A Stochastic Model for the Transition to Strong Convection

SAMUEL N. STECHMANN

Department of Mathematics, and Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, Los Angeles, California

J. DAVID NEELIN

Department of Atmospheric and Oceanic Sciences, and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, Los Angeles, California

(Manuscript received 7 January 2011, in final form 5 April 2011)

ABSTRACT

A simple stochastic model is designed and analyzed in order to further understand the transition to strong convection. The transition has been characterized recently in observational data by an array of statistical measures, including (i) a sharp transition in mean precipitation, and a peak in precipitation variance, at a critical value of column water vapor (CWV), (ii) an approximate power law in the probability density of precipitation event size, (iii) exponential tails in the probability density of CWV values, when conditioned on either precipitating or nonprecipitating locations, and (iv) long and short autocorrelation times of CWV and precipitation, respectively, with approximately exponential and power-law decays in their autocorrelation functions, respectively. The stochastic model presented here captures these four statistical features in time series of CWV and precipitation at a single location. In addition, analytic solutions are given for the exponential tails, which directly relates the tails to model parameters. The model parameterization includes three stochastic components: a stochastic trigger turns the convection on and off (a two-state Markov jump process), and stochastic closures represent variability in precipitation and in "external" forcing (Gaussian white noise). This stochastic external forcing is seen to be crucial for obtaining extreme precipitation events with high CWV and long lifetimes, because it can occasionally compensate for the heavy precipitation and encourage more of it. This stochastic model can also be seen as a simplified stochastic convective parameterization, and it demonstrates simple ways to turn a deterministic parameterization—the trigger and/or closure—into a stochastic one.

1. Introduction

The relationship between precipitation and water vapor is crucial for predictions of precipitation. In the context of convective parameterizations, for instance, one attempts to represent the statistics of unresolved moist convection in terms of the resolved water vapor and thermodynamic state. While the main statistic of interest has traditionally been mean values, it is becoming increasingly apparent that a wide range of statistics may be valuable. Several recent studies have documented an array of precipitation and water vapor statistics, and, taken together, these statistics can be seen as a characterization of the transition to strong convection.

One aspect of the transition to strong convection is an increase in precipitation with increasing column water vapor (CWV). Bretherton et al. (2004) found such an increase in satellite microwave data on daily and monthly time scales, and subsequent work has examined data on shorter time scales. Peters and Neelin (2006) and Neelin et al. (2009) showed that there is a critical value q_c of the CWV q beyond which the precipitation increases rapidly as an approximate power law $\langle P \rangle \sim (q - q_c)^{\beta}$, for $q > q_c$. Since the observed value of β is less than 1, the rate of increase in precipitation diminishes for higher values of q. Also, the precipitation variance has a strong peak at the critical value q_c. This is one piece of the transition to strong convection: mean precipitation is small for low values of CWV and large for high values of CWV, and a sharp, highly variable transition occurs near a critical value of CWV.

Another aspect of the transition to strong convection is the occurrence of extreme events. In this context, an "extreme" event could mean high CWV, high precipitation

Corresponding author address: Samuel N. Stechmann, Department of Mathematics, University of Wisconsin—Madison, Madison, WI 53706.
E-mail: stechmann@wisc.edu

DOI: 10.1175/JAS-D-11-028.1

rate, and/or high total precipitation; and the importance of these events depends on their frequency of occurrence. Peters and Neelin (2006) and Neelin et al. (2009) showed that high CWV values—above q_c—usually occur simultaneously with high precipitation rates. Furthermore, these extreme events occur more frequently than would be expected from Gaussian statistics, in the sense that the frequency of occurrence of high CWV values has a long tail with exponential decay as CWV increases. Another measure of extreme events is the total precipitation in a precipitation event. It has been shown that this measure, also called the ''event size,'' has a probability density function (PDF) with smaller events having higher probabilities (Peters et al. 2002; Neelin et al. 2008; Peters and Neelin 2009). However, the PDF has a long tail with power-law decay as event size increases, which is another indication that extreme events occur somewhat frequently and provides another way to quantify their frequency.

Finally, the transition to strong convection also includes interesting temporal variability. In particular, many studies have shown a temporal relationship with increases in lower-tropospheric water vapor leading increases in precipitation (Sherwood 1999; Sherwood and Wahrlich 1999; Sobel et al. 2004; Mapes et al. 2006, 2009). In addition to lower-tropospheric water vapor, CWV increases have also been shown to lead precipitation increases (Holloway and Neelin 2010). In this role of a convective precursor, CWV is likely acting as a surrogate for lower-tropospheric water vapor, since it has been shown that CWV variance corresponds mainly to moisture variance in the lower free troposphere around 800 hPa (Holloway and Neelin 2009). These results quantify the probability of precipitation and show that there tends to be a time lag before the initiation of the next precipitation event, even in periods of high CWV. This suggests that precipitation might not always initiate immediately when CWV exceeds some critical value, as might be expected from the notion of conditional instability. Likewise, for parameterizations of precipitation and convection, it may not be appropriate to model the onset of precipitation as a fixed critical threshold of CWV. Instead, it may be appropriate to model onset as a stochastic switch that has a mean ''critical value'' as in the notion of conditional instability, but which also has some random variance about this mean critical value. Such a model would be a ''stochastic trigger'' for the onset of precipitation/convection.

The observations described above should be useful in guiding and constraining the development of convective parameterizations: can one design a parameterization that captures the observational features described above? Given the nature of the observations, including the prominence of extreme events, one might suspect that a stochastic parameterization is appropriate. In this direction, one might ask: what is the simplest model that can capture all of the local features of the transition to strong convection described above?

The first purpose of this paper is to present such a simple stochastic model and to use it to lend insight into the statistics of extreme precipitation events. The second purpose of the present paper is to discuss this simple stochastic model in the context of stochastic convective parameterizations. As will be discussed below, the model designed here can be thought of in a number of ways: for instance, as a simplified version of some existing stochastic parameterizations, as a prototype around which a more comprehensive parameterization could be built, or as a simple way to ''stochasticize'' an existing deterministic parameterization (Lin and Neelin 2000; Bright and Mullen 2002; Majda and Khouider 2002; Majda et al. 2008; Majda and Stechmann 2008; Neelin et al. 2008; Plant and Craig 2008; Khouider et al. 2010). In all of these contexts, a key strength of the present model is its simplicity. As will be shown below, the model is simple enough that one can often directly relate the model parameters to the behavior of the model, which then makes it possible to choose model parameters based on observational data.

While the statistics described above are a significant portion of the transition to strong convection, there are other interesting statistics that will not be examined here. Of particular note are statistics of spatial variability of clouds and cloud clusters (Houze and Cheng 1977; Lovejoy 1982; Mapes and Houze 1993; Neggers et al. 2003; Nesbitt et al. 2006; Peters et al. 2009). The model of the present paper will include a single column and will not explicitly include spatial variability or correlations between columns. On a related note, one aspect of the observed spatial variability is long-range spatial correlations, which in turn are one of many aspects of the transition to strong convection that resemble continuous phase transitions and/or self-organized criticality from statistical physics (Peters et al. 2002; Peters and Neelin 2006; Neelin et al. 2009; Peters and Neelin 2009; Peters et al. 2009, 2010). On the other hand, Muller et al. (2009) offer a closely related (although differently phrased) interpretation for some aspects of the transition to strong convection (the mean and variance of precipitation as a function of CWV), making separate assumptions for the stochastic variations across the transition. The present paper takes the pragmatic point of view that the essential question is: what observed characteristics can be imitated in simpler systems, both for understanding and for use in climate model parameterizations?

The paper is organized as follows. The stochastic model is formulated in section 2. Numerical simulations and

simple analytic solutions are presented and compared with observational data in section 3. Implications for convective parameterizations are discussed in section 4. Finally, conclusions are summarized in section 5.

2. A stochastic model for tropical precipitation

What is the simplest model that can capture the observations of CWV and precipitation as described in section 1? The desired model should include, at a minimum, a time-varying degree of freedom that represents the CWV of a single column, which is denoted here as $q(t)$. Notice that this model does not explicitly include spatial variability or correlations between CWV of different columns. As described at the end of section 1, other interesting aspects can be seen in observed relationships of neighboring locations or clusters of locations on mesoscales (Peters et al. 2009), but these aspects will not be included in the model here. See section 4 for further discussion.

The time evolution of the CWV is then given by the differential equation

$$\frac{dq}{dt} = S, \tag{1}$$

where the source S is specified below. Three aspects of S will be defined stochastically below: a stochastic trigger for the onset and demise of convection (section 2a), a stochastic closure for precipitation (section 2b), and a stochastic closure for "other" forcing (also in section 2b). This "other forcing" will represent some forcing that would be accounted for by resolved dynamics if a large-scale model with spatial variability were used; but it also might represent some forcing that would be left unresolved, as discussed further below.

a. Stochastic trigger

What conditions should describe the onset of precipitation? Or, in other words, under what condition should the precipitation parameterization be turned on? One commonly used "trigger" condition is to turn on precipitation when $q > q_c$ for some critical threshold value and to turn off precipitation when $q < q_c$. This is a *deterministic* trigger condition. On the other hand, there are many unresolved factors, besides the resolved CWV q, that affect the onset of precipitation; hence, it may be best to model the onset (and demise) of precipitation as a *stochastic* switch or trigger (Lin and Neelin 2000, 2002; Majda and Khouider 2002; Khouider et al. 2003). In this spirit, a stochastic trigger is used here to turn the precipitation parameterization on and off. Motivated by the stochastic models of Majda and Khouider (2002) and Khouider et al. (2003), a Markov jump process $\sigma(t)$ is used to indicate whether the system is in the nonprecipitating state ($\sigma = 0$) or the precipitating state ($\sigma = 1$) at each time t (Lawler 1995; Gardiner 2004):

$$\sigma(t) = \begin{cases} 0 \Leftrightarrow \text{non-precipitating} \\ 1 \Leftrightarrow \text{precipitating} \end{cases}. \tag{2}$$

In (1) S will then take a different form depending on whether the system is in a nonprecipitating state or a precipitating state:

$$S = \begin{cases} S_0 & \text{if non-precipitating} \\ -S_1 & \text{if precipitating} \end{cases}, \tag{3}$$

where closures for S_0 and S_1 are given below in section 2b.

