Institute of
Hydrology

Natural Environment Research Council

A. R. Black
IH (Scotland)

# Hydrological Frequency Analysis Package
# (HYFAP)

## CONTENTS

# 1 INTRODUCTION

## 1.1 Structure of the manual

Section 1 (this section) gives a short description of the HYFAP package, information about the hardware requirements and installation instructions.

Section 2 of the manual explains how to start the HYFAP package and describes how to use the menu system. It also gives information about error messages and the function of particular keyboard operations.

Section 3 explains how to perform a frequency analysis with HYFAP, with the aid of a tutorial example.

Appendix A gives additional information about the installation procedures.

Appendix B provides some background information about frequency analysis in general and the facilities provided by HYFAP in particular.

Appendix C contains a number of key references.

## 1.2 Conventions used in the manual

Keys on the computer keyboard are indicated by using the name of the key in angle brackets. For example, the key marked F1 on the keyboard is represented in the manual by <F1>. The Enter key (sometimes marked Return, CR or ↵ ) is indicated by <Enter>.

HYFAP is controlled by selecting choices from menus displayed on the screen. Menu choices are indicated in the manual by square brackets, e.g., [Exit to DOS].

Any input that the user is required to type is represented by bold characters. For example, to start the program the user must type **HYFAP** then press the <Enter> key.

Section 3 of the manual contains tutorial examples. These are shown in italics.

## 1.3 HYFAP overview

The Hydrological Frequency Analysis Package, HYFAP, is designed to perform a frequency analysis on a time series of annual maximum event data. It can estimate parameters for six different probability distributions using up to three parameter estimation procedures, and can also apply user-defined dimensionless regional frequency curves to the site data. The package will provide estimates of the magnitudes of events with particular return periods (together with their standard error of estimate), or alternatively the return periods associated with specified magnitudes. The data and fitted frequency distributions can be plotted the on the computer screen or on a printer or plotter.

Data can be read in from a file and edited within HYFAP, or entered directly from the keyboard. The period of record to be analysed can be selected by the user, and analysis can be performed on standardised data.

HYFAP is menu-driven and provides context sensitive help at every stage.

## 1.4 Hardware requirements

HYFAP is designed to run on an IBM PC, AT or PS/2, or similar computer, with a hard drive, 640k RAM, a graphics monitor and adaptor card (CGA, EGA or VGA). The hard disk drive will need 1Mb of free space.

## 1.5 Installation

HYFAP Version 2 is distributed on two 720k, $3^1/_2$" disks or three 360k, $5^1/_4$" disks.

HYFAP uses a protection system to safeguard your rights as a registered user. Before you can run HYFAP you must transfer the protection system to the hard disk drive where the HYFAP program files will reside.

Insert the HYFAP program disk 1 into a floppy disk drive, log on to that drive (e.g., to log on to drive B, type **B:**<Enter>) and then type:

**TRANSFER source drive: destination drive:** <Enter>

where source drive is the disk drive containing program disk 1 and destination drive is the designation of the hard drive you want to install HYFAP to. For example, to transfer the copy protection from floppy drive B to hard drive D you would type the command:

**TRANSFER A: D:**

You are only allowed a single installation from your original program disk 1. If you need to move HYFAP to a different computer see Appendix A.

After you have installed the protection system you can transfer the HYFAP program files to your hard drive. An installation program is provided to do this for you. With the program disk 1 still in the floppy disk drive type:

**HYFAPINS** <Enter>

This will run the installation program. During installation you can press the <F1> key to display a simple help facility. Follow the installation procedure instructions on the screen:

i. Select the source disk drive (where Program Disk 1 is located) and the destination disk drive (the hard disk where you want HYFAP to be installed).

ii. Select the graphics drivers for the display and hardcopy devices.

The installation procedure will create a directory called \HYFAP on the selected hard drive and copy all the necessary files from the floppy disks into it. Finally a configuration file will be created in the HYFAP directory to enable the program to plot and print graphics.

# 2 OPERATION

## 2.1 Introduction

HYFAP is a menu driven program which is user friendly, with context sensitive help built in. To start the HYFAP program, log onto the hard disk drive containing the HYFAP directory by typing **C:** <Enter> (if drive C is being used), then **CD \HYFAP** <Enter> and finally **HYFAP** <Enter>. You can run HYFAP on computers with only 512k RAM by typing **HYFAP /N** <Enter>. If you use this option the computer will not be able to display or print graphical output.

If your computer does not have sufficient memory for graphical displays a message will prompt you to use the /N option.

## 2.2 Basic operation

When the program is run it will check to see that the copy protection system is installed on the hard drive containing the HYFAP program files. If the protection check is satisfactory a title screen is displayed and you will be prompted to press the <Enter> key to start the procedures.

The screen will then display a bar menu on the top line and a function key menu on the bottom line. The top bar menu has seven options:

[■] [Exit] [Input] [Fitting] [Output] [Graphics] [Help]

The function key menu has three choices:

[Esc=Cancel] [F1=Help] [F5=Action]

If a mouse is installed, a mouse cursor will also be displayed on the screen. Menu options may be selected by using any of the following methods:

i. If a mouse is installed you can position the mouse cursor anywhere on the text of a menu option and then click on the left mouse button.

ii. The arrow keys (usually located on the right side of the keyboard) can be used to move the menu highlight bar to the required option, which can then be selected by pressing the <Enter> key.

iii. Each of the menu choices displays a highlighted letter. Selections can be made by typing the highlighted letter.

When a selection is made from any of the options on the top menu bar, a "drop down" menu is displayed which gives further choices. The [■] option provides three choices. The first one, [About], will display a panel of information about the version of HYFAP installed on your computer. The second option toggles the clock display on and off and the third, [Cancel], clears the drop down menu display. The <Esc> key can also be used to cancel the drop down menu display. All of the drop down menus offer [Cancel] as an option and the manual omits further mention of them in the following descriptions.

The [Exit] option allows the user to terminate the HYFAP program and return control to DOS.

The [Input] option provides two choices for entering data. Data may be input from a disk file or from the keyboard. Both methods prompt the user for more details about data standardization and return periods for the data fitting. For more information about the data standardization and return periods please refer to section 3.3.

The [Fitting] option displays all the possible fitting methods provide by HYFAP. Selecting one will cause a fitting result to be displayed on the screen. The selected fitting method will be retained until another is chosen or the user exits from the program.

Selecting [Reset] will clear any previous fitting information.

The [Output] option has two associated choices, [Text output] and [Graphics output]. If [Text output] is selected the fitting parameters and results will be written to a disk file in ASCII format. The user will be prompted to assign a file name. Selecting [Graphics output] will allow a graphics output device to be chosen for subsequent graphic displays. The selectable output devices are screen, printer or plotter, with screen as the default if no other choice is made. This version of HYFAP does not automatically check that the correct device drivers were installed during the installation procedure, or whether the selected device is attached to the computer and is ready to recieve data.

The [Graphics] option allows users the choice of displaying a time series bar graph or a frequency analysis fitting graph. The display will be sent to the output device selected through the [Output] menu option. The frequency analysis fitting graph will be plotted using the method selected in the [Fitting] option.

The [Help] option gives access to all the context sensitive help screens. These are the screens that are displayed when the <F1> key is pressed.

## 2.3 Exit to DOS

To exit to DOS select the [Exit] option on the top menu bar and then select [Exit to DOS]. This will allow the user to exit the program and return to DOS mode.

## 2.4 Operational information

Selecting an option on the top bar menu (other than [▪], [Exit] and [Help]) will cause an error message to be displayed if HYFAP has not been given any data to work with via the [Input] option.

When [Input] from a file has been selected, the data file name may be chosen from a displayed list of available data files. The list of displayed files can be restricted by specifying a file name mask using the standard DOS conventions. A file may also be read by typing its directory location and name, again using standard DOS conventions. An error message will be displayed if the file can not be found in the specified directory.

The data selected must be a minimum of three years or an error message will be displayed.

If the [Graphics] menu option has been selected and the graphics driver has not been installed, or an ASCII character image file is not found, an error message will be displayed.

# 3 USING HYFAP: WITH A TUTORIAL EXAMPLE

## 3.1 Introduction

The aim of a frequency analysis is to determine the relationship between the size of an event and the probability of that magnitude event being exceeded in the future. This is done by estimating the parameters of a probability distribution from a sample of recorded data. It is assumed that the estimated probability distribution provides a reasonable approximation to the "true" probability distribution which threw up the observed sample.

The HYFAP package can be used to estimate the parameters of a number of probability distributions from a sample of annual maximum event magnitudes, which is comprised of the largest event recorded in each year. This manual does not attempt to provide a guide to frequency analysis, and the reader is referred to the UK Flood Studies Report (NERC, 1975) and Cunnane (1989) for comprehensive reviews and recommendations. There are, however, several important points which must be made before embarking upon the detailed description of the HYFAP package.

Firstly, a large number of probability distributions have been recommended for frequency analysis, and many different methods are used in different countries. There is no single probability distribution and method of application which will be best in all circumstances, although some methods are more likely to be appropriate than others. The majority of probability distributions recommended for practical frequency analysis have two or three parameters. Three-parameter distributions are more flexible than two-parameter distributions, and have greater potential for describing the underlying "true" magnitude-frequency relationship. However, a longer record is usually needed to produce an estimate with a given degree of reliability using a three-parameter distribution than with a two-parameter distribution, although differences in the statistical characteristics of distributions and methods for estimating parameters make comparisons difficult (see Cunnane, 1989).

