nature climate change

SUPPLEMENTARY INFORMATION

DOI: 10.1038/NCLIMATE1907

Changes in rainfall seasonality in the tropics

Xue Feng[1], Amilcare Porporato[1,2] *, and Ignacio Rodriguez-Iturbe[3]

Supplementary information

[1] Department of Civil and Environmental Engineering, Duke University, North Carolina, USA

[2] Nicholas School of the Environment and Earth Sciences, Duke University, North Carolina, USA

[3] Department of Civil and Environmental Engineering, Princeton University, New Jersey, USA

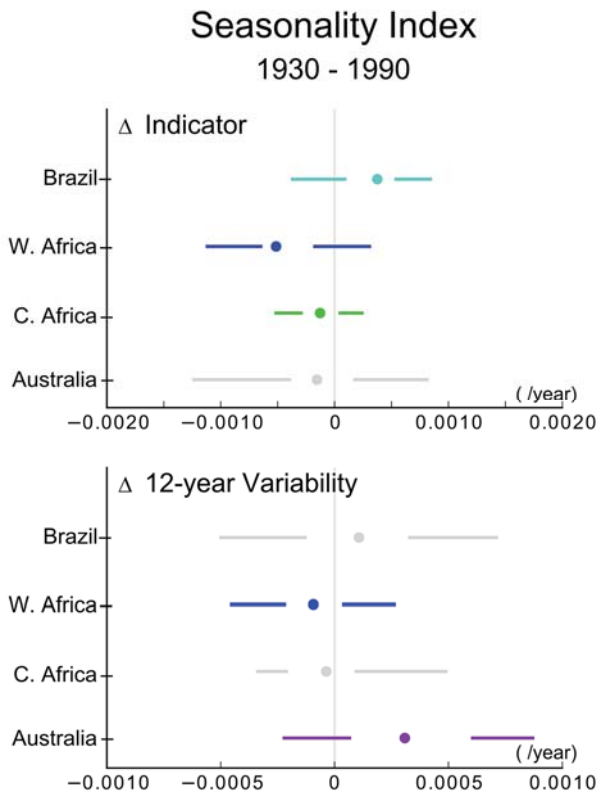
* Corresponding author: amilcare.porporato@duke.edu, (919) 660-5511

1. Additional analyses of the seasonality index

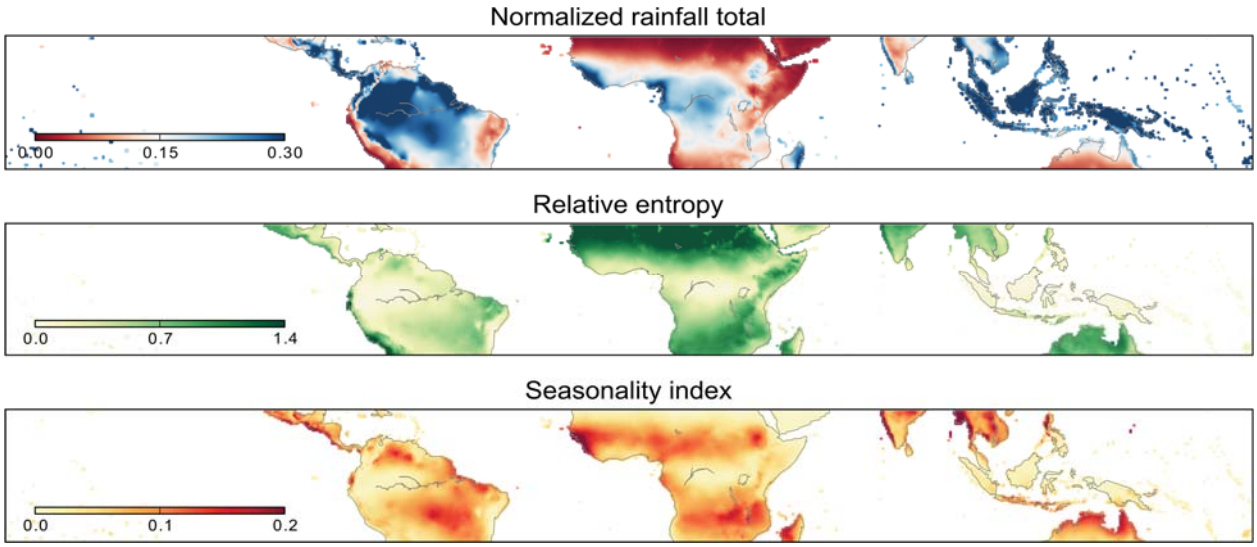
This section presents additional analyses of the seasonality index and its trends over the past century using monthly rainfall data from stations in highly seasonal regions (Supplementary Figure S1) as well as monthly gridded data over the entire tropics (Supplementary Figures S2 and S3).

