST gradient.

One may argue that the capability of the SMCM to produce sensible results is given by its design principles, *e.g.* prescribing certain transition timescales, assuming tropical convection to be dependent on two predictors only or coupling it to a simple two-layer atmospheric model. In fact, a comparison of the SMCM simulated cloud area fractions to observational data is still outstanding. It is the aim of this study to use the SMCM in a diagnostic fashion by forcing it with an observed large-scale state to investigate the feasibility of using its underlying stochastic concept for convective parametrizations in full GCMs.

b. Two datasets of observed large-scale atmospheric state over tropical areas

We utilise two datasets comprising various quantities describing the large-scale atmospheric state over a tropical location for the purpose of this study. One dataset covers a $\approx 190 \times 190\text{km}^2$ pentagon-shaped area centered over Darwin, Australia (Jakob et al. 2011), investigated during the TWP-ICE campaign (Tropical Warm Pool - International Cloud Experiment, May et al. 2008). The size of the area is chosen to approximately represent that of a typical GCM grid-box and the grid-box mean values of atmospheric variables are computed using a variational analysis after Zhang and Lin (1997). This variational analysis is applied to a large part of three consecutive wet seasons (2004/2005, 2005/2006, 2006/2007). Over northern Australia, the wet season is defined as the time period between September of one year and April of the following year. The dataset and its documentation can be obtained via the Atmospheric Radiation Measurement (ARM) Climate Research Facility's website (http://www.arm.gov/data/pi/46) and we use all available data for the analysis presented here. Atmospheric variables are available every 6 hours. Information on clouds and precipitation is retrieved from radar observations by the C-band polarimetric (CPOL) research radar (Keenan et al. 1998) located at Gunn Point and operated by the Australian Bureau of Meteorology. From those data, rain area fractions attributable to either stratiform or convective precipitation are determined after Steiner et al. (1995) and used as a proxy for stratiform and convective cloud fractions (Kumar et al. 2012). Convective clouds are separated into congestus and deep convection according to cloud top height (CTH): convective clouds having CTHs of less than 7 km are classified as congestus whereas clouds having higher CTHs are classified as deep convective clouds (V.V. Kumar, personal communication, 2012). The dataset encompasses the period of the TWP-ICE campaign (May et al. 2008) which took place in the same area during January and February 2006. The collected data of meteorological regimes encountered during TWP-ICE have already proven to be very valuable for the evaluation of GCM convective parametrizations (*e.g.* Lin et al. 2012).

The second dataset represents the large-scale atmospheric state over Kwajalein and is obtained by applying the same variational analysis as is used for the Darwin dataset. Convective and stratiform precipitation area fractions are also calculated according to Steiner et al. (1995), congestus area fractions are however not available because the radar data available to us only consists of horizontal 2D-scans. The Kwajalein dataset covers a shorter time period (May 2008 – Jan 2009) and was produced to match the observation intensive period of the YOTC (Year Of Tropical Convection, Waliser and Moncrieff 2007) project. For better comparability, the Kwajalein data is derived for an area identical to the pentagon-shaped one over Darwin.

We use both datasets in this study to show that the functional dependency of tropical convection on a given large-scale atmospheric state is similar for both locations although they are subject to distinctly different boundary conditions, *e.g.* land-sea distribution or monsoonal forcing.

To illustrate the multitude of meteorological regimes found in the datasets, we show the time series of selected atmospheric parameters for the time period of 10 Nov 2005 – 18 April 2006 over Darwin in Fig. 1. It is evident that apart from the variability during the TWP-ICE period (19 Jan 2006 – 28 Feb 2006, May et al. 2008), the snapshot shown in Fig. 1 alone contains a number of evident meteorological-regime changes which result in distinctly different cloud populations. Characterising the middle-troposphere level, the time series of relative humidity qualitatively exemplifies "wet" periods around 20 January 2006 or 1 April 2006 (among others) and "dry" periods around 25 November 2005 or 1 March 2006 (among others) of the time series. As shown in the plot of derived convective and stratiform cloud fractions, the above mentioned wet and dry periods are each associated with specific cloud regimes: the wet periods are generally associated with higher cloud fractions compared to the dry periods. Stratiform clouds exhibit the highest cloud area fractions, with deep convective cloud fraction being about an order of magnitude less and congestus cloud fraction being again an order of magnitude less than the latter. It must be noted that the derived cloud area fractions are representative for precipitating clouds only. However, this does not present a serious issue, *i.e.* fractions of tropical congestus, deep convective or stratiform clouds derived from the scanning rain radar compare very well to those derived from a vertically pointing cloud radar (V. Kumar, pers. communication, 2012).

It should be mentioned at this point that the observational data we are comparing the SMCM-simulated cloud fractions to are also subject to uncertainties and give room for interpretation. The most prominent uncertainty is of course the estimation of rain rates from radar echoes, which is not all too straight forward itself, and the subsequent assumption that the area of a particular type of rainfall (derived after Steiner et al. 1995) is equal to the cloud fraction of that particular cloud type. Therefore, this analysis is limited to precipitating clouds only. Also, land surface characteristics of the geographical area covered by the large-scale observational dataset used in this study are far from homogeneous. The CPOL radar at Gunn Point covers both water and land surfaces, with some of the land surface areas being subject to a pronounced convective diurnal cycle which results in some of the deepest convection on the planet (Keenan et al. 1990; Crook 2001). As these events are locally driven, environmental conditions leading to their initiation cannot be represented in the observational dataset. This uncertainty in environmental conditions obviously does not apply to the Kwajalein data.

c. Deriving model forcing parameters from the observations

The evolution of the cloud ensemble as simulated by the SMCM with respect to the large scale atmospheric state is designed to be dependent on two predictors. One parameter is used as a proxy for the environment's potential to develop and sustain convection (C in the following) and the other one is used as a proxy for mid-tropospheric dryness (D in the following). Here, the underlying assumption is that convection is initiated/sustained and hindered/depleted by high values of C and D, respectively. Because we aim to use the SMCM in a diagnostic manner by forcing it with an observed large scale atmospheric state, we have to derive C and D from the available observational data. This requires to adapt the formulas for calculation C and D as given in Khouider et al. (2010) as these are defined to be used for a large scale state given by the simple two-layer model (Majda and Shefter 2001; Khouider and Majda 2006).

As mentioned above, C and D are used as proxies for the convective potential of the tropospheric column and mid-tropospheric dryness, respectively. In the original SMCM these quantities are scaled to vary roughly between 0 and 2. For the evaluation of the SMCM, we derive a total of six (instead of just two) forcing predictors. We proceed in this way because there may exist a multitude of possible predictor constellations for adequately describing the dependency of tropical convection on the large scale atmospheric state.

1) C – a proxy value for convective activity

In the original formulation given in Khouider et al. (2010), C is given by the scaled convective available potential energy (CAPE) ($\mathrm{C_C}$ in the following). CAPE corresponding to the time series shown in Fig. 1 yields values in the range from 0 – 1700 [J/kg]; we therefore scale the CAPE values by 1000 [J/kg] to achieve the desired range of $\mathrm{C_C} \in [0;2]$.

As it has been argued before that CAPE alone may not be a good proxy for characterising the occurrence of tropical convection (*e.g.* Sherwood 1999), we also define additional versions of C, represented by scaled values of either the ratio of low-level CAPE (LCAPE), *i.e.* CAPE integrated only to the freezing level, to total CAPE ($\mathrm{C_{rC}}$), or large scale vertical velocity at 500 hPa ω_{500} ($\mathrm{C_\omega}$):

$$
\begin{aligned}
\mathrm{C_{rC}} &= 2\left(\frac{\mathrm{LCAPE}}{\mathrm{CAPE}}\right) \\
\mathrm{C_\omega} &= -\left(\frac{1}{10}\ \mathrm{hPa^{-1}\ hr}\right)\omega_{500}, \quad \omega_{500} < 0
\end{aligned}
\tag{1}
$$

The choice to investigate the proxies $\mathrm{C_C}$ and $\mathrm{C_\omega}$ is relatively intuitive and straight forward, whereas the choice of $\mathrm{C_{rC}}$ warrants explanation. Khouider et al. (2010) found that assuming congestus activity being positively related to LCAPE (derived from a two-layer atmospheric model) rather than total CAPE improves the SMCM-modelled variability. However, our observations show that LCAPE alone is roughly constant throughout the whole observational period and it is only the ratio to total CAPE resembling some relationship with observed convection. For illustrative purposes, we show the time series of C for the subset of the data shown in Fig. 1 in the top two panels of Fig. 2.

Recalling the preceding short analysis of "wet" and "dry" periods (Sec. b), the pattern of C_C (2, top panel) reveals no evident correlation to these periods. The relatively high values of C_C during the first 40 days of the time series should yield intense convective activity, however, the observed cloud fractions do not support this. Furthermore, the wetter periods are characterised by low C_C values throughout. However, especially stratiform cloud fraction, most probably originating from deep convection, is high during these periods. This supports a separate analysis of the present dataset which indeed suggests that in the area of interest, convective precipitation shows no significant correlation with CAPE (Jakob et al. 2011). In fact, CAPE has been shown to be approximately anti-correlated with precipitation for a region in relatively close proximity to the area covered by our dataset (McBride and Frank 1999).

C_{rC} exhibits large values when convective activity is high (cf. Figs. 1 and 2), implying that in situations of intense convection, total CAPE is dominated by the contribution coming from below the freezing level. Because low-level CAPE itself does not vary very much, it is the lack of contributions to total CAPE coming from above the freezing level which make up for high values of C_{rC}, consistent with the findings of McBride and Frank (1999) who concluded that high values of CAPE are dominated by contributions from above 600 hPa. High values of C_{rC} thus imply that during periods of intense convection, such as those shown in Fig. 1, the specific heating profile of stratiform precipitation, *i.e.* latent heating of the upper troposphere and evaporative cooling of the lower troposphere (*e.g.* Houze 1997), serves to adjust the lapse-rate towards the moist adiabat. However, it is the occurrence of convection itself which may enforce high values of C_{rC}, resulting in possible ambiguities when attempting to use it as a predictor for convection.

