The universal scaling characteristics of tropical oceanic rain clusters

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2016JD025921

Key Points:
- The size and rainfall distributions of observed tropical rain clusters appear to be universal across the oceans
- Scaling exponents of cluster rain rate and cluster area are related in a way that is consistent with self-organized critical phenomena
- Cluster areal rain efficiency at the mesoscale increases with cluster area but is practically constant at larger scales

Correspondence to:
C.-K. Teo,
ckteo@suss.edu.sg

Citation:
Teo, C.-K., H.-N. Huynh, T.-Y. Koh, K. K. W. Cheung, B. Legras, L. Y. Chew, and L. Norford (2017), The universal scaling characteristics of tropical oceanic rain clusters, *J. Geophys. Res. Atmos.*, *122*, 5582–5599, doi:10.1002/2016JD025921.

Received 12 SEP 2016
Accepted 9 MAY 2017
Accepted article online 11 MAY 2017
Published online 5 JUN 2017

The universal scaling characteristics of tropical oceanic rain clusters

C.-K. Teo[1], H.-N. Huynh[2,3,4], T.-Y. Koh[5], K. K. W. Cheung[6], B. Legras[7], L. Y. Chew[2,8], and L. Norford[9,10]

[1]Centre for Applied Research, SIM University, Singapore, [2]Complexity Institute, Nanyang Technological University, Singapore, [3]Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore, [4]Department of Mathematics, Imperial College London, London, UK, [5]UniSIM College, SIM University, Singapore, [6]Department of Environmental Sciences, Macquarie University, Sydney, New South Wales, Australia, [7]Laboratoire de Météorologie Dynamique, CNRS and Ecole Normale Supérieure, Paris, France, [8]School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, [9]CENSAM, Singapore-MIT Alliance for Research and Technology, Singapore, [10]Department of Architecture, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

Abstract Using multiyear satellite rainfall estimates, the distributions of the area, and the total rain rate of rain clusters over the equatorial Indian, Pacific, and Atlantic Oceans was found to exhibit a power law $f_S(s) \sim s^{-\zeta_S}$, in which S represents either the cluster area or the cluster total rain rate and $f_S(s)$ denotes the probability density function of finding an event of size s. The scaling exponents ζ_S were estimated to be 1.66 ± 0.06 and 1.48 ± 0.13 for the cluster area and cluster total rain rate, respectively. The two exponents were further found to be related via the expected total rain rate given a cluster area. These results suggest that convection over the tropical oceans is organized into rain clusters with universal scaling properties. They are also related through a simple scaling relation consistent with classical self-organized critical phenomena. The results from this study suggest that mesoscale rain clusters tend to grow by increasing in size and intensity, while larger clusters tend to grow by self-organizing without intensification.

1. Introduction

Despite the advancement in understanding tropical deep convection (see, for example, *Houze* [2004] for a review), a complete understanding of how tropical deep convection self-organizes into clusters across a wide range of spatial scales remains elusive. Evidences of self-similarity in observed tropical convection abound. For example, *Mapes et al.* [2006] showed that convectively coupled tropical synoptic-scale waves exhibit similar life cycles to mesoscale convective systems. Elsewhere, the power law nature in the probability distributions of observed cloud and rain cluster geometrical measures appears to be ubiquitous [e.g., *Lovejoy*, 1982; *Cahalan and Joseph*, 1989; *Peters et al.*, 2009, 2012; *Wood and Field*, 2011; *Traxl et al.*, 2016], suggesting scale invariance in the geometries of cloud and rain clusters.

Some important understanding of the mechanisms behind the self-organization of tropical convection has been gained through idealized numerical radiative-convective equilibrium (RCE) simulations [e.g., *Bretherton et al.*, 2005; *Muller and Held*, 2012; *Wing and Emanuel*, 2014; *Muller and Bony*, 2015]. Recently, *Khairoutdinov and Emanuel* [2010] demonstrated through simulated RCE that disaggregated tropical convection tends to self-aggregate into a single large cluster above a certain sea surface temperature (SST). They argued that such self-aggregation can be regarded as a form of phase transition; the tropical convection in the real atmosphere could be a self-organized critical phenomenon [*Bak et al.*, 1987] through regulating the SST around a critical value via radiative effects in the presence or absence of cloud cover. The conjecture of tropical convection as a self-organized critical phenomenon has been put forth earlier from observational studies of oceanic rainfall by *Peters and Neelin* [2006]. They suggest that convective precipitation exhibits self-organized criticality (SOC) with the column water vapor self-regulating by precipitation.

Motivated by the above background work, we decided to test the hypothesis that the rain clusters over the tropical oceans exhibit SOC by investigating the scaling behavior of the rain cluster. Our work primarily seeks evidence for the universality in the scaling of deep convection which underlies the concept of self-organized criticality (SOC). It is for this reason that we focused on the tropical oceanic basins, where the surface boundary conditions are relatively homogenous, in the hope to minimize any organizing effects heterogeneous surfaces have on deep convection. Land surfaces have uneven terrain, soil moisture, roughness, and

urbanization elements which are all known to organize rainfall. Coastlines also organize mesoscale convection in land-sea breeze systems, and at large scales, they have a multifractal nature which may complicate the scaling behavior further. We believe that such boundary-induced effects would potentially mask the universal character of the otherwise self-organizing rain clusters. In contrast, the oceanic basins are large enough for us to observe rain clusters across a wide range of spatial scales, and consequently, we should be able to better characterize the scaling regime or regimes.

We focused our study primarily on regions where deep convection is frequent, namely, the Indo-Pacific Warm Pool and the Intertropical Convergence Zone (ITCZ) over the east Pacific Ocean and the Atlantic Ocean. These regions are where deep convection frequently self-organizes into mesoscale convective systems. In the context of critical phenomena, the common tendency for tropical convective systems to aggregate and self-organize [*Mapes*, 1993] over a domain with homogeneous environmental condition, such as uniform sea-surface temperature, leads to the universally emergent statistical character of tropical convection across different oceanic basins. In particular, we hypothesize that the scaling exponents and their relations for convective rain clusters are the same, regardless of the dynamical details of individual rain clusters and their geographic location. Although it has been reported that the prominent stratocumulus decks can organize themselves into mesoscale cells [*Bretherton et al.*, 2010], the underlying mechanisms in the organization of these shallow convective rain clusters are different from that of rain clusters that involve deep convection in their life cycle. So as a first investigation, we confined our analysis to tropical oceanic regions with frequent deep convection and excluded those regions where the convection is predominantly shallow, for example, south of the equator over the east Pacific.

Specifically, we examined the probability distributions of the cluster observables using the simple scaling [*Christensen and Moloney*, 2005 p. 273]:

$$f_S(s) \propto s^{-\zeta_S} G_S\left(\frac{s}{\xi_S}\right); \quad s > s_0 \tag{1}$$

where f_S is the probability density function of a given observable of the cluster S, ζ_S is the scaling exponent, and G_S is the scaling function accounting for the departure from a strict power law nature due to the finite system size. G_S decreases monotonically for $s/\xi_S > 1$ and is a constant when $s/\xi_S \ll 1$. Between the upper cutoff ξ_S and lower cutoff s_0 the distribution is a power law $f_S(s) \sim s^{-\zeta_S}$. For such distributions, ξ_S is the only characteristic scale that in the simple scaling hypothesis depends on the size of the domain in which the system is confined [e.g., *Pruessner*, 2012, p. 27]. For tropical convection, the equivalent domain size would ideally correspond to an outer length scale in which the tropical convective clusters are embedded. That this length scale would likely be controlled by the larger-scale atmospheric circulation and the sea surface temperature means that ξ_S for the different oceanic basins might not be the same. But when we analyze the cluster statistics sampled from a domain with size smaller than this natural outer length scale, the power law scaling will be limited not by the outer length scale but the finite size of the sampled domain.

We adopt the following notation for the remaining of this paper: A and R denote the random variables representing the rain cluster area and cluster total rain rate, respectively (see section 3 for the precise definitions of the two quantities). Symbols a and r represent sample values of A and R, respectively. The conditional mean of R for a given cluster area a is denoted as $\langle R|\, a\rangle$. For brevity the symbol S (s) shall be used to represent A (a) or R (r) whenever the discussion happens to apply for both the variables.

2. A SOC Theory for Rainfall

The two cluster variables A and R that are readily observed from satellite data were chosen for analysis. They were selected because (1) the distribution of A contains information about the spatial variability of radiative fluxes and (2) the distribution of R is related to the spatial variability of the vertical fluxes of moist static energy from the boundary layer. Cloud-radiative feedback and boundary layer moist static energy fluxes are found to be important in determining the self-aggregation of convection [*Muller and Held*, 2012]. In our SOC paradigm for rainfall, R corresponds to a measure of the "avalanche intensity" of the tropical atmosphere-ocean system, releasing the potential instability of the atmosphere built-up as a result of the radiative and sensible heat fluxes on a timescale significantly longer than the convective "charging" timescale. A corresponds to a measure of the "avalanche area."

Specifically, we sought the evidence for universality in the rain clusters by showing that ζ_S is the same across the different oceanic basins, S being either A or R. A further relation between the scaling exponents for distributions of A and R under the SOC paradigm can also be established through the scaling properties of $\langle R|a\rangle$. For the tropical rain clusters, the following model was found to be suitable (section 5):

$$\langle R|a\rangle = \begin{cases} ka^{\beta} & a \le \alpha \\ k\alpha^{\beta-\chi}a^{\chi} & a > \alpha \end{cases} \tag{2}$$

where k, β, and χ are positive numbers and $\beta \neq \chi$. The α in equation (2) is known as the crossover cluster size since the exponent in equation (2) takes the value of either β or χ depending on whether a is larger or smaller than α. By assuming that (1) ξ_R scales with the conditional average rain rate at a given crossover cluster size, i.e., $\langle R|\alpha\rangle$, and (2) α scales with ξ_A, we were able to establish the following relation:

$$\beta = \frac{\zeta_A - 1}{\zeta_R - 1} \tag{3}$$

Details of the proof can be found in Appendix A1. Equation (3) is a well-known relation for SOC models for the special case where there is no crossover scale in the scaling exponent in $\langle R|a\rangle$ [see, for example, *Pruessner*, 2012, pp. 41–43].

