AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE

10.1002/2014JD022680

The consequences of a local approach in statistical models of convection on its large-scale coherence

Jackson Tan[1,2], Christian Jakob[1], and Todd P. Lane[3]

[1] ARC Centre of Excellence for Climate System Science and School of Earth, Atmosphere and Environment, Monash University, Melbourne, Vic, Australia, [2] Now at NASA Wallops Flight Facility, Wallops Island, Virginia, USA [3] ARC Centre of Excellence for Climate System Science and School of Earth Sciences, University of Melbourne, Melbourne, Vic, Australia

Key Points:
- Observed convection has coherence beyond the grid box and the time step
- A model local in space and time does not produce such coherence
- Incorporating spatial and temporal memory recovered some of this coherence

Correspondence to:
J. Tan,
jackson.tan@nasa.gov

Citation:
Tan, J., C. Jakob, and T. P. Lane (2015), The consequences of a local approach in statistical models of convection on its large-scale coherence, *J. Geophys. Res. Atmos.*, *120*, 931–944, doi:10.1002/2014JD022680.

Received 12 OCT 2014
Accepted 11 JAN 2015
Accepted article online 15 JAN 2015
Published online 7 FEB 2015

Abstract Organized tropical convection is a crucial mechanism in the climate system, but its representation in climate models through parametrization schemes has numerous shortcomings. One of these shortcomings is that they are deterministic despite the statistical nature of the relationship they are representing. Several attempts at devising a stochastic parametrization scheme have been made, many of which assume a local approach, that is, one in which the convection in a grid box is determined without consideration of the previous time steps and the surrounding boxes. This study seeks to explore the effect of this assumption on the coherence of convection using cloud regimes, which represent various modes of tropical convection. First, we analyze the coherence of observed convection beyond the typical size of a model grid box and time step. Then, we evaluate the consequences of the local assumption on this coherence in simple statistical models. Cloud regimes in the real world show high degrees of coherence, manifesting in their lifetimes, areas, and inter-regime relationships. However, in a local statistical model, they are too small, too short-lived, and have incorrect relationships between each other. This can be improved by incorporating time memory and spatial dependence in the modeling. Our results imply that a local approach to a statistical representation of convection is not viable, and a statistical model must account for nonlocal influence in order to have large-scale convective coherence that more closely resembles the real world.

1. Introduction

Atmospheric convection is a dominant process in the tropics, critical to tropical cloud cover and the planetary circulation, and thus strongly influences variables such as precipitation and global energy balance. The phenomena for which convection plays a major role can range from a single-cell thunderstorm of several kilometers in horizontal size to the Madden-Julian Oscillation that has a circulation with a zonal extent greater than 10,000 km. This unique feature for convection to organize across such scales underscores its significance in Earth's weather and climate and the challenge in understanding it.

Despite its importance, the representation of convection in general circulation models (GCMs) remains a challenge. As GCMs typically have grid boxes that are too large to explicitly resolve convection, parametrization schemes are needed to predict the ensemble properties of convection based on the resolved large-scale environment. However, current convective parametrization schemes have numerous shortcomings, and many model errors have been attributed to deficiencies in the representation of convective processes. These errors include biases in tropical clouds and precipitation [*Soden and Held*, 2006; *Bauer et al.*, 2011; *Rossow et al.*, 2013], poor simulation of tropical variability on different timescales [e.g., *Yang and Slingo*, 2001; *Neale and Slingo*, 2003; *Lin et al.*, 2006], and lack of large-scale organization of convection [*Moncrieff et al.*, 2012].

As the task of parametrization schemes is to connect the statistical properties of subgrid-scale processes to grid-scale parameters such as convergence and humidity, they should naturally be stochastic or have a component that is so. At the same time, the variability associated with stochastic methods can be used to represent uncertainties in weather and climate predictions [*Palmer et al.*, 2005]. However, convective parametrization schemes in GCMs are deterministic, i.e., assuming a one-to-one relationship between the collective convective field in a grid box and a particular configuration of large-scale conditions. Recently, many separate efforts toward more stochastic parametrizations of convection have been made to improve model performance. This began with *Buizza et al.* [1999] with the introduction of stochastic variability in

an ensemble forecast system by multiplying the parametrized tendencies with random numbers that are spatially and temporally autocorrelated. Since then, numerous attempts have been made, mostly by modifying a model component or input with stochastic elements, such as in *Lin and Neelin*, [2000, 2002], *Majda and Khouider* [2002], *Khouider et al.* [2003], *Lin* [2003], *Plant and Craig* [2008], *Ball and Plant* [2008], *Teixeira and Reynolds* [2008], *Tompkins and Berner* [2008], *Keane and Plant* [2012], *Bengtsson et al.* [2013], *Frenkel et al.* [2013], and *Deng et al.* [2015]. Approaching the same question from another angle, *Khouider et al.* [2010] and *Khouider* [2014] developed a stochastic multicloud model by framing conceptual convective modes into a stochastic structure using Markov chains, with further investigation of its tuning and implementation in *Frenkel et al.* [2013], *Peters et al.* [2013], and *Dorrestijn et al.* [2013]. These stochastic parametrizations generally improved the results of the parent models, particularly in simulating the variability of quantities associated with convection such as the Madden-Julian Oscillation.

At the same time, many parametrization schemes, including stochastic ones, assume a local (or diagnostic) approach. *A local approach is one in which convection in one grid box is diagnosed from the large-scale environment without consideration of the previous time step and the surrounding grid boxes*. In other words, these parametrization schemes are not concerned with the spatiotemporal coherence of convection above the grid-scale, assuming instead that it will emerge through the large-scale variables through which convection is determined. Several of the stochastic parametrization schemes mentioned previously have prognostic elements that permits influence from previous time steps, but only a few have an explicit consideration of information in neighboring grids [*Majda and Khouider*, 2002; *Khouider et al.*, 2003; *Bengtsson et al.*, 2013; *Dorrestijn et al.*, 2013; *Khouider*, 2014]. It has been long recognized that a stochastically parametrized climate model possesses red noise [*Hasselmann*, 1976], which means that the subgrid variations are autocorrelated. Some of the stochastic schemes draw random variables from a red noise distribution to characterize this aspect of the relationship [e.g., *Buizza et al.*, 1999; *Lin and Neelin*, 2000; *Keane and Plant*, 2012], but it is not clear if this approach has an impact on the large-scale organization of the convective field.

The goal of this study is to examine the consequences of the local approach on modeling convection in a simple statistical framework. This is achieved in two parts. First, we analyze the large-scale coherence of tropical convection using quantities such as lifetimes and contiguous areas. Second, we explore how much of this coherence is reproduced in simple models: a local statistical model, a deterministic model, a Markov-chain model, and a Markov-chain model that takes into consideration its surroundings. All models are driven by large-scale variables from reanalysis data based on relationships derived from observations.

2. Data and Methods

2.1. ISCCP Cloud Regimes

The International Satellite Cloud Climatology Project (ISCCP) D1 data set gives statistical descriptions of clouds within 280 km × 280 km equal-area grids in the form of joint-histograms of their cloud top pressures and optical thickness since 1983 [*Rossow and Schiffer*, 1999]. These joint histograms are composed of satellite pixel measurements with an approximate horizontal resolution of 5 km at nadir from a network of geostationary and polar-orbiting satellites. In *Jakob and Tselioudis* [2003], the *k*-means clustering algorithm is applied to the joint-histograms to identify repeating patterns. In *Rossow et al.* [2005], this method is further improved upon by a set of criteria to determine the number of clusters. This clustering technique offers an objective categorization of cloud fields into distinct cloud regimes (also called weather states), which have been examined in terms of their meteorological conditions [*Jakob et al.*, 2005; *Jakob and Schumacher*, 2008], precipitation [*Lee et al.*, 2013; *Rossow et al.*, 2013], large-scale environment [*Tan et al.*, 2013; *Handlos and Back*, 2014], and heating profiles [*Stachnik et al.*, 2013; *Li et al.*, 2013]. These regimes have been applied in observations to study the Madden-Julian Oscillation [*Chen and Del Genio*, 2009; *Tromeur and Rossow*, 2010], trends in organized deep convection [*Tselioudis et al.*, 2010], cloud radiative feedback [*Oreopoulos and Rossow*, 2011], African Easterly Waves [*Mekonnen and Rossow*, 2011], dominant modes of climate variability [*Tselioudis and Rossow*, 2011], and precipitation extremes [*Rossow et al.*, 2013]. They have also been used to evaluate the performance of climate models [*Gordon et al.*, 2005; *Williams and Tselioudis*, 2007; *Williams and Webb*, 2009; *Tsushima et al.*, 2013]. Recent work has also extended these regimes to the Southern Ocean [*Haynes et al.*, 2011; *Mason et al.*, 2014] or even globally [*Tselioudis et al.*, 2013].

