POTSDAM-INSTITUT FÜR KLIMAFOLGENFORSCHUNG

AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2016GL070692

D. Traxl and N. Boers contributed equally to this work.

Key Points:
- Spatiotemporal integration of high-resolution data to measure the size distribution of rainfall clusters around the globe
- Extreme rainfall cluster sizes over the global oceans are distributed according to an exponentially truncated power law
- Generative model emphasizes the role of land masses for the truncation of extreme rainfall cluster size distribution

Correspondence to:
D. Traxl and N. Boers,
dominik.traxl@posteo.org;
boers@pik-potsdam.de

Citation:
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.

Received 2 AUG 2016
Accepted 5 SEP 2016
Accepted article online 16 SEP 2016
Published online 23 SEP 2016

The size distribution of spatiotemporal extreme rainfall clusters around the globe

D. Traxl[1,2,3], N. Boers[3,4], A. Rheinwalt[3,5], B. Goswami[3,5], and J. Kurths[1,3,6]

[1]Department of Physics, Humboldt Universität zu Berlin, Berlin, Germany, [2]Bernstein Center for Computational Neuroscience, Berlin, Germany, [3]Potsdam Institute for Climate Impact Research, Potsdam, Germany, [4]Geosciences Department and Laboratoire de Météorologie Dynamique, Ecole Normale Supérieure, Paris, France, [5]Institute of Earth and Environmental Science, University of Potsdam, Potsdam, Germany, [6]Department of Control Theory, Nizhny Novgorod State University, Nizhny Novgorod, Russia

Abstract The scaling behavior of rainfall has been extensively studied both in terms of event magnitudes and in terms of spatial extents of the events. Different heavy-tailed distributions have been proposed as candidates for both instances, but statistically rigorous treatments are rare. Here we combine the domains of event magnitudes and event area sizes by a spatiotemporal integration of 3-hourly rain rates corresponding to extreme events derived from the quasi-global high-resolution rainfall product Tropical Rainfall Measuring Mission 3B42. A maximum likelihood evaluation reveals that the distribution of spatiotemporally integrated extreme rainfall cluster sizes over the oceans is best described by a truncated power law, calling into question previous statements about scale-free distributions. The observed subpower law behavior of the distribution's tail is evaluated with a simple generative model, which indicates that the exponential truncation of an otherwise scale-free spatiotemporal cluster size distribution over the oceans could be explained by the existence of land masses on the globe.

1. Introduction

The spatial and temporal scaling behavior of convection and rainfall has attracted considerable attention in the physical and atmospheric sciences during the past decades. The spatial size distributions of single rainfall events and clouds have been thoroughly analyzed on the basis of various data sets (without considering temporal extents of the events), and these distributions are mostly assumed to be best approximated by a lognormal distribution [e.g., *López*, 1977; *Houze and Cheng*, 1977; *Cheng and Houze*, 1979; *Williams and Houze*, 1987]. However, recent studies [*Mapes and Houze*, 1993; *Nesbitt et al.*, 2006] have called this into question, and a power law-type behavior has been proposed as an alternative [e.g., *Cahalan and Joseph*, 1989; *Neggers et al.*, 2003]. Similarly, the scaling properties of rainfall event magnitudes without considering spatial and temporal extents of the events have been studied extensively [e.g., *Papalexiou and Koutsoyiannis*, 2013; *Serinaldi and Kilsby*, 2014], and the number of rainfall events as a function of the event magnitude has been found to exhibit a scale-free range over several orders of magnitude, hinting at similarities to nonequilibrium relaxation processes such as earthquakes or avalanches [*Peters et al.*, 2002; *Dickman*, 2003; *Peters et al.*, 2010]. Indeed, strong empirical evidence has been reported that rainfall might be a real-world example of self-organized criticality [*Andrade et al.*, 1998; *Peters and Neelin*, 2006, 2009].

However, existing studies investigating the spatial and temporal scaling behavior of rainfall do not perform statistically rigorous comparisons of the proposed power law to alternative heavy-tailed distributions, which may attain shapes that are very hard to distinguish from a true power law [*Clauset et al.*, 2009; *Virkar and Clauset*, 2014]. Whether the relative event magnitude or area size frequencies actually follow a power law distribution has therefore not been rigorously assessed to date. In fact, both the area size and the magnitude distributions have been suggested to exhibit tails which decay faster than that of a power law distribution fitted to the observed values [*Peters et al.*, 2012]. Additionally, the scaling characteristics of rainfall have been investigated either in the spatial domain or in the magnitude domain. For instance, *Peters et al.* [2012] analyzed the frequency distribution of instantaneous rainfall rates integrated over the spatial extents of the corresponding rainfall cluster—defined there as the set of connected pixels experiencing significant rainfall—but only

for single time slices, thus excluding the temporal extents of the events. Also for this quantity, the tail of the frequency distribution decays faster than that of a corresponding power law distribution [*Peters et al.*, 2012].

Here we analyze the frequency distribution of the total water volume precipitated in spatiotemporally extended extreme rainfall events. The distributional characteristics of the spatiotemporally integrated water volumes, i.e., the total *cluster sizes*, have previously been proposed to follow a scale-free distribution even if this does not hold true for neither the area size, the single-site event magnitudes, nor the spatially integrated event magnitudes [*Peters et al.*, 2012]. We use the satellite-derived, gauge-calibrated rainfall data set Tropical Rainfall Measuring Mission (TRMM) 3B42 (V7), available at 3-hourly temporal resolution on a regular 0.25° grid covering Earth's surface from 50°N to 50°S and study the spatiotemporal cluster sizes with respect to possible differences over the global oceans and land masses. This is further motivated by the fact that the largest rainfall events at the Earth's surface are thunderstorms in the form of mesoscale convective systems and hurricanes (typhoons over the NW Pacific) [*Maddox*, 1980; *Goldenberg*, 2001; *Zipser et al.*, 2006], which are — in addition to their spatial sizes — characterized by their outstanding temporal persistence.