From a dynamical perspective, it is well known that large-scale vertical ascent, and thus moisture convergence, is associated with and facilitates the development of deep convection (cf. the recent study of Hohenegger and Stevens 2012). Like the convective area fractions shown in Fig. 1, the time series of C_ω also appears highly intermittent and seems to very closely follow the former. This is especially true for the first ≈40 days of the time series in which the observed stratiform and convective cloud fractions are relatively low. During that particular period, C_ω shows relatively small values with higher ones occurring sparsely, indicating a weakly but somewhat constantly forced convective regime. However, ambiguities in establishing sound cause-and-effect relationships between C and convection are apparent for C_ω, which is directly related to large-scale convergence which can in turn be considered as both a cause and consequence of convective heating. In fact, discussion of these ambiguities is one of the most persistent issues in the meteorological community. Ambiguities may also arise from the method to derive C_ω itself. Vertical pressure velocity ω is the key parameter obtained from the variational analysis used to derive the large scale atmospheric state we use here. The variational analysis itself is constrained by total areal rainfall itself, thus ω is somewhat tuned to match observed rain rates. However, because we use area fractions, and not rain rates, of convective and stratiform rain in our analysis, the causal link to the data processing in the variational analysis is weak.

2) D – a proxy for mid-tropospheric dryness

In the original formulation of the SMCM, the proxy for mid-tropospheric dryness D_{θ_e} is given by

$$D_{\theta_e} = \frac{\theta_{e,\mathrm{BL}} - \theta_{e,\mathrm{m}}}{15\mathrm{K}}, \tag{2}$$

with $\theta_{e,\mathrm{BL}}$ being the equivalent potential temperature in the boundary layer, $\theta_{e,\mathrm{m}}$ the equivalent potential temperature in the mid-troposphere and 15 K a climatological mean scaling factor (Khouider and Majda 2006). Here, the underlying assumption is that the difference between the equivalent temperatures as given in Eq. 2 is large when the middle troposphere is dry compared to the boundary layer. For the calculation of D_{θ_e} from the observed large scale state, we define $\theta_{e,\mathrm{BL}}$ and $\theta_{e,\mathrm{m}}$ as the equivalent potential temperatures at 1000 hPa and 500 hPa, respectively. To yield the desired range of $D_{\theta_e} \in [0;2]$, we use a scaling factor of 10 K instead of 15 K.

Additional to the original formulation of D, we introduce a simpler proxy for representing the mid-tropospheric dryness by use of the relative humidity at 500 hPa. Then, D_{RH} is given by

$$D_{\mathrm{RH}} = 2 \cdot (1 - \mathrm{RH}_{500}), \tag{3}$$

with $\mathrm{RH}_{500} \in [0;1]$. The resulting time series of D calculated with both methods are shown in Fig. 2 (bottom).

Unlike the time series of C, the ones for D show a very high level of agreement. It is just for two short time periods where the values of D_{θ_e} and D_{RH} disagree significantly, namely around 5 February 2005 and 10 April 2006 of the time series displayed in Fig. 1. These periods are relatively dry compared to the rest of the time series, with low values of relative humidity reaching down into the boundary layer. For these two cases, relatively high values of D_{RH} indicate a "dry" case, whereas the low (or even negative) values of D_{θ_e} indicate a rather "wet" case. This is because low values of θ_e occur throughout the tropospheric column down to the surface, thereby not yielding the anticipated large difference between θ_e at 1000 and 500 hPa. Defining D_{θ_e} by Eq. 2 therefore poses a limitation for running the SMCM when using observational data. As D_{RH} agrees very well with D_{θ_e} throughout the rest of the time series, we will use D_{RH} for all further analyses presented in this study. Also, Khouider et al. (2010) used D_{θ_e} simply because it is more convenient in the context of the two-layer model.

3. The observed mean convective state at Darwin and Kwajalein

Before assessing whether the mathematical framework of the SMCM is suitable for reproducing observed convective behavior of tropical convection, we first analyse the observations laid out in Sec. b in a manner suitable for direct comparison with SMCM output. Given the specific values of the forcing parameters C and D (cf. Sec. c), the birth-death process used in the SMCM yields stationary cloud fraction distributions of every cloud type. Hence, it is possible to calculate a 2-d histogram of the stationary cloud fraction as a function of C and D. Examples of such equilibrium cloud fraction distributions for a given set of transition timescales are given in Khouider et al. (2010). Here, we therefore calculate joint histograms of observed convective and stratiform cloud fractions in the parameter space of observed values of C and D to enable a straightforward comparison between observed and modelled convective behavior.

We show such joint histograms of mean observed cloud fractions for three sets of forcing parameters, as well as their standard deviations and number of measurements, in Figs. 3 - 5, for Darwin and Kwajalein. In the three sets of forcing parameters, the mid-tropospheric dryness parameter is represented by D_{RH} and the convection parameter C is represent by either C_C, C_{rC} or C_ω. Because of the observational limitations mentioned above, we only analyse deep convective and stratiform cloud fractions and neglect congestus clouds in the context of this study.

We only discuss the results for Darwin in detail. Generally, the data for Kwajalein show the same relationships as for Darwin, but with less frequent high values of the C parameter and generally smaller stratiform cloud fractions. The important finding to keep in mind is that convective and stratiform cloud area fractions show very similar behavior at both locations given a particular large scale atmospheric state, justifying using the observations from both locations together to investigate cloud fractions simulated by the SMCM.

When we stratify the observational data using C_C as indicator for convective activity (cf. Fig. 3), we obtain maximum area fractions for both cloud types for some of the smallest values of C_C and D_{RH}, indicating relatively high convective activity for small values of CAPE and a moist middle troposphere. Most observations fall into a range spanning the lower half of both parameter ranges, also resulting in the lowest cloud area fraction variability, *i.e.* relative standard deviation, in that range. Similar results are presented in McBride and Frank (1999) who found an inverse relationship between CAPE and precipitation when analysing data obtained during active and break monsoon periods for a location in the Gulf of Carpentaria.

When stratifying the observations according to either one of the other two choices for C (cf. Figs. 4 and 5), we obtain a completely different functional dependency of convective and stratiform cloud fractions on C and D. Using C_{rC} and C_ω as choices for C lead to

i) maximum values for both cloud area fractions for highest values of C,

ii) high and low cloud area fraction variability for low and high values of C, respectively,

iii) a sharp increase in cloud area fractions above a certain value of C

iv) most observations for low values of C spanning a wide range of D_{RH}-values.

The results give valuable insight into tropical convective behavior. For weak forcing of convective activity, *i.e.* small values of C, average cloud area fractions are small but exhibit large variability, indicating a somewhat stochastic behavior. This is particularly interesting because a large part of the observations yield such weak forcing which would normally act to reduce sample variability. The stronger the forcing of convective activity gets, the less observations are registered per bin, suggestive of an expected increase in sample variability. However, cloud area fraction variability is lowest for strong forcing of convection, suggesting a more and more deterministic behavior of convection with increasing forcing, in line with other results derived from the same dataset (Jakob et al. 2011). Physically, this implies that as forcing is weak, convection occurs more randomly in the domain, inducing large-scale convergence itself which then enables stronger convective features to form. These results however do not support the idea that the stochastic component of unresolved subgrid-scale processes scales linearly with their mean response as put forward in earlier studies (*e.g.* Buizza et al. 1999; Shutts and Palmer 2007). The sharp increase in cloud area fraction above a certain value of C is consistent with the "threshold-behavior" of convection as laid out in *e.g.* Peters and Neelin (2006). Furthermore, the histograms we show in Figs. 4 and 5 indicate that at least for these two choices of C, deep convective as well as stratiform area fractions are anti-correlated with dryness at mid-levels, broadly consistent with earlier findings from observational studies (Redelsperger et al. 2002; Derbyshire et al. 2004; Takemi et al. 2004; Takayabu et al. 2010).

The increase in cloud area fractions also appears to occur rapidly above a certain value of C, supporting earlier findings of critical behavior in tropical convection (*e.g.* Peters and Neelin 2006). We also note that regimes exhibiting both a strong forcing of convection and a dry middle troposphere basically do not exist at the locations considered in this study. This may be obvious, but such a result is not apparent from Fig. 3 where there still exist a quite large number of measurements yielding a combination of a dry middle troposphere and high values of C_C.

4. Reproducing observed convective behavior using the SMCM

a. *Adjusting the model parameters*

The equilibrium cloud fractions of the multistate Markov chain used in the SMCM are calculated by analytically determining its stationary equilibrium distribution (cf. Khouider et al. (2010) for details). In this case, the equilibrium distribution is represented by area fractions for each of the four allowed states of the Markov chain, *i.e.* either clear sky, congestus, deep convection or stratiform clouds. The sum of all four area fractions for each pair of discrete C and D values is 1 and the distribution of area fractions among the four states can be adjusted by manipulating the transition timescales associated with the transition from one state to another.

In previous publications, the transition timescales used in the SMCM were chosen in an either ad-hoc, but physically meaningful manner (Khouider et al. 2010, KBM10) or to improve the intermittency of the simulated convection in idealised experiments (Frenkel et al. 2012, FMK12). Here we use observations to gauge the applicability of the chosen timescales to represent observed convective behavior. For reference purposes, we show the joint histograms of the analytically derived equilibrium deep convective area fractions for the transition timescales introduced in KBM10 and FMK12 (cf. Tab. 1) in Fig. 6. These joint histograms clearly indicate that the previously used transition timescales are not suited for reproducing the statistics of observed convection laid out in Sec. 3 for several reasons. First, the transitions used in case 1 of KBM10 and in FMK12 yield equilibrium deep convective area fractions about an order of magnitude larger than those observed. Second, the transition timescales used in case 2 of KBM10 result in a deep convective area distribution unsuitable for reproducing observed behavior.

To obtain a model which is most suitable for reproducing the observed convective behavior, we systematically adjust the transition timescales until we arrive at a close visual match between the analytical equilibrium solution of the SMCM and the observed mean deep convective cloud fractions for each convective proxy (C_{C}, C_{rC}, C_{ω}) for Darwin shown in Figs. 3 - 5 (we only use data for Darwin here to test the robustness of the adjusted transition timescales by applying it to the Kwajalein data in the next section). This close match should ideally agree to the general cloud fraction distribution in C-D-space in both magnitude and shape. Additionally, the equilibrium area fraction calculated for the mean observed C and D values (black dots in Figs. 3 – 5) should also match closely. The second requirement achieves a tuning of the model to the "mean observed climate", thus yielding an optimal representation of observed tropical convective cloud distribution – given that the cloud-type relationships imposed in the SMCM correspond to those in nature. We find that it proves difficult to adequately satisfy both conditions, leading to a trade-off of getting either the mean climate or the maxima right. In general, we focus on arriving at the correct mean climate cloud fractions as this is of higher relevance regarding a possible future implementation into GCMs. The final "best-fit" transition timescales for each convective proxy C are listed in Tab. 1 and a comparison of modeled equilibrium- and observed mean deep convective area fractions as f(C,D) is displayed in Fig. 7.

