

Contents of DISTRIB Help System

What is a Distribution?

Types of Distributions

Using DISTRIB

References

About SMADA and Programs

Registering DISTRIB

Types of Distributions

Normal

Log Normal

3 Parameter Log Normal

Pearson

Log Pearson

Gumbel

Methods of Analysis

Maximum Likelihood

Method of Moments

Using DISTRIB

Using Files

DISTRIB Files

Importing Data

Editing Data and Options

Editing Actual Data

Level of Confidence

Plot Text

Types of Analysis

Summary Comparison

Predictive Analysis

Comparison Analysis

Plotting Analysis

Comparison Plots

Histograms

What is a Distribution?

Understanding distributions can best be done by example. So...

In the study of hydrology it is very advantageous to predict a flood. It is, however, almost impossible to predict if a flood will occur, say, next year. Instead we try to predict the probability of a flood. If a flood has occurred 4 times in the last 100 years then we can very simply state that there is probably a 1 in 25 chance that one will occur next year. However, is the last 100 years representative of the probability? What if one of the floods occurred last year. If we were predicting a flood based on 100 years the year before the flood we would have thought a flood occurred every 33 years and therefore there was a 1 in 33 chance of the flood. To alleviate these inaccuracies why don't we analyze **all** storms that occurred in the last 100 years. Only those storms which are the largest will cause a flood, but now our analysis is based on a lot more data. IN FACT we can probably use only data from the last 10 years to predict a flood even if a flood has not occurred in the last 10 years. Why? Because the data will probably fit a distribution. Rainfall data typically fits any number of distributions and fit them well.

We run DISTRIB and we find our data fits a 2 parameter log normal distribution very nicely. Odds are any storms in the future will also fit nicely. We can predict storms which occur every 100 storms, every 1000 storms, etc.. We have fit our data to a distribution and we can make predictions based on that distribution. You may be familiar with the most common distribution - the normal distribution You know, the one that is shaped like a typical bell curve. Well, surprise! there are many distributions which fit many different types of data. This program will help you fit your data to the distributions available in the program.

Normal Distribution

The probability density function for the normal distribution is:

where: μ = mean of the population of x
 σ = variance of the population of x

The Assumptions of the Normal Distribution are:

1. The variable is continuous
2. Consecutive Values are Independent
3. Probabilities are Stable

The **Prediction** for any value of a normal distribution can be found from:

where K (frequency factor) is the standard normal deviate.

The **Standard Error** of a normal distribution can be found from:

$$\sqrt{\frac{\sigma^2}{n}}$$

Various methods for the estimation of the delta parameter exist. The results for the Prediction and Standard Error using the maximum likelihood and method of moments for the Normal distribution are identical.

Log Normal Distribution (2 Parameter Log Normal)

The probability density function for the Log Normal Distribution is:

where: $y = \ln(x)$ - natural log of x
 μ_y = mean of the population y
 σ_y = variance of the population of y

The Assumptions of the Normal Distribution are:

1. The variable is continuous
2. Consecutive Values are Independent
3. Probabilities are Stable
4. All variables are non-zero

A number of methods have been used to handle 0.0 values. DISTRIB does not perform any of these conversions for you and may crash if you attempt to fit 0 data to a log distribution. To alleviate this problem you may:

1. Add 1.0 to all data
2. Add a small positive value to all data.
3. Substitute 1.0 in place of all 0 data.
4. Substitute a small positive number in the place of all zero readings.
5. Ignore all zero observations.
6. Consider the probability distribution as the sum of the probability mass at 0.0 and a probability distribution over the remainder of the range. This method is described in Jennings and Benson.

3 Parameter Log Normal

The probability density function for the 3 parameter log normal distribution is:

where: $y = \ln(x-a)$ - natural log of $(x-a)$
 μ_y = mean of the population y
 σ_y = variance of the population of y

Pearson Distribution

The Pearson distribution is represented by following probability density function. The mode of this function is a $x = 0$. This equation is a selective case of the three parameter gamma distribution.