We use maximum likelihood estimation (MLE) and maximum likelihood ratio (MLR) comparison tests between several plausible heavy-tailed candidate distributions and find that the total cluster size distributions over the oceans (land masses) are best described by an exponentially truncated power law (stretched exponential) although they appear to be scale-free over several orders of magnitude. With the help of a simple generative model, we propose the existence of land masses as a possible explanation of the subpower law behavior of these distributions.

2. Data

We employ the Tropical Rainfall Measuring Mission (TRMM) 3B42 V7 dataset [*Huffman et al.*, 2007], with 3-hourly temporal resolution for the time period from 1998 to 2014. The data set is spatially gridded at a resolution of 0.25° × 0.25° ranging from 50°S to 50°N. Each of the $N = 46.752 \cdot 1440 \cdot 400 \approx 2.69 \cdot 10^{10}$ data point consists of the time of the measurement t_i, the geographical location given by a tuple of coordinates $(\mathrm{lon}_i, \mathrm{lat}_i)$, and the average rainfall rate during a 3 h time window r_i.

We extract extreme rainfall events from the data by considering only those measurements above the 90th percentile of so-called *wet times* (defined as data points with rainfall rates $r \geq 0.1$ mm/h). This wet-time threshold is employed to assure that only data points with significant rainfall are used to compute the distributional characteristics we are interested in here [e.g., *Huffman et al.*, 2007; *Scheel et al.*, 2011; *Chen et al.*, 2013; *Zulkafli et al.*, 2014]. The 90th percentile is chosen in agreement with the definition of extreme rainfall events in the Intergovernmental Panel on Climate Change report [*Intergovernmental Panel on Climate Change*, 2012] (see Figure 1a for the threshold values at each geographical location). This results in $n \approx 2.16 \times 10^8$ extreme events, which we partition into spatiotemporal clusters as described in section 3.

3. Methods

We first specify our concept of spatiotemporal clusters and their sizes (section 3.1). Thereafter, we introduce the different candidate distributions and the elements of Bayesian parameter inference used to fit candidate distributions to the observed histograms of cluster sizes (section 3.2). Finally, we propose a minimal generative model which reproduces the observed truncated power law behavior for the cluster size distribution over the global oceans (section 3.3).

3.1. Spatiotemporal Clusters and Their Sizes

A spatiotemporal cluster of extreme rainfall is defined as the union of nearest neighbors of extreme rainfall events in the discrete space-time grid prescribed by the resolutions of the TRMM data set. In order to detect the clusters, we first enumerate the given longitude, latitude and time coordinates. Thereby, we can associate every event with discrete space-time coordinates, $(\mathrm{lon}_i, \mathrm{lat}_i, t_i) \leftrightarrow (x_i, y_i, t_i)$. Every event can have up to 26 possible neighbors (eight neighbors in the time slice of the measurement, t_i, and nine neighbors in each of the time slices $t_i - 1$ and $t_i + 1$). Having linked all pairs of neighboring extreme rainfall events, we find the clusters by identifying them as the connected components of the graph $G = (V, E)$, where the set of nodes V is given by the events and the set of edges E by the links between pairs of neighboring events (see Figure 1b for an illustration of this definition and *Traxl et al.* [2016] for details of the graph methodology used to determine the spatiotemporal clusters). We find a total of $n^C \approx 1.42 \times 10^7$ spatiotemporal clusters. For each of these