Secondly, the records available to the frequency analyst tend to be short (they rarely cover more than 50 years), and the estimation of the frequency of large events therefore involves considerable extrapolation beyond the limits of recorded experience. It is quite common for different frequency estimation procedures to give similar results for small and medium sized events, but to produce very different estimates for larger, rarer events. Small sample sizes will cause considerable uncertainty in the estimated parameters of a probability distribution, and will make it difficult to discriminate between alternative procedures. It is also assumed that the characteristics of extreme events in the future will be the same as those estimated from the short record.

The results of a frequency analysis of data from a single site need to be treated with caution because of a possible combination of data error (due to inaccuracies in the observed data), model error (caused, for example, by the choice of probability distribution and the assumption that the future will be like the past) and parameter error (resulting from the estimation of model parameters from short records which may be subject to a high degree of measurement error). In most cases it is preferable to undertake a regional frequency analysis, where information from many gauged sites in a region is combined to produce more reliable estimates at a particular location (see Cunnane, 1989). Regional frequency analyses cannot be performed with the HYFAP package itself, but the package does allow a regional frequency relationship to be plotted against site data and at-site frequency estimates. It is the user's responsibility to combine the site and regional information in the most effective way.

It is important to understand what else the package does not do:

i. HYFAP uses only annual maximum data. It cannot perform a frequency analysis using all the peaks over a threshold.

ii. HYFAP is designed for the analysis of annual maxima. It is not suitable for annual minimum frequency analysis (for example, for droughts and low flows).

iii. Floods with a magnitude equal to zero are ignored by the package, although it does indicate how many zeroes have been omitted.

iv. HYFAP cannot incorporate "historical" information into the analysis (for example, that a notable flood occurred 27 years before the start of a record).

v. The package does not recommend the "best" method for use in a particular circumstance. This decision is left to the user but some helpful notes are provided in Appendix B.3.

## 3.2 Frequency analysis with HYFAP

A typical frequency analysis with HYFAP will involve the following stages:

i. Enter annual maximum data.

ii. Plot a time series of the data to check, visually, for any unusual features (such as a progressive trend or a jump).

iii. Fit a number of probability distributions.

iv. Plot the different estimated distributions against the observed data.

v. Print out a tabular summary of the analysis.

The rest of this section describes the facilities of the HYFAP package with the aid of a tutorial example. The example uses data from the River Nowater at Aircastle (a gauging station with 25 years of annual maximum flood data) and the Styx at Charon's Bridge (with 31 years of data). The data are held in the file EXAMPLE.DBS supplied with the distribution disks.

## 3.3 Entering data into HYFAP

To enter data choose the [Input] option from the menu bar along the top of the screen. Another menu drops down offering three options. One of these is to [Cancel] and go back to the top menu. The other two options permit the entry of data from a file or from the keyboard. Figure 3.1 gives a flow chart of the actions involved in entering data. The tutorial shows how to read data from a file and how to enter it through the keyboard. Go back to the input stage if you want to do the "keyboard entry" tutorial after the "file entry" tutorial.

## 3.4 Entering data from a file

When the user chooses to input from a file, a box appears in the centre of the screen listing all the files in the current directory matching the default pattern (*.DBS). The search directory can be changed by typing **D**, and then filling in the directory name in the box indicated. The default pattern can be changed (for example to *.TXT) by typing **P** and entering the new pattern. You can enter the name of the desired file directly by first typing **N** and then filling in the name in the highlighed box.

Choose the data file from amongst those listed using the cursor or mouse, and make the selection with the <Enter> key. The format of the input file is:

| | | |
|---|---|---|
| A10, A40 | Number | Name |
| I10, F10.3 | Year | Magnitude |
| .. | | |
| .. | | |
| I10 | -1 | (end of station flag) |

The package can only handle records less than 200 years long. (Note: a file with the correct format can be extracted automatically from the Institute of Hydrology's HYDATA package using HYSTAT - see the HYSTAT manual.)

If the input file contains more than one station, a box appears listing the station names. Select a station using the cursor or mouse and the <Enter> key.

*Tutorial: Select the file EXAMPLE.DBS: a panel will be displayed showing that the file contains two stations. Choose the catchment NOWATER AT AIRCASTLE.*

The package then offers the opportunity to EDIT THE DATA. If the opportunity is taken, a panel appears showing the station details and listing the annual data. Select a year to edit using the arrow keys or the mouse (the panel of data can be scrolled up and down), and then press the <Enter> key. You can now change the selected value. It is not possible to add data to the beginning or end of the record, but it is possible to shorten the record by replacing data by the missing value code *m*. Individual values can be ignored in an analysis by replacing them temporarily with *m*. Press the <F2> key

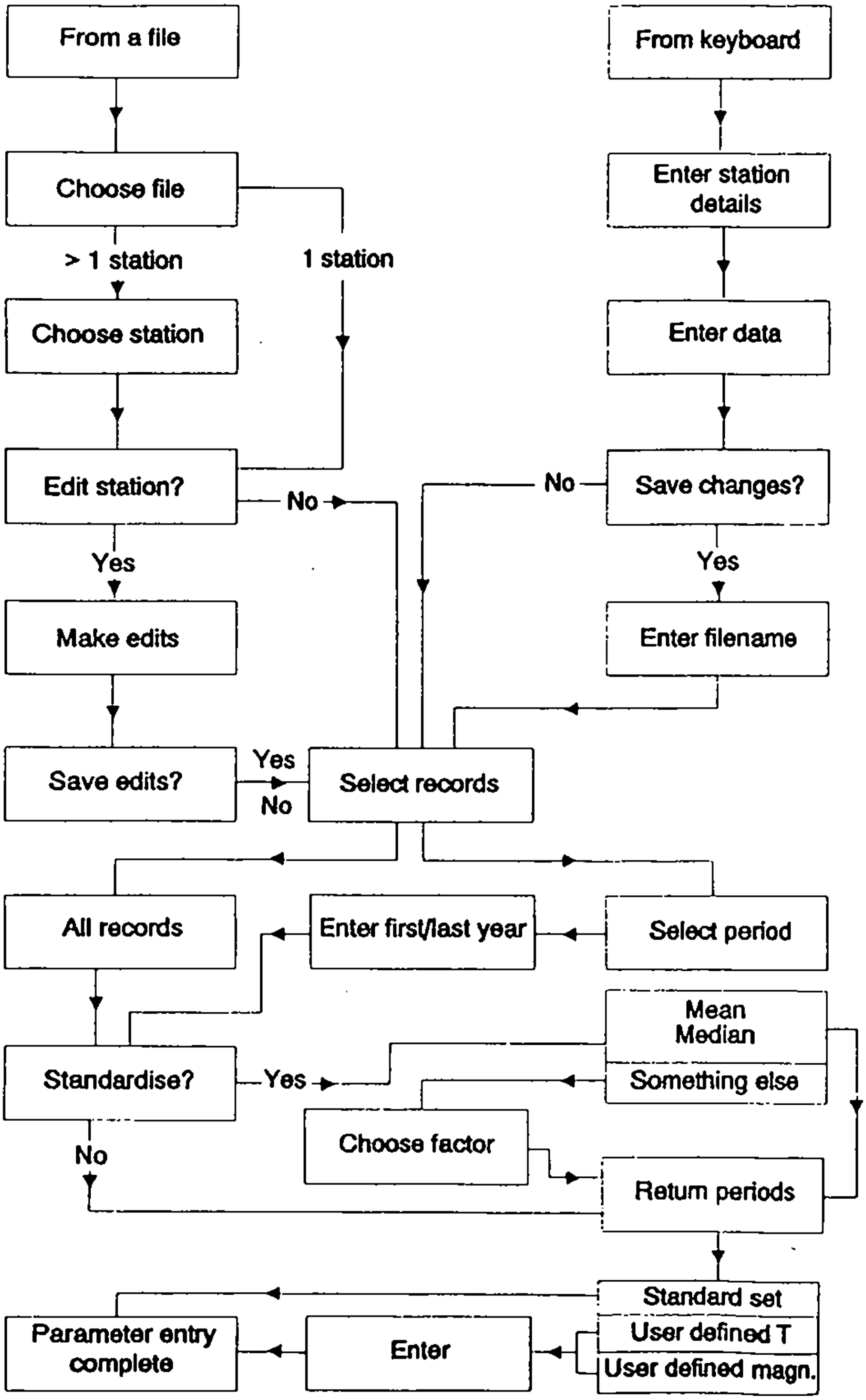


Figure 3.1: Data entry into HYFAP

to leave the panel and save the edited data back in the original data file, or press <Esc> to leave the panel without automatically saving the edits. If <Esc> is pressed the program asks if the data are to be saved.

