Some Probabilistic Methods In Civil and Environmental Engineering

1. Background

A. Chances Are

It is very likely that in your experience you have used at least some of the methods presented in this course. For the most part they have been around for some time and have proven their usefulness. But a few may be new to you and may inspire you to look farther into them.

B. The Opposite Methods

Methods of problem solving may be broadly categorized in several ways. Two of the most useful are to distinguish between deterministic methods and probabilistic methods.

Structural calculations offer a good example of the deterministic approach. In those calculations we make use of known or measurable properties of materials and the geometry of the structure being designed or analyzed. The behavior of the structure can be predicted with good accuracy once we know these things and how they are related.

But what about the loads we apply to the structure? There we must make some guesses, or predictions, which are far less accurate than we might like. Wind and seismic loads are two good examples of loads we cannot predict with great certainty. Rather, we depend on the probabilistic approach. We can gather historical data on winds, for example, and come up with conservative estimates for the greatest wind load our structure may need to resist.

Much of Civil and Environmental engineering practice is like this. We need to combine both kinds of problem solving because much of what we deal with is imprecise, random and only somewhat predictable.

2. Learning Objectives

This course has three principle learning objectives.

A. Gaining Confidence in Using Statistical Approaches
B. Developing Your Own Solutions
C. Learning New Methods

If, along the way you learn, or refresh you memory, of statistics that too will be a benefit in your day to day work. The author is not an expert statistician; nor need you be to gain from this course. A working knowledge of statistics is all that is needed.

3. **Why Use Probability?**

Faced with any engineering problem we must first decide how best to go about solving it. Very often there are only a few choices available to us. We must also recognize that gathering the data we need to solve the problem takes time and costs money so we frequently seek a solution which requires the minimum of data. As an example, the Rational Method of storm drain design has survived over a century mainly because it is simple, quick, and requires very little data. It is NOT because it is accurate, or theoretically sound, that it persists.

Probabilistic methods are frequently the method of choice because the data is available. Weather and streamflow records are examples where the data has been gathered by others and made available to us for little or no cost. Studying that data can lead us to answers more quickly than most other methods. Even though the answer we get is not very precise it is often acceptable. Where speed is as, or more, important than precision then statistical approaches make sense.

4. How Many Such Methods are There?

The number of available statistical methods far exceeds the limitations of this course. Only a few can be presented here and these are mostly simple examples but they may suggest others you will fins useful in your work.

A. Conventional Methods

Regression Analysis

One of the commonest ways of analyzing statistical or experimental data is regression analysis. Stated as simply as possible, this just means fitting the data to a mathematical expression. Usually, we have some idea that the data will fit some "formula" and often we even know what kind of formula to expect. When we plot the data, for example, we can often see a pattern. If the pattern looks like a straight line then we conclude the relationship is probably linear and we look for the equation of the straight line that fits the data most closely.

Other kinds of relationships can also be discerned. Here are some of the most common:

Linear: $y = ax + b$

Quadratic: $y = ax^2 + bx + c$

Multivariate: $y = axz + bz + cx + d$

Exponential: $y = ae^x + b$, e = the base of natural logarithms 2.7182

Periodic: $y = a \sin bx + c$

Miscellaneous: $y = a \sin z + be^{x} + cy + d$

In all these equations , y is the independent variable. The variables, x,z, etc. are dependent variables and a,b,c,d,f and so forth are coefficients to be determined by regression analysis.

You will see examples of each of these below.

Before leaving this subject, we should make mention of one particular type of regression analysis. The Log-Pearson, Type III distribution is used so widely in hydrology that it has become nearly a standard and is required for most work done for the federal government in the U.S. The flood frequency analysis discussed below is a product of this type of analysis.

C. Inspired Guesswork

Much of what we do in our work requires us to make guesses and then check them against reality. The Hardy Cross method of analyzing water pipe networks is one such example. We make initial guesses as to how flows will divide at a node and then check that guess against the fact that we know pressure drops must be equal around a pipe loop. The more experienced we are at these initial guesses the better our chances of arriving at a good final answer.

The Rosetta Stone represents some inspired guess work on the part of French and English linguists.

The Rosetta Stone was found in 1799 by French soldiers of Napoleon in Egypt, near the town of Rosetta (also known as Rashid). The Stone was a marker placed at a temple site by King Ptolemy V Epiphanes in 196 B.C. The Stone was the key in deciphering the three languages of ancient Egypt. The tablet contains the same inscription in ancient Greek, demotic script, and hieroglyphs,

Both Greek and demotic writings were well known so some of the translation went quickly and easily. But certain parts, those parts contained in Cartouches, did not seem to make sense. But, in 1822 the French Egyptologist Jean-Francois Champollion deciphererd the names of the Pharos which was the key to a nearly complete translation.

The Cartouche of Ahknentomen 12

Champollion had made a guess that the cartouche was used for the names of special people; Pharos and other high ranking personages. That guess proved to be true and the remainder of the mystery quickly unraveled.

Engineers too must often make guesses. When they do however, they cannot proceed as though their guess is correct but must check it against reality. Many iterative solutions in engineering are based on this idea.

5. What are Some of the Best Methods?

In the discussion which follows you will find examples of some, though hardly all, of the useful approaches to solving questions which are not completely predictable.

A. Flood Frequency Analysis

"Estimates of the magnitude and frequency of flood-peak discharges and flood hydrographs are used for a variety of purposes, such as the design of bridges, culverts, and flood-control structures, and for the management and regulation of flood plains. These estimates are often needed at ungaged sites where no observed flood data are available.

"To provide simple methods of estimating flood-peak discharges, the U.S. Geological Survey (USGS) has developed and published regression equations for every State, the Commonwealth of Puerto Rico, and a number of metropolitan areas in the United States. These equations have been compiled into the National Flood Frequency (NFF) Program.

"NFF requires user input of physical and climatic characteristics used as independent variables in the equations. Manual or automated methods for measuring the input parameters are described in documentation provided for each state.

"Output is provided in the NFF user interface, and includes input parameters, peak-flow estimates, standard errors, and equivalent years of record. This output can be saved to a text file or printed. Hydrographs and frequency plots are presented in separate windows. The graphs and data used to create them can be saved to bitmap files or printed.

So, what does this all look like ? Here are some regression equations developed by the USGS for urban areas:

"The following seven-parameter equations and definitions are excerpted from Sauer and others (1983). The equations are based on multiple regression analysis of urban flood-frequency data from 199 urbanized basins,

$$UQ_2 = 2.35 A^{0.41} SL^{0.17} (RI_2 + 3)^{2.04} (ST + 8)^{-0.65} (13 - BDF)^{-0.32} IA^{0.15} RQ_2^{\ 0.47}$$

standard error of estimate is 38 percent

$$UQ_5 = 2.70 A^{0.35} SL^{0.16} (RI_2 + 3)^{1.86} (ST + 8)^{-0.59} (13 - BDF)^{-0.31} IA^{0.11} RQ_5^{\ 0.54}$$

standard error of estimate is 37 percent

$$UQ_{10} = 2.99 A^{0.32} SL^{0.15} (RI_2 + 3)^{1.75} (ST + 8)^{-0.57} (13 - BDF)^{-0.30} IA^{0.09} RQ_2^{\ 0.58}$$

standard error of estimate is 38 percent

$$UQ_{25} = 2.78 A^{0.31} SL^{0.15} (RI_2 + 3)^{1.76} (ST + 8)^{-0.55} (13 - BDF)^{-0.29} IA^{0.07} RQ_{25}^{\ 0.60}$$

standard error of estimate is 40 percent

$$UQ_{50} = 2.67 A^{0.29} SL^{0.15} (RI_2 + 3)^{1.75} (ST + 8)^{-0.55} (13 - BDF)^{-0.28} IA^{0.06} RQ_2^{\ 0.62}$$

standard error of estimate is 42 percent

$$UQ_{100} = 2.50 A^{0.29} SL^{0.15} (RI_2 + 3)^{1.76} (ST + 8)^{-0.52} (13 - BDF)^{-0.32} IA^{0.06} RQ_{100}^{\ 0.63}$$

standard error of estimate is 44 percent

$$UQ_{500} = 2.27 A^{0.39} SL^{0.16} (RI_2 + 3)^{1.86} (ST + 8)^{-0.54} (13 - BDF)^{-0.27} IA^{0.16} RQ_{500}^{\ 0.63}$$

standard error of estimate is 49 percent where

UQ2, UQ5,... UQ500 are the urban peak discharges, in cubic feet per second (ft3/s), for the 2-, 5-, ... 500-year recurrence intervals;