In particular, *Tan et al.* [2013] concluded that these cloud regimes are informative proxies for the various modes of convection and are empirical candidates for the building blocks of tropical convection. A further

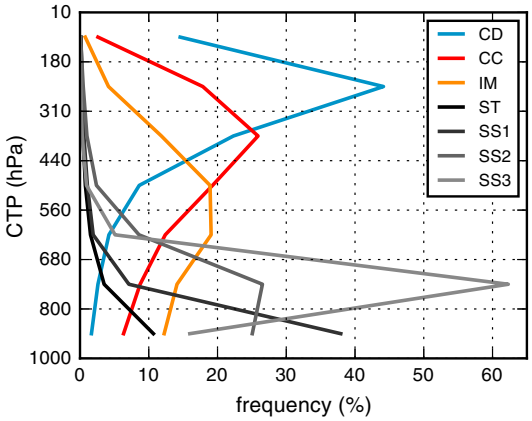

Figure 1. The centroids of the seven IR-only cloud regimes as a function of cloud top pressure (CTP). Their frequencies of occurrence (FOCs) are 0.061 (CD), 0.102 (CC), 0.146 (IM), 0.357 (ST), 0.165 (SS1), 0.118 (SS2), and 0.051 (SS3).

extension to these cloud regimes in *Tan and Jakob* [2013] was able to identify the major cloud regimes at a higher time resolution. This improvement allowed the cloud regimes to be resolved at the 3 h resolution native to ISCCP D1 and at all times of the day, instead of what was previously a daytime-averaged resolution or daytime-only availability. As we will be statistically modeling the regimes based on their relationships with large-scale variables, the time step of the models is thus tied to the time resolution of the regimes. In any model, the time step has to be sufficiently short to minimize numerical errors. Therefore, it is advantageous for our purpose to use these new IR-only regimes, named because they rely only on the infrared (IR) channel of the satellites, to represent the major types of tropical convection. To avoid confusion, the original cloud regimes will henceforth be called "IR-VIS regimes" to indicate their dependence on both the IR and visible (VIS) channels, and the term "(cloud) regimes" without a prefix is assumed to be this IR-only version.

Between ±35° latitudes in the period from July 1983 to December 2009, there are seven cloud regimes, with their cluster centroids (means of the regime members) shown in Figure 1. There are originally eight IR-VIS regimes in the same region, but the IR-VIS regime with a high occurrence of very thin cirrus cloud has to be discarded in the derivation of the IR-only regimes. This is because the thin cirrus signal is chiefly a product of a correction algorithm based on the VIS channel and hence absent in the IR-only cloud profiles. For the geographical distributions of these seven regimes, see Figure S4 in the supporting information for *Tan and Jakob* [2013]. The precipitation distributions and large-scale environment of the IR-only regimes [*Tan*, 2014] are strikingly similar to those for the IR-VIS regimes in *Tan et al.* [2013], thus providing further confidence that this temporal improvement does not come at a significant cost of errors in identification. Of the greatest interest to us is the CD regime, which represents organized deep convection. It has a prevalence of towering cumulus and deep stratiform clouds and has the highest incidence of clouds with high tops. Two other regimes with substantial signals of deep convection are the CC and IM regimes, which represent deep convection that is of a less organized nature. The CC regime has a predominance of cirrus clouds, while the IM regime has a signature suggestive of cumulus congestus clouds. The four remaining regimes all characterize convectively suppressed atmospheric conditions and have cloud tops that are generally below 680 hPa. The ST regime describes trade cumulus clouds and fair weather conditions and has the highest frequency of occurrence. The other three regimes are dominated by stratocumulus clouds at various stages of their development, occurring primarily over the oceans. Although these suppressed regimes are important features of the tropical environment, our focus here is on deep convection, so emphasis is given to the three regimes with deep convective signals. Indeed, for the remainder of the paper, we will combine the SS regimes into one SS regime when plotting them in the diagrams.

All regimes are derived in grid boxes of equal area, but we interpolate them to 2.5° × 2.5° grids using the nearest-neighbor technique to simplify our analysis; i.e., a regime value in the latitude-longitude grid is set to that of the nearest grid point in the original equal-area grid. We exclude the outermost 5° latitudes at both ends so as to, first, deal with latitude boundary cutoffs when analyzing the spatial properties of the regimes (section 3) and, second, avoid boundary artifacts when we statistically model the regimes based on their surrounding grid boxes (section 4). Finally, the results shown in this section are constrained to the year 2007 excluding the first month. This is so as to coincide with the period of analysis for the statistical models. A separate analysis over longer periods showed very little differences (not included here), so the conclusions from this 1 year result are robust.

The aim of this paper is to focus on how deep convection as represented by the cloud regimes form spatial and temporal structures *beyond* a typical grid box and time step. Contrast this with the different degrees of organization *within* the grid box as represented by the various regimes, ranging from organized deep convection in the CD regime to the scattered trade cumulus clouds in the ST regime. To avoid confusion,

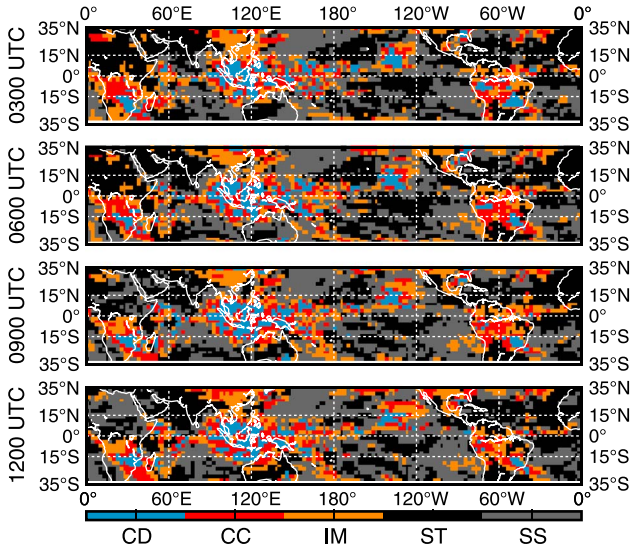

Figure 2. Snapshots of the observed IR-only regime field for four consecutive time slices on 1 January 2007.

we use the term *coherence* to describe structures coarser than the resolution of the dataset (i.e., 2.5° and 3 h) and the term *organization* to indicate the nature of convection within the grid box as represented by the regimes.

3. Coherence of Convection in the Real World

Figure 2 shows snapshots of the observed regime fields for four consecutive time slices of 3 h separation. The most striking feature of Figure 2 is the spatial coherence of the regimes, forming contiguous areas (CAs) that span over several grid boxes. These CAs are also persistent through the 12 h period, as clearly exemplified by the CA of the CD regime in the central Pacific around 15°N. This is an illustration of temporal coherence. Such spatiotemporal coherence can be quantified by examining the lifetimes of the regimes in every grid box and the sizes of their CAs at every time step.

The lifetime of an occurrence of a regime is the time between its emergence in a grid box to its extinction. We collect all occurrences of all regimes in grid boxes between ±30° latitudes and analyze the distributions of lifetimes through histograms of their durations. We define duration as proportional to histogram counts in each bin multiplied by the bin value, normalized such that duration sums to 3 h (one time step) over all bins and regimes. Using duration instead of the histogram counts takes in account the fact that long-lived occurrences of a regime may occur for the same length of time as short-lived occurrences even if they are fewer in number; e.g., a regime that has 1 occurrence lasting 100 time steps is equal in duration to a regime that has 100 occurrences with a lifetime of 1 time step each. This avoids conveying the impression that the chance of encountering a long-lived occurrence is lower than it actually is. Duration can be interpreted as the likelihood of getting a particular regime with a particular lifetime, if we were to randomly sample a grid box at a particular time step. Note that the computation of lifetimes is an Eulerian approach; i.e., it does not distinguish, for example, between the genuine extinction of convection versus the propagation of convection out of a grid box.

