European Geosciences Union
General Assembly 2009

Vienna, Austria, 19 - 24 April 2009

Session: S8.1/AS4.1/NH1.2/NP3.6
Precipitation: from measurement to modelling and application in catchment hydrology

An all-timescales rainfall probability distribution

S.M. Papalexiou and D. Koutsoyiannis
Department of Water Resources and Environmental Engineering, National Technical University of Athens, Greece (www.itia.ntua.gr)

1. Abstract

The selection of a probability distribution for rainfall intensity at many different timescales simultaneously is of primary interest and importance as typically the hydraulic design strongly depends on the rainfall model choice. It is well known that the rainfall distribution may have a long tail, is highly skewed at fine timescales and tends to normality as the timescale increases. This behaviour, explained by the maximum entropy principle (and for large timescales also by the central limit theorem), indicates that the construction of a "universal" probability distribution, capable to adequately describe the rainfall in all timescales, is a difficult task. A search in hydrological literature confirms this argument, as many different distributions have been proposed as appropriate models for different timescales or even for the same timescale, such as Normal, Skew-Normal, two- and three-parameter Log-Normal, Log-Normal mixtures, Generalized Logistic, Pearson Type III, Log-Pearson Type III, Wakeby, Generalized Pareto, Weibull, three- and four-parameter Kappa distribution, and many more. Here we study a single flexible four-parameter distribution for rainfall intensity (the JH distribution) and derive its basic statistics. This distribution incorporates as special cases many other well known distributions, and is capable of describing rainfall in a great range of timescales. Furthermore, we demonstrate the excellent fitting performance of the distribution in various rainfall samples from different areas and for timescales varying from sub-hourly to annual.

2. Distribution's basic characteristics

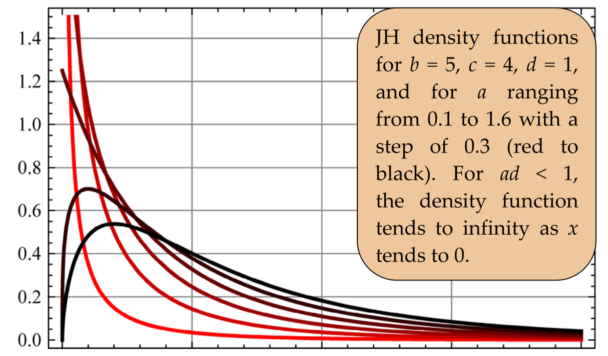
The JH distribution (Papalexiou and Koutsoyiannis, 2008), is an extremely versatile distribution. Its basic statistical characteristics are presented here.

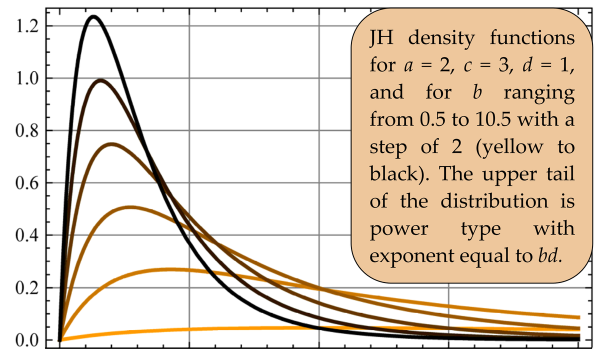
Density function	$f_X(x) = \frac{d}{c\,B(a,b)}\left(\frac{x}{c}\right)^{ad-1}\left[\left(\frac{x}{c}\right)^{d}+1\right]^{-(a+b)}$, where $B(a,b)=\int_0^1 t^{a-1}(1-t)^{b-1}\,dt$
Distribution function	$F_X(x) = I_{\frac{1}{1+\left(\frac{x}{c}\right)^{-d}}}(a,b)$, where $I_z(a,b)=\frac{B_z(a,b)}{B(a,b)}$ and $B_z(a,b)=\int_0^z t^{a-1}(1-t)^{b-1}\,dt$
Quantile function	$Q_X(u) = c\left(\frac{1}{1-I_u^{-1}(a,b)}-1\right)^{\frac{1}{d}}$, where $I_z^{-1}(a,b)$ the inverse of $I_z(a,b)$
Variable range	$X \in \mathbb{R}^+$
Parameter constrains	$a,b,c,d \in \mathbb{R}^+$
q^{th} raw moment	$m_q = \frac{c^q}{B(a,b)}B\left(a+\frac{q}{d}, b-\frac{q}{d}\right)$
Moment existence	$m_q < \infty$ if $b\,d > q$
Asymptotic behaviour	$\Pr\{X > x\} \sim \frac{c^{bd}}{b\,B(a,b)}x^{-bd}$
Shape characteristics	$a\,d < 1 \Rightarrow f_X(0) = \infty$, J shaped density $a\,d = 1 \Rightarrow f_X(0) = \frac{d}{c\,B(a,b)}$ $a\,d > 1 \Rightarrow f_X(0) = 0$, Bell shaped density