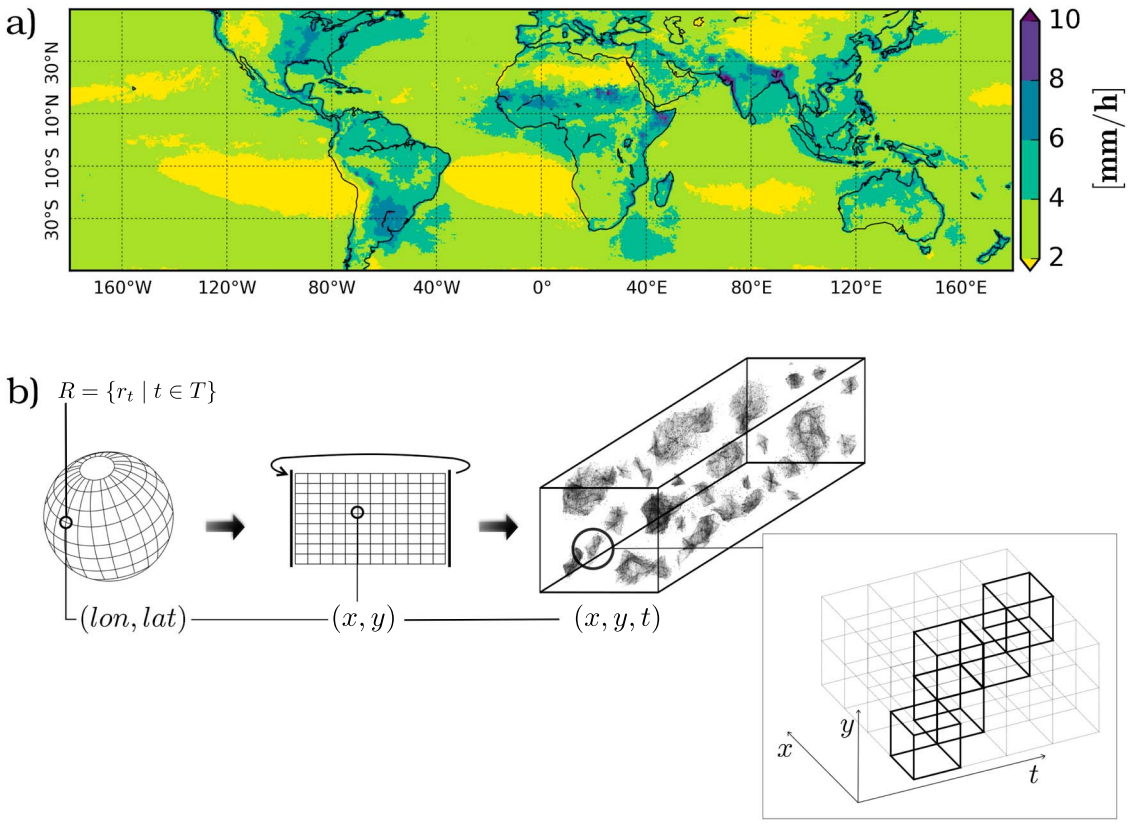

Figure 1. Extreme rainfall thresholds and sketch of spatiotemporal cluster detection. (a) The 90th percentile threshold values (in mm/h) for each geographical location. Only rainfall events with rainfall rates above these thresholds are considered in this study. (b) In order to partition the extreme rainfall events into clusters, we first enumerate the given longitude, latitude, and time coordinates, allowing us to associate every event with discrete space-time coordinates, $(\text{lon}_i, \text{lat}_i, t_i) \leftrightarrow (x_i, y_i, t_i)$. Each event can have up to 26 possible neighbors (eight neighbors in the time slice of the measurement, t_i, and nine neighbors in each of the time slices $t_i - 1$ and $t_i + 1$). Linking all pairs of neighboring extreme rainfall events, we find the clusters by identifying them as the connected components of the graph $G = (V, E)$, where the set of nodes V is comprised of the events themselves and the set of edges E indicates whether events are neighbors on the space-time grid. An exemplary cluster is depicted on the very right in Figure 1b.

clusters, we compute the total volume of water precipitated (of dimension: length^3), i.e., the cluster size s of cluster C, by

$$s = \sum_C a_i \cdot r_i \cdot 3h, \tag{1}$$

where a_i is the surface area of a rainfall event, given by

$$a_i = (111\text{km})^2 \cdot (0.25)^2 \cdot \cos\left(\frac{2\pi}{360°} \cdot \text{lat}_i\right). \tag{2}$$

This definition of cluster size follows the definition of "event size" in *Peters et al.* [2012]. Note that an alternative choice of defining the size of clusters would be to simply take the number of events belonging to a cluster. In fact, for the clusters obtained as described above, the total volume of precipitated water is highly correlated to the number of events ($r_{\text{Pearson}} = 0.98$). However, we chose the total volume of water as the metric for cluster sizes, since it is more accurate.

The set of all clusters is then partitioned into three groups: clusters that precipitated mainly above ocean, mainly above land, and a group for the remaining clusters. For this, we first use a land-ocean mask to classify every rainfall event as an ocean or a land event. We then decide that a cluster belongs to the ocean (land) group if more than 90% of its constituent events are ocean (land) events, leading to $n^C_{\text{ocean}} \approx 8.68 \times 10^6$ ($n^C_{\text{land}} \approx 5.20 \times 10^6$) clusters. The remaining cluster are attributed to the mixed group, with $n^C_{\text{mixed}} \approx 3.55 \times 10^5$ elements. Note that the results presented below are insensitive to changing this parameter, e.g., from 90% to 100%.

3.2. Estimation of the Distributions of Spatiotemporal Extreme Rainfall Clusters

For all clusters combined, the ocean and land groups separated, and the generative model introduced in the next section, we fit a set of candidate distributions to the cluster size distributions. These candidates include a power law (PL), a truncated power law (TPL), a stretched exponential (SEXP), and a lognormal (LN), given by

$$\text{PL: } f(x) = \frac{(\alpha - 1)}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-\alpha} \tag{3}$$

$$\text{TPL: } f(x) = \frac{\lambda^{1-\alpha}}{\Gamma\left(1 - \alpha, \lambda x_{\min}\right)} x^{-\alpha} e^{-\lambda x} \tag{4}$$

$$\text{SEXP: } f(x) = \beta \lambda x^{\beta - 1} e^{-\lambda \left(x^{\beta} - x_{\min}^{\beta} \right)} \tag{5}$$

$$\text{LN: } f(x) = \sqrt{\frac{2}{\pi \sigma^2}} \left[\operatorname{erfc}\left(\frac{\ln x_{\min} - \mu}{\sqrt{2}\sigma} \right) \right]^{-1} \frac{1}{x} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2} \right) \tag{6}$$

In a first step, optimal functional forms for these distributions with respect to the observed cluster sizes are determined by MLE [*Clauset et al.*, 2009]: For each proposed candidate ρ, the likelihood of its parameters $\mathcal{P}$, given the set of observed cluster size values $\mathcal{C}$ is maximized. Under the assumption of flat priors $P(\rho, \mathcal{P})$, Bayes' Theorem assures that the parameters determined via this optimization are the most likely parameters given the observed data $\mathcal{C}$:

$$P(\rho, \mathcal{P}|\mathcal{C}) = \frac{P(\mathcal{C}|\rho, \mathcal{P})\, P(\rho, \mathcal{P})}{P(\mathcal{C})}\ , \tag{7}$$

where $P(\mathcal{C})$ is unknown and in practical terms impossible to compute. The likelihood of the parameters $\mathcal{P}$ given the data $\mathcal{C}$ is defined as $\mathcal{L}_{\mathcal{C}}(\rho, \mathcal{P}) \equiv P(\mathcal{C}|\rho, \mathcal{P})$, and assuming that $P(\rho, \mathcal{P})$ is noninformative (i.e., a flat distribution), we have $P(\rho, \mathcal{P}|\mathcal{C}) \propto \mathcal{L}_{\mathcal{C}}(\rho, \mathcal{P})$. We are thus left with the following optimization problem:

$$\mathcal{P}^* = \arg\max_{\mathcal{P}} \mathcal{L}_{\mathcal{C}}(\rho, \mathcal{P}) = \arg\max_{\mathcal{P}} \prod_{i=1}^{n^{\mathcal{C}}} \rho(s_i; \mathcal{P})\ , \tag{8}$$

where $n^{\mathcal{C}}$ denotes the number of clusters and s_i their respective spatiotemporally integrated size. The optimal parameters thus determined for the four candidate distributions are listed in the legends of Figures 2a–2c.

