

ARTICLE OPEN

Seasonality of Taylor's law of fluctuation scaling in all-India daily rainfall

Michael K. Tippett^{1*} and Joel E. Cohen^{2,3,4}

Here we show that the variance over time of all-India daily rainfall (AIR) can be related to the mean over time of AIR by a seasonally varying power law. Outside of the peak monsoon months of July and August, AIR variance increases in proportion to a positive power of mean daily rainfall. During July and August, monthly averages of AIR show little association with the corresponding variances. This power-law relationship of temporal variance to temporal mean is known in biological sciences as Taylor's law (TL) and in physical sciences as fluctuation scaling. We explain the seasonal variation in TL qualitatively by independent and identically distributed random sampling. Accounting for day-to-day correlation in AIR is sufficient to match quantitatively the observed power-law behavior. Our findings provide a quantitative month-specific assessment of the variability of AIR that could prove useful for the design of crop insurance and social safety nets for the large fraction of the population of the Indian subcontinent that depends on rainfed agriculture.

npj Climate and Atmospheric Science (2020)3:3; <https://doi.org/10.1038/s41612-019-0104-6>

INTRODUCTION

The Indian monsoon is a major determinant of the freshwater supply of the Indian subcontinent. Its variability from year-to-year influences natural environments, agricultural production, food prices, human migration from rural to urban areas, and economic and political stability in the region.¹ Predicting the date of onset and the magnitude of precipitation of the Indian monsoon has challenged meteorologists since the nineteenth century, and has stimulated the development of novel statistical methods.² Much research has focused on seasonal and monthly averages of rainfall although societal impacts are more closely related to rainfall that occurs on shorter time periods.³ Understanding of the temporal variability of the rainfall in the region remains incomplete.

"For some purposes averages are adequate; for other purposes full data on their range of variation are indispensable."⁴ The relationship of variance to mean is called a variance function in statistics and is called fluctuation scaling in physical sciences. Here we examine the variance function of all-India daily rainfall (AIR). Variance functions have been examined for annual, summer monsoon, and monthly total rainfall across stations.⁵ We show that this variance function is remarkably well approximated by a seasonally varying power-law relationship:

$$\text{variance}(\text{set of observations}) = a \times [\text{mean}(\text{set of observations})]^b. \quad (1)$$

Several ecologists and statisticians, including Taylor,⁶ in the first half of the twentieth century found that this power-law relationship described remarkably well several spatial distributions of insects and plants. Subsequently, the power-law form of the variance function was named Taylor's law (hereafter TL), and TL has been confirmed for a very wide variety of data on the spatial and temporal variability of hundreds of species, including human populations,⁷ as well as physical, financial, and meteorological phenomena.^{8–11}

RESULTS

We found that TL is a useful simple approximation to the observed variance function of AIR in two ways. Among days of the year from January 1 to December 31, the mean and variance (over the 116 years 1901–2016) of AIR on each calendar date separately scaled according to TL with exponent that was slightly less than one (Fig. 1). A TL exponent of one is consistent with Poisson processes with rates that vary from day to day of the year. The behavior here is more complex, with hysteresis that is particularly evident between spring and fall when mean values of AIR are similar but variances are higher in fall.

Among years from 1901 to 2016, the monthly mean and variance of AIR scaled according to TL with exponent that varied by month (Fig. 2). This scaling for each month separately is a hierarchical temporal TL: e.g., for January, the 116-year daily time series was divided into month-long blocks (e.g., January 1901, January 1902, etc.), and for each January of each year, the monthly mean and monthly variance were calculated over the days within that month.

The exponent b (the slope on the log–log coordinates in Fig. 2) was statistically indistinguishable from zero in the peak monsoon months of July and August. Consequently, monthly averages of AIR during these peak monsoon months provided little information about the corresponding variances. In all other months, b did not fall significantly outside the interval [1, 2], as in many other empirical tests of TL. Consequently, the coefficient of variation of daily rainfall in any given month diminished, although the variance increased, with increasing mean daily rainfall. After July and August, the next lowest TL exponents were in the monsoon onset and withdrawal months June and September with values of 1.32 and 0.726, respectively. A quadratic version of TL scaling (see Methods) with negative quadratic coefficient fitted the data better for the months of October and November, with the improvement in fit being most evident for the years with least rainfall in those months (Fig. 2).

¹Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY 10027, USA. ²Laboratory of Populations, Rockefeller University, New York, NY 10065, USA. ³Earth Institute and Department of Statistics, Columbia University, New York, NY 10027, USA. ⁴Department of Statistics, University of Chicago, Chicago, IL 60637, USA. *email: mkt14@columbia.edu

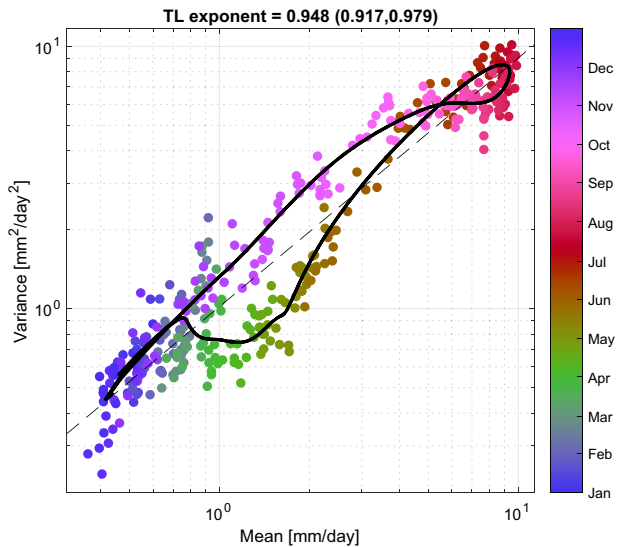


Fig. 1 Mean (mm/day) and variance (mm/day)² of all-India daily rainfall for each day of the year (color-coded) 1901–2016, on log-log coordinates. TL exponent b and 95% confidence interval are given in the title. The estimate of the proportionality constant a is 1.018 with 95% confidence interval (0.980, 1.056). The solid black curve results from LOESS (locally estimated scatterplot smoothing) of the abscissa and ordinate data.

How can the dependence of the TL exponent b on calendar month be understood? Stochastic multiplicative growth can interpret TL.¹² We tested (see Methods) the monthly averages, variances, 10th percentiles, and 90th percentiles for linear and quadratic trends during 1901–2016, and found no compelling evidence for trends, so we rejected the stochastic multiplicative growth model as an explanation for TL here.

The gamma distribution was previously used to fit wet-day Indian station data.¹³ The fitted gamma distribution matched the mean well in all months, overestimated July–September variances, and underestimated (overestimated) the skewness in December–March (June–November) (Fig. 3a–d). Random sampling from a gamma distribution fitted to AIR with parameters that vary by month had median TL exponent of 2 (Supplementary Fig. 7) as predicted analytically.¹⁴ The gamma distribution failed to fit the AIR data.

Random sampling of skewed distributions can interpret TL.¹⁴ In that case, the TL exponent, denoted b_{CX15} , is approximated by the skewness of the data divided by its coefficient of variation (see the Cohen-Xu formula in Eq. (3) of the Methods). The monthly statistics of AIR showed that mean AIR was highest in July (Fig. 3a); variance was highest in June and July (Fig. 3b); the coefficient of variation was lowest in July and August (Fig. 3c); and the skewness was highest in winter months and lowest in July and August (Fig. 3d). Consequently, the seasonal variation of b_{CX15} , computed separately for each calendar month, qualitatively resembled the TL monthly exponent of AIR but overestimated it for most months (Fig. 4).

The validity of the b_{CX15} approximation for iid data was confirmed by independent-day sampling (see Methods) of the data, as 95% confidence intervals of the TL exponent of 1000 realizations of independent-day sampling included the estimates from¹⁴. We interpret this result as a numerical confirmation of the Cohen-Xu formula in Eq. 3 for iid data. On the other hand, we interpret the result that the observed TL exponent is often outside the 95% confidence intervals of the TL exponent of 1000 realizations of independent-day sampling as reason to reject the null hypothesis of iid data.

