GEOPHYSICAL RESEARCH LETTERS, VOL. 32, L14812, doi:10.1029/2005GL022465, 2005

A fundamental probability distribution for heavy rainfall

P. S. Wilson and R. Toumi

Department of Physics, Imperial College London, London, UK

Received 17 January 2005; revised 13 June 2005; accepted 21 June 2005; published 22 July 2005.

[1] There is currently no physical understanding of the statistics of heavy precipitation. These statistics are, however, central to diagnosing climate change and making weather risk assessments. This work derives a fundamental rainfall distribution. Interpretation of the water balance equation gives a simple expression for precipitation as the product of mass flux or advected mass, specific humidity and precipitation efficiency. Statistical theory predicts that the tail of the distribution of the product of these three random variables will have a stretched exponential form with a shape parameter of 2/3. This is verified for a global daily precipitation data set. The stretched exponential tail explains the apparent 'heavy' tailed behaviour of precipitation under standard assumptions used in extreme value theory. The novel implications for climate change are that the stretched exponential shape is unlikely to change, although the scale may, and precipitation efficiency is important in understanding future changes in heavy precipitation. **Citation:** Wilson, P. S., and R. Toumi (2005), A fundamental probability distribution for heavy rainfall, *Geophys. Res. Lett.*, *32*, L14812, doi:10.1029/2005GL022465.

1. Introduction

[2] A host of statistical distributions (e.g. gamma [*Groisman et al.*, 1999], log-normal [*Biondini*, 1976], mixed exponential [*Woolhiser and Roldan*, 1982]) are used to empirically approximate daily precipitation totals. The physical justification for these distributions is however unclear and hence their applicability to unmeasured or future extremes and their structural stability under climate change is questionable [*Allen and Ingram*, 2002; *Trenberth*, 1999; *Trenberth et al.*, 2003]. Heavy precipitation events are known to be dominated by moisture advection from the surrounding regions [*Trenberth*, 1999; *Trenberth et al.*, 2003] with a negligible contribution from local evaporation and relatively little contribution from change in the total column atmospheric water content. Local precipitation can therefore be expressed as the column integrated moisture flux. Applying a two layer model of the atmosphere with low level convergence up to the top of the moist layer, z_m, gives local precipitation as [*Stevens and Lindzen*, 1978]

$$R \simeq -\int_0^{z_m} \overline{\nabla_H \cdot (q\rho\mathbf{v})} dz = \int_0^{z_m} \frac{\overline{\partial q\rho\mathbf{w}}}{\partial z} dz = \overline{(q\rho\mathbf{w})}_{z_m} \qquad (1)$$

where R is the precipitation rate, q is specific humidity or mass mixing ratio, ρ is density, $\mathbf{v}$ is the horizontal velocity and $\mathbf{w}$, as we consider only convergence, is the upward vertical velocity. The over-bar represents a temporal average such that $\overline{\rho\mathbf{w}}$ is the mean upward mass flux. Extending this model to include, for example, upper level divergence and increases in moisture storage, the actual precipitation rate can then be expressed as

$$R = \overline{\kappa(q\rho\mathbf{w})}_{z_m} \qquad (2)$$

where κ is the instantaneous precipitation efficiency and represents the fraction of the vertical moisture flux at z_m which is precipitated out [*Fankhauser*, 1988]. Studies using equations 1 and 2 in, for example, evaluating extra-tropical storms [*Palmen*, 1958], orographic precipitation [*Alpert*, 1986], latent heating [*Magagi and Barros*, 2004] and probable maximum precipitation [*Abbs*, 1999] have shown good results by assuming independent variables. Precipitation can therefore be expressed as proportional to the product of just 3 independent variables; the mass flux, the specific humidity and the precipitation efficiency. The accumulated precipitation total, R_{acc}, for a given storm can also be described as a triple product: $R_{acc} = \overline{\kappa}\,\overline{q}m$, where m is the mass of air advected into the column and pushed through the moist level. Both precipitation rate and total precipitation can be thought of as an expression of a multiplicative cascade, where m (and the vertical mass flux) is the relaxation to the initial mass imbalance of the whole system, $\overline{q}$ and $\overline{\kappa}$ are rainfall specific independent fractions.