3. Data and Methods

We used the 3-hourly rainfall intensity observations from the Tropical Rainfall Measuring Mission (TRMM) 3B42 [*Huffman et al.*, 2007] from 1998 to 2012 which have a spatial resolution of 0.25° × 0.25°. We focused our attention on four equatorial regions IO (60°E–90°E, 20°S–10°N), WP (150°E–180°, 15°S–15°N), EP (110°W–140°W, 15°S–15°N), and AO (15°W–45°W, 10°S–20°N) covering the major ocean basins as shown in Figure 1. These regions were equal in size, each with a length (Z) of 120 TRMM grid points in both the longitudinal and latitudinal directions covering an area of approximately 9×10^6 km^2. We tried as much as possible to locate the study regions away from land or coastal seas where the topography and coastlines can organize localized mesoscale weather systems [*Teo et al.*, 2011] that could impose a preferential scale that is more related to the geometry of the coastlines or topography.

For each available time slice of the data, we first identified the oceanic rain clusters within each focus region. In the interest of clarity, the term "focus region" is used interchangeably with "domain of length $Z = 120$" or "120 × 120 domain" in the following discussion. A cluster is defined as a set of rainy grid boxes where every grid box from the set shares at least one common edge with another rainy grid box within the set. In our work a grid box is considered rainy if the observed rain intensity is greater than zero. The area of the rain cluster A is simply the area covered by all grid boxes in the set, while the corresponding total cluster rain rate R is the sum of the individual rain rates of the cluster grid boxes. (Note that the quantity R was named "event power" in *Peters et al.* [2012].) A is measured in units equal to the constant area ΔA of a grid box (ignoring the slight distortion due to the curvature of the Earth near the equator), and R is measured in units of $\Delta A \times 0.01$ mm/h (as TRMM 3B42 data are discretized in intervals of 0.01 mm/h). The few rain rates in the data set that were not multiples of 0.01 mm/h were rounded up to the nearest multiple of 0.01 mm/h. In the following discussion, A and R are dimensionless quantities, normalized by ΔA and $\Delta A \times 0.01$ mm/h, respectively.

Due to the temporal correlation between rain clusters that was estimated to be significant up to a time interval of 5 days for all the focus regions, the cluster data for each focus region were further partitioned into 40 nonoverlapping data subsets. The details of how the temporal correlations between the rain clusters were calculated can be found in Appendix A2. Each of these data subsets contained cluster data sampled 5 days apart at the same coordinated universal time (UTC). To mitigate complicating influences of land and predominantly shallow convective regions, we only admitted those focus regions with less than 10% land coverage and if their mean cluster rain rates calculated from each of the 40 data subsets were all larger than a given threshold. This threshold is empirically defined as $M(Z = 120) - S(Z = 120)$, where $M(Z = 120)$ is the ensemble mean of the average cluster rain rates for the focus region and $S(Z = 120)$ is the corresponding ensemble standard deviation. Details of the selection criteria which the domains were subjected to can be found in Appendix A3.

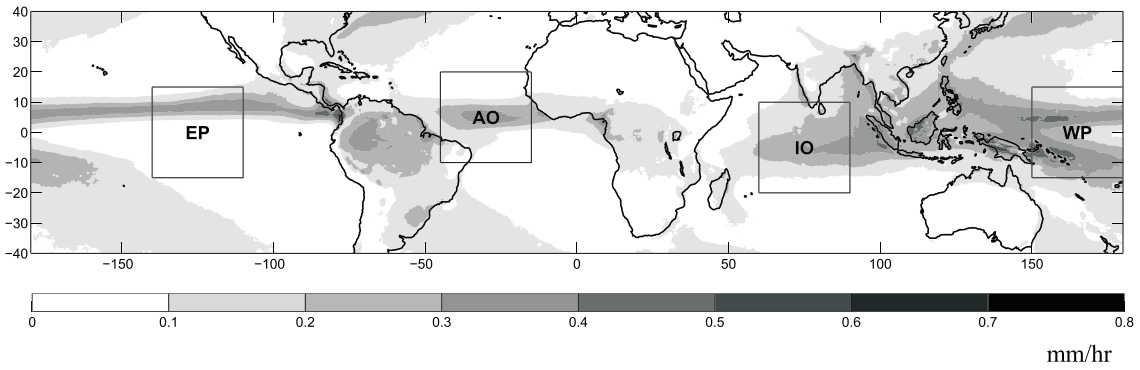

Figure 1. Focus regions where the rainfall cluster distributions were analyzed. The 1998–2012 average rainfall rates obtained from TRMM 3B42 are shown as shaded contours.

Due to the dependence of the scaling exponents on the domain sizes in which the clusters were sampled, we further divided each focus region (including EP) into nonoverlapping squares of length $Z = 24$, 30, 40, and 60 to investigate this dependence. The process of testing for deep convective oceanic domains as well as identifying the rain clusters described for the $Z = 120$ domains was repeated for all these smaller domains. Note that although EP was excluded for analysis for $Z = 120$, it was included for analysis for the smaller Zs. This was because the (smaller) domains within the focus region where the ITCZ is located could be admitted based on our selection criteria. Thereafter, the ensemble probability distribution for a given Z (for $Z < 120$) was computed among all the deep convective oceanic domains of the same size for each focus region. The ensemble probability for a $Z \times Z$ domain for a given focus region was computed by averaging the probability distribution of the $Z \times Z$ domains for that focus region that passed the selection criteria. For example, the ensemble probability distribution for $Z = 60$ domain for EP was found by averaging the sample probability distributions over the two 60×60 domains that passed the selection criteria. Obviously, the ensemble probability distribution for each of the 120×120 domain (the focus region) is just the sample probability distribution since there is only one such domain per focus region.

Figure 2 shows the domains for different Z, admitted by the selection process. To gauge how good our selection criteria were in excluding predominantly shallow convective domains, we estimated the average fractional areal coverage of clouds with cloud tops between 7 km and 15 km in height, as a proxy to the frequency of deep convection. Thirteen years (2000–2012) of the annually averaged areal cloud cover retrieved from measurements made by the Multiangle Imaging Spectroradiometer (MISR) [*Di Girolamo et al.*, 2010] on board the Terra satellite were used in computing the required high cloud fraction shown in Figure 2. The selection criteria performed reasonably well for all Z, admitting most of the domains over the IO and WP where deep convection is frequent, while excluding a few of the southernmost domains for $Z = 24$ in IO. Likewise, the ITCZs in the EP and AO were admitted for analysis as expected, and the predominantly shallow convective areas outside the ITCZs were excluded. The exclusion of the 120×120 domain for the EP is reasonable, since most of the area south of the equator was dominated by rain clusters associated with low-level clouds.

The spatial patterns of the domains selected remained relatively robust to the variation of the selection criteria: if we relaxed the selection criteria to include a domain for analysis so long as 50% instead of 100% of the mean cluster rain rates calculated from each of the 40 data subsets were larger than the same threshold, some of the rejected domains, like the south-western IO domain, and the south western AO domain at $Z = 60$, as well as those at the north-eastern edges of the WP at $Z = 24$, would be included (cf. Figure 2). However, the selected and rejected EP domains at all Z remained unchanged. All results and conclusions presented in the later sections remained virtually unchanged under the relaxed criteria, demonstrating the robustness of our analysis to the selection criteria. Henceforth, we present only the results for the more stringent selection criteria.

Figure 3 shows an example of the resulting ensemble distributions of A and R for the IO for the different domain sizes Z. The dependence of the scaling exponents, which correspond to the gradients of the linear region of the distributions on the double logarithmic plots, can be visually discerned for both A and R.