As expected from the observed mean cloud fractions as f(C,D), we find that matching the SMCM-modelled equilibrium cloud fractions to the mean CAPE-stratified observed cloud fractions results in starkly different timescales compared to the other three convection proxies (Tab. 1). However, all three sets of best-fit transition timescales preserve an important constraint laid out in KBM10, namely that cloud decay acts on identical or longer timescales than cloud formation. It must be kept in mind that these best-fit timescales were found by visually matching the joint histograms of modeled and observed area fractions, though.

The joint histograms displayed in Fig. 7 indicate that each of the three analytical equilibrium deep convective area distributions corresponding to the "best-fit" transition timescales in Tab. 1 has some difficulty in reproducing certain aspects of the corresponding observations at Darwin. For every version of C, the model overestimates deep convective area fraction for almost the entire range of considered combinations of C and D.

This overestimation is highest when using C_{rC} to stratify the observations, however the overall functional relationship is captured (cf. Fig. 4). Using observations stratified by C_C to adjust the transition timescales yields higher modeled area fractions at nearly every considered C,D pair, with the degree of overestimation showing no functional dependence on C and D. Using C_ω, the SMCM's equilibrium distribution resembles the functional dependency of the observations well. Furthermore, the relative difference of modeled versus observed area fractions shows an evident dependency on C and D. The model over- and underestimates deep convective area fractions for low and high values of C, respectively. This transition from over- to underestimating the area fractions appears systematic and gradual – a promising result in terms of possible future model adjustments (see below). The modeled joint histograms in Fig. 7 however do not show the capability of the SMCM concept to reproduce observed temporally resolved tropical convection; they are merely analytical solutions of the SMCM's internal birth-death process.

We conjecture that the main reason why the SMCM over- and underestimates deep convective area fraction for low and high values of C_ω (and C_{rC}), respectively, is not a matter of finding the correct transition timescales or of ill-formulated "transition rules", but due to the functional dependency of transition rates on C and D. Khouider et al. (2010) formulate this dependency as

$$\Gamma(x) = 1 - e^{-x}, \quad x \in [0; 2] \tag{4}$$

with x being either C or D and Eq. 4 being directly linked to transition rates R, *e.g.*

$$R_{ab} \propto \Gamma(\mathrm{C})\Gamma(\mathrm{D}), \tag{5}$$

being the transition rate R from cloud state a to b. This formulation leads pronounced changes in transition rates for small values of C or D with the response becoming less strong with increasing values of C and D. Therefore, the SMCM in its original formulation is not designed to reproduce the sharp increase in observed cloud fractions shown in Figs. 4 and 5 for higher values of C. Alternative formulations of $\Gamma(x)$ could be sought to improve the SMCM's capability to reproduce observed cloud are fraction distributions. This will be investigated in future research.

b. Applying the SMCM to observations

In this section, we use the three sets of observation-derived parameters discussed in Secs. c and 3 in combination with the "best-fit" transition timescales shown in Tab. 1 to perform simulations with the SMCM. We first quantitatively discuss the temporally resolved reproduction of cloud area fractions compared to observations in Sec. 1 and then carry out a more thorough statistical analysis in Sec. 2.

1) SMCM-modeled temporally resolved tropical convection

We use the subsets of the data from the Darwin and Kwajalein locations introduced in Sec. b to compare the time series of observed cloud area fractions to those modelled by the SMCM for illustrative purposes. As we obtained the "best-fit" transition timescales shown in Tab. 1 from analysing just Darwin data, application of these timescales to Kwajalein provides a strong test for our method. We force the SMCM with each of the three combinations of $\mathrm{C_C}$, $\mathrm{C_{rC}}$ and $\mathrm{C_\omega}$ with $\mathrm{D_{RH}}$. The internal model time step is set to 5 minutes. The 6-hourly observations were linearly interpolated to match the model time step. The subgrid-scale lattice of the SMCM is set up to have 20×20 sites. As the whole domain covers an area of $\approx 190 \times 190$ km^2, each lattice site thus has an edge length of about 10 km. There is currently no fixed spatial scale for an individual lattice point considered in the formulation of the SMCM. Preliminary analysis shows that an increase in lattice sites, and the reduction of lattice size going with it, reduces the simulated temporal variability compared to observations but has no effect on correlations. From a GCM parameterization perspective, a high number of lattice points with fixed spatial scale per GCM grid box would lead to increasing convective variability with increasing resolution, thus yielding a more realistic representation of convection compared to current deterministic schemes.

The resulting modelled time series of deep convective cloud area fractions for Darwin and Kwajalein are shown in Figs. 8 and 9, with the observed time series included for reference purposes. We show neither observed and modelled congestus nor stratiform cloud fractions because our main interest lies in assessing the representation of deep convection as this is our current target for GCM convection parametrizations.

We first consider the observed and modeled deep convective area fractions over Darwin shown in Fig. 8 as we have adjusted the model parameters of the SMCM specifically for this location. Forcing the SMCM with C_C results in more or less constant convective cloud area fractions showing no resemblance of the different regimes found in the observations. Due to the non-negative and mostly non-zero values of the C_C timeseries (cf. Fig. 2), the SMCM cannot reproduce the intermittency of cloud area fractions found in the observations. The same issue is apparent when forcing the SMCM with C_{rC}. However, periods of higher modelled deep convective cloud fraction seem to loosely correspond to periods of higher observed fractions, giving slightly more confidence in using C_{rC} over C_C.

The results from using C_ω to force the SMCM show substantially more agreement with the observations, with C_ω leading to more variability during periods of low convective activity, especially during the first month or so of the considered time period. Despite these encouraging results, the issues raised towards the end of Sec. 4 are apparent. For periods of weak forcing, the SMCM produces too high a deep convective cloud fraction whereas cloud fractions during strongly forced periods are substantially underestimated compared to observations. This is exactly what is to be expected from the modelled equilibrium cloud fractions shown in Fig. 7.

The observed and modeled time series of deep convective area fraction for the Kwajalein area (Fig. 9) generally show the same behavior as the ones for the Darwin area (Fig. 8). Especially the over- and underestimation of deep convective area fractions for small and large values of C_ω, respectively, is evident. Nevertheless, C_ω proves to be the parameter of choice for reproducing deep convective features over Kwajalein with the SMCM. Considering that we did not use the Kwajalein data to adjust the transition timescales in the SMCM, this result confirms the findings presented in Sec. 3, namely that convection over Kwajalein shows similar functional dependencies to the large scale environment as convection over Darwin. Furthermore, this result indicates that at least in the framework of the SMCM, tropical convection acts on similar timescales for both tropical locations considered here. It is however important to keep in mind the possible ambiguities when attempting to establish cause-and-effect relationships between the large-scale state and convection when using C_ω (cf. Sec. 3).

2) Statistics of SMCM-modeled versus observed tropical convection

We now analyse the SMCM-modeled tropical convection to quantify the capability of the SMCM framework to reproduce the observed statistical properties of deep convective and stratiform area fractions laid out in Sec. 3 as well as the actual stochasticity of the modeled convection. For the sake of brevity, we limit this analysis to experiments in which convection in the SMCM is determined by C_ω. We choose to do so because the SMCM-versions using the two other parameters C_{C} and C_{rC} were shown unsuitable for reproducing basic temporal behavior of convection (cf. Sec. 1).

Similar to the analysis of observed convection presented in Sec. 3, we stratify the modeled time series of deep convective and stratiform area fractions by the values of C_ω and D_{RH} used for forcing the model. To ensure comparability with the observations, we average the modeled area fractions over 6-hour periods centered over each time step of the observed large scale atmospheric state. Similar to the histograms shown in Figs. 3 – 5, we show the results obtained for Darwin and Kwajalein separately in Fig. 10, again providing a test for the validity of the chosen transition time scales for both locations.

As expected, the joint histogram of SMCM-modeled deep convective area fractions obtained from the modeled time series of the Darwin location very much resemble that of the analytically derived equilibrium area fraction for the same set of transition time scales (Fig. 7, bottom). These statistics of the modeled time series more clearly reveal the shortcomings of the SMCM framework in reproducing observed convection already mentioned in Secs. a and 1. The order of magnitude of deep convective area fraction is generally well captured, with the SMCM over- and underestimating area fractions for weak- and strong convective forcing, respectively. The same also holds for the simulated stratiform cloud fractions for the Darwin area, which we show here for illustrative purposes, mainly to highlight that the transition time scales we determined in Sec. a also yield sensible values for that cloud type. More importantly, the sample standard deviations of deep convective and stratiform area fractions of the modeled time series show similar behavior compared to those of the observations, *i.e.* area fractions show higher and lower variability for weaker and stronger convective forcing, respectively. The modeled time series underestimate the degree of variability throughout, though. So for the Darwin area, the SMCM framework is suitable for reproducing observed behavior of tropical convection, both in terms of deep convective and stratiform cloud area fractions and variability, as a function of the observed large scale environment.

For the Kwajalein area, the joint histograms in Fig. 10 lead us to similar conclusions, thereby supporting the applicability of the SMCM framework to both tropical locations considered here. However, due to the sparse sampling of strong convective forcing over Kwajalein, the overestimation of cloud area fractions for weak convective forcing dominates the statistics. As mentioned in Sec. a, the sometimes substantial overestimation of cloud area fractions could be mediated by using alternative formulations of Eq. 4, which will be a topic of future research.

5. Summary and Conclusions

This study was driven by the need for alternatives to the mostly deterministic convection parametrizations used in general circulation models (GCMs). For this, we first determined statistics of observed tropical convection over Darwin and Kwajalein stratified by environmental conditions. Then, we used these observed statistics to investigate whether the underlying framework of the Stochastic MultiCloud Model (SMCM Khouider et al. 2010) is suitable for reproducing observed tropical convection – a prerequisite to using the underlying stochastic framework of the SMCM in a GCM convection parametrization.