Figure 3a shows the distributions of the durations of the observed regimes. A regime's duration is clearly influenced by its frequency of occurrence (FOC), with the durations of most common regimes (ST and SS) above the least common regime (CD) for lifetimes below 40 h. Of all the convective regimes, the CD regime has lower durations at shorter lifetimes, which means that it has a smaller chance than the CC or IM regimes of a shorter-lived occurrence. The CC and IM regimes have durations that are different by a magnitude proportional to their FOCs (FOC_{CC} / FOC_{IM} = 0.70) but otherwise similar in distribution profiles. This suggests that these two regimes of less organized deep convection have similar temporal coherence. Of all the regimes, the ST regime has the highest duration after a lifetime of 15 h, an outcome of both its high FOC and the persistent nature of clear sky or fair weather conditions. An interesting observation in Figure 3a is the roughly constant durations at short lifetimes for the CD and ST regimes. That is to say, when these two regimes occur, the probability of them with a lifetime of several time steps is comparable to the probability of a lifetime of just one time step.

The regular angle grid of our regime fields allows an easy identification of CAs of regimes at any time step, with contiguous defined as grid boxes with the same regime sharing a common border. The sizes of these CAs can then be quantified and converted to units of km^2 depending on the latitudes of their grid boxes. We consider all CAs with at least one grid box within 30° latitude but not if it touches the 35° boundary. The sizes of the CAs are presented as coverage, which is the histogram counts multiplied by bin value and then normalized such that the sum over all bins and regimes is the total area between ±30° latitudes. Therefore,

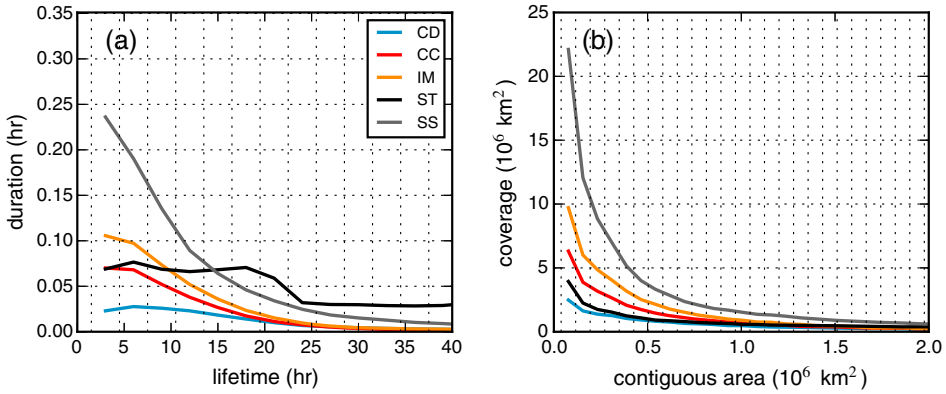

Figure 3. (a) Distributions of the durations of the regimes as a function of lifetimes. (b) Distributions of the coverages of the regimes as a function of contiguous areas. The vertical grid lines demarcate the histogram bin edges (3 h for duration; roughly one grid box for coverage).

analogous to duration, coverage accounts for the fact that large CAs are fewer in number than small CAs for the same cumulative area and can be interpreted as the likelihood of getting a CA in a particular regime of a particular size, if we were to randomly sample a grid box at any time step.

Figure 3b shows the distributions of the coverage of the observed regimes. The coverage of all regimes declines monotonically with larger CAs. This is unlike the temporal coherence of the regimes (Figure 3a), though this difference possibly stems from the contrast between the spatial and temporal resolutions of the IR-only regime dataset. At a 280 km and 3 h resolution, the ratio of space to time resolution is 26 m s^{-1}, which is faster than typical propagation speeds of convective systems. This means that the decay in duration per 3 h is at a lower rate than the decay in coverage per 280 × 280 km^2.

Comparing the regimes in Figure 3b, the CD and ST regimes have the lowest coverages at smaller CAs, which means that the typical areas of their CAs are largest. For the latter, this is remarkable since coverage is a direct function of the regime's FOC. Similar to Figure 3a, the coverage profiles between the CC and IM regimes differ only by a scaling proportional to their FOC ratios. At higher CAs, the differences in coverage between regimes are hard to distinguish in Figure 3b, but above 3 × 10^6 km^2, the ST regime dominates (not shown). With the probabilistic interpretation in mind, the relative chances of getting a large CA as opposed to a small CA is low for the CC, IM, and SS regimes but high for the CD and ST regimes, again a signature of the persistence of these two regimes.

Returning to Figure 2, the central Pacific CA of the CD regime pointed out earlier is surrounded by the CC and IM regimes, which represent deep convection as well but of weaker organization. This hints at a spatiotemporal relationship between regimes of closer convective characters. In other words, the regime of organized deep convection is surrounded by convective regimes of weaker organization, both in space and in time. Such an inter-regime relationship can be studied by counting the transitions and the neighbors of the regimes.

Transition counts record the regimes in each grid box at one time step and the following time step, for all grid boxes within ±30° latitudes. Neighbor counts record the regimes in the eight grid boxes surrounding a regime in a grid box, for all grid boxes within ±30° latitudes (including grid boxes bordering ±30°). We present this information in absolute counts instead of normalizing against the FOCs of the regimes, as normalization will disrupt any symmetry (or lack of) in the relationships.

Figure 4a displays the transition counts in a matrix diagram. The highest counts occur along the diagonal, indicating that all regimes are more likely to persist than to switch to any one other regime. This is another manifestation of temporal coherence. Furthermore, transitions between regimes of sharply contrasting convective nature are unlikely; e.g., CD→CC is more likely than CD→ST. Both observations support the notion of a temporal relationship between regimes of similar convective nature. In addition, the matrix is approximately though not exactly symmetric within the bin thresholds, suggesting that the number of transitions between different convective states as represented by these regimes are of similar orders of

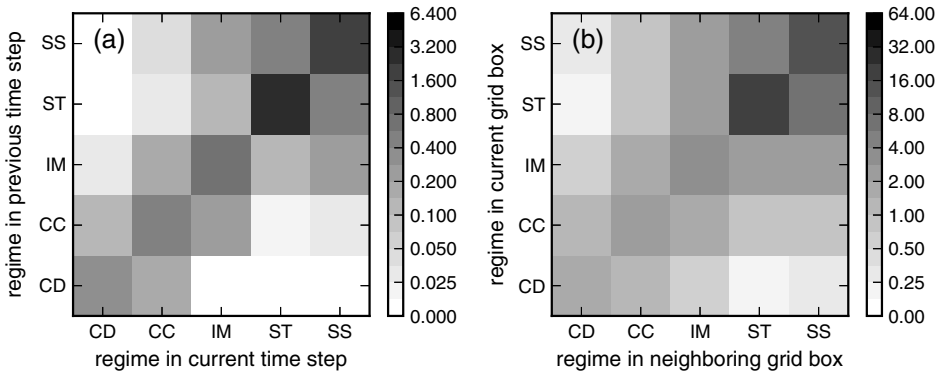

Figure 4. (a) Transition counts and (b) neighbor counts of the observed regimes, in units of millions of counts. Note the geometric scaling of the grey shading.

magnitude. Consistent with Figure 3, the CC and IM regimes differ by a ratio similar to that of their FOCs (CC→CC: 0.503; IM→IM: 0.710; ratio = 0.71).

Figure 4b displays the neighbor counts in a similar matrix plot. Again, the highest counts occur mostly along the diagonal, revealing that regimes tend to have themselves as neighbors more than any one other regime. This is a sign of spatial coherence. Counts between regimes of similar convective nature are higher, but the off-diagonal terms are not as low relative to the diagonal terms as is the case for transition counts. This disparity between transition and neighbor counts is a consequence of the ratio in space to time resolution as alluded to previously. Indeed, when using the VIS-IR regimes of daily time resolution, the transition counts are more similar to the neighbor counts (not shown). Figure 4b also shows a symmetry in the matrix, but in this case it is a natural consequence of the methodology: a grid box on the left will count the grid box on the right, and vice versa. Again, in line with previous observations, the CC and IM regimes differ proportionally to their FOCs ratio (CC→CC: 2.66; IM→IM: 3.70; ratio = 0.72).