*Tutorial: Chose the EDIT option, and look at the data. A screen-full of data will appear. Edit the data by changing, for example, the value of 59.38 in 1963 to 89.38. Leave the edit panel using the <Esc> key.*

## 3.5 Entering data from the keyboard

The package first asks for the following station details:

- station name
- station number
- first and last year of record

A panel is then displayed showing the station details together with a list of years from in the record from the first to the last. The code letter *m* appears against each year, indicating a missing value. To enter values, select the required year (using the cursor keys or a mouse), press the <Enter> key, and type in the value. Press the <Enter> key again to record the entry. Enter **m** if the value for a year is missing.

Press the <F2> key to save the data and type in a file name when prompted.

*Tutorial: Choose the "ENTER DATA FROM KEYBOARD" option, and fill in the station details. Then simply type in the data. Save your data in a file.*

## 3.6 Defining records, standardising and selecting

The remaining questions are asked regardless of whether the data are entered from a file or the keyboard.

Select period of record

It is possible to perform the frequency analysis on the entire period of record available, or on a subset. If the subset option is selected, the package prompts for the start and end year required. HYFAP checks that the end year is later than the start year, and will not accept periods less than three years (although it would not of course be wise to do a frequency analysis with such a short record).

*Tutorial: Choose the "SELECT PERIOD" option, and enter a new start and end year for analysis.*

Standardise the data

It is possible to perform a frequency analysis with standardised data. This may be desirable in order to compare frequency curves estimated at a number of different sites in a region. If the standardisation option is chosen the package offers three alternatives:

- standardise by the mean
- standardise by the median
- standardise by something else

The box prints the sample mean and median for information. The third alternative allows the data to be standardised by, for example, the previously estimated 5 year magnitude, or by some multiple of the mean or median. The Flood Studies Report (NERC, 1975) recommends, for example, that the site average be estimated from 1.07 times the median if the sample contains an outlier more than 3 times the median. HYFAP does not directly warn of the presence of outliers but it is possible to identify them from the package output.

*Tutorial: Choose the "STANDARDISE" option, and look at the alternatives offered. If you are using the original NOWATER AT AIRCASTLE data, the mean should be 79.60, and the median 77.73. Standardise the data by the mean (Option 1).*

Select return periods

The package can be used either to estimate the magnitudes of particular return period events or the return periods of specified magnitudes. Three different options are available.

i. Choose the "standard set" of return periods. The package will estimate the 2, 5, 10, 25, 50, 100 and 200 year events.

ii. Define a number of return periods to estimate. If this option is selected the user is invited to enter up to 10 return periods (which need not be whole years: 2.5 is allowed, for example). Enter 0, or simply press <Enter>, to stop adding return periods.

*Tutorial: Select this option, and enter a selection of return periods, such as 2.5, 5, 10, 25 and 50 years.*

iii. Define a number of magnitudes for which return period estimates are required. If this option is selected the user is invited to enter up to 10 magnitudes. Press <Enter> when all the desired magnitudes have been entered.

The selected data are written to a working file FFAPTEMP.DAT, which is then read during the subsequent operations.

## 3.7 Performing a frequency analysis

Choose the FITTING option from the menu bar along the top of the screen. A menu will drop down listing a number of alternative frequency estimation procedures, as shown in Table 3.1. Appendix B gives more information about the various procedures and their strengths and weaknesses. Select the method desired with the cursor or mouse and the <Enter> key. The results will be displayed in a panel. Press any key to clear the panel and chose another fitting procedure. You can do as many fits as you like in one session.

The tutorial uses the data from the Nowater at Aircastle, with no changes, no standardisation and using the standard set of return periods. Go back to the [Input] option if your data set is by now different.

```
  1.  EV1 Moments
  2.  EV1 MLE
  3.  EV1 PWM
  4.  GEV MLE
  5.  GEV PWM
  6.  2-Par Lognormal (mom)
  7.  2-Par Lognormal (MLE)
  8.  3-Par Lognormal (mom)
  9.  3-Par Lognormal (MLE)
 10.  3-Par Lognormal (PWM)
 11.  Pearson-3 (mom)...
 12.  Log-Pearson-3 (mom)...
 13.  Regional curve
 14.  Reset Fitting
 15.  Cancel [Esc]
```

Table 3.1: The fitting options available within HYFAP

## 3.8 Single station frequency analysis

Methods 1 to 10 are invoked through just one press of the <Enter> key. Methods 11 and 12, however, both require the answer to a subsidiary question:

Skewness correction:
    No correction
    Simple correction
    WRC correction
    Bobee and Robitaille

The question refers to the method used to correct sample skewness when estimating model parameters (as outlined in Appendix B).

*Tutorial: Select the GEV-PWM estimation procedure.*

Table 3.2 shows the panel presented when the GEV-PWM procedure (Method 5) is applied with the data from the River Nowater (and when the magnitudes of a set of return periods are required). The panel summarises the data and shows the length of record used. A message would be displayed if the data had been standardised. The panel will also indicate how many zeroes were omitted from the analysis, if any.

```
      12345         Nowater at Aircastle
Number of years            :  25
Fitting procedure          :  GEV-PWM

  u      =      64.651     a      =      24.927
  k      =      -.022
Return period         Magn.        S.E.
      2.               73.82        6.35
      5.              102.67        9.40
     10.              122.17       13.08
     25.              147.28       21.08
     50.     *        166.25       29.73
    100.     *        185.38       40.66
    200.     *        204.73       53.88
```

Table 3.2: GEV-PWM applied to the Nowater at Aircastle

The package presents the model parameters, in this case u, a and k, which are summarised in Appendix B. The package also calculates the magnitude corresponding to each return period event and the standard error of estimate at each return period (except when using method 10, the 3-parameter lognormal distribution estimated by probability-weighted moments). The standard error gives an indication of the precision of the estimated magnitude with a given return period (see Appendix B), but it is appropriate here to emphasise two warnings:

i. The standard error indicates only the parameter error. It does not allow for the fact that the model selected may be incorrect.

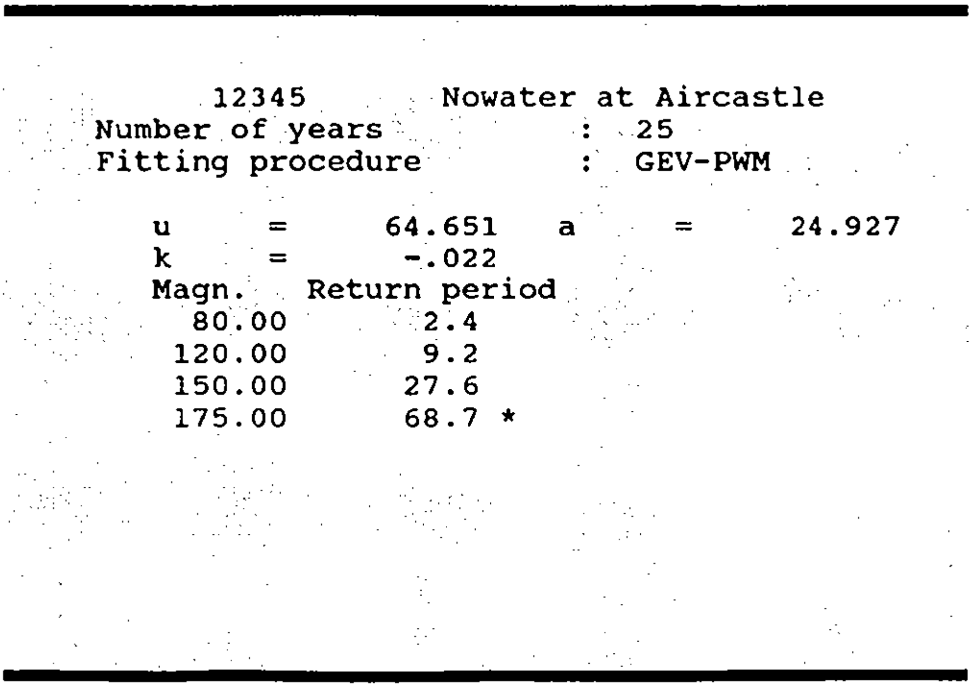
ii. It is difficult to estimate confidence intervals around the estimated magnitude with the standard error. The assumption that the sampling distribution of the T-year event is normal is inappropriate, particularly at high return periods (Appendix B).

If it is not possible to estimate standard errors, the program will display a value of -9.9.

Frequency analysis generally involves the extrapolation to rare events larger than those experienced in the data sample, and is therefore inherently risky. It is impossible to give definitive rules limiting the most extreme return period which can be estimated from a short sample. The reliability of an estimate falls off rapidly once the return period of interest is greater than the record length (with the rate of decline depending on how well the distribution fits the data). As a visual indication of increased unreliability, the HYFAP package uses an asterisk to label estimated magnitudes with a return period more than twice the record length, as shown in Table 3.2. It must be emphasised that reliability declines continuously as return period increases, and that any rule about the allocation of asterisks must be arbitrary. Warnings also appear if the record is deemed to be "short" (less than 10 years when a two-parameter distribution is used and less than 15 years if a three-parameter distribution is fitted).