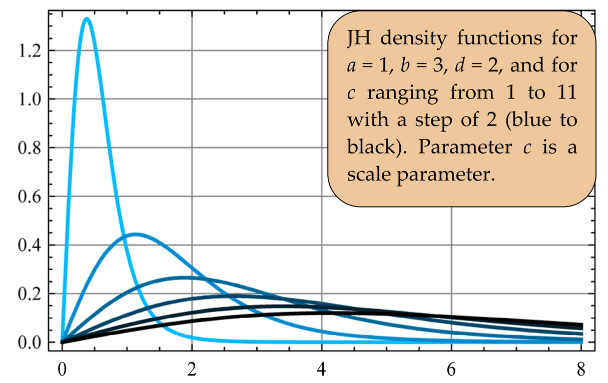
3. Special cases of the JH distribution

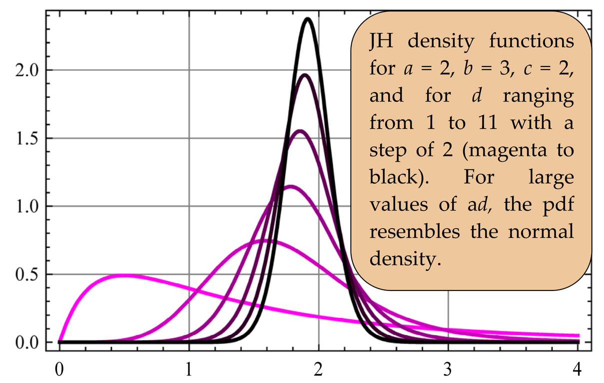
Name	Probability density	Comments
Beta prime	$f(x)=\frac{x^{\alpha-1}(1+x)^{-\alpha-\beta}}{B(\alpha,\beta)}$, $x\ge0,\ \alpha>0,\ \beta>0$	This is a simple transformation of the Beta distribution in order to support the whole real axis.
Burr-type III	$f(x)=\alpha\beta x^{-\beta-1}(1+x^{-\beta})^{-\alpha-1}$, $x\ge0,\ \alpha>0,\ \beta>0$	A flexible distribution introduced by Burr (1942) with many application (see also Rodriguez, 1977).
Burr-Type VII	$f(x)=\alpha\beta x^{\beta-1}(1+x^{\beta})^{-\alpha-1}$, $x\ge0,\ \alpha>0,\ \beta>0$	A simple transformation of the Burr-type III introduced by Burr (1942).
F	$f(x)=\frac{m^{\frac{m}{2}}n^{\frac{n}{2}}}{B\left(\frac{n}{2},\frac{m}{2}\right)}x^{\frac{n}{2}-1}\left(1+\frac{n}{m}x\right)^{-\frac{n+m}{2}}$, $x\ge0,\ m\in\mathbb{N}^*,\ n\in\mathbb{N}^*$	The well-known F-distribution with n and m degrees of freedom used in classical statistics.
Pareto	$f(x)=\frac{1}{\alpha}\left(1+\frac{\beta x}{\alpha}\right)^{-\frac{1}{\beta}-1}$, $x\ge0,\ \alpha>0,\ \beta>0$	The celebrated power-type Pareto distribution with many applications in economy, geophysics and other scientific fields.
Log-Logistic	$f(x)=\frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1}\left[1+\left(\frac{x}{\alpha}\right)^{\beta}\right]^{-2}$, $x\ge0,\ \alpha>0,\ \beta>0$	A similar in shape distribution with the log-normal distribution but with heavier tails.
q-Gamma	$f(x)=\frac{\Gamma[1/(q-1)](q-1)^{\alpha}}{\beta^{\alpha}\Gamma(\alpha)\Gamma[1/(q-1)-\alpha]}x^{\alpha-1}\left[1-(1-q)\frac{x}{\beta}\right]^{1/(1-q)}$, $x\ge0,\ q>1$	Derived by Queiros (2006) based on a possible dynamical scenario using the compounding technique (Beck and Cohen, 2003).
q-Student	$f(x)=\frac{\sqrt{2\beta}\,\Gamma[(1+m)/2]}{\sqrt{\pi(1+m)}\,\Gamma(m/2)}\left(1+\frac{2\beta x^2}{1+m}\right)^{-(1+m)/2}$, $x\in\mathbb{R},\ \beta>0,\ m=\frac{3-q}{q-1},\ 1<q<3$	Derived using the non-extensive entropy by De Souza and Tsallis (1997). Similar with the Student's t.
q-Weibul	$f(x)=\frac{\beta(2-q)}{\alpha^{\beta}}x^{\beta-1}\left[1-(1-q)\left(\frac{x}{\alpha}\right)^{\beta}\right]^{1/(1-q)}$, $x\ge0,\ q>1$	Similar with the Burr-type III but derived using the compounding technique (Picoli et al., 2003).
Student's t	$f(t)=\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}\left(1+\frac{t^2}{v}\right)^{-(v+1)/2}$, $t\ge0,\ v\in\mathbb{N}^*$	The famous Student's t distribution that is the basis of the Student's t-tests.

