



IDF Curve Tool — Formulas and Methodology

Updated Edition

This document gives the mathematical details of every step the IDF Curve tool performs, from raw rainfall input through the construction of intensity-duration-frequency (IDF) curves and the closed-form equations that summarize them. It is intended as a technical companion to the user-facing Tutorial.

The exposition follows the order in which the tool processes data. Each section is self-contained: variables are defined where they first appear, source papers are cited, and one-line caveats are placed under each metric so a reader can see immediately what each number does and does not say.

1. Time Scales and Aggregation

Let the input series be $x(t)$ at time step Δt (5 min, 10 min, 15 min, 30 min, or 60 min for sub-daily input; one day for daily input). The tool aggregates $x(t)$ into a set of derived durations $D = \{d_1, d_2, \dots\}$. For 5-minute input, the standard duration set is:

$$D = \{5, 10, 15, 30 \text{ min}, 1, 2, 3, 6, 9, 12, 18, 24 \text{ h}\}.$$

For hourly input, the sub-hourly entries are omitted and the set starts at 1 hour. For daily input, the set is $\{1, 2, 3, 4, 5, 6\}$ days.

1.1 Rolling-window aggregation

For each duration d , the aggregated series is the rolling sum over a window of size $w = d/\Delta t$:

$$X_d(t) = x(t) + x(t-\Delta t) + \dots + x(t-(w-1)\Delta t).$$

For example, with hourly input ($\Delta t = 1$ h) and a 3-hour aggregation ($w = 3$), the rolling sums are:

$$X_{3h}(03:00) = x(01:00) + x(02:00) + x(03:00),$$

$$X_{3h}(04:00) = x(02:00) + x(03:00) + x(04:00),$$

...

The rolling form makes sure every possible storm window of length d is examined, so the maximum is computed without missing any candidate.

1.2 Intensity

After aggregation, the rolling sum X_d is converted to an intensity (e.g., mm/h) by dividing by d :

$$i_d(t) = X_d(t) / d.$$

2. Maximum Calculation Options



After computing the duration-d rolling intensity series, the tool extracts one maximum per period to form the sample used in frequency analysis. The choice of period is configurable.

2.1 Year type

A period is delimited by a starting month $s \in \{1, \dots, 12\}$. The default is the calendar year ($s = 1$). The water-year convention common in hydrology uses $s = 10$ (October). All other starting months are also supported.

Each period is labeled by the year in which it *ends*. So with $s = 10$, the period October 2024 – September 2025 is labeled 2025 (USGS water-year 2025).

2.2 Maximum types

The tool produces one of the following maximum series, indexed by year y :

- **Yearly maximum:** $M_{d,y} = \max\{i_d(t) : t \in \text{year } y\}$.
- **Monthly maximum:** $M_{d,y,m} = \max$ over the chosen month m within year y .
- **Seasonal maximum:** max over a three-month season (see Table 1 below).
- **Custom months:** max over a user-defined subset of months.

The yearly maximum is the most common choice for IDF analysis; it yields the Annual Maximum Series (AMS) used in extreme-value statistics.

2.3 Reference: twelve predefined seasons

When "Seasonal maximum" is selected, the tool offers the twelve overlapping three-month windows defined by their initials.

| Code | Months |
|------|--|
| DJF | December, January, February (Winter) |
| JFM | January, February, March (Winter to Spring transition) |
| FMA | February, March, April (Early Spring) |
| MAM | March, April, May (Spring) |
| AMJ | April, May, June (Late Spring) |
| MJJ | May, June, July (Early Summer) |
| JJA | June, July, August (Summer) |
| JAS | July, August, September (Summer to Autumn transition) |
| ASO | August, September, October (Early Autumn) |
| SON | September, October, November (Autumn) |
| OND | October, November, December (Late Autumn) |



| Code | Months |
|------|---|
| NDJ | November, December, January (Autumn to Winter transition) |

3. Rainfall Intensity Table

The output of Sections 1 and 2 is the Rainfall Intensity Table: rows indexed by year (or by month/season), columns indexed by duration, cells containing the maximum intensity for that combination. The table is shown on the Input Data tab of the tool and can be exported to Excel. It is the input to the frequency analysis of Section 4.