We investigated the dependency of tropical convection, given by the fractional area coverage with deep convective or stratiform clouds, on a set of two proxy values obtained from the observed large-scale atmospheric state (derived by means of variational analysis (Jakob et al. 2011)). One proxy (C) represents the ability of the atmospheric column to initiate/sustain convection whereas the second proxy (D) represents mid-tropospheric dryness. As there exists no generally accepted theory of which environmental conditions actually lead to tropical convection, we used three different formulations for C: CAPE, the ratio of low-level CAPE (CAPE integrated up to the freezing level, LCAPE) to CAPE and vertical velocity at 500 hPa. D is obtained from relative humidity at 500 hPa.

We found that the relationship of observed cloud area fractions with CAPE is very different compared to the other two C-proxies. We find highest deep convective and stratiform cloud area fractions for low values of CAPE, supporting earlier findings that CAPE is approximately anti-correlated with tropical precipitation (McBride and Frank 1999). On the other hand, deep convective and stratiform cloud area fractions are positively correlated with the other two C-proxies. The cloud area fraction distributions as function of C and D also revealed that for those two C-proxies,

i) high and low cloud area fraction variability occurs for low and high values of C, respectively, implying that convection appears more random under weakly forced conditions and gets more and more deterministic with increasing forcing (consistent with earlier findings from the same dataset, Jakob et al. 2011), thus contradicting the idea that the stochastic component of unresolved subgrid-scale processes scales linearly with their mean response (*e.g.* Buizza et al. 1999; Shutts and Palmer 2007),

ii) cloud area fractions increase sharply above a certain value of C, consistent with earlier reports on critical behavior of tropical convection (*e.g.* Peters and Neelin 2006),

iii) cloud area fractions show identical relationships to environmental conditions for both locations (Darwin and Kwajalein), albeit starkly different boundary conditions (*e.g.* land-sea distribution, monsoonal forcing),

iv) deep convective and stratiform cloud area fractions are anti-correlated with mid-tropospheric dryness (consistent with Redelsperger et al. 2002; Derbyshire et al. 2004; Takemi et al. 2004; Takayabu et al. 2010).

By design, the SMCM has a stationary equilibrium cloud area fraction distribution. By adjusting this distribution to the mean observed cloud area fractions, we tuned the SMCM for it to potentially reproduce the observed convection most closely. It proved difficult to exactly match the mean observed cloud area fraction distribution as f(C,D), especially for the data stratified by CAPE. Generally, the SMCM yields too high and too low a cloud fraction for weak and strong large-scale forcing, respectively. We found that the values of the tuning parameters leading to a sensible match to the observed convection also respect the general rules for cloud transition probabilities laid out in Khouider et al. (2010) – an overall very encouraging result.

Using the parameter-adjusted SMCM, we simulated convective area fractions using the time series of the observed large-scale state. We thus applied the SMCM in a diagnostic fashion and found that the modelled area fractions of deep convective and stratiform clouds compare better to observations when using the convection proxies related to convergence, *i.e.* vertical velocity at 500 hPa, rather than those related to stability, *i.e.* total CAPE and the ratio of low-level to total CAPE. This is most probably related to the non-intermittent and positive-definite nature of the latter proxies which does not allow for simulation of the intermittent cloud features found in the observations.

When using the convergence-based convection proxy to force the SMCM to generate time series of tropical convection, we found that the framework of the SMCM is capable of reproducing the overall functional relationships as well as the statistics of observed tropical convection well. In particular, the SMCM-modeled tropical convection also shows higher variability in weakly forced conditions compared to stronger forced conditions. The degree of variability is underestimated compared to observations, though. We conjecture that the variability of the modeled convection would be higher if the SMCM were used in a prognostic framework rather than the diagnostic framework we applied it to in this study. Furthermore, the 6-hourly time step of the observed large-scale state that we employ here may smear out part of the convective-scale variability, thus possibly constraining the stochastic process employed in the SMCM too strongly.

We acknowledge that there do exist ambiguities in establishing sound cause-and-effect relationships when attempting to relate tropical convection to large-scale convergence. We will investigate whether convergence serves as adequate predictor in a prognostic framework, rather than a diagnostic one as applied in this study, in upcoming work. Furthermore, future work will investigate the sensitivity of modeled cloud fractions to the number of sub-grid lattice sites, *i.e.* attaching spatial and temporal scales to the simulated processes.

This study has shown that the stochastic concept behind the SMCM has potential to underpin novel convection parametrizations in GCMs. As mass-flux convection parametrizations need to predict the vertical mass-flux at cloud base, the concept of the SMCM would yield the area and the updraft velocity could be given by another adequate formulation, *e.g.* such as that introduced in Jakob and Siebesma (2003). Ultimately, future efforts will converge towards implementing a prototype version of such a parametrization into a full GCM.

Acknowledgments.

The research of L.D. is supported by the Office of Science (BER), U.S. Department of Energy under grant DE-FG02-09ER64742. The research of B.K. is supported by the Natural Sciences and Engineering Council of Canada.

REFERENCES

Arakawa, A., 2004: The cumulus parameterization problem: Past, present, and future. *J. Climate*, **17 (13)**, 2493–2525.

Bright, D. and S. Mullen, 2002: Short-range ensemble forecasts of precipitation during the southwest monsoon. *Weather Forecast.*, **17 (5)**, 1080–1100.

Buizza, R., M. Milleer, and T. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Q. J. Roy. Meteor. Soc.*, **125 (560)**, 2887–2908.

Charney, J., et al., 1979: *Carbon Dioxide and Climate:A Scientific Assessment*. The National Academies Press.

Crook, N., 2001: Understanding Hector: The dynamics of island thunderstorms. *Mon. Wea. Rev.*, **129 (6)**, 1550–1563.

Dai, A., 2006: Precipitation characteristics in eighteen coupled climate models. *J. Climate*, **19 (18)**, 4605–4630, doi:10.1175/JCLI3884.1.

Derbyshire, S., I. Beau, P. Bechtold, J. Grandpeix, J. Piriou, J. Redelsperger, and P. Soares, 2004: Sensitivity of moist convection to environmental humidity. *Quart. J. Roy. Meteor. Soc.*, **130 (604)**, 3055–3079.

Dorrestijn, J., D. Crommelin, A. Siebesma, and H. Jonker, 2012: Stochastic parameterization of shallow cumulus convection estimated from high-resolution model data. *Theor. Comp. Fluid Dyn.*, 1–16, doi:10.1007/s00162-012-0281-y.

Frenkel, Y., A. Majda, and B. Khouider, 2012: Using the Stochastic Multicloud Model to Improve Tropical Convective Parameterization: A Paradigm Example. *J. Atmos. Sci.*, **69**, 1080–1105, doi:10.1175/JAS-D-11-0148.1.

Frenkel, Y., A. Majda, and B. Khouider, 2013: Stochastic and Deterministic Multicloud parameterizations for tropical convection. *Clim. Dynam.*, in press.

Gillespie, D., 1975: An Exact Method for Numerically Simulating the Stochastic Coalescence Process in a Cloud. *J. Atmos. Sci.*, **32**, 1977–1989.

Groenemeijer, P. and G. C. Craig, 2012: Ensemble forecasting with a stochastic convective parametrization based on equilibrium statistics. *Atmos. Chem. Phys.*, **12 (10)**, 4555–4565, doi:10.5194/acp-12-4555-2012.

Hohenegger, C. and B. Stevens, 2012: Preconditioning deep convection with cumulus congestus. *J. Atmos. Sci.*, doi:10.1175/JAS-D-12-089.1, in press.

Houze, R., 1997: Stratiform precipitation in regions of convection: A meteorological paradox ? *Bull. Amer. Meteor. Soc.*, **78**, 2179–2196.

Jakob, C., L. Davies, V. Kumar, and P. May, 2011: Representing convection in models – How stochastic does it need to be? *Proceedings of the ECMWF Workshop on "Representing Model Uncertainty and Error in Weather and Climate Prediction"*, ECMWF.

Jakob, C. and A. Siebesma, 2003: A New Subcloud Model for Mass-Flux Convection Schemes: Influence on Triggering, Updraft properties, and Model Climate. *Mon. Wea. Rev.*, **131 (11)**, 2765–2778.

Keenan, T., K. Glasson, F. Cummings, T. Bird, J. Keeler, and J. Lutz, 1998: The BMRC/NCAR C-band polarimetric (C-Pol) radar system. *J. Atmos. Ocean Tech.*, **15 (4)**, 871–886.

Keenan, T. D., B. R. Morton, X. S. Zhang, and K. Nyguen, 1990: Some characteristics of thunderstorms over Bathurst and Melville Islands near Darwin, Australia. *Q.J.R. Meteorol. Soc.*, **116 (495)**, 1153–1172, doi:10.1002/qj.49711649508.

Khouider, B., J. Biello, and A. Majda, 2010: A stochastic multicloud model for tropical convection. *Commun. Math. Sci.*, **8 (1)**, 187–216.

Khouider, B. and A. Majda, 2006: A simple multicloud parameterization for convectively coupled tropical waves. Part I: Linear analysis. *J. Atmos. Sci.*, **63 (4)**, 1308–1323, doi:10.1175/JAS3677.1.

Khouider, B. and A. Majda, 2008a: Equatorial convectively coupled waves in a simple multicloud model. *J. Atmos. Sci.*, **65 (11)**, 3376–3397.

Khouider, B. and A. Majda, 2008b: Multicloud models for organized tropical convection: Enhanced congestus heating. *J. Atmos. Sci.*, **65 (3)**, 895–914.

Khouider, B., A. Majda, and M. Katsoulakis, 2003: Coarse-grained stochastic models for tropical convection and climate. *Proc. Natl. Acad. Sci.*, **100 (21)**, 11 941–11 946.

Klocke, D., R. Pincus, and J. Quaas, 2011: On constraining estimates of climate sensitivity with present-day observations through model weighting. *J. Climate*, **24 (23)**, 6092–6099, doi:10.1175/2011JCLI4193.1.

Kumar, V., A. Protat, P. May, C. Jakob, G. Penide, S. Kumar, and L. Davies, 2012: On the effects of large-scale environment and surface types on convective cloud characteristics over Darwin, Australia. *Mon. Wea. Rev.*, doi:10.1175/MWR-D-12-00160.1, in press.