*Tutorial: Press <Esc> to clear the panel, and choose another frequency fitting procedure.*

Table 3.3 shows the results panel when the return periods of a number of user-defined magnitudes are required. Again, an asterisk marks a return period of more than twice the record length.



```
           12345           Nowater at Aircastle
     Number of years            :  25
     Fitting procedure          :  GEV-PWM

        u     =      64.651    a     =      24.927
        k     =      -.022
        Magn.   Return period
          80.00       2.4
         120.00       9.2
         150.00      27.6
         175.00      68.7 *
```

Table 3.3: GEV-PWM applied to the Nowater at Aircastle: user-defined magnitudes

The frequency analysis omits zero values. If a record does contain zeroes, the estimated probability of a particular magnitude event has a slightly different meaning. It is a conditional probability, showing the probability of an event occurring given that an event which was greater than zero occurred. This conditional probability can be converted to an annual probability using:

$$\text{Probability of event occurring in any year} = \text{Conditional probability} * \lambda$$

where $\lambda$ is the ratio of years with events to the total record length. This can be restated as:

$$T = T_{withzero} / \lambda$$

where $T_{withzero}$ is the conditional return period (as tabulated in the panel) and T is the unconditional return period. For example, if there are 3 zeroes in a sample of size 20, $\lambda$ is 0.85 and the conditional 10 year return period event becomes the unconditional 11.77 year event (in other words, there is a 1 in 10 chance that an event is greater than the specified magnitude given that it is not zero, and a return period of 11.77 years between events).

Occasionally it will not be possible to estimate the parameters of a particular probability distribution using a particular method from a sample. In such circumstances the results panel explains that there has been a fitting failure and no return period estimates can be calculated. Every effort has been made to ensure that HYFAP can recover from fitting failures but it is possible that the failure will be so drastic as to cause the program to crash.

## 3.9 Regional frequency analysis

Although the HYFAP package does not perform a regional frequency analysis, it is possible to apply a dimensionless regional frequency curve to the site data. The dimensionless regional curve shows the magnitude of a particular return period event as a multiple of some index. The index is estimated from the sample data.

If Method 13 is selected, the package first prompts for the name of a file containing the regional frequency curve information. The file is then read and the regional curves contained listed. The user is then prompted to select from the list provided using the cursor or mouse and the <Enter> key.

The file FSR.REG, provided on the distribution disks, gives the 10 regional flood frequency curves presented in the UK Flood Studies Report, together with the Great Britain and Ireland national curves (from Flood Studies Supplementary Report 14: Institute of Hydrology (1983)).

The regional frequency file (such as FSR.REG) has the following format:

| | |
|---|---|
| Data set name | A20 |
| No. of curves NC, no. of return periods NR, code | free format |
| NR return periods | free format |
| Region name, NR growth factors<br>(repeated NC times) | A10,10F6.3 |

The code can take three values:

code = 1: curve shows T-year event as multiple of the mean;

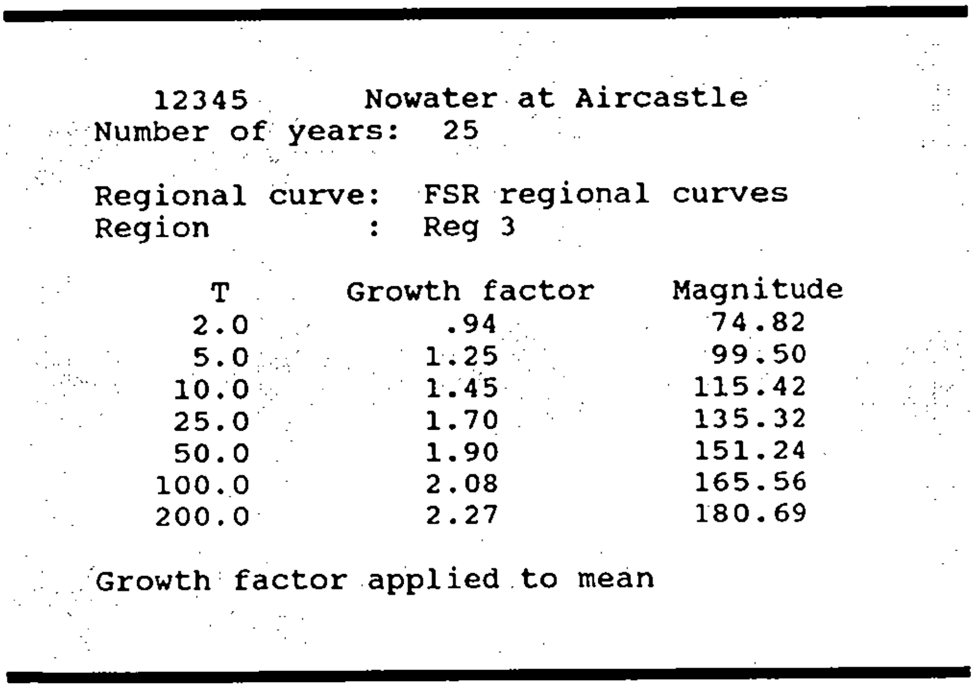
code = 2: curve shows T-year event as multiple of the median;
code = 3: curve shows T-year event as multiple of the mean, or as a multiple of 1.07 times the median if the sample contains an outlier more than three times the median. This is the rule used in the UK Flood Studies Report (NERC, 1975).

The file can be created using a text editor. Unfortunately HYFAP cannot check that the file is correctly formatted and may fail if there is an error.

*Tutorial: Select the regional estimation procedure, choose the file FSR.REG, and select region 3.*

Table 3.4 shows the panel resulting from the application of the regional curve for Region 3 to the data from the River Nowater.

It is only possible to apply three regional frequency curves in any one analysis.



```
   12345      Nowater at Aircastle
Number of years:  25

Regional curve:  FSR regional curves
Region        :  Reg 3

      T       Growth factor    Magnitude
     2.0           .94           74.82
     5.0          1.25           99.50
    10.0          1.45          115.42
    25.0          1.70          135.32
    50.0          1.90          151.24
   100.0          2.08          165.56
   200.0          2.27          180.69

Growth factor applied to mean
```

Table 3.4: Regional frequency curve applied to the Nowater at Aircastle

## 3.10 Reset: remove all fits

Whenever a new fitting procedure is applied, the results are added to those from earlier analyses. To "clear" the memory, chose [Reset] from the menu of fitting procedures. This may be useful if the user wants to start again and concentrate on a small number of alternative methods for plotting (Section 3.5) or tabulating (Section 3.6).

## 3.11 Producing a plot

HYFAP can produce both a plot of the time series of the annual maximum data and a frequency plot, showing the relationship between magnitude and return period. Both types of plot are obtained via the [Graphics] option along the top of the screen. By default, HYFAP assumes that you want the plot on the screen. To change to plot on a printer you need to select the [Output] option from the top menu, and then choose [Graph output]. You can then select from the screen, a printer or a plotter. HYFAP will give a warning if an attempt is made to produce a plot on a device for which a graphics driver has not been installed. If such a warning appears you must [Exit] HYFAP and go through the installation process (Chapter 2) again.

The user has considerable flexibility in defining the characteristics of a graph plot.

## 3.12 Producing a time series plot

The time series plot allows a quick check of the input data. It may be possible to spot discontinuities and trends. Select the [Graphics] option from the top menu and then choose the [Time series] option.

A panel will appear showing the lower and upper Y axis limits, the first and last year along the horizontal axis, and the default label for the Y axis. The user can change any of these characteristics and enter a title for the graph which will be printed along the top above the station name. Note that the window in the panel only shows a portion of the Y axis label; use the cursor arrows to expose the rest.

*Tutorial: Choose the TIME SERIES option from the GRAPHICS menu. Change the upper limit of the Y axis to be 250, change the first and last years to be 1960 and 1985 respectively, and add a title to the plot.*

Figure 3.2 shows the resulting time series plot. The name of the owner of the HYFAP package and the date of the plot are shown along the bottom of plots produced on either a printer or a plotter.