We considered two ways in which the data violate the iid assumption of the Cohen-Xu formula. The first is that data from different days are dependent. The summer monsoon is famously marked by “break” periods of 3–8 days of dry conditions and “active” periods of 3–4 days of wet conditions.¹⁵ Temporal autocorrelation (i.e., correlation from 1 day to the next or neighboring days) has been documented in an AIR index during the monsoon season.¹³ We confirmed that the AIR autocorrelation was statistically significant for lag 1 on all days of the year, for lag 2 on most days of the year, and for particularly long lags in the monsoon onset month of June and to a lesser extent during the monsoon withdrawal months of September–November (Supplementary Fig. 2a). The second way in which the data violate the iid assumption of the Cohen-Xu formula is that data from different years have different distributions. Seasonal totals of AIR tend to be lower during El Niño years and higher during La Niña years.^{16,17} Differences in seasonal means can also be viewed as a form of dependence since a low-AIR value observed today makes it more likely that this month is a low-AIR month, which means a low-AIR value is more likely tomorrow. We tested by ANOVA whether averages of AIR computed over windows ranging from 3 to 31 days varied significantly from year to year. Year-to-year differences in averages of AIR were statistically significant for windows >7 days during most of the year, highest in early June near the monsoon onset date, and also large during the monsoon withdrawal months of September–November (Supplementary Fig. 2b). Monsoon onset dates vary substantially from year to year, and the standard deviation of the onset date is ~8 days.¹⁸

Having found two ways in which the data violate the iid assumption of the Cohen-Xu formula, we designed sampling strategies to test whether they are quantitatively sufficient to account for the discrepancy between the observed TL exponents and those of the Cohen-Xu formula. We used block sampling to test the impact of autocorrelation up to 7 days on the TL exponent. Block sampling represents random variation among years and correlation in rainfall in segments of successive days averaging just over 7 days in length (see Methods for details). Confidence intervals for the TL exponent under block sampling contained the ordinary least-squares (OLS) point estimates of the TL exponent in all 12 months. We used tercile sampling to measure the impact of different years having different means, assuming that different days have independent amounts of rainfall. The confidence intervals of the TL exponent b under tercile sampling contained the OLS point estimates of the slope from the daily AIR during 10 months (Fig. 4). In July and October, the tercile sampling confidence intervals fell below the OLS point estimates of the slope. We used tercile block sampling to test the combined impact on the TL exponent of different years having different mean rainfall and of autocorrelation in the rainfall of successive days up to 7 days. Tercile block sampling gave confidence intervals for the TL exponent b that contained the OLS point estimates for all months. Unlike block sampling, tercile block sampling allowed for inter-annual differences in the monthly average. The variation during the annual cycle of the monthly TL exponent b of AIR can be partially explained by the different values of Niño 3.4 among years (Fig. 4). When years were stratified into terciles according to the level of Niño 3.4, the TL exponent agreed with that estimated from the daily AIR except during January, March, and April, when the 95% interval from Niño 3.4 sampling fell (slightly) above the OLS point estimate of b . Block sampling was the best explanation of the seasonality in b in the sense that its intervals contained all point estimates and its assumptions were minimal (Fig. 4).

DISCUSSION

We identified patterns in daily AIR not previously recognized and explained them. We showed that, outside of the summer, each

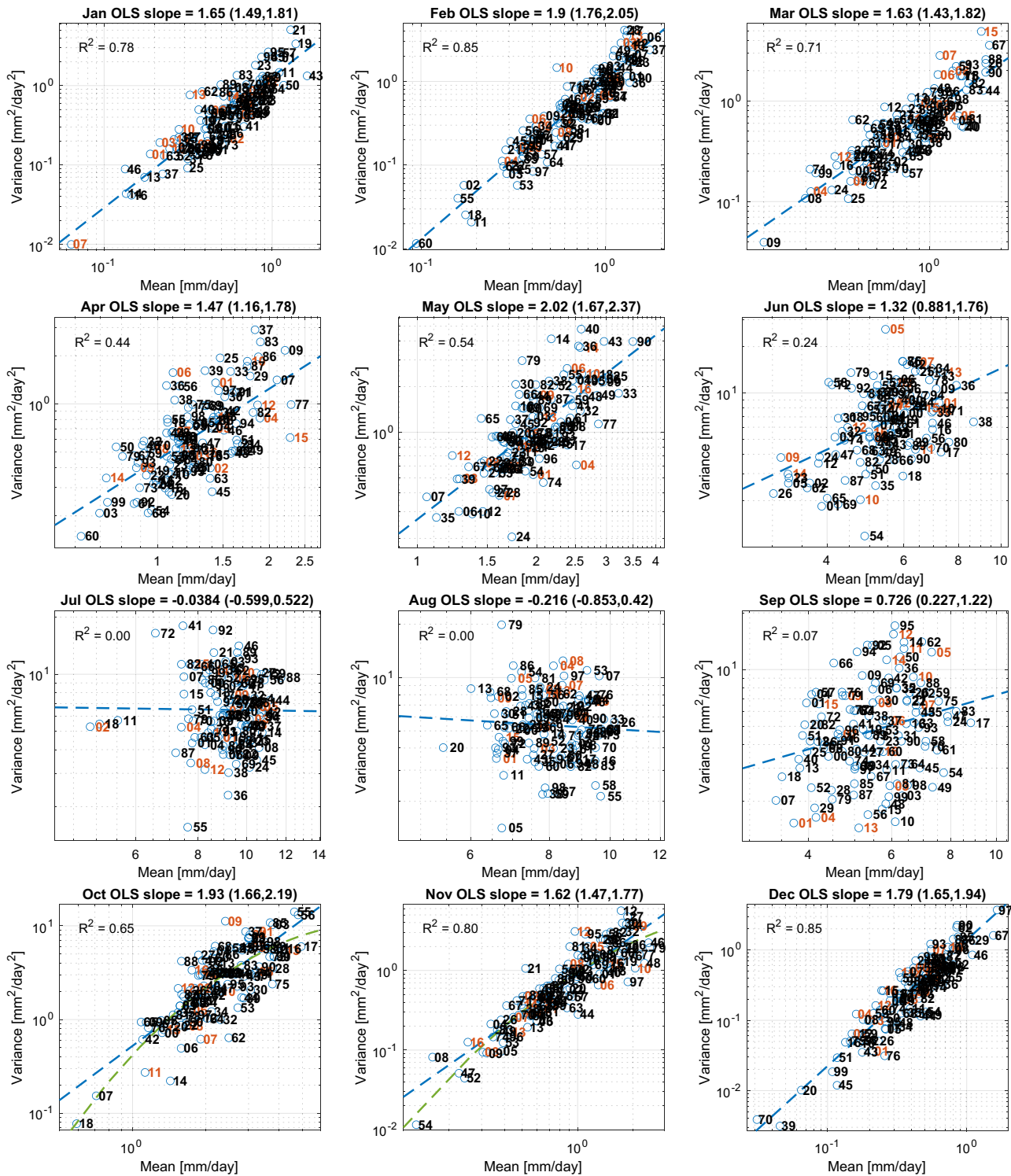


Fig. 2 Mean (mm/day) and variance (mm/day)² of all-India daily rainfall for each month 1901–2016, on log-log coordinates. Each panel is a calendar month. The two-digit markers indicate the year with red for 2001–2016. Slopes of best-fit lines (blue dashes) and 95% confidence intervals estimated by ordinary least-squares (OLS) for each calendar month are given in the left corner. Best-fit OLS quadratic curves (green dashes) are shown for the months of October and November; for these months, the quadratic curve fit the data statistically significantly better (at the 0.05 significance level) than the fitted linear model.

month's variance of daily rainfall varied across years approximately as a power function of each month's mean of daily rainfall, in accordance with TL. The exponent of TL varied seasonally from values between 1 and 2 outside the monsoon season to values

statistically indistinguishable from zero during the months of peak rainfall, July and August. The lack of TL scaling in summer months means that monthly averages of AIR do little to constrain the within-month variance of AIR. We documented the temporal