In this section, we investigated the lifetimes, sizes of the contiguous areas (CAs), as well as the transitions and neighbors of the regimes. Generally, there are fewer occurrences which are long-lived and fewer CAs which are large. The CD and ST regimes are more coherent in their spatiotemporal properties than the CC, IM, and SS regimes. The CC and IM regimes demonstrate similar coherence, which is consistent with the view that they both represent less organized deep convection. Lastly, regimes are more likely to transition to, or have as neighbors, regimes of the same or similar convective nature.

4. Coherence in Statistical Models

Tan et al. [2013] used three large-scale variables that identify the key aspects of the large-scale state associated with tropical convection to establish the relationship between the large-scale conditions and the IR-VIS regimes. These variables are saturation fraction (r), modified K-index (K), and grid-mean vertical motion at 600 hPa (ω), calculated from the European Centre for Medium-Range Weather Forecasts Interim Re-Analysis [*Dee et al.*, 2011]. Saturation fraction r is defined as $\int q\,\mathrm{d}p / \int q_s\,\mathrm{d}p$, where q and q_s are the specific humidity and saturated specific humidity, respectively, at pressure level p. Unlike other measures of humidity, r is relatively independent of sea surface temperature while exhibiting a strong relationship with precipitation that is invariant over different ocean basins [*Bretherton et al.*, 2004]. A value of 0 is associated with dry conditions while a value of 1 indicates a saturated column. K (unit: Kelvin) is defined as

$$K = \frac{T_{1000} + T_{850}}{2} - T_{500} + \frac{T_{d_{1000}} + T_{d_{850}}}{2} - T_{700} + T_{d_{700}},$$

where T_p and T_{d_p} are temperature and dew point at p hPa, respectively [*Charba*, 1977]. It provides a simple measure of the convective instability and is employed in the weather forecast community as an index for the likely presence of convective showers and thunderstorms. K typically ranges from −40 K to 50 K, with a high value indicating an unstable environment. *Davies et al.* [2013] has evaluated K in detail over Darwin and found that it has some predictive skill for convective precipitation. ω (unit: Pa / s) describes the dynamical environment of the cloud regimes. A negative ω reveals an environment with large-scale ascending motion,

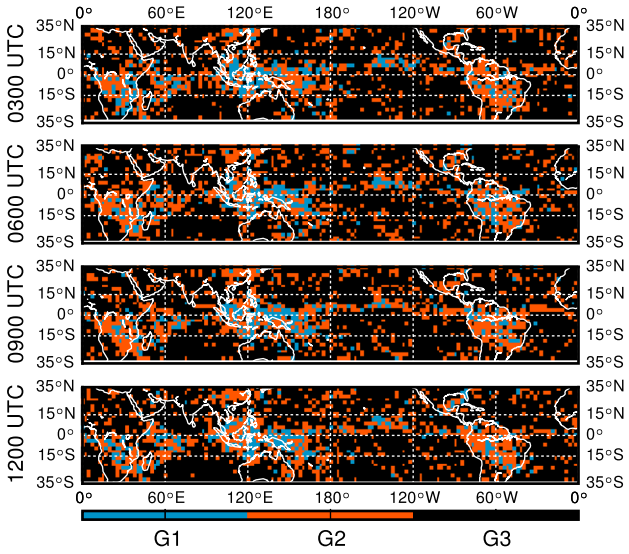

Figure 5. Snapshots of the local model regime group field for four consecutive time slices on 1 January 2007.

a condition we expect to be associated with deep convection in the atmosphere; positive ω, on the other hand, is indicative of large-scale subsidence.

Since the convective state of the atmosphere in a grid box must reflect the large-scale condition, we will be using these variables and their relationship to construct a range of statistical models of the regimes. First, we will need to extend their resolution to match the IR-only regimes. This requires a linear interpolation of the variables from their native 1.5° and 6 h resolution to the 2.5° and 3 h resolution of the IR-only regimes. At the same time, to reduce the substantial overlap between the regimes, we group the regimes based on their convective nature into three distinct groups: regime group G1, comprising solely the CD regime; regime group G2, consisting of the CC and IM regimes; and regime group G3, composed of the ST and SS regimes. Finally, we composite the variables and the regime groups to construct the relationship between the large-scale and the convective state.

The knowledge of the relationship between the regime groups and the three large-scale variables allows us to predict the regime in a grid box given a configuration of the variables in a range of statistical models of varying complexity. One potential issue is the use of the same set of information to construct the relationship and to predict the regimes based on the large-scale variables. To avoid this duplicated use of data, we set 2004 to 2006 as the "training period" to establish the relationship (which is sufficiently robust even with one year's data), and 2007 as the "modeling period," coinciding with the analysis of the real world. All statistical models are developed separately for ocean grid boxes and land grid boxes. Each model is run for 20 realizations with different initial random number seeds, all of which are used to calculate the measures of coherence employed in section 3, but we exclude the first month of 2007 to allow the model outputs to shed any semblance to the reality that may be used to initialize it.

Such a modeling approach makes two assumptions. First, the relationship between the regime groups and the variables are assumed to be globally consistent. Departure from this assumption may arise when the regime actually represents cloud fields that are highly disparate in their convective character yet similar in their IR signatures. For example, ISCCP IR-VIS joint-histograms which suggest midlatitude frontal clouds may be assigned to the CD regime. However, errors of this nature are mitigated by its relative rarity to organized deep convection in the region considered ($\pm$30° latitudes). Second, the lack of a feedback from the convection to the large-scale environment is assumed to be inconsequential to the characteristics of convection which we are focusing on. This shortcoming is inevitable as we are using an observation data set to conduct our modeling, and to separate the "control" and "feedback" of convection is quite simply beyond the capability of such a framework. Above all, it is imperative to understand that we are not ultimately designing a statistical model that can accurately reproduce the regimes given the large-scale state. This study seeks to understand the necessary rules of statistical models of convection and is in no way crafting a candidate for an ideal convective parametrization scheme.

4.1. Local Statistical Model

A simple statistical model is one that reads the large-scale variable in each grid box at each time step, diagnoses the probabilities, and determines the regime group independently for every grid box and time step. We call such a model a *local model*, as the probability of each regime is determined locally (or diagnostically) for each grid box, without any regard for its history or surrounding. In this model, the relationship between these regime groups and the large-scale variables is sorted into a multivariate

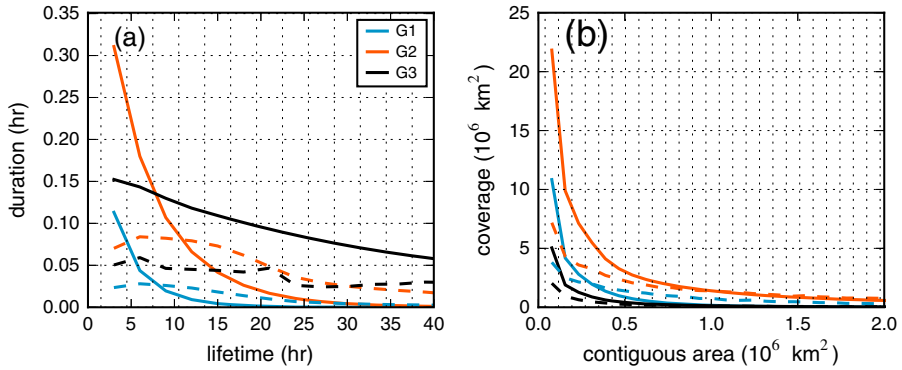

Figure 6. (a) Distributions of the durations of the regime group as a function of lifetimes in the local model (solid lines). (b) Distributions of the coverages of the regime groups as a function of contiguous areas in the local model (solid lines). The dashed lines in both diagrams are that for the observed regime groups. The vertical grid lines demarcate the histogram bin edges.

histogram with 10 bins in each dimension (10^3 bins in total). The bin edges are chosen so as to have an approximately equal number of total counts in every bin. The histogram counts in each bin can then be normalized to obtain a set of probabilities for G1, G2, and G3. With this histogram, the essential steps of the models are to take the input of three variables for each grid box, identify the bin it belongs to, and randomly assigning a regime based on those probabilities.