4. Distribution's flexibility


JH density functions for b = 5, c = 4, d = 1, and for a ranging from 0.1 to 1.6 with a step of 0.3 (red to black). For ad < 1, the density function tends to infinity as x tends to 0.


JH density functions for a = 2, c = 3, d = 1, and for b ranging from 0.5 to 10.5 with a step of 2 (yellow to black). The upper tail of the distribution is power type with exponent equal to bd.


JH density functions for a = 1, b = 3, d = 2, and for c ranging from 1 to 11 with a step of 2 (blue to black). Parameter c is a scale parameter.


JH density functions for a = 2, b = 3, c = 2, and for d ranging from 1 to 11 with a step of 2 (magenta to black). For large values of ad, the pdf resembles the normal density.

5. Parameter estimation

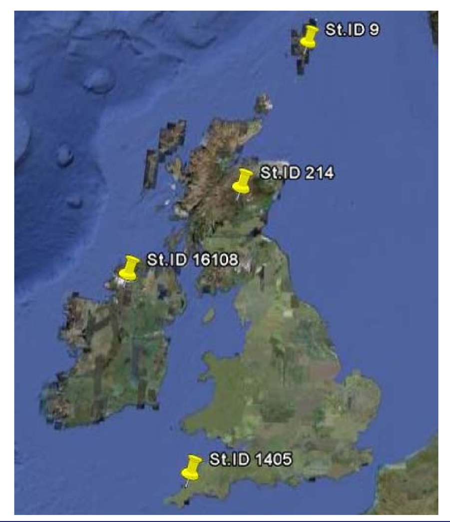
The JH distribution is an extremely versatile power-type, distribution. In its general case, it incorporates four parameters that can be estimated using several different methods. Unfortunately, explicit expressions do not exist—at least for the general four-parameter case—and thus, numerical techniques are necessary. Methods we have tested include:

1. The maximum likelihood estimation (MLE). ML estimates may be obtained by numerically minimizing the log-likelihood function for the observed sample.
2. The least square estimation (LSE). LS estimates may be obtained by numerically minimizing the square error between the empirical distribution function and the theoretical distribution function.
3. The classical method of moments. In general, this method should be avoided in cases of highly skewed samples that are expected to follow a power-type distribution. It is clear, that the sample estimates of higher moments, i.e., third and fourth, based on such samples are extremely sensitive and thus may lead to bad estimates. Furthermore, if the tail is $\Pr\{X > x\} \sim x^{-bd}$ with scaling exponent $bd \ge 3$, then the higher theoretical moments do not exist, whereas finite sample estimates are always obtained.
4. The method of fractional moments. In order to avoid using higher-order moments, fractional moments may be used of orders, e.g., 1/2, 1, 3/2, 2, for which closed expressions exist. The method consists of equating the fractional theoretical moments with their corresponding sample estimates, and solving the resulting equations for the unknown parameters, or minimizing the sum of square errors thereof.
5. The method we used to estimate the JH parameters in this study consists of two parts. First, we estimate the scaling exponent, i.e., $bd = p$, and second, we use the method of fractional moments for orders 2/3, 4/3, and 2 under the constrain $bd = p$.