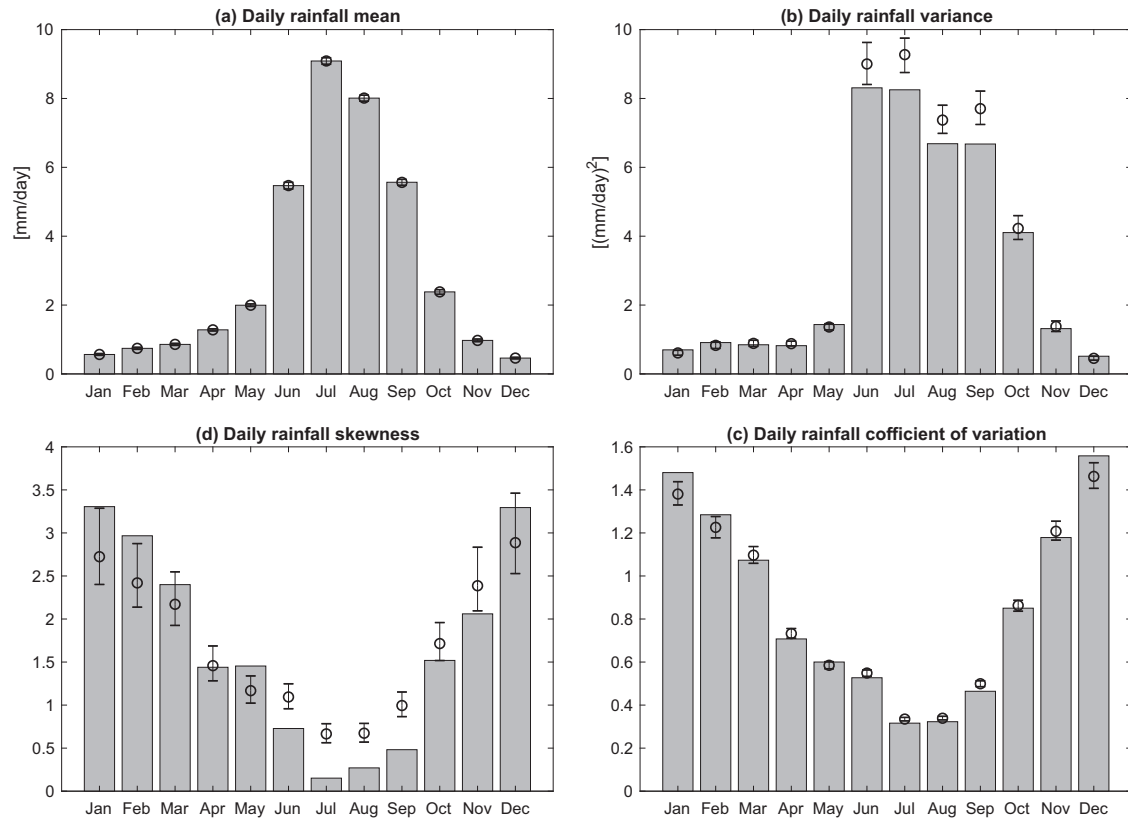


Fig. 3 All-India daily rainfall by calendar month. **a** Mean, **b** variance, **c** coefficient of variation, and **d** skewness. Black error bars indicate the 95% intervals of the corresponding values for samples from a gamma distribution fitted by maximum likelihood to the monthly data.

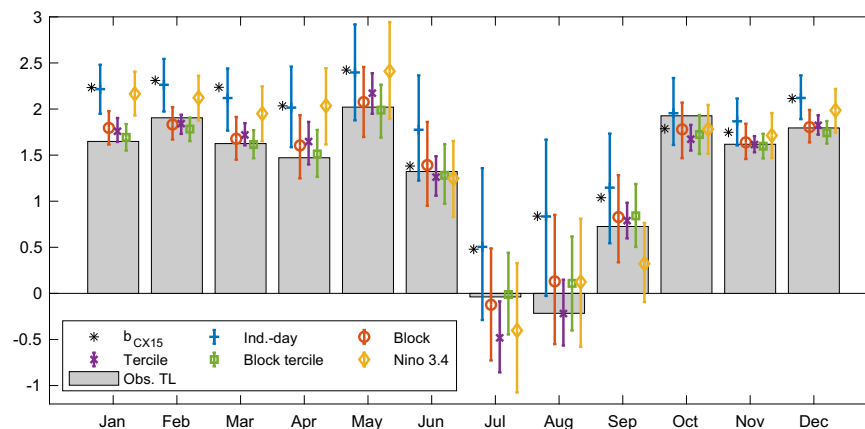


Fig. 4 Taylor exponent by calendar month (gray bars). Median and 2.5 to 97.5 percentile error bars of Taylor exponents for independent-day sampling (blue), block sampling (red), tercile sampling (purple), tercile block sampling (green), and Niño 3.4 sampling (yellow). Black asterisks are the point estimates of the exponent b_{CX15} from ref. ¹⁴ and are within the 95% intervals of the independent-day samples, as predicted analytically.

variance, temporal skewness, and temporal autocorrelation of AIR month by month and the inter-annual variation of AIR in random blocks of days with mean length just over a week. We showed that the variation by month of the TL exponent for AIR is accounted for by dependence within random blocks of days sampled from a skewed distribution across years, but is not well described by a gamma distribution.

A question for future research is whether TL holds for precipitation on smaller spatial scales. This issue is of particular interest because societal impacts depend directly on rainfall at

specific locations rather than on a country-wide average. At least for the summer monsoon, seasonal rainfall amounts at the sub-divisional level are correlated with the all-India total, but the association is not perfect.¹⁹ Summer monsoon rainfall time series from sub-divisions that are near each other tend to be strongly and positively correlated.²⁰ We have examined the spatial variation of the TL exponent for rainfall among days of the years (analogous to the TL exponent near one in Fig. 1) and found a TL exponent with coherent spatial variation and values between 1 and 2 (not shown). The linkages between the spatio-temporal

variability of the TL exponent for rainfall among years and rainfall processes remain to be investigated.

The TL scaling found in AIR has potential forecast applications. During non-monsoon months, mean is related to variance. If a physics-based model or statistical model can predict the mean rainfall, TL could be used to infer the variance. TL scaling also provides a new diagnostic that could be used to assess the realism of physics-based computer models that are used to make seasonal forecasts and climate change projections, an application which is similar to the use of TL scaling as a metric for selecting among demographic projections.²¹

METHODS

Data

We used two observational data sets: the Niño 3.4 index and daily AIR. The Niño 3.4 index is the average of tropical Pacific sea surface temperature (SST) in the latitude-longitude box 5N–5S, 170W–120W and here was computed from version 5 of the Extended Reconstructed Sea Surface Temperature data set.²²

AIR values over the period 1901–2016 were computed as area-weighted averages of gridded ($0.25^\circ \times 0.25^\circ$) daily rainfall data.^{23,24} The interval from 1 January 1901 to 31 December 2016, included 42,369 days, of which 29 days fell on February 29. For some calculations, the latter 29 days were omitted for computational convenience, leaving 42,340 calendar dates and AIR values.

AIR expressed in mm/day and percentile rank showed strong seasonality and regularity from year to year (Supplementary Fig. 1). The highest values were found in the monsoon months of June–September. Prior to testing TL, we checked AIR for trends. Specifically, we plotted, separately for each month, the temporal mean, temporal variance, 10th percentile, and 90th percentile of that month's AIR over the 116 years (Supplementary Figs. 3–6). We also estimated the parameters and 95% confidence intervals of the ordinary least-squares linear and quadratic regressions of these four monthly statistics as functions of year. Our null hypotheses were that the slope of every linear regression, and the slope and quadratic coefficient of every quadratic regression, were zero. Five of the null hypothesis models were rejected at the nominal 5% level while $4.8 = 0.05 \times 2 \times 12 \times 4$ cases were expected to be nominally significant by random sampling alone, and these cases of apparent significance were sporadically distributed across months and models. We concluded that our tests found no evidence for a trend across these 116 years in the monthly means, variances, 10th percentiles, and 90th percentiles of AIR.

Taylor's law

To describe TL more precisely, let Θ be an arbitrary (discrete or continuous, finite or infinite) non-empty set with typical element θ . We interpret each $\theta \in \Theta$ as the label of a probability distribution. Let $X(\theta) := \{X(\theta): \theta \in \Theta\}$ be a family of nonnegative real-valued random variables with finite, positive mean $\mu(\theta)$ and variance $\sigma^2(\theta)$ for each $\theta \in \Theta$, i.e., such that $\mu(\theta) := E(X(\theta)) \in (0, \infty)$ and $\sigma^2(\theta) := E[(X(\theta) - \mu(\theta))^2] \in (0, \infty)$, where E denotes expectation. Then we say that TL holds for the family $X(\theta)$ if and only if there exist a positive real constant a and a real constant b , both independent of $\theta \in \Theta$, such that for all $\theta \in \Theta$, $\sigma^2(\theta) = a[\mu(\theta)]^b$. Here the value of θ corresponds to, or is the label of, a “set of observations” in Eq. (1). In our empirical example, suppose that AIR values 1901–2016 are arranged in the form of a table with 365 rows, one for each day of the year (neglecting February 29), and 116 columns, one for each year from 1901 through 2016. The first row contains 116 January 1 values, and the last row contains 116 December 31 values (Supplementary Fig. 1). We constructed sets of observations based on rows and based on columns. Taking each row to be a set of observations gives 365 (neglecting February 29) means and variances, as shown in Fig. 1. Here each θ labels a day of the year. Alternatively, taking each column to be a set of observations gives 116 annual means and annual variances, and each θ labels a year. Here we refined annual analysis to monthly analysis. By taking the first 31 rows of each column to be a set of observations, that is, each January's collection of 31 days, we constructed the scatter plots in as in Fig. 2a (and similarly for each other month). In Fig. 2a, each θ labels the January of a particular year.