■



where δ = difference between mean and mode ($\delta = \mu - X_m$)

X_m = mode of population x

α = Scale parameter of distribution

p_0 = value of $p_x(x)$ at mode

Log Pearson Distribution

Substituting $y=\ln(x)$ for x in the Pearson distribution gives the Log Pearson type III distribution;



where δ_y = difference between mean and mode ($\delta = \mu_y - Y_m$)

Y_m = mode of population y

α = Scale parameter of distribution

p_{y_0} = value of $p_x(y)$ at mode

Gumbel Distribution

The Gumbel distribution (also referred to as Fisher-Tippett Type I, Double Exponential, Gumbel Type I, and Gumbel Extremal distribution) is characterized by the probability density function;

■ ■ ■

where: α = Scale parameter of the distribution
 β = Location parameter of the distribution

Standard Error of Distribution

The standard error can be calculated by the equation.

$$\sqrt{\frac{SSE}{n}}$$

Where SSE = the sum squared error difference between the actual and predicted data.
n = Number of Points.

Standard Error of Prediction

The Standard Error of the Prediction is calculated specific to a prediction at a certain probability. It is important to the calculation of a confidence interval. This statistic is output in the probability analysis and the Comparison Analysis. This statistic should not be confused with the Standard Error of Distribution which is for the entire distribution.

Kolmogorov - Smirnov Statistic

The Kolmogorov-Smirnov Statistic is found by the equation:

$$D = \max_x |P(x) - P_0(x)|$$

where $P(x)$ = Cumulative Probability of of sample events at class limit

$P_0(x)$ = Actual Cumulative Number of Events at class Limit

DISTRIB uses a continuous form of this equation allowing for the calculation of $P(x)$ and $P_0(x)$ using the following equations:

$$P(x) = \frac{i}{n}$$

where i = Actual cumulative number of events with a probability less than predicted for interval.

j = Predicted number of events so $P_0 = 1/n, 2/n, 3/n \dots$

n = Total Number of Events

Chi-Squared Statistic

The Chi- Squared Statistic is calculated by:

where O = Number of Observed data in interval
E = Expected Number of Data in Interval
k = number of intervals

DISTRIB uses a continuous form of this equation where the number of intervals is also the same as number of points ($k = n$). To standardize the data with this form of the chi-squared statistic the calculated value is divided by n before reporting. To get the actual statistic multiply the reported statistic by n. The chi-squared statistic can be calculated based on any number of intervals. To perform by hand use the following steps:

1. Choose a number of intervals which divides evenly by n (i.e. with 60 points choose 5, 6, 10, or 12 intervals).
2. Run a Comparison Analysis.
3. Divide the predicted data into the number intervals (Mark the Output).
4. Count the number of actual data which fit into the divided intervals.
5. Use the equation as shown above.
6. If still confused buy the documentation for an example.

Confidence Interval

A confidence interval for any prediction can be obtained by using the equation

where: X_t = Prediction Value of Event

S_t = Standard Error of Event

t = Standard Normal Deviate Corresponding to Confidence Level

The t statistic will be found from the level of confidence. The level of confidence can be changed but **it must be between 0.50 and 0.999**. The level of confidence can be changed using: Edit, Level of Confidence.

Weibull Probability

Weibull Probability is calculated by the equation:

where i = sorted number of data (1,2, ...,n)
 n = number of data points

A weibull probability can be calculated for every data point in an array of data. The data must be sorted prior (DISTRIB will automatically sort the data).

Summary Analysis

The Summary Analysis allows you to compare the different methods of analysisHLP_DISTRIBUTIONS. An analysis comparison between the predicted and the actual data is performed for all types of deifined distibutions. Three statistics allow you to compare the accuracy of the different methods. The three statistics are;

Standard Error

Kolmogorov-Smirnov Statistic

Chi-Squared Statistic

Predictive Analysis

The Predictive Analysis allows you to compare prediction methods for any probability which you enter. You must enter a probability between 0.0 and 1.0: **0.0 and 1.0 are not valid probabilities**. The output of this routine will be the value associated with the given probability for all distribution methods which converge. If convergence is not achieved then a message stating such will be posted in the view box. The results of this analysis can be printed from the view or copied to the clipboard. The standard error for the probability is also given. A confidence interval of the prediction can also be calculated.