Supplementary Table S1: Characteristics and references of the 2 rainfall datasets used in the text.

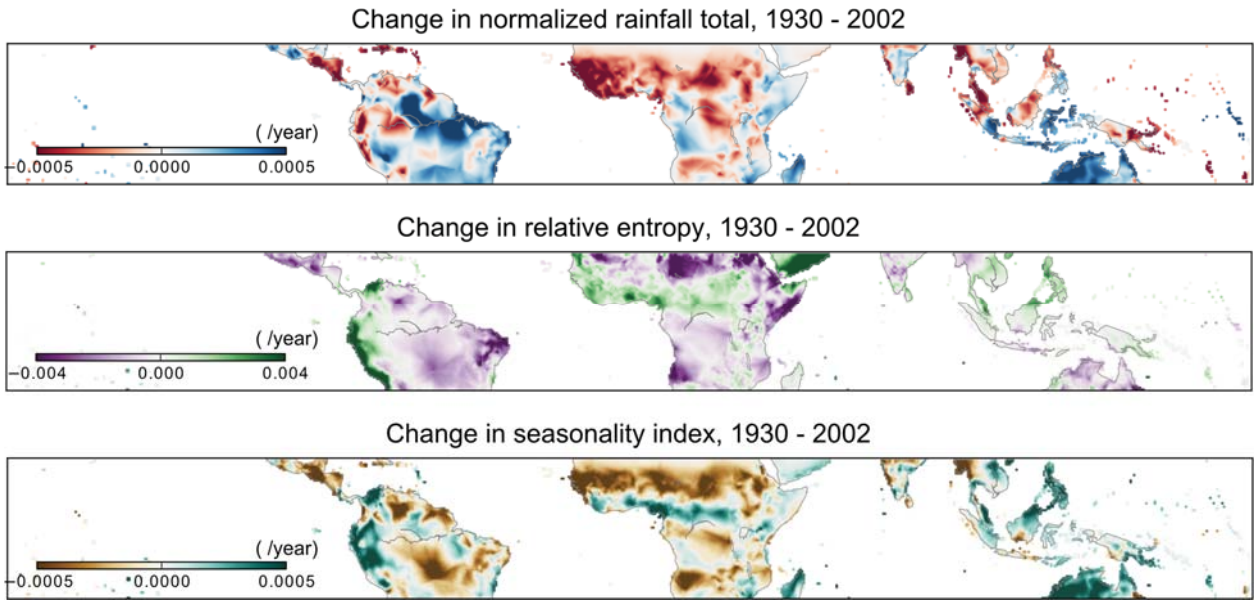
Series	Period of Record	Gauge/ Gridded	Reference	Used for
GHCN-Monthly	Varies	Gauge	Vose et al. (2002)[1]	Figures 1, 3, 4, S1, S4, S5 Tables S1-S9
CRU TS 2.1	1901-2002	Gridded	Mitchell et al. (2005)[2]	Figures 2, S2, S3 Table S1



Supplementary Figure S1: Trends in the annual seasonality index (top) and its variability (bottom) for each region over 1930-1990. The ensembles contain stations from the GHCN-Monthly dataset used to produce Figure 4 in the text. See Methods for details on trend and variability analysis.



Supplementary Figure S2: Normalized rainfall total ($\bar{R}/\bar{R}_{\max}$), relative entropy ($\bar{D}$), and seasonality index ($\bar{S}$) over the tropics (20°S to 20°N). Produced with the CRU TS 2.1 gridded dataset averaged over 1930-2002.



Supplementary Figure S3: Trends in the normalized rainfall total ($\bar{R}/\bar{R}_{\max}$), relative entropy ($\bar{D}$), and seasonality index ($\bar{S}$) over the tropics (20°S to 20°N). The trends are calculated over 1930-2002 using the CRU TS 2.1 gridded dataset.

2. Mean, extreme, and p-values of the boxplot ensembles (for annual rainfall, centroid, and spread)

The mean, the maximum/minimum, and the p-values of the boxplot ensembles are tabulated here for the annual rainfall, the centroid, and the spread as compiled in Figure 4 of the text.

Supplementary Table S2: Mean values of the boxplot ensembles (from Figure 4) for the changes in the indicators (top 5 rows) and their 12-year variability (bottom 5 rows) for each region, in the annual rainfall (magnitude), the centroid (timing), and the spread (duration).