2. Theory

[3] The idealised nature of the model outlined above, i.e. defined on z_m and averaged over time, limits the availability of observations to characterise the distributions of $\overline{\kappa}$, $\overline{q}$ and $\overline{\rho\mathbf{w}}$ (or m). However, as averaged quantities, assuming sufficiently light tails and a sufficient averaging period, the distributions of these variables can, as a first approximation for the high magnitudes relevant to heavy rainfall, be predicted by the central limit theorem to be Gaussian. Scale normalising the variables, precipitation can be expressed as proportional to the product $n_{\overline{\kappa}} n_{\overline{q}} n_{\overline{\rho\mathbf{w}}}$ where n_i are unit normal variables. *Frisch and Sornette* [1997] have shown that the probability density of the product of a finite number of independent random variables will be approximately of the stretched exponential form in the upper right tail of the distribution, i.e. for very large values. *Laherrere and Sornette* [1998] discuss a number of applications of this concept. *Frisch and Sornette* [1997] show that for large values the probability of the product is controlled by realisations where all terms in the product are of the same order. The probability of precipitation is therefore, to a leading order, the joint probability of 3 random variables all having common values, $R^{1/3}$. For the three Gaussian independent variables in question this equates to a product, such that $P(R) \sim [P(n = R^{1/3})]$

0094-8276/05/2005GL022465$05.00

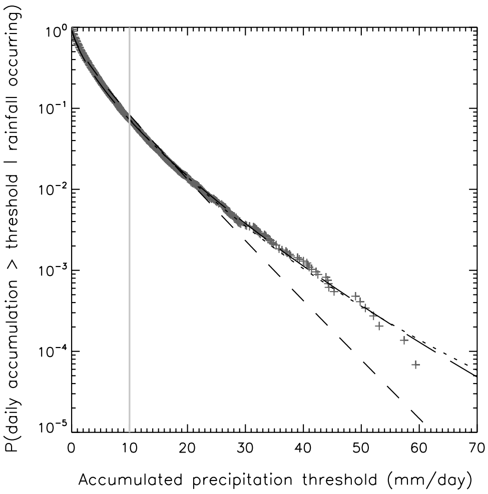

Figure 1. Probability distribution of daily precipitation for Cambridge Botanic Gardens (52.2°N, 0.13°W) 1898–1999. The empirical cumulative distribution (+) and the gamma distribution, fitted to all data, (dashed line) are shown. Fits of precipitation greater than 10 mm/day to both the stretched exponential (long dashed line) and generalised Pareto (dotted line) distributions are also shown. The stretched exponential, linear least squares estimated, shape parameter, c, is 0.695 ± 0.001. The maximum likelihood estimate of the generalised Pareto shape parameter, ξ, is 0.11 ± 0.03.

where $P(n = R^{1/3}) \propto \exp(-(R^{1/3})^2)$. Integrating the resultant probability density the cumulative distribution of heavy precipitation, as the threshold $r \to \infty$, is to a leading order,

$$P(R > r) = \exp\left[-\left(\frac{r}{R_0}\right)^c\right] \tag{3}$$

with shape parameter $c = 2/3$ and scale parameter, R_0, a real number. The form of equation 3 with $c < 1$ is known as a stretched exponential distribution. We describe this distribution as fundamental in the sense that it is based on the concept of individual rain events and assumes that the underlying three variables are Gaussian and independent. On shorter and much longer timescales changes in the atmospheric water content could invalidate equation 1 and the Gaussian assumption may not hold. The well known inverse relationship between duration and intensity [*Barry and Chorley*, 1992] indicates that, for heavy rainfall, daily rainfall can be used as a proxy. Daily totals have therefore been used in data analysis [*Groisman et al.*, 1999; *Karl et al.*, 1995; *Suppiah and Hennessy*, 1998], risk assessments and modelling studies [*Allen and Ingram*, 2002; *Frei et al.*, 1998; *Wilby and Wigley*, 2002]. We examine a global data set of daily precipitation where the time average in equations 1–2 is taken over a single day.