A is the contributing drainage area, in square miles, as determined from the best available topographic maps; in urban areas, drainage systems sometimes cross topographic divides. Such drainage changes should be accounted for when computing A;

SL is the main channel slope, in feet per mile (ft/mi), measured between points that are 10 percent and 85 percent of the main channel length upstream from the study site (for sites where SL is greater than 70 ft/mi, 70 ft/mi is used in the equations);

RI2 is the rainfall, in inches (in) for the 2-hour, 2-year recurrence interval, determined from U.S. Weather Bureau (USWB) Technical Paper 40 (1961) (eastern USA), or from NOAA Atlas 2 (Miller and others, 1973) (western USA);

ST is basin storage, the percentage of the drainage basin occupied by lakes, reservoirs, swamps, and wetlands; in-channel storage of a temporary nature, resulting from detention ponds or roadway embankments, should not be included in the computation of ST;

BDF is the basin development factor, an index of the prevalence of the urban drainage improvements;

IA is the percentage of the drainage basin occupied by impervious surfaces, such as houses, buildings, streets, and parking lots; and

RQT, are the peak discharges, in cubic feet per second, for an equivalent rural drainage basin in the same hydro-logic area as the urban basin, for a recurrence interval of T years; equivalent rural peak discharges are computed from the rural equations for the appropriate State, in the NFF program, and are automatically transferred to the urban computations.

The basin development factor (**BDF**) is a highly significant variable in the equations, and provides a measure of the efficiency of the drainage basin. It can easily be determined from drainage maps and field inspections of the drainage basin. The basin is first divided into upper, middle, and lower thirds on a drainage map…. Each third should contain about one-third of the contributing drainage area, and stream lengths of two or more streams should be approximately the same in each third. However, stream lengths of different thirds can be different.

"……..Within each third of the basin, four characteristics of the drainage system must be evaluated and assigned a code of 0 or 1. Summation of the 12 codes (four codes in each third of the basin) yields the BDF. The following guidelines should not be considered as requiring precise measurements. A certain amount of subjectivity will necessarily be involved, and field checking should be performed to obtain the best estimates.

1. Channel improvements..If channel improvements such as straightening, enlarging, deepening, and clearing are prevalent for the main drainage channels and principal tributaries (those that drain directly into the main channel), then a code of 1 is assigned. To be considered prevalent, at least 50 percent of the main drainage channels and principal tributaries must be improved to some degree over natural conditions. If channel improvements are not prevalent, then a code of 0 is assigned.

2. Channel linings..If more than 50 percent of the length of the main channels and principal tributaries has been lined with an impervious surface, such as concrete, then a code of 1 is assigned to this characteristic; otherwise, a code of 0 is assigned. The presence of channel linings would obviously indicate the presence of channel improvements as well. Therefore, this is an added factor and indicates a more highly developed drainage system.

3. Storm drains or storm sewers..Storm drains are defined as those enclosed drainage structures (usually pipes), commonly used on the secondary tributaries where the drainage is received directly from streets or parking lots. Many of

these drains empty into open channels; however, in some basins they empty into channels enclosed as box and pipe culverts. Where more than 50 percent of the secondary tributaries within a subarea (third) consists of storm drains, then a code of 1 is assigned to this aspect; otherwise, a code of 0 is assigned.

4. Curb-and-gutter streets..If more than 50 percent of the subarea (third) is urbanized (covered with residential, commercial, and/or industrial development), and if more than 50 percent of the streets and highways in the subarea are constructed with curbs and gutters, then a code of 1 is be assigned to this aspect; otherwise, a code of 0 is assigned. Drainage from curb- and-gutter streets commonly empties into storm drains.

"Estimates of urban flood-frequency values should not be made using the seven-parameter equations under certain conditions. For instance, the equations should not be used for basins where flow is controlled by reservoirs, or where detention storage is used to reduce flood peaks. The equations also should not be used if the rural equations for the region of interest contain independent variables, such as basin development factor, percentage of impervious area, percentage of urban development, or an urbanization index……"

LOCAL URBAN EQUATIONS

"The NFF Program includes additional equations for some cities and metropolitan areas that were developed for local use in those designated areas only. These local urban can be used in lieu of the nationwide urban equations, or they can be used for comparative purposes. It would be highly coincidental for the local equations and the nation-wide equations to give identical results. Therefore, the user should compare results of the two (or more) sets of urban equations, and compare the urban results to the equivalent rural results. Ultimately, it is the user's decision as to which urban results to use.

"The local urban equations are described in this report in the individual summaries of State flood-frequency techniques for States that use the same equations as those that appeared in the previous version of NFF. The local urban equations are described in fact sheets for States that have updated either their rural or urban equations since the previous version of NFF was released (Jennings and others, 1994). In addition, some of the rural reports contain estimation techniques for urban watersheds. Several of the rural reports suggest the use of the nationwide equations given by Sauer and others (1983) and described above.