		Magnitude (mm/year)	**Timing** (days/year)	**Duration** (days/year)
Δ Indicators	Northeast Brazil	6.7	0.045	0.092
	Western Africa	-4.3	-0.015	-0.034
	Central Africa	0.12	-0.0095	0.037
	Northern Australia	1.2	-0.0081	0.026
		Magnitude (mm/year)	**Timing** (days/year)	**Duration** (days/year)
Δ 12-year variability	Northeast Brazil	3.0	0.034	0.0035
	Western Africa	-0.61	0.010	0.020
	Central Africa	0.32	0.033	-0.010
	Northern Australia	3.2	-0.011	-0.037

Supplementary Table S3: Minimum and maximum values of the boxplot ensembles (from Figure 4) for the changes in the indicators (top 5 rows) and their 12-year variability (bottom 5 rows) for each region. The indicators are the annual rainfall (magnitude), the centroid (timing), and the spread (duration).

		Magnitude (mm/year)	**Timing** (days/year)	**Duration** (days/year)
Δ Indicators	Northeast Brazil	0.52 — 22	-0.026 — 0.13	-0.097 — 0.27
	Western Africa	-18 — 0.069	-0.069 — 0.035	-0.14 — 0.073
	Central Africa	-4.3 — 3.1	-0.082 — 0.051	-0.036 — 0.11
	Northern Australia	-4.1 — 6.4	-0.16 — 0.070	-0.14 — 0.23
		Magnitude (mm/year)	**Timing** (days/year)	**Duration** (days/year)
Δ 12-year variability	Northeast Brazil	-0.81 — 9.1	-0.17 — 0.13	-0.15 — 0.13
	Western Africa	-4.2 — 5.7	-0.065 — 0.069	-0.075 — 0.091
	Central Africa	-2.4 — 2.8	-0.027 — 0.14	-0.094 — 0.054
	Northern Australia	-1.6 — 11	-0.098 — 0.12	-0.16 — 0.099

Supplementary Table S4: P-values of the boxplot ensembles (from Figure 4) for the changes in the indicators (top 5 rows) and their 12-year variability (bottom 5 rows) using the Wilcoxon signed-rank test with a null hypothesis that the mean value of the ensemble is zero. The p-values are associated with the ensembles of the annual rainfall (magnitude), centroid (timing), and spread (duration).

		Magnitude	Timing	Duration
Δ Indicators	Northeast Brazil	1.2 E-06	4.1 E-06	2.1 E-04
	Western Africa	5.5 E-10	1.8 E-04	2.9 E-06
	Central Africa	4.0 E-01	2.7 E-02	1.9 E-05
	Northern Australia	4.4 E-03	7.2 E-01	7.3 E-02
		Magnitude	**Timing**	**Duration**
Δ 12-year variability	Northeast Brazil	2.6 E-06	5.7 E-03	7.8 E-01
	Western Africa	9.3 E-05	3.7 E-02	1.8 E-04
	Central Africa	1.9 E-01	7.4 E-05	1.3 E-01
	Northern Australia	2.0 E-07	1.3 E-01	1.8 E-03

3. Decomposition of seasonality through demodulation and entropic spread

This section introduces a complementary set of indicators (the demodulated amplitude, the demodulated phase, and the entropic spread) which can be used in conjunction with those introduced in the text (the annual rainfall, the centroid, and the spread) to describe the magnitude, timing, and duration of the annual seasonality. As can be seen in Supplementary Figure S4, these indicators corroborate the results already presented in Figure 4.

3.1 Entropic spread (duration)

The entropic spread E_k is a measure based on information theory for the support[3] of the monthly rainfall distribution in each year, $r_{k,m}$ (for each hydrological year k and month m). Thus, it is a measure of the duration of the rainy season and is defined as follows: the information entropy[3] is first calculated for each year from the *observed* monthly rainfall distribution $p_{k,m}$ as $H_k = -\sum_{m=1}^{12} p_{k,m} \log_2 p_{k,m}$. In the limiting case when rainfall is evenly distributed year round, the information entropy is $\log_2 n$, where $n = 12$ is the number of possible values (months) in the annual rainfall distribution. We can then calculate the *effective* number of values for our observed distribution using ${n_k}' = 2^{H_k}$, which will be less than 12 unless the distribution is perfectly uniform. By substituting ${n_k}'$ in place of n into the formula for the variance of the discrete uniform distribution, $\frac{n^2-1}{12}$, and taking its square root, the entropic spread for each year k is then derived as

$$E_k = \sqrt{\frac{2^{2H_k}-1}{12}}.$$

3.2 Demodulation amplitude (magnitude) and phase (timing)

Demodulation[4,5] is a localized harmonic analysis in which the amplitude and phase of a chosen frequency component of the spectral representation (e.g., the annual cycle) are estimated at every time step (e.g., month) of the demodulated time series. It is especially useful when a time series contains a periodic component (as in our case) whose amplitude and phase are slowly changing over time.