4. Frequency Analysis: Distributions

Frequency analysis fits a probability distribution $F(x; \theta)$ to the maximum series at each duration d . The T -year quantile is the inverse cumulative distribution function evaluated at the non-exceedance probability:

$$p = 1 - 1/T, \quad x_T = F^{-1}(p; \hat{\theta}).$$

Here $\hat{\theta}$ is the parameter estimate computed from the sample at duration d . The tool offers eight distributions / methods.

4.1 Empirical Plotting Position (EPP)

Non-parametric. Let $x_{(1)} \geq x_{(2)} \geq \dots \geq x_{(n)}$ be the sample sorted in descending order. The exceedance probability assigned to $x_{(i)}$ uses the Weibull plotting position:

$$p_{exc}(x_{(i)}) = i/(n + 1), \quad T = 1/p_{exc}.$$

A T -year quantile is read directly off the sorted data at the rank whose Weibull p matches $1 - 1/T$. No distribution is fitted, so EPP is fast but bound to the sample.

Caveat. EPP cannot estimate quantiles for $T > n + 1$; it is purely interpolative within the observed range.

4.2 Extreme Value Type I (Gumbel)

Two parameters: location μ and scale $\beta > 0$.

$$F(x; \mu, \beta) = \exp\{ -\exp[-(x - \mu)/\beta] \}.$$

Parameter estimation (method of moments). For sample mean \bar{x} and sample standard deviation s :

$$\hat{\beta} = s \cdot \sqrt{6} / \pi, \quad \hat{\mu} = \bar{x} - 0.5772 \cdot \hat{\beta}.$$

Quantile. For non-exceedance probability p :

$$x_T = \hat{\mu} - \hat{\beta} \cdot \ln(-\ln p).$$



The constant 0.5772... is the Euler–Mascheroni constant γ . The same estimator (mean-based, not median-based) is used by the IDF curve, the bootstrap confidence band, the GoF Table, and the Specs Dist dialog, so the four are mutually consistent.

Caveat. Gumbel assumes a light upper tail and can underestimate large quantiles when the data has heavier tails; consider GEV in that case.

4.3 Gamma

Two parameters: shape $k > 0$ and scale $\theta > 0$. The probability density is:

$$f(x; k, \theta) = x^{k-1} \exp(-x/\theta) / [\Gamma(k) \cdot \theta^k], \quad x > 0.$$

Parameters are estimated by maximum likelihood using the Accord library (the special function $\Gamma(\cdot)$ is computed via the Lanczos approximation). The quantile $x_T = F^{-1}(p; \hat{k}, \hat{\theta})$ is also from Accord.

4.4 Exponential

Single parameter: rate $\lambda > 0$.

$$f(x; \lambda) = \lambda \exp(-\lambda x), \quad x \geq 0.$$

Maximum-likelihood estimator: $\hat{\lambda} = 1/\bar{x}$. Quantile: $x_T = -\ln(1 - p)/\hat{\lambda}$.

4.5 Log-Normal

Let $Y = \ln X$. Assume $Y \sim \text{Normal}(\mu, \sigma^2)$. Then X has density:

$$f(x; \mu, \sigma) = 1 / (x\sigma\sqrt{2\pi}) \cdot \exp\{ -(\ln x - \mu)^2 / (2\sigma^2) \}.$$

Maximum-likelihood estimators are the sample mean and sample standard deviation of $\ln X$. The quantile is $x_T = \exp(\hat{\mu} + \hat{\sigma} \cdot z_p)$, where z_p is the standard normal p -quantile.