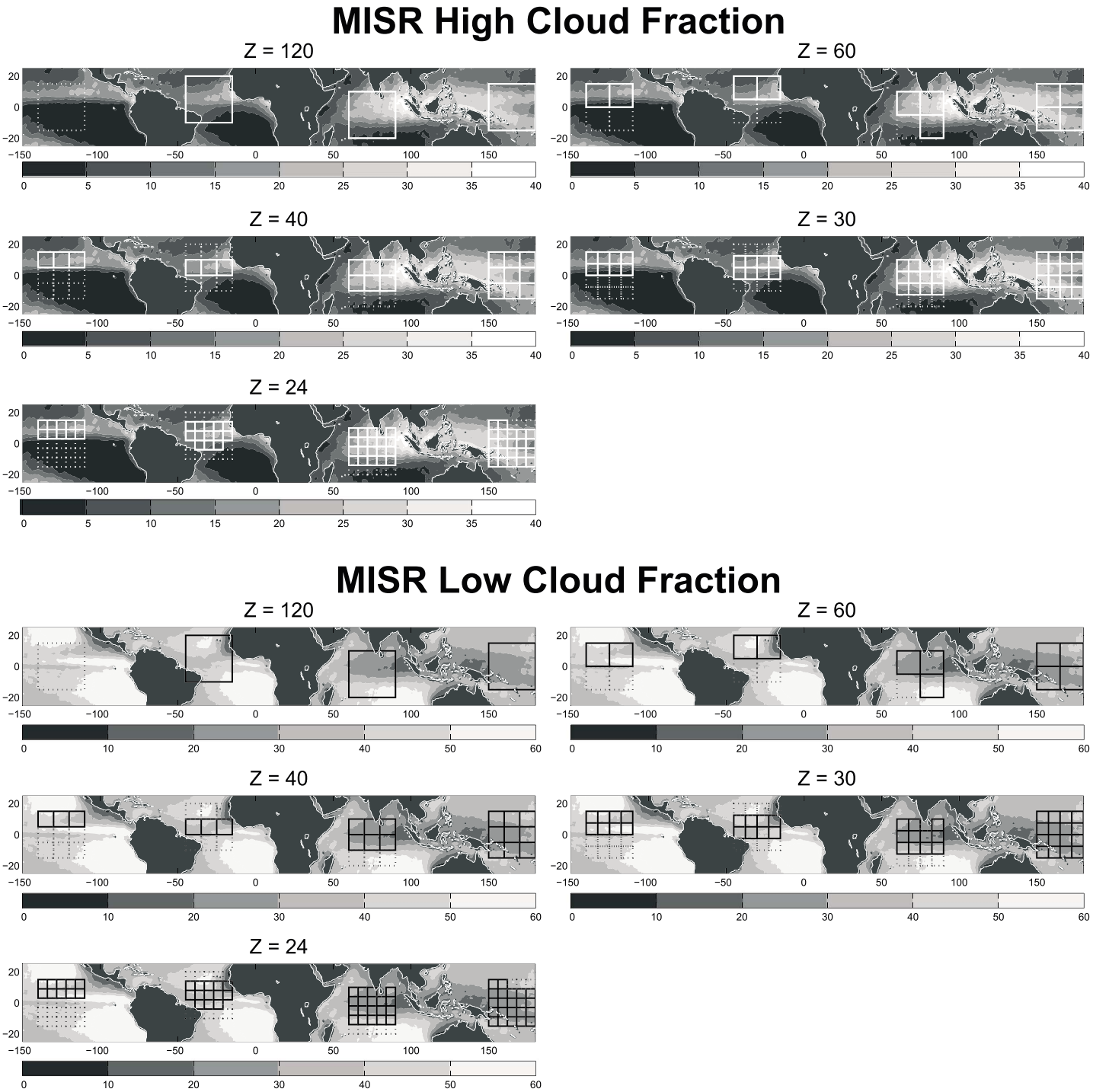

Figure 2. Selected domains for different Z together with the climatological (top) high and (bottom) low cloud amount statistics from satellite observations. The boxes with solid (dotted) edges are domains that passed (failed) the selection criteria. The contours are the mean of the annually averaged high cloud fractions, reported as percentages, as observed by the MISR from 2000–2012. The annually averaged high (low) cloud fraction is defined as the annually averaged spatial cloud cover in a 0.5° × 0.5° grid box, expressed as percentages, for clouds with cloud top heights between 7 km and 15 km (500 m and 3 km). Note that the contour intervals for the two panels are different.

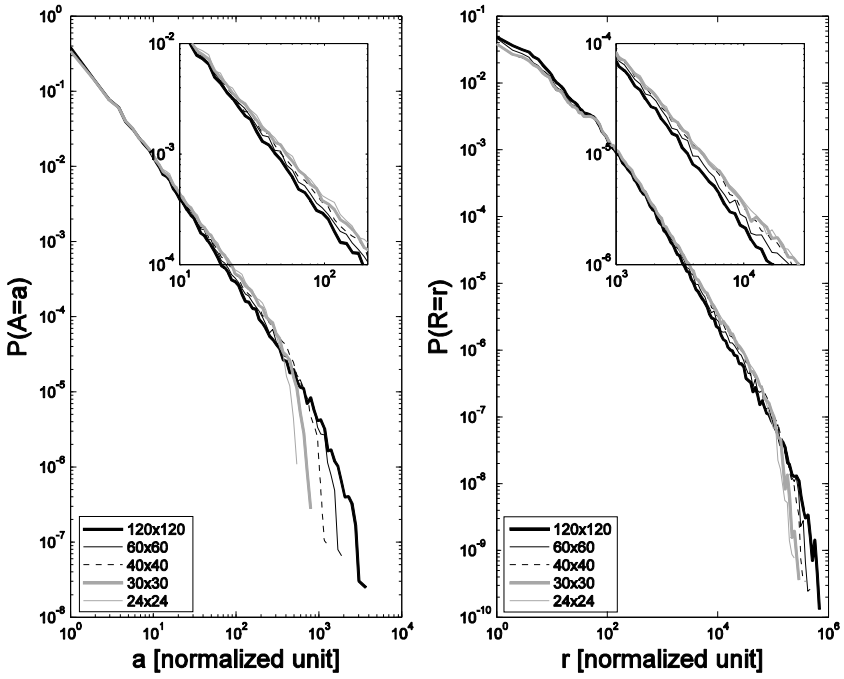

Figure 3. Double logarithmic plots of the ensemble probability distribution of (left) rain cluster area A and (right) total cluster rain rate R over the focus region IO of Figure 1. A and R are normalized by ΔA and $\Delta A \times 0.01$ mm/h, respectively. Each curve represents the probability for the ensemble of all nonoverlapping samples of one domain size $Z \times Z$ (cf. the legend). The inserts are magnified views that emphasize the dependence of the scaling exponent on the domain size.

4. Universality of Cluster Distributions

For each of the individual data subsets of a $Z \times Z$ domain, we modeled the distribution of A and R from the observations as

$$P(S = s) = C_s s^{-\tau_s} \exp[-(s/S_c)^{\nu_s}]; \quad s \in \{S_0, S_0 + 1, S_0 + 2, \cdots, S_m\} \tag{4}$$

where $P(S = s)$ is the probability of the discrete random variable S taking a positive value s, C_s is the appropriate normalizing constant, and τ_S, ν_S, and S_c are positive numbers to be determined empirically. Discrete random variables were used since both A and R obtained from the TRMM data were discrete as noted previously. In equation (4), $S_0 \geq 1$; A_m is the largest possible cluster area which is the domain area Z^2; R_m is the largest possible cluster rain rate which is the saturation value for TRMM's rainfall measurement multiplied by the A_m. The exponential term on the right-hand side (RHS) of equation (4) is one of the simplest form that accounts for the rapid departure of the probability from a power law due to the global constraint imposed by the finite nature of S_m as seen in Figure 3. Note that S_0 has to be determined empirically as it is not known a priori; below this scale the cluster distribution is no longer associated with the actual organized deep convective clusters but is sensitive to the myriad details of mesoscale convective dynamics. For instance, observed rain clusters with very small r of values of 1 or 2, corresponding to 0.01 or 0.02 mm/h × (0.25° × 0.25°), may actually be associated with the nonuniform filling of the cluster area by smaller regions of higher rain rates, or with shallow-convective or stratiform rain.

For each focus region we estimated τ_A and τ_R for the different domain sizes Z by fitting the model parameters in equation (4) to the cluster data using the maximum likelihood method [e.g., *Papoulis and Pillai*, 2001]. The details of the maximum likelihood method used in estimating the model parameters are described in Appendix A4. For each Z, estimates of the scaling exponents were obtained for a range of different S_0 values to account for the uncertainty in the S_0. This was done as follows: For each S_0, 40 separate estimates for the scaling exponent were derived from the 40 nonoverlapping cluster data time series (section 2), and from these estimates, a best estimate $\tau_S(Z, S_0)$ was obtained by the method described in Appendix A4. τ_S is plotted in Figures 4a–4d with each data point corresponding to the estimate of the scaling exponent (for A or for R) for one (Z, S_0) pair. The dependence of τ_S on Z is a consequence of both the finite size of the domain as well as

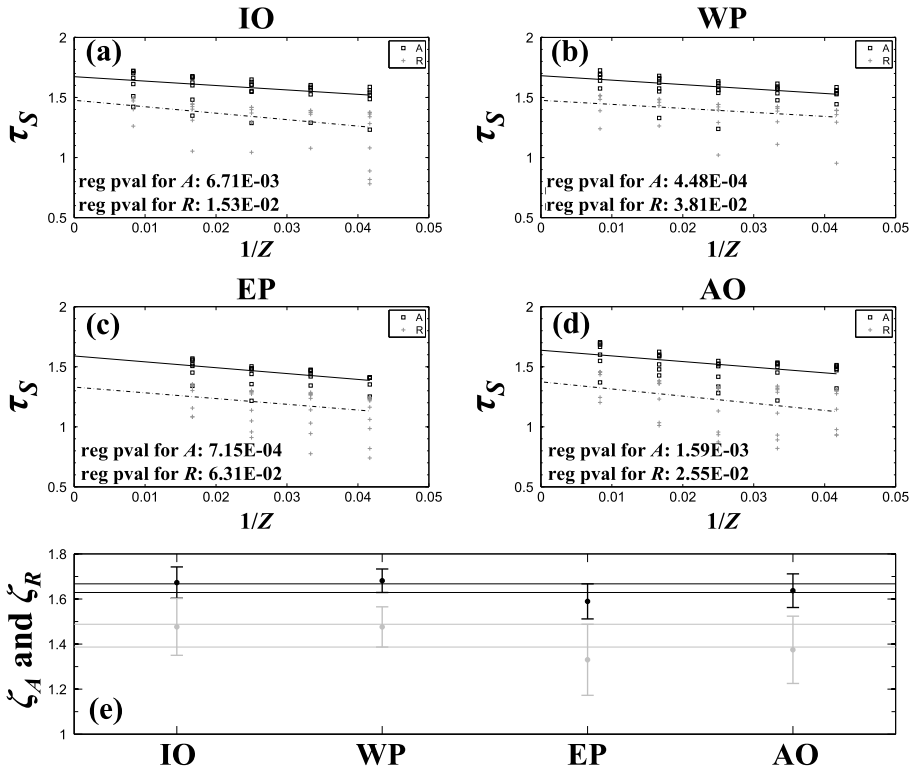

Figure 4. (a–d) Robust iterative least squares linear regression with Huber weights of scaling exponents versus $1/Z$ for the different focus regions for rain cluster size A (black open squares) and total cluster rain rate R (gray crosses). Each data point in the plots is the averaged scaling exponent $\tau_S(Z, S_0)$, S representing A or R. (e) Estimate for ζ_A (black) and ζ_R (gray) for individual focus regions. The point estimates are the regression intercepts of Figures 4a–4d. The horizontal lines show the range of values where the separate estimates for all focus regions overlap. The error bars represent the 95% confidence level of the estimates.

the finite resolution of the satellite data; in theory, the scaling exponent of the cluster area or rain rate will emerge in the thermodynamic limit, i.e., as the resolution and/or the domain size increases indefinitely. To this end, we modeled the scaling exponent-domain size dependence by a linear regression of τ_S with $1/Z$ as shown in the figures; the intercept of the regression ζ_S is the domain size-independent estimate of the scaling exponent. A robust iterative least squares method [*Holland and Welsch*, 1977] was used in the regression instead of the ordinary least squares (OLS) regression due to the presence of influential data points particularly for R. Some of the important features of the robust regression method are highlighted in Appendix A5.