To specify the dynamics of $\sigma(t)$, we must specify the probability of a jump from $\sigma = 0$ to $\sigma = 1$ (and vice versa); these jumps correspond to the onset (and demise) of precipitation events. An intuitive way to think of the random jumps is the following. If $\sigma(t) = 0$ at some time t, then the probability that precipitation turns on—that is, the probability that $\sigma(t + \Delta t) = 1$—is given approximately by $r_{01}\Delta t$, where $r_{01}[q(t)]$ is the transition rate for the $0 \rightarrow 1$ transition from nonprecipitating to precipitating. A transition rate $r_{10}[q(t)]$ similarly defines the probability for the $1 \rightarrow 0$ transition from precipitating to nonprecipitating. The stochastic trigger is therefore specified once $r_{01}(q)$ and $r_{10}(q)$ are specified as functions of q. The functions used here are shown in Fig. 1. They have been defined to represent three atmospheric regimes: a low-CWV regime, a high-CWV regime, and a middle regime in between these two. If the atmosphere is sufficiently dry, then transitions from the nonprecipitating state to the precipitating state are essentially not allowed (by defining r_{01} to be nearly zero), while transitions from the precipitating state to the nonprecipitating state are allowed (by defining r_{10} to be positive). The opposite situation occurs if the atmosphere is sufficiently moist; and, for moderate CWV values, both transitions are allowed. Each of these functions shown in Fig. 1 is a stretched and shifted version of the tanh function:

$$f(q) = f_{-\infty} + (f_{+\infty} - f_{-\infty})\frac{1}{2}\left[1 + \tanh\left(\frac{q - q_{\text{mid}}}{q_{\text{width}}}\right)\right]. \tag{4}$$

The four parameters that determine this function are the asymptotic values of f at $q = \pm\infty$, $f_{\pm\infty}$; the location of the

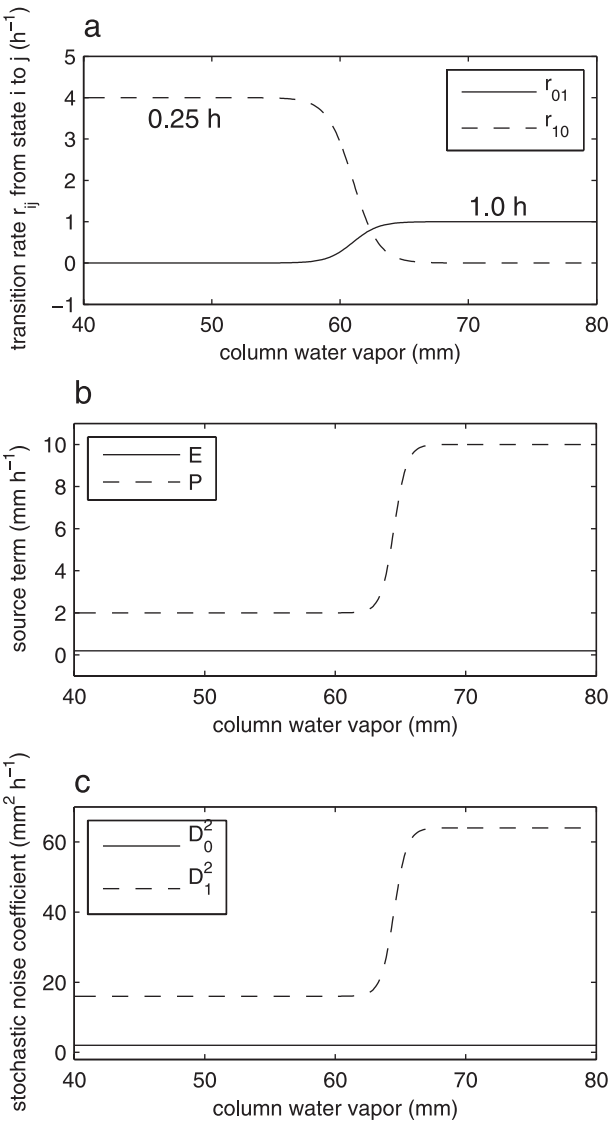

FIG. 1. Model parameters as functions of CWV. (a) Transition rates $r_{01}(q)$ and $r_{10}(q)$ for the stochastic jump process (trigger) $\sigma(t)$. (b) Deterministic parts $E(q)$ and $P(q)$ of the source term from (5). (c) Variances $D_0^2(q)$ and $D_1^2(q)$ of the stochastic part of the source term from (5).

transition q_{mid}; and the "width" of the transition q_{width}. The values of these parameters are given in Table 1 for each of the functions from Fig. 1. These particular values were chosen based on a combination of theory and observations, but a discussion of this is delayed until section 4, after the theory is presented in section 3b. Also note that, for any single feature of interest, only a subset of these parameters will be relevant; however, for achieving the full set of features of interest, the full model and parameter set are described in this section.

TABLE 1. Parameter values that determine the functions in Fig. 1 using the generic form of tanh in (4). Values of CWV q are given in millimeters, and units of other quantities are shown in the first column.

f	$f_{-\infty}$	$f_{+\infty}$	q_{mid}	q_{width}
r_{01} (h^{-1})	0	1.0	61.0	2.0
r_{10} (h^{-1})	4.0	0	63.0	2.0
E (mm h^{-1})	0.2	0.2	N/A	N/A
P (mm h^{-1})	2.0	10.0	64.5	1.0
D_0^2 (mm^2 h^{-1})	2.0	2.0	N/A	N/A
D_1^2 (mm^2 h^{-1})	16.0	64.04	64.5	1.0
D_E^2 (mm^2 h^{-1})	16.0	64.00	64.5	1.0
D_P^2 (mm^2 h^{-1})	0.0	0.04	64.5	1.0

b. Stochastic closures for precipitation and forcing

If the value of $\sigma(t)$ is given as described above, then the model (1) is fully specified once the sources S_0 and S_1 are chosen. Since these sources are meant to represent a variety of different moistening processes, we choose stochastic closures/forcing of the form

$$S_0 = E(q) + D_0(q)\eta(t),$$
$$S_1 = P(q) + D_1(q)\eta(t), \tag{5}$$

where $\eta(t)$ represents Gaussian white noise with mean 0 and variance 1.

Physically, S_0 is a parameterization of moistening processes such as moisture convergence from mesoscale and synoptic-scale waves; this is represented by a mean moistening $E(q)$ (for "evaporation," even though it parameterizes moistening in a broader sense) and by a stochastic noise term $D_0(q)\eta(t)$, which could actually be either positive or negative, corresponding to moisture convergence or divergence, respectively, by the large-scale flow. Similarly, $P(q)$ is a parameterization of mean precipitation, and $D_1(q)\eta(t)$ actually represents a combination of two stochastic noise terms:

$$S_1 = [P(q) + D_P(q)\eta_P(t)] + D_F(q)\eta_F(t). \tag{6}$$

In this form, S_1 has been split into a contribution from precipitation (in square brackets) and a contribution from "external" forcing $D_F(q)\eta(t)$, which represents processes such as moisture convergence/divergence from mesoscale and synoptic-scale waves. One would expect much of this external forcing to be accounted for by resolved dynamics in large-scale models with spatial variability, although it is possible that resolved dynamics would not account for the full intensity of $D_F\eta_F(t)$, as discussed further below. The definition (6) separates $D_1(q)$ to emphasize that it includes a contribution $D_P(q)$ due to variance in the precipitation rate

and a contribution $D_F(q)$ due to variance in the external forcing, where $D_1^2 = D_P^2 + D_F^2$. This also makes it clear that the precipitation rate will be defined as

$$\text{precip}(t) = \begin{cases} 0 & \text{if} \quad \sigma(t) = 0 \\ P(q) + D_P(q)\eta_P(t) & \text{if} \quad \sigma(t) = 1 \end{cases}. \tag{7}$$

If analyzed carefully in observations, the variability of these physical processes might best be described by red noise or even by Stratonovich white noise; here, however, to keep the model as simple as possible, $\eta(t)$ is taken to be white noise in the Ito sense.

The forms of $E(q)$, $P(q)$, $D_E(q)$, and $D_P(q)$ are shown in Fig. 1, and they are in the form of the tanh function from (4) with parameter values given in Table 1. These particular values were chosen based on a combination of theory and observations, but a discussion of this is delayed until section 4, after the theory is presented in section 3b. Also, the form of the tanh function was chosen partly to facilitate analytical work in particular regimes of interest. The nonprecipitating state has constant parameters $E = 0.2$ mm h^{-1} and $D_0^2 = 2.0$ mm^2 h^{-1} that do not depend on q. On the other hand, the precipitating state has parameters $P(q)$ and $D_1^2(q)$ that vary with CWV q. As was the case for the transition rates $r_{01}(q)$ and $r_{10}(q)$, the variations of $P(q)$ and $D_1(q)$ define three atmospheric regimes: a low-CWV regime, a high-CWV regime, and a middle regime in between these two. For simplicity, the function values are roughly constant within each of the low-CWV and high-CWV regimes, and the functions have smooth, yet somewhat sharp, transitions between these two regimes. The low-CWV regime is characterized by a weaker precipitation rate of 2 mm h^{-1} and a weaker stochastic variance of 16.0 mm^2 h^{-1}; this could represent either the early stage of development of a convective system or the final, dying stage of a convective system with stratiform rain; the two alternatives would be distinguished by the previous history of $q(t)$. In contrast, the high-CWV regime is characterized by a stronger precipitation rate of 10 mm h^{-1} and a stronger stochastic variance of 64.0 mm^2 h^{-1}; this could represent a fully developed convective system with heavy rainfall and with intense variations in moisture convergence/divergence from updraft/downdraft circulations and mesoscale and synoptic-scale waves. The particular value of 10 mm h^{-1} was chosen based on the microwave estimate of observed rainfall (Hilburn and Wentz 2008) used in Neelin et al. (2009) (noting that high-end rain rate estimates differ in amplitude among observational products). We emphasize that the intense stochastic forcing $D_1(q)\eta(t)$ could be either positive or negative; if positive, it represents a moisture source from, for instance, mesoscale/synoptic-scale moisture convergence, and it could possibly be so intense that it compensates (or overcompensates) the moisture loss from precipitation.