In theoretical examples, a Poisson distribution with mean $\theta \in \Theta := (0, \infty)$ has variance θ , so TL holds for this family of Poisson random variables with $a = 1$, $b = 1$. An exponential distribution with mean $\theta \in \Theta := (0, \infty)$ has

variance θ^2 , so TL holds for this family of exponential random variables with $a = 1$, $b = 2$. The gamma distribution with shape parameter $k > 0$ and scale parameter $\xi > 0$ has probability density function $x^{k-1}e^{-x/\xi}/[\Gamma(k)\xi^k]$, mean $k\xi$, and variance $k\xi^2$. For fixed k and varying ξ , TL holds with $a = 1/k$, $b = 2$. For varying k and fixed ξ , TL holds with $a = \xi$, $b = 1$. This example shows that the same family of distributions may give rise to different sets of TL parameters a , b depending on which parameter(s) vary and which are held constant.

By the strong law of large numbers and Slutsky's theorem, if TL holds for the moments of $X(\theta)$, then asymptotically for large samples of observations of each $X(\theta)$, (1) or equivalently

$$\log(\text{variance}(\text{set of observations})) = b \log(\text{mean}(\text{set of observations})) + \log a \quad (2)$$

must hold approximately for the sample means and sample variances. The error term in empirical tests of (2), i.e., the deviation from exact agreement, depends on $X(\theta)$. Prior empirical tests of TL often assumed without testing that the error term in empirical tests of (2) was independently and identically distributed (iid) according to a normal distribution with constant variance for all $\theta \in \Theta$. We found that the autocorrelation of the residuals in Fig. 2 up to lag 20, separately for each of the 12 months, all fell within the 95% confidence intervals around no autocorrelation, except for 5 (sporadically) of the 240 values. Therefore we followed the convention of assuming that the error term in empirical tests of (2) was iid according to a normal distribution with constant variance for all $\theta \in \Theta$. We estimated the TL exponent b and its 95% confidence intervals using ordinary least-squares fits to the logarithms of means and variances of daily AIR.

Some models of TL

TL can arise from many distinct stochastic processes in addition to the parametric families of distributions described above.^{8,11} We consider here two models that make assumptions that we can test against the data. This approach goes beyond fitting a power law to the empirically estimated variance function by trying to identify underlying statistical processes. The two models posit stochastic multiplicative growth¹² and random sampling of skewed distributions.¹⁴

The model of stochastic multiplicative growth assumes $X(\theta) \in (0, \infty)$ with probability 1, i.e., $\Pr\{X(\theta) \in (0, \infty)\} = 1$, and for $\theta = 1, 2, \dots$, $X(\theta) = Y(\theta) \times X(\theta - 1)$, where the stationary ergodic stochastic process $Y(\cdot)$ satisfies $\Pr\{Y(\theta) \in (0, \infty)\} = 1$ for all θ . Here the index θ labels discrete time, and $Y(\theta)$ is the multiplicative growth rate of $X(\cdot)$ from time $\theta - 1$ to θ . When $Y(\cdot)$ is iid or Markovian, the mean and variance of $X(\theta)$ asymptotically grow or decline exponentially, except in critical cases, and the TL exponent b depends on the relation of the exponential growth rates of the mean and the variance.^{12,25} An initial test of this model is to examine time series of observations of $X(\theta)$ for exponentially increasing or decreasing trend. In the absence of an asymptotically strictly increasing or strictly decreasing exponential trend, this model does not hold.

Another simple model for the origin of TL assumes that all sampled observations are iid from a fixed nonnegative distribution with finite positive mean and variance and with finite non-zero skewness, and observations are randomly allocated to blocks (sets of observations) labeled $\theta = 1, 2, \dots, m$.¹⁴ Asymptotically, when the number n_θ of observations in each block θ becomes large, the sample means and sample variances of the blocks converge to TL with estimated exponent b that converges in probability to the value given by the Cohen-Xu formula¹⁴

$$b_{\text{CX15}} = \frac{\gamma_1}{\text{CV}}, \quad (3)$$

where γ_1 is the skewness and CV is the coefficient of variation (standard deviation divided by mean) of the random variable. Here, blocks are constructed using calendar months. For instance, January AIR values could be grouped in blocks by year with $\theta = 1901, \dots, 2016$ labeling the years, and, for each of these blocks, with $n_\theta = 31$ observations (days). For gamma-distributed data of the same shape (i.e., with fixed k), $b_{\text{CX15}} = 2$ regardless of the parameters of the gamma distribution.

The data would reject this “null” model (called “null” because it assumes no mechanism other than iid random sampling) if the data were dependent, or if there were significant differences among blocks of observations, e.g., most obviously, if there were significant differences of mean precipitation from month to month or from year to year when only a single month is considered. Rejection of this model would indicate that some other statistical process must be invoked to explain TL when a power law is a good description of the variance function.

Analysis of variance

Analysis of variance (ANOVA) tests for significant year-to-year differences in monthly averages by comparing the between-month variance to the within-month variance. Under the null hypothesis of independence and no difference in means or variances, the between-month variance should equal the within-month variance scaled by the number of days in the month. In ANOVA, the test statistic F is

$$F = \frac{\text{between-month variance}}{\frac{1}{W-1} \text{within-month variance}}, \quad (4)$$

where W is the number of days in a month. Values of F far from 1 favor rejection of the null hypothesis of no difference in means, assuming equal variances. Here, we generalize that monthly analysis to averaging windows of size W centered on each calendar day (Supplementary Fig. 2b). The variance ratio F is

$$F = \frac{\frac{1}{Y-1} \sum_{y=1}^Y \left(\frac{1}{W} \sum_{d=1}^W R_{dy} - \frac{1}{WY} \sum_{y=1}^Y \sum_{d=1}^W R_{d'y} \right)^2}{\frac{1}{W-1} \frac{1}{Y} \sum_{y=1}^Y \frac{1}{W-1} \sum_{d=1}^W \left(R_{dy} - \frac{1}{W} \sum_{d'=1}^W R_{d'y} \right)^2}, \quad (5)$$

where Y is the number of years of data, here 116 (the first and last years are lost owing to windows that extend into the previous year or into the subsequent year), and R_{dy} is the rainfall in year y on day d of the window centered on a given calendar day. In the case of Gaussian iid data, $F \sim F(Y-1, Y(W-1))$, and the critical value of F at the 1% level for the smallest window size $W=3$ is 1.45 and for the largest window size $W=31$ is 1.34.

By F test, variance ratios were highly statistically significantly different from one for all averaging windows centered on all days of the year (Supplementary Fig. 2b); the smallest variance ratio was 2.79, which is much higher than the critical values. The expected variance ratio for iid Gaussian data is 1. Here significance testing based on critical values of the F -distribution confirmed that the daily rainfalls were sequentially dependent. The permutation test with block sampling accounts for the relatively short autocorrelation previously detected and corresponds to the null hypothesis of serially correlated (from day to day) but identically (from year to year) distributed data.

Sampling

To identify factors such as dependence and year-to-year changes in monthly means that might explain the observed behavior of the monthly TL exponent b over the annual cycle, we constructed 12×116 artificial months of pseudo rainfall (the same size as the AIR data) by sampling the daily AIR in five ways described below. We then calculated the monthly mean and monthly variance of pseudo rainfall and estimated the TL exponent b as we did for the AIR data. For each sampling method, we repeated the process of sampling and computing the TL exponent 1000 times. For each sampling method, we compared the TL exponent from the AIR data with the 1000 values from the samples. If the TL exponent from the AIR data fell outside the 2.5 to 97.5 percentiles of the 1000 values from the sampled data, we rejected the model associated with the sampling strategy.

1. **Independent-day sampling** preserves the day of the year. The years are permuted randomly. For example, the pseudo rainfall on 1 January 1901, is chosen at random from the AIR on January 1 of all years; and the same for every day and year. The pseudo rainfall on 29 February 1904 is chosen from other February 29's in leap years and from March 1's in non-leap years. This sampling was used to test the null hypothesis that all-India rainfall on a given day of a given month is iid across years, and rainfall from different (even adjacent) days is independent.
2. **Block sampling** is the same as independent-day sampling except days in the same segment are drawn from the same year. Each 365-day year is partitioned into 52 segments of random length. The start dates of the 52 segments consist of the 12 dates 1 January, 1 February, ..., 1 December, and of 40 dates chosen randomly without replacement from the remaining 353 dates. The end date of each segment is the day before the next start date. The average length of the segments is just over 7 days. As a consequence of this construction, no segment intersects with more than 1 month. The pseudo rainfall for the first segment of 1901 is chosen at random from one of the 116 years. For instance, if the first segment of 1901

were January 1–6, the January 1–6 AIR values from 1983 might be chosen as the pseudo rainfall for the first segment of 1901. In the same way, subsequent 1901 segments would then be populated with values on the same days from randomly selected years. An independent partitioning is used for 1902 and for each year of the simulation. Block sampling was used to test the null hypothesis that random intervals of AIR within a month are iid across years: specifically that within each month, intervals containing a random number of days less than 1 month are randomly sampled from all years, allowing for correlation in rainfall in segments of successive days averaging just over 7 days in length. Within each block, all the autocorrelation of AIR is preserved. Across blocks, pseudo rainfall values are independent. Consequently, block sampling gives within-month autocorrelation which is slightly weaker than that in Supplementary Fig. 2a and which is zero by construction at the end of each month.