3. Observations

[4] For illustrative purposes Figure 1 shows the probability distribution for daily precipitation totals from Cambridge in the UK. The stretched exponential distribution, with an estimated shape parameter of about 0.7, is shown to be consistent with the generalised Pareto distribution (the general asymptotic tail form for any distribution) and is clearly superior to the underestimation of the observed tail by the most widely used gamma distribution. Figure 2 shows the global variation in the stretched exponential shape parameter fitted to summer and winter heavy precipitation from selected stations of the Global Daily Climatology Network. The mean shape parameter is 0.66 for the summer hemisphere and 0.67 for the winter hemisphere with a standard deviation of 0.16 and 0.18 respectively. The annual global mean shape parameter is 0.66 with a standard deviation of 0.12. We find the same mean shape parameter for different (90th, 95th or 99th percentile) definitions of heavy rainfall. Importantly there is no systematic latitudinal variation (see Figure 3a), although geographical variation is evident. Apart from numerical uncertainties in the fitting, there are a number of potential mechanisms to explain any departures of the estimated shape parameter from 2/3. Our theoretical analysis can now aid in understanding the relevant processes.

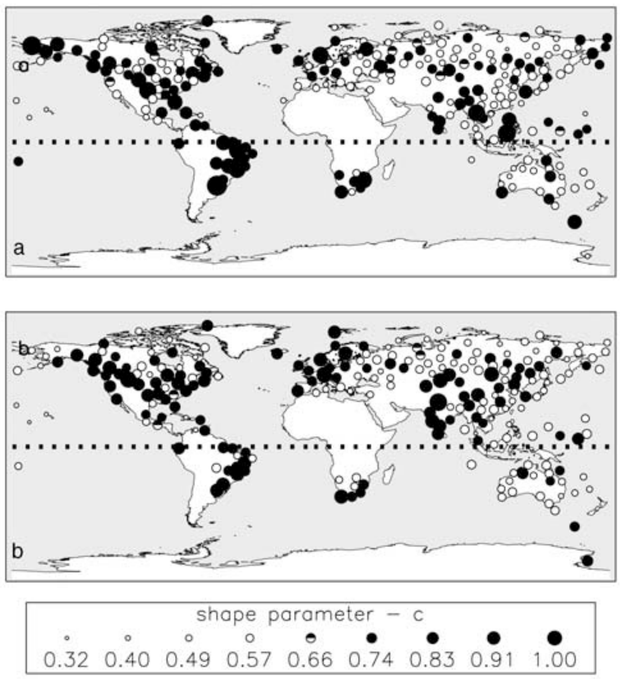

Figure 2. Global variation in the stretched exponential shape parameter. Data from 270 GDCN (http://www.ncdc.noaa.gov/gdcn.html) stations selected on the condition of more than 30 years of daily precipitation data and at least 5° separation. Shape parameter, c, is estimated by linear least squares regression for daily totals with probability less than 5% conditional on there being at least 10 unique points for (a) June through August and (b) December through February. The null hypothesis that the events are distributed according to the stretched exponential distribution cannot be rejected (defined as the χ^2 test statistic outside 2 standard deviations) for 86% of the seasonal data and 71% of the full records, increasing to the expected 95% when accounting for non-stationarity by testing individual non-overlapping 5 year segments.