4.6 Weibull

Two parameters: shape $k > 0$ and scale $\lambda > 0$.

$$f(x; k, \lambda) = (k/\lambda) (x/\lambda)^{k-1} \exp[-(x/\lambda)^k], \quad x \geq 0.$$

Parameters are estimated by maximum likelihood (via MathNet.Numerics for the fit, then wrapped in Accord for the quantile evaluation). The quantile has the closed form:

$$x_T = \hat{\lambda} \cdot [-\ln(1 - p)]^{1/\hat{k}}$$

4.7 Generalize Extreme Value (GEV) — NEW

Three parameters: location ξ , scale $\alpha > 0$, and shape k . The cumulative distribution is, for $k \neq 0$:

$$F(x; \xi, \alpha, k) = \exp\{ -[1 - k(x - \xi)/\alpha]^{1/k} \},$$

valid where $1 - k(x - \xi)/\alpha > 0$. When $k = 0$ the limit recovers Gumbel.



Parameter estimation (Hosking 1990 L-moments). Sample L-moments are defined from the order statistics. The first three are:

$$\hat{\lambda}_1 = \bar{x}, \quad \hat{\lambda}_2 = (1/2) E[X_{(2)} - X_{(1)}], \quad \hat{\tau}_3 = \hat{\lambda}_3 / \hat{\lambda}_2.$$

In the tool these are computed using the unbiased estimator of Hosking & Wallis (1997) via running sums over the sorted sample.

The shape parameter k is obtained from the L-skewness $\hat{\tau}_3$ using an initial approximation followed by Newton refinement:

$$c = 2/(3 + \hat{\tau}_3) - \ln 2 / \ln 3, \quad k^{(0)} = 7.8590 c + 2.9554 c^2.$$

Newton iterates solve $\tau_3(k) - \hat{\tau}_3 = 0$, where $\tau_3(k)$ is the theoretical L-skewness of GEV at shape k . With \hat{k} in hand:

$$\alpha = \hat{\lambda}_2 k / [(1 - 2^{-k}) \Gamma(1 + k)], \quad \xi = \hat{\lambda}_1 + \alpha (\Gamma(1 + k) - 1) / k.$$

Quantile. For non-exceedance probability p (using $k \neq 0$):

$$x_T = \xi + (\alpha / \hat{k}) \cdot [1 - (-\ln p)^k].$$

When $|\hat{k}| < 10^{-7}$ the Gumbel limit is used instead. Reference: Hosking, J.R.M. (1990) and Hosking & Wallis (1997).

Caveat. Some moments of GEV are finite only for restricted shape: mean for $k < 1$, variance for $k < 1/2$, mode for $k > -1$. Specs Dist flags moments outside these ranges as undefined.

4.8 Log-Pearson Type III (LP3) — NEW

Three parameters specified on the logarithmic scale. Let $Y = \log^{10} X$. Then Y is assumed to follow a Pearson III distribution.

Sample log-moments. Compute on $y_i = \log^{10} x_i$:

$$\hat{\mu} = \bar{y}, \quad \hat{\sigma}^2 = (1/(n - 1)) \cdot \Sigma(y_i - \bar{y})^2, \quad \hat{g} = n \cdot \Sigma(y_i - \bar{y})^3 / [(n - 1)(n - 2) \hat{\sigma}^3].$$

Kite frequency factor. For non-exceedance probability p , let $z = \Phi^{-1}(p)$ and $\kappa = \hat{g}/6$. The Kite (Chow) frequency factor is:

$$K(p, \hat{g}) = z + (z^2 - 1)\kappa + (z^3 - 6z)\kappa^2/3 - (z^2 - 1)\kappa^3 + z\kappa^4 + \kappa^5/3.$$

Quantile. Apply the Kite factor on the log scale, then back-transform:

$$x_T = 10^{\hat{\mu} + K(p, \hat{g}) \cdot \hat{\sigma}}.$$

Reference: U.S. Geological Survey *Bulletin 17B* (1982) and its successor *Bulletin 17C* (2019). The tool implements the Bulletin 17B Kite-factor expansion, the same formula used by the USACE HEC-SSP software and by most state-level flood frequency packages in the United States.