3. **Tercile sampling** is the same as independent-day sampling but all days in a month are drawn from years in which the corresponding month is in the same tercile category. The tercile category of a month is determined by summing its daily AIR values and comparing that sum to the sum for the same month in all other years. For example, if January in a particular year falls in the lowest tercile of summed January rainfalls, then, for that year, January 1 is simulated by random sampling of the roughly one-third of January 1's in all years in the lowest tercile; and likewise for January 2; and likewise if January in a particular year falls in the middle tercile or the highest tercile. To simulate 1 January 1901, the AIR on January 1 is chosen at random from all January 1's of the years in the same tercile as January 1901; while the simulated rainfall of 1 February 1901, is chosen at random from all February 1's of the years in the same tercile as February 1901, which may differ from the tercile of January 1901. This sampling was used to test the null hypothesis that all-India rainfall on a given day of a given month is iid across years within the same tercile, allowing for differences among terciles, and rainfall from different (even adjacent) days is independent.
4. **Niño 3.4 sampling** is the same as tercile sampling except that the pseudo rainfalls on all days in a given month are drawn from years in which the corresponding month is in the same Niño 3.4 tercile as the given month. The Niño tercile of a month is determined by comparing its Niño 3.4 value to the Niño 3.4 value of the same month in all other years. This sampling was used to test the null hypothesis that all-India rainfall on a given day of a given month is iid across years within the same Niño 3.4 tercile, allowing for differences among terciles, and rainfall from different (even adjacent) days is independent.
5. **Block tercile sampling** is block sampling with the additional constraint that all segments in a month are drawn from years in the same tercile category.

Block sampling with fixed block length introduces non-stationarity into stationary data, because the joint distribution of adjacent values is not invariant to shifts. For instance, suppose that a 7-day block were used, and pseudo rainfall data were constructed by sampling January 1–7 from random years and January 8–14 independently from other random years. Such sampling would result in January 6 and 7 being dependent, and January 7 and 8 being independent. Therefore, we used block sampling with random block length to preserve day-to-day correlation in the pseudo rainfall.²⁶

Estimating the variance function: testing for linear and quadratic trends

We estimated linear and quadratic versions of TL using OLS. The quadratic version of TL²⁷ has the form $\log \text{variance} = a + b \log \text{mean} + c(\log \text{mean})^2$. We judged the quadratic version of TL to be a better fit to the data when the reduction in sum squared error was statistically significant by F test at the 0.05 significance level.

Gamma distribution goodness of fit

Here we fitted a gamma distribution by maximum likelihood estimation to AIR of each month separately, using all years of AIR data. A sample of the same size as the AIR data was drawn from the gamma distribution with the fitted parameters. Monthly means, variances, skewness, and TL exponents

were computed for each month. This sampling was repeated 1000 times and 2.5 to 97.5 percentiles were used to form the error bars in Fig. 3.

DATA AVAILABILITY

The gridded rainfall data are available from the India Meteorological Department http://www.imd.gov.in/advertisements/20160219_advnt_12.pdf. Version 5 of the Extended Reconstructed Sea Surface Temperature data set is available from the IRI Data Library <http://iridl.ldeo.columbia.edu/SOURCES/NOAA/NCDC/ERSST/version5/sst>.

Received: 19 August 2019; Accepted: 12 December 2019;
Published online: 17 January 2020

REFERENCES

- Gadgil, S. & Gadgil, S. The Indian monsoon, GDP and agriculture. *Economic Political Weekly* **41**, 4887–4895 (2006).
- Katz, R. W. Sir Gilbert Walker and a connection between El Niño and statistics. *Stat. Sci.* **17**, 97–112 (2002).
- Gadgil, S., Rao, P. S. & Rao, K. N. Use of climate information for farm-level decision making: rainfed groundnut in southern India. *Agricultural Systems* **74**, 431–457 (2002).
- Usher, A. P. *A History of Mechanical Inventions* (Harvard University Press, 1929), 1954 edn.
- Singh, N. & Mulye, S. S. On the relations of the rainfall variability and distribution with the mean rainfall over India. *Theor. Appl. Climatol.* **44**, 209–221 (1991).
- Taylor, L. R. Aggregation, variance and the mean. *Nature* **189**, 732–735 (1961).
- Cohen, J. E., Xu, M. & Brunborg, H. Taylor's law applies to spatial variation in a human population. *Genus* **69**, 25–60 (2013).
- Eisler, Z., Bartos, I. & Kertész, J. Fluctuation scaling in complex systems: Taylor's law and beyond. *Adv. Phys.* **57**, 89–142 (2008).
- Tippett, M. K. & Cohen, J. E. Tornado outbreak variability follows Taylor's power law of fluctuation scaling and increases dramatically with severity. *Nat. Commun.* **7**, 10668 (2016).
- Tippett, M. K., Lepore, C. & Cohen, J. E. More tornadoes in the most extreme U.S. tornado outbreaks. *Science* **354**, 1419–1423 (2016).
- Taylor, R. A. J. *Taylor's Power Law: Order and Pattern in Nature* (Elsevier Academic Press, Cambridge, MA, 2019).
- Cohen, J. E., Xu, M. & Schuster, W. S. F. Stochastic multiplicative population growth predicts and interprets Taylor's power law of fluctuation scaling. *Proc. R. Soc. B* **280**, 20122955 (2013).
- Stephenson, D. B. et al. Extreme daily rainfall events and their impact on ensemble forecasts of the Indian monsoon. *Mon. Weather Rev.* **127**, 1954–1966 (1999).
- Cohen, J. E. & Xu, M. Random sampling of skewed distributions implies Taylor's power law of fluctuation scaling. *Proc. Natl Acad. Sci.* **112**, 7749 (2015).
- Rajeevan, M., Gadgil, S. & Bhate, J. Active and break spells of the Indian summer monsoon. *J. Earth Syst. Sci.* **119**, 229–247 (2010).
- Kumar, K. K., Rajagopalan, B., Hoerling, M., Bates, G. & Cane, M. Unraveling the mystery of Indian monsoon failure during El Niño. *Science* **314**, 115–119 (2006).
- Kucharski, F. & Abid, M.A. Interannual variability of the Indian monsoon and its link to ENSO, <https://doi.org/10.1093/acrefore/9780190228620.013.615>.
- Pai, D. S. & Nair, R. M. Summer monsoon onset over Kerala: new definition and prediction. *J. Earth Syst. Sci.* **118**, 123–135 (2009).
- Mooley, D. A. & Parthasarathy, B. Indian summer monsoon and El Niño. *Pure Appl. Geophys.* **121**, 339–352 (1983).
- Parthasarathy, B. Interannual and long-term variability of Indian summer monsoon rainfall. *Proc. Indian Acad. Sci.—Earth Planet. Sci.* **93**, 371–385 (1984).
- Xu, M., Brunborg, H. & Cohen, J. E. Evaluating multi-regional population projections with Taylor's law of mean-variance scaling and its generalisation. *J. Popul. Res.* **34**, 79–99 (2017).
- Huang, B. et al. Extended reconstructed sea surface temperature, version 5 (ERSSTv5): upgrades, validations, and intercomparisons. *J. Clim.* **30**, 8179–8205 (2017).
- Pai, D. et al. Development of a new high spatial resolution (0.25° × 0.25°) long period (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. *Mausam* **65**, 1–18 (2014).
- Pai, D. S., Sridhar, L., Badwaik, M. R. & Rajeevan, M. Analysis of the daily rainfall events over India using a new long period (1901–2010) high resolution (0.25° × 0.25°) gridded rainfall data set. *Clim. Dyn.* **45**, 755–776 (2015).
- Cohen, J. E. Stochastic population dynamics in a Markovian environment implies Taylor's power law of fluctuation scaling. *Theor. Popul. Biol.* **93**, 30–37 (2014).
- Politis, D. N. & Romano, J. P. The stationary bootstrap. *J. Am. Stat. Assoc.* **89**, 1303–1313 (1994).
- Taylor, L., Woiwod, I. & Perry, J. The density-dependence of spatial behaviour and the rarity of randomness. *J. Anim. Ecol.* **47**, 383–406 (1978).

AUTHOR CONTRIBUTIONS

M.K.T. and J.E.C. designed the analysis. M.K.T. carried out the calculations. M.K.T. and J.E.C. wrote the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at <https://doi.org/10.1038/s41612-019-0104-6>.