To find numerical solutions of this stochastic model, the following method is used. A uniform time step of 0.01 h is chosen to be comparable with the 1-min resolution of the observational data analyzed by Holloway and Neelin (2010). At each time step, the model variables are updated in two stages: (i) the CWV $q(t)$ is updated to $q(t + \Delta t)$ using $dq/dt = S$ with the value of $\sigma(t)$ fixed, and (ii) $\sigma(t)$ is updated to $\sigma(t + \Delta t)$ with the value of $q(t + \Delta t)$ fixed. For (i), the stochastic differential equation is advanced in time using the Euler–Maruyama method (Higham 2001; Gardiner 2004). For (ii), the probability of switching states in a time interval Δt is $1 - e^{-\gamma \Delta t}$, where r is either r_{01} or r_{10}. Pseudorandom numbers are generated using the Mersenne Twister algorithm (Matsumoto and Nishimura 1998), and random variables distributed uniformly on the interval (0, 1] are converted to Gaussian distributed random variables using the Box–Muller method (Ross 1998). The model is advanced for 4×10^6 time steps, or approximately 4.5 yr, which is comparable to the longest time series of these fields, with similar temporal resolution, in currently available observational records. An even longer simulation of approximately 13.5 yr was used for comparison (see Figs. 3 and 4). Such a long time series is necessary because one focus of this paper is extreme events, which occur somewhat rarely, and many samples of these events must be gathered in order to accurately compute their statistics.

3. Properties of the transition in numerical simulations and analytic solutions

In section 3a, it is shown that the simple model from section 2 can capture all of the locally analyzed observations of the transition to strong convection, as described in section 1. Also, in section 3b, simple analytic formulas are derived for some features of the model behavior; these formulas indicate how model behavior depends on parameter choices, which lessens the need for exhaustive numerical sensitivity studies.

a. Numerical simulations

Using the model formulation of section 2, a time series of CWV $q(t)$ and $\sigma(t)$ is created. Figure 2 shows a sample of this time series over different intervals of time, zooming in from the broad picture to detailed events, showing 100- (Fig. 2a), (b) 5- (Fig. 2b), and 0.5-day intervals (Figs. 2c,d). In Fig. 2a, the 100-day interval shows that the CWV can take values ranging from 30 to 80 mm, although the extreme values occur somewhat

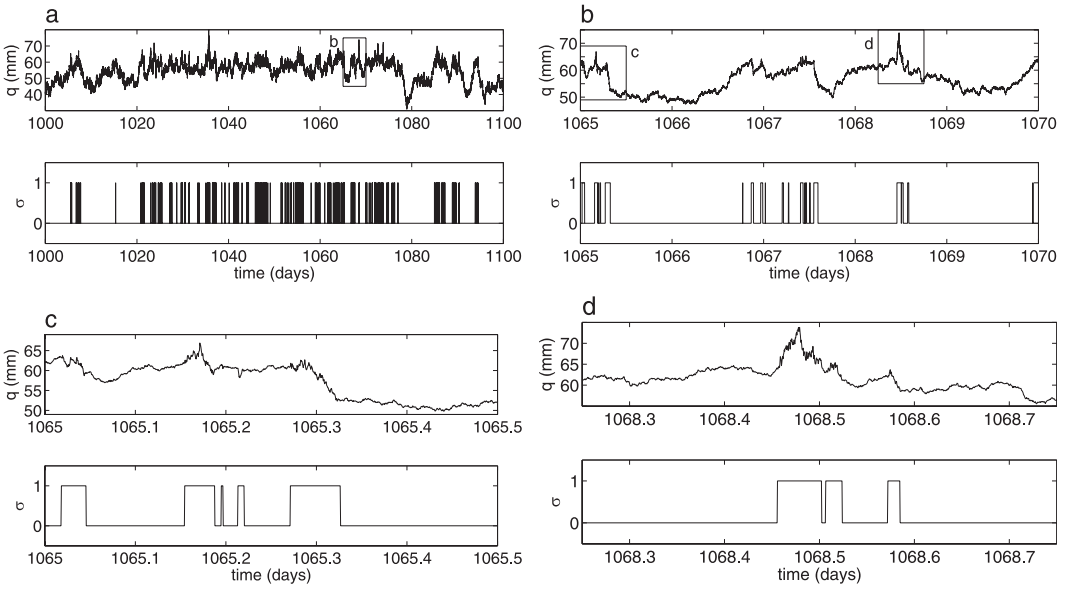

FIG. 2. Sample time series of $q(t)$ and $\sigma(t)$. Temporal variations are shown over intervals of (a) 100, (b) 5, and (c),(d) 0.5 days. The box in (a) shows the data used in (b), and the two boxes in (b) show the data used in (c) and (d). Compare (b) with the observational data in Fig. 2 of Holloway and Neelin (2010).

rarely. An individual precipitation event begins when σ jumps from 0 to 1 and ends when σ jumps from 1 to 0, although individual events occur on time scales much shorter than 100 days and are therefore hard to identify in Fig. 2a. Nevertheless, this plot shows that precipitation events occur irregularly, and there are several periods of up to 1 week when there is no precipitation at all (when $\sigma = 0$). Figure 2b zooms in on a 5-day interval from Fig. 2a. This plot shows the synoptic-scale and mesoscale variability, and it is comparable to the sample time series from observational data in Holloway and Neelin (2010). Again, precipitation events occur irregularly, and there are periods of 1–2 days without precipitation as well as periods with clusters of several precipitation events. Figures 2c and 2d zoom in on two different clusters of precipitation events; these plots show five and three events, respectively.

In Fig. 2, the life cycles of the precipitation events are highly variable, yet some of them appear to follow similar patterns. One might naively expect that precipitation events always follow the pattern of the third event in Fig. 2d: initially, there is a period of slow moistening until the atmosphere is moist enough to allow convection to initiate; then precipitation begins and rapidly dries the atmosphere; and finally, after precipitation has sufficiently reduced the moisture level, the precipitation event ends. While this does occur for some events in Figs. 2c and 2d, there are other events that do not follow this simple pattern, and those events tend to occur with high CWV values and tend to have long lifetimes. Two examples are the second event in Fig. 2c and the first event in Fig. 2d. These events follow a different pattern: precipitation begins, but CWV *increases* after the initiation of convection; after initially increasing, the CWV eventually begins to decrease; finally, after the CWV decreases sufficiently, the precipitation event ends. The key difference here is that these events include a period where CWV increases at the same time as heavy precipitation. This requires significant moistening from $D_F(q)\eta(t)$ from (5) and (6). Hence, these extreme precipitation events represent events occurring as part of mesoscale convective systems, hurricanes, or other sources of intense variations in moistening/drying, presumably from moisture convergence/divergence. It is these extreme precipitation events that will strongly factor into some of the statistical measures described below. We note that, in the observed system or a full atmospheric model, there would be feedback between the convective heating and moisture convergence that is not included here.

The autocorrelation function for the time series data is shown in Fig. 3a for CWV and for precipitation. This is in broad agreement with the autocorrelations from observational data in Fig. 1 of Holloway and Neelin (2010): the CWV is well correlated with itself at much later times of roughly 1 day, whereas the precipitation has appreciable autocorrelation only for roughly 1 h. This is further confirmation of the realism of the variability in the model, in addition to visual comparison of time series of $q(t)$.