All the regressions are significant at the 5% level except for the regression line of R for EP. The P value for the regression of R for EP is slightly larger than 0.05 (Figure 4c), likely a result of a lack of scaling exponent estimates at $Z = 120$; repeating the regression for IO, WP, and AO while excluding the scaling exponents estimates for $Z = 120$ (not shown) gave P values greater than 0.05 as well. This provides strong evidence that the linear model is a good model for EP, and as such, we included the original regression estimates for ζ_R for further analysis. The theoretical limits of the scaling exponents ζ_A and ζ_R estimated for each focus region are shown in Figure 4e. There is agreement in both estimates for ζ_A and ζ_R across the various ocean basins (cf. the horizontal lines in Figure 4e) demonstrating clearly the universality of oceanic rainfall scaling. We next estimated the universal scaling exponents for A and R by repeating the procedures described in the last section using the observed clusters from all the focus regions combined and obtained the values below:

$$\begin{aligned} \zeta_A &= 1.66 \pm 0.06 \\ \zeta_R &= 1.48 \pm 0.13 \end{aligned} \tag{5}$$

where the 95% confidence intervals of the estimates are cited. Both regressions for the combined regions are significant at the 5% level.

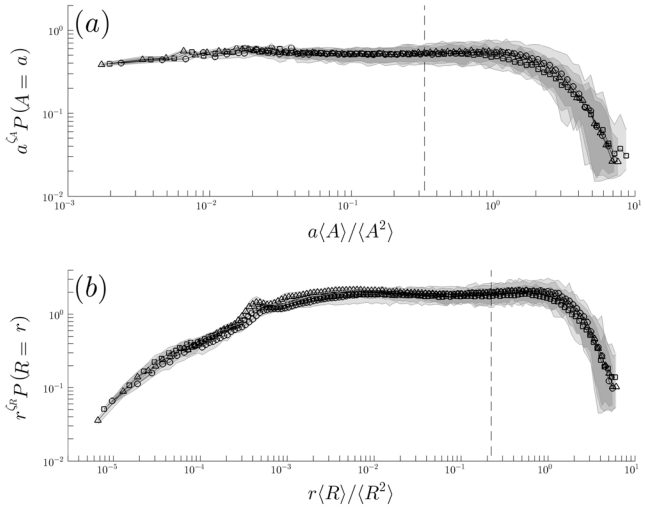

Figure 5. Inferred scaling function of the (a) cluster size and (b) cluster total rain rate for different oceans for clusters identified at domain size $Z = 120$ (black) for IO (squares), WP (triangles), and AO (circles). The bounds of the shaded area represent the 2.5% and the 97.5% of the 40 estimated scaling function values. The vertical dashed lines in both plots mark the rescaled system sizes ξ_A and ξ_R.

It is of interest to examine our scaling exponent estimates in the context of other relevant scaling estimates of tropical rain clusters in the literature, namely, those reported by *Peters et al.* [2009, 2012] hereafter abbreviated as P09 and P12, respectively, and *Traxl et al.* [2016].

Our estimates in equation (5) appear to be smaller in comparison to reported estimates for ζ_A of about 2–2.05 [P09; P12] and for ζ_R of about 1.87 [P12]. The rainfall products used by P09 and P12 are based on the TRMM Precipitation Radar (PR) which has a narrow swath width, unlike the IR instruments which the TRMM 3B42 product is also based on in addition to the PR. This raises the possibility that the discrepancies between our estimates and those of P09 and P12 could be due, at least in part, to the different aspect ratio of the domain from which the rain clusters were identified. To test this idea, we analyzed rain clusters from the WP region in domains of size 120 × 12 TRMM 3B42 grid points, mimicking the narrow swath of the PR used by P09 and P12. Similar plots to those in Figure 2 of the sample distributions of A and R from these elongated domains (not shown) were made. There were no discernable difference in the scaling exponent estimates from those in equation (5) within the error margins, suggesting that the differences in the aspect ratio of the domains are not the main cause for our departure from P09 and P12.

P09 showed that ζ_A depends on the total column water vapor content within the rain cluster, with the scale-free regime being most pronounced for rain clusters with the critical column water vapor content. They found that ζ_A is about 2 for these critical clusters and is *larger* for clusters with water vapor content below the critical value. Although the rain clusters used in our analysis were not conditioned on their water vapor content, our estimate for ζ_A is smaller, not larger, than 2 as one might expect following P09. Thus, water vapor content is unlikely to be the factor causing our departure from P09.

We suspect that the difference between our values and those reported by P09 and P12 to arise from the different sensitivity characteristics of rainfall products used. The TRMM 3B42 rainfall estimates that we used were based in part on passive radiometric sensing in the infrared bands. The infrared bands have limited accuracy in detecting finer rain features as they essentially sense cloud top height from which rainfall is estimated [*Arkin and Meisner*, 1987]. On the other hand, the PR is known to miss large areas of light rain due to its relatively low detection sensitivity [*Schumacher and Houze*, 2000]. With imperfect instrument observations, it is hard to ascertain which set of estimates is more reliable, and more future work is necessary to settle this question.

Traxl et al. [2016] reported a value of 1.712 for the scaling exponent analogous to ζ_R, but for *spatial-temporal* oceanic clusters (i.e., clusters in space and time, in contrast with our definition of spatial clusters) retrieved from TRMM 3B42. Unfortunately, it is difficult to draw any meaningful comparison with our estimates since their clusters included a temporal dimension. Future theoretical studies comparing the scaling of spatial and spatiotemporal clusters seem to be warranted at this juncture.

Strictly speaking, universality in SOC, and critical phenomena in general, also requires the scaling function $G_S(u)$ (equation (1)) to be identical [*Hinrichsen*, 2000; *Pruessner*, 2012, p. 20]. Evidence supporting the claim that the scaling functions across the different ocean basins are identical comes from Figure 5. There is a good data

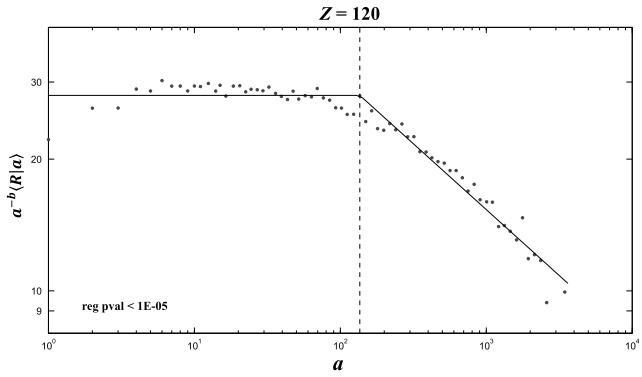

Figure 6. The observed conditional average of cluster total rain rate with different cluster area computed by combining the data observed over the focus regions with domain size $Z = 120$. The solid line shows the optimal piecewise regression with a single breakpoint represented by the vertical dashed line. Note that for clarity of presentation, we multiplied the factor a^{-b} (see equation (6) in the main text) to the ordinate using the regressed value of b.

collapse for the inferred scaling functions G_A and G_R for the cluster observed from the largest domain ($Z = 120$) for IO, WP, and AO. The corresponding scaling functions for EP at Z = 120 were not assessed due to uncertainty about its clusters' statistical homogeneity over the $Z = 120$ domain noted earlier (see section 2). Since the actual system sizes ξ_S associated with each focus region are unknown, A and R in the abscissa of the plots are rescaled by their respective moment ratios $\langle S^2\rangle/\langle S\rangle$, taking advantage of the fact that $\xi_S \propto \langle S^2\rangle/\langle S\rangle$ if $f_S(s)$ were to follow equation (1) [*Peters et al.*, 2010]. The data points plotted are the average of $s^{\zeta_S}P(S = s)$ computed from the 40 cluster data subsets using the estimated values of ζ_A and ζ_R (equation (5)). $\langle S\rangle$ ($\langle S^2\rangle$) was obtained by averaging the different estimates of $\langle S\rangle$ ($\langle S^2\rangle$) from the 40 different cluster data subsets. G_A and G_R similarly inferred from clusters from the smaller domain sizes showed reasonable collapse across different ocean basins but exhibit discernable difference across the largest and smallest domain (not shown). The mismatch of the inferred scaling functions across the different domain sizes is likely a result of increasing finite size effects on the cluster distributions for the smaller domain sizes. As such, we believe that the inferred scaling functions from the largest domains ($Z = 120$) for the different focus regions would be the most representative of the actual scaling function, which in theory would be defined in the thermodynamic limit $Z \to \infty$. We note from Figure 5b that G_R is not a constant for $r\langle R\rangle/\langle R^2\rangle \lesssim 10^{-3}$. This would imply that the probability density function of R is nonscaling for small enough r. Indeed, the deviation from power law can be visually discerned in Figure 3 for $r \lesssim 100$ for IO which corresponds to its best estimate of r_0, the smallest scale s_0 where the scaling ansatz (equation (1)) is valid when $S = R$.

5. Scaling Relations

The observed $\langle R|a\rangle$ was computed by combining the cluster data in all focus regions. This was done by first calculating the conditional mean cluster rain rate for a set of cluster area intervals for the individual $Z \times Z$ domains (that are land-free and dominated by deep convection) disregarding which focus region the domain belongs to. Subsequently, the ensemble average of the conditional mean cluster rain rate at each cluster area interval was used as an estimate of $\langle R|a\rangle$ for each Z. We found that the conditional mean can be modeled as a power law with a crossover exponent as illustrated by Figure 6:

$$\langle R|a\rangle = \begin{cases} ka^b & 1 \le a \le a_1 \\ ka_1^{b-c}a^c & a_1 < a < Z^2 \end{cases} \tag{6}$$

Note that the observed $\langle R|a = 1\rangle$ in Figure 6 deviated from equation (6) due to the misidentification of random grid-point noises as rainy grid-points. The values for the crossover cluster size a_1 and the exponents b and c were determined by performing piecewise linear least squares regression with the cluster data. Since a_1 must be known before determining b and c in the regression, we repeated the regression varying a_1 across the range of observed cluster size and adopted the regression with the

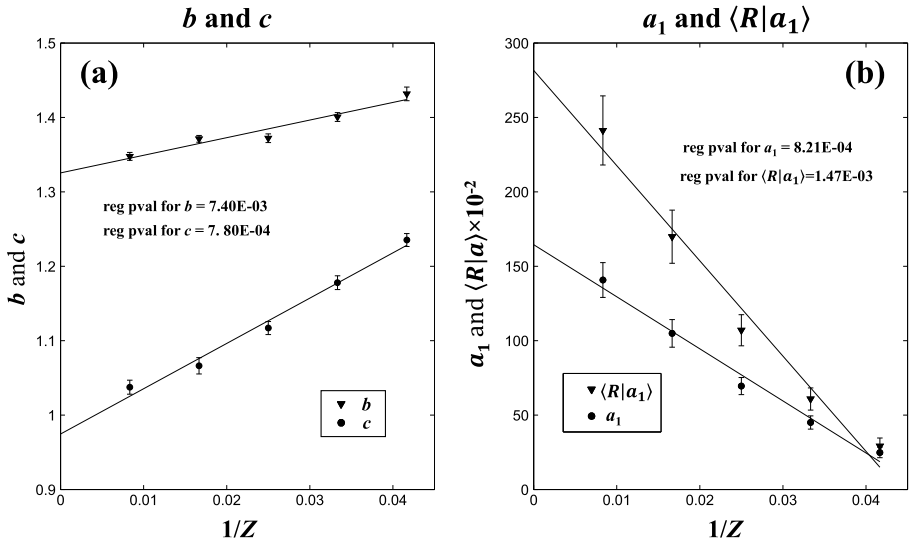

Figure 7. The regression of (left) b, c, (right) a_1 and $\langle R|a\rangle$ against the inverse of domain size from which the respective asymptotes β, χ, α, and λ were estimated as the regression intercepts.

least root-mean-square errors as the optimal solution. The technical details of the piecewise regression can be found in Appendix A6.