Demodulation consists of multiplying the time series by 2 sinusoidal functions and then applying a low pass filter to isolate the low frequencies from high frequency fluctuations. Given a signal x_t (for example, rainfall, with subscript t denoting its time dependency) comprised of a mean that varies over long term, $\bar{x}_t$, a periodic component (which we assume to be the seasonal signal varying at the annual scale ω_0) with slow changing amplitude A_t and phase θ_t, and a high frequency noise term ε_t (e.g., $x_t = \bar{x}_t + A_t \cos[\omega_0 t + \theta_t] + \varepsilon_t$), we first write x_t into its equivalent complex formulation, then multiply it by a complex exponential to cancel its known periodic frequency (ω_0), and call the result y_t:

$$x_t \cdot e^{-i\omega_0 t} \rightarrow y_t = \bar{x}_t\, e^{-i\omega_0 t} + \frac{1}{2} A_t \left\{ e^{i\,\theta_t} + e^{-i\left(2\,\omega_0 t + \theta_t\right)} \right\} + \varepsilon_t e^{-i\omega_0 t}.$$

We then extract the second, low frequency term, $\frac{1}{2}A_t e^{i\,\theta_t}$, using a low pass filter F over y_t.

$$F[y_t] \approx \frac{1}{2}A_t e^{i\,\theta_t} = \frac{1}{2}A_t\cos[\theta_t] + i\,\frac{1}{2}A_t\sin[\theta_t].$$

Since this term is made up of the amplitude $A_t$and phase θ_t of the periodic component, we can designate $u_t = \frac{1}{2}A_t\cos[\theta_t]$ and $v_t = \frac{1}{2}A_t\sin[\theta_t]$, and extract the amplitude and phase from its real and imaginary parts using the following formulas:

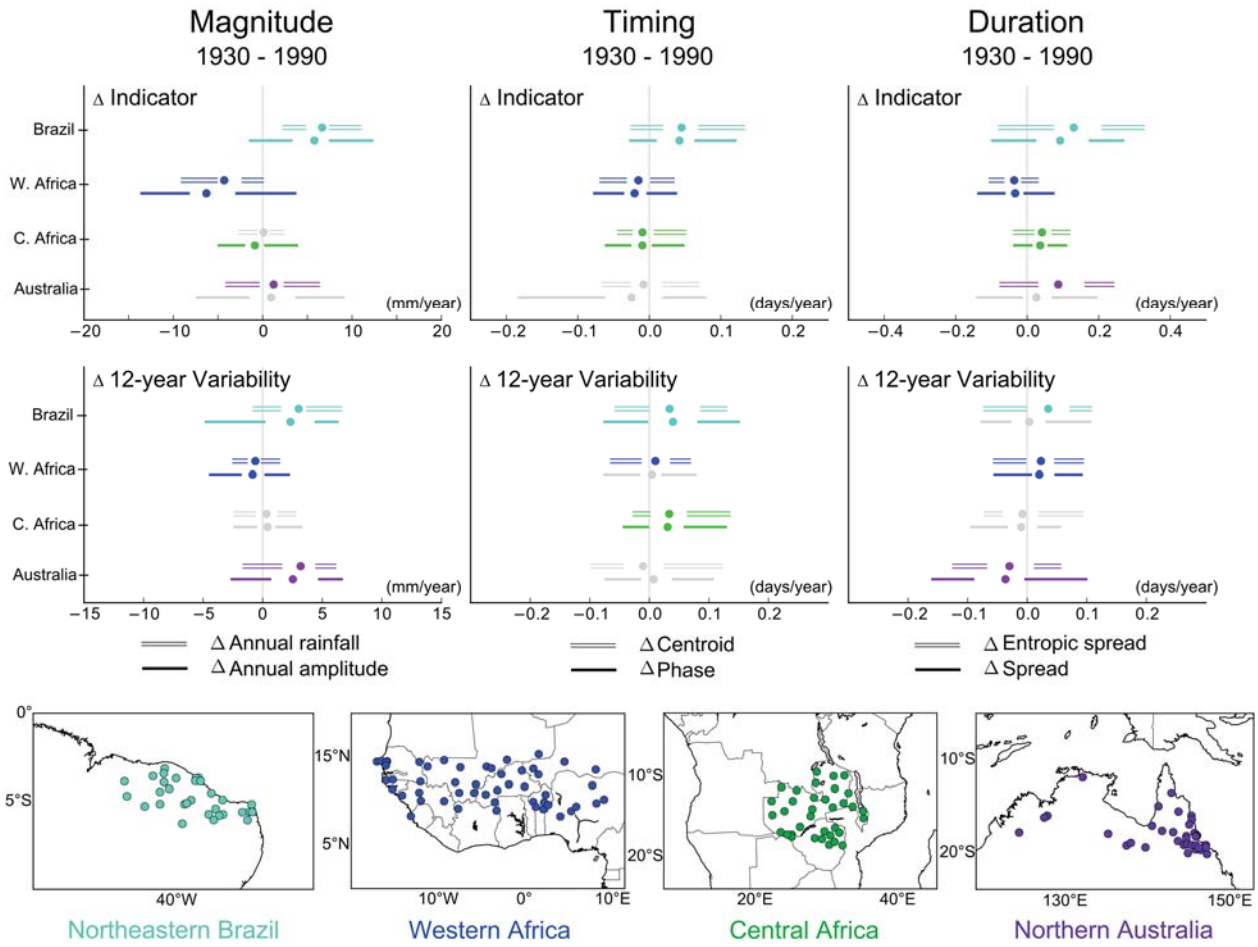
$$A_t = 2\sqrt{{u_t}^2 + {v_t}^2} \qquad \theta_t = \arctan\left[\frac{v_t}{u_t}\right].$$