Lin, J. and J. Neelin, 2000: Influence of a stochastic moist convective parameterization on tropical climate variability. *Geophys. Res. Lett.*, **27 (22)**, 3691–3694.

Lin, J. and J. Neelin, 2003: Toward stochastic deep convective parameterization in general circulation models. *Geophys. Res. Lett*, **30 (4)**, 1162, doi:10.1029/2002GL016203.

Lin, Y., et al., 2012: TWP-ICE global atmospheric model intercomparison: Convection responsiveness and resolution impact. *J. Geophys. Res.*, **117 (D9)**, D09 111, doi:10.1029/2011JD017018.

Majda, A., C. Franzke, and B. Khouider, 2008: An applied mathematics perspective on stochastic modelling for climate. *Philos. T. R. Soc. A*, **366 (1875)**, 2427–2453.

Majda, A. and B. Khouider, 2002: Stochastic and mesoscopic models for tropical convection. *Proc.Nat. Acad. Sci.*, **99 (3)**, 1123–1128.

Majda, A. and M. Shefter, 2001: Models for stratiform instability and convectively coupled waves. *J. Atmos. Sci.*, **58 (12)**, 1567–1584.

May, P., J. Mather, G. Vaughan, C. Jakob, G. McFarquhar, K. Bower, and G. Mace, 2008: The tropical warm pool international cloud experiment. *Bull. Am. Meteorol. Soc*, **89 (5)**, 629–646, doi:10.1175/BAMS-89-5-629.

McBride, J. and W. Frank, 1999: Relationships between Stability and Monsoon Convection. *J. Atmos. Sci.*, **56 (1)**, 24–36.

Meehl, G., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. Mitchell, R. Stouffer, and K. Taylor, 2007: The WCRP CMIP3 multi-model dataset: A new era in climate change research. *B. Am. Meteorol. Soc.*, **88**, 1383–1394.

Murphy, J., D. Sexton, D. Barnett, G. Jones, M. Webb, M. Collins, and D. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430 (7001)**, 768–772, doi:10.1038/nature02771.

Palmer, T., 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Q. J. Roy. Meteor. Soc.*, **127 (572)**, 279–304.

Peters, O. and J. Neelin, 2006: Critical phenomena in atmospheric precipitation. *Nat. Phys.*, **2 (6)**, 393–396, doi:10.1038/nphys314.

Pincus, R., C. Batstone, R. Hofmann, K. Taylor, and P. Glecker, 2008: Evaluating the present-day simulation of clouds, precipitation, and radiation in climate models. *J. Geophys. Res.*, **113**, D14 209, doi:10.1029/2007JD009334.

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

Randall, D., et al., 2007: Climate Models and Their Evaluation. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Redelsperger, J., D. Parsons, and F. Guichard, 2002: Recovery processes and factors limiting cloud-top height following the arrival of a dry intrusion observed during TOGA COARE. *J. Atmos. Sci.*, **59 (16)**, 2438–2457.

Sherwood, S., 1999: Convective precursors and predictability in the tropical western pacific. *Mon. Wea. Rev.*, **127 (12)**, 2977–2991.

Shutts, G. and T. Palmer, 2007: Convective forcing fluctuations in a cloud-resolving model: Relevance to the stochastic parameterization problem. *J. Climate*, **20 (2)**, 187–202.

Steiner, M., R. Houze, and S. Yuter, 1995: Climatological characterisation of three-dimensional storm structure from operational radar and rain gauge data. *J. Appl. Meteorol.*, **34 (9)**, 1978–2007.

Takayabu, Y., S. Shige, W. Tao, and N. Hirota, 2010: Shallow and deep latent heating modes over tropical oceans observed with TRMM PR spectral latent heating data. *J. Climate*, **23 (8)**, 2030–2046, doi:10.1175/2009JCLI3110.1.

Takemi, T., O. Hirayama, and C. Liu, 2004: Factors responsible for the vertical development of tropical oceanic cumulus convection. *Geophys. Res. Lett.*, **31 (11)**, L11 109.

Teixeira, J. and C. Reynolds, 2008: Stochastic nature of physical parameterizations in ensemble prediction: A stochastic convection approach. *Mon. Wea. Rev.*, **136 (2)**, 483–496.

Tompkins, A. and J. Berner, 2008: A stochastic convective approach to account for model uncertainty due to unresolved humidity variability. *J. Geophys. Res.*, **113 (D18)**, D18 101, doi:10.1029/2007JD009284.

Waliser, D. and M. Moncrieff, 2007: Year of Tropical Convection-A joint WCRP-THORPEX activity to address the challenge of tropical convection. *GEWEX News*, **17 (2)**, 8–9.

Zhang, M. and J. Lin, 1997: Constrained variational analysis of sounding data based on column-integrated budgets of mass, heat, moisture, and momentum: Approach and application to ARM measurements. *J. Atmos. Sci.*, **54 (11)**, 1503–1524.

List of Tables

TABLE 1. Transition timescales in [hours] as used in the SMCM. The three leftmost columns contain the transition timescales introduced in previous studies (KBM10,FMK12), yielding the equilibrium deep convective area fraction distributions in Fig. 6. The three rightmost columns contain the visually derived "best fitting" transition timescales for each of the three convection proxies leading to the modeled equilibrium cloud fractions in Fig. 7.

	KBM10		FMK12	this study		
Process	case 1	case 2		C_C	C_{rC}	C_ω
formation of congestus (τ_{01})	1	3	1	1	1	1
decay of congestus (τ_{10})	5	2	1	1	1.2	1.2
conversion of congestus to deep (τ_{12})	1	2	1	3	1.2	1.2
formation of deep (τ_{02})	2	5	3	4	2.2	2.2
conversion of deep to stratiform (τ_{23})	3	0.5	3	0.13	0.16	0.16
decay of deep (τ_{20})	5	5	3	5	2.2	2.4
decay of stratiform (τ_{30})	5	24	5	5	4	4

List of Figures

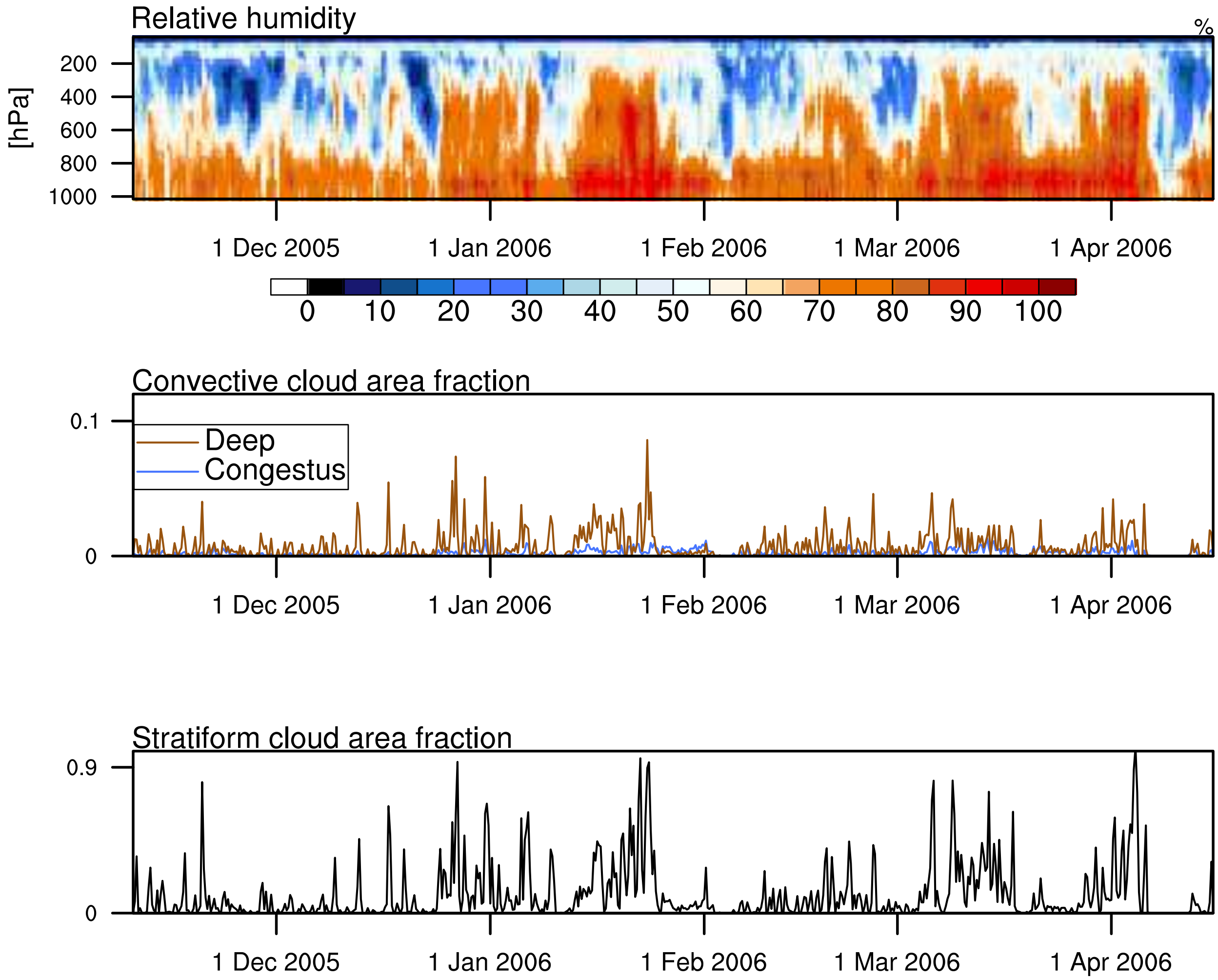

FIG. 1. Subset of the dataset comprising the atmospheric large scale state over Darwin as used in this study. Time series covering the time period from 10 Nov 2005 – 15 Apr 2006 showing vertically resolved relative humidity (top) as well as convective (middle) and stratiform (bottom) cloud fractions obtained from a scanning rain radar situated at Darwin, Australia (bottom). See text for details.