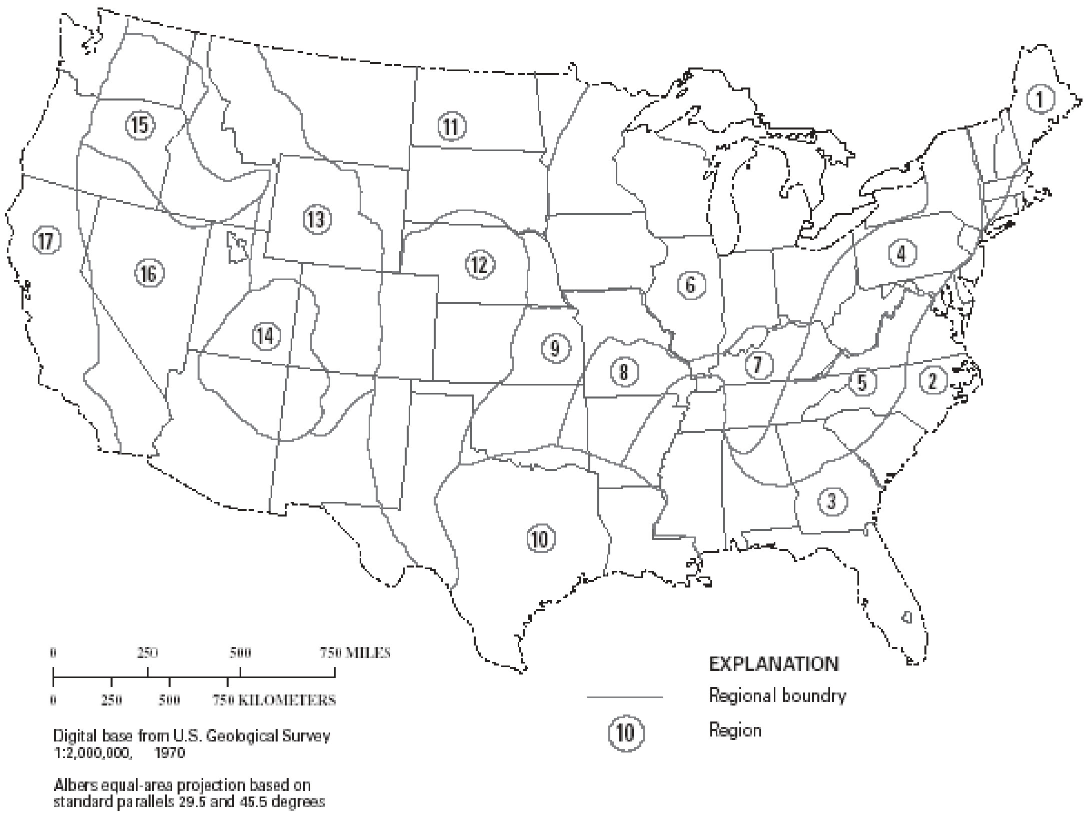
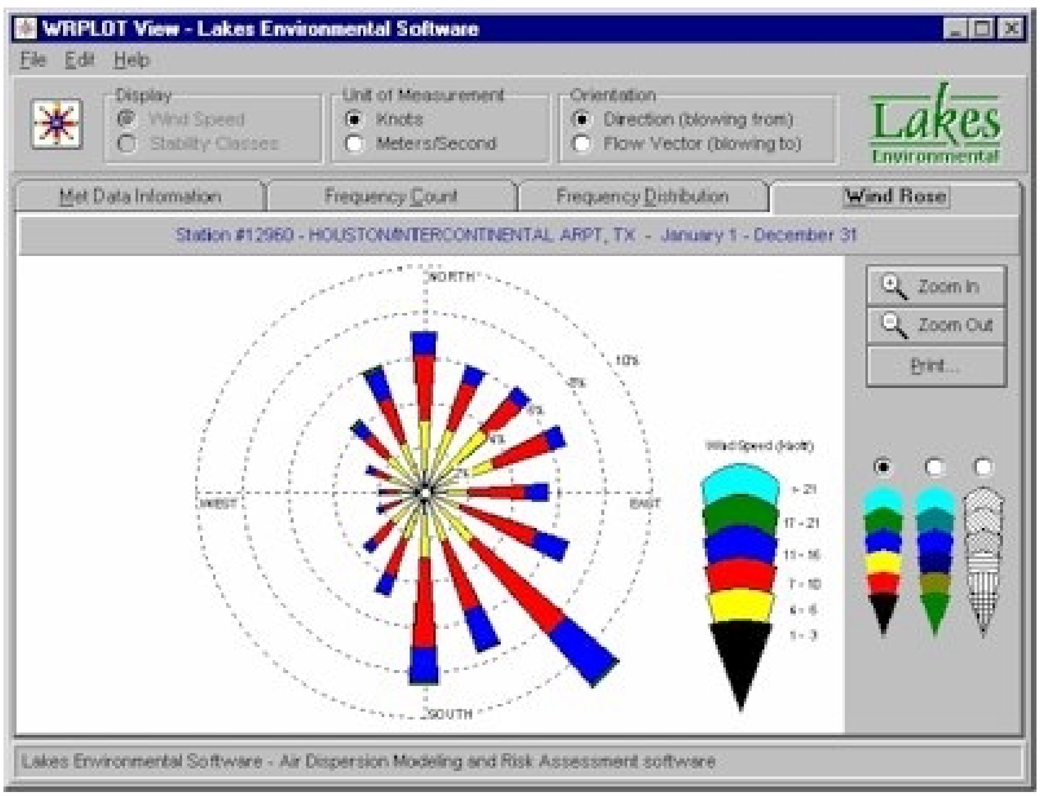


Figure 3. Map of the conterminous United States showing flood-region boundaries (from Crippen and Bue, 1977).

B. Wind Direction and Speed

Airport designers, and those who build wind driven electrical generators, have a considerable need for wind direction and speed data. Fortunately, this information is readily available but needs to be organized in a way that is useful for design. The wind rose is the device most often used.



The graphic above is a screen shot from a commercial wind rose plotter. From it , a designer can easily see how best to orient a runway or runways at a proposed airport, for example. In this case a runway oriented NW to SE and a second runway oriented North-South would probably best serve by being available for takeoff or landing a very large percentage of the time.

If your task is to "aim" a wind driven generator, then SE would also be the direction of choice. The program also allows you to calculate the energy of the wind from any direction from which you might be able to make a reasonable forecast of annual power generated.

C. Seismic Forces

Earthquakes, like weather, occur when they want to; not when it might be convenient for us. The map below shows probable peak ground accelerations at various locations in Oregon.

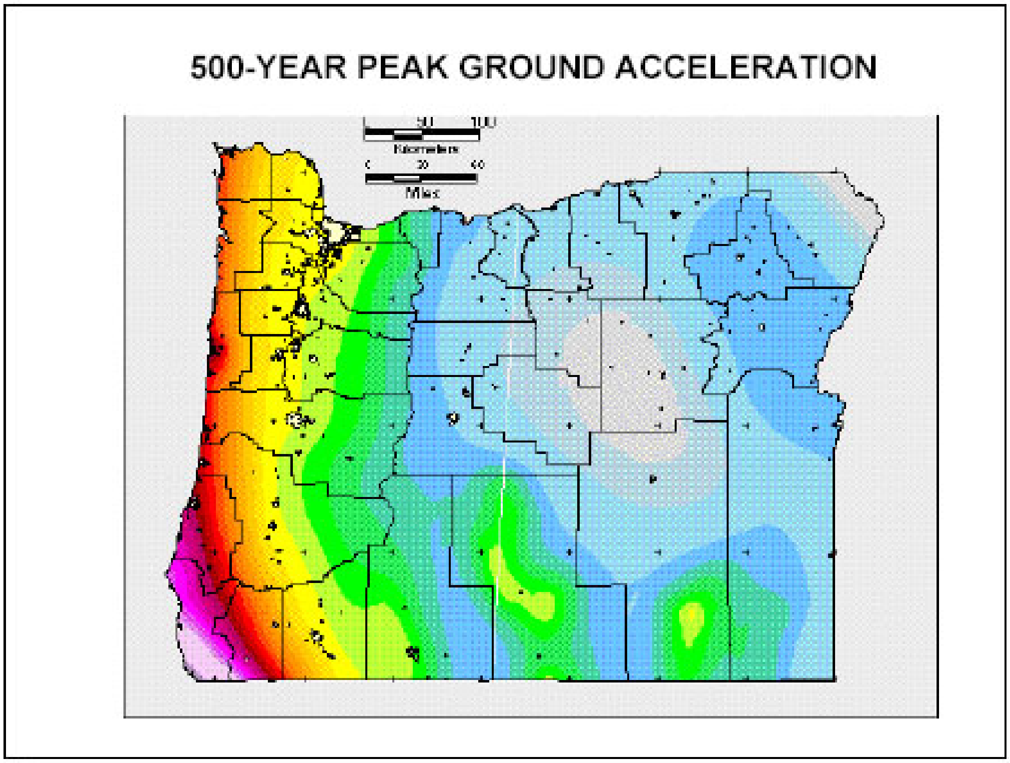


Figure 6

Note that these are based on a statistical analysis very much like rainfall, runoff and other natural phenomena we cannot predict. You will also notice that the accelerations are called "500 Year" peaks. As you'll see in the next section, this terminology can be misleading.

A map similar to this should be available for your area of the world.

D. What Probability of What Flood ?

An Implied Level of Protection

We often speak of the selected return period for design as "the 50 year flood" or "the 100 year flood." Many who hear this understand it to mean that this is a flood expected to occur once every 50 or 100 years. By speaking this way we imply a level of protection which is seldom achieved. For example, if we design for a 100 year storm event many people would believe that would mean the system should function for 100 years without ever being over capacity.

Unfortunately, this is not true. Because flood probabilities are based on historical records, their accuracy depends on the length and completeness of those records. Many reporting stations have only a few years of record so that the probabilities calculated from them are less reliable than stations with a record of 40, 50 or more years. Recently NOAA has updated their precipitation frequency records adding an additional 30 years of record to the data. This should improve the accuracy of the published data and make our estimates more reliable. Even so, they will remain "estimates" only. That is, they are simplified ways of stating the probability of an unpredictable event occurring.

