

**Lecturer In Charge: Dr. J.A. Awomeso**

**Dept of Water Res. Magt. & Agromet**

**UNAAB. Abeokuta. Ogun State**

**Nigeria**

Course Title: WMA 409

Week 8-12

## **Application of Statistical methods in Hydrology**

### **A.**

- The types of analysis to be considered are:
  - ❖ Computation of basic statistics
  - ❖ Empirical frequency distributions and cumulative frequency distributions (flow duration curves)
  - ❖ Fitting of theoretical frequency distributions
  - ❖ Time series analysis
    - Moving averages
    - Mass curves
    - Residual mass curves
    - Balances
  - ❖ Regression/relation curves
  - ❖ Double mass analysis
  - ❖ Series homogeneity tests
  - ❖ Rainfall runoff simulation

## **2 Computation of basic statistic**

**Basic statistics are widely required for validation and reporting. The following are commonly used:**

- Arithmetic mean  
✓

$$\bar{X} = \frac{1}{N} \sum_{i=1}^n X_i$$

- Median- the median value of a ranked series  $X_i$
- Mode- the value of  $X$  which occurs with greatest frequency or the middle value of the class with greatest frequency
- Standard deviation- the root mean squared deviation  $S_x$ :

$$S_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{N-1}}$$

- Skewness or the extent to which the data deviate from a symmetrical distribution.

$$C_x = \frac{N}{(N-1)(N-2)} \frac{\sum_{i=1}^n (x_i - \bar{X})(x_i - \bar{X})^3}{S_x^3}$$

- Kurtosis or peakedness of a distribution

$$K_x = \frac{(N^2 - 2N + 3)}{(N-1)(N-2)(N-3)} \frac{\sum_{i=1}^n (x_i - \bar{X})^4}{S_x^3}$$

### 3. Empirical frequency distributions (flow duration curves)

**A popular method of studying the variability of stream flow is through flow duration curves which can be regarded as a standards reporting output from hydrological data processing. Some of their uses are:-**

- In evaluation dependable flows in the planning of water resources engineering projects
- In evaluating the characteristics of the hydropower potential of a river
- In assessing the effects of river regulation and abstractions on river ecology
- In the design of drainage systems
- In flood control studies
- In computing the sediment load and dissolved solids load of a river
- In comparing with adjacent catchments.

A flow-duration curve is a plot of discharge against the percentage of time the flow was equaled or exceeded. This may also be referred to as a cumulative discharge frequency curve and it is usually applied to daily mean discharges.

#### **4. Fitting of Frequency distributions**

##### **4.1 General description**

The fitting of frequency distribution to time sequences of stream flow data is widespread whether for annual or monthly means or for extreme values of annual maxima or minima. The principle of such fitting is that the parameters of the distribution are estimated from the available sample of data, which is assumed to be representative of the population of such data. These parameters can then be used to generate a theoretical frequency curve from which discharges with given probability of occurrence (exceedence or non-exceedence) can be computed. Generically, the parameters are known as location, scale and shape parameters which are equivalent for the normal distribution to:

- Location parameter    mean (first moment)
- Scale parameter        standard deviation (second moment)
- Shape parameter        skewness (third moment)

Different parameters from mean, standard deviation and skewness are used in other distributions. Frequency distributions for data averaged over long periods such as annual are often normally distributed and can be fitted with a symmetrical normal distribution, using just the mean and standard deviation to define the distribution.

Data become increasingly skewed with shorter durations and need a third parameter to define the relationship. Even so, the relationship tends to fit least well at the extremes of the data which are often of greatest interest. This may imply that the chosen frequency distribution does not perfectly represent the population of data and that the resulting estimates may be biased.

Normal or log-normal distributions are recommended for distributions of mean annual flow.

##### **4.2 frequency distributions of extremes**

Theoretical frequency distributions are most commonly applied to extremes of time series, either of floods or droughts. The following series are required:

- Maximum of a series: The maximum instantaneous discharge value of an annual series or of a month or season may be selected. All values (peaks) over a specified threshold may also be selected. In addition to instantaneous values maximum daily means may also be used for analysis.
- Minimum of a series: with respect to minimum the daily mean or period mean is usually selected rather than a instantaneous value which may be unduly influenced by data error or a short lived regulation effect.