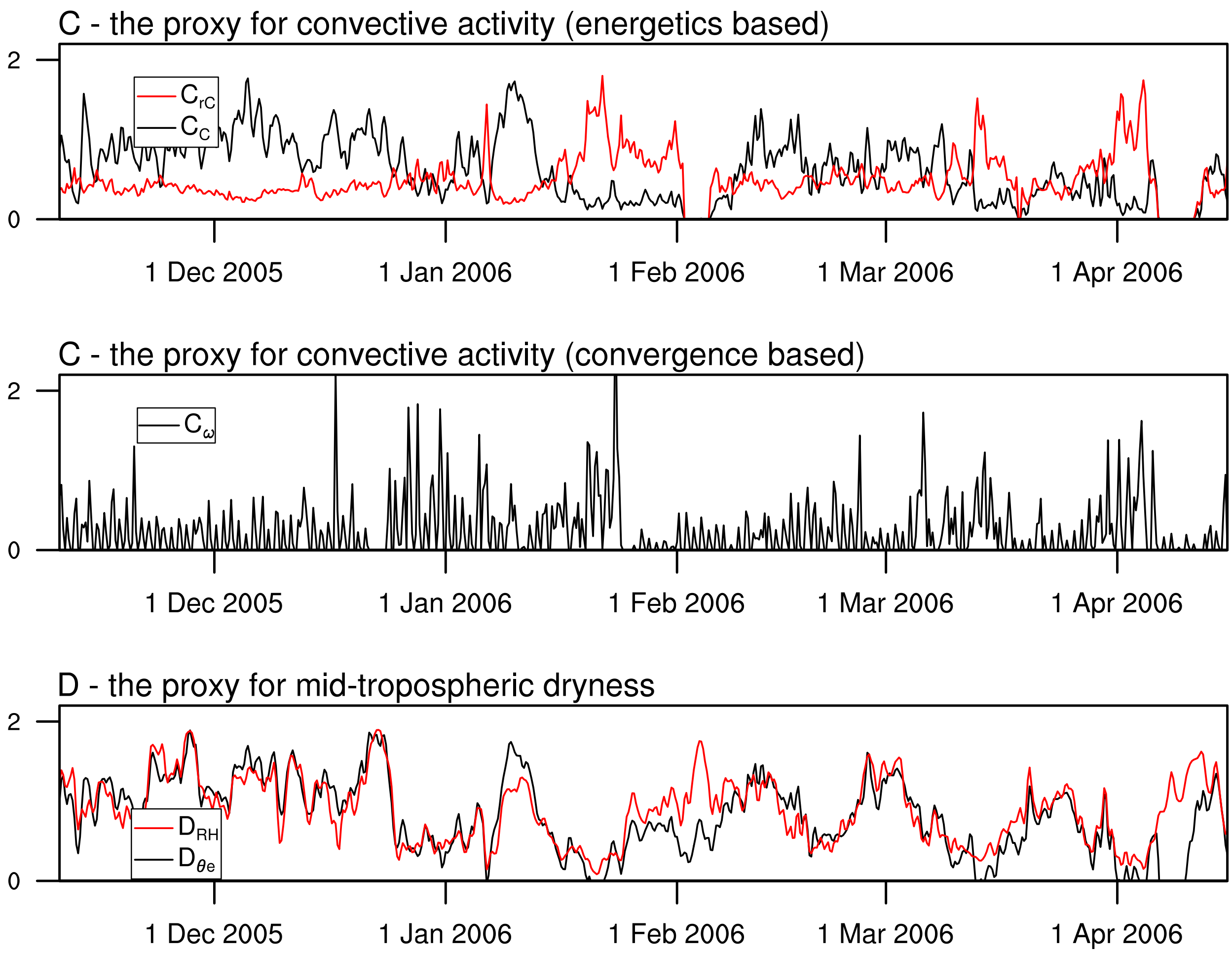

FIG. 2. Time series of model forcing predictors obtained from the large scale state shown in Fig. 1. The top two panels show values for C, *i.e.* the proxy for convective activity. The bottom panel shows values for D, *i.e.* the proxy for mid-tropospheric dryness. See text for calculation of the predictors.

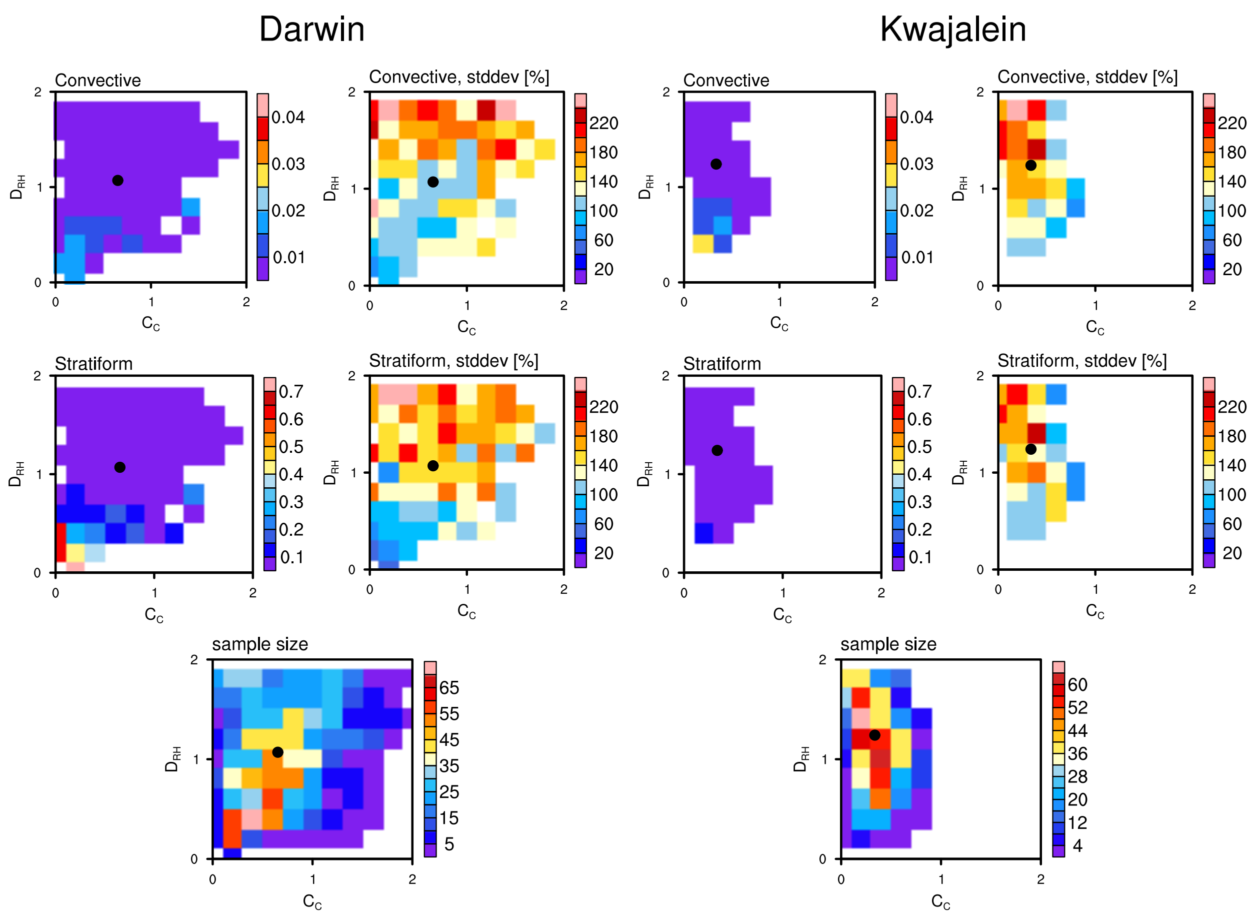

FIG. 3. Joint histogram of observed cloud area fractions and relative standard deviations as function of large scale variables C_C and D_{RH} at the Darwin (left two columns) and the Kwajalein (right two columns) sites. Only pixels having more than 5 observations are shown. Top: deep convective clouds, middle: stratiform clouds, bottom: sample size per bin. The black markers denote the mean values of C_C and D_{RH}.

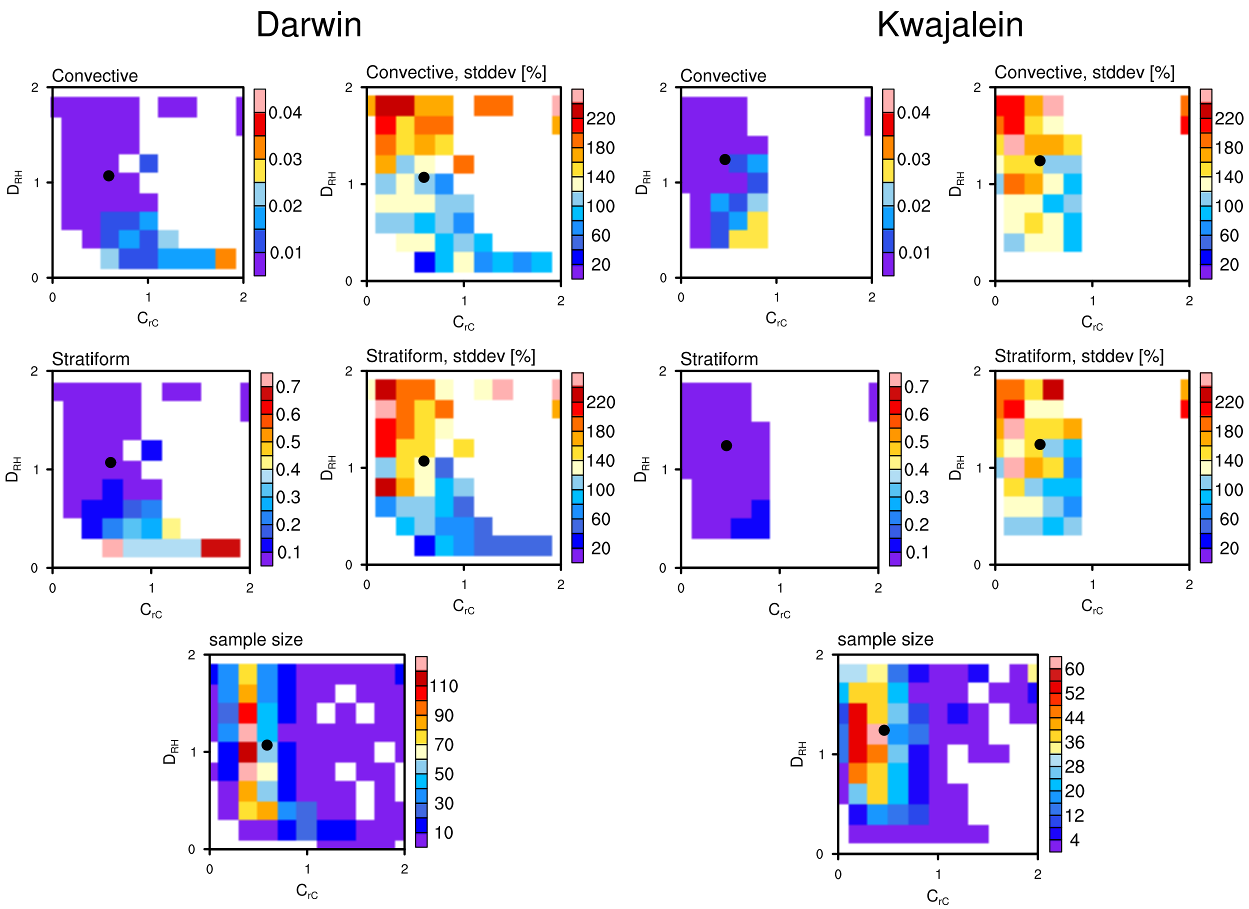

FIG. 4. Joint histogram of observed cloud area fractions and relative standard deviations as function of large scale variables C_{rC} and D_{RH} at the Darwin (left two columns) and the Kwajalein (right two columns) sites. Only pixels having more than 5 observations are shown. Top: deep convective clouds, middle: stratiform clouds, bottom: sample size per bin. The black markers denote the mean values of C_{rC} and D_{RH}.