A possibly better way of viewing these events, and speaking about them, is to refer to them by their annual probability of occurrence. The so called 100 Year storm is, by definition, the storm which has a 1 percent probability of occurring in any one year. If we want to know how frequently such a storm might occur over a longer period of time, that probability can be calculated by:

$P_x = 1 - (1-1/N)^x$

Where: P_x is the probability of occurrence in x number of years

1/N= the Probability of Occurrence in any one year

For example, if we want to calculate the probability of occurrence of the 100 year flood over 100 years the calculations would be:

$P_x = 1 - (1-1/100)^{100}$

$P_x = 1 - (1- .01)^{100}$

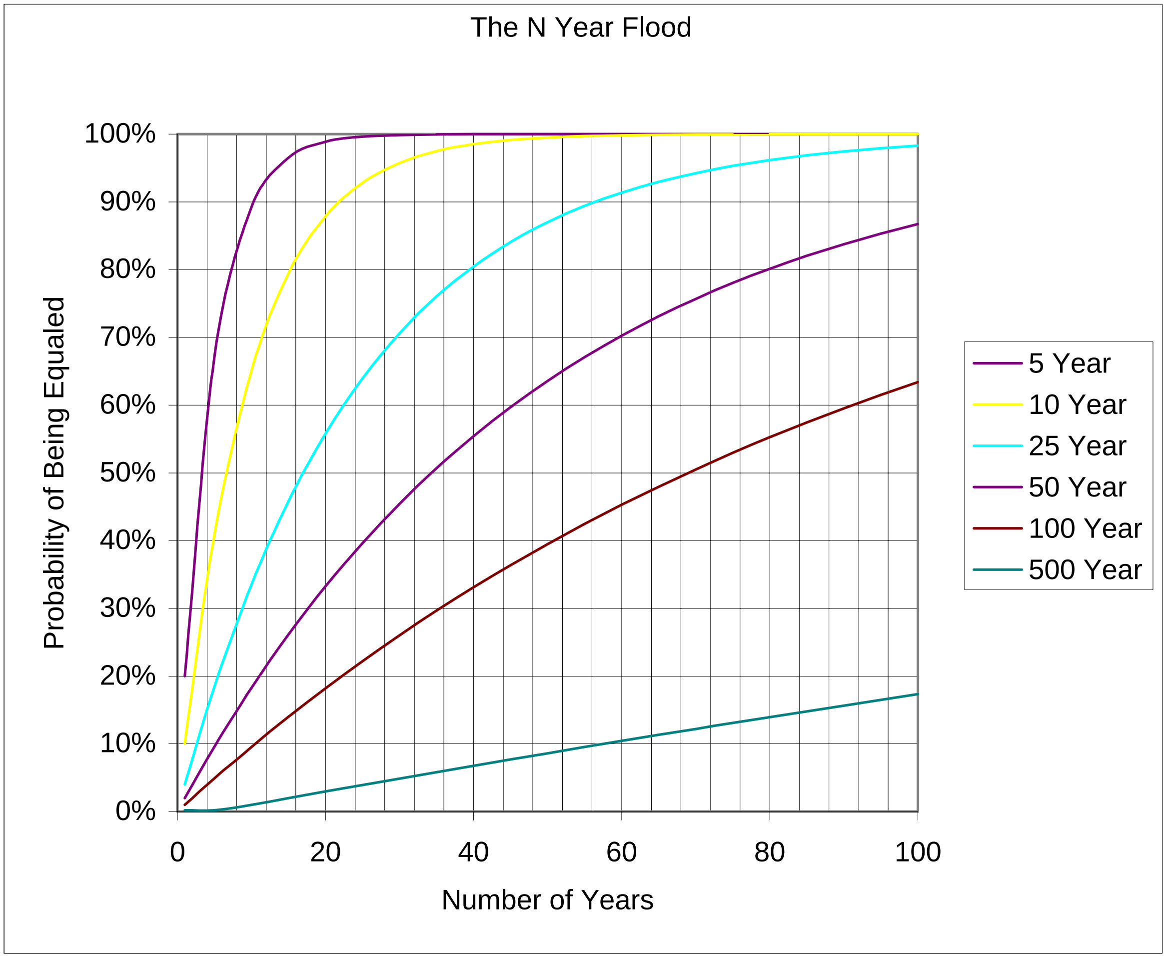
$P_x = 1 - (0.99)^{100}$

$P_x = 1- 0.366$

$P_x = 0.634$

In other words, there is a 63 percent probability that the 1 Percent storm will occur one or more times over the next 100 years.

This kind of calculation can be done for any selected range of frequencies and time periods. This has been done in the figure below.



Probability of the N Year Event Occurring in x Years

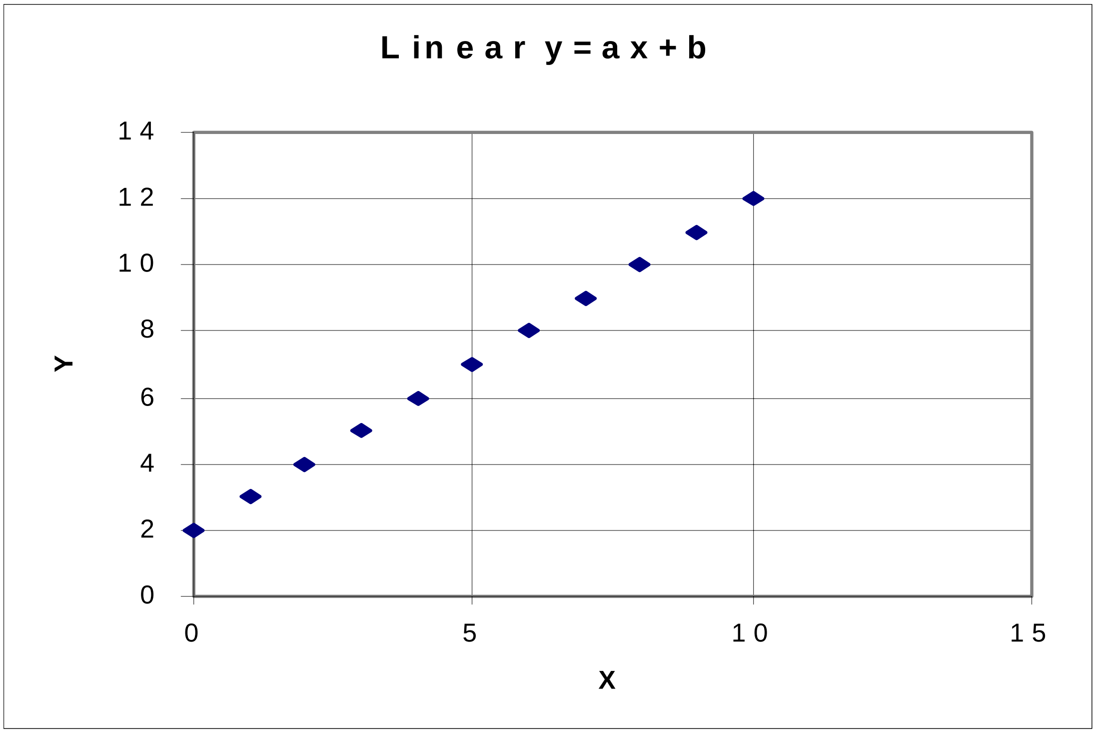
E. Patterns of Plausible Inference

Looking for patterns becomes second nature to engineers and scientists. These patterns, which a logician would call patterns of plausible inference, are not proof of a cause and effect relationship. But they are useful nonetheless because they point the way to such relationships.