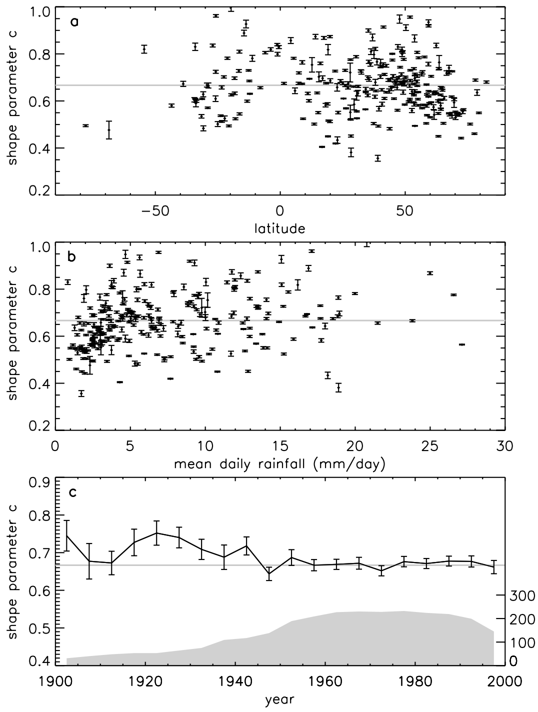

Figure 3. Latitudinal, scale and temporal variation of the stretched exponential shape parameter, *c*. (a) and (b) show, respectively, the variation of the individual station all period estimate of *c* with latitude and mean daily precipitation conditional on precipitation occurring. The error bars for individual stations are the standard error on the linear fit. (c) shows the temporal variation of the mean shape parameter, *c*, from stations with data during non-overlapping 5 year periods. Error bars are the standard error on the mean. The number of stations reporting in each 5 year period is shown as the grey shaded region on the alternate axis. In all figures *c* is estimated for heavy precipitation defined as daily totals with probability less than 5%. The expected mean shape parameter, 0.66, is given as a grey line.

Non-stationarity in the scale parameter, R_0, will give an estimate of *c* less than 2/3 and multiple events contributing to the daily totals can result in $c > 2/3$ [*Sornette*, 2000]. These mechanisms can potentially explain the element of coherent geographical variation in the shape parameter seen in Figure 2, for example, the known non-stationarity in Australian records [*Suppiah and Hennessy*, 1998]. However, for individual non-overlapping 5 year segments of the data, which can be considered approximately stationary, the estimated shape parameters are within the uncertainty range of $c = 2/3$ (estimated via Monte Carlo simulations). Some heavy events may extend over consecutive days. The mean shape parameter is slightly larger at 0.71 and 0.70 with standard deviations of 0.20 and 0.24 for summer and winter hemisphere two day totals respectively. This small increase is expected from the increased potential for multiple independent events contributing to the totals.

[5] To further demonstrate the robustness of the shape parameter, Figure 3b shows the variation of *c* with mean daily precipitation and Figure 3c shows the temporal variation of the mean shape parameter from the non-overlapping 5 year segments. There is no significant long term trend and no systematic variation of the shape of the tail of the distribution with the mean daily precipitation. The robustness of the shape parameter may now seem unsurprising given the physical basis of moisture conservation and its prediction of the stretched exponential form. We also note that simulations of the triple product can be fitted to the widely used gamma distribution with parameters consistent with those estimated for daily rainfall where the gamma shape parameter has also been shown to have invariant properties [*Groisman et al.*, 1999]. However, the gamma distribution has a known tendency to underestimate heavy events (see Figure 1) and there is no theoretical basis for the estimated parameters. Equation 3 however, for the first time, quantitatively predicts the form of the extreme tail.

[6] It is well known that the variation in humidity and atmospheric circulation are critical in determining the occurrence of heavy precipitation. Our analysis puts a further constraint on interpreting heavy precipitation events. Describing rainfall distributions as a product of just two Gaussian variables (e.g. just humidity and vertical mass flux) is excluded from the rainfall analysis since the stretched exponential shape parameter, *c*, is less than one for all stations examined. This implies a critical role for the efficiency parameter (as the third variable) to explain the observed statistics. The precipitation efficiency is controlled by cloud microphysics, entrainment, detrainment, the water holding capacity of the column, and large scale divergence above the moist level [*Trenberth et al.*, 2003]. We assert that predicting the future evolution of the precipitation efficiency is perhaps at least as important as predicting

