

1 EVT for hydrology

Probability methods have often been used for solving many kinds of hydrological problems. The principal objective of these methods is to find a probabilistic distributions, or models, that are capable of explaining the observed characteristics of physical phenomenon of interest.

2 Hydrometry

Hydrometry: measuring hydrological phenomena

- How to measure (areal) rainfall?
- How to measure runoff?

3 One-hundred-year floods

A one-hundred-year flood is calculated to be the level of flood water expected to be equaled or exceeded every 100 years on average. The 100-year flood is more accurately referred to as the 1% annual exceedance probability flood, since it is a flood that has a 1% chance of being equaled or exceeded in any single year. Similarly, a flood level expected to be equaled or exceeded every 10 years on average is known as a ***ten-year flood***.

Based on the expected flood water level, a predicted area of inundation can be mapped out. This floodplain map figures very importantly in building permits, environmental regulations, and flood insurance.

4 Time of concentration

Time of concentration is a concept used in hydrology to measure the response of a watershed to a rain event. It is defined as the time needed for water to flow from the most remote point in a watershed to the watershed outlet.

It is a function of the topography, geology, and land use within the watershed. Time of concentration is useful in predicting flow rates that would result from hypothetical storms, which are based on statistically-derived return periods.

For many (often economic) reasons, it is important for engineers and hydrologists to be able to accurately predict the response of a watershed to a given rain event. This

can be important for infrastructure development (design of bridges, culverts, etc.) and management, as well as to assess flood risk.

5 Flood Risk Management

5.1 Indicative Floodplain Map

A map that delineates the areas estimated to be at risk of flooding during an event of specified flood probability. Being indicative, such maps only give an indication of the areas at risk but, due to the scale and complexity of the exercise, cannot be relied upon to give precise information in relation to individual sites.

5.2 Likelihood of flooding

A general concept relating to the chance of an event occurring. Likelihood is generally expressed as a probability or a frequency of a flood of a given magnitude or severity occurring or being exceeded in any given year. It is based on the average frequency estimated, measured or extrapolated from records over a large number of years and is usually expressed as the chance of a particular flood level being exceeded in any one year. For example, a 1 in 100 or 1% flood is that which would, on average, be expected to occur once in 100 years, though it could happen at any time.

5.3 Flood zones

The flood zones are defined on the basis of the probability of flooding from rivers and the sea. Because of the generally more dynamic nature of coastal flooding compared to river flooding, a lower probability of coastal flooding is used to define the highest-risk zone.

- Zone A is at highest risk and has a 1 in 100 (or 1%) chance of flooding in any one year from rivers and a 1 in 200 (or 0.5%) chance of flooding from the sea.
- Zone B is at moderate risk of flooding from rivers and the sea and its outer limit is defined by a 1 in 1000 (or 0.1%) chance of flooding in any one year.
- Zone C is the low risk area, with a less than 1 in 1000 ($< 0.1\%$) chance of flooding from rivers, estuaries or the sea in any one year.

The definition of these zones does not, however, take account of the potential for flooding from other sources, such as ground water or artificial drainage systems. Flooding from these sources could occur in any of the zones and as such should always be considered, regardless of zone.

6 How is flood risk measured?

Flood risk is a combination of the likelihood of occurrence and the consequences of a flood occurring. This is normally expressed as:

$$\text{Flood risk} = \text{probability} \times \text{consequences} .$$

Probability is difficult to estimate because it has to take account of the uncertainty of hydrological predictions based on the analysis of many years of flow records. Consequences are also complex to measure in terms of the potential loss of life, damage to property etc, which depend on the vulnerability of the land-use and property affected by the flood.

7 Cumulative frequency analysis

Cumulative frequency analysis is the analysis of the frequency of occurrence of values of a phenomenon less than a reference value. The phenomenon may be time or space dependent. Cumulative frequency is also called frequency of non-exceedance.

Cumulative frequency analysis is done to obtain insight into how often a certain phenomenon (feature) is below a certain value.

This may help in describing or explaining a situation in which the phenomenon is involved, or in planning interventions, for example in flood protection.

8 Extreme Value Theory

Extreme Value Theory is mathematical study of extreme values. Extreme value theory is a branch of statistics dealing with the extreme deviations from the median of probability distributions. The general theory sets out to assess the type of probability distributions generated by processes. Extreme value theory is important for assessing risk for highly unusual events, such as 100-year floods.

The field of extreme value theory was pioneered by Leonard Tippett (1902 – 1985). Tippett was employed by the British Cotton Industry Research Association, where he worked to make cotton thread stronger. In his studies, he realized that the strength of a thread was controlled by the strength of its weakest fibers.