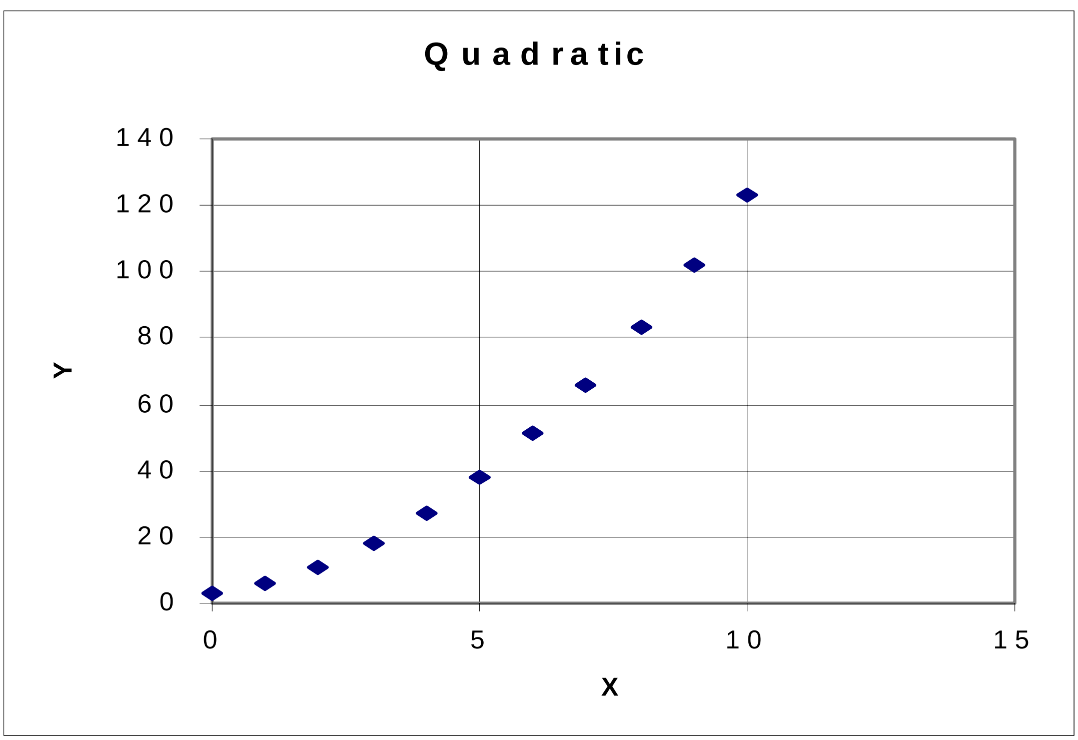
These patterns are most easily visualized by plotting graphs of them.

Return your attention now to the types of equations often encountered in engineering problems, namely:

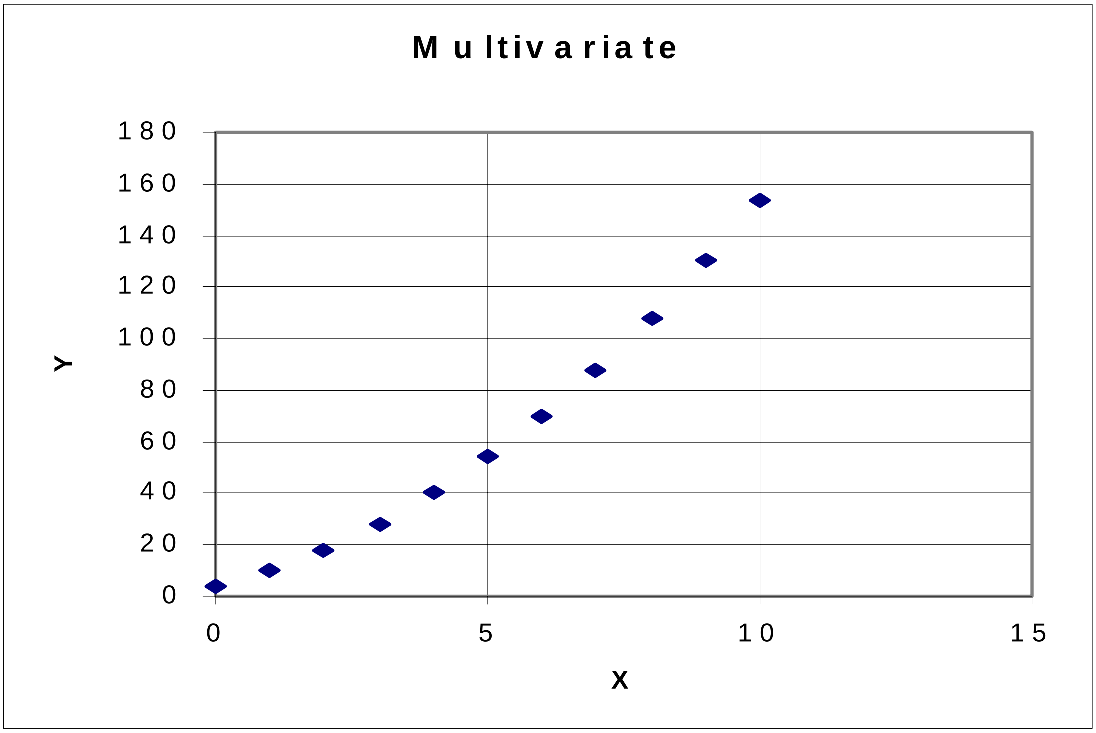
Linear: $y = ax + b$



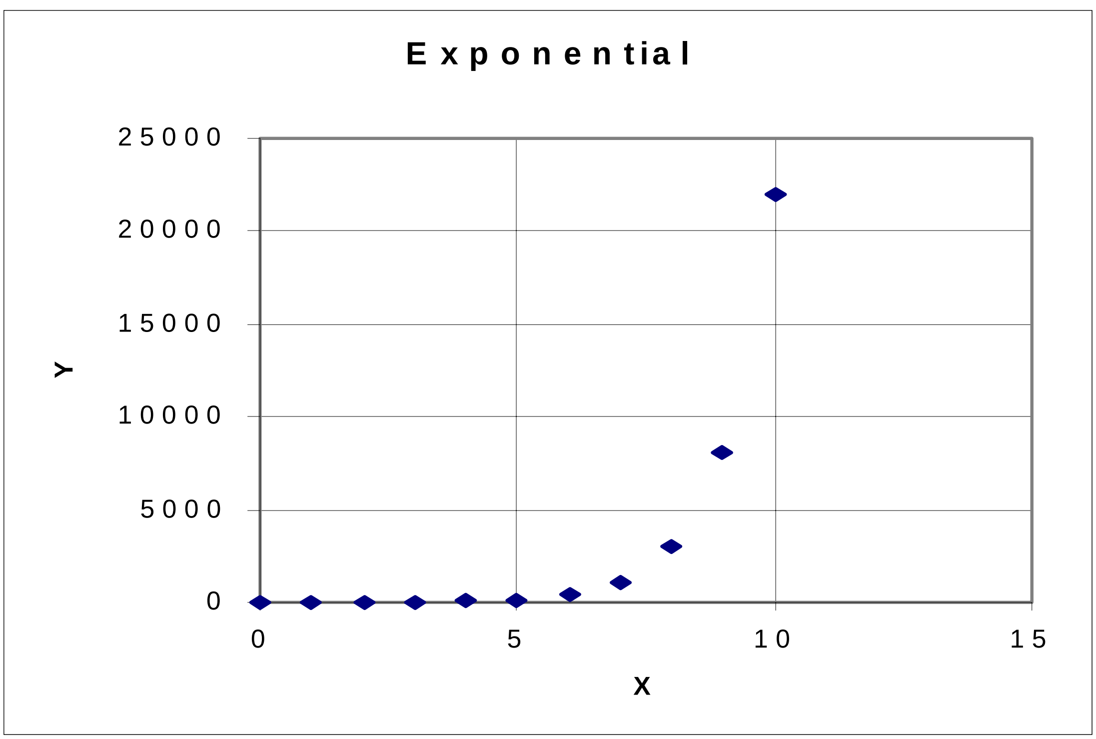
Quadratic: $y = ax^2 + bx + c$

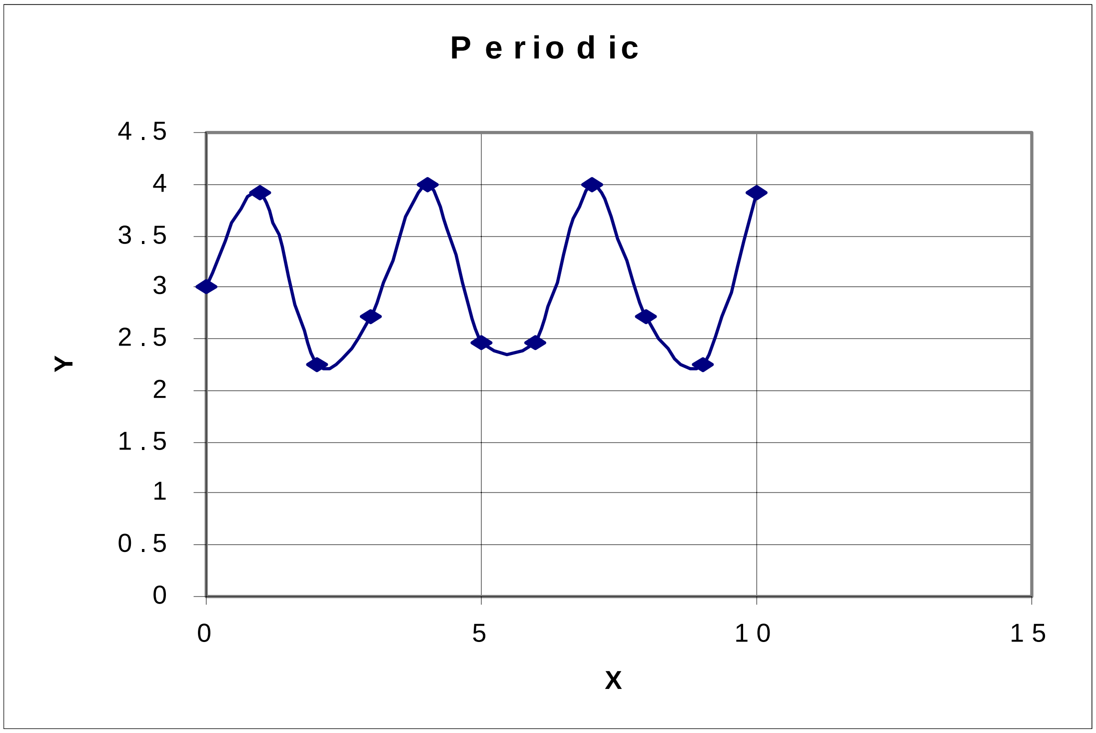


Multivariate: $y = axz + bz + cx + d$



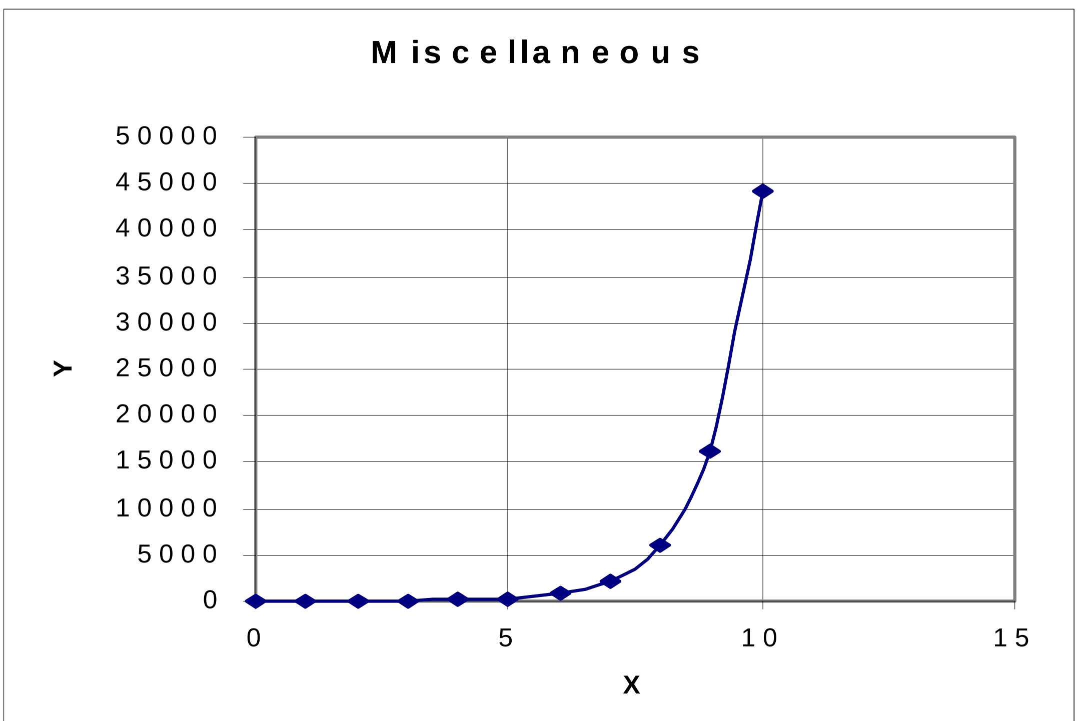
Exponential: $y = ae^{x} + b$, e = the base of natural logarithms 2.7182





Periodic: $y = a \sin bx + c$

Miscellaneous: $y = a \sin z + be^{x} + cy + d$



In addition to all of these, engineers and scientists often try to "linearize" data by plotting it on semi-logarithmic or log-log scales. Very often when this is done the plot will approximate a straight line and the equation for that line can then be used to derive at least an approximate relationship between the variables.

F. What are the Chances of Failure

Here's an example to illustrate how easily we can fool ourselves. The County Engineer of Imaginary County (a very conservative fellow) has arbitrarily decreed criteria for stormwater detention facilities as follows:

Design Inflow: 50 year storm event, developed conditions
Allowable Outflow: 10 year storm event, natural conditions
Design Life: 25 years
Risk of overload during design life: 5 percent

Before we begin to design this detention pond let's calculate the risk of overload during the design life using the equation introduced above:

$$P_x = 1 - (1-1/N)^x$$

In Imaginary County we have N= 50 years and x = 25 years. Substituting we have:

$$P_{25} = 1 - (1-1/50)^{25}$$

P_{25}= 0.396 or 39.6 percent

This fails to meet the County's risk criteria of 5 percent by quite a bit. In fact, it suggests that the pond could "fail" almost 40 percent of the times when it is most needed.

Now we might ask ourselves what storm event must we really design for to meet the 5 percent requirement? We can use the same equation but this time solve for N, setting Px = 0.05.

$$P_x = 1 - (1-1/N)^x$$

$$0.05= 1-(1-1/N)^{25}$$

……………………………………………..

N = 488 years !

Conservative indeed !

All structures and systems built by man ultimately fail. But knowing how and when that failure is likely to occur is invaluable information and gives us a truer assessment of how well we have designed.

G. Lies, Damned Lies, and Statistics (and Baseball)

Perhaps no group of sports fans love statistics more than baseball fans. The Oakland Athletics have turned this interest into a kind of science which has, in recent years, led to their on-field success despite a very small (by baseball standards) budget.

In his book Moneyball: The Art of Winning an Unfair Game, Michael Lewis describes the insightful use of statistics by A's General Manager Billy Beane.

> " By finding undervalued players and by not paying the market price for superstars. Billy Beane makes his living off the misperceptions of baseball players that other general managers have. He is always selling players at a high price and buying them too cheaply.
>
> " The Oakland A's built a model to explain where runs came from. Lots of people outside baseball did this, but people inside baseball didn't do it. In their model, they said walks, singles, doubles, and triples each have a certain effect on run production. They assigned weight to each kind of event. They tested this model. If they have so many of this or that, they have so many runs by the end of the year. These various components have extreme value. Walks are a lot more valuable than people thought. The A's can find players who are otherwise unexceptional, except for walking, and insert them into the lineup. It's analysis, not numbers alone.

But statistics too can be badly interpreted. Worse yet, as in much advertising, they can mislead to the point of being outright lies.

No engineer can afford to mislead himself or herself , the client, or some plan reviewer at a regulatory agency. Doing so can lead only to mistrust, lost future work and even lawsuits.

H. Low Flows Matter Too

Frequency of Low and High River Flows in NW and SE Britain

Number of days per year when rivers have low water flows (below those occurring 90% of the time)
Number of days per year when rivers have high water flows (above those occurring 90% of the time)

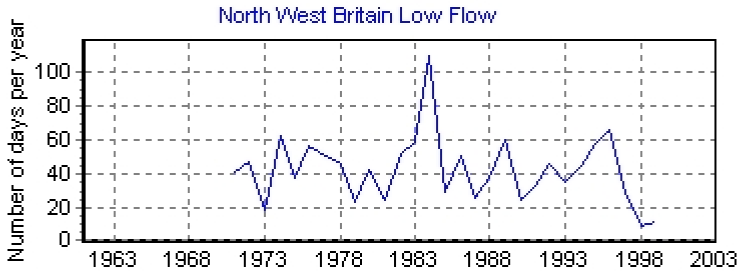
Low and high river flows are obviously important. Low flows not only threaten water supplies, they also reduce effluent dilution, increase the likelihood of algal blooms and damage wetlands and aquatic habitats. …….