For the low pass filter F we chose the Hanning window

$$F[y_t] = \frac{1}{2p}\left(\sum_{j=-p+1}^{p-1} y_{t+j} + \frac{1}{2}y_{t-p} + \frac{1}{2}y_{t+p}\right)$$

to be applied twice over y_t. We chose the length of the moving average $2p$ to be 12 for both the first and second application.

To be compatible in length with the other indicator series, the monthly time series for the demodulated amplitude is summed over each hydrological year *k* and sampled annually at the beginning of the year, resulting in the annual amplitude A_k. Likewise, the demodulated phase is sampled annually, resulting in θ_k.



Supplementary Figure S4: Changes in rainfall magnitude, timing, and duration in seasonality hotspots using 2 indicator series for each dimension of rainfall seasonality. The same analysis used to produce Figure 4 is carried out for each dimension with an additional set of indicators: demodulated amplitude, demodulated phase, and entropic spread. The layout follows that of Figure 4 in the text.

Supplementary Table S5: Mean values of the boxplot ensembles (from Supplementary Figure S4) for the changes in the indicators (top 5 rows) and their 12-year variability (bottom 5 rows) for each region, in the demodulated amplitude (magnitude), the demodulated phase (timing), and the entropic spread (duration).

		Magnitude (mm/year)	**Timing** (days/year)	**Duration** (days/year)
Δ Indicators	Northeast Brazil	5.8	0.042	0.13
	Western Africa	-6.3	-0.021	-0.037
	Central Africa	-0.87	-0.0098	0.042
	Northern Australia	0.94	-0.025	0.087
		Magnitude (mm/year)	**Timing** (days/year)	**Duration** (days/year)
Δ 12-year variability	Northeast Brazil	2.3	0.040	0.036
	Western Africa	-0.85	0.0044	0.023
	Central Africa	0.40	0.031	-0.0080
	Northern Australia	2.5	0.0072	-0.030

Supplementary Table S6: Minimum and maximum values of the boxplot ensembles (from Supplementary Figure S4) for the changes in the indicators (top 5 rows) and their 12-year variability (bottom 5 rows) for each region. The indicators are the demodulated amplitude (magnitude), the demodulated phase (timing), and the entropic spread (duration).

		Magnitude (mm/year)	**Timing** (days/year)	**Duration** (days/year)
Δ Indicators	Northeast Brazil	-1.4 — 23	-0.026 — 0.12	-0.080 — 0.33
	Western Africa	-26 — 3.6	-0.077 — 0.037	-0.11 — 0.065
	Central Africa	-6.5 — 3.8	-0.081 — 0.067	-0.039 — 0.12
	Northern Australia	-9.9 — 9.0	-0.18 — 0.078	-0.076 — 0.24
		Magnitude (mm/year)	**Timing** (days/year)	**Duration** (days/year)
Δ 12-year variability	Northeast Brazil	-4.8 — 15	-0.0752—0.150	-0.073 — 0.11
	Western Africa	-7.3 — 7.0	-0.0751—0.0769	-0.057 — 0.094
	Central Africa	-2.3 — 5.1	-0.0426—0.128	-0.072 — 0.093
	Northern Australia	-2.6 — 6.6	-0.0731—0.106	-0.13 — 0.056