Comparison Analysis

The Comparison Analysis will calculate a prediction and a standard error for every point within an existing data set. A Distribution Type must be specified along with a method of analysis. If the method chosen converges then an output of the actual and predicted data along with the standard error and a confidence value. **Before plotting the analysis you MUST perform a comparison analysis.** The probabilities for the individual points are calculated by Weibull equation.

Maximum Likelihood Estimation

The maximum likelihood estimates the distribution parameters such that product of the likelihoods of the individual events (L) is maximized. In terms of an equation this becomes an estimation of α, β, \dots such that

is maximized.

Method of Moments Estimation

The method of moments uses the calculation of the r th moment about the origin of a distribution.

The probability function $p(x)$ is then directly substituted into the equation and the distribution parameters are solved for directly. For example in the case of a Normal Distribution it can be shown that distribution parameters solve into the form.

$\sqrt{\quad}$

Opening and Saving DISTRIB (*.DST) Files

DISTRIB opens and saves all files with a DST extension. These files are saved in a binary format. Only files saved with DISTRIB in a binary format can be retrieved by DISTRIB. The DISTRIB files will contain all the data and the last comparison analysis performed. If plot titles were entered they will also be included in the file. All DST files will have EXACTLY the same file size. Data can also be imported.

Importing Data

Data can be imported. The data must be in ASCII format with one number per line. There should be no extra spaces. For example if you wish to import a set of data with 9 pieces of data then the file would look like;

***** Beginning of File - Do Not Put this Line in Your File *****

11.23

12.54

14.65

16.87

9.54

11.76

14.32

19.54

12.87

***** End of File Do Not Put this Line in Your File *****

You can create this file using Windows NotePad or any ASCII text editor, be sure there are no extra spaces at the end of your file. Once a file is imported it can be saved in DST format for later use.

Editing Data

Data can be edited using the edit data option. To move from edit field to edit field use the tab key. If no data exists you will first be asked how many data points exist. DISTRIB does not eliminate 0.0 or negative data points, however these data will not allow any type of log transformation distribution to be performed. Use the **next page button** to move from data entry page to the next.

Registration Information for SMADA and available programs

ORDER FORM FOR PROGRAMS

Please include the following information

NAME: _____

ADDRESS: _____

Program	Price per Copy	6% FL Tax	P&H	Unit Price	Copies Ordered	Total Price (Unit Price * Copies)
SMADA (DOS)	82.55	4.95	2.50			
RETEN(DOS)	82.55	4.95	2.50			
OPSEW(DOS)	82.55	4.95	2.50			
REGRESS/HC6	25.95	1.55	2.50			
DISTRIB(DOS)						
HYDGEN (WIN)	25.95	1.55	2.50			
DISTRIB(WIN)	25.95	1.55	2.50			
PONDER(WIN)	25.95	1.55	2.50			
Non-Florida resident exclude 6% sales tax.				TOTAL DUE		_____

Make Check Payable to: **UCF Bookstore**

or pay by credit card:

Please Check One MC _____ VISA _____ Discover _____
Account Number _____ Expiration _____

Signature _____

You will receive the latest update of the program and extensive documentation including example problems and theoretical basis of the program.

Please **Mail Payment to:**

Stormwater Software
University of Central Florida Bookstore
Alafaya Trail
Orlando, FL 32816-2444

Please Mail all other Correspondence/Suggestions to:

attn: M.P. Wanielista
CEE Department
University of Central Florida
Orlando, Florida 32816-2450

Information About Programs Currently Available:

SMADA(DOS) - Stormwater Management and Design Aid

This program assists the user in the creation of rainfall hydrographs, it allows the user to interactively create, save, and retrieve watershed, rainfall, and hydrograph files. It allows for multiple methods of hydrograph generation. The program also allows the user to perform pond, canal, and pipe routing of this hydrograph.