As in the case when estimating the scaling exponents (section 4), b, c, a_1, and $\langle R|a\rangle$ were found to be dependent on the domain size Z. We therefore performed linear regression for each variables b, c, and a_1 against $1/Z$ and took the corresponding regression intercepts as their corresponding domain size-independent estimates. The regressions of these variables are shown in Figure 7 and the estimated asymptotes ($Z \to \infty$) for b, c, a_1, and $\langle R|a_1\rangle$, denoted by β, χ, α, and λ, respectively, are

$$\begin{aligned} \beta &= 1.33 \pm 0.03 \\ \chi &= 0.97 \pm 0.04 \\ \alpha &= (1.6 \pm 0.2) \times 10^2 \Delta A \\ \lambda &= (2.8 \pm 0.5)\ \mathrm{cm/hr} \cdot \Delta A \end{aligned} \tag{7}$$

where $\Delta A = 773\ \mathrm{km}^2$, ignoring the curvature of the Earth's surface in the tropics. For rain clusters smaller than α (corresponding to an actual length scale ~320 km) the analysis interestingly suggests the cluster areal rain efficiency, defined as $d\langle R|a\rangle/da$, in mesoscale clusters increases as the cluster grows larger ($d\langle R|a\rangle/da \propto a^{\beta-1} \simeq a^{1/3}$ if we take $\beta \simeq 4/3$). In contrast, by taking $\chi \simeq 1$ for rain clusters larger than α, the cluster areal rain efficiency practically saturates to the constant λ/α (corresponding to 1.71 mm/h from our analysis).

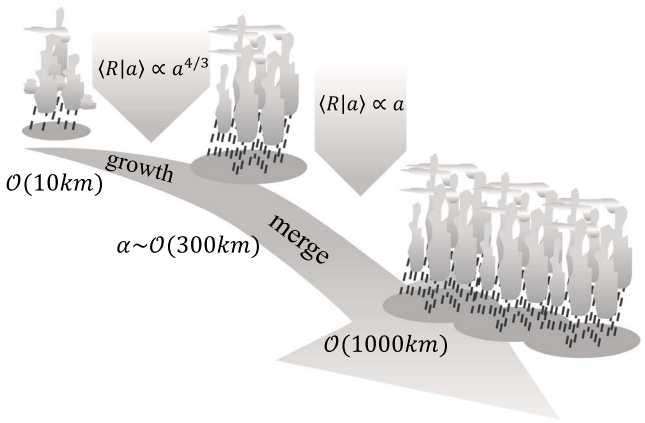

Figure 8. A schematic depicting the possible reasons behind the relation between clusters size and rain rate. The opposite process of breakup is also possible.

One way to understand these results is to suppose that below 320 km, clusters tend to grow by increasing the size and intensity of individual storms. Above 320 km, clusters tend to grow into superclusters by organizing themselves together without much intensification as depicted in

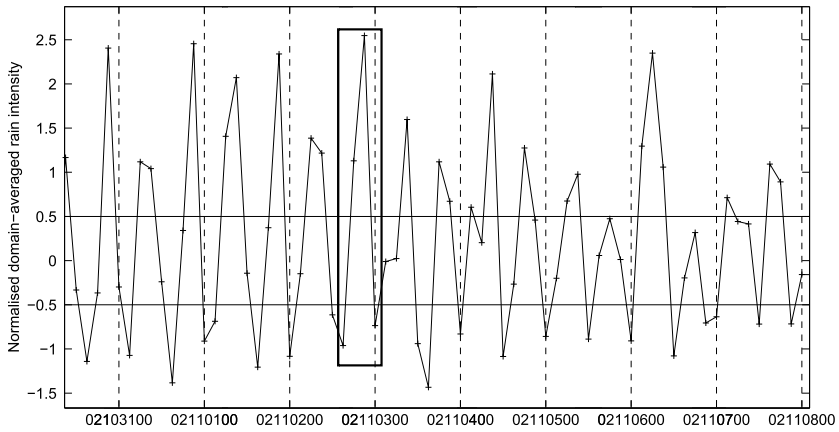

Figure 9. Time series of the normalized domain-mean cluster rain intensity for IO. The vertical dashed lines mark the time stated in the abscissa in the two-digit year-month-day:hour format in UTC. The time interval between data points (+) is 3 h. The horizontal lines denote half a standard deviation above and below the long-term mean of the time series.

Figure 8. Similarly, the breakup of clusters in the manner opposite to what is illustrated in Figure 8 is also possible.

To provide some motivation for our speculation as to how the rain clusters evolve, we computed the domain-averaged cluster rain intensity, $\left\langle \frac{R}{a} \right\rangle$, over the 120 × 120 domain in IO at 3-hourly intervals of the TRMM 3B42 data. Figure 9 shows the normalized, 30 day high band-pass filtered $\left\langle \frac{R}{a} \right\rangle$ time series over several days in November 2002 for discussion. The variation of the cluster intensity shown is fairly representative of the rest of the time series, with $\left\langle \frac{R}{a} \right\rangle$ peaking in the afternoon around 09 UTC (about 14 local standard time (LST) over IO) and during the early night around 21 UTC (about 02 LST over IO). The timescale of the intensification of the oceanic clusters is about 6 h. Such double peaks in the rainfall diurnal cycle over the Indian Ocean were also observed in other studies [e.g., *Yang and Smith*, 2006]. The average cluster area for the set of time slices of which $\left\langle \frac{R}{a} \right\rangle$ is smaller (greater) than half a standard deviation from the long-term mean was found to be 15 ΔA (24 ΔA), consistent with the idea of clusters growing in size while intensifying with time. Of course, Figure 9 is also consistent with the breakup of clusters in the direction opposite to that depicted in Figure 8.

The scaling relation of equation (3) is consistent with our observations since our point estimates for ζ_A and ζ_R and β from the satellite data can be approximated below as

$$\begin{aligned} \zeta_A &\simeq \frac{5}{3} \\ \zeta_R &\simeq \frac{3}{2} \\ \beta \simeq \frac{4}{3} &= \frac{\frac{5}{3}-1}{\frac{3}{2}-1} \end{aligned} \tag{8}$$

6. Concluding Remarks

We derived estimates for the scaling exponents for the cluster area and cluster rain rate distributions for tropical oceanic rainfall. To account for the effects of the finite domain size on the scaling of these rain clusters, we performed the analysis for different domain size Z in each oceanic region. By assuming a linear relation between the scaling exponents and $1/Z$, we were able to estimate the "true" values of the exponents in the limit as the sampled domain becomes unbounded. Our investigation suggests that rain clusters over the tropical oceans possess a few simple universal characteristics: (1) The cluster area and its total rain rate exhibit scaling behaviors and their scaling exponents were estimated to be ζ_A = 1.66 ± 0.06 and

ζ_R = 1.48 ± 0.13, respectively. (2) The two exponents ζ_A and ζ_R are related to the exponent β of the conditional average cluster rain rate for a given cluster area in a form consistent with the SOC paradigm (equation (3)) if the fractions 5/3, 3/2, and 4/3 are adopted for ζ_A, ζ_R and β, respectively. Both ζ_A and ζ_R estimated in this work are smaller than those obtained from analyses based on TRMM PR rainfall estimates reported in *Peters et al.* [2012]. Unfortunately, that piece of work only investigated the western Pacific Ocean and so they did not show that the probability distributions of A (or R) are universal in the sense that they collapse onto the same curve when the system is rescaled by its characteristic scale, ξ_A (or ξ_R), which would take on different values for different ocean basins. Sensitivity tests to the domain size were also not carried out in that paper.

At this juncture, explanations for the observed cluster area-rain rate relation (equation (2)) are needed: (1) What is the reason behind the emergence of a crossover cluster size α as a cluster grows? Is it related to intrinsic convective self-organization dynamics or is it imposed by synoptic-scale dynamics such as the Madden-Julian Oscillation [*Madden and Julian*, 1994] or equatorial waves? (2) Why does the cluster areal rain efficiency for mesoscale clusters (i.e., clusters with area less than α) increase with area? (3) Why is there a "maximal" rain intensity λ/α, what determines it and why is λ/α proportional to $\alpha^{\beta-1}$? Investigations directed to these questions would complement ongoing research into organized tropical convection.

Why tropical convective rainfall is an SOC as demonstrated in this paper needs to be better understood. Developing an SOC model based on the known physics of moist convection could be a way forward to elucidate the mechanisms that cause the tropical oceanic rain to exhibit SOC. *Stechmann and Neelin* [2014] have recently proposed a first-passage process model, a prototype SOC based on simple moist physics, for the time evolution of convective precipitation. Interestingly, they reported a scaling exponent of 3/2 for the event size, the temporal analogue to ζ_R of spatial rain clusters, which agrees with our estimate of ζ_R. However, their model also produced the same value of 3/2 for the scaling exponent for the event duration, the analogue to ζ_A of spatial rain clusters (cf. our estimate of 5/3 for ζ_A). The equivalence in their scaling exponent for the event size and the event duration is perhaps unsurprising, since the rain intensity was independent of the event duration in their model, leading to the event size being proportional to the event duration on average. A similar stochastic SOC model for spatial rain clusters could be developed, perhaps by taking into account the conditional rain rate $<R|A>$ identified in this paper which determines the rain intensity. The scaling exponents of this model can then be compared directly against the observed estimates reported here and elsewhere. Other stochastic rain models that account for the spatial aggregation of convection have been developed in the past [*Randall and Huffman*, 1980], but it is not known whether or not they exhibit SOC.