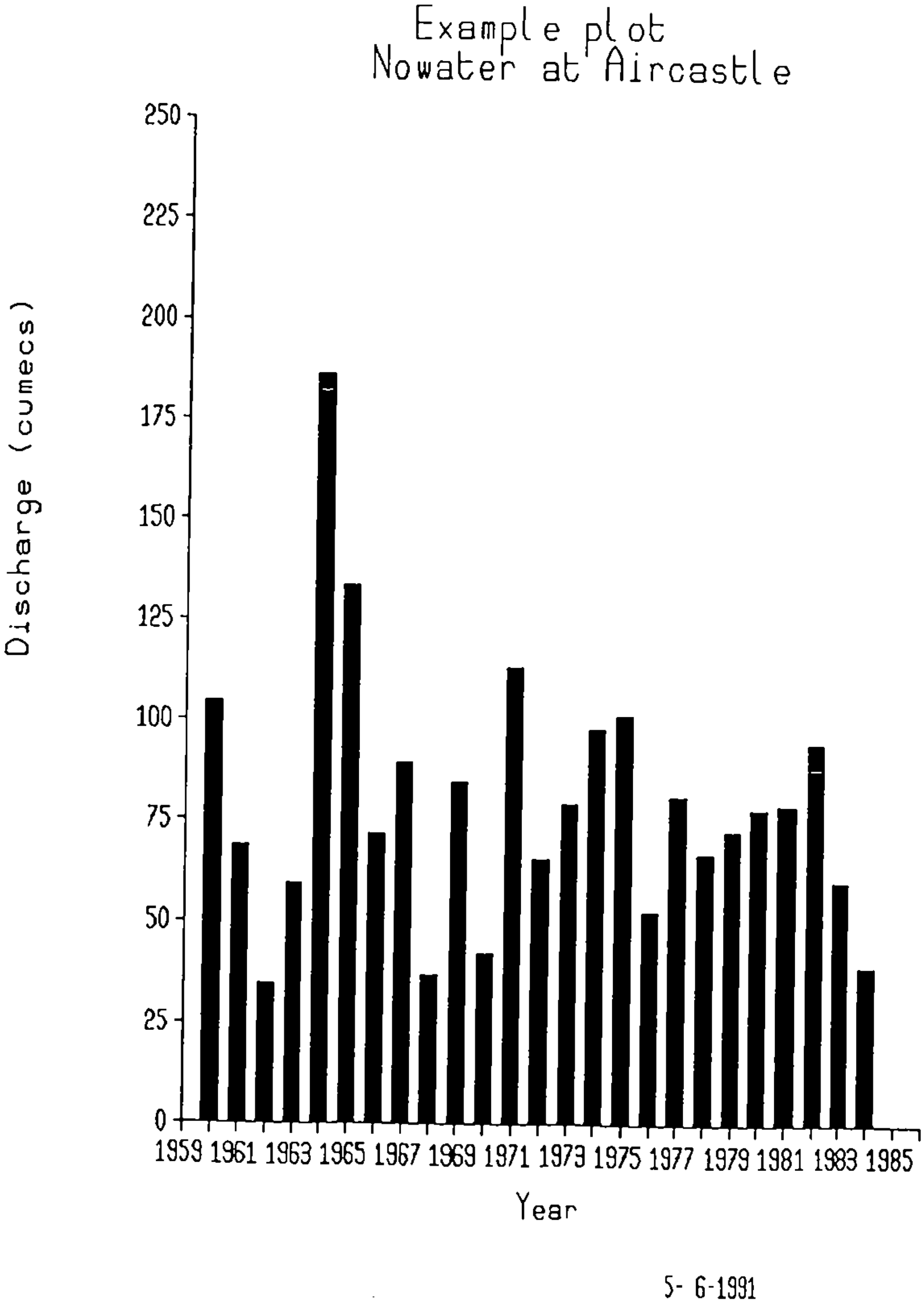


Figure 3.2: Time series plot

## 3.13 Producing a frequency plot

A frequency plot is often the end result of a frequency analysis. It shows the relationship of magnitude to return period and can be used to estimate graphically the magnitude of any return period event (and vice versa). A plot is produced by selecting first the [Graphics] option from the top of the screen and then by choosing the [Frequency plot] option.

The package first asks for the highest return period to be shown on the graph. This defines the limits of the horizontal axis, and only four options are available: 100, 200, 500 and 1000 years. In most cases 200 years will be adequate.

The package then displays a similar menu to that for the time series plot, but without giving the opportunity to change the horizontal limits of the plot.

*Tutorial: Select the FREQUENCY PLOT option. Choose 200 years return period, change the upper limit of the Y axis to 250, and add a title to the plot.*

Figure 3.3 shows the resulting frequency plot. There are a number of important points:

i. The graph plots linear magnitude against a transformation of probability. The transformation used is:

$$Y = -\ln(-\ln(1-1/T))$$

where Y is the "linear reduced variate" and T is the return period. A Gumbel (or EV1) distribution plots as a straight line with this transformation, and Y is also known as the "Gumbel reduced variate". The plot is in effect being produced on Gumbel paper. A return period scale is added to the graph.

ii. The package uses the Gringorten plotting position (Gringorten, 1963; NERC, 1975) to plot the observed data. The Gringorten plotting position has the form:

$$F_i = \frac{(i - 0.44)}{(N + 0.12)}$$

where i is the rank of the event (with the smallest of rank 1), N is the sample size and $F_i$ is the probability of an event being smaller than the i rank event. The return period T is equal to

$$T = \frac{1}{1 - F_i}$$

iii. The fitted frequency relationships are shown as dashed lines beyond a return period equal to twice the record length. This is an arbitrary device to

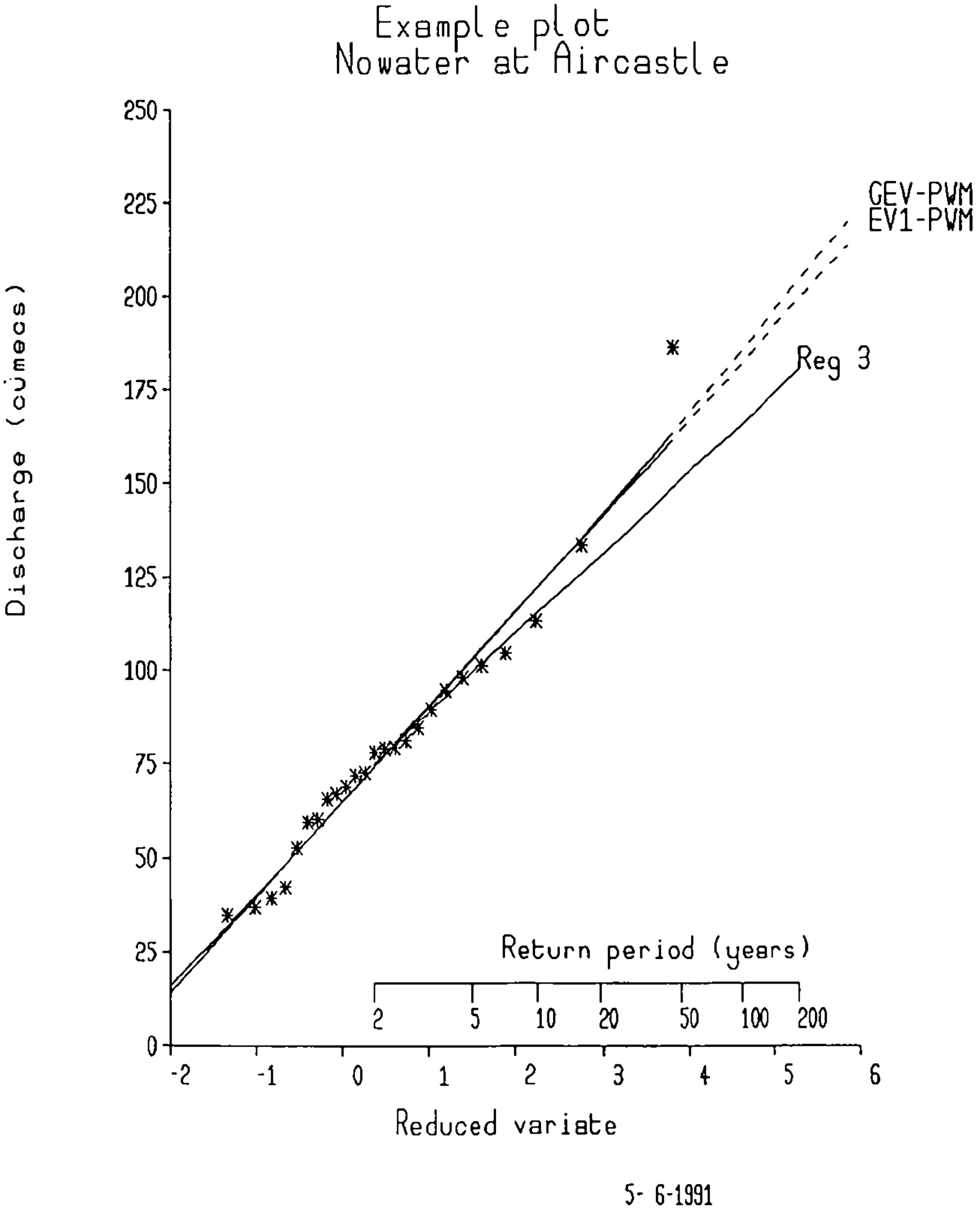


Figure 3.3: Frequency plot

emphasise the uncertainties involved in extrapolating beyond the record length available.

iv. It is possible to plot all the frequency curves on the graph at the same time, but it becomes difficult to distinguish the different methods when too many are plotted.

## 3.14 Producing a summary table

The HYFAP package can produce a summary file containing the results from all the methods applied during a session. Select the [Output] option from the top menu and then choose the option [Text output]. The package will prompt for a file name. The file created will be a standard ASCII file and can be either printed or read into a word processing package for further manipulation.

*Tutorial: Create an output file containing the results of a number of frequency estimation procedures applied to the Nowater at Aircastle.*

Table 3.5 shows the contents of the output file from the analysis of the Nowater data. The file summarises the input data (period of record, mean, maximum, coefficient of variation and skewness), lists the fitted parameters and tabulates estimated magnitudes, standard errors and return periods. The file also lists the methods which could not be applied to the data.

## 3.15 Finishing the analysis

At any point in the analysis it is possible to go back to the beginning and enter new data through the [Input] menu. The package does not check that results have been written to an output file before allowing the user to analyse another data set.