Correspondence and requests for materials should be addressed to M.K.T.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

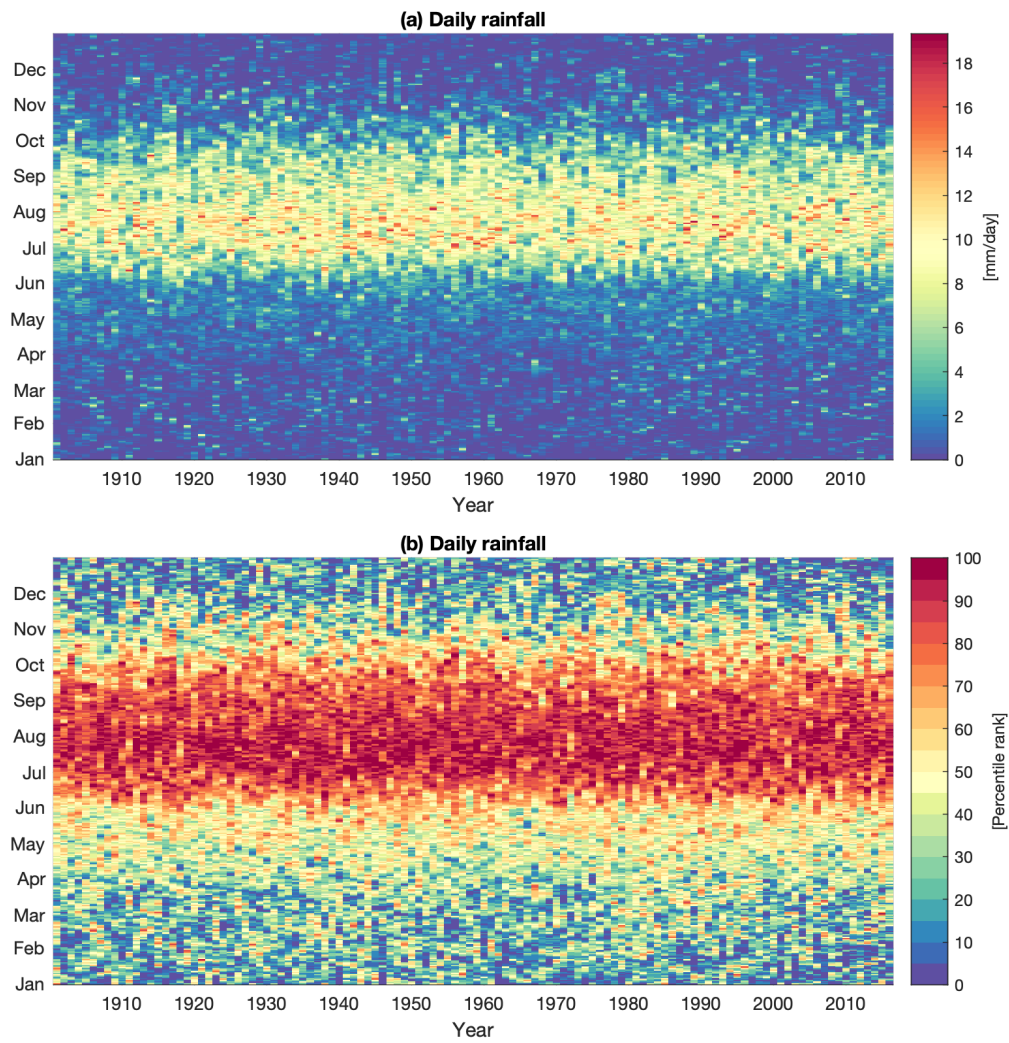


Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

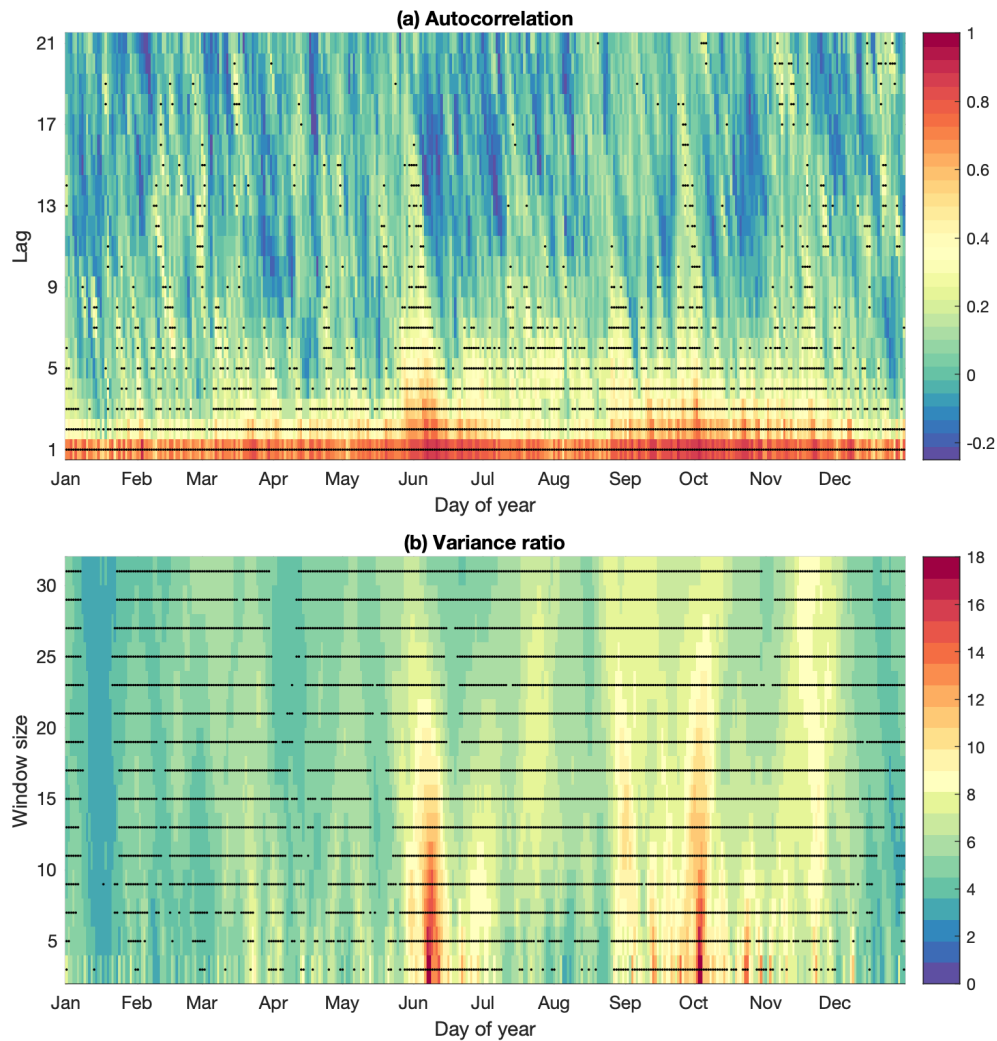
© The Author(s) 2020

Supplementary Information

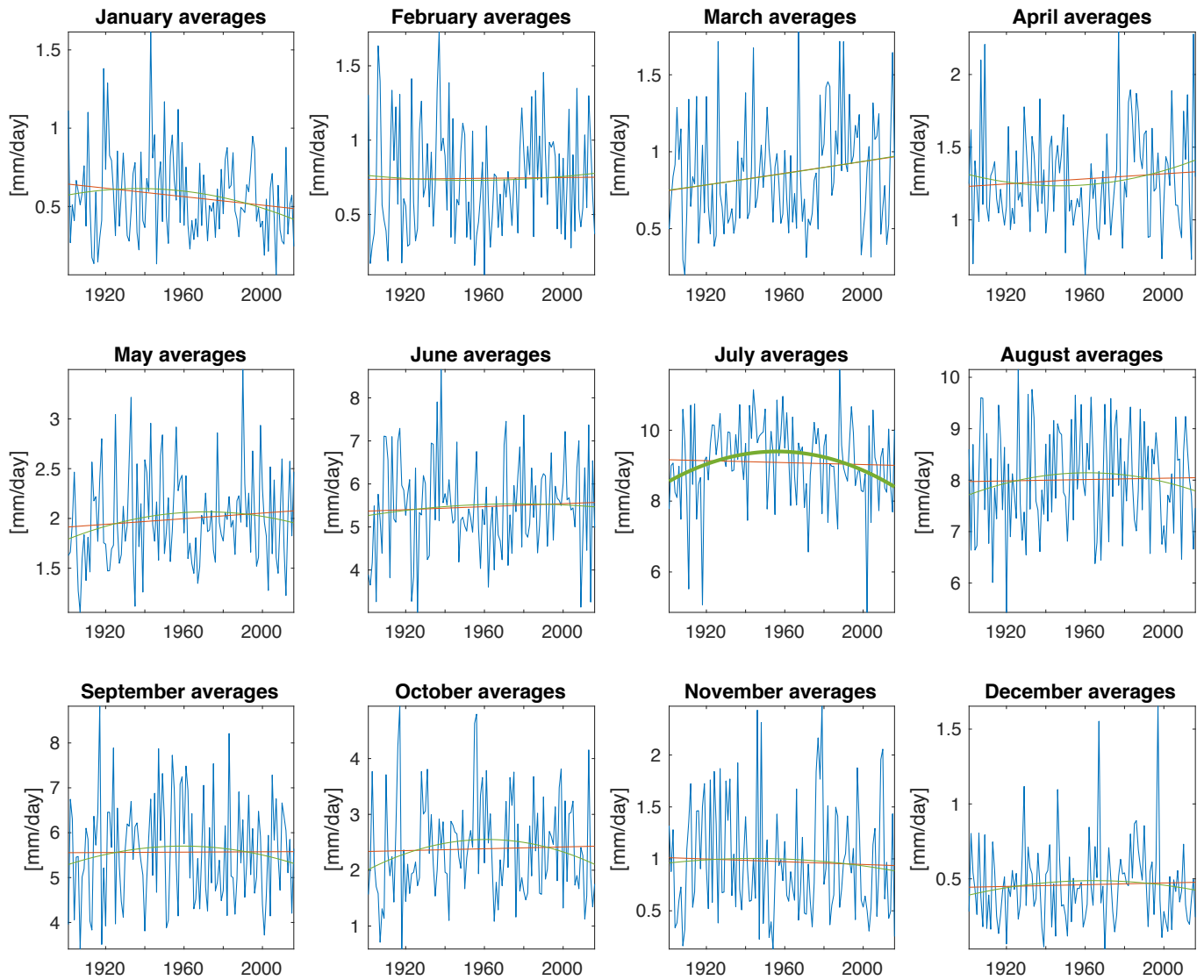
Supplementary Figures 1–7.