Recent RCE experiments using atmospheric general circulation models (AGCMs) coupled with slab ocean models conducted by *Reed et al.* [2015] seem to lend support to the conjecture that the real-world RCE maintains itself between the disorganized convective state and the state of convective self-aggregation through self-regulating the SST. They reported that self-aggregated convective clusters observed in an AGCM under a prescribed uniform SST boundary condition were absent when the AGCM was coupled to an interactive slab ocean model. The SST in their ocean-atmosphere coupled experiment at RCE was cooler than the prescribed SST, at which convective self-aggregation was observed in their atmosphere-only experiment. In view of the observational evidence that the column water vapor is the self-tuning parameter for the SOC in deep convection [*Peters and Neelin*, 2006; *Peters et al*., 2009], the apparent self-regulatory role of the SST in convective self-organization seen in RCE experiments such as those mentioned above could be an extension of the relation between SST and the atmospheric water vapor content. The deep convective regions, on which our present analysis is based, are regions of warm SST and high moisture content in the free troposphere as well [*Kanemaru and Masunaga*, 2013]. Simulations with an idealized cloud-resolving model coupled with a slab ocean model configured for RCE experiments could potentially be useful in elucidating the role of SST and column water vapor in convective self-organization. If a critical state exists and can be self-maintained in the modeled RCE, it would demonstrate that SOC behavior, such as the scaling exponents and relations analyzed in this paper, can be achieved in the modeled ocean–atmosphere system. Due to the relatively low demand on computational resources, well-defined boundary conditions and domain size, a systematic investigation can be conducted to address the above questions.

Appendix A

A1. Scaling Relations for A and R With Crossover in the Exponent in $\langle R|\, a\rangle$

The probability density function (pdf) of the cluster area A and total rain rate R are assumed to be simple scaling:

$$f_S(s) = \begin{cases} c_S s^{-\zeta_S} G_S\left(\frac{s}{\xi_S}\right) & s \geq s_0 \\ q_S(s) & 0 \leq s < s_0 \end{cases} \tag{A1}$$

where $s \in \{a, r\}$, G_S is the finite scaling function and $q_S(s)$ is the nonscaling portion of the pdf and c_S does not depend on ξ_S. $\langle R|\, a\rangle$ is given by equation (2) in the main text. We further assume that

$$\alpha = h_A \xi_A \text{ and } \xi_R \propto \langle R|\xi_A\rangle \Rightarrow \xi_R = h_R \xi_A^{\beta} \tag{A2}$$

based on the following assertions:

1. The crossover is fundamentally an emergent behavior independent of the mesoscale workings of the convective clusters, and the crossover scale has some functional relations with the characteristic spatial scale, ξ_A, which depends on the basin-wide dynamics (e.g., SST pattern and the larger-scale atmospheric circulation of the Walker cell).
2. ξ_R and ξ_A can be loosely interpreted respectively as the "largest" cluster rain rate and the "largest" cluster size that could be found within the system. Thus, it is reasonable to relate the "largest" cluster rain rate to the "largest" cluster size in the system.

The simple proportional relations between the variables are chosen for the lack of reasons to require more complicated relations in our theoretical constructs.

Using equation (A1), the mean cluster rain rate can be shown to be

$$\langle R\rangle = w_1^{(R)}(r_0) + c_R \xi_R^{2-\zeta_R} g_{1-\zeta_R}^{(R)}\left(\frac{r_0}{\xi_R}, \infty\right); \;\; 1 < \zeta_R < 2 \tag{A3}$$

where $w_\eta^{(S)}(s_0) \overset{\text{def}}{=} \int_0^{s_0} s^\eta q_S(s) ds$ and $g_\eta^{(S)}(x, y) \overset{\text{def}}{=} \int_x^y u^\eta G_S(u) du$ for $S \in \{A, R\}$.

Since $\langle R\rangle = \int_0^\infty f_A(a)\langle R|a\rangle da$, we can also compute the mean cluster rain rate using equation (2) in the main text and $f_A(a)$ given by equation (A1) noting that $a_0 < \alpha$, as

$$\begin{aligned} \langle R\rangle &= k\int_0^{a_0} a^\beta q_A(a) da + kc_A \int_{a_0}^{\alpha} a^{\beta-\zeta_A} G_A\left(\frac{a}{\xi_A}\right) da \\ &\quad + kc_A a_1^{\beta-\chi} \int_\alpha^\infty a^{\chi-\zeta_A} G_A\left(\frac{a}{\xi_A}\right) da \\ &= k w_\beta^{(A)}(a_0) + kc_A \xi_A^{\beta-\zeta_A+1} g_{\beta-\zeta_A}^{(A)}\left(\frac{a_0}{\xi_A}, \frac{\alpha}{\xi_A}\right) \\ &\quad + kc_A a_1^{\beta-\chi} \xi_A^{\chi-\zeta_A+1} g_{\chi-\zeta_A}^{(A)}\left(\frac{\alpha}{\xi_A}, \infty\right) \end{aligned} \tag{A4}$$

for $\beta - \zeta_A + 1 > 0$, $-1 < \chi - \xi_A < 0$.

Equating equations (A3) and (A4) and using equation (A2), we obtain

$$\xi_A^{\beta(1-\zeta_R)+\zeta_A-1} = \frac{\frac{k w_\beta^{(A)}(a_0)}{\xi_A^{\beta-\zeta_A+1}} + kc_A\left[g_{b-\zeta_A}^{(A)}\left(\frac{a_0}{\xi_A}, h_A\right) + h_A^{\beta-\chi} g_{\chi-\zeta_A}^{(A)}(h_A, \infty)\right]}{\frac{w_1^{(R)}(r_0)}{\xi_A^{\beta(2-\zeta_R)}} + c_R h_R^{2-\zeta_R} g_{1-\zeta_R}^{(R)}\left(\frac{r_0}{h_R \xi_A^b}, \infty\right)} \tag{A5}$$

If $w_\beta^{(S)}(s_0)$ is bounded above, the limit of the RHS of equation (A5) exist as $\xi_A \to \infty$ $\left(= \frac{kc_A\left[g_{\beta-\zeta_A}^{(A)}(0, h_A) + h_A^{\beta-\chi} g_{\chi-\zeta_A}^{(A)}(h_A, \infty)\right]}{c_R h_R^{2-\zeta_R} g_{1-\zeta_R}^{(R)}(0, \infty)} > 0\right)$. This implies the exponent of the ξ_A term in the LHS of equation (A5) must be zero:

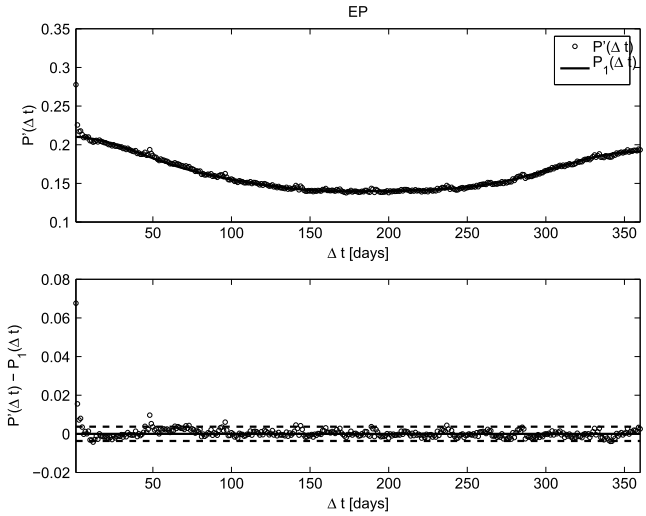

Figure A1. $P'(\Delta t)$ for EP estimated from the gridded rainfall data from the satellite observations. (top) $P'(\Delta t)$ estimated for lags with sampling interval of 1 day for a maximum lag of 365 days (open circles). The bold line is the estimated secular variation. (bottom) Residuals of $P'(\Delta t)$ after removing the secular variation showing that $P'(\Delta t)$ becomes indistinguishable from the background variability (demarcated by the dotted lines) after 5 days. The 95% of the residual are bounded between the dotted lines by definition.

$$\beta(1-\zeta_R)+\zeta_A-1=0 \tag{A6}$$

A2. Subsets of Cluster Data

The entire cluster data were partitioned into subsets each of which contains rainfall that are approximately statistically independent in time. The decorrelation timescale of the rain clusters was estimated separately for each focus region by the conditional probability for a rainy grid-point to be raining after a time lag Δt:

$$P(\Delta t) \triangleq Prob(q(t+\Delta t)0 \mid q(t)>0) \tag{A7}$$

where $q(t)$ is the rainfall intensity for a grid point within the 120×120 grid points of a given focus region at time t. The lag interval used in our calculation was 1 day. The largest lag available for $P(\Delta t)$ is more than 10 years. We used $P(\Delta t)$ as a proxy of cluster correlation instead of the standard autocorrelation of $q(t)$ since the latter statistic would be influenced by the autocorrelation of nonrainy grid points (i.e., grid points that are not raining at t and $t + \Delta t$), which was not desirable for our purpose.

Figure A1 illustrates how the decorrelation timescale was estimated from $P(\Delta t)$ using the focus region EP as an example. Visual inspection of the time series of $P(\Delta t)$ indicated the presence of a pronounced secular variation with a 1 year period for all focus regions. This 1 year trend exists because of the strong annual rain cycle in these regions; the probability of a TRMM grid point being rainy would likely to be the highest during the rainy seasons. We estimated the secular variation for each region separately following the below procedure:

1. First estimate any long-term trend, $P_0(\Delta t)$, in the decade-long lag time series of $P(\Delta t)$ for a given region by a polynomial regression of degree 2 using the averaged $P(\Delta t)$ across lag bins with bin width of 1 year. Then we removed the estimated long term trend from the lag time series, i.e. $P'(\Delta t) = P(\Delta t) - P_0(\Delta t)$.
2. Calculate the averaged 1 year secular variation by averaging the detrended lag time series $P'(\Delta t)$ in the following manner:

$$P_1(\Delta t)=\frac{\sum_{j=0}^{N} P'(365j+\Delta t)}{N-1};\Delta t=0,1,2,\cdots,364 \text{ days} \tag{A8}$$

where 365 $N + \Delta t$ is the largest available lag in the lag time series for a given Δt. $P_1(\Delta t)$ represents the contribution to $P'(\Delta t)$ by the annual rain cycle. For lag larger than a year $P_1(\Delta t) = P_1(\Delta t \bmod 365)$.