To leave the HYFAP package select [Exit] from the top menu, and choose the [Exit to DOS] option.

```
=================================================================

         STATION       12345          Nowater at Aircastle
         DATA CHARACTERISTICS

         Period of record used from 1960 to 1984

         Number of years... 25
         mean......        79.6    max. flood...     186.1 in 1964
         CV.........        .40    skewness.....     1.259
 Ranked data are
   1: 1964  186.1    2: 1965  133.4    3: 1971  113.1    4: 1960  104.5
   5: 1975  101.1    6: 1974   97.8    7: 1982   94.2    8: 1967   89.2
   9: 1969   84.5   10: 1977   81.1   11: 1973   79.1   12: 1981   78.8
  13: 1980   77.7   14: 1979   72.4   15: 1966   71.6   16: 1961   68.5
  17: 1978   66.9   18: 1972   65.4   19: 1983   60.2   20: 1963   59.4
  21: 1976   52.5   22: 1970   42.0   23: 1984   39.0   24: 1968   36.8
  25: 1962   34.7

-----------------------------------------------------------------

         PARAMETERS OF FITTED DISTRIBUTIONS


 EV1-PWM       u   =    64.906   a   =   25.454
 GEV-PWM       u   =    64.651   a   =   24.927   k   =      -.022
-----------------------------------------------------------------

         FLOOD RETURN PERIOD ESTIMATES

      T              EV1-PWM   GEV-PWM
     2.               74.2      73.8
     5.              103.1     102.7
    10.              122.2     122.2
    25.              146.3     147.3
    50.     *        164.2     166.2
   100.     *        182.0     185.4
   200.     *        199.7     204.7

         Standard errors

      T              EV1-PWM   GEV-PWM
     2.                5.3       6.3
     5.                7.8       9.4
    10.               10.6      13.1
    25.               14.6      21.1
    50.     *         17.7      29.7
   100.     *         20.8      40.7
   200.     *         24.0      53.9

   * indicates return period more than twice record length
```

```
------------------------------------------------------------------

   REGIONAL FREQUENCY ANALYSIS

   FSR regional curves
           Reg 3

        T         Q(T)
       2.0        74.8
       5.0        99.5
      10.0       115.4
      25.0       135.3
      50.0       151.2
     100.0       165.6
     200.0       180.7

   FSR regional curves  Reg 3      Growth factors applied to mean

------------------------------------------------------------------
      Institute of Hydrology, Wallingford    31  6-1991
```

Table 3.5: Output file

# APPENDIX A: HYFAP INSTALLATION

## A.1 The copy protection system

HYFAP incorporates a copy protection system to prevent unauthorised use. The protection system must be transferred to the hard disk drive where the HYFAP package will be installed. The method is described in Section 1.5 of this manual. You are only allowed to install HYFAP on a single hard disk drive, but the protection system can be un-installed if you want move HYFAP to another computer or disk drive (see below for details).

The protection system should not affect any normal disk management operations. The HYFAP program and data files can be freely moved to any disk directory on the same hard drive (i.e., other than the default one, \HYFAP, set up by the installation procedure). Backup copies of the program and data files may be made if required. The hard disk may also be optimised (by using, for example, the optimisers supplied with Norton Utilities™ or PC Tools™) without affecting the copy protection system.

If it becomes necessary to move the HYFAP package to another computer, or to a different hard drive, first use the batch file called TRANSFER.BAT on the HYFAP program disk 1 to move the protection system back to that floppy disk, using similar commands to those required to install the system. For example, if HYFAP is installed on hard drive C you could put the original HYFAP program disk 1 in floppy drive A and type the command:

```
TRANSFER C: A: <Enter>
```

A message will be displayed on the screen when the protection has been successfully removed from the hard disk. HYFAP may now be installed on a different computer or hard disk.

## A.2 The installation program

The installation program, HYFAPINS, described in Section 1.5 is used to copy the HYFAP program and sample data files to a hard disk. It is also used to install the graphic device drivers used to display graphical output. If the type of graphics monitor, printer or plotter attached to your computer is changed, you will need to run HYFAPINS again. Note that in this case you do not need to TRANSFER the copy protection system again.

# APPENDIX B: FREQUENCY ESTIMATION PROCEDURES IN HYFAP

## B.1 INTRODUCTION

The objective of this appendix is to provide some background details on the facilities provided by the HYFAP package. It is assumed that the reader is aware of the basics of frequency analysis. For more information on frequency analysis in general, see the UK Flood Studies Report (NERC, 1975) and Cunnane (1989).

Section B.2 gives some general information on parameter estimation procedures, standard errors and assessing the goodness-of-fit.

Section B.3 summarises the procedures contained within HYFAP.

## B.2 GENERAL INFORMATION

### B.2.1 What HYFAP can and cannot do

HYFAP is designed to perform a single-site frequency analysis on a series of annual maximum magnitudes. A previously-defined regional frequency curve can be applied to the site data.

It cannot:

Apply methods which utilise all the values over a threshold, i.e. a "Peaks over a Threshold" analysis.

Include historical information, such as the fact that a large event occurred some time before the instrumental record began.

Be applied with series of annual minimum events, such as low flows.

Correct automatically for the presence of zero values (but see Section 3.4.2 in the manual);

Advise on the "best" method.

### B.2.2 Parameter estimation

HYFAP provides three different methods of estimating parameters for six probability distributions, although not all the methods are used with all the probability distributions.

i. Method of moments (MOM)

The parameters of a probability distribution can be expressed in terms of the moments of the distribution, such as the mean, variance and skewness. If these moments can be estimated from a sample, it is therefore possible to estimate distribution parameters. The method is well-used (NERC, 1975) and intuitively straightforward. However, estimates of higher moments, such as skewness, made from small samples may be very uncertain, and the method does not give very reliable estimates for distributions with three or more parameters.

ii. Method of maximum likelihood (MLE)

The probability of obtaining a particular sample from a probability distribution with specified parameters can be determined from the likelihood function:

$$L = \prod_{i=1}^{N} f(x_i|\theta)$$

where f( ) is the probability density function, $f(x_i|\theta)$ is the probability of an event of magnitude $x_i$ occurring given distribution parameters $\theta$, and N is the sample size ($\prod$ denotes a product of terms). The parameters which give the largest value of this likelihood function are assumed to be those which "most likely" generated the observed sample, and hence are the maximum likelihood estimates of the distribution parameters. When sample sizes are large, maximum likelihood estimation procedures give the "best" parameter estimates of all possible methods. For smaller sample sizes (less than, say, 50) other methods may give parameter estimates that are statistically more efficient (NERC, 1975). Maximum likelihood estimation procedures are based on function minimisation, and tend to involve iterative solutions. They are therefore difficult to implement, and are usually impractical to apply by hand.

iii. Method of probability-weighted moments (PWM)

The method of probability-weighted moments is a more recent innovation, and was introduced to hydrology by Greenwood et al. (1979). In a sense, it is a generalisation of the more conventional method of moments. The general form for a probability-weighted moment is:

$$M_{r,s,t} = \int x^r F(x)^s (1-F(x))^t f(x)\,dx$$

where f(x) is the probability density function and F(x) is the cumulative distribution function.

If s and t are zero, this reduces to the expression for conventional moments. Probability-weighted moments are analagous to conventional moments, but are defined either by

$$M_{1,0,t} = \int x(1-F(x))^{t} f(x)\,dx$$

or

$$M_{1,s,0} = \int x F(x)^{s} f(x)\,dx$$

The choice is related to the particular probability distribution being used. The parameters can be expressed in terms of an equal number of probability-weighted moments. The latter are estimated from the annual maximum series, $x_i$, by

$$M_{1,0,t} = \frac{1}{N}\sum_{i=1}^{N} [\binom{N-i}{t} / \binom{N-1}{t}]\, x_i$$

or

$$M_{1,s,0} = \frac{1}{N}\sum_{i=1}^{N} [\binom{i-1}{s} / \binom{N-1}{s}]\, x_i$$

where the $x_i$ are ranked in descending order of magnitude.

Probability-weighted parameter estimation procedures tend to be easier to apply than maximum likelihood methods. They generally involve little or no iteration and are more "robust" than conventional moments methods. The uncertainties in sample estimates of high conventional moments arise largely because the observed values are raised to a high power (three in the case of skewness). The derived moments are therefore sensitive to fairly small changes in data values, particularly where these are unusually large or small. Probability-weighted moments are all linear functions of the sample data and are thus much less sensitive to idiosyncracies of the data. PWM-estimation procedures have also been found to be more efficient than MLE procedures for the sample sizes commonly available in hydrology (Hosking et al., 1985).

Other parameter estimation procedures which have been used in frequency analysis include graphical methods and the method of maximum entropy. These are not included in HYFAP.

### B.2.3 Standard errors

An estimate of the magnitude of, for example, the 100-year event made from a sample is not exact. A different sample from the same population would give a different estimate. There is a sampling distribution of possible 100-year magnitudes that could be estimated from a sample of a given length drawn from the same parent population. The standard deviation of this sampling distribution is known as the standard error of the quantile estimate, and gives an indication of the precision of the estimated magnitude with a given return period.