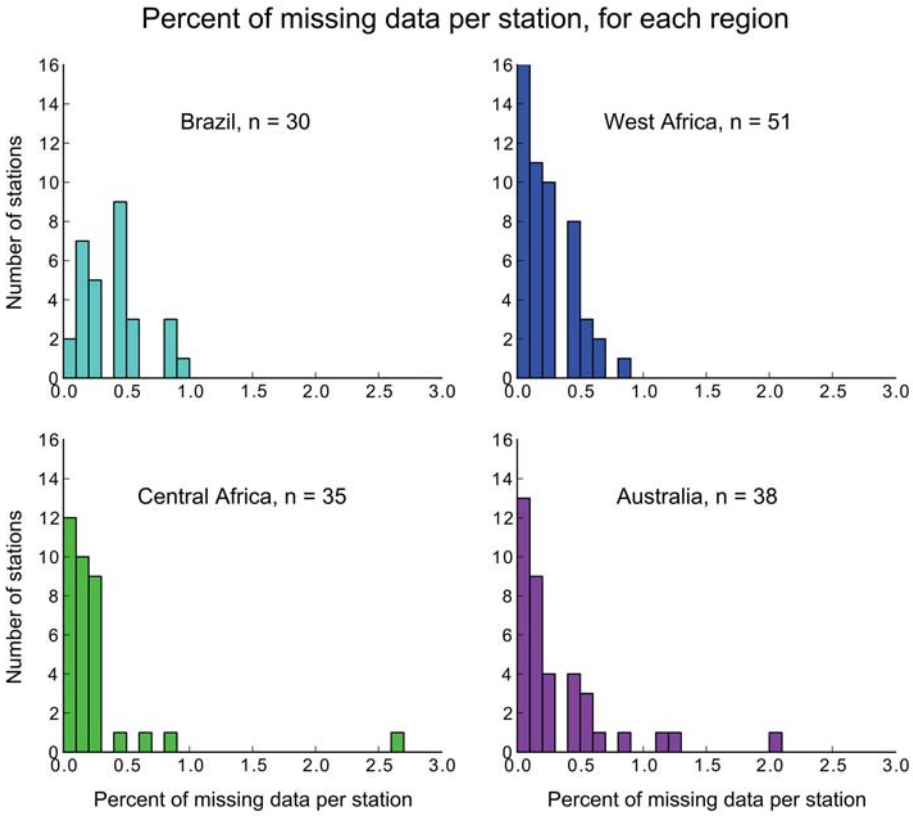
Supplementary Table S7: P-values of the boxplot ensembles (from Figure 4) for the changes in the indicators (top 5 rows) and their 12-year variability (bottom 5 rows) using the Wilcoxon signed-rank test with a null hypothesis that the mean value of the ensemble is zero. The p-values are associated with the ensembles of the demodulated amplitude (magnitude), the demodulated phase (timing), and the entropic spread (duration).

		Magnitude	Timing	Duration
Δ Indicators	Northeast Brazil	2.8 E-06	2.1 E-05	1.2 E-05
	Western Africa	2.0 E-09	1.4 E-05	5.6 E-07
	Central Africa	2.7 E-02	3.8 E-02	1.1 E-05
	Northern Australia	1.1 E-01	5.5 E-02	2.7 E-05
		Magnitude	**Timing**	**Duration**
Δ 12-year variability	Northeast Brazil	7.0 E-04	1.2 E-03	5.6 E-04
	Western Africa	9.7 E-04	3.2 E-01	8.6 E-05
	Central Africa	1.7 E-01	7.4 E-04	2.3 E-01
	Northern Australia	7.7 E-06	4.0 E-01	2.5 E-03

4. Sensitivity analysis

4.1 Missing data

To test the sensitivity of our results to the presence of missing data in the monthly rainfall series (Supplementary Figure S5), we compute the mean percentage change (%) of the trend values (as presented in Supplementary Tables S2 and S5) averaged over each regional ensemble. Each station was subjected to 1000 iterations of random gaps punched along its rainfall time series according to the number and sizes of the original gaps present. For example, if a station contained 2 consecutively missing months in 1935 and 1 missing month in 1978, then those "original" gaps were first filled by linear interpolation (see Methods) and then 2 gaps, of length 2 and 1 month, were randomly punctured anywhere along the rainfall series. Then the series filled with the "original" and the "random" gaps were processed simultaneously to produce the 6 indicators of their seasonality components. The percent difference in their trends are determined and averaged over 1000 iterations for each station and over all stations used to produce Figure 4 and Supplementary Figure S4. The results, presented below in Supplementary Tables S8 and S9, show the largest observed mean percentage change in Central Africa, at –10.4%, for the spread.



Supplementary Figure S5: Distribution in each region of the percent of missing data per station used in the time series analysis of Figure 4 and Supplementary Figure 4, over the 720 months between 1930-1990. All stations contain less than 3% missing data.

Supplementary Table S8: Sensitivity of trends in the indicators (top 5 rows) and their 12-year variability (bottom 5 rows) to randomly punctured gaps for the annual rainfall (magnitude), centroid (timing), and spread (duration).