Caveat. Sample log-skew g is an unstable statistic for short records. Bulletin 17C uses a regional skew adjustment; the desktop tool does not, so LP3 should be treated cautiously for $n < 30$.



5. Goodness-of-Fit Metrics

The Goodness-of-Fit (GoF) Table compares all seven distributions on the data at a chosen duration. Four metrics are computed for each.

5.1 Log-likelihood

For a fitted distribution with parameters θ and PDF f , the log-likelihood is:

$$\ell(\theta) = \sum_i \ln f(x_i; \theta).$$

Higher is better. Used as a building block of AIC.

5.2 Akaike Information Criterion (AIC)

k is the number of distribution parameters.

$$AIC = 2k - 2\ell(\theta).$$

Lower is better. AIC trades fit quality against complexity: a distribution with one more parameter must reduce -2ℓ by at least 2 to be preferred. The differences $\Delta AIC = AIC - AIC_{\min}$ are widely interpreted following Burnham & Anderson (2002): $\Delta AIC < 2$ essentially tied; $\Delta AIC > 10$ strong evidence against the higher model.

Caveat. AIC is a relative measure; it tells you which fit is best in this list, not whether that best is good in absolute terms.

5.3 Kolmogorov–Smirnov statistic

Let F be the fitted CDF and F_n the empirical CDF. With $x_{(i)}$ the order statistics:

$$D = \max_i \max(|F(x_{(i)}) - i/n|, |F(x_{(i)}) - (i-1)/n|).$$

Smaller D = closer fit at the worst-fitting point. The KS statistic itself does not penalize parameter count, so two distributions with the same D but different k can have very different AICs.

5.4 KS p-value (Stephens 1970)

For finite n , the KS p-value is approximated using the Stephens (1970) correction:

$$\lambda = (\sqrt{n} + 0.12 + 0.11/\sqrt{n}) \cdot D.$$

The asymptotic Kolmogorov tail distribution is then summed:

$$p = 2 \sum_{j=1}^{\infty} (-1)^{j-1} \exp(-2j^2\lambda^2).$$

The series converges geometrically; the tool truncates at the first term smaller than 10^{-12} (about 5–10 terms in practice). This is the same formula used internally by Accord's KolmogorovSmirnovTest, so the GoF Table and any external KS check on the same fit will agree.



Caveat. When parameters are estimated from the same sample being tested, the KS p-value is conservatively large. The GoF Table uses it for ranking, not for formal hypothesis testing.

6. Confidence Bands via Bootstrap

When "Show 90% CI" is enabled, each return-period curve is surrounded by a 90% bootstrap confidence band. The procedure is non-parametric resampling on the annual maximum series.

6.1 Procedure

Let $x = (x_1, \dots, x_n)$ be the AMS at duration d . Fix Bootstrap $N = B$ (default 1000). For each $b = 1, \dots, B$:

1. x^*_b is a sample of size n drawn *with replacement* from x .
2. Refit the chosen distribution on x^*_b to obtain new parameters θ^*_b .
3. Compute the quantile $q^*_b(T)$ for each selected return period.

The bootstrap distribution $\{q^*_1(T), \dots, q^*_B(T)\}$ summarizes sampling uncertainty in the fitted quantile. The 90% band at T uses the 5th and 95th sample percentiles:

$$\text{Band}(T) = [q^*(T; \alpha/2), q^*(T; 1 - \alpha/2)], \quad \text{with } \alpha = 0.10.$$

Failed refits (NaN or non-positive intensities) are dropped, not zero-filled, so an unstable bootstrap on a small sample does not corrupt the percentiles.

Caveat. Confidence bands describe sampling uncertainty in the fitted quantile estimator, not predictive uncertainty for an individual future storm. An observed annual maximum can lie outside the band even when the fit is correct.