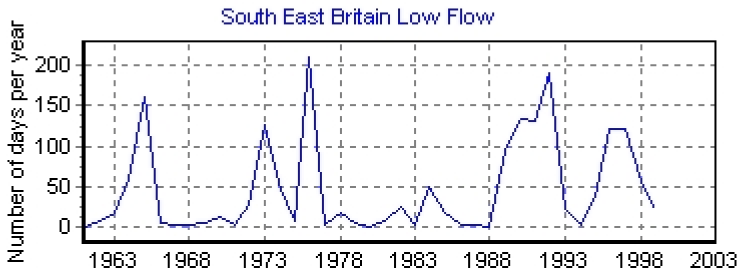
River flows are not related to rainfall alone. Flows in most rivers draining impermeable catchments in NW Britain decline rapidly when there is little rain, whereas, natural groundwater inflows sustain many rivers in SE Britain; however, when groundwater levels are low, river flows may be depressed for lengthy periods. In many catchments, climate-driven changes in the duration of low flows can be difficult to distinguish from changes resulting from land use change, inter-basin transfers, abstractions and effluent returns, and low flow augmentation programmes.

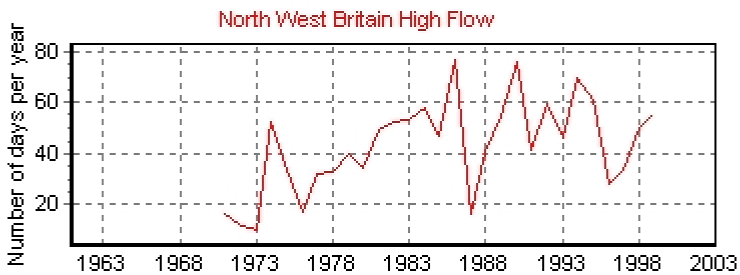
Low river flows can be especially protracted in SE Britain where - in permeable catchments - they are often the result of very low groundwater levels. Records for the River Thames and elsewhere reveal no long-term trends but show extended periods of very low flows (e.g. 1890-1910 and in the 1930s). Low flow episodes were also very prevalent in the 1990s but exceptional autumn and early winter rainfall in 2000 heralded an exceptionally sustained period during which flows remained above the low flow threshold.

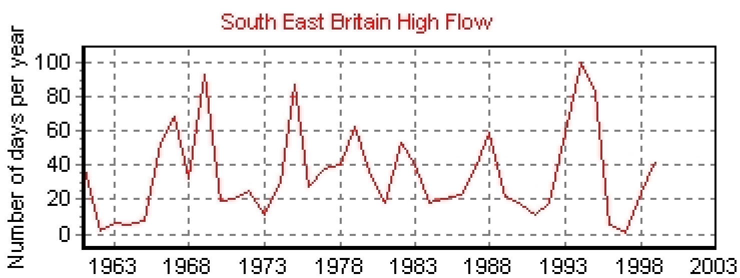
In NW Britain, where groundwater makes only a modest contribution to flows, the number of low flow days show less variability from year to year. Again no compelling trend is evident but low flows have recently been less prevalent than in the in dry 1970s. If the UK climate becomes drier and warmer, the frequency of low flows may be expected to increase. However, a similar outcome could result if rainfall increases modestly but summers become drier and evaporative demands increase. In the latter circumstances, substantially different responses may be expected in rivers draining permeable catchments (where summer flows will be sustained by increased aquifer recharge in the winter) compared with those draining impermeable catchments.

Below are some typical high and low flow charts covering several years of record on streams in the UK.









I. Measures of Central Tendency

In all statistics, we seek to bring some order out of chaos. One of the best ways to do that is to list our data in some order and then see if it is "clustered" around some central value. Three measures of that central value are:

The mean, the mode and the median.

The mean we all know by its alter ego, the average. Mathematically it is defined as the sum of all the values divided by the number of those values. Of the three it is probably the most useful to engineers and scientists,

The statistical median is the middle number of a group of numbers that have been arranged in order by size. If there is an even number of terms, the median is the mean of the two middle numbers:

To find the median of a group of numbers:

- Arrange the numbers in order by size
- If there is an odd number of terms, the median is the center term.
- If there is an even number of terms, add the two middle terms and divide by 2.

The statistical mode is the number that occurs most frequently in a set of numbers.

To find the mode of a group of numbers:

- Arrange the numbers in order by size.
- Determine the number of instances of each numerical value.
- The numerical value that has the most instances is the mode.
- There may be more than one mode when two or more numbers have an equal number of instances and this is also the maximum instances
- A mode does not exist if no number has more than one instance.

Example: The mode of 2, 4, 5, 5, 5, 7, 8, 8, 9, 12 is 5.

J. Bits and PCs 

By their nature, statistical methods require number crunching. Usually, the more numbers, the better. A 100 year stream flow record will yield better predictions than a shorter one, for example. For this reason, the personal computer development over the last twenty five years has been a large factor in the increasing use of probabilistic methods.

There are many specialized statistical analysis programs on the market but for engineers the best general purpose program remains the spreadsheet. Spreadsheets are the largest selling category of software; and for good reason. No other tool can do so many things so well.

In this brief course we'll make mention of only one such tool; Microsoft's Excel. The following quote from the Excel Help File briefly describes many of that programs features.

"Ways to analyze statistics

"Microsoft Excel provides a set of data analysis tools called the Analysis ToolPak that you can use to save steps when you develop complex statistical or engineering analyses. You provide the data and parameters for each analysis; the tool uses the appropriate statistical or engineering macro functions and then displays the results in an output table. Some tools generate charts in addition to output tables.
"To view a list of available analysis tools, click Data Analysis on the Tools menu. If the Data Analysis command is not on the Tools menu, run the Setup program to install the Analysis ToolPak. After you install the Analysis ToolPak, you must select it in the Add-In Manager. …..

"To use these tools, you need to be familiar with the specific area of statistics or engineering that you want to develop analyses for.

"Note Microsoft Excel provides many other statistical, financial, and engineering worksheet functions. To see a list of available worksheet functions, click Edit Formula on the formula bar, and then click the down-arrow in Insert Function ."

A. A Must Have Add-in

The program XLXTRFUN.XLL © is one you must add to your collection of very useful tools. It provides many functions Excel leaves out including; interpolation, extrapolation, curve fitting, calculation of polynomials which fit a given curve, differentiation, and more.

B. How xlxtrfun.xll works

The program works very much like those mini Windows programs called ".dll " s, or dynamic linked libraries. Written by Scott Allen Rauch, it is shareware and may be downloaded free at his website:

http://www.netrax.net/~jdavita/XlXtrFun/XlXtrFun.htm

C. Example Interpolation using Interp

- *When you want to use this function and accept the default values for the optional parameters, use the Interp function. Interp uses the same numerical routines as this function but assumes the default values for the optional parameters; i.e., Extrapolate? = FALSE, Parabolic? = TRUE, Averaging? = TRUE, and SmoothingPower = 1.*
- *As an alternative to entering TRUE or FALSE for the parameters Extrapolate?, Parabolic?, and Averaging?, 0 (zero) is equivalent to FALSE, and any non-zero number is equivalent to TRUE.*

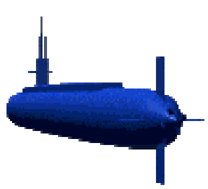
*If you have Excel, c*lick on the icon below to see the "interp" function used to calculate the probability of the 50 year flood occurring in 27 years. Note especially cell F39 which uses the simpler form of interpolation.

Click on the icon below to view the soreadsheet.

Microsoft Excel Worksheet

INTERP.XLS (1 page)

K. Operations Research



Operations research had its beginnings during World War II. In the battle of the North Atlantic Allied shipping was menaced by the German U-boat fleet, at times with devastating effect. A complete history may be found in, Methods of Operations Research, by Phillip M. Moore and George E. Kimball. A few excerpts from that book will help you to understand what Operations Research is, if you have not encountered it before. You may think of it as game theory for the deadly game of warfare.