The likelihood of each candidate, evaluated with the respective MLE-optimal parameters, is then compared by means of a MLR comparison test. The Neyman-Pearson lemma assures that this is the most efficient statistical test possible to compare between two candidate distributions [*Neyman and Pearson*, 1933]. Setting $L_{\mathcal{C}}(\rho) = \max_{\mathcal{P}} \mathcal{L}_{\mathcal{C}}(\rho, \mathcal{P})$, we compute for two candidates ρ_1 and ρ_2 the log ratio

$$\mathcal{R}_{\mathcal{C}}(\rho_1, \rho_2) = \log \frac{L_{\mathcal{C}}(\rho_1)}{L_{\mathcal{C}}(\rho_2)}\ . \tag{9}$$

If $R_{\mathcal{C}}(\rho_1, \rho_2) > 0$ (< 0), we conclude that ρ_1 is a more (less) likely model of the observed cluster size distribution than ρ_2. A test of statistical significance for the values of $\mathcal{R}$ can be derived from the central limit theorem (see *Clauset et al.* [2009] for details, in particular for cases where the two distributions to be compared are nested versions of each other).

3.3. Generative Model for Spatiotemporal Cluster Sizes

Here we introduce a generative model to test the hypothesis that the subpower law behavior of observed spatiotemporal extreme rainfall cluster size distributions is due to the existence of land masses on Earth. The model is motivated by the assumption of a scale-free (i.e., PL) distribution of rainfall cluster sizes, whose truncation is caused by the fact that hurricanes end prematurely as soon as they hit the coast. The model is designed as follows: A synthetical storm with a lifetime drawn from a PL distribution is placed on a random pixel of a cage consisting of 1000×1000 pixels. The temporal evolution of the storm is prescribed by randomly selecting a neighboring pixel at each time step, where the selection probabilities depend on the previous direction of movement. Moving in the same direction as before has a higher probability ($p = 0.95$) than turning left

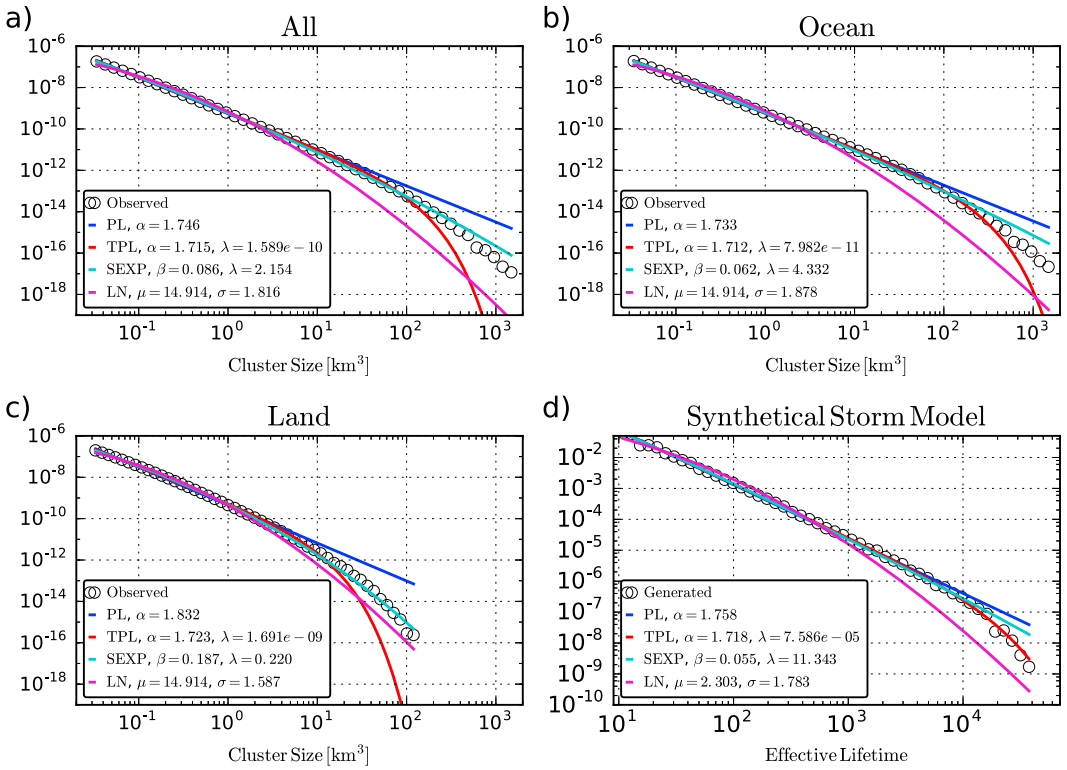

Figure 2. Histograms of the observed spatiotemporal extreme rainfall cluster sizes and the effective lifetimes of the generative model introduced in section 3.3 (circles), as well as the corresponding MLE-optimized fits of the proposed candidate distributions (lines). In all panels, blue lines indicate the optimal power law (PL) fits, red lines the truncated power law (TPL) fits, cyan lines the stretched exponential (SEXP) fits, and magenta lines the lognormal (LN) fits. The respective optimal parameters of the MLE fits are stated in the legends. (a) Histogram of observed cluster sizes for all clusters combined ($n^C \approx 1.42 \cdot 10^7$) and MLE-fitted cluster size distributions. (b) Histogram of observed cluster sizes for the subset of ocean clusters ($n^C_{\text{ocean}} \approx 8.68 \cdot 10^6$) and MLE-fitted cluster size distributions for this subset. (c) Histogram of observed cluster sizes for the subset of land clusters ($n^C_{\text{land}} \approx 5.20 \cdot 10^6$), and MLE-fitted cluster size distributions for this subset. (d) Histogram of effective lifetimes obtained from the generative model introduced in section 3.3 and corresponding MLE-fitted lifetime distributions. Note that the units are arbitrary in this case.