		Magnitude %	Timing %	Duration %
Δ Indicators	Northeast Brazil	1.53	-3.19	-0.03
	Western Africa	-0.44	0.32	1.63
	Central Africa	-0.72	-5.34	0.83
	Northern Australia	-0.15	0.09	0.60
		Magnitude %	**Timing %**	**Duration %**
Δ 12-year variability	Northeast Brazil	0.01	1.27	-0.05
	Western Africa	-0.11	1.28	-0.16
	Central Africa	-0.11	0.24	-10.39
	Northern Australia	-0.28	-4.16	-0.28

Supplementary Table S9: Sensitivity of trends in the indicators (top 5 rows) and their 12-year variability (bottom 5 rows) to randomly punctured gaps for the demodulated amplitude (magnitude), demodulated phase (timing), and the entropic spread (duration).

		Magnitude %	Timing %	Duration %
Δ Indicators	Northeast Brazil	-1.12	0.47	1.23
	Western Africa	0.20	-0.15	-0.21
	Central Africa	-0.36	0.06	3.73
	Northern Australia	1.54	-1.51	0.94
		Magnitude %	**Timing %**	**Duration %**
Δ 12-year variability	Northeast Brazil	1.28	0.20	8.08
	Western Africa	0.04	1.47	0.59
	Central Africa	-0.19	0.32	0.04
	Northern Australia	-0.14	0.01	-0.09

4.2 Aggregation of data from the daily to the monthly scale

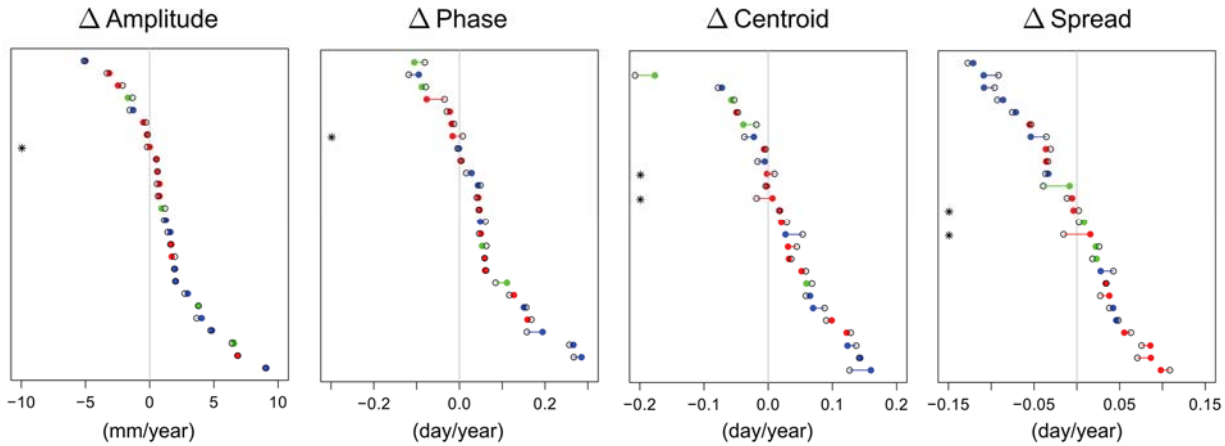
Here we show that the fractional monthly trends we observe in the indicators using monthly rainfall series provide significant information and are relatively unaffected by the scale at which we conducted our analyses. Using daily rainfall data gathered from 26 stations from the Daily Global Historical Climatology Network[6], we calculate the indicator trends (and also their variability trends) from daily series and their equivalent monthly series. While each station's trend has an error margin associated with the aggregation of data from the daily to the monthly level, by inspection these errors have only marginal effects on the indicator trends (Supplementary Figures S7 and S8).

To quantify the trends that can result from aggregation errors, we adopt a Bayesian approach using a simple linear regression model:

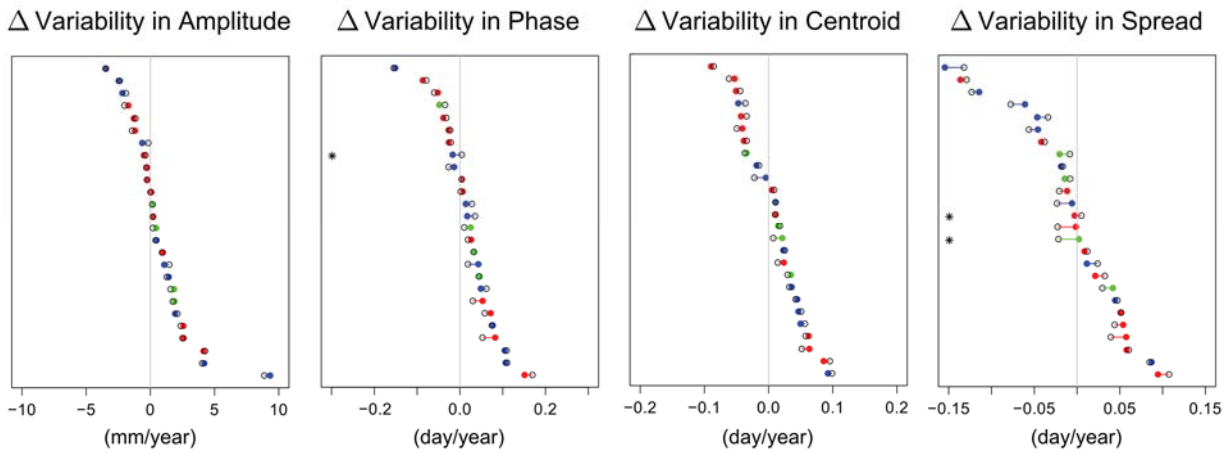
$$y_i = \alpha + \beta x_i + \varepsilon_i,$$