"In (World War II) operational analysts were to be found at work in strange places and under unlikely circumstances. Mathematicians discussed gunnery problems with British soldiers in Burma.; chemists did bomb damage assessments with economist colleagues …outside London; generals conferred about tank strategy ….with biochemists and lawyers….. "

"What scientists brought to the operational problems - apart from specialized knowledge- was the scientific outlook. This was in fact their major contribution. They tended to think anew, to suspect preconceptions, to act only on evidence. Their indispensable tool was the mathematics of probability and they made use of its subtlest theories and most powerful techniques…………

" Just as with every other field of applied science, the improvement of operations of war by the application of scientific analysis requires a certain flair which comes with practice but which is difficult to put into words…………

" In order to make a start in so complex a subject one must ruthlessly strip away details…. and arrive at a few broad, very approximate 'constants of the operation .'

" It is well to emphasize that these constants which measure the operation are useful even though they are extremely approximate; it might almost be said that they are more valuable *because* they are very approximate. This is because successful application of operations research usually results in improvements by factors of 3 or 10 or more..… Many operations are ineffectively compared….because of a single faulty component; inadequate training of crews, or incorrect use of equipment, or inadequate equipment……. In our first study of any operation we are looking for these large factors of possible improvement. They can be discovered if the constants of the operation are given only one significant figure, and any greater accuracy simply adds unessential detail.

"One might term this type of thinking 'hemibel thinking'. A bel is defined as a unit in a logarithmic scale corresponding to a factor of 10. Consequently a hemibel corresponds to a factor of the square root of 10, or approximately 3."

You've probably been faced with similar tasks. Handed a set of partial, error ridden, unreliable data we are tempted to throw up our hands and believe the task is impossible. But "hemibel thinking" can often save the day by allowing us to draw the most useful information out of the most unpromising material. Don't expect great accuracy but think in terms of orders of magnitude. The results may surprise you.

L. Economies of Scale

Here is a text book definition:

"We see the last generation of economic growth theorists creating a new version of (Adam) Smith's virtuous circle. But Smith's own approach was not entirely abandoned. It was developed early in the twentieth century by such great British economists as Alfred Marshall, A. C. Pigou, and Nicholas Kaldor.

"For these economists, the key concept of economic growth is "economies of scale." Suppose all inputs are increased in the same proportion. For example, we increase the quantity of land, labor and capital in use, each by 20%. Will output increase by the same proportion -- by 20% -- or by more, or by less? Notice that the principle of diminishing returns does not answer that question, since it **assumes** that one input is fixed (and so the fixed input cannot increase in proportion to the rest). When it is possible to increase all inputs in proportion, the principle of diminishing returns is not applicable.
Here is the terminology:

Economies of scale or increasing returns to scale

output increases more than in proportion to inputs

Constant returns to scale

ouput increases just in proportion to inputs

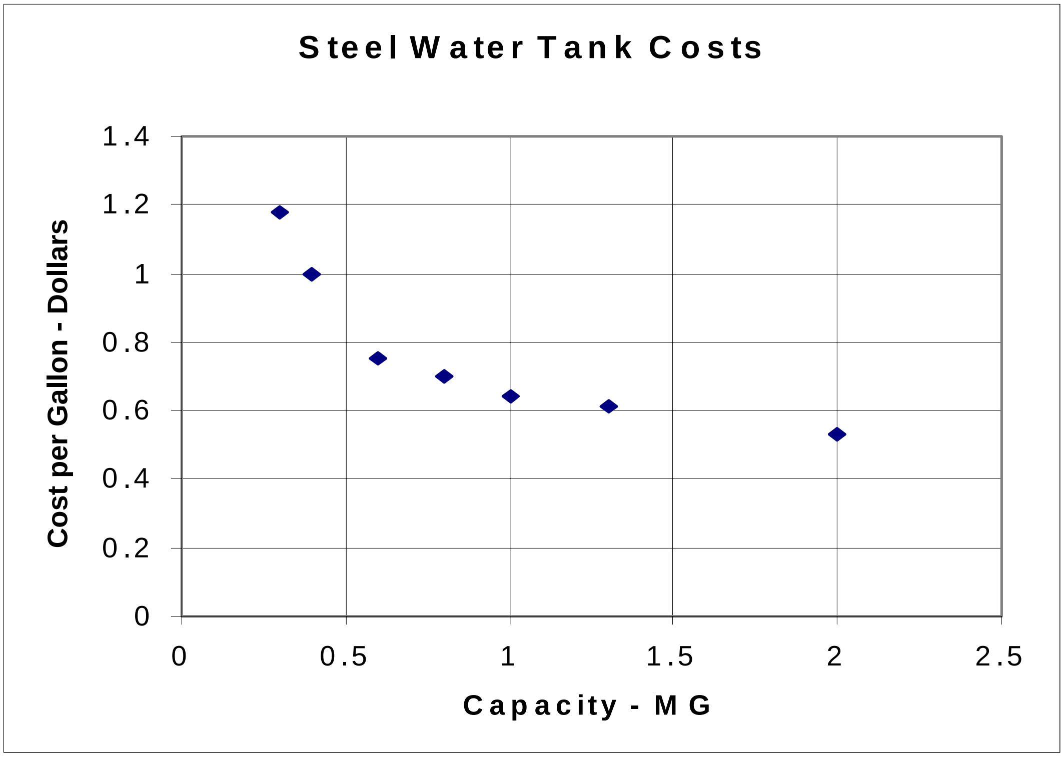
Diseconomies of scale or decreasing returns to scale

output increases less than in proportion to inputs

"Any of the three is possible, and perhaps all could be applied in particular cases. But the British economists mentioned above suggested that economies of scale would be the most important in general. They reasoned from Smith's principle of the division of labor. They reasoned that as in industry or an industrial economy grows larger, the larger size would allow for increased division of labor. This would increase the productivity of labor, and so output would increase more than proportionately to input.

"If economies of scale exist in the economy as a whole, and if they are quite strong, they might overcome the limited quantity of land and allow continuous increases in the division of labor, the productivity of labor, and economic growth. In this way, Smith's insight on division of labor was extended into the twentieth century.

Economies of scale pervade the construction industry and can usually be estimated, at least roughly, based on past experience. Reductions in the cost of producing a product as the scale of output increases is common to many kinds of work. An illustration of this is shown by the plot below of the cost of steel water tanks versus tank capacity.



This pattern is often repeated in many areas of engineering where the unit cost of some product or process varies inversely as some power of the scale factor.

6. π **Summary**

Dealing with uncertainty in engineering has been, and always will be a challenge. We can never predict certain things with absolute certainty so statistical methods will always be with us. In this course you have seen only a few of those methods but it is hoped these examples will inspire you to build your own library and seek out other examples and sources of data.

References

Documentation of the NFF program is contained in:

Ries, K.G., III, and Crouse, M.Y., 2002, The National Flood Frequency Program, Version 3: A Computer Program for Estimating Magnitude and Frequency of Floods for Ungaged Sites: U.S. Geological Survey Water-Resources Investigations Report 02-4168, 42 p.

"Documentation of the equations, maps, and other information needed to solve the regression equations for individual States is provided on line through links from the NFF web page (http://water.usgs.gov/software/nff.html) to the complete original State reports, fact sheets, or still-current pages from the report for the previous version of NFF:

Jennings, M.E., Thomas, W.O., Jr., and Riggs, H.C., 1994, Nationwide Summary of U.S. Geological Survey Regional Regression Equations for Estimating Magnitude and Frequency of Floods for Ungaged Sites, 1993: U.S. Geological Survey Water-Resources Investigations Report 94-4002, 196 p.

The World of Mathematics by James R. Newman, Simon and Schuster, 1956

Hydrology, US Department of Transportation, Federal Highway Administration, 1984

Standard Handbook of Engineering Calculations, Tyler G. Hicks, Editor, Mc Graw-Hill, Inc. 1994