Figure 5 shows snapshots of the regime group field in the first four time steps of the model output. In comparison with Figure 2, the regime group field is clearly too noisy, with far too many CAs with sizes of only one grid box. Likewise, across time steps, there is no persistence in the regime groups. Overall, the regime group field "flickers" with each time step.

Unsurprisingly, the durations and the coverages of the regime groups are too short lived and too small (Figure 6). Temporally, the modeled regime groups do not possess the distinctive feature of an increase in duration when going from a 3 h lifetime to higher lifetimes (Figure 6a). Note that the observed G2 too has this feature (dotted line in Figure 6a), unlike the CC and IM regimes which forms it (Figure 3). This is simply because a transition of, say, CC → IM is now G2 → G2 and thus considered here as the same occurrence. Likewise, G3 does not discriminate between the transitions amongst the ST, SS1, SS2, and SS3 regimes, leading to an increased likelihood of very long-lived occurrences. Spatially, the coverages of small CAs is too high and the coverage of large CAs is too small. We caution readers in interpreting the coverage of G3: as it occurs more than half the time, it tends to form very large CAs (as evident in Figure 5), such that more than 95% of CAs are invalid due to the 35° latitude boundary condition.

Further evidence of the local model's failure in reproducing coherence is given in Figure 7, wherein the transition and neighbor counts of the regime groups are tabulated. These counts compare remarkably poor to the real world (Figure 8), violating in particular the observation of the maximum along the diagonal

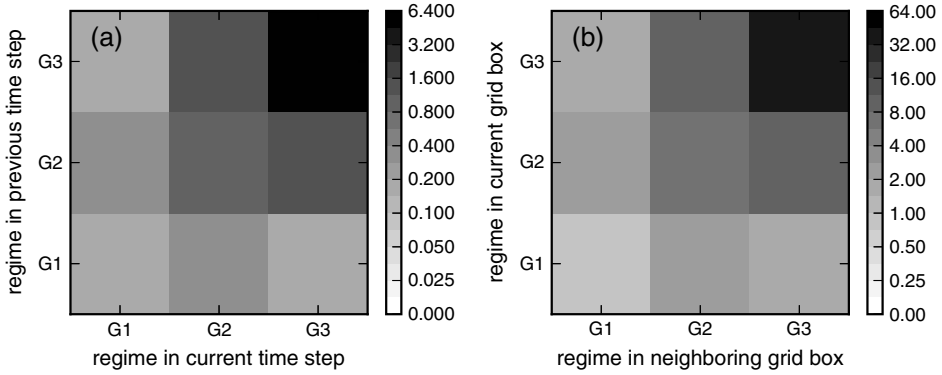

Figure 7. (a) Transition counts and (b) neighbor counts of the regime groups in the local model, in units of millions of counts.

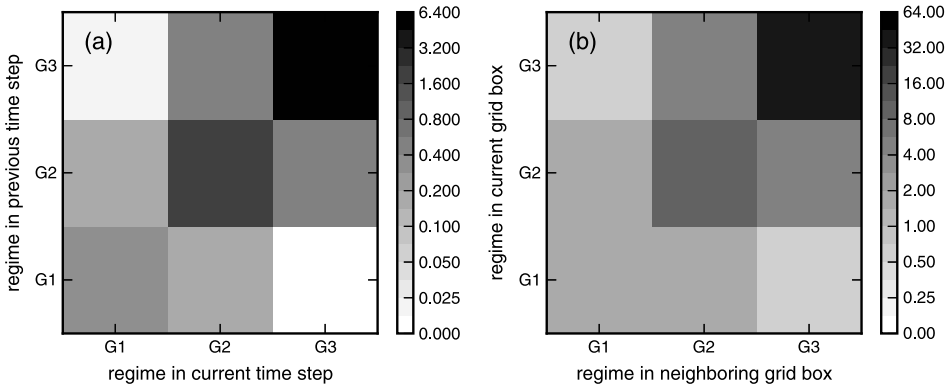

Figure 8. (a) Transition counts and (b) neighbor counts of the observed regime groups, in units of millions of counts.

(section 3). This means that a regime in a particular grid box at a particular time step is more likely to change to another regime and has as neighbors another regime than itself. This runs contrary to the persistence of these regimes that we have seen in section 3.

Therefore, the local statistical model, which does not account for nonlocal influences, fails to reproduce the coherence observed in regime groups of the real world. Perhaps this is not too surprising, as the random component in the model adds noise into the system. Indeed, many stochastic parametrization schemes were introduced with the idea of injecting the process with variability (see section 1).

Current parametrization schemes in climate models do not suffer this coherence problem as they are deterministic. To investigate this further, we constructed a simple *deterministic model* of the regime groups by first normalizing each variable (taking care to "orient" the variable from least to most convective) and averaging into one value in each grid box. We then assign a regime group to the grid boxes based on this averaged variable, with the "most convective" grid boxes to G1 in proportion to its observed FOC, and then the "next most convective" to G2, and the remainder to G3. This is done separately for 15° latitudes and 15°–35° latitudes. For 15°, the highest 8.7th percentile of the averaged variable will be assigned G1, the next 28.8th to G2, and the remainder to G3; for 15°–35°, the highest 4.1th precentile goes to G1, the next 21.7th percentile to G2, and the rest to G3.

Snapshots of this deterministic model are shown in Figure 9. The coherence in this model is simply too strong. The CAs are too large, replicating their spatial coherence from the smoothness of the large-scale variables. The geographical distributions of the regime groups are also substantially different from the real world (not shown). Furthermore, there is a fundamental problem with this model. The local statistical model, by construction, possesses the correct variance in the relationships between the regime groups and the variables. The deterministic model, on the other hand, eliminates this spread in favor of coherence, which is hardly a satisfactory resolution to the problem. What is needed is a way to encode memory into the statistical model so as to remain truthful to the statistical behavior of the relationship and yet allow coherence to manifest organically.

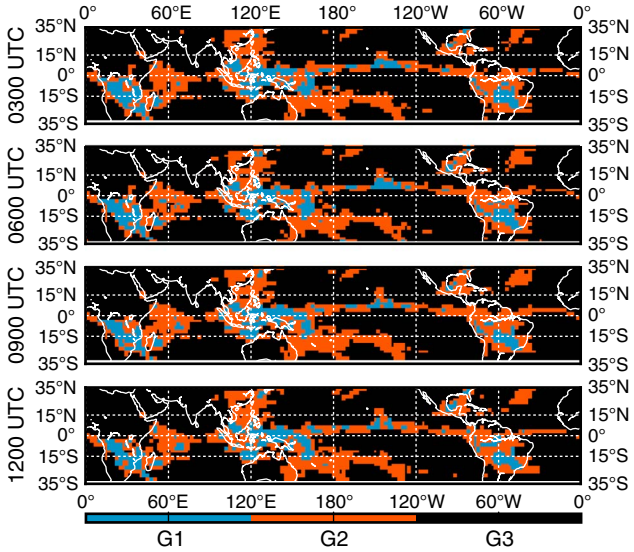

Figure 9. Snapshots of the deterministic model regime group field for four consecutive time slices on 1 January 2007.