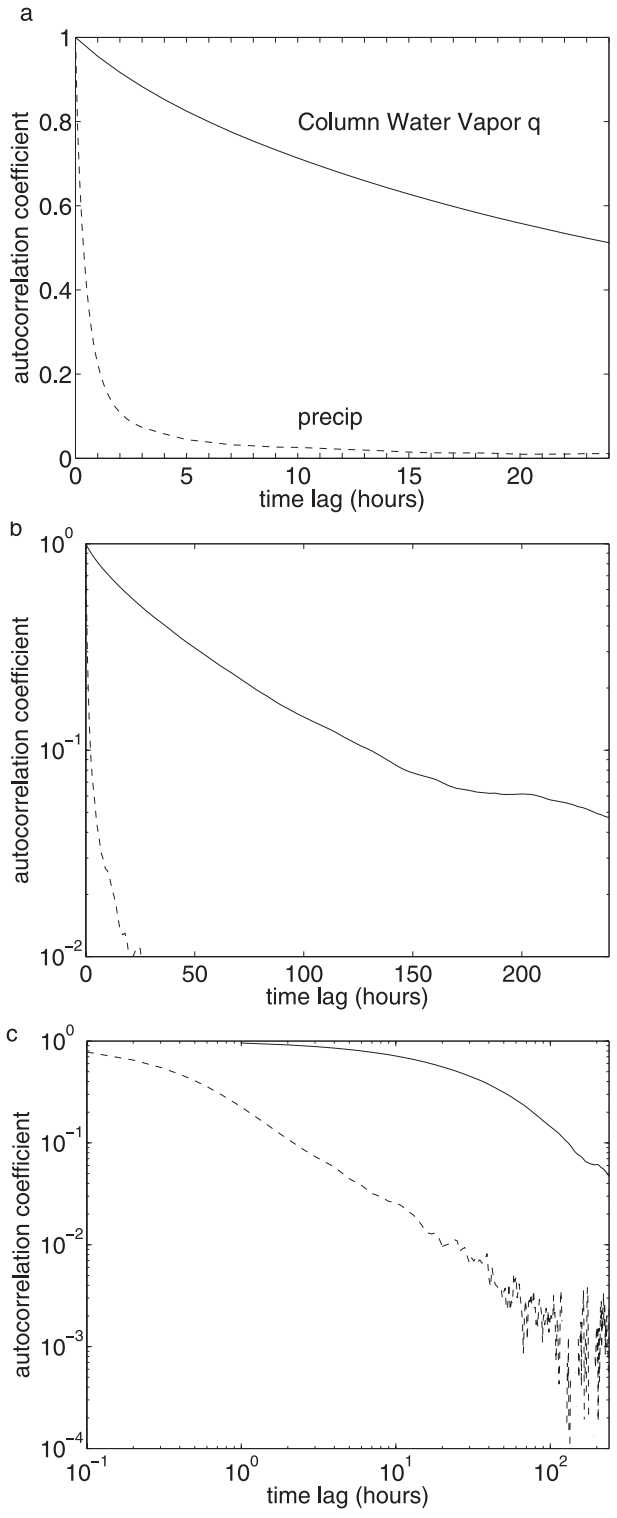

FIG. 3. Autocorrelation as a function of time lag for column water vapor (solid) and precipitation (dashed). (a) Lags of 0–24 h, with linear–linear axis scaling. (b) Lags of –240 h, with log–linear axis scaling. (c) Lags of 0.1–240 h, with log–log axis scaling. Compare to the observational data in Fig. 1 of Holloway and Neelin (2010).

Furthermore, Figs. 3b and 3c indicate the functional forms of the autocorrelation functions on mesoscale and synoptic time scales. First, the CWV autocorrelation function is shown on a log–linear plot and has a roughly, but not precisely, exponential form for lags of about 1–6 days. Note that this stochastic model does not necessarily have to have an exponential CWV autocorrelation function, yet this form (approximately) arises. Second, the precipitation autocorrelation function is shown on a log–log plot and has a roughly power-law form for lags of about 30 min–4 days. The exponent of this power law is roughly 1, which is fairly close to that estimated from optical gauge data (Holloway and Neelin 2010) of roughly 0.8. Hence, while the precipitation autocorrelation time appears to be short, there are also long-range correlations out to at least 4 days. The short time-scale end of this range, near 30 min, may be set here by model parameters. A likely candidate is the time scale $D_1^2/P^2 \approx 0.6$ h, which arises from the precipitation scale P in the precipitating range combined with a moisture scale D_1^2/P (which arises in both core and tail of the precipitating regime, discussed in section 3b below). Beyond this, there is a range for which there appears to be no clear time scale—a circumstance that tends to lead to power-law behavior as summarized in Peters and Neelin (2009).

To further isolate the behavior seen in Fig. 3, several sensitivity studies (not shown) were carried out. Two cases led to much more rapid decay of the precipitation autocorrelation with no power-law range: a two-state jump process alone or the full model with all white noise shut off (the latter also yields more rapid exponential decay of CWV autocorrelation than in Fig. 3). On the other hand, in a version of the model with the stochastic jumps replaced by a deterministic Heaviside function at $q = 61$ mm, the approximate exponential and power-law decays of Figs. 3b and 3c were still produced. Thus, the source of the power-law range in the autocorrelation clearly involves the rapid switches of precipitation interacting with time-dependent moisture variations, forced by the noise.

Figure 4 shows the number of occurrences of the total precipitation that falls in individual precipitation events, defined as an uninterrupted sequence of time steps with $\sigma = 1$. For instance, Fig. 2c shows five events and Fig. 2d shows three events. Figure 4 shows that smaller events are most common, and the PDF roughly follows two power laws, one each for events smaller and larger than about 10 mm. Numerical sensitivity studies (not shown) indicate that long, approximately power-law tails are roughly representative of typical behavior, and they can even be seen when the stochastic jumps are replaced by a deterministic Heaviside function; that is, while these

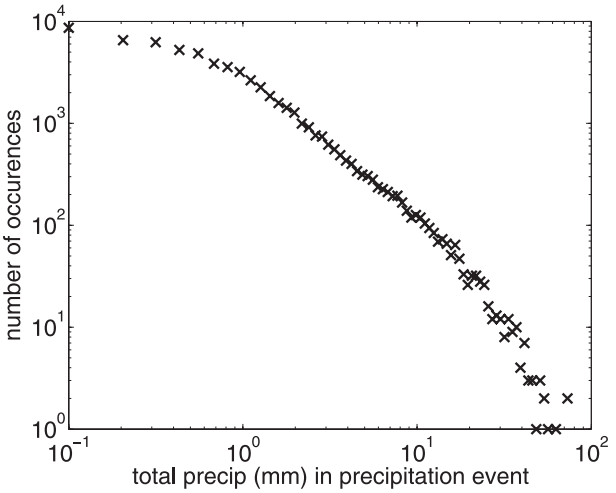

FIG. 4. Density function of total precipitation of precipitation events. Compare to the power laws seen in Fig. 6b of Neelin et al. (2008), in Fig. 1 of Peters and Neelin (2009), and in Figs. 2 and 3a of Peters et al. (2010), for heavy events with total size of greater than 1 mm.

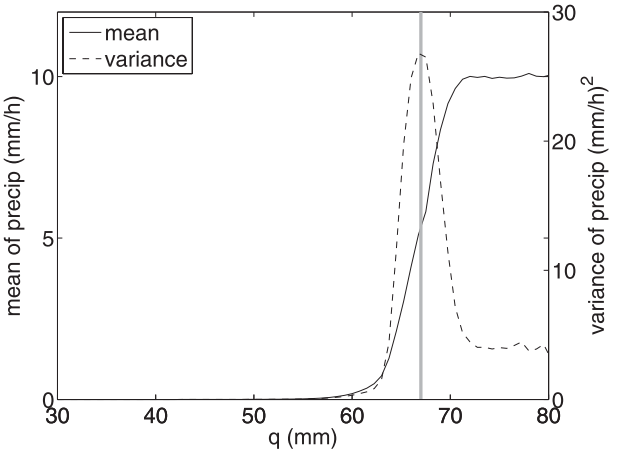

FIG. 5. Mean and variance of precipitation conditioned on the value of q. The thick gray line displays the value of q, 67 mm, at which the precipitation variance is maximum. Compare to the observational data in Fig. 1 of Peters and Neelin (2006) and Figs. 1, 2, and 4 of Neelin et al. (2009).

data do not always bear a precise linear fit, they are always slowly decaying with significant contributions from extreme precipitation events. This is in qualitative agreement with the observational data presented in Fig. 6b of Neelin et al. (2008), in Fig. 1 of Peters and Neelin (2009), and in Figs. 2 and 3a of Peters et al. (2010).

In fact, the approximate power-law decay in Fig. 4 also bears quantitative comparison with observations. For the heavy precipitation events of greater than about 10 mm, the occurrence frequency in Fig. 4 falls by about 2 factors of 10 over a range of about 1 factor of 10 in total precipitation. This 2-to-1 ratio is also approximately seen in Neelin et al. (2008) and Peters and Neelin (2009) for events with similar total size of greater than 10 and 1 mm, respectively: the PDF falls by about 4 factors of 10 over a range of about 2 factors of 10 in event size in both Neelin et al. (2008) and Peters and Neelin (2009). Notice that this is different from a second power law seen in the observations for smaller event sizes with total precipitation of 0.001–1.0 mm, which is emphasized in Neelin et al. (2008), Peters and Neelin (2009), and Peters et al. (2010) by the power-law fit lines. For these lighter precipitation events, the power law is flatter, as is represented somewhat here in Fig. 4 for lighter events 1–10 mm in size (although the slope appears to slightly differ from observations). Here, however, the emphasis is on the heavy precipitation events, for which the power-law decay carries significance as a measure of extreme events and their frequency: a power-law decay indicates a relatively high frequency of occurrence of extreme, high-total-precipitation events (relatively high frequency in comparison to Gaussian, or even exponential, decay).

Figure 5 shows the mean and variance of precipitation for each value of q. Near $q \approx 67$ mm, there is a rapid increase of the mean and a sharp peak in the variance, in qualitative agreement with the observational data in Figs. 1 and 4 of Neelin et al. (2009). Note that the precise functional form of the mean here does not appear to be a power law for high CWV, as it is in observations (Peters and Neelin 2006); such a detailed comparison is not the focus here. Nevertheless, a power law could actually be achieved in this model by giving $P(q)$ a power-law form for high CWV instead of the simple tanh function; this is because the mean precipitation is essentially equal to $P(q)$ for high CWV, where σ is essentially always equal to 1. The mean and variance were also previously recovered by the model of Muller et al. (2009), which is a stochastic two-layer model for a single column. The other local features of the transition to strong convection, however, were not demonstrated in Muller et al. (2009); the model did not include evolution in time.