In general terms, the standard error of a quantile estimate is approximated by:

$$se(X_T) \; = \; \left[ \sum_{r=1}^{P} \sum_{s=1}^{P} \frac{dX_T}{d\theta_r} \frac{dX_T}{d\theta_s} cov(\theta_r, \theta_s) \right]^{1/2}$$

where P is the number of parameters and $cov(\theta_r,\theta_s)$ is the variance-covariance matrix between the P parameters. The variance-covariance matrix is specific to a given probability distribution and parameter estimation procedure, and in most cases has to be estimated asymptotically (in other words, by assuming samples are infinitely large).

There are three types of error associated with the estimation of an event with a particular return period:

i. Measurement error due to errors in the sample data.

ii. Model error caused, for example, by the use of an inappropriate probability distribution and model of the rate of occurrence of events over time.

iii. Parameter error due to the estimation of model parameters from a single, possibly short, sample.

It is important to remember that the calculated standard error of an estimate reflects only the last of these three sources of error.

Confidence intervals for an estimate with a given return period T are occasionally calculated from the standard error of estimate, assuming that the sampling distribution of estimates of the T-year event is normal:

$$\hat{X}_T - z_p se(\hat{X}_T) \quad \le \quad X_T \quad \le \quad \hat{X}_T + z_p se(\hat{X}_T)$$

where $z_p$ is the standard normal deviate with exceedance probability, or degree of confidence, p. Unfortunately, this assumption is rarely met in practice. The sampling distribution of the T-year event is highly skewed, particularly for short records, for high return periods and for parent populations with high values of skewness. Upper confidence intervals may be considerably underestimated if it is assumed that the sampling distribution is normal.

### B.2.4 Goodness-of-fit

The HYFAP package does not give a measure of the goodness-of-fit of a particular frequency curve. This is because it is difficult to objectively evaluate the fits of different distributions. There are two basic groups of methods which could, in principle, be used:

i. Statistical goodness-of-fit tests. A variety of statistical tests has been proposed, including the chi-squared test, the Kolmogorov-Smirnov test (see NERC, 1975) and, most recently, the Anderson-Darling test (Ahmad et al., 1988). There are two main problems:

a. The tests tend to concentrate on the fit of the probability distribution in the centre of the data region, and do not pay much attention to the extremes. Unfortunately, it is the extremes which are of greatest interest to a frequency analyst and where alternative estimation procedures show the greatest differences. The Anderson-Darling test, however, has been designed to provide greater discrimination at the extremes.

b. Each test involves the comparison of a calculated statistic and a tabulated critical value to determine the degree of difference between the observed and estimated frequencies. The test would show, for example, that the observed and estimated frequencies are not significantly different at the 5% level. Unfortunately, the critical values depend on the probability distribution and parameter estimation procedure used. The values tabulated in statistical tables assume that the true probability distribution is known. This is not the case in frequency analysis when the parameters must be estimated from the sample. It is therefore not possible to compare the magnitudes of a given test statistic calculated for different frequency estimation procedures, and it would be necessary to compare the statistical significance of each statistic. Tables of critical values are not widely available at present, although Ahmad et al. (1988) produced critical values for the application of the Anderson-Darling test in the special case when frequencies are estimated using the Generalised Extreme Value distribution with parameters determined by the method of probability-weighted moments.

ii. Graphical comparisons. This group of methods relies upon a graphical plot of the fitted frequency curve against the observed data. The fit may be evaluated visually or by some numerical measure, such as the sum of squares of the differences between observed and estimated events (NERC, 1975). The major problem with this approach is that it is necessary to assign an appropriate plotting position to each observation. It may be possible to claim that the fitted frequency distribution does not go through the points because the points are in the wrong place. There are two complications in assigning a plotting position to a given observation:

a. The appropriate plotting position for a given rank observation depends on the probability distribution used (see NERC (1975) and Cunnane (1978) for details about unbiased plotting positions). Fortunately, the differences between various probability distributions tend to be small for all but the highest rank events. Note that the HYFAP package uses the Gringorten plotting position (see Section 3.13 of the manual) which is an approximation to the unbiased plotting position for the EV1 (i.e. Gumbel) distribution.

b. The position at which a point is plotted depends on the period of record used. The largest event in a sample of 27 years, for example, might still be the largest event if the record spanned 50 years and should, therefore, be plotted somewhere to the right of its current position. Alternatively, the 28th year of record might produce a larger event. The previous maximum would therefore plot further to the left as the second largest in a sample of 28.

The observed data will also be associated with a (perhaps unknown) degree of measurement error, and each point should therefore perhaps be better represented as a blurry blob.

Given the above difficulties, the best advice at present is simply to look at a plot of the estimated frequency curves and the observed data, remembering that the points (particularly the large ones) might be poorly located. It is also strongly recommended that return period estimates made at one site are compared with estimates made at other sites in the region (after dividing through by the site mean annual maximum to facilitate comparisons), and that estimates based on site data alone are compared with estimates derived from a regional frequency analysis.

## B.3 PROCEDURES AVAILABLE WITHIN HYFAP

### B.3.1 Extreme Value Type 1 (EV1 or Gumbel) distribution

The Extreme Value Type 1 distribution is a member of the extreme value family of distributions, as proposed by Gumbel (1941). It has two parameters and has been widely used in frequency analysis, partly because of its ease of application and partly due to the theoretical arguments Gumbel developed to support his family of distributions.

Distribution function:

The distribution function has the form:

$$F(x) = \exp\{-\exp\{-(\frac{x-u}{\alpha})\}\}$$

where u is a location parameter and $\alpha$ is a scale parameter. The skewness of the distribution is fixed at 1.14.

The magnitude of an event with return period T can be estimated from:

$$x_T = u+\alpha y_T$$

where $y_T$, the "Gumbel reduced variate" is calculated from:

$$y_T = -\ln(-\ln(1-1/T))$$

Parameter estimation procedures used:

Three parameter estimation procedures are available within HYFAP:

i. Method of moments (after NERC, 1975)

ii. Maximum likelihood estimation (after NERC, 1975)

iii. Probability-weighted moments (after Landwehr et al., 1979)

Comments:

Landwehr et al. (1979) compared the performance of the methods of moments, maximum likelihood and probability-weighted moments using computer simulation experiments, and concluded that the PWM procedure gave the best estimates.

The fixed skewness of 1.14 is lower than skewness values frequently observed in real data, and the distribution often produces low estimates of high return period events.

### B.3.2 Generalised Extreme Value distribution (GEV)

Distribution function:

Jenkinson (1955) developed the Generalised Extreme Value (GEV) distribution to bring together a number of distributions in Gumbel's extreme value family. The distribution function is:

$$F(X) = \exp-[1-k(\frac{X-u}{\alpha})]^{1/k}$$

and the magnitude with Gumbel reduced variate y can be determined from:

$$x = u+\alpha(\frac{1-e^{-ky}}{k})$$

$$\text{where} \quad u+\frac{\alpha}{k} \leq x < \infty \quad \text{if } k<0$$

$$-\infty < x \leq u+\frac{\alpha}{k} \quad \text{if } k>0$$

The parameter u is a location parameter, $\alpha$ is a scale parameter, and k controls the shape of the distribution. If k is zero, the GEV reduces to an EV1 distribution; if k is negative, the GEV is an Extreme Value Type 2 distribution with no upper limit; and, if k is positive, the GEV becomes an EV3 distribution with an upper bound of $u+\alpha/k$.

Parameter estimation procedures used:

Two parameter estimation procedures are available within HYFAP:

i. Maximum likelihood estimation (after Hosking, 1985)

ii. Method of probability-weighted moments (after Hosking et al., 1985)

Comments:

Hosking et al. (1985) used simulation experiments to compare the performance of MLE and PWM estimation procedures, and concluded that the PWM procedure gave better estimates for samples of less than 50 drawn from GEV distributions with "flood-like" parameters.

The GEV distribution is a flexible distribution which has been found to fit flood and rainfall extremes in a variety of environments. A good first attempt is therefore to estimate GEV parameters by PWM.

### B.3.3 Two-parameter lognormal distribution (LN2)

Distribution function:

The normal distribution is rarely appropriate for the analysis of environmental extremes, but the lognormal distribution has been widely used. The probability density function of the two-parameter lognormal distribution is:

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

is a location parameter, and $\sigma$ is a scale parameter.

The magnitude of the event with return period $T=1/p$ can be estimated from:

$$\ln x_T = \mu+\sigma Z_p$$

where $z_p$, is the standard normal variate with exceedance probability p.

Parameter estimation procedures used:

Two parameter estimation procedures are available within HYFAP:

i. Method of moments (after NERC, 1975)

ii. Maximum likelihood estimation (after NERC, 1975)

Comments:

The procedure is simple to apply but, because there are only two parameters, the distribution does not always reproduce the observed data well.

### B.3.4 Three-parameter lognormal distribution (LN3)

Distribution function:

The three-parameter lognormal distribution is a further generalisation of the two-parameter distribution, and takes the form:

$$f(x) \quad = \quad \frac{1}{(x-\gamma)\,\sigma\sqrt{2\pi}}\exp\{-\frac{1}{2}(\frac{\ln(x-\gamma)-\mu}{\sigma})^{2}\}$$

and $\sigma$ are location and scale parameters respectively, and $\gamma$ controls the shape of the distribution.

The magnitude of the event with return period T=1/p can be determined from:

$$\ln(x_T-\gamma) \quad = \quad \mu+\sigma Z_P$$

where $z_p$ is the standard normal variate with exceedance probability p.