6. Datasets of the study

St. ID	Station name	Area	Country	Latitude	Longitude	Elevation	Temporal resolution	Start date	End date
NOA	Nat. Observatory	Athens	Greece	37°58'	23°43'	107 m	60 min	01/01/1927	31/12/1996
1405	St. Mawgan	Cornwall	UK	50°26	4°59'	103 m	10 min	25/06/1987	27/05/2003
16108	Ardeemore	Tyrone	UK	54°40'	7°54'	253 m	10 min	01/08/1986	01/04/2005
214	Faskally	Perthshire	UK	56°46'	3°46'	94 m	10 min	03/03/1987	29/03/2003
9	Lerwick	Shetland	UK	60°89'	1°10'	82 m	10 min	18/09/1986	31/12/2002

One hourly rainfall dataset from the National Observatory of Athens, Greece, and four datasets from different regions of the UK, available from the British Atmospheric Data Centre (BACD) are studied. The UK data, originally available as tipping bucket measurements, were converted in 10-min temporal resolution. We aggregated each dataset over several timescales, estimated basic statistics and fitted the JH distribution to assess its performance.

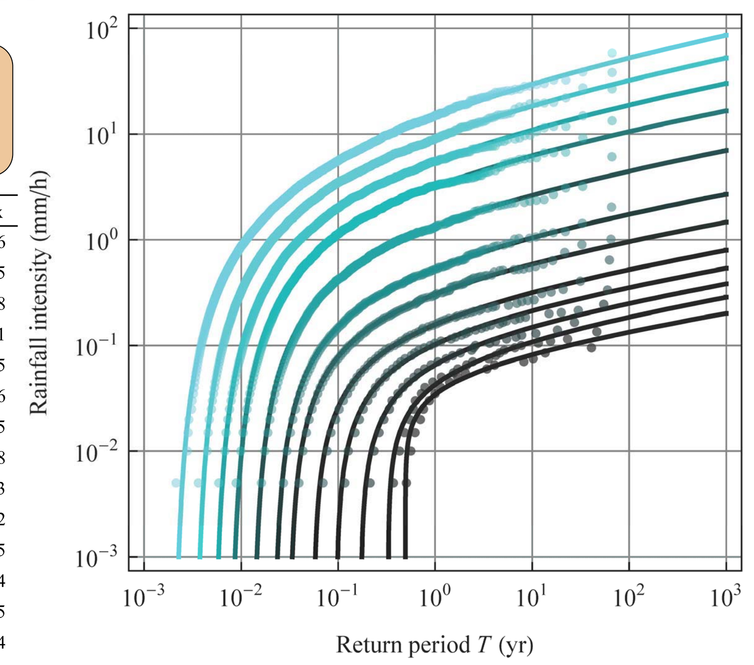


7. Station ID GR-NOA

Statistics of the aggregated (average) rainfall intensity over several timescales. The vast variation of statistics across the timescales is clearly manifested.

Scale	Pr{wet}	Mean	[a]SD	[b]C_V	[c]C_S	[d]C_K	Max
1 h	0.06	0.70	1.71	2.43	7.36	106.89	58.56
2 h	0.07	0.58	1.30	2.24	6.53	87.48	38.25
3 h	0.08	0.51	1.08	2.10	6.34	96.70	33.98
4 h	0.09	0.45	0.92	2.05	5.88	82.05	26.71
6 h	0.10	0.38	0.74	1.96	5.47	68.99	19.25
8 h	0.12	0.33	0.62	1.89	4.73	47.16	13.36
12 h	0.14	0.28	0.50	1.80	5.25	67.87	12.15
24 h	0.21	0.18	0.31	1.69	4.24	40.41	6.08
3 d	0.38	0.10	0.15	1.46	3.43	23.46	2.03
6 d	0.53	0.07	0.09	1.31	2.87	16.49	1.02
15 d	0.74	0.05	0.06	1.13	2.52	15.80	0.65
30 d	0.86	0.04	0.04	1.00	1.66	7.17	0.34
60 d	0.96	0.04	0.03	0.92	1.49	6.80	0.25
120 d	1.00	0.04	0.03	0.76	0.85	3.33	0.14
180 d	1.00	0.04	0.02	0.57	0.45	2.51	0.10
365 d	1.00	0.04	0.01	0.28	0.21	2.42	0.07

[a]Standard deviation
[b]Coefficient of variation
[c]Coefficient of skewness
[d]Coefficient of kurtosis