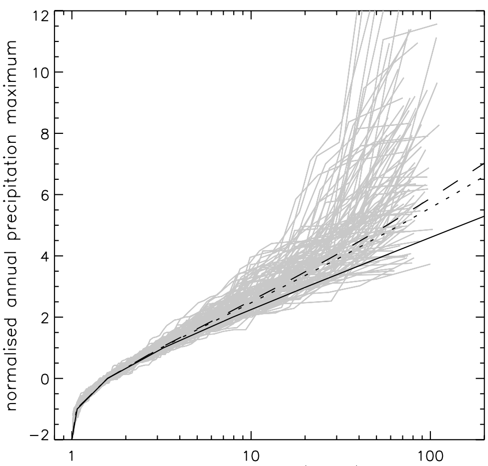

Figure 4. Annual daily maximum precipitation return periods. Grey lines - observed maxima from 102 daily station records from the UK with a minimum of 50 years data (average length 89 years) and less than 15% missing. Dashed line - GEV distribution for the mean maximum likelihood estimated shape parameter (ξ = 0.1) from all stations. Solid line - standard Gumbel form (ξ = 0). Dotted line - predicted GEV distribution from the mean of 100 simulations of annual maxima from 200 year daily time series of 365 independent stretched exponential variables with $c = 2/3$. Annual maxima are normalised by both scale and location.

humidity changes. This is a severe challenge as it is poorly quantified even for our current climate. It may also be worth noting that similar distributions can be expected for other flux related variables (e.g. pollution exposure).

4. Extremes

[7] The physical justification for the stretched exponential distribution has significant implications for the formal treatment of heavy precipitation through extreme value analysis. Extreme value theory predicts that the largest observation from any large sample will belong to the generalised extreme value (GEV) distribution and the magnitude of exceedance of a high threshold will be distributed according to the generalised Pareto (GP) distribution [*Coles*, 2001]. In both cases the distribution shape parameter, ξ, is critical in extrapolating to future extremes and is dependent solely on the tail form of the underlying distribution. For $\xi = 0$, describing the classic Gumbel distribution, the underlying distribution is 'light' tailed while for $\xi > 0$ the underlying distribution has 'heavy' tails. In Figure 1, the GP distribution fitted to daily precipitation accumulations greater than 10 mm for Cambridge is shown to have $\xi = 0.11 \pm 0.03$, consistent with the extensive evidence that precipitation is heavy tailed [*Koutsoyiannis*, 2004]. The stretched exponential distribution is however formally attracted to the $\xi = 0$ limit. The disparity arises due to the slow convergence to the $\xi = 0$ limit where, under standard practical approximations to the asymptotic theory, $\xi > 0$ is appropriate for stretched exponentials with $c < 1$ [*Cook and Harris*, 2004].

[8] Figure 4 shows the normalised fit to a GEV distribution for long daily precipitation records from the UK. The mean maximum likelihood estimate of ξ is 0.1 with a standard deviation of 0.1. Annual maxima from simulations of daily time-series of independent stretched exponential variables with $c = 2/3$ give a mean estimate of $\xi = 0.08$ with a standard deviation of 0.05 for 100% rain days increasing to $\xi = 0.11$ with a standard deviation 0.05 for 27% rain days. The prediction of the stretched exponential tail for the daily distribution of precipitation thus also predicts the apparent 'heavy' tails of observed precipitation extremes. This has a significant effect on, for example, risk assessment. A Gumbel, $\xi = 0$, prediction of a one in a hundred year event changes into a much more likely one in 40 year event for $\xi = 0.1$.

5. Conclusion

[9] The origin of the tail of the distribution of daily precipitation totals has been demonstrated to be the result of the multiplicative properties of the vertical moisture flux and the precipitation efficiency. A direct consequence of this distributional form is the apparent 'heavy' tails observed in extreme value analysis of precipitation. Based on a physical model, supported by a global analysis, we would expect the stretched exponential shape of the tail of the rainfall distributions to be largely unaffected by climate change. The scale of extreme events will however be determined by multiplicative changes in the magnitudes of vertical mass flux, specific humidity and the precipitation efficiency. Our analysis is therefore consistent with projected and observed increases in heavy rainfall events [*Groisman et al.*, 2005].

[10] **Acknowledgments.** This work was supported by a NERC CASE award with the Environment Agency of England and Wales. Precipitation records for the UK have been provided by the UK Meteorological Office through the British Atmospheric Data Centre. Global precipitation records from the Global Daily Climatology Network have been provided by the KNMI (Koninklijk Nederlands Meteorologisch Instituut) web site.