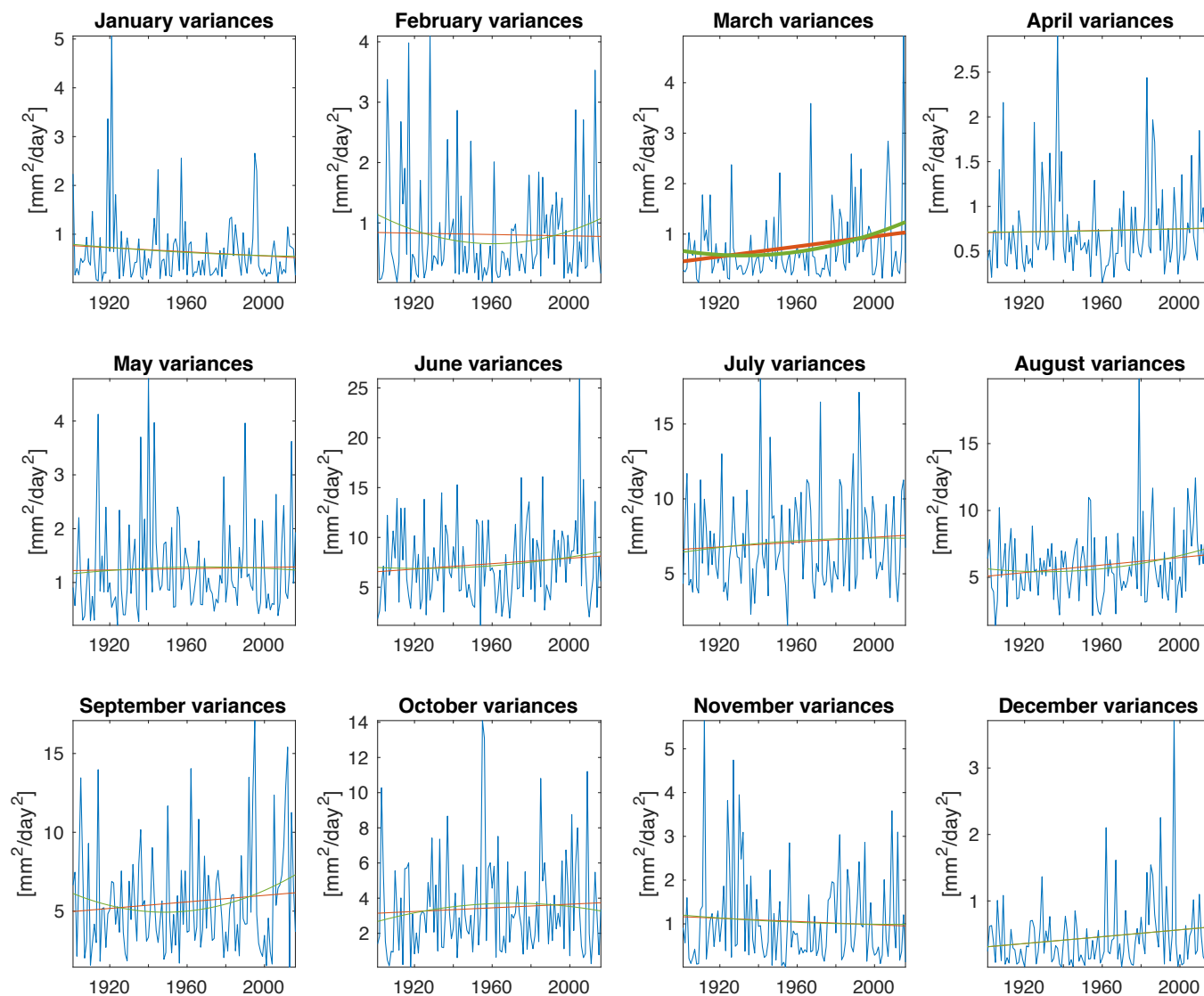
Supplementary Figure 1. All-India daily rainfall 1901–2016 expressed in units of (a) mm/day and (b) percentile rank. Percentile rank is relative to all the data.



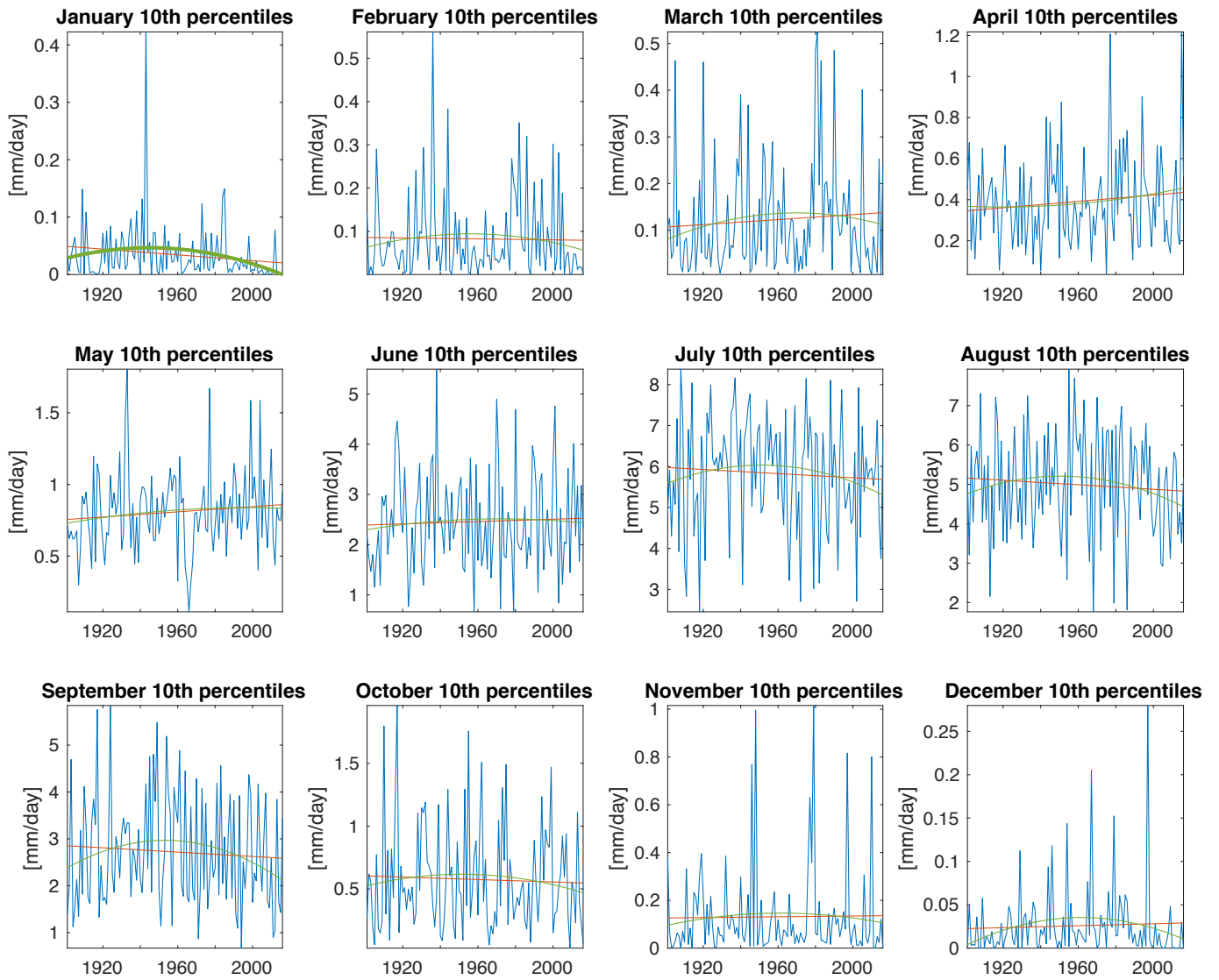
Supplementary Figure 2. (a) Autocorrelation of all-India rainfall for each day of the year (x-axis) for lags 1–21 (y-axis). (b) Variance ratio defined as the sample variance of the sample mean divided by the sample mean of the sample variance multiplied by the length of the window (in days) centered on the day of the year for windows of length 3, 5, . . . , 31. Values in panel (a) larger in magnitude than 0.187 (sample size 115; one year is lost because of windows that start in December and go into the next year) are statistically significant at the 0.05 level and marked by a black dot. Values in panel (b) that are statistically significant at the 0.05 level by permutation test with block sampling are marked by a black dot.