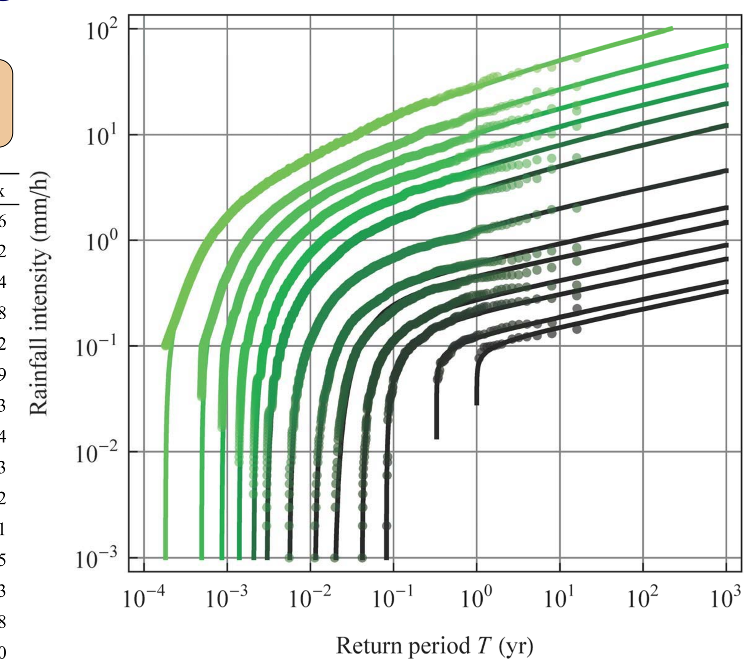


Empirical distributions of the aggregated rainfall intensity over several timescales and the fitted theoretical JH distributions. Timescales from above to below: 1 h, 2 h, 4 h, 8 h, 24 h, 3 d, 6 d, 15 d, 30 d, 60 d, 120 d, 180 d.

8. Station ID UK-1405

Statistics of the aggregated rainfall intensity over several timescales.

Scale	Pr{wet}	Mean	SD	C_V	C_S	C_K	Max
10 min	0.11	1.04	1.69	1.63	6.14	78.93	53.06
20 min	0.11	0.99	1.48	1.50	5.41	67.82	44.22
30 min	0.12	0.94	1.33	1.41	4.40	40.64	28.64
1 h	0.13	0.83	1.09	1.32	3.72	28.96	18.98
2 h	0.16	0.66	0.86	1.30	3.45	27.03	16.72
3 h	0.19	0.56	0.73	1.30	3.56	33.81	15.89
4 h	0.22	0.49	0.65	1.30	3.26	25.15	12.03
6 h	0.27	0.41	0.52	1.27	2.94	18.61	8.04
8 h	0.31	0.36	0.46	1.28	2.86	16.38	6.03
12 h	0.37	0.30	0.36	1.19	2.38	11.43	4.02
24 h	0.49	0.22	0.24	1.08	1.99	8.63	2.01
3 d	0.73	0.15	0.14	0.96	1.36	4.74	0.85
6 d	0.85	0.13	0.11	0.88	1.16	4.00	0.63
15 d	0.98	0.11	0.08	0.76	0.98	3.49	0.38
30 d	1.00	0.11	0.07	0.60	0.68	2.87	0.30
60 d	1.00	0.11	0.05	0.46	0.53	3.26	0.26
120 d	1.00	0.11	0.04	0.33	0.57	3.48	0.23
180 d	1.00	0.11	0.03	0.26	0.07	2.61	0.18
365 d	1.00	0.11	0.02	0.16	0.11	2.52	0.14