($p = 0.04$) or right ($p = 0.01$), and the probability to move backward is zero. The storm keeps moving either until its predetermined (PL-distributed) lifetime ends or until it hits the boundary of the cage, which immediately ends the lifetime of the storm. With each time step, the storm grows by one unit; hence, its "size" is proportional to its lifetime. The experiment is simulated with a large number of storms (10.000), for each of which we record its effective lifetime. The resulting distribution of lifetimes is shown in Figure 2d and discussed in section 4.

The inertia of the storm's movement scheme imposed by the dynamic selection probabilities ensures more realistic storm tracks: First, a storm is very unlikely to move backward (hence $p = 0$ for turning backward).

Table 1. MLR Test Results for the Comparisons Between All Considered Candidate Distributions (PL, TPL, SEXP, and LN) for the Observed Cluster Size Distributions ("All" Clusters, "Ocean" Clusters, and "Land" Clusters) and the Lifetime Distribution of the Synthetical Storm Model ("Model")[a]

	$R_C(\rho_{\text{TPL}}, \rho_{\text{PL}})$	$R_C(\rho_{\text{SEXP}}, \rho_{\text{PL}})$	$R_C(\rho_{\text{SEXP}}, \rho_{\text{TPL}})$	$R_C(\rho_{\text{LN}}, \rho_{\text{PL}})$	$R_C(\rho_{\text{LN}}, \rho_{\text{TPL}})$	$R_C(\rho_{\text{LN}}, \rho_{\text{SEXP}})$
All	14778	22252	7474	−154656	−169435	−176908
Ocean	5366	3751	−1615	−129157	−134523	−132908
Land	19865	24404	4539	−8909	−28774	−33312
Model	339	191	−148	−4038	−4377	−4229

[a]We recall from section 3.2 that a distribution ρ_1 is more (less) likely than another distribution ρ_2 if $R_C(\rho_1, \rho_2) > 0$ (< 0). All corresponding p values are smaller than 10^{-30}.