References

Abbs, D. J. (1999), A numerical modelling study to investigate the assumptions used in the calculation of probable maximum precipitation, *Water Resour. Res.*, *35*, 785–796.
Allen, M. R., and W. J. Ingram (2002), Constraints on future changes in climate and the hydrologic cycle, *Nature*, *419*, 224–232.
Alpert, P. (1986), Mesoscale indexing of the distribution of orographic precipitation over high mountains, *J. Clim. Appl. Meteorol.*, *25*, 532–545.
Barry, R. G., and R. J. Chorley (1992), *Atmosphere, Weather and Climate*, Routledge, Boca Raton, Fla.
Biondini, R. (1976), Cloud motion and rainfall statistics, *J. Appl. Meteorol.*, *15*(3), 205–224.
Coles, S. (2001), *An Introduction to Statistical Modelling of Extreme Values*, Springer, New York.
Cook, N. J., and R. I. Harris (2004), Exact and general FT1 penultimate distributions of extreme wind speeds drawn from tail-equivalent Weibull parents, *Struct. Safety*, *26*(4), 391–420.
Fankhauser, J. C. (1988), Estimates of thunderstorm precipitation efficiency from field-measurements in CCOPE, *Mon. Weather Rev.*, *116*(3), 663–684.
Frei, C., C. Schar, D. Luthi, and H. C. Davies (1998), Heavy precipitation processes in a warmer climate, *Geophys. Res. Lett.*, *25*(9), 1431–1434.
Frisch, U., and D. Sornette (1997), Extreme deviations and applications, *J. Phys. I*, *7*(9), 1155–1171.
Groisman, P. Y. A., et al. (1999), Changes in the probability distribution of heavy precipitation: Important indicators of climate change, *Clim. Change*, *42*, 243–283.
Groisman, P. Y. A., et al. (2005), Trends in intense precipitation in the climate record, *J. Clim.*, *18*(9), 1326–1350.
Karl, T. R., R. W. Knight, and N. Plummer (1995), Trends in high frequency climate variability in the twentieth century, *Nature*, *377*, 217–220.
Koutsoyiannis, D. (2004), Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records, *Hydrolog. Sci. J.*, *49*(4), 591–610.
Laherrere, J., and D. Sornette (1998), Stretched exponential distributions in nature and economy: Fat tails with characteristic scales, *Eur. Phys. J. B*, *2*, 525–539.
Magagi, R., and A. P. Barros (2004), Estimation of latent heating of rainfall during the onset of the Indian monsoon using TRMM PR and radiosonde data, *J. Appl. Meteorol.*, *43*, 328–349.
Palmen, E. (1958), Vertical circulation and release of kinetic energy during the development of hurricane Hazel into a extratropical storm, *Tellus*, *10*(1), 1–23.
Sornette, D. (2000), *Critical Phenomena in Natural Sciences*, Springer, New York.
Stevens, D. E., and R. S. Lindzen (1978), Tropical wave-CISK with a moisture budget and cumulus friction, *J. Atmos. Sci.*, *35*, 940–961.
Suppiah, R., and K. J. Hennessy (1998), Trends in total rainfall, heavy rain events and number of dry days in Australia, 1910–1990, *Int. J. Climatol.*, *10*, 1141–1164.
Trenberth, K. E. (1999), Conceptual framework for changes of extremes of the hydrological cycle with climate change, *Clim. Change*, *42*, 327–339.
Trenberth, K. E., A. G. Dai, R. M. Rasmussen, and D. B. Parsons (2003), The changing character of precipitation, *Bull. Am. Meteorol. Soc.*, *84*(9), 1205–1217.
Wilby, R. L., and T. M. L. Wigley (2002), Future changes in the distribution of daily precipitation totals across North America, *Geophys. Res. Lett.*, *29*(7), 1135, doi:10.1029/2001GL013048.
Woolhiser, D. A., and R. J. Roldan (1982), Stochastic daily precipitation models: 2. A comparison of distribution of amounts, *Water Resour. Res.*, *18*, 1461–1468.

R. Toumi and P. S. Wilson, Department of Physics, Imperial College London, Prince Consort Road, London SW7 2BZ, UK. (r.toumi@imperial.ac.uk)