Which physical processes and which aspect of the model create the variance peak in Fig. 5? Notice that the precipitation, as defined in (7), is affected by both $\sigma(t)$ and $D_P\eta(t)$. Is the variance peak in Fig. 5 due to the variance in $\sigma(t)$, or to the Gaussian white noise fluctuations $D_P\eta(t)$, or both? It is, in fact, $\sigma(t)$ that creates the variance peak. This can be seen directly from Table 1: since $D_P^2 = 0.04\,\mathrm{mm^2\,h^{-1}}$ at most, and since the time step is 0.01 h, the variance due to these fluctuations is at most 4 $\mathrm{mm^2\,h^{-2}}$. This would occur if $\sigma = 1$ always for a particular value of q, and, in fact, it is the asymptotic value of the precipitation variance in Fig. 5 for large values of q. However, the peak variance is much larger

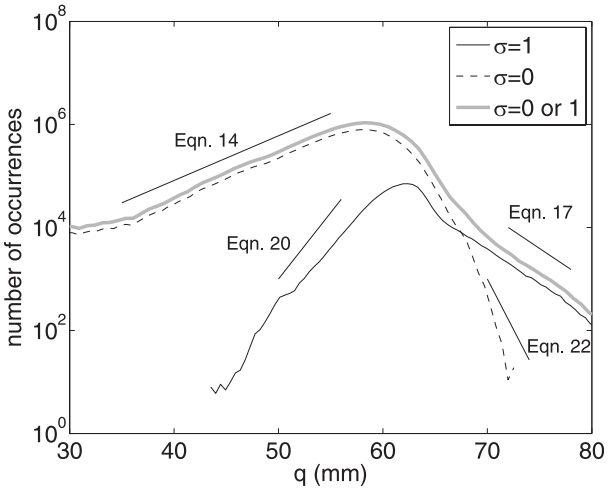

FIG. 6. Density function of q conditioned on the value of σ [i.e., conditioned on precipitating points ($\sigma = 1$) or on nonprecipitating points ($\sigma = 0$) or on all points ($\sigma = 0$ or 1)]. Line segments display theoretical slopes, whose formulas given in the annotated equations. Compare to the observational data in Figs. 5 and 9 of Neelin et al. (2009).

than this and is therefore due to the variability in $\sigma(t)$. In terms of physical processes, this means that the variance peak is caused not by the precipitation fluctuations within an individual event, but by the onset, and demise, of precipitation events (and all of the processes—cold pools, gravity waves, etc.—that contribute to the onset and demise). This is one example of the important role of $\sigma(t)$.

Figure 6 shows the number of occurrences of $\sigma = 0$ and $\sigma = 1$ for each value of q. This figure can be compared with the observational data shown in Figs. 6, 9, and 10 of Neelin et al. (2009), where $\sigma = 0$ corresponds to nonprecipitating points and $\sigma = 1$ corresponds to precipitating points. The model results here are in qualitative agreement with the observational data in Neelin et al. (2009): the PDFs have roughly exponential tails; overall, the nonprecipitating state occurs more frequently than the precipitating state; on the other hand, for sufficiently moist q values, the precipitating state occurs more frequently than the nonprecipitating state; and the most probable q value is slightly drier than the q value where nonprecipitating and precipitating states are equally probable (which is close to the q value of peak variance from Fig. 5). In addition, the exponential decay rates in Fig. 6 are in agreement with simple theoretical predictions, as indicated by the thin gray line segments in Fig. 6, and as described next.

b. Analytic solutions and interpretations

The exponential tails in Fig. 6 can be understood in terms of simple stochastic processes with simple physical interpretations. Several cases are described in this subsection. First, the general equation is described for the PDF $\mathbf{p}(q, t) = [p_0(q, t), p_1(q, t)]^{\mathrm{T}}$ of the stochastic process of section 2. Next, specific limiting forms of the general equation are considered and solved analytically. Each of these specific limits applies to one of the exponential tails seen in Fig. 6. Finally, some alternative simplified cases are discussed.

1) GENERAL MASTER–FOKKER–PLANCK EQUATION

Two types of stochastic processes enter into the model formulated in section 2: (i) the stochastic closure/forcing as a drift–diffusion process, whose PDF can be described by a Fokker–Planck equation, and (ii) the stochastic trigger as a Markov jump process, whose PDF can be described by a Master equation (Lawler 1995; Gardiner 2004). The PDF of interest here is a q-dependent vector PDF $\mathbf{p}(q, t) = [p_0(q, t), p_1(q, t)]^{\mathrm{T}}$, where $p_j(q, t)$ is the probability that $\sigma(t) = j$ for $j = 0, 1$.

First consider each of (i) and (ii) in isolation from the other, from which the full dynamics can be built by adding these two contributions. For the drift–diffusion process (i), in isolation from (ii), the evolution is given by the Fokker–Planck equations

$$\partial_t p_0 + \partial_q (E p_0) = \partial_q^2 \left(\frac{1}{2} D_0^2 p_0 \right), \tag{8}$$

$$\partial_t p_1 - \partial_q (P p_1) = \partial_q^2 \left(\frac{1}{2} D_1^2 p_1 \right). \tag{9}$$

On the other hand, for the jump process (ii), in isolation from (i), the evolution is given by a Master equation (Lawler 1995; Gardiner 2004):

$$\partial_t \begin{pmatrix} p_0 \\ p_1 \end{pmatrix} = \begin{pmatrix} -r_{01} & r_{10} \\ r_{01} & -r_{10} \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \end{pmatrix}. \tag{10}$$

For the combination of (i) and (ii), the evolution is given by a combined Master–Fokker–Planck equation, which is a combination of (8), (9), and (10):

$$\partial_t \begin{pmatrix} p_0 \\ p_1 \end{pmatrix} + \partial_q \left[\begin{pmatrix} E & 0 \\ 0 & -P \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \end{pmatrix} \right] = \frac{1}{2} \partial_q^2 \left[\begin{pmatrix} D_0^2 & 0 \\ 0 & D_1^2 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \end{pmatrix} \right] + \begin{pmatrix} -r_{01} & r_{10} \\ r_{01} & -r_{10} \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \end{pmatrix}. \tag{11}$$

Since we are interested in understanding the stationary PDF in Fig. 6, we use the steady-state version of this equation:

$$\partial_q\left[\begin{pmatrix} E & 0 \\ 0 & -P \end{pmatrix}\begin{pmatrix} p_0 \\ p_1 \end{pmatrix}\right] = \frac{1}{2}\partial_q^2\left[\begin{pmatrix} D_0^2 & 0 \\ 0 & D_1^2 \end{pmatrix}\begin{pmatrix} p_0 \\ p_1 \end{pmatrix}\right] + \begin{pmatrix} -r_{01} & r_{10} \\ r_{01} & -r_{10} \end{pmatrix}\begin{pmatrix} p_0 \\ p_1 \end{pmatrix}. \quad (12)$$

While this Master–Fokker–Planck equation is too complicated to solve analytically, some limiting cases can be solved analytically for each of the exponential tails in Fig. 6, each of which corresponds to a different atmospheric regime.

2) NONPRECIPITATING, LOW-CWV REGIME

Consider the nonprecipitating, low-CWV regime of the general steady Master–Fokker–Planck equation in (12). In other words, assume that the system essentially remains in the nonprecipitating state $\sigma = 0$ (so that $p_1 \approx 0$). Also, assume that the CWV is low (less than roughly 60 mm) so that so that the probability of precipitation onset is low: $r_{01} \approx 0$. In this case, the stationary PDF $p_0(q)$ satisfies an approximate limiting form of (12):

$$E\partial_q p_0 = \frac{D_0^2}{2}\partial_q^2 p_0, \quad (13)$$

where E and D_0 take the constant, low-CWV limiting values from Fig. 1 and Table 1. This Fokker–Planck equation has solutions of the form

$$p_0(q) = A\exp\left(\frac{2E}{D_0^2}q\right), \quad (14)$$

which is valid in the low-CWV regime. This process gives intuitive insight into the dominant physical mechanisms in this case, as supported by the good agreement with the model simulation: in Fig. 6 the slope of the thin gray line from $q = 35$ to 55 mm is given by $2E/D_0^2$. Note that this limiting case is as if the CWV were evolving according to

$$\frac{dq}{dt} = E + D_0\eta(t), \quad (15)$$

where a suitable boundary condition would be needed to keep the system in the low-CWV regime and to produce a stationary PDF. For the full model here, there is a matching condition to the other CWV/precipitation regimes; for an even simpler model presented below in section 3b, piecewise matching is not necessary because the solutions extend across several regimes. Also note that in observations this part of the distribution (low CWV, nonprecipitating) can vary considerably. The dependence of E on q can be significant in affecting this, but the details of this part of the distribution are not the focus here.

3) PRECIPITATING, HIGH-CWV REGIME

Similarly, one can consider the precipitating case where it is assumed that $\sigma = 1$ always and that the CWV is high (greater than roughly 70 mm). In other words, it is assumed that $p_0 \approx 0$ and $r_{10} \approx 0$. In this case, the stationary PDF $p_1(q)$ satisfies an approximate limiting form of (12):

$$-P\partial_q p_1 = \frac{D_1^2}{2}\partial_q^2 p_1, \quad (16)$$

where it is understood that, in this approximate scenario, P and D_1 have constant values given by their asymptotic values at high CWV. This Fokker–Planck equation has solutions of the form

$$p_1(q) = A\exp\left(-\frac{2P}{D_1^2}q\right), \quad (17)$$

which is valid in the high-CWV regime. This is in good agreement with the model simulation, as displayed in Fig. 6 by the thin gray line from $q = 72$ to 78 mm. This limiting case is as if the CWV were evolving according to

$$\frac{dq}{dt} = -P + D_1\eta(t). \quad (18)$$

This corresponds to a precipitating column with white noise forcing, and, similar to (15), a boundary or matching condition would be needed at the low-CWV end of this regime in order to obtain a stationary PDF. Note that any slowly varying functions $P(q)$ and $D_1(q)$ should lead to approximately exponential solutions, even if they are not precisely constant functions. In fact, while it was assumed that P and D_1 were constant for the derivation presented, a similar derivation would apply for general functions $P(q)$ and $D_1(q)$, in which case the solutions would not be exactly exponential.