4.2. Markov Chain Model

A natural progression from the local model is to employ Markov chain techniques in the modeling process. In this extension, we are seeking the

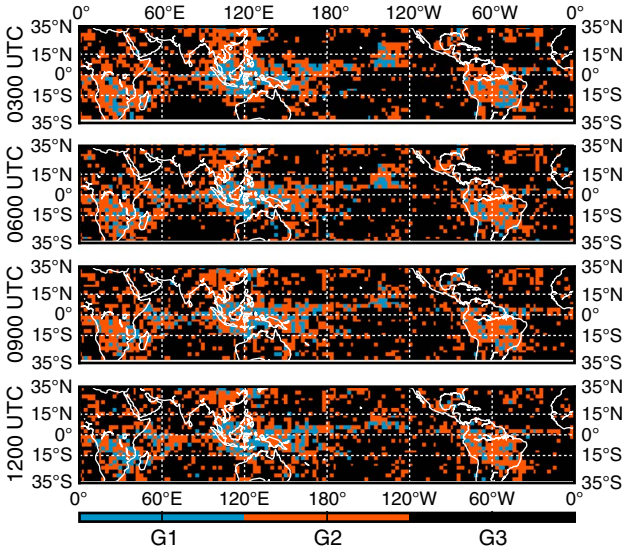

Figure 10. Snapshots of the Markov chain model regime group field for four consecutive time slices on 1 January 2007.

transition probabilities from one regime group to another, conditioned upon the large-scale environment. Specifically, the relationship between the regime groups and the large-scale variables are sorted into a multivariate histogram of the three variables and the regime group in the same grid box in the previous time step (resulting in $10^3 \times 3$ bins). The probabilities that drive this *Markov chain model* are thus a function of the three variables and the regime group in the previous time step. Therefore, a connection between consecutive time steps is established. Note that the model is still spatially independent.

Figure 10 shows snapshots of the model output, beginning with the second time step (the first time step is an exact copy of the real world to initialize the model). While the regime group field is still noisy in its spatial coherence, there is arguably some improvement in terms of its persistence over time steps. This becomes clearer in its duration (Figure 11a). A striking recovery of the temporal coherence is evident, with the durations of the regime groups comparable to observation. Likewise, the transition counts of the regime groups returned to the pattern expected from observation (Figure 11b). In fact, Figure 11b is indistinguishable from Figure 8a, which means that the model transition counts are identical to the real world to the limit of the histogram bins. Therefore, the inclusion of a 3 h Markovian memory is sufficient to recover the temporal coherence of the regime groups. The spatial coherence of the Markov chain model regime groups are unaffected by the addition of temporal memory, so we do not present those measures here.

The encouraging success of the Markov chain model inspired us to seek an equivalent approach for spatial coherence. Unfortunately, this proved to be a much more challenging task. To begin with, the unidirection of time cannot be applied to the spatial dimension. A grid box on the left will influence and be influenced by the grid box on its right, whereas a grid box in the past will only influence and cannot be influenced by a grid box in the future. Moreover, we will need to consider all surrounding grid boxes, unlike the temporal case with only one grid box. Any strategies at encoding spatial memory will need to surmount these two obstacles.

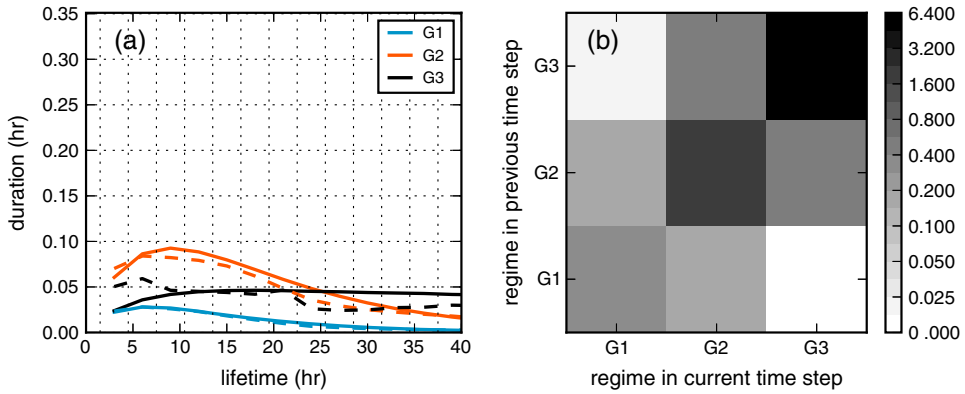

Figure 11. (a) Distributions of the durations of the regime group as a function of lifetimes in the Markov chain model. (b) Distributions of the coverages of the regime groups as a function of contiguous areas in the Markov chain model. The dashed lines in both diagrams are that for the observed regime groups. The vertical grid lines demarcate the histogram bin edges.

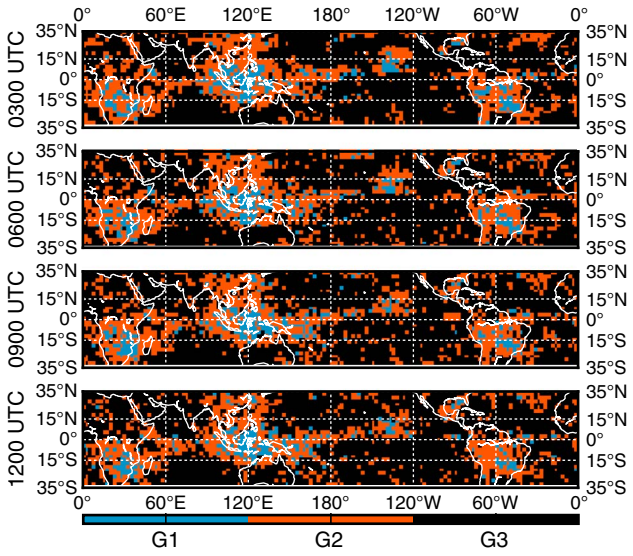

Figure 12. Snapshots of the spatiotemporal Markov chain model regime group field for four consecutive time slices on 1 January 2007.

We make such one attempt here by taking into account the eight surrounding grid boxes of the previous time step for every grid box in the Markov chain model. This is similar to the approach of considering neighboring lattice sites in *Majda and Khouider* [2002], *Khouider et al.* [2003], and *Khouider* [2014], but here we formulate the spatial influence across grid boxes using observed behavior. First, we compile statistics from 2004 to 2006, calculating the probability of a grid box to be in a regime group given the number of each regime group in the surrounding eight grid boxes 3 h earlier. For example, a grid box surrounded by 8 G1 previously has a probability of 0.98 to be G1 and 0.02 to be G2; a grid box surrounded by 5 G1, 2 G2, and 1 G3 in the previous time step has a probability of 0.76, 0.23, and 0.01 to be G1, G2, and G3 in this time step, respectively. Therefore, by knowing the eight surrounding grid boxes in the previous time step, we can produce an additional set of probabilities that characterizes spatial influence. Together with the Markov chain set of probabilities, we can merge the two sets by multiplying them element-wise and then normalizing them. (that is, the probability of regime group g is $p_g = \left(p_g^m \times p_g^s\right) / \sum_g \left(p_g^m \times p_g^s\right)$, where p_g^m and p_g^s are the probabilities of regime group g based on the Markov chain and surrounding grid boxes, respectively.)

Figure 12 shows snapshots of the regime group field generated by this surrounding Markov chain model. Comparing it with Figures 5 and 10, there is a slight decrease in the number of CAs one grid box in size as well as a perceptible increase in the sizes of larger CA and appearing less "broken up." However, the model still produces excessive one-grid-box CAs in contrast to Figure 2. Looking at their coverage in Figure 13a, the decrease in one-grid-box CAs and the improvement in high CA coverages (especially for G1) is noticeable in comparison to Figure 6 but clearly insufficient to match the observations. Likewise, the neighbor counts return an improved result (Figure 13b) but still clearly different from observation (Figure 8). It is plausible that averaging the two sets of probabilities with greater weight given to the surrounding influence may improve spatial coherence, but this comes at a cost of an accuracy in the regime group FOC since the

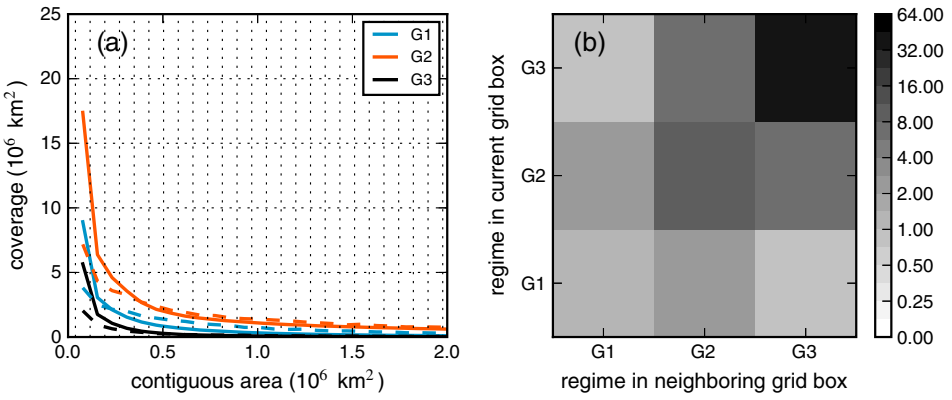

Figure 13. (a) Distributions of the durations of the regime group as a function of lifetimes in the spatiotemporal Markov chain model. (b) Distributions of the coverages of the regime groups as a function of contiguous areas in the spatiotemporal Markov chain model. The dashed lines in both diagrams are that for the observed regime groups. The vertical grid lines demarcate the histogram bin edges.

probabilities from the surrounding grid boxes do not preserve the ratios of the FOCs. The temporal coherence of this model is not adversely affected by the inclusion of spatial memory.