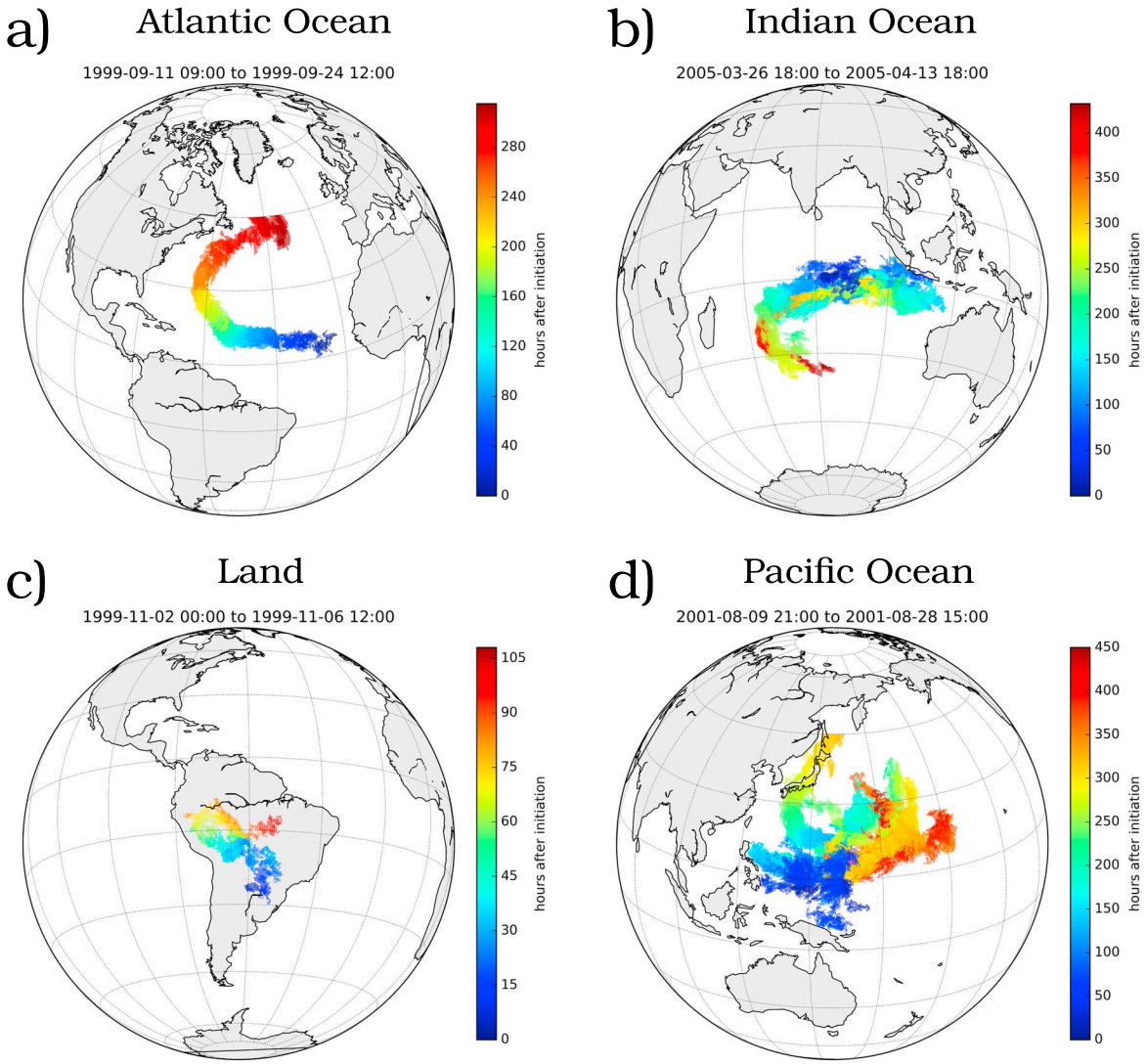

Figure 3. Largest clusters of the land and different ocean groups. For each panel, each colored grid cell has received at least one event above the 90th percentile belonging to the respective cluster. The colors indicate the last time (in units of hours) a given grid cell is hit by the cluster, relative to its initiation time (see titles of the respective panels). (a) The largest cluster that precipitated above the Atlantic ocean. (b) The largest cluster that precipitated above the Indian ocean. (c) The largest cluster that precipitated above land. (d) The largest cluster that precipitated above the Pacific ocean.

Second, the symmetry of moving left or right is broken in order to mimic the influence of the Coriolis force. However, we found that changing the selection probabilities or removing the inertia entirely ($p = 0.25$ for each direction) does not lead to qualitatively different lifetime distributions.

The assumptions behind this generative mechanism is that large storms such as hurricanes typically initiate over the ocean and that their lifetimes over the oceans would be PL distributed due to the abundant energy provided by the ocean. However, most large storms eventually hit the coast, where the lack of available energy for their persistence causes them to die out [*Whitaker and Davis*, 1994; *Briegel and Frank*, 1997; *Raymond and Sessions*, 2007; *Nolan et al.*, 2007]. The lifetimes can be approximately taken to be proportional to the sizes of the storms in terms of total precipitated water. Therefore, this model is suitable to investigate how the boundaries of the oceans (i.e., the coasts) would impact a cluster size distribution prescribed as a PL.

4. Results

The observed spatiotemporally integrated extreme rainfall cluster sizes for all clusters (Figure 2a) as well as for clusters over the global oceans (Figure 2b) extend beyond 10^3 km^3 of precipitated water. In contrast, the cluster size distribution over the global land masses only reaches sizes up to 10^2 km^3 (Figure 2c). Out of the

four distributions proposed as candidates for the cluster sizes (PL, TPL, SEXP, and LN), the SEXP wins the MLR comparison test (see section 3.2) for the combination of all clusters. The TPL wins for the subset of clusters over the oceans. Note that visually, the SEXP may appear to be the better model. This is, however, only caused by the fact that the log-log representation of the probability density functions emphasizes the tail of distributions rather than the comparably smaller cluster sizes, which have higher probability weights by several orders of magnitude. For clusters over the global land masses, the SEXP is again the most likely candidate distribution. For comparison, the generative model of synthetical storms, with PL-distributed lifetimes that are put into a cage with absorbing boundaries, leads to a TPL distribution of the effective lifetimes as the most likely candidate (Table 1).

A visual comparison of all four histograms with their respective optimal PL fits (a straight line in the log-log plots of Figure 2) may suggest a scale-free distribution at least over several orders of magnitude (approximately: for all clusters up to 20 km^3, for ocean clusters up to 40 km^3, for land clusters up to 3 km^3, and for the generative model up to 4000 units). However, the PL does not win the MLR comparison even when only considering cluster sizes and lifetimes below these values. Toward larger cluster sizes, all observed histograms clearly show dampened tails as compared to the PL, albeit to different extents.

To get a visual impression of clusters at the tail end of the distributions, we show the time evolution of the largest cluster observed over land, as well as the largest clusters over the Atlantic, Pacific, and Indian Ocean (Figure 3). The partitioning of extreme rainfall events into clusters of (spatiotemporally) connected neighboring pixels can lead to clusters with very clear trajectories (Figures 3a and 3c) and also to rather scattered clusters without a clear propagation direction (Figures 3b and 3d). Amongst the largest clusters over the oceans we found — by visual inspection — a large number of hurricanes (typhoons), such as the one over the Atlantic (Figure 3a). The largest land cluster in the data (Figure 3c) has been recently discussed by *Boers et al.* [2014, 2015] and *Traxl et al.* [2016].

5. Discussion

Our results indicate that none of the observed spatiotemporally integrated extreme rainfall cluster size distributions are described well by a PL. This is in contrast to propositions of scale-free distributions in the literature [*Cahalan and Joseph*, 1989; *Peters et al.*, 2002; *Dickman*, 2003; *Neggers et al.*, 2003] but also corroborates earlier results obtained for cluster sizes defined by either a temporal or a spatial integration of rain rates [*Peters et al.*, 2012]. Although the size distributions of rainfall clusters over the lands and oceans are visually similar to the PL fits over several orders of magnitude, both of them decay faster than the corresponding PL at the tails of the distributions. The MLR test results (SEXP for all clusters combined, TPL for ocean clusters, and SEXP for land clusters) suggest that the influence of the subsets of land and mixed clusters on the size distribution of all clusters combined leads to a SEXP also for all clusters. However, clusters larger than 10^2 km^3 occur almost exclusively over the oceans, and this is also the domain where the dampening of the distribution as compared to a corresponding PL becomes most apparent (Figure 2a).