3. Finally obtain the "smoothed" secular variation using a low-pass Lanczos filter on $P_1(\Delta t)$ with a cutoff period of 4-months. Figure A1 (top) shows $P'(\Delta t)$ up to a lag of a year for EP and the estimated secular variation $P_1(\Delta t)$.

Subsequently, the departure of $P'(\Delta t)$ was obtained from the estimated secular variation. The decorrelation time of a rainy grid point was then estimated to be the minimum lag where $P'(\Delta t)$ became indistinguishable from the background variability (Figure A1 (bottom)). The background variability is defined as the symmetric range about zero, which contained 95% of the departures of $P'(\Delta t)$ from the secular variation. Repeating the analyses for the other focus areas gave a decorrelation timescale of 5 days as well. The cluster data were subsequently divided into nonoverlapping subsets with each set containing rain clusters observed at the

Table A1. Number of $Z \times Z$ Domains Within Each Focus Region That Satisfy the Two Conditions Described in Appendix A3[a]

	Z				
	120 [1]	60 [4]	40 [9]	30 [16]	24 [25]
IO	1	3	6	11	19
WP	1	4	9	16	22
EP	0	2	3	8	10
AO	1	2	3	8	13

[a]The numbers in the square brackets beside the Z values are the maximum number of $Z \times Z$ domains that we can have for each focus region.

same time in coordinate universal time (UTC) spaced 5 days apart. For a given focus region 40 subsets (five subsets for each for the eight different TRMM observation time of the day, 00 UTC, 03 UTC,…, 21UTC) were therefore available for analysis for each $Z \times Z$ domain.

A3. Selecting Deep Convective Cluster Domains

There are 40 different data series in the TRMM data set (from eight starting times a day over the first five consecutive days, with data points in each series spaced 5 days apart). To be admitted for further analysis, each $Z \times Z$ domain covering the focus regions (e.g., one of the four domains with $Z = 60$ over IO) was tested for two conditions:

1. The land fraction coverage over the domain must be less than 10% of the domain area, Z^2.
2. Denote $m_{j,k}$ (Z) as the average cluster rain rate for the jth $Z \times Z$ domain satisfying condition (1) in the kth data series. Suppose there are J such domains over all ocean basins. For example, $J = 16$ if all 60×60 domains in the four ocean basins satisfy condition (1). We require that $m_{j,k}$ $(Z) > M(Z) - S(Z)$ for all k, where $M(Z)$ and $S(Z)$ are respectively the average and standard deviation of the set $\{m_{j,k}\ (Z) \mid j = 1, \ldots, J;\ k = 1, \ldots, 40\}$.

The motivation of the first condition is in the main text and needs no further elaboration. The second condition was applied so as to broadly exclude those regions where shallow convection dominates. The number of domains that satisfy the two conditions is summarized in Table A1. Most of the domains over IO and WP were included for subsequent analysis. For EP, the largest domain ($Z = 120$) and all domains south of ~5°N were excluded as they failed to satisfy condition (2); the domains that remained were those in the East Pacific ITCZ (cf. Figure 2), evidently where frequent deep convection dominates. Similar for AO, the smaller domains away from the Atlantic ITCZ were excluded. Thus, condition (2) is a simple but effective empirical rule to identify regions of frequent deep convection.

For brevity, we shall call the domains that satisfy the above two conditions as "valid domains."

A4. Estimating the Scaling Exponents

For a given focus region the unknown probability model parameters for the cluster area and cluster rain rate for each valid domain size Z were determined using the maximum likelihood method (MLM) [*Papoulis and Pillai*, 2001]. The log likelihood function associated with equation (4) in the main text is proportional to

$$F(\tau_S, \nu_S, S_c) = -\ln C_S - \tau_S \langle \ln s \rangle - \gamma_S \langle s^{\nu_S} \rangle \tag{A9}$$

where s stands for a or r and $\langle \cdot \rangle$ denotes the sample mean. The MLM estimates of γ_S and ν_S were then used to estimate $S_c = \gamma_S^{-1/\nu_S}$. In equation (A9), we estimated the model parameters by maximizing F using the interior-point optimization method [*Waltz et al.*, 2006] available in standard computational package subjected to the constraints that all the parameters were bounded below by zero.

For each focus region and domain size Z, we fitted the model parameters $[\tau_S, \nu_S, S_c]$ to the observed ensemble distributions using the MLM estimates for $s \geq S_0$ using a range of S_0 values to account for the uncertainty in S_0. The selected S_0 values were 1 to 64 for A and 1 to 1024 for R with equal binary logarithmic intervals (e.g., the A_0 tested were 1, 2, 4, 8, 16, 32, and 64). The upper limit of the A_0 being tested was chosen so that it was about 2 orders of magnitude less than the smallest area of the domain to be analyzed (i.e., $24 \times 24 = 576$). For R, preliminary inspection of their sample probability distribution for the different domain size suggests that the upper limit of the scaling regime is in the order of 10^5. We therefore selected the upper limit of R_0 tested to be ~10^3 so as to have at least 2 orders of magnitude of separation in any detected scaling regime. Estimation of the model parameters was repeated for all the cluster data subsets resulting in 40 sets of model parameters for each (S_0,Z) pair per focus region. Out of these 40 potential τ_S, the optimal set of scaling exponents were subsequently selected as follows: (1) The scaling exponent corresponding to those fitted probability distributions that exhibited a scaling regime less than an order in S (i.e., $S_c/S_0 < 10$) was discarded.

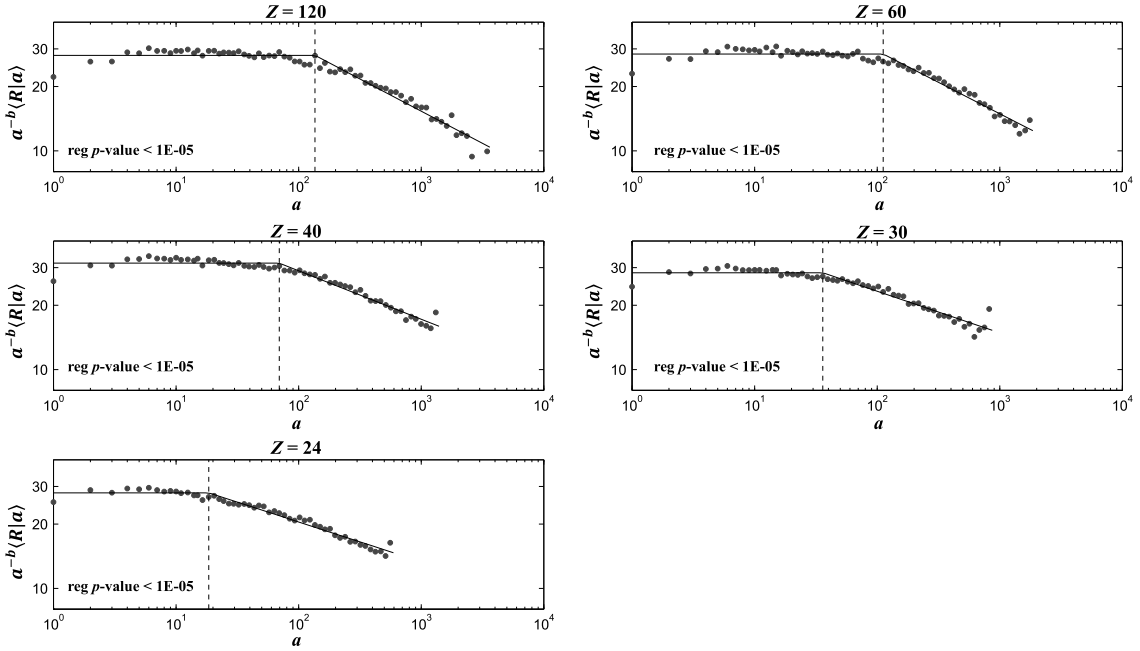

Figure A2. Same as Figure 6 but for different domain sizes Z.

(2) From the remaining scaling exponents, we chose the best fitted model parameters by keeping only those scaling exponents if their corresponding Kolmogorov-Smirnov statistics [e.g., *Sachs*, 1984, p. 330] were less than the 25th percentile. Finally, we obtained a single representative scaling exponent corresponding to each S_0 tested by averaging across those τ_S that satisfy the above two selection criteria.

A5. Robust Iteratively Least Squares Linear Regression

Here we highlight the main features of the robust iteratively reweighted least squares (IRLS) linear regression used to obtain the scaling exponents ζ_A and ζ_R (Figure 4). Technical details of the robust IRLS can be found in *Holland and Welsch* [1977] and *Street et al.* [1988]. Given a set of n predictor $\{x_i|i=1,\cdots,n\}$ and the corresponding set of response $\{y_i|i=1,\cdots,n\}$, the method solves for the regression coefficient $\mathbf{b}=(b_0\ b_1)^T$ where $y_i=b_0+b_1x_i+\varepsilon_i$, $\{\varepsilon_i|i=1,\cdots,n\}$ being the identical and independent errors, by solving the system of equations:

$$\sum_{i=1}^{n} w_i e_i \boldsymbol{x}_i^T = 0 \tag{A10}$$

where $\boldsymbol{x}_i=(1\ x_i)^T$; $e_i = y_i - \boldsymbol{x}_i^T\mathbf{b}$; $\mathbf{0}$ is the zero vector and the weights w_i used in our work are the Huber weights:

$$w_i = \begin{cases} 1 & \text{for } |e_i| \le k \\ \dfrac{k}{|e_i|} & \text{for } |e_i| > k \end{cases} \tag{A11}$$

where $k = 1.345$. This value of k gives coefficient estimates that are approximately 95% as statistically efficient as the ordinary least squares estimates, provided the response has a normal distribution with no outliers. Since the weights in equation (A10) depend on the residuals $\{e_i\}$ which in turn depends on the estimate of $\mathbf{b}$, equation (A11) needs to be solved iteratively from an initial guess of $\mathbf{b}$ from ordinary least squares regression.