In this section, we used a variety of statistical models to investigate the effects of memory on spatiotemporal coherence. A local model simply does not reproduce the correct coherence. When time memory is included into the modeling process via Markov chains, the temporal coherence improved tremendously. However, spatial coherence proved to be more challenging, as our strategy of using the surrounding grid boxes in the previous time step could only partially regain the coherence lost due to the statistical noise.

5. Discussion and Conclusion

We demonstrated that various states of convection, as represented by the IR-only cloud regimes, have coherent features on scales beyond a typical grid box and time step of a climate model. This coherence manifests both in space and in time, as shown by the regimes' lifetimes, contiguous areas (CAs), transition counts, and neighbor counts. Such spatiotemporal coherence beyond the elementary resolution is not present in the local statistical model. This is despite the fact that this model is driven by the coherent large-scale environment using rules derived in observations. Integrating time memory into the statistical model recovered much of the temporal coherence. Including spatial influence is a more challenging task, as demonstrated by our partially successful attempt when surrounding grid boxes were considered.

The fact that we are attempting to relate statistical processes to resolved quantities necessitates a stochastic approach. Or, at the very least, parametrization schemes need to have a stochastic component to express the spread associated with this statistical relationship. In a way, the deficiency of our deterministic model is a reflection of this need within our simplified framework. However, our results imply that a parametrization scheme that is both stochastic and local (or diagnostic) will not possess sufficient degrees of coherence. While our study is fairly simplistic due to the limitations of using an observational approach, these conclusions should be applicable to any stochastic attempts to modeling convection. Therefore, the introduction of stochastic elements into parametrization will need to be implemented carefully as an approach that is ignorant of previous time steps and surrounding grid boxes will lead to a breakdown of large-scale coherence in convection. Any stochastic parametrization of convection will likely need to advance beyond such a local framework if it desires to maintain a large-scale coherence of convection that resembles the real world. Such a conclusion is consistent, for example, with *Buizza et al.* [1999] in the need to add spatial and temporal memory to the random number field used in their stochastic parametrisation.

One potential caveat of our study is that the convective nature of the cloud regimes is determined through its cloud signature, which means that convection and clouds are inseparable in our analysis. As such, it is not clear if our conclusion on the need for a nonlocal approach applies to the convective closure or the cloud scheme of the parametrization, or both. This insight cannot be provided within the framework of this study. Nonetheless, our study suggests that both spatial and temporal memory are needed in the combined package of convection and clouds.

There is also the possibility that more complicated but fundamentally local statistical techniques in modeling convection may allow for coherence to a realistic degree. For example, one can envisage a "partially statistical" model which is a blend of our local model and our deterministic model. While this may produce coherence that resembles the real world, there is the problem of reduced variability, as we have alluded to in section 4. Coherence is seldom if at all a focus of stochastic parametrization attempts, so it is unknown if these schemes can retain the right amounts of coherence and variability.

In *Bengtsson et al.* [2013], a regional climate model of resolution 5.5 km and 120 s was coupled to a cellular automaton, which activates when a certain threshold for convective available potential energy is exceeded. It is notable for some memory effects such as different lifetimes on the leading and trailing edges of a squall line and a propagation of the cellular automaton cells against the mean flow. This suggests that the intergrid spatial memory introduced into the model boosted the spatial coherence of convection, producing more realistic results in contrast to a reference model without spatial memory. Such a result is consistent with our argument that considering nonlocal influence is necessary for spatiotemporal coherence, though one should exercise caution in making quantitative comparisons due to the disparity in space and time scales between the studies. While the degree of spatiotemporal memory may vary depending on the resolution or the model complexity, it is very likely that our conclusions—i.e., memory is needed for coherence—apply.

In conclusion, we showed that convection has coherence beyond the grid box and the time step in the real world. In particular, the cloud regime representing organized deep convection lives longer and forms larger contiguous areas than the other deep convective regimes that are less organized. However, when we modeled convection in simple statistical models, a model that is local in space and time is bereft of such coherence. Incorporating information from previous time steps and surrounding grid boxes reclaimed some of this coherence. Given that convective parametrization schemes have to be stochastic by nature, our conclusions imply that a local approach is not viable to attain the observed level of convective coherence.

Acknowledgments
The authors thank Kais Hamza and Karsten Peters for discussions on the study and three anonymous reviewers for their helpful comments. This project is funded under the Australian Research Council Centre of Excellence for Climate System Science (CE110001028). J. T. acknowledges support from Monash University Postgraduate Publication Award. Data and codes pertaining to this paper can be obtained by contacting the corresponding author (jackson.tan@nasa.gov).

References

Ball, M. A., and R. S. Plant (2008), Comparison of stochastic parametrization approaches in a single-column model, *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.*, *366*(1875), 2603–2621, doi:10.1098/rsta.2008.0050.

Bauer, P., G. Ohring, C. Kummerow, and T. Auligne (2011), Assimilating satellite observations of clouds and precipitation into NWP models, *Bull. Am. Meteorol. Soc.*, *92*(6), ES25–ES28, doi:10.1175/2011BAMS3182.1.

Bengtsson, L., M. Steinheimer, P. Bechtold, and J.-F. Geleyn (2013), A stochastic parametrization for deep convection using cellular automata, *Q. J. R. Meteorol. Soc.*, *139*(675), 1533–1543, doi:10.1002/qj.2108.

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

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

Charba, J. P. (1977), Operational system for predicting thunderstorms two to six hours in advance, *Tech. Memo. NWS TDL-64*, NOAA, Silver Spring, Md.

Chen, Y., and A. D. Del Genio (2009), Evaluation of tropical cloud regimes in observations and a general circulation model, *Clim. Dyn.*, *32*(2–3), 355–369, doi:10.1007/s00382-008-0386-6.

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.

Dee, D. P., et al. (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, *137*(656), 553–597, doi:10.1002/qj.828.

Deng, Q., B. Khouider, and A. J. 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.

Dorrestijn, J., D. T. Crommelin, J. A. Biello, and S. J. Boing (2013), A data-driven multi-cloud model for stochastic parametrization of deep convection, *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.*, *371*(1991), 20120374, doi:10.1098/rsta.2012.0374.

Frenkel, Y., A. J. Majda, and B. Khouider (2013), Stochastic and deterministic multicloud parameterizations for tropical convection, *Clim. Dyn.*, *41*(5–6), 1527–1551, doi:10.1007/s00382-013-1678-z.

Gordon, N. D., J. R. Norris, C. P. Weaver, and S. A. Klein (2005), Cluster analysis of cloud regimes and characteristic dynamics of midlatitude synoptic systems in observations and a model, *J. Geophys. Res.*, *110*, D15S17, doi:10.1029/2004JD005027.

Handlos, Z. J., and L. E. Back (2014), Estimating vertical motion profile shape within tropical weather states over the oceans, *J. Clim.*, *27*(20), 7667–7686, doi:10.1175/JCLI-D-13-00602.1.

Hasselmann, K. (1976), Stochastic climate models part I. Theory, *Tellus*, *28*(6), 473–485.

Haynes, J. M., C. Jakob, W. B. Rossow, G. Tselioudis, and J. Brown (2011), Major characteristics of southern ocean cloud regimes and their effects on the energy budget, *J. Clim.*, *24*(19), 5061–5080, doi:10.1175/2011JCLI4052.1.

Jakob, C., and C. Schumacher (2008), Precipitation and latent heating characteristics of the major tropical western pacific cloud regimes, *J. Clim.*, *21*(17), 4348–4364, doi:10.1175/2008JCLI2122.1.