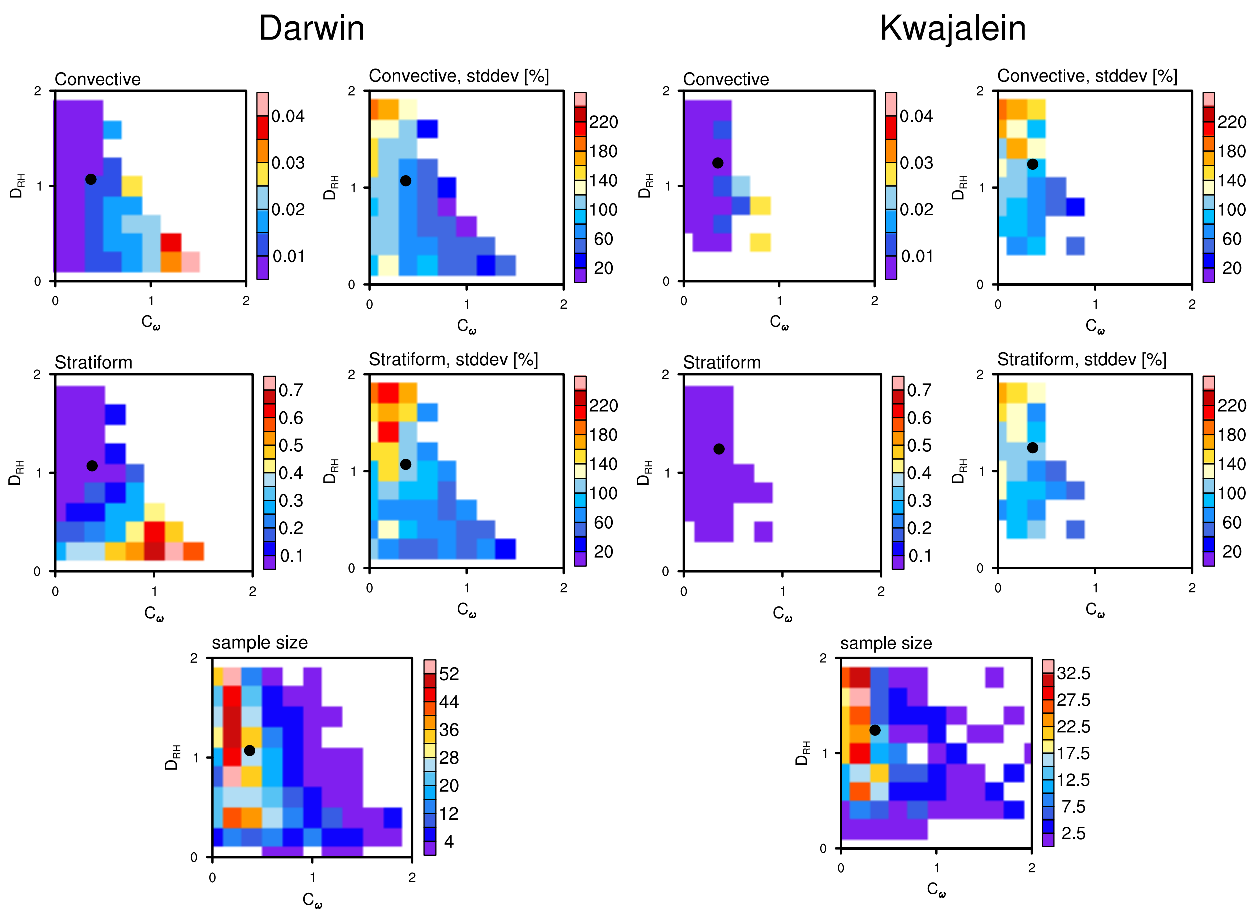

FIG. 5. Joint histogram of observed cloud area fractions and relative standard deviations as function of large scale variables C_ω and D_{RH} at the Darwin (left two columns) and the Kwajalein (right two columns) sites. Only pixels having more than 5 observations are shown. Top: deep convective clouds, middle: stratiform clouds, bottom: sample size per bin. The black markers denote the mean values of C_ω and D_{RH}.

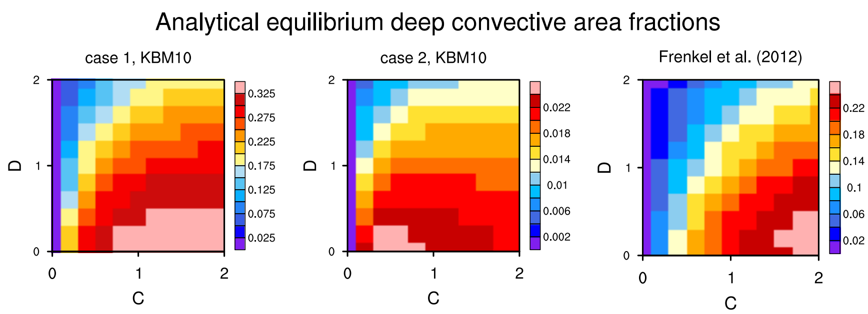

FIG. 6. Analytical equilibrium deep convective area fraction of the SMCM's birth-death process given the two sets of transition timescales introduced in KBM10 and FMK12 (Tab. 1). Left and middle: case 1 and 2 timescales of KBM10, respectively. Right: timescales used in FMK12. For the two cases of KBM10, the transition from deep convective to stratiform area depends on C. See text and Khouider et al. (2010) for details regarding the calculation of equilibrium area fractions.

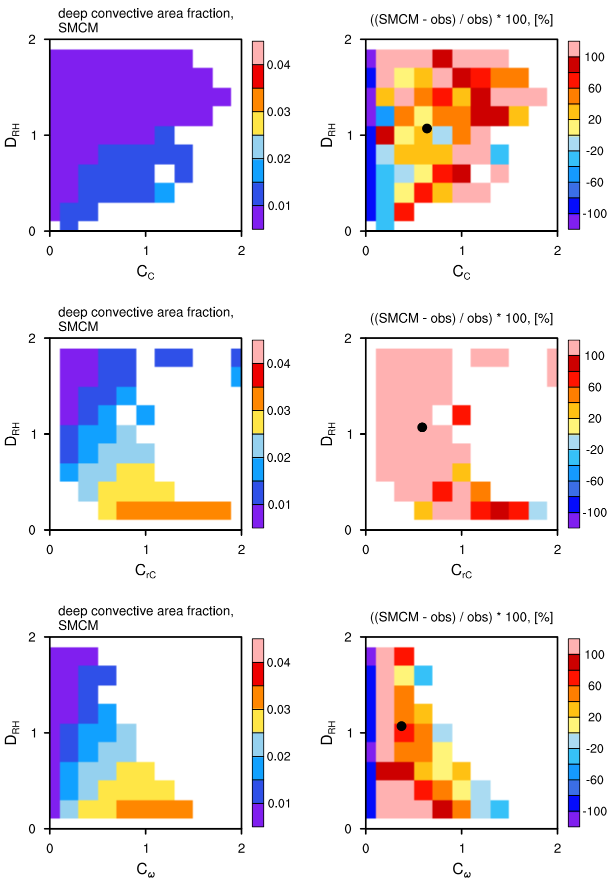

FIG. 7. Joint histograms of analytically computed equilibrium deep convective area fractions of the SMCM (left column) and the relative difference to observed mean deep convective area fractions at Darwin (right column) as function of large scale variables C_C (top), C_{rC} (middle) and C_ω (bottom) and D_{RH}. SMCM-modeled cloud fractions for each version of C correspond to the transition timescales shown in Tab. 1. Only histogram boxes having more than 5 observations are shown. The markers denote the mean observed values of C_C, C_{rC} and C_ω and D_{RH} at Darwin, respectively.