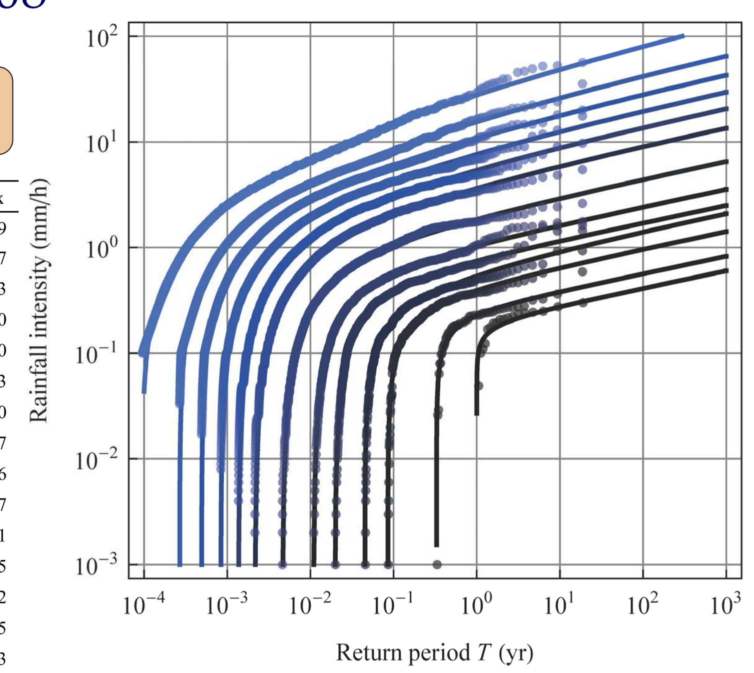


Empirical distributions of the aggregated rainfall intensity over several timescales and the fitted theoretical JH distributions. Timescales from above to below: 10 min, 30 min, 1 h, 2 h, 4 h, 8 h, 24 h, 3 d, 6 d, 15 d, 30 d, 120 d, 365 d.

9. Station ID UK-16108

Statistics of the aggregated rainfall intensity over several timescales.

Scale	Pr{wet}	Mean	SD	C_V	C_S	C_K	Max
10 min	0.20	0.98	1.47	1.49	6.16	99.20	56.49
20 min	0.21	0.95	1.31	1.38	5.08	71.97	52.87
30 min	0.21	0.92	1.21	1.31	4.19	42.59	35.43
1 h	0.23	0.84	1.04	1.23	3.37	25.30	20.00
2 h	0.27	0.72	0.87	1.20	3.05	22.22	17.60
3 h	0.31	0.64	0.77	1.20	2.79	17.65	12.93
4 h	0.34	0.58	0.71	1.21	2.67	14.77	9.70
6 h	0.39	0.51	0.60	1.19	2.46	12.04	6.47
8 h	0.43	0.46	0.54	1.18	2.32	10.52	5.46
12 h	0.49	0.40	0.46	1.14	2.22	9.92	4.07
24 h	0.60	0.33	0.35	1.06	1.85	7.32	2.61
3 d	0.75	0.26	0.24	0.93	1.70	7.88	1.75
6 d	0.83	0.24	0.20	0.85	1.78	10.13	1.62
15 d	0.91	0.22	0.17	0.76	2.33	16.10	1.45
30 d	0.96	0.20	0.13	0.66	1.62	9.22	0.93
60 d	0.99	0.20	0.12	0.62	1.91	12.54	0.93
120 d	1.00	0.20	0.09	0.47	1.10	7.64	0.59
180 d	1.00	0.20	0.08	0.41	0.32	4.57	0.45
365 d	1.00	0.19	0.06	0.33	-0.63	2.81	0.30

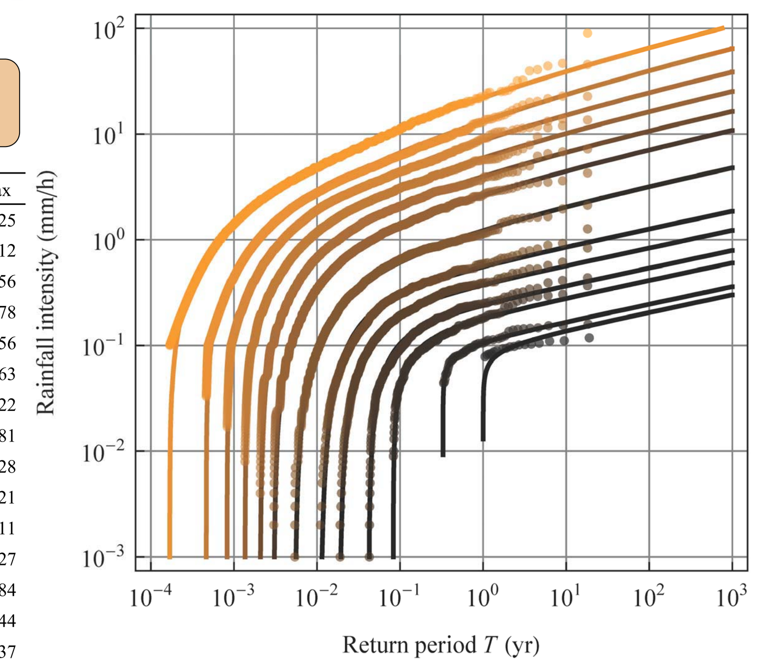


Empirical distributions of the aggregated rainfall intensity over several timescales and the fitted theoretical JH distributions. Timescales from above to below: 10 min, 30 min, 1 h, 2 h, 4 h, 8 h, 24 h, 3 d, 6 d, 15 d, 30 d, 120 d, 365 d.