Jakob, C., and G. Tselioudis (2003), Objective identification of cloud regimes in the Tropical Western Pacific, *Geophys. Res. Lett.*, *30*(21), 2082, doi:10.1029/2003GL018367.

Jakob, C., G. Tselioudis, and T. Hume (2005), The radiative, cloud, and thermodynamic properties of the major tropical western Pacific cloud regimes, *J. Clim.*, *18*(8), 1203–1215.

Keane, R. J., and R. S. Plant (2012), Large-scale length and time-scales for use with stochastic convective parametrization, *Q. J. R. Meteorol. Soc.*, *138*(666), 1150–1164, doi:10.1002/qj.992.

Khouider, B. (2014), A coarse grained stochastic multi-type particle interacting model for tropical convection: Nearest neighbour interactions, *Commun. Math. Sci.*, *12*(8), 1379–1407, doi:10.4310/CMS.2014.v12.n8.a1.

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

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

Lee, D., L. Oreopoulos, G. J. Huffman, W. B. Rossow, and I.-S. Kang (2013), The precipitation characteristics of ISCCP tropical weather states, *J. Clim.*, *26*(3), 772–788, doi:10.1175/JCLI-D-11-00718.1.

Li, W., C. Schumacher, and S. A. McFarlane (2013), Radiative heating of the ISCCP upper level cloud regimes and its impact on the large-scale tropical circulation, *J. Geophys. Res. Atmos.*, *118*, 592–604, doi:10.1002/jgrd.50114.

Lin, J.-L., et al. (2006), Tropical intraseasonal variability in 14 IPCC AR4 climate models. Part I: Convective signals, *J. Clim.*, *19*(12), 2665–2690.

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

Lin, J. W.-B., and J. D. Neelin (2000), Influence of a stochastic moist convective parameterization on tropical climate variability, *Geophys. Res. Lett.*, *27*(22), 3691–3694, doi:10.1029/2000GL011964.

Lin, J. W.-B., and J. D. Neelin (2002), Considerations for stochastic convective parameterization, *J. Atmos. Sci.*, *59*(5), 959–975.

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

Mason, S., C. Jakob, A. Protat, and J. Delanoë (2014), Characterizing observed midtopped cloud regimes associated with southern ocean shortwave radiation biases, *J. Clim.*, *27*(16), 6189–6203, doi:10.1175/JCLI-D-14-00139.1.

Mekonnen, A., and W. B. Rossow (2011), The interaction between deep convection and easterly waves over tropical north africa: A weather state perspective, *J. Clim.*, *24*(16), 4276–4294, doi:10.1175/2011JCLI3900.1.

Moncrieff, M. W., D. E. Waliser, M. J. Miller, M. A. Shapiro, G. R. Asrar, and J. Caughey (2012), Multiscale convective organization and the YOTC virtual global field campaign, *Bull. Am. Meteorol. Soc.*, *93*(8), 1171–1187, doi:10.1175/BAMS-D-11-00233.1.

Neale, R., and J. Slingo (2003), The maritime continent and its role in the global climate: A GCM study, *J. Clim.*, *16*(5), 834–848.

Oreopoulos, L., and W. B. Rossow (2011), The cloud radiative effects of International Satellite Cloud Climatology Project weather states, *J. Geophys. Res.*, *116*, D12202, doi:10.1029/2010JD015472.

Palmer, T., G. Shutts, R. Hagedorn, F. Doblas-Reyes, T. Jung, and M. Leutbecher (2005), Representing model uncertainty in weather and climate prediction, *Annu. Rev. Earth Planet. Sci.*, *33*(1), 163–193, doi:10.1146/annurev.earth.33.092203.122552.

Peters, K., C. Jakob, L. Davies, B. Khouider, and A. J. Majda (2013), Stochastic behavior of tropical convection in observations and a multicloud model, *J. Atmos. Sci.*, *70*(11), 3556–3575, doi:10.1175/JAS-D-13-031.1.

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

Rossow, W. B., and R. A. Schiffer (1999), Advances in understanding clouds from ISCCP, *Bull. Am. Meteorol. Soc.*, *80*(11), 2261–2287, doi:10.1175/1520-0477(1999)080<2261:AIUCFI>2.0.CO;2.

Rossow, W. B., G. Tselioudis, A. Polak, and C. Jakob (2005), Tropical climate described as a distribution of weather states indicated by distinct mesoscale cloud property mixtures, *Geophys. Res. Lett.*, *32*, L21812, doi:10.1029/2005GL024584.

Rossow, W. B., A. Mekonnen, C. Pearl, and W. Goncalves (2013), Tropical precipitation extremes, *J. Clim.*, *26*(4), 1457–1466, doi:10.1175/JCLI-D-11-00725.1.

Soden, B. J., and I. M. Held (2006), An assessment of climate feedbacks in coupled ocean-atmosphere models, *J. Clim.*, *19*(14), 3354–3360, doi:10.1175/JCLI3799.1.

Stachnik, J. P., C. Schumacher, and P. E. Ciesielski (2013), Total heating characteristics of the ISCCP tropical and subtropical cloud regimes, *J. Clim.*, *26*(18), 7097–7116, doi:10.1175/JCLI-D-12-00673.1.

Tan, J. (2014), The role of large-scale organisation of convection for tropical weather and climate, PhD thesis, Monash Univ., Melbourne, Australia.

Tan, J., and C. Jakob (2013), A three-hourly data set of the state of tropical convection based on cloud regimes, *Geophys. Res. Lett.*, *40*, 1415–1419, doi:10.1002/grl.50294.

Tan, J., C. Jakob, and T. P. Lane (2013), On the identification of the large-scale properties of tropical convection using cloud regimes, *J. Clim.*, *26*(17), 6618–6632, doi:10.1175/JCLI-D-12-00624.1.

Teixeira, J. A., and C. A. Reynolds (2008), Stochastic nature of physical parameterizations in ensemble prediction: A stochastic convection approach, *Mon. Weather Rev.*, *136*(2), 483–496, doi:10.1175/2007MWR1870.1.

Tompkins, A. M., and J. Berner (2008), A stochastic convective approach to account for model uncertainty due to unresolved humidity variability, *J. Geophys. Res.*, *113*, D18101, doi:10.1029/2007JD009284.

Tromeur, E., and W. B. Rossow (2010), Interaction of tropical deep convection with the large-scale circulation in the MJO, *J. Clim.*, *23*(7), 1837–1853, doi:10.1175/2009JCLI3240.1.

Tselioudis, G., and W. B. Rossow (2011), Time scales of variability of the tropical atmosphere derived from cloud-defined weather states, *J. Clim.*, *24*(3), 602–608, doi:10.1175/2010JCLI3574.1.

Tselioudis, G., E. Tromeur, W. B. Rossow, and C. S. Zerefos (2010), Decadal changes in tropical convection suggest effects on stratospheric water vapor, *Geophys. Res. Lett.*, *37*, L14806, doi:10.1029/2010GL044092.

Tselioudis, G., W. Rossow, Y. Zhang, and D. Konsta (2013), Global weather states and their properties from passive and active satellite cloud retrievals, *J. Clim.*, *26*(19), 7734–7746, doi:10.1175/JCLI-D-13-00024.1.

Tsushima, Y., M. A. Ringer, M. J. Webb, and K. D. Williams (2013), Quantitative evaluation of the seasonal variations in climate model cloud regimes, *Clim. Dyn.*, *41*(9–10), 2679–2696, doi:10.1007/s00382-012-1609-4.

Williams, K. D., and G. Tselioudis (2007), GCM intercomparison of global cloud regimes: Present-day evaluation and climate change response, *Clim. Dyn.*, *29*(2–3), 231–250, doi:10.1007/s00382-007-0232-2.

Williams, K. D., and M. J. Webb (2009), A quantitative performance assessment of cloud regimes in climate models, *Clim. Dyn.*, *33*(1), 141–157, doi:10.1007/s00382-008-0443-1.

Yang, G.-Y., and J. Slingo (2001), The diurnal cycle in the tropics, *Mon. Weather Rev.*, *129*(4), 784–801, doi:10.1175/1520-0493(2001)129<0784:TDCITT>2.0.CO;2.