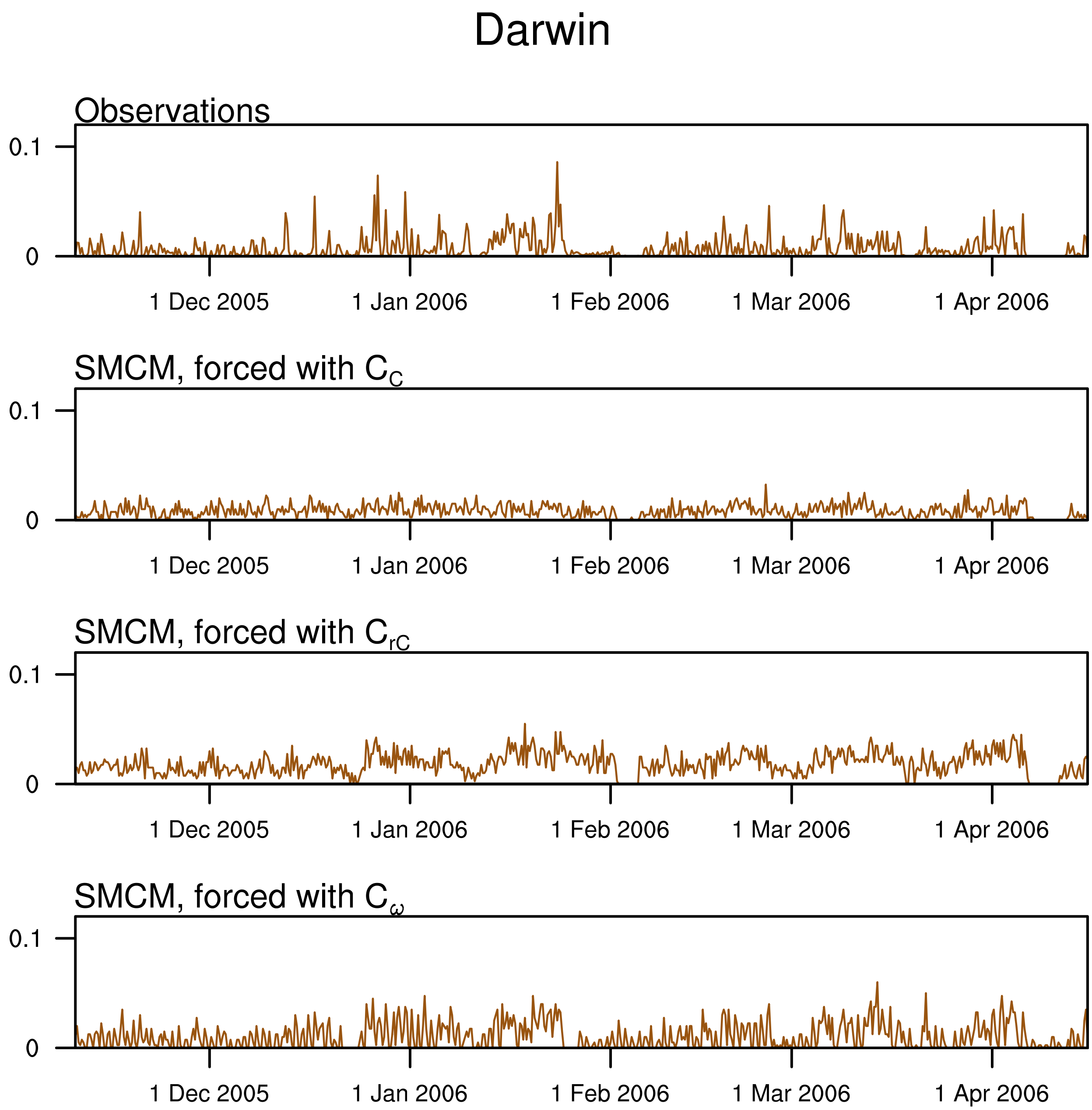

Fig. 8. Observed and SMCM-modeled time series of deep convective area fraction over Darwin during the time period of 10 Nov 2005 – 18 April 2006. SMCM-modelled time series are obtained by forcing the SMCM with the observed C and D parameters introduced in Sec. c and the transition timescales shown in Tab. 1. Results indicate one possible solution of the stochastic modelling approach.

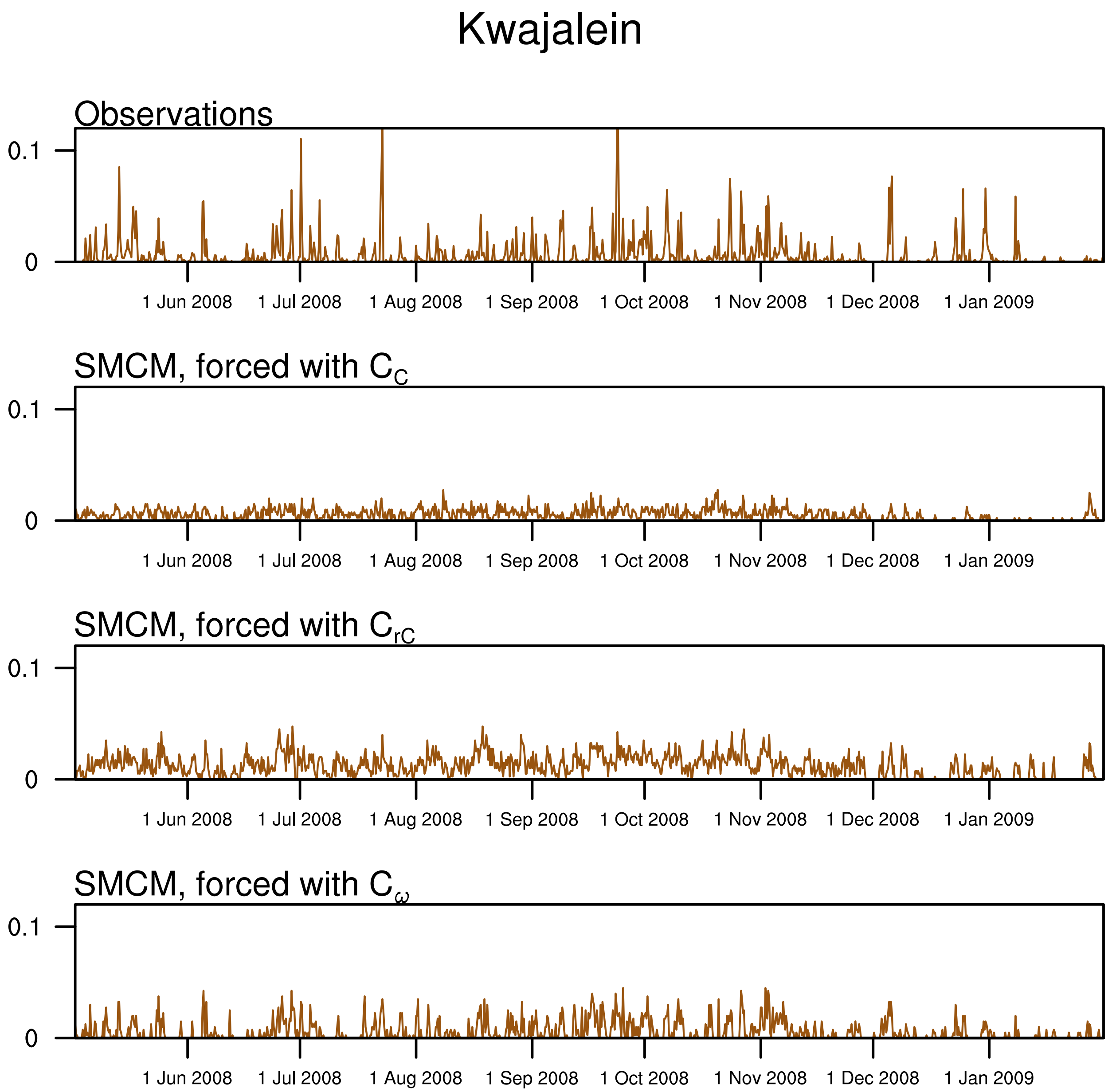

FIG. 9. Observed and SMCM-modeled time series of deep convective area fraction over Kwajalein during the time period of 2 May 2008 - 31 January 2009. SMCM-modelled time series are obtained by forcing the SMCM with the observed C and D parameters introduced in Sec. c and the transition timescales shown in Tab. 1. Results indicate one possible solution of the stochastic modelling approach.

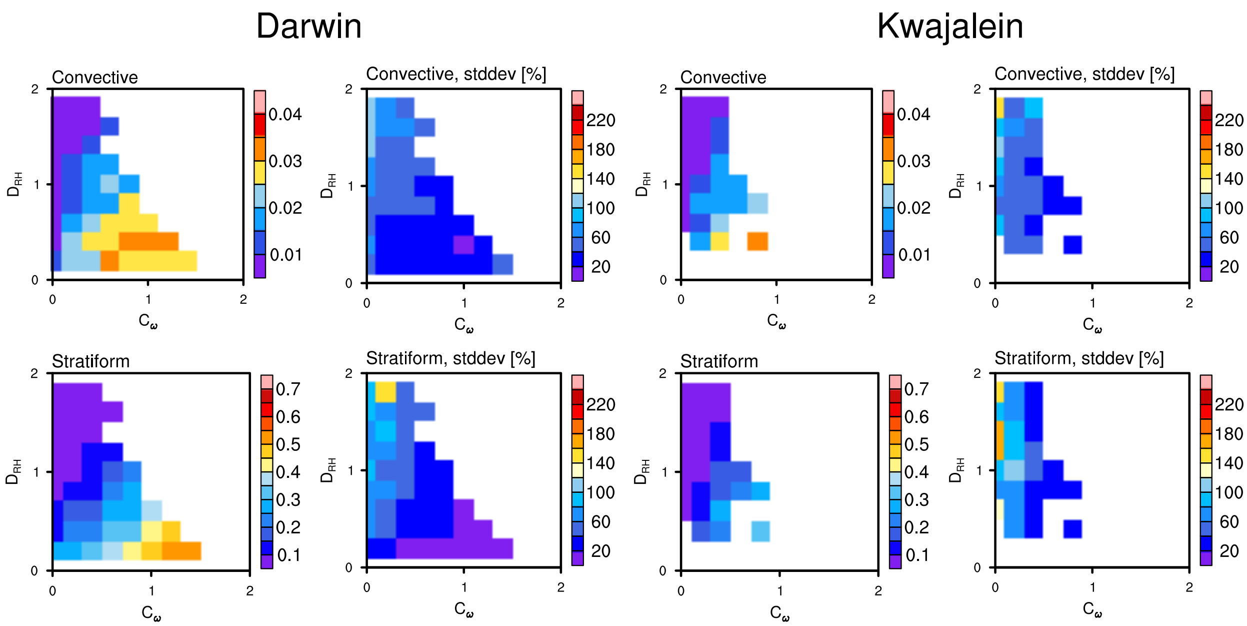

FIG. 10. Joint histogram of modeled cloud area fractions and relative standard deviations as function of large scale variables C_ω and $D_{\rm RH}$ at the Darwin (left two columns) and the Kwajalein (right two columns) sites derived from sampling the modeled cloud area fraction time series using all the available forcing data from observations (cf. Sec. b) and the transition time scales from Tab. 1. Only pixels having more than 5 observations are shown. Top row: deep convective clouds, middle row: stratiform clouds. Sample size per bin and color scales are the same as shown in Fig. 5.