where $\{x_i, y_i\}$ are ordered pairs of years and indicator errors (the difference between results from using the monthly vs. daily series), and ε_i are independently drawn from $Normal(0, \sigma^2)$, whose variance σ^2 is estimated from a fitted normal distribution. The posterior distribution of the slope β (using a noninformative flat prior) gives us information about possible trends that can result from the aggregation of data. For most stations, the trends in the errors are an order of magnitude *below* those produced from the monthly series, lending support to real changes observed outside of aggregation uncertainties (Supplementary Figures S7, S8, and Table S10).

Supplementary Figure S6: Locations of stations from the GHCN-Daily dataset used for quantifying errors due to aggregation of daily to monthly rainfall.



Supplementary Figure S7: Ordered pairs of points from 26 stations (from the GHCN-Daily dataset, locations shown in Supplementary Figure S6) of trends for the annual indicator series (shown in each panel) calculated from the observed daily series (filled dot) and an equivalent monthly series (empty dot). Point pairs that span between negative and the positive trends are marked with a star at the left side of the panel. Dot colors correspond to the region from which the original data is collected (Red = Australia, Blue = Northeast Brazil, Green = Africa, see Figure L4).



Supplementary Figure S8: Ordered pairs of points from 26 stations (from the GHCN-Daily dataset, locations shown in Supplementary Figure S6) of trends for the 12-year variability in the annual indicator series (shown in each panel) calculated from the observed daily series (filled dot) and an equivalent monthly series (empty dot). Point pairs that span between negative and the positive trends are marked with a star at the left side of the panel. Dot colors correspond to the region from which the original data is collected (Red = Australia, Blue = Northeast Brazil, Green = Africa, see Figure L4).

Supplementary Table S10: 95% sample quantile range for the posterior mean of (1) the trend in the difference of annual indicators and (2) the trend in the difference of the 12-year variability of annual indicators when they are calculated using the observed daily series or (a) its equivalent aggregated monthly series, (b) a synthesized daily series in which the monthly total is placed on the mid-month day, and (c) a synthesized daily series in which the monthly total is distributed uniformly over all days in the month. "Same" denotes unchanged ranges between method (b) and (c).

	1. Posterior trend			**2. Posterior trend in the variability**		
	a. Monthly	b. Midmonth	c. Uniform	a. Monthly	b. Midmonth	c. Uniform
Amplitude (mm/year)	(–0.31, 0.35)	--	--	(–0.36, 0.42)	--	--
Phase (day/year)	(–0.031, 0.037)	--	--	(–0.019, 0.026)	--	--
Centroid (day/year)	(–0.023, 0.032)	(–0.022, 0.031)	same	(–0.010, 0.015)	(–0.0097, 0.015)	same
Spread (day/year)	(–0.018, 0.031)	(–0.017, 0.031)	(–0.018, 0.030)	(–0.016, 0.022)	(–0.017, 0.022)	same

References:

1. Vose, R. S. *et al.* The Global Historical Climatology Network: long-term monthly temperature, precipitation, sea level pressure, and station pressure data. (1992).

2. Mitchell, T. D. & Jones, P. D. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology* **25**, 693–712 (2005).

3. Cover, T. M. & Thomas, J. A. *Elements of Information Theory*. (John Wiley & Sons, Inc.: Hoboken, New Jersey, 2006).

4. Rodriguez-Iturbe, I., Dawdy, D. R. & Garcia, L. E. Adequacy of Markovian Models with Cyclic Components for Stochastic Streamflow Simulation. *Water Resources Research* **7**, 1127 (1971).

5. Bloomfield, P. *Fourier Analysis of Time Series: An Introduction*. (John Wiley & Sons, Inc.: New York, 2000).

6. Tank, A. K. & Coauthors. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *International Journal of Climatology* **22**, 1441–1453 (2002).