7. IDF Equation Fitting

Once the IDF curves are constructed, the tool fits closed-form intensity equations that summarize them. Two equation forms are produced together. In both, intensity is in the units of the input data, duration t is in hours, and return period T is in years.

7.1 Sherman (per return period)

One fit per selected return period T :

$$i(t; T) = a / (t + b)^n.$$

The fit minimizes the sum of squared residuals in the original intensity units, not in log space. Linearization on log scale gives:

$$\ln i = \ln a - n \cdot \ln(t + b),$$

which is linear in $\ln a$ and n for any fixed b . The tool searches for b in $[10^{-6}, 10 \cdot t_{max}]$ using golden-section optimization, computing the linear-regression best $(\ln a, n)$ for each candidate b and selecting the b that minimizes the original-units SSE.



7.2 Koutsoyiannis (global)

A single fit covering all selected return periods together:

$$i(t, T) = a \cdot T^m / (t + b)^n.$$

Linearization gives:

$$\ln i = \ln a + m \cdot \ln T - n \cdot \ln(t + b),$$

which for fixed b is a three-parameter linear regression of $\ln i$ on the design matrix $X = [1, \ln T, \ln(t + b)]$. The normal equations $X^T X \beta = X^T y$ are solved by Gaussian elimination with partial pivoting on the 3x3 system. The same golden-section search on b finds the global minimum-SSE fit. Reference: Koutsoyiannis, Kozonis & Manetas (1998).

7.3 Fit quality

Two metrics are reported for both Sherman and Koutsoyiannis:

$$R^2 = 1 - \text{SSE}/\text{SST}, \quad \text{SST} = \sum (i_j - \bar{i})^2.$$

$$\text{RMSE} = \sqrt{(\text{SSE} / N)},$$

where N is the number of (duration, T) pairs used in the fit. R^2 close to 1 indicates the closed-form equation tracks the fitted IDF curves well. RMSE is in the original intensity units.

Caveat. R^2 and RMSE measure how well the equation reproduces the fitted IDF curves, not how well those curves predict future rainfall. The latter depends on the underlying distribution choice and the sample size.

8. References

Bulletin 17B (1982). *Guidelines for Determining Flood Flow Frequency*. Interagency Advisory Committee on Water Data, U.S. Geological Survey, Reston, Virginia.

Bulletin 17C (England, J.F., Cohn, T.A., Faber, B.A., Stedinger, J.R., Thomas Jr., W.O., Veilleux, A.G., Kiang, J.E. & Mason Jr., R.R.) (2019). *Guidelines for Determining Flood Flow Frequency*. U.S. Geological Survey Techniques and Methods, book 4, chap. B5.

Burnham, K.P. & Anderson, D.R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd ed.). Springer-Verlag.

Hosking, J.R.M. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society B*, 52(1), 105–124.

Hosking, J.R.M. & Wallis, J.R. (1997). *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press.

Koutsoyiannis, D., Kozonis, D. & Manetas, A. (1998). A mathematical framework for studying rainfall intensity-duration-frequency relationships. *Journal of Hydrology*, 206, 118–135.

Sherman, C.W. (1931). Frequency and intensity of excessive rainfalls at Boston, Massachusetts. *Transactions of the ASCE*, 95, 951–960.



Stephens, M.A. (1970). Use of the Kolmogorov–Smirnov, Cramér-von Mises and related statistics without extensive tables. *Journal of the Royal Statistical Society B*, 32(1), 115–122.

Software libraries used: Accord.NET Framework (KS, AIC, distribution fitting for Gamma, Exponential, Log-Normal, Weibull, Gumbel) (accord-framework.net); MathNet.Numerics (sample statistics, special functions).

9. Further Resources

Tool website: https://agrimetsoft.com/idf_curve

Tutorial videos: <http://www.youtube.com/AgriMetSoft>

User-facing workflow descriptions, warnings, and step-by-step instructions are in the companion document "Tutorial.docx".