This case lends insight into the ''extreme'' precipitation events that occur at high CWV values and have large precipitation rates in the mean. The analytic solution suggests that the frequency of occurrence of these events decays exponentially as q increases, with a decay rate of $2P/D_1^2$. Since $D_1^2 = D_P^2 + D_F^2 \approx D_F^2$, this decay is determined by the mean P and the variance D_F^2 of external moistening/drying processes. If D_F^2 of this external

forcing were small (relative to P), then the exponential decay would be very rapid, and large values of CWV would not be observed in nature. However, that is not the case: since D_F^2 is somewhat comparable to P in magnitude, a slower exponential decay with a long tail is observed, and large values of CWV are observed relatively frequently.

4) PRECIPITATING, LOW-CWV REGIME

Consider now the case of low CWV (less than roughly 60 mm). For such low CWV values, the probability of precipitation onset is essentially zero. In (12), this corresponds to the approximation $r_{01} \approx 0$, which decouples the equation for p_1:

$$-P\partial_q p_1 = \frac{D_1^2}{2}\partial_q^2 p_1 - r_{10} p_1. \tag{19}$$

This equation has exponential solutions for the precipitating state of the form

$$p_1(q) = A\exp(m_1 q), \quad m_1 = \frac{-P + \sqrt{P^2 + 2D_1^2 r_{10}}}{D_1^2} > 0. \tag{20}$$

This curve is shown in Fig. 6 by the thin gray line from $q = 50$ to 56 mm, and it is in good agreement with p_1 from the model simulation. This case corresponds to the PDF of precipitation events at the end of their lifetimes: it is a competition among $P(q)$, $D_F(q)\eta(t)$, and the event-ending stochastic jump from $\sigma = 1$ to $\sigma = 0$. Also notice that this case is *evidence of stochastic hysteresis in the life cycle of some precipitation events*. Since the probability of precipitation onset is essentially zero for roughly $q <$ 58 mm, the events that contribute to this tail likely began at high CWV values and ended at low CWV values. Such behavior is made possible by $\sigma(t)$, which allows precipitation events to continue, with some probability, even after CWV has been reduced significantly. This may represent stratiform rain at the end of a precipitation event.

5) NONPRECIPITATING, HIGH-CWV REGIME

Similarly, consider the case of high CWV (greater than roughly 70 mm). For such high CWV values, the probability that a precipitation event ends is essentially zero. In (12), this corresponds to the approximation $r_{10} \approx 0$, which decouples the equation for p_0:

$$E\partial_q p_0 = \frac{D_0^2}{2}\partial_q^2 p_0 - r_{01} p_0. \tag{21}$$

This equation has exponential solutions of the form

$$p_0(q) = A\exp(-m_0 q), \quad m_0 = \frac{-E + \sqrt{E^2 + 2D_0^2 r_{01}}}{D_0^2} > 0. \tag{22}$$

This curve is shown in Fig. 6 by the steep thin gray line from $q = 64$ to 68 mm, and it is in good agreement with p_0 from the model simulation. This is relatively good agreement with the model simulation in light of the scarce samples of these events. This case corresponds to the PDF of a nonprecipitating column just before the onset of precipitation: it is a competition among mean moistening E, $D_0\eta(t)$, and the event-beginning jump from $\sigma = 0$ to $\sigma = 1$.

6) INCORRECT CASE THAT SHOWS THE IMPORTANCE OF STOCHASTIC VARIANCE

Note that a slightly simpler case than the combined Master–Fokker–Planck equation would also give exponential tails. If the $\partial_q^2\mathbf{p}$ diffusion part of (12) is ignored, the result is a system with the competing effects of deterministic drift and stochastic jumps. While a simplified version of this system also has stationary PDF solutions with exponential tails, the analytic solutions do not agree with the numerical solutions in Fig. 6. The slopes that would result would be $m_1 = r_{10}/P$ and $m_0 = r_{01}/E$, and their numerical values would be roughly 4 and 5 times steeper than those shown in Fig. 6. In other words, the precipitating state would be seen much more rarely at low CWV values, and the nonprecipitating state would be seen much more rarely at high CWV values. This is because the stochastic forcing $D_i(q)\eta(t)$ (associated with moisture convergence variations) plays an important role in forcing the CWV, in addition to the mean components $P(q)$ and $E(p)$; and the interaction of this with the jump process is important in (20) and (22).

7) SIMPLIFIED CASE WITHOUT EXPLICIT JUMPS

Another simple case lends further insight into the form of the PDFs in Fig. 6, which display cores that appear approximately Gaussian and tails that appear approximately exponential. Suppose the CWV evolves by the equation

$$\frac{dq}{dt} = \overline{S}(q) + D\eta(t), \tag{23}$$

where $\overline{S}(q)$ changes continuously from $+E$ at low CWV to $-P$ at high CWV. This q-dependent drift $\overline{S}(q)$ is meant to implicitly include the effect of the stochastic jumps. In this case, the Fokker–Planck equation for $p(q, t)$ is

$$\partial_t p + \partial_q(\overline{S}p) = \left(\frac{D^2}{2}\right)\partial_q^2 p, \tag{24}$$

which has stationary solutions that satisfy

$$\partial_q \log p = \frac{2}{D^2}\overline{S}(q) \tag{25}$$

and thus

$$p(q) = A\exp\left[\frac{2}{D^2}\int_{q_{\rm ref}}^{q} \overline{S}(q')dq'\right], \tag{26}$$

where $q_{\rm ref}$ is a reference value that can conveniently be taken equal to the midpoint $q_{\rm mid}$ of the transition region.

As a simple, concrete example, $\overline{S}(q)$ could be chosen to have the piecewise linear form

$$\overline{S}(q) = \begin{cases} +E_* & \text{if} \quad -\infty < q < -q_* \\ S_a - \dfrac{S_*}{q_*}q & \text{if} \quad -q_* < q < +q_*\,, \\ -P_* & \text{if} \quad +q_* < q < +\infty \end{cases} \tag{27}$$

where q is given relative to $q_{\rm mid}$, and $2q_*$ is the width of the transition region. This replaces the effect of $\sigma(t)$ by a linear transition from constant evaporation E_* to constant precipitation P_*, with $S_* = (P_* + E_*)/2$. The asymmetry between E_* and the larger P_* is measured by $S_a = (P_* - E_*)/2$.

Because (25) gives the slope of the PDF on the log plot, as in Fig. 6, it provides a straightforward view of the relation between core and tail regimes. Below the transition region, the slope is constant—that is, the PDF is exponential with a decay scale E_*/D^2 that corresponds to (14), except that we have neglected the variation of D from the nonprecipitating to precipitating regime. Above the transition region, the PDF is likewise exponential with slope $-P_*/D^2$ that corresponds to (17). Within the transition region, the slope varies linearly (i.e., $\log p$ has a quadratic term). If there were no asymmetry between E_* and P_* (i.e., $S_a = 0$), there would be an exactly "Gaussian" core in the transition region:

$$p(q) = A\exp\left(-\frac{S_*}{D^2 q_*}q^2\right), \quad -q_* < q < +q_*. \tag{28}$$

In presence of asymmetry, this is modified by an exponential term, corresponding to the tilt of the quadratic core in plots of $\log p$. From (25), it may be seen that in this solution the slope of the exponential tail (in $\log p$) necessarily matches the slope of the core at the boundary of the transition region when $\overline{S}(q)$ is continuous. Note that the moisture scale associated with this Gaussian is approximately the same as the moisture scale of the exponential tail (17) in the precipitating high-CWV regime D_1^2/P.

This simple case provides a prototype for the PDF of all points (precipitating and nonprecipitating) in Fig. 6. It illustrates one simple way of producing a distribution with exponential tails and a (modified) Gaussian core, in which $\overline{S}$ is approximately independent of q in each of the tail regions, changing sign smoothly in the core region. Although the simple case can yield asymmetry between the two tails, it is missing some features. Comparison of this simple case to (14) and (17) shows that the increase in the stochastic noise coefficient $D_1(q)$ in the full model must be important to the angle between the core region and the tail at high CWV, a feature seen in observations. The simple case does not distinguish precipitating and nonprecipitating points, but it can be useful as a foil to ask what features depend on the Markov jump process in the more complex case of the PDF of precipitating points. Consider the transition from (9), valid in the high CWV regime, to (19), valid in the low CWV regime. In (9), the balance is between drift and diffusion terms, as in the simple solution above. The increase of r_{10} as one transitions to low CWV creates a core that involves a three-way balance between it and the drift and the diffusion terms, transitioning to the exponential solution (20) where the coefficients become constant.

PDFs similar to this simple case have recently been identified for several tracers, including CWV, in data from observations, models, and reanalysis (Neelin et al. 2010; Lintner et al. 2011). We note that PDFs of similar form can be produced purely from passive scalar advection–diffusion [see Bourlioux and Majda (2002) and related work]. In that case the core–tail relations depend on the characteristics of a flow field, not on explicit tracer sinks as in the solutions here. In either case, a representation of vertical motion/moisture convergence drives the excursions of moisture, and the question is whether it is limited by an actual sink (such as P) or by interplay with a cross-flow as in Bourlioux and Majda (2002). Both are potentially realistic factors in more complex models or nature, and a more unified understanding of these processes in a simple system that contains both processes will be of interest.

Simple stochastic models provide formulas for the exponential tails in terms of model parameters. This suggests that the observed exponential PDF from Peters and Neelin (2006) and Neelin et al. (2009) could be used to constrain the parameters of a stochastic convective parameterization. In fact, many of the parameter choices used here were motivated by this consideration. This topic will be discussed further in section 4.

4. Implications for convective parameterizations

We now discuss the stochastic model from three perspectives: as a simplified version of some existing stochastic parameterizations, as a prototype around which a more comprehensive parameterization could be built, and as a simple way to ''stochasticize'' an existing deterministic parameterization.