A6. Modeling the Conditional Mean Cluster Rain Rate

For each valid $Z \times Z$ domain (with no differentiation of which focus region the domain belongs to), we first obtained the conditional rain rate by averaging the cluster rain rate for clusters with area that fall within an equal logarithmic bin interval $[1.1^j\ 1.1^{j+1})$, $j = 0, 1, 2, 3, \ldots$. After that, the ensemble average of the (binned) conditional mean cluster rain rate among all the domains of size Z was obtained. Preliminary inspection of the

logarithmic plots of the ensemble conditional mean cluster rain rate versus cluster area (e.g., Figure A2) suggested that the conditional means can be modeled by two linear regimes of different gradient across the range of cluster size with equation (6) in the main text. Note that k in the RHS of equation (6) has the meaning of the mean cluster rainfall for the smallest cluster resolvable ($a = 1$) by TRMM 3B42. For each domain size Z, we estimated b and c by performing an ordinary least squares (OLS) piecewise linear regression with one breakpoint (a_1) on the logarithm of the computed conditional mean cluster rain rate and the logarithm of the cluster area. In our work, the regression was performed by recasting the logarithm of equation (6) as an OLS regression with two variables:

$$\log\langle R|a\rangle = cz_2 + bz_1 + \log k \tag{A12}$$

where $z_1 = \log\frac{a}{a_1} J\left(\log\frac{a_1}{a}\right) + \log a_1$, $z_2 = \log\frac{a}{a_1}\left[1 - J\left(\log\frac{a_1}{a}\right)\right]$, $J(x) = \begin{cases} 0 & x \geq 0 \\ 1 & x < 0 \end{cases}$.

OLS regression was chosen in this case as opposed to the robust technique employed for regression analysis for the scaling exponents because, compared to the latter case, there were no obvious outliers that have sufficient leverage to appreciably affect the estimation of the regression coefficients. As the break point a_1 was needed to be specified before the linear regression with equation (A12), we repeated the regression using the geometric mean of the bin intervals (except the first and last bin where the model represented by equation (6) in the main text would not be valid) as candidate a_1 values. The optimal set of estimates for b, c, and a_1 were subsequently identified from the regression with the smallest regression root-mean-square error. An example of the optimal set of regressions is shown in Figure A2. The above mentioned procedure was repeated for each of the 40 nonoverlapping cluster data subsets for a given domain size Z. The mean of the resulting 40 estimates of b, c, and a_1 were then taken to the best estimate for b, c, and a_1 for the given Z.

Acknowledgments

We thank Matthew R. Igel and another anonymous reviewer for their suggestions and feedback for the manuscript. H.N.H. acknowledges the support of the A*STAR International Fellowship. The TRMM 3B42 is freely available from the website hosted by the NASA Goddard Earth Sciences Data and Information Services Center, GES-DISC, http://disc.sci.gsfc.nasa.gov. The MISR data were obtained from the NASA Langley Research Center Atmospheric Science Data Center. This research is in collaboration with the Center for Environmental Sensing and Modeling (CENSAM) and is supported by the National Research Foundation Singapore under its Campus for Research Excellence and Technological Enterprise program. CENSAM is an interdisciplinary research group of the Singapore MIT Alliance for Research and Technology.

References

Arkin, P. A., and B. N. Meisner (1987), The relationship between large-scale convective rainfall and cold cloud over the western hemisphere during 1982-84, *Mon. Weather Rev.*, *115*(1), 51–74.

Bak, P., C. Tang, and K. Wiesenfeld (1987), Self-organized criticality: An explanation of the 1/f noise, *Phys. Rev. Lett.*, *59*(4), 381.

Bretherton, C. S., P. N. Blossey, and M. Khairoutdinov (2005), An energy-balance analysis of deep convective self-aggregation above uniform SST, *J. Atmos. Sci.*, *62*(12), 4273–4292.

Bretherton, C. S., R. Wood, R. C. George, D. Leon, G. Allen, and X. Zheng (2010), Southeast Pacific stratocumulus clouds, precipitation and boundary layer structure sampled along 20 S during VOCALS-REx, *Atmos. Chem. Phys.*, *10*(21), 10639–10654.

Cahalan, R. F., and J. H. Joseph (1989), Fractal statistics of cloud fields, *Mon. Weather Rev.*, *117*(2), 261–272.

Christensen, K., and N. R. Moloney (2005) *Complexity and Criticality*, p. 273, Imperial College Press, London.

Di Girolamo, L., A. Menzies, G. Zhao, K. Mueller, C. Moroney, and D. J. Diner (2010), Level 3 cloud fraction by altitude algorithm theoretical basis, JPL Tech. Doc. D-62358, 18 pp.

Hinrichsen, H. (2000), Nonequilibrium critical phenomena and phase transitions into absorbing states, *Adv. Phys.*, *49*(7), 815–958.

Holland, P. W., and R. E. Welsch (1977), Robust regression using iteratively reweighted least-squares, *Commun. Stat. Theor. M.*, *6*(9), 813–827.

Houze, R. A., Jr. (2004), Mesoscale convective systems, *Rev. Geophys.*, *42*, RG4003, doi:10.1029/2004RG000150.

Huffman, G. J., D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong, K. P. Bowman, and E. F. Stocker (2007), The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, *J. Hydrometeorol.*, *8*(1), 38–55.

Kanemaru, K., and H. Masunaga (2013), A satellite study of the relationship between sea surface temperature and column water vapor over tropical and subtropical oceans, *J. Clim.*, *26*(12), 4204–4218.

Khairoutdinov, M. F., and K. A. Emanuel (2010), Aggregated convection and the regulation of tropical climate, extended abstracts, 29th Conf. on Hurricanes and Tropical Meteorology, Tucson.

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

Madden, R. A., and P. R. Julian (1994), Observations of the 40-50-day tropical oscillation—A review, *Mon. Weather Rev.*, *122*(5), 814–837.

Mapes, B. E. (1993), Gregarious tropical convection, *J. Atmos. Sci.*, *50*(13), 2026–2037.

Mapes, B., S. Tulich, J. Lin, and P. Zuidema (2006), The mesoscale convection life cycle: Building block or prototype for large-scale tropical waves?, *Dyn. Atmos. Oceans*, *42*(1), 3–29.

Muller, C., and I. M. Held (2012), Detailed investigation of the self-aggregation of convection in cloud-resolving simulations, *J. Atmos. Sci.*, *69*(8), 2551–2565.

Muller, C., and S. Bony (2015), What favors convective aggregation and why?, *Geophys. Res. Lett.*, *42*, 5626–5634, doi:10.1002/2015GL064260.

Papoulis, A., and U. Pillai (2001), *Probability, Random Variables and Stochastic Processes*, 4th ed. pp. 320–322, McGraw-Hill, New York.

Peters, O., and J. D. Neelin (2006), Critical phenomena in atmospheric precipitation, *Nat. Phys.*, *2*(6), 393–396.

Peters, O., J. D. Neelin, and S. W. Nesbitt (2009), Mesoscale convective systems and critical clusters, *J. Atmos. Sci.*, *66*(9), 2913–2924.

Peters, O., A. Deluca, Á. Corral, J. D. Neelin, and C. E. Holloway (2010), Universality of rain event size distributions, *J. Stat. Mech. Theor. Exp.*, *2010*(11), P11030.

Peters, O., K. Christensen, and J. D. Neelin (2012), Rainfall and dragon-kings, *Eur. Phys. Spec. Top.*, *205*(1), 147–158.

Pruessner, G. (2012), *Self-Organised Criticality: Theory, Models and Characterisation*, pp. 20–43, Cambridge Univ. Press, New York.

Randall, D. A., and G. J. Huffman (1980), A stochastic model of cumulus clumping, *J. Atmos. Sci.*, *37*(9), 2068–2078.

Reed, K. A., B. Medeiros, J. T. Bacmeister, and P. H. Lauritzen (2015), Global radiative–convective equilibrium in the Community Atmosphere Model, version 5, *J. Atmos. Sci.*, *72*(5), 2183–2197.

Sachs, L. (1984), *Applied Statistics*, 2nd ed., *Springer Series in Statistics*, p. 330, Springer, New York, Translated from the fifth German edition by Z. Reynarowych.

Schumacher, C., and R. A. Houze Jr. (2000), Comparison of radar data from the TRMM satellite and Kwajalein oceanic validation site, *J. Appl. Meteorol.*, *39*(12), 2151–2164.

Stechmann, S. N., and J. D. Neelin (2014), First-passage-time prototypes for precipitation statistics, *J. Atmos. Sci.*, *71*(9), 3269–3291.

Street, J. O., R. J. Carroll, and D. Ruppert (1988), A note on computing robust regression estimates via iteratively reweighted least squares, *Am. Stat.*, *42*(2), 152–154.

Traxl, D., N. Boers, A. Rheinwalt, B. Goswami, and J. Kurths (2016), The size distribution of spatiotemporal extreme rainfall clusters around the globe, *Geophys. Res. Lett.*, *43*, 9939–9947, doi:10.1002/2016GL070692.

Teo, C.-K., T.-Y. Koh, J. C.-F. Lo, and C. B. Bhatt (2011), Principal component analysis of observed and modeled diurnal rainfall in the Maritime Continent, *J. Clim.*, *24*(17), 4662–4675.

Waltz, R. A., J. L. Morales, J. Nocedal, and D. Orban (2006), An interior algorithm for nonlinear optimization that combines line search and trust region steps, *Math. Prog.*, *107*(3), 391–408.

Wing, A. A., and K. A. Emanuel (2014), Physical mechanisms controlling self-aggregation of convection in idealized numerical modelling simulations, *J. Adv. Model. Earth. Syst.*, *6*(1), 59–74.

Wood, R., and P. R. Field (2011), The distribution of cloud horizontal sizes, *J. Clim.*, *24*(18), 4800–4816.

Yang, S., and E. A. Smith (2006), Mechanisms for diurnal variability of global tropical rainfall observed from TRMM, *J. Clim.*, *19*(20), 5190–5226.