OPSEW(DOS) - OPTimal SEWER design package

OPSEW is a program designed to allow the user to interactively design simple sewer networks. It will print out the Florida DOT spreadsheet used in sewer permits. It can handle partial, full and surcharged flow. It will perform a number of hydraulic calculations, calculate hydraulic grades based on any number of assumptions, and will output to screen, printer, or disk.

RETEN(DOS) - Pollutant, Reuse, Exfiltration, Swale design

Contains design routines including the following:

- Pollutant loading for multiple landuse watersheds
- Design and analysis of exfiltration system
- Design and analysis of water reuse systems
- Design and analysis of swales
- Use of Stochastic and Runoff Volume Formulas for retention design

DISTRIB (DOS) - Fits data to the following random distributions

- Normal
- 2 Parameter Log Normal
- 3 Parameter Log Normal
- Gumbel
- Pearson
- Log Pearson
- USWRC distribution

Calculates Chi-square, Kolmogorov-Smirnov, and Random Statistics. Compares statistical rankings, performs probability based predictions. Plots data and distribution. Requires VGA Graphics.

REGRESS (DOS) - Performs a regression analysis using a variety of models.

It is limited to 2 parameter regressions. performs power and exponential transformations, analysis of regression coefficients, plotting of fit, plot and confidence intervals.

HC6 (DOS) - User interactive Hardy Cross Analysis.

This program allows the user to interactively enter the hardy cross analysis data and to save this to a data file. It also allows for fixed hydraulic grades, pump additions and fixed pipes flows. HC6 allows for up to 1000 nodes.

HYDGEN (WINDOWS) -

Part of the SMADA for Windows series. This program allows the user to generate and plot hydrographs using watershed and rainfall information. Rainfall information can be generated using dimensionless curves, user input, or text import. Watershed infiltration can be estimated using Soil Conservation Service Curve Number (SCS-CN) information or Horton Equation Information. Hydrograph Generation can be performed with SCS procedures, Continuous Convolution, or Santa Barbara Urban Hydrograph methods. Pollutographs can be generated to allow estimation of pollutant loads from hydrograph information. Fully compatible with PONDER (WINDOWS) and all other SMADA for Windows programs.

PONDER (WINDOWS) -

Part of the SMADA for Windows series. This program allows the user to create a detention pond for pollution control. Pond stage storage discharge relationships can be calculated from stage area, stage volume, and weir information. Hydrographs can be easily routed through ponds to generate inflow-outflow hydrographs. Pollutographs can also be routed through ponds. The shareware version of PONDER comes with registration of

HYDGEN.

DISTRIB(WINDOWS) -

The Windows counterpart to DOS version of DISTRIB. DISTRIB was originally written to help water management engineers create rainfall isopleths. Thousands of pieces of rainfall data from hundreds of sites were analyzed for their best distribution fit and from this data the 10 year, 25 year, 50 year, and 100 year return period rainfalls were determined. These rainfalls were used to determine generate the isopleths. DISTRIB has also proven useful in analyzing stream flow data. The distribution analyzed are the same as the DOS version with the exception of the USWRC method. The programmers are amenable to adding distributions upon request.

About the Authors

Ron Eaglin (as of the time of this writing) is a graduate student in Environmental Engineering on that long slow path towards a doctorate. He has worked as a computer consultant, mechanical engineer, nuclear engineer, and a freelance(not the program) coder (that's someone who actually enjoys writing code). He is always looking for good ideas for code and loves to hear from users of his programs with suggestions and ideas. He can be reached by mail at:

Ron Eaglin
3931 Whittington Drive
Orlando Florida, 32817

Dr. M. P. Wanielista is a professor of engineering at the University of Central Florida. He is the primary researcher in Water Resources.

Linda Dansereau - Linda specializes in water resources and writes many of these fantastic help systems. She is a Leo. Need we say more?