Several existing stochastic parameterizations include the two stochastic elements used here—stochastic closure and trigger—but they are in slightly different forms, and they are not explicitly identified using this terminology. The distinction between the closure and the trigger can be blurry in some circumstances, but it is still used here as a useful organizing principle. The two examples considered here include the models of Lin and Neelin (2000, 2002, 2003) and the models of Majda and Khouider (2002), Khouider et al. (2003), Katsoulakis et al. (2006), Majda et al. (2008), Majda and Stechmann (2008), and Khouider et al. (2010). One of the main differences between these two examples is in the type of stochastic process that is used: Gaussian processes in the former and Markov jump processes in the latter.

In the first example models, in the notation of Lin and Neelin (2000), the convective heating Q_c is given by

$$Q_c \propto \frac{1}{\tau_c}\,\mathcal{H}(C_1+\xi)(C_1+\xi), \tag{29}$$

where τ_c is the convective relaxation time scale, $\mathcal{H}$ is the Heaviside function, C_1 is a measure of the convective available potential energy (CAPE), and ξ is Gaussian red (not necessarily white) noise. This is a simplified, stochastic version of the parameterization of Betts and Miller (1986), which is recovered by setting $\xi = 0$. While the distinction between closure and trigger is somewhat blurry here, this parameterization does include both aspects.

In the second example models, in the notation of Majda and Khouider (2002) and Khouider et al. (2003), each column is broken into many subcolumns. Each subcolumn x is assigned a stochastic jump process $\sigma_I(x, t)$ that can take the value $\sigma_I = 0$ or $\sigma_I = 1$. In each column $[X, X + \Delta X)$, the subgrid scales can be represented by an average over l subcolumns:

$$\overline{\sigma}_I(X,t) = \frac{1}{l}\sum_{x\in[X,X+\Delta X)} \sigma_I(x,t). \tag{30}$$

This could allow, for instance, the possibility of having convection in a fraction of the column, since $\overline{\sigma}_I$ can take any of the $l + 1$ values from $0, 1/l, 2/l, \ldots 1$. Again, while the distinction between closure and trigger is somewhat blurry here, this parameterization does include both aspects.

The model of the present paper is a combination of the methods from these two examples. The Gaussian stochastic closures/forcings $D_i(q)\eta(t)$ are motivated by the first set of example models above, and $\sigma(t)$ is motivated by the second set of example models above.

As a second perspective, one could view the present paper's stochastic model as a prototype around which a more comprehensive parameterization could be built. In this respect, the key aspects of this model are that it is simple and that the model parameters can easily be constrained by observations (Neelin et al. 2008). Indeed, in this paper, the values of several model parameters were chosen based on the observations of Neelin et al. (2009) described in section 1 and based on the simple theory from section 3b. The values of $P(+\infty)$ and $D_P(+\infty)$ were chosen based on the mean and variance of precipitation from the observational record. The value of $D_1(+\infty)$ was chosen based on the theoretical exponential tail $p_1(q) = A\exp[-(2P/D_1^2)q]$ and the exponential tail from the observational data. The values of E and D_0 were partly based on the theoretical exponential tail $p_0(q) = A\exp[+(2E/D_0^2)q]$ and the exponential tail from the observational data. The values of r_{01} and r_{10} were partly based on the theoretical exponential tails $p_0(q) = A\exp[-m_0 q]$ and $p_1(q) = A\exp[+m_1 q]$ and the exponential tails from the observational data. And the tanh form of the transition rates was chosen with a sharp transition near $q \approx 60$–64 mm to be in agreement with the critical CWV from the observational record (Neelin et al. 2009).

Furthermore, there are even more observational statistics available that could be used to constrain extensions of this stochastic parameterization. For instance, this model did not include the effects of atmospheric temperature, which have been documented by Neelin et al. (2009). The model also did not include details of the vertical structure of the atmospheric thermodynamic state, and it did not include details of spatial variability, even though several interesting statistics have been documented by Holloway and Neelin (2009) and Peters et al. (2009) for the former and latter, respectively. [Methods for including vertical structure and spatial variability in simple stochastic models have also been introduced by the stochastic multicloud model of Khouider et al. (2010) and the coarse-grained stochastic models of Khouider et al. (2003).] Given this suite of observational constraints, the strength of the model presented here is that it takes a form that is easily related to statistics from observations: parameters can be constrained a priori via simple formulas, rather than a posteriori via numerous lengthy simulations over a large parameter space. However, it remains to be

seen whether this type of theory-based parameter tuning can be done reliably in a more complex situation with full vertical resolution and/or spatial variability. In addition, while many features were constrained here by the observational record, some features were not. For instance, no simple theory was given for the power laws in the autocorrelation functions in Fig. 3 or the event-size PDF in Fig. 4, although mechanisms have been identifiable in similar systems (Sornette 2004). If a theory for these approximate power laws can be formulated in future work, then this would provide further observational constraints for model parameters.

Finally, as a third perspective, one could view the present paper's stochastic model as a simple way to "stochasticize" an existing deterministic convective parameterization. This could be done by adding a stochastic component to either the convective closure or trigger (or both). The approach of "stochasticizing" a convective parameterization has been taken in earlier work, in which many different approaches have been presented (Lin and Neelin 2000; Bright and Mullen 2002; Lin and Neelin 2002; Majda and Khouider 2002; Khouider et al. 2003; Lin and Neelin 2003; Katsoulakis et al. 2006; Song et al. 2007; Majda et al. 2008; Majda and Stechmann 2008; Plant and Craig 2008; Khouider et al. 2010).

Here, as in Majda and Khouider (2002) and other work, we advocate $\sigma(t)$ as a simple way to implement a stochastic trigger with several important features. Any deterministic trigger condition of the form $\mathcal{H}(q - q_c)$ could be turned into a stochastic trigger by creating a $\sigma(t)$, as in this paper, to replace $\mathcal{H}$. This type of stochastic trigger was crucial for the strong peak in precipitation variance in Fig. 5, and it may play a role in reproducing the event-size PDF in Fig. 4. It also allows precipitation to occur occasionally at moderate or low CWV values, which reproduces the exponential tail $p_1(q) = A \exp[+m_1 q]$ from Fig. 6, in agreement with the observation record (Neelin et al. 2009); and it allows the analytic representation of $p_1(q) = A \exp[+m_1 q]$ as described in section 3b.

In addition, the results here also demonstrated the importance of intense stochastic "external" forcing $D_F\eta(t)$. This was crucial for obtaining extreme precipitation events with high CWV and long lifetimes, as it could occasionally compensate for the heavy precipitation and encourage more of it. These extreme events could be seen in individual instances in Figs. 2c and 2d and in statistical measures such as the event-size PDF in Fig. 4 and the exponential tail at high CWV from (17) and Fig. 6. In an atmospheric simulation with spatial variability, one would expect much of the stochastic external forcing to be represented by resolved dynamics. However, it is also possible that the resolved variance would not account for the full intensity of $D_F\eta(t)$. A similar parameterization could be used to account for unresolved variance, as in Lin and Neelin (2000, 2002, 2003).

5. Concluding discussion

A simple stochastic model was designed and studied in order to further understand the transition to strong convection. The transition has recently been characterized by an array of statistical measures (Peters and Neelin 2006; Neelin et al. 2008, 2009; Peters and Neelin 2009; Holloway and Neelin 2010), which were summarized in section 1. It was shown that the simple model from section 2 captures the following four local statistical features from the observations:

1) autocorrelation functions of CWV and precipitation with long and short autocorrelation times, respectively, and with approximate exponential and power-law decays, respectively [compare Fig. 3 here with Fig. 1 of Holloway and Neelin (2010)];
2) approximate power laws in PDF of precipitation event size [compare Fig. 4 here with Fig. 6b of Neelin et al. (2008), with Fig. 1 of Peters and Neelin (2009), and with Figs. 2 and 3a of Peters et al. (2010)];
3) a sharp transition in the mean of precipitation and a peak in variance of precipitation, near a critical CWV value [compare Fig. 5 here with Figs. 1 and 4 of Neelin et al. (2009)]; and
4) exponential tails in PDFs of CWV values conditioned on precipitating points $p_1(q)$ and nonprecipitating points $p_0(q)$, and maxima in PDFs located at CWV values just below the critical CWV value [compare Fig. 6 here with Figs. 5 and 9 of Neelin et al. (2009)].

Furthermore, the slopes of the exponential tail of the PDFs were predicted with simple analytic solutions (section 3b). This directly ties model behavior to model parameter values. With knowledge of the PDFs from observational data (Neelin et al. 2009), it provides a way to constrain model parameters via simple formulas, rather than via model simulations over parameter space. At the same time, however, the PDF slopes are built into the model in some sense, whereas one might like to have the slopes arise from some fundamental physical postulates.

Given the agreement of the model results with the observed statistics, further insight can be gained into the nature of the observed statistics. For instance, the peak in precipitation variance was tied to stochastic transitions in $\sigma(t)$ between the precipitating and nonprecipitating states, rather than to fluctuations in the precipitation

rate from $D_P\eta(t)$. As another example, evidence of stochastic hysteresis was seen in the low-CWV exponential tail in the CWV PDF of precipitating points. These precipitating points likely began their life cycle at higher CWV values and continued to exist even after CWV was reduced significantly. This behavior is made possible by the stochastic switch $\sigma(t)$. On the other hand, the stochastic switch is not essential to the power law in precipitation autocorrelation, which can also be reproduced with a deterministic switch. As yet another example, the stochastic "external" forcing $D_F\eta(t)$ was seen to be crucial for obtaining extreme precipitation events. This $D_F\eta(t)$ is meant to represent the various moistening/drying processes associated with deep moist convection, such as moisture convergence/divergence from mesoscale and synoptic-scale waves, updraft/downdraft circulations, etc. These mechanisms occasionally were strong enough to compensate for the drying from precipitation and to allow extreme precipitation events with high CWV values and long lifetimes. One question that remains is whether, in large-scale models with spatial variability, the resolved dynamics would account for the full intensity of $D_F\eta(t)$. In nature these extreme events would typically be associated with upward motion and moisture convergence in systems such as tropical waves, mesoscale convective systems, hurricanes, etc.