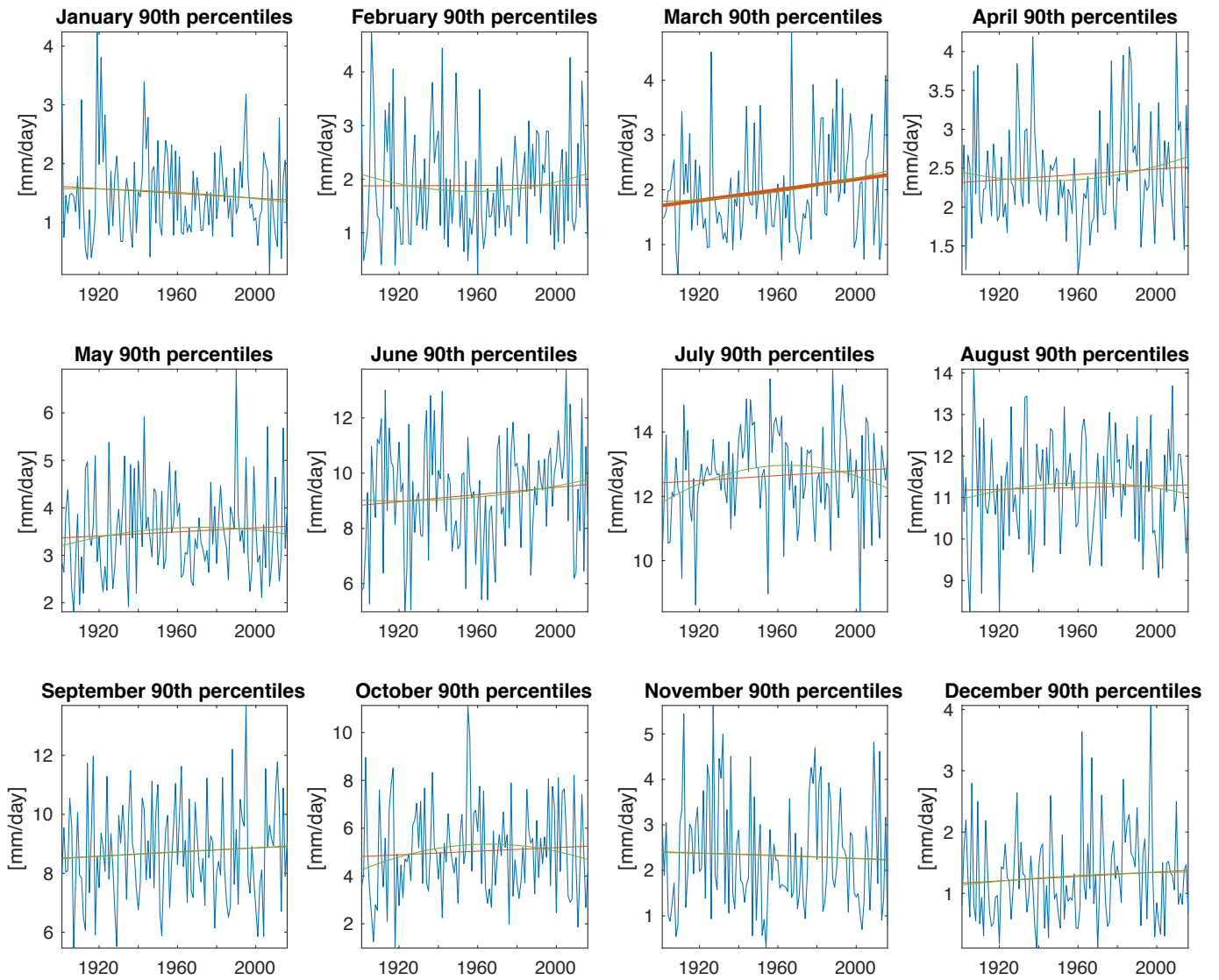
Supplementary Figure 3. Averages of January, February, . . . , and December daily AIR, 1901–2016, along with linear (red) and quadratic (green) trend curves. Statistically significant trends are indicated by thick curves, i.e., only July’s quadratic (green) trend curve differs significantly from a horizontal straight line in this figure.



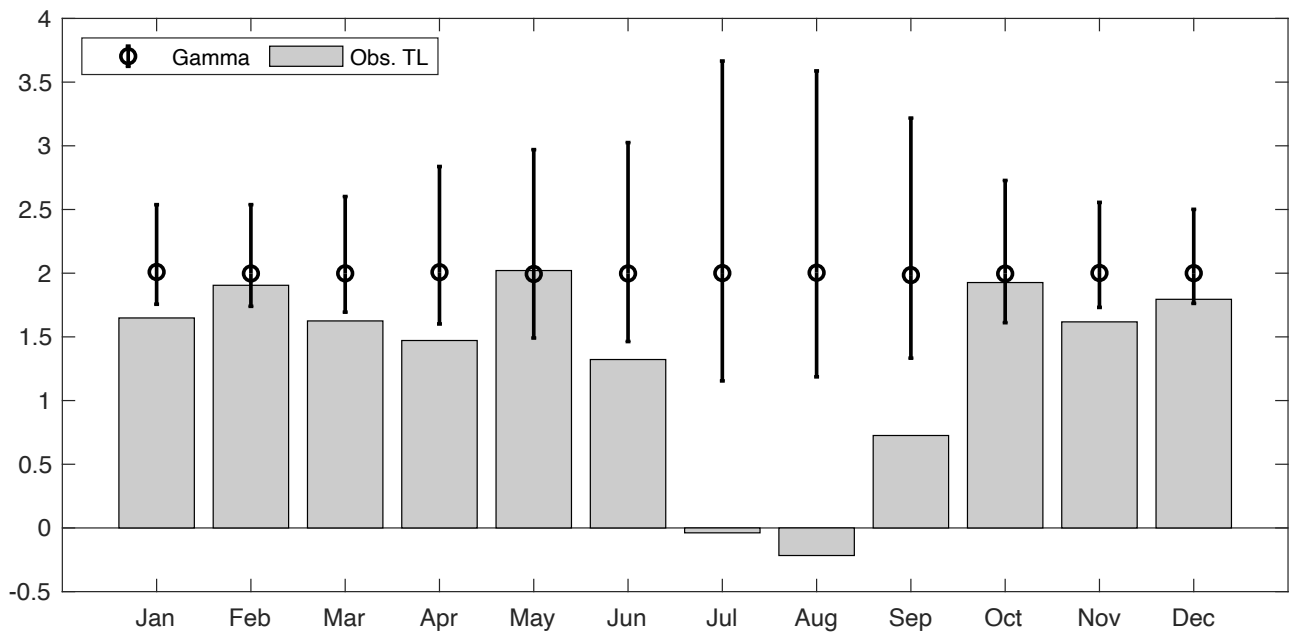
Supplementary Figure 4. Variances of January, February, . . . , and December daily AIR, 1901–2016, along with linear (red) and quadratic (green) trend curves. Statistically significant trends in March are indicated by thick curves.



Supplementary Figure 5. 10th percentiles of January, February, . . . , and December daily AIR, 1901–2016, along with linear (red) and quadratic (green) trend curves. A statistically significant quadratic trend in January is indicated by a thick curve.



Supplementary Figure 6. 90th percentiles of January, February, . . . , and December daily AIR, 1901–2016, along with linear (red) and quadratic (green) trend curves. A statistically significant linear trend in March is indicated by a thick line.



Supplementary Figure 7. Taylor exponent b by calendar month (gray bars) and error bars that indicate the 95% intervals of the corresponding values for samples from a gamma distribution fitted by maximum likelihood to the monthly data.