The object of flood frequency analysis is to assess the magnitude of a flood of given probability or return period of occurrence. Return period is the reciprocal of probability and may also be defined as the average interval between floods of a specified magnitude.

A large number of different or related flood frequency distributions have been devised for extreme value analysis. These include:

- Normal and log-normal distributions and 3-parameter log-normal
- Pearson type III or (gamma distribution)
- Log-Pearson Type II
- Extreme value type I (Gumbel), II or III and General extreme value (GEV)
- Logistic and general logistic
- Goodrich/Weibull distribution
- Exponential distribution
- Pareto distribution

Different distributions fit best to different individual data sets but if it is assumed that the parent population is of single distribution of all stations, then a regional best distribution may be recommended. A typical graphical output of flood frequency distribution is shown in fig.4

It is clear that there is no single distribution that represents equally the population of annual floods at all stations, and one has to use judgment as to which to use in a particular location depending on experience of flood frequency distributions in the surrounding region and the physical characteristics of the catchment. No recommendation is therefore made here.

A standard statistic which characterizes the flood potential of a catchment and has been used as a 'index flood' in regional analysis is the mean annual flood, which is simply the mean of the maximum instantaneous floods in each year. This can be derived from the data or from distribution fitting. An alternative index

flood is the median annual maximum, similarly derived. These may be used in reporting of general catchment data.

Flood frequency analysis may be considered a specialist application required for project design and is not a standard part of data processing or validation. Detailed descriptions of the mathematical functions and application procedures are not described here. They can be found in standard mathematical and hydrological text or in the HYMOS manual.

## 5. Time series analysis

Time series analysis may be used to test the variability, homogeneity or trend of a streamflow series or simply to give an insight into the characteristics of the series as graphically displayed. The following are described here:

- Moving averages
- Residual series
- Residual mass curves
- Balances

### 5.1 Moving averages

To investigate the long term variability or trends in series, moving average curves are useful. A moving average series  $Y_i$  of series  $X$  is derived as follows:

$Y_i$

Where averaging takes place over  $2M+1$  elements. The original series can be plotted together with the moving average series.

### 5.2 Mass Curves and residual mass curves.

These methods are usually applied to monthly data for the analysis of droughts.

For mass curves, the sequence of cumulative monthly totals are plotted against time. This tends to give a rather unwieldy diagram for long time series and should not be used. Residual mass curves or simply residual series are an alternative procedure and has the advantage of smaller numbers to plot. An example is shown in fig.6 with respect to reservoir design (fig.7), each flow value in the record is reduced by the mean flow and the accumulated residuals plotted against time. A line such as AB drawn tangential to the peaks of the residual mass curve would represent a residual cumulative constant yield that would require a reservoir of capacity CD to fulfil the yield, starting with the reservoir full at A and ending full at B.

### 5.3 Run length and run sum characteristics

Related properties of time series which are used in drought analysis are run-length and run-sum. Consider the time series  $X_1 \dots\dots\dots X_n$  and a constant demand level  $y$  as shown in fig.8. A negative run occurs when  $X_1$  is less than  $y$  consecutively during one or more time intervals. Similarly a positive run occurs when  $x_1$  is consecutively greater than  $y$ . a run can be defined by its length, its sum or its intensity. The means, standard deviation and the maximum of run length and run sum are important characteristics of the time series.

#### 5.4 Storage analysis

Use of sequent peak algorithm can be made for computing water shortage or equivalently the storage requirements without running dry for various draft levels from the reservoir.

#### B. Theoretical Distribution of Floods

The primary objectives of flood frequency analysis are to determine the return periods and then to estimate the magnitude of events for design return periods beyond the recorded range. The three types of distribution commonly used are:

- log Pearson type III Distribution.
- Extreme Value type I (Gumbel) Distribution
- Log Normal Distribution

Detailed explanation of Analytical and Graphical solution of Extreme Value Distribution.

Design Flood from Flood Frequency Analysis.

The probability of an event being equated or exceeded is employed by hydrologists in designing water resources structures.

Discussion of the probability technique for estimating the magnitude and the frequency of the hydrologic event for the safety design of flood control structures.