The stochastic model was also discussed in the context of stochastic convective parameterizations in section 4. The model presented here has three stochastic components, each of which is treated separately: a stochastic trigger and stochastic closures for precipitation and for forcing. The stochastic trigger is a two-state Markov jump process $\sigma(t)$, which takes the value $\sigma = 0$ when there is no convection and $\sigma = 1$ when there is convection. The transition rates for $\sigma(t)$ depend on the current atmospheric thermodynamic state through the CWV: $r_{01}[q(t)]$ and $r_{10}[q(t)]$. On the other hand, the stochastic closures for precipitation and external forcing take the form $D(q)\eta(t)$, where $\eta(t)$ is Gaussian white noise and $D(q)$ is a CWV-dependent variance function. These stochastic components—trigger and closures—are partly motivated by the earlier work of Majda and Khouider (2002), Khouider et al. (2003, 2010), and Lin and Neelin (2000, 2002, 2003), respectively. Each of these stochastic components offers a simple way to stochasticize an existing deterministic convective parameterization. The simplicity of these stochastic components offers some important advantages: it allows the separate roles of the stochastic jumps and the Gaussian process to be distinguished, and it allows model parameters to be tied directly to observed statistics.

The stochastic model presented here could also be used as a prototype around which a more comprehensive parameterization could be built. For instance, this model did not include the effects of atmospheric temperature, whose effects on the observed statistics have been documented by Neelin et al. (2009). The model also did not include details of the vertical structure of the atmospheric thermodynamic state, and it did not include details of spatial variability, even though several interesting statistics have been documented by Holloway and Neelin (2009) and Peters et al. (2009) for the former and latter, respectively. Methods for including vertical structure and spatial variability in simple stochastic models have also been introduced by the stochastic multicloud model of Khouider et al. (2010) and the coarse-grained stochastic models of Khouider et al. (2003) and Katsoulakis et al. (2006). A combination of these ideas should be a fruitful direction of future research.

Acknowledgments. The authors thank A. Majda for helpful discussion and A. Sobel, O. Peters, and an anonymous reviewer for comments that improved the presentation of the paper. The research of S.N.S. was partially supported by a NOAA Climate and Global Change Postdoctoral Fellowship, a NSF Mathematical Sciences Postdoctoral Research Fellowship, and a start-up grant from the University of Wisconsin—Madison. J.D.N. was supported by NOAA Grants NA08OAR4310882 and NA11OAR4310099 and NSF Grants ATM-0645200 and AGS-1102838.

REFERENCES

Betts, A. K., and M. J. Miller, 1986: A new convective adjustment scheme. Part II: Single column tests using GATE wave, BOMEX, ATEX and arctic air-mass data sets. *Quart. J. Roy. Meteor. Soc.,* **112,** 693–709.

Bourlioux, A., and A. J. Majda, 2002: Elementary models with probability distribution function intermittency for passive scalars with a mean gradient. *Phys. Fluids,* **14,** 881–897.

Bretherton, C. S., M. E. Peters, and L. E. Back, 2004: Relationships between water vapor path and precipitation over the tropical oceans. *J. Climate,* **15,** 1517–1528.

Bright, D., and S. Mullen, 2002: Short-range ensemble forecasts of precipitation during the southwest monsoon. *Wea. Forecasting,* **17,** 1080–1100.

Gardiner, C. W., 2004: *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. 3rd ed. Springer Series in Synergetics, Vol. 13. Springer–Verlag, 415 pp.

Higham, D., 2001: An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev.,* **43,** 525–546.

Hilburn, K. A., and F. J. Wentz, 2008: Intercalibrated passive microwave rain products from the unified microwave ocean retrieval algorithm (UMORA). *J. Appl. Meteor. Climatol.,* **47,** 778–794.

Holloway, C. E., and J. D. Neelin, 2009: Moisture vertical structure, column water vapor, and tropical deep convection. *J. Atmos. Sci.,* **66,** 1665–1683.

——, and ——, 2010: Temporal relations of column water vapor and tropical precipitation. *J. Atmos. Sci.,* **67,** 1091–1105.

Houze, R. A., Jr., and C.-P. Cheng, 1977: Radar characteristics of tropical convection observed during GATE: Mean properties and trends over the summer season. *Mon. Wea. Rev.,* **105,** 964–980.

Katsoulakis, M., A. Majda, and A. Sopasakis, 2006: Intermittency, metastability and coarse graining for coupled deterministic–stochastic lattice systems. *Nonlinearity,* **19,** 1021–1047.

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

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

Lawler, G., 1995: *Introduction to Stochastic Processes*. Chapman & Hall/CRC, 176 pp.

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

——, and ——, 2002: Considerations for stochastic convective parameterization. *J. Atmos. Sci.,* **59,** 959–975.

——, and ——, 2003: Toward stochastic deep convective parameterization in general circulation models. *Geophys. Res. Lett.,* **30,** 1162, doi:10.1029/2002GL016203.

Lintner, B. R., C. E. Holloway, and J. D. Neelin, 2011: Column water vapor statistics and their relationship to deep convection and vertical and horizontal circulation and moisture structure at Nauru. *J. Climate,* **24,** 5454–5466.

Lovejoy, S., 1982: Area–perimeter relation for rain and cloud areas. *Science,* **216,** 185–187.

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

——, and S. N. Stechmann, 2008: Stochastic models for convective momentum transport. *Proc. Natl. Acad. Sci. USA,* **105,** 17 614–17 619.

——, C. Franzke, and B. Khouider, 2008: An applied mathematics perspective on stochastic modeling for climate. *Philos. Trans. Roy. Soc. London,* **366A,** 2427–2455.

Mapes, B. E., and R. A. Houze Jr., 1993: Cloud clusters and superclusters over the oceanic warm pool. *Mon. Wea. Rev.,* **121,** 1398–1416.

——, S. Tulich, J.-L. Lin, and P. Zuidema, 2006: The mesoscale convection life cycle: Building block or prototype for large-scale tropical waves? *Dyn. Atmos. Oceans,* **42,** 3–29.

——, R. Milliff, and J. Morzel, 2009: Composite life cycle of maritime tropical mesoscale convective systems in scatterometer and microwave satellite observations. *J. Atmos. Sci.,* **66,** 199–208.

Matsumoto, M., and T. Nishimura, 1998: Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul.,* **8,** 3–30.

Muller, C., L. Back, P. O'Gorman, and K. Emanuel, 2009: A model for the relationship between tropical precipitation and column water vapor. *Geophys. Res. Lett.,* **36,** L16804, doi:10.1029/2009GL039667.

Neelin, J. D., O. Peters, J.W.-B. Lin, K. Hales, and C. E. Holloway, 2008: Rethinking convective quasi-equilibrium: Observational constraints for stochastic convective schemes in climate models. *Philos. Trans. Roy. Soc. London,* **366A,** 2581–2604.

——, ——, and K. Hales, 2009: The transition to strong convection. *J. Atmos. Sci.,* **66,** 2367–2384.

——, B. R. Lintner, B. Tian, Q. Li, L. Zhang, P. K. Patra, M. T. Chahine, and S. N. Stechmann, 2010: Long tails in deep columns of natural and anthropogenic tropospheric tracers. *Geophys. Res. Lett.,* **37,** L05804, doi:10.1029/2009GL041726.

Neggers, R. A. J., H. J. J. Jonker, and A. P. Siebesma, 2003: Size statistics of cumulus cloud populations in large-eddy simulations. *J. Atmos. Sci.,* **60,** 1060–1074.

Nesbitt, S. W., R. Cifelli, and S. A. Rutledge, 2006: Storm morphology and rainfall characteristics of TRMM precipitation features. *Mon. Wea. Rev.,* **134,** 2702–2721.

Peters, O., and J. D. Neelin, 2006: Critical phenomena in atmospheric precipitation. *Nat. Phys.,* **2,** 393–396.

——, and ——, 2009: Atmospheric convection as a continuous phase transition: Further evidence. *Int. J. Mod. Phys.,* **23B,** 5453–5465.

——, C. Hertlein, and K. Christensen, 2002: A complexity view of rainfall. *Phys. Rev. Lett.,* **88,** 018701, doi:10.1103/PhysRevLett.88.018701.

——, J. D. Neelin, and S. W. Nesbitt, 2009: Mesoscale convective systems and critical clusters. *J. Atmos. Sci.,* **66,** 2913–2924.

——, A. Deluca, A. Corral, J. D. Neelin, and C. E. Holloway, 2010: Universality of rain event size distributions. *J. Stat. Mech.,* P11030, doi:10.1088/1742-5468/2010/11/P11030.

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

Ross, S., 1998: *A First Course in Probability*. 5th ed. Prentice Hall, 514 pp.

Sherwood, S. C., 1999: Convective precursors and predictability in the tropical western Pacific. *Mon. Wea. Rev.,* **127,** 2977–2991.

——, and R. Wahrlich, 1999: Observed evolution of tropical deep convective events and their environment. *J. Atmos. Sci.,* **127,** 1777–1795.

Sobel, A. H., S. E. Yuter, C. S. Bretherton, and G. N. Kiladis, 2004: Large-scale meteorology and deep convection during TRMM KWAJEX. *Mon. Wea. Rev.,* **132,** 422–444.

Song, Y., C. Wikle, C. Anderson, and S. Lack, 2007: Bayesian estimation of stochastic parameterizations in a numerical weather forecasting model. *Mon. Wea. Rev.,* **135,** 4045–4059.

Sornette, D., 2004: *Critical Phenomena in Natural Sciences*. Springer Verlag, 528 pp.