The proposed generative model (section 3.3) tries to explain the observed behavior by postulating that the ocean cluster sizes would indeed follow a scale-free (i.e., PL) distribution on a planet without land masses and that the exponential truncation occurs primarily due to the fact that hurricanes end prematurely as soon as they hit the coast. Physically, this hypothesis can be motivated as follows: While the specific mechanisms of cyclogenesis are still not entirely understood, it is clear that cyclones can only maintain themselves over the oceans, under the conditions (among others) of sufficiently warm sea surface temperatures and strong, moist convection, which guarantee that enough latent heat can be released to the atmosphere to fuel the storms. Shortly after their landfall, the cyclones die because these conditions are no longer fulfilled [*Whitaker and Davis*, 1994; *Briegel and Frank*, 1997; *Raymond and Sessions*, 2007; *Nolan et al.*, 2007]. In the generative model, storms with a relatively longer lifetime have a higher probability of hitting a border of the cage at some point in their life span, which leads to the observed truncation of the effective lifetime distribution (Figure 2d). The analogy to the ocean cluster size distribution is that it would in fact also follow a power law, if there were no land masses on the planet. Hence, rainfall clusters could freely move, like the synthetical storms without a cage, on this imaginary aqua planet, with their size distribution predetermined by a power law. The fact that there are land masses on Earth would then lead to an exponential truncation of the PL distribution (Figure 2b), as it is the case for the synthetical storms of the generative model (Figure 2d).

We note that the proposed hypotheses correspond to a very simplified view of the complex physical processes involved in the formation and maintenance of cyclones. We do not propose that the only relevant physics behind the exponential truncation of extreme rainfall cluster size distributions are given by the stated hypotheses. It is, however, remarkable that the generative model is — despite its simplicity — capable of reproducing the observed statistical characteristics of rainfall cluster sizes over the oceans. From a purely statistical point of view, our hypothesis can hence not be rejected.

6. Conclusion

Based on the high-resolution TRMM 3B42 data set, we have provided statistically rigorous evidence that the spatiotemporally integrated size distribution of extreme rainfall clusters does not — as previously suggested — follow a power law. Instead, we find that the size distribution of rainfall clusters over the oceans (land masses) is best approximated by an exponentially truncated power law (stretched exponential), leading to a stretched exponential as the most likely candidate distribution for all clusters combined (land and ocean together). We hypothesize that the size distribution of extreme rainfall clusters over the oceans could, in principle, follow a scale-free distribution on a planet without land masses and that the exponential truncation of the observed distribution is caused by the presence of land masses. Physically, this is motivated by the fact that the conditions for cyclogenesis are not met over land. To test this hypothesis, we proposed a simple generative model of synthetic storms with power law distributed lifetimes, evolving in a finite spatial area with absorbing boundaries. This simple model reproduces the exponentially truncated power law observed for extreme rainfall clusters over the oceans, indicating that the proposed hypothesis suffices to explain the distributional characteristics discovered here.

Acknowledgments
This paper was developed within the scope of the IRTG 1740/TRP 2011/50151-0, funded by the DFG/FAPESP. N.B. acknowledges funding by the Alexander von Humboldt Foundation and the German Federal Ministry for Education and Research. We thank Bodo Bookhagen for the very helpful comments and discussions. The data employed in this study (TRMM 3B42 V7) can be retrieved at http://mirador.gsfc.nasa.gov/.

References

Andrade, R., H. Schellnhuber, and M. Claussen (1998), Analysis of rainfall records: Possible relation to self-organized criticality, *Physica A*, *254*(3–4), 557–568, doi:10.1016/S0378-4371(98)00057-0.

Boers, N., B. Bookhagen, H. M. J. Barbosa, N. Marwan, J. Kurths, and J. Marengo (2014), Prediction of extreme floods in the Eastern Central Andes based on a complex network approach, *Nat. Commun.*, *5*, 5199, doi:10.1038/ncomms6199.

Boers, N., H. M. J. Barbosa, B. Bookhagen, J. A. Marengo, N. Marwan, and J. Kurths (2015), Propagation of strong rainfall events from Southeastern South America to the Central Andes, *J. Clim.*, *28*(19), 7641–7658, doi:10.1175/JCLI-D-15-0137.1.

Briegel, L. M., and W. M. Frank (1997), Large-scale influences on tropical cyclogenesis in the Western North Pacific, *Mon. Weather Rev.*, *125*, 1397–1413, doi:10.1175/1520-0493(1997)125<1397:LSIOTC>2.0.CO;2.

Cahalan, R. F., and J. H. Joseph (1989), Fractal statistics of cloud fields, *Mon. Weather Rev.*, *117*, 261, doi:10.1175/1520-0493(1989)117<0261:FSOCF>2.0.CO;2.

Chen, S., et al. (2013), Evaluation of the successive V6 and V7 TRMM multisatellite precipitation analysis over the Continental United States, *Water Resour. Res.*, *49*, 8174–8186, doi:10.1002/2012WR012795.

Cheng, C.-P., and R. A. Houze Jr. (1979), The distribution of convective and mesoscale precipitation in GATE radar echo patterns, *Mon. Weather Rev.*, *107*(10), 1370–1381, doi:10.1175/1520-0493(1979)107<1370:TDOCAM>2.0.CO;2.

Clauset, A., C. Shalizi, and M. Newman (2009), Power-law distributions in empirical data, *SIAM Rev.*, *51*(4), 661–703.

Dickman, R. (2003), Rain, power laws, and advection, *Phys. Rev. Lett.*, *90*(10), 108701, doi:10.1103/PhysRevLett.90.108701.

Goldenberg, S. B. (2001), The recent increase in Atlantic Hurricane activity: Causes and implications, *Science*, *293*(5529), 474–479, doi:10.1126/science.1060040.

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. Weather Rev.*, *105*, 964–980, doi:10.1175/1520-0493(1977)105<0964:RCOTCO>2.0.CO;2.