With the help of R. A. Fisher, Tippet obtained three asymptotic limits describing the distributions of extremes. The German mathematician Emil Julius Gumbel codified this theory in his 1958 book ***Statistics of Extremes***, including the Gumbel distributions that bear his name.

9 Frequency Analysis

- Frequency Analysis (*FA*): Probabilistic description of hydrological extremes
- Extraction of extreme variables from data
- Choice of a distribution
- Parameter estimation
- Quantiles + uncertainty

10 Probability Distribution used for Extreme Value Theory

- The GEV distribution
- The Pareto distribution

10.1 The Generalized Extreme Value (GEV) Distribution

The generalized extreme value (GEV) distribution is a family of continuous probability distributions developed within extreme value theory to combine the Gumbel, Frchet and Weibull families also known as type I, II and III extreme value distributions.

By the extreme value theorem the GEV distribution is the limit distribution of properly normalized maxima of a sequence of independent and identically distributed random variables. Because of this, the GEV distribution is used as an approximation to model the maxima of long (finite) sequences of random variables.

In some fields of application the generalized extreme value distribution is known as the FisherTippett distribution, named after R. A. Fisher and L. H. C. Tippett. However usage of this name is sometimes restricted to mean the special case of the Gumbel distribution.

10.2 The Pareto Distribution

In hydrology the Pareto distribution is applied to extreme events such as annually maximum one-day rainfalls and river discharges. The blue picture illustrates an example of fitting the Pareto distribution to ranked annually maximum one-day rainfalls showing also the 90% confidence belt based on the binomial distribution. The rainfall data are represented by plotting positions as part of the *cumulative frequency analysis*.

10.3 The Gumbel Distribution

The extreme value type I distribution has two forms. One is based on the smallest extreme and the other is based on the largest extreme. We call these the minimum and maximum cases, respectively. Formulas and plots for both cases are given. The extreme value type I distribution is also referred to as the Gumbel distribution.

11 using R to study Extreme Value Theory

The statistical programming environment R supports many of the most frequently used EVT analyses.

11.1 Probability distributions supported by R

- The Gumbell dsitribution
- Generalized pareto distribution (GPD)

11.2 Ocmulgee data set

The ocmulgee data frame has 40 rows and 2 columns. The columns contain maximum annual flood discharges, in units of 1000 cubed feet per second, from the Ocmulgee River in Georgia, USA at Hawkinsville (upstream) and Macon (downstream), for the years 1910 to 1949. The row names give the years of observation.

11.3 EVIR

Functions for extreme value theory, which may be divided into the following groups;

- exploratory data analysis,
- block maxima,
- peaks over thresholds (univariate and bivariate),
- point processes,
- gev/gpd distributions.

12 The GWSDAT R package

The GroundWater Spatio-Temporal Data Analysis Tool, or GWSDAT, has been developed by Shell Global Solutions to facilitate the analysis and reporting of trends in groundwater monitoring data. GWSDAT automatically uploads data from MS Excel to generate a user- friendly interface, which allows users to scroll through the groundwater monitoring history of a site, exporting trend plots and graphics as required. The underlying statistical calculations and graphical output are generated using the open-source statistical program R (www.r-project.org)

13 Analyzing Geospatial data

Methods for visualising and handling Spatial data include the R packages *deldir*, *sp*, *splancs* and *maptools*.

- *hydroTSM* - contains functions for management, analysis, interpolation and plotting of time series used in hydrology and related environmental sciences. In particular, this package is highly oriented to hydrological modelling tasks. The focus of this package has been put in providing a collection of tools useful for the daily work of hydrologists.
- *sp* - A package that provides classes and methods for spatial data. The classes document where the spatial location information resides, for 2D or 3D data. Utility functions are provided, e.g. for plotting data as maps, spatial selection, as well as methods for retrieving coordinates, for subsetting, print, summary, etc.
- *maptools* - Set of tools for manipulating and reading geographic data, in particular ESRI shapefiles; C code used from shapelib. It includes binary access to GSHHS shoreline files. The package also provides interface wrappers for exchanging spatial objects with packages such as PBSmapping, spatstat, maps, RArcInfo, Stata tmap, WinBUGS, Mondrian, and others.
- *evd* - this package extends simulation, distribution, quantile and density functions to univariate and multivariate parametric extreme value distributions, and provides fitting functions which calculate maximum likelihood estimates for univariate and bivariate maxima models, and for univariate and bivariate threshold models.

14 Contour Lines

A contour line (also isoline or isarithm) of a function of two variables is a curve along which the function has a constant value. In cartography, a contour line (often just called a "contour") joins points of equal elevation (height) above a given level, such as mean sea level. A contour map is a map illustrated with contour lines, for example a topographic map, which thus shows valleys and hills, and the steepness of slopes. The contour interval of a contour map is the difference in elevation between successive contour lines.