10. Station ID UK-214

Statistics of the aggregated rainfall intensity over several timescales.

Scale	Pr{wet}	Mean	SD	C_V	C_S	C_K	Max
10 min	0.11	0.85	1.32	1.56	8.82	283.34	90.25
20 min	0.12	0.81	1.17	1.44	5.77	87.65	45.12
30 min	0.12	0.78	1.08	1.39	6.01	115.11	45.56
1 h	0.14	0.69	0.90	1.30	4.13	41.68	22.78
2 h	0.17	0.57	0.72	1.27	3.63	30.94	13.56
3 h	0.20	0.49	0.62	1.27	3.25	22.66	9.63
4 h	0.22	0.43	0.55	1.27	3.02	18.43	7.22
6 h	0.26	0.36	0.46	1.28	2.78	14.49	4.81
8 h	0.30	0.32	0.41	1.29	2.77	14.26	4.28
12 h	0.37	0.26	0.33	1.27	2.55	11.97	3.21
24 h	0.51	0.19	0.23	1.21	2.32	10.75	2.11
3 d	0.75	0.13	0.14	1.08	1.97	8.95	1.27
6 d	0.88	0.11	0.10	0.96	1.63	6.97	0.84
15 d	0.98	0.10	0.08	0.77	1.24	4.72	0.44
30 d	0.99	0.10	0.06	0.63	1.39	5.63	0.37
60 d	1.00	0.10	0.04	0.45	0.45	2.47	0.20
120 d	1.00	0.10	0.03	0.31	0.38	2.42	0.16
180 d	1.00	0.10	0.02	0.25	0.41	3.25	0.16
365 d	1.00	0.10	0.01	0.12	0.19	2.21	0.12

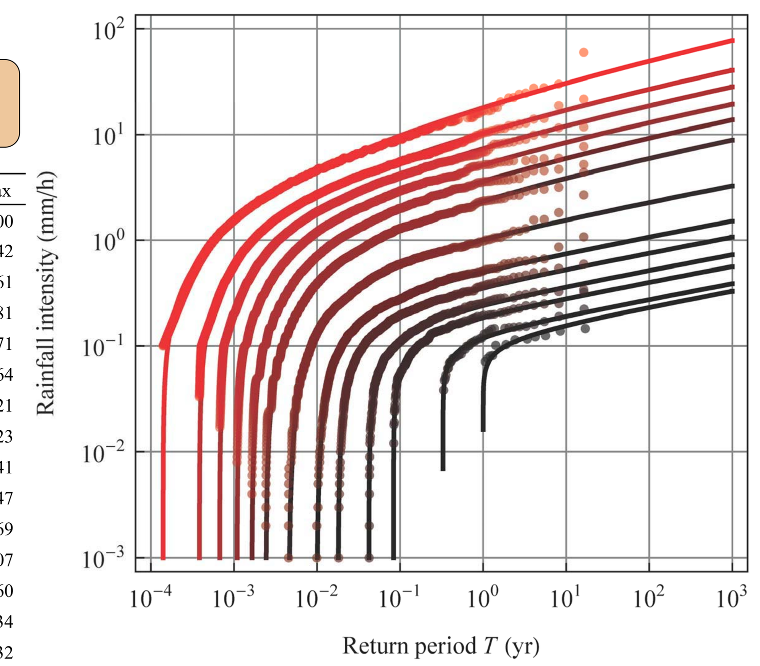


Empirical distributions of the aggregated rainfall intensity over several timescales and the fitted theoretical JH distributions. Timescales from above to below: 10 min, 30 min, 1 h, 2 h, 4 h, 8 h, 24 h, 3 d, 6 d, 15 d, 30 d, 120 d, 365 d.

11. Station ID UK-9

Statistics of the aggregated rainfall intensity over several timescales.