Huffman, G., D. Bolvin, E. Nelkin, D. Wolff, R. Adler, G. Gu, Y. Hong, K. Bowman, and E. Stocker (2007), The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, *J. Hydrometeorol.*, *8*(1), 38–55, doi:10.1175/JHM560.1.

Intergovernmental Panel on Climate Change (2012), Glossary of terms, in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation - A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC)*, edited by C. B. Field et al., pp. 555–564, Cambridge Univ. Press, Cambridge, U K., and New York.

López, R. E. (1977), The lognormal distribution and cumulus cloud populations, *Mon. Weather Rev.*, *105*, 865–872, doi:10.1175/1520-0493(1977)105<0865:TLDACC>2.0.CO;2.

Maddox, R. A. (1980), Mesoscale convective complexes, *Bull. Am. Meteorol. Soc.*, *61*(11), 1374–1387.

Mapes, B. E., and R. A. Houze Jr. (1993), Cloud clusters and superclusters over the oceanic warm pool, *Mon. Weather Rev.*, *121*, 1398–1415, doi:10.1175/1520-0493(1993)121<1398:CCASOT>2.0.CO;2.

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*(8), 1060–1074, doi:10.1175/1520-0469(2003)60<1060:SSOCCP>2.0.CO;2.

Nesbitt, S. W., R. Cifelli, and S. A. Rutledge (2006), Storm morphology and rainfall characteristics of TRMM precipitation features, *Mon. Weather Rev.*, *134*(10), 2702–2721, doi:10.1175/MWR3200.1.

Neyman, J., and E. S. Pearson (1933), On the problem of the most efficient tests of statistical hypotheses, *Philos. Trans. R. Soc. A*, *231*, 289–337, doi:10.1098/rsta.1933.0009.

Nolan, D. S., E. D. Rappin, and K. A. Emanuel (2007), Tropical cyclogenesis sensitivity to environmental parameters in radiative-convective equilibrium, *Q. J. R. Meteorol. Soc.*, *133*(629), 2085–2107, doi:10.1002/qj.170.

Papalexiou, S. M., and D. Koutsoyiannis (2013), Battle of extreme value distributions: A global survey on extreme daily rainfall, *Water Resour. Res.*, *49*, 187–201, doi:10.1029/2012WR012557.

Peters, O., and J. D. Neelin (2006), Critical phenomena in atmospheric precipitation, *Nat. Phys.*, *2*, 393–396, doi:10.1038/nphys314.
Peters, O., and J. D. Neelin (2009), Atmospheric convection as a continuous phase transition: Further evidence, *Int. J. Mod. Phys. B*, *23*(28–29), 5453–5465, doi:10.1142/S0217979209063778.
Peters, O., C. Hertlein, and K. Christensen (2002), A complexity view of rainfall, *Phys. Rev. Lett.*, *88*(1), 018701, doi:10.1103/PhysRevLett.88.018701.
Peters, O., A. Deluca, A. Corral, J. D. Neelin, and C. E. Holloway (2010), Universality of rain event size distributions, *J. Stat. Mech: Theory Exp.*, *2010*, P11030, doi:10.1088/1742-5468/2010/11/P11030.
Peters, O., K. Christensen, and J. D. Neelin (2012), Rainfall and dragon-kings, *Eur. Phys. Spec. Top.*, *205*(1), 147–158, doi:10.1140/epjst/e2012-01567-5.
Raymond, D. J., and S. L. Sessions (2007), Evolution of convection during tropical cyclogenesis, *Geophys. Res. Lett.*, *34*, L06811, doi:10.1029/2006GL028607.
Scheel, M. L. M., M. Rohrer, C. Huggel, D. Santos Villar, E. Silvestre, and G. J. Huffman (2011), Evaluation of TRMM Multi-satellite Precipitation Analysis (TMPA) performance in the Central Andes region and its dependency on spatial and temporal resolution, *Hydrol. Earth Syst. Sci.*, *15*(8), 2649–2663, doi:10.5194/hess-15-2649-2011.
Serinaldi, F., and C. G. Kilsby (2014), Rainfall extremes: Toward reconciliation after the battle of distributions, *Water Resour. Res.*, *50*(1), 336–352, doi:10.1002/2013WR014211.
Traxl, D., N. Boers, and J. Kurths (2016), Deep graphs: A general framework to represent and analyze heterogeneous complex systems across scales, *Chaos*, *26*, 065303.
Virkar, Y., and A. Clauset (2014), Power-law distributions in binned empirical data, *Ann. Appl. Stat.*, *8*(1), 89–119, doi:10.1214/13-AOAS710.
Whitaker, J., and C. Davis (1994), Cyclogenesis in a saturated environment, *J. Atmos. Sci.*, *51*, 889–908, doi:10.1175/1520-0469(1994)051<0889:CIASE>2.0.CO;2.
Williams, M., and R. A. Houze Jr. (1987), Satellite-observed characteristics of winter monsoon cloud clusters, *Mon. Weather Rev.*, *115*, 505–519, doi:10.1175/1520-0493(1987)115<0505:SOCOWM>2.0.CO;2.
Zipser, E. J., D. J. Cecil, C. Liu, S. W. Nesbitt, and D. P. Yorty (2006), Where are the most intense thunderstorms on Earth?, *Bull. Am. Meteorol. Soc.*, *87*(8), 1057–1071, doi:10.1175/BAMS-87-8-1057.
Zulkafli, Z., W. Buytaert, C. Onof, B. Manz, E. Tarnavsky, W. Lavado, and J.-L. Guyot (2014), A comparative performance analysis of TRMM 3B42 (TMPA) versions 6 and 7 for hydrological applications over Andean-Amazon river basins, *J. Hydrometeorol.*, *15*(2), 581–592, doi:10.1175/JHM-D-13-094.1.