Scale	Pr{wet}	Mean	SD	C_V	C_S	C_K	Max
10 min	0.14	0.79	1.10	1.39	5.67	116.51	60.00
20 min	0.14	0.75	0.98	1.30	4.12	44.96	32.42
30 min	0.15	0.72	0.90	1.26	3.71	32.34	21.61
1 h	0.17	0.63	0.77	1.21	3.19	20.83	11.81
2 h	0.21	0.51	0.62	1.21	2.84	15.33	7.71
3 h	0.25	0.43	0.54	1.24	2.86	15.49	6.64
4 h	0.28	0.38	0.48	1.24	2.79	14.44	5.21
6 h	0.33	0.32	0.40	1.25	2.74	13.82	4.23
8 h	0.38	0.28	0.35	1.22	2.71	14.66	4.41
12 h	0.45	0.24	0.28	1.19	2.62	14.74	3.47
24 h	0.60	0.18	0.19	1.10	2.49	16.11	2.69
3 d	0.83	0.13	0.12	0.90	1.69	8.09	1.07
6 d	0.91	0.12	0.09	0.78	1.25	5.24	0.60
15 d	0.97	0.11	0.07	0.63	0.86	3.46	0.34
30 d	0.98	0.11	0.06	0.51	0.71	3.37	0.32
60 d	0.99	0.11	0.05	0.44	0.69	2.89	0.26
120 d	1.00	0.11	0.04	0.38	0.57	2.91	0.22
180 d	1.00	0.11	0.03	0.31	0.51	2.43	0.19
365 d	1.00	0.11	0.02	0.20	-0.07	2.26	0.15



Empirical distributions of the aggregated rainfall intensity over several timescales and the fitted theoretical JH distributions. Timescales from above to below: 10 min, 30 min, 1 h, 2 h, 4 h, 8 h, 24 h, 3 d, 6 d, 15 d, 30 d, 120 d, 365 d.

12. Conclusions

- A single flexible four-parameter distribution for rainfall intensity (the JH distribution) is studied and its basic statistics are derived.
- The distribution is very flexible with a power-type tail while its density varies from J shaped, unbounded in zero, to bell shaped that resembles the Gaussian curve.
- This distribution incorporates as special cases many other well-known distributions, e.g., Burr, Beta prime, Student, F and q-distributions.
- The closed forms of the expressions of the distribution's moments for integer and fractional order, allows a robust parameter estimation by replacing the highly uncertain high order moments by lower order fractional moments.
- The application of the distribution with rainfall data from different locations in Europe, shows an excellent performance in describing rainfall for different climates and for a wide range of timescales (from sub-hourly to annual).
- The excellent performance at multiple timescales suggests that the distribution may be used in the construction of theoretically consistent ombrian curves (also known as IDF curves).

References

Beck, C., and E. G. D. Cohen (2003), Superstatistics, *Physica A: Statistical Mechanics and its Applications*, *322*, 267-275, doi:10.1016/S0378-4371(03)00019-0.
British Atmospheric Data Centre (2006), UK Meteorological Office. MIDAS Land Surface Stations data (1853-current), [online] Available from: http://badc.nerc.ac.uk/data/ukmo-midas/ (Accessed 8 April 2009)
Burr, I. W. (1942), Cumulative frequency functions, *The Annals of Mathematical Statistics*, *13*, 215-232.
De Souza, A. M. C., and C. Tsallis (1997), Student's t-and r-distributions: Unified derivation from an entropic variational principle, *PHYSICA A*, *236*, 52-57.
Papalexiou, S.M, and D. Koutsoyiannis (2008), Ombrian curves in a maximum entropy framework, in *European Geosciences Union General Assembly 2008*, p. 00702. [online] Available from: http://www.itia.ntua.gr/en/docinfo/851/ (Accessed 8 April 2009)
Picoli, S., R. S. Mendes, and L. C. Malacarne (2003), q-exponential, Weibull, and q-Weibull distributions: an empirical analysis, *Physica A: Statistical Mechanics and its Applications*, *324*(3-4), 678-688, doi:10.1016/S0378-4371(03)00071-2.
Queiros, S. M. D. (2004), On a possible dynamical scenario leading to a generalised Gamma distribution, *physics/0411111*. [online] Available from: http://arxiv.org/abs/physics/0411111 (Accessed 8 April 2009)
Rodriguez, R. N. (1977), A guide to the Burr type XII distributions, *Biometrika*, *64*(1), 129-134, doi:10.1093/biomet/64.1.129.
UK Meteorological Office. MIDAS Land Surface Stations data (1853-current), British Atmospheric Data Centre, 2006. [online] Available from http://badc.nerc.ac.uk/data/ukmo-midas/ (Accessed 8 April 2009).