Parameter estimation procedures used:

Three parameter estimation procedures are available within HYFAP:

i. Method of moments (after Kite, 1977)

ii. Maximum likelihood estimation (after Kite, 1977)

iii. Probability-weighted moments (after Hosking, 1986)

Note that standard errors not available for LN3-PWM.

Comments:

The three-parameter lognormal distribution is more flexible than the two-parameter version, and LN3-PWM often provides acceptable fits not too different from GEV-PWM.

## B.3.5 Pearson Type III distribution (P3)

Distribution function:

The Pearson Type III distribution is one of a large family of probability distributions developed by Pearson (NERC, 1975). Both the exponential and gamma distributions are special cases of the Pearson Type III distribution.

The distribution function is:

$$f(x) = \frac{(x-x_0)^{\gamma-1}\exp\{\frac{-(x-x_0)}{\beta}\}}{\beta^{\gamma}\Gamma(\gamma)}$$

$x_o$, ß and $\gamma$ control location, scale and shape respectively. $x_o$ constitutes a lower bound. It is difficult to calculate exactly by hand the magnitude of the event with a return period T. The simplest approach is to estimate the T-year event from the sample mean, standard deviation and a factor $K_T$:

$$x_T = \bar{x} + \sigma K_T$$

The factor $K_T$ varies with return period T and sample skewness g, and can be read from tables (US Water Resources Council (1981); Kite (1977) Table 9-2; Linsley et al. (1982) Table 13-4, for example). Sample skewness g can be calculated directly from the sample, or estimated from the parameter $\gamma$ using:

$$g = \frac{2}{\sqrt{\gamma}}$$

Parameter estimation procedures used:

Although it is possible to estimate Pearson Type III parameters by maximum likelihood and, indeed, probability-weighted moments, the HYFAP package only includes the method of moments (as contained in Kite, 1977). The method of moments, however, requires an estimate of sample skewness. Small sample estimates can be very biased and a number of methods have been proposed for correcting for this bias. Apart from the option of making no correction, three alternatives are available. The first is to use the procedure recommended by the US Water Resources Council (1981). The second alternative uses Kite's (1977) suggested bias correction, and the third uses a more complicated correction given by Bobee and Robitaille (1975). The different correction methods produce little difference in estimated quantiles.

Comments:

The Pearson Type III distribution is difficult to apply and, because sample skewness is very uncertain, may give unreliable results.

### B.3.6 Log-Pearson Type III distribution (LP3)

Distribution function:

The Log-Pearson Type III distribution is a refinement to the Pearson Type III distribution. The distribution function is:

$$f(x) = \frac{(\ln x - x_0)^{\gamma-1}\exp\left(\frac{-(\ln x - x_0)}{\beta}\right)}{\beta^{\gamma}\Gamma(\gamma)}$$

$x_0$, ß and $\gamma$ control location, scale and shape respectively. The lower bound of the distribution is $\exp(x_0)$.

Estimates of the magnitude of the event with a T-year return period can be obtained as for the Pearson Type III distribution.

Parameter estimation procedures used:

A method of moments estimation procedure is provided (from Kite, 1977), with options for correcting sample skewness as outlined in Section B.3.5.

Comments:

The Log-Pearson Type III distribution has been recommended for use by federal agencies in the United States (US Water Resources Council, 1981), and has been widely applied. It does, however, have a number of weaknesses:

i. model parameters rely on estimates of site skewness (although the US Water Resources Council did recommend that regional skewness estimates be used to modify site estimates).

ii. The method is difficult to apply by hand.

iii. It has been claimed (e.g., Wallis, 1973) that the Log-Pearson Type III distribution reproduces very poorly the characteristics of real flood data.

The implied lower bound is frequently calculated to be higher than some of the smaller annual maximum events in the data series.

### B.3.7 Regional frequency analysis

While the HYFAP package does not perform a regional frequency analysis, it is possible to apply a regional frequency curve (showing the T-year event as a multiple of some index) to the site data. This is strongly recommended, particularly where the site record is "short" and the required return period is appreciably larger than the available record length. An adopted frequency curve might be some form of compromise between a probability distribution applied to site data for short return periods (up to, say, the period of record), and a regional frequency curve used for longer return periods. This merging of site and regional information is difficult to automate and, to encourage thoughtful estimates, it is preferable that this final step is carried out by the user.

## APPENDIX C: REFERENCES

Ahmad, M.I., Sinclair, C.D. and Spurr, B.D. (1988)
Assessment of flood frequency models using EDF statistics. Water Resources Research 24(8), 1323-1328.

Bobee, B. and Robitaille, R. (1975)
Correction for bias in the estimation of the coefficient of skewness. Water Resources Research 11(6), 851-854.

Cunnane, C. (1978)
Unbiased plotting positions - a review. Journal of Hydrology 37(3/4), 205-222.

Cunnane, C. (1989)
Statistical distributions for flood frequency analysis. WMO Operational Hydrology Report no. 33. WMO: Geneva.

Greenwood, J.A., Landwehr, J.M., Matalas, N.C. and Wallis, J.R. (1979)
Probability-weighted moments: definition and relation to parameters of distributions expressible in inverse form. Water Resources Research 15(5), 1049-1054.

Gringorten, I.I. (1963)
A plotting rule for extreme probability paper. Journal of Geophysical Research 68, 813-814.

Gumbel, E.J. (1941)
The return period of flood flows. Annals of Mathematical Statistics 12(2), 163-190.

Hosking, J.R.M. (1985)
Maximum likelihood estimation of the generalised extreme value distribution. Applied Statistics 34(3), 301-310.

Hosking, J.R.M. (1986)
The theory of probability-weighted moments. IBM Math. Research Report RC12210. Yorktown Heights, New York. 160pp.

Hosking, J.R.M., Wallis, J.R. and Wood, E.F. (1985)
Estimation of the generalised extreme value distribution by the method of probability-weighted moments. Technometrics 27(3), 251-261.

Institute of Hydrology (1983)
Review of regional growth curves. Flood Studies Supplementary Report 14. Institute of Hydrology.

## TERRESTRIAL AND FRESHWATER SCIENCES DIRECTORATE

Natural Environment Research Council
Polaris House
North Star Avenue
Swindon SN2 1EU
Tel: 0793 411500 Fax: 411502
Telex: 444293 ENVRE G

## INSTITUTE OF FRESHWATER ECOLOGY

**Windermere Laboratory**
Far Sawrey, Ambleside
Cumbria LA22 0LP
Tel: 09662 2468 Fax: 6914
Telex: 94070416 WIND G

**River Laboratory**
East Stoke, Wareham
Dorset BH20 6BB
Tel: 0929 462314 Fax: 462180
Telex: 94070672 WARE G

**Edinburgh Laboratory**
Bush Estate, Penicuik
Midlothian, EH26 0QB
Tel: 031 445 4343 Fax: 3943
Telex: 72579 BUSITE G

**Eastern Rivers Laboratory**
Monks Wood Experimental Station
Abbots Ripton, Huntingdon
Cambs PE17 2LS
Tel: 04873 381 Fax: 467
Telex: 32416 MONITE G

**Teesdale Laboratory**
c/o Northumbrian Water
Lartington Treatment Works
Lartington, Barnard Castle
Co Durham DL12 9DW
Tel: 0833 50600 Fax: 50827

## INSTITUTE OF HYDROLOGY

Maclean Building
Crowmarsh Gifford
Wallingford
Oxon OX10 8BB
Tel: 0491 38800 Fax: 32256
Telex: 849365 HYDROL G

**Plynlimon Office**
Staylittle, Llanbrynmair
Powys SY19 7DB
Tel: 05516 652 Fax: 441

**Balquhidder Office**
Tulloch Lodge
Balquhidder, Lochearnhead
Perthshire FK19 8PQ
Tel: 08774 257

## INSTITUTE OF TERRESTRIAL ECOLOGY (NORTH)

**Edinburgh Research Station**
Bush Estate, Penicuik
Midlothian EH26 0QB
Tel: 031 445 4343 Fax: 3943
Telex: 72579 BUSITE G

**Banchory Research Station**
Hill of Brathens, Glassel
Banchory, Kincardineshire, AB31 4BY
Tel: 03302 3434 Fax: 3303

**Merlewood Research Station**
Grange-over-Sands, Cumbria LA11 6JU
Tel: 05395 32264 Fax: 34705
Telex: 65102 MERITE G

## INSTITUTE OF TERRESTRIAL ECOLOGY (SOUTH)

**Monks Wood Experimental Station**
Abbots Ripton, Huntingdon
Cambs PE17 2LS
Tel: 04873 381 Fax: 467
Telex: 32416 MONITE G

**Bangor Research Unit**
University College of North Wales
Deinol Road, Bangor LL57 2UW
Tel: 0248 370045 Fax: 355365
Telex: 61224 BANITE G

**Furzebrook Research Station**
Wareham, Dorset BH20 5AS
Tel: 0929 551518/9 Fax: 551087

## INSTITUTE OF VIROLOGY AND ENVIRONMENTAL MICROBIOLOGY

Mansfield Road, Oxford OX1 3SR
Tel: 0865 512361 Fax: 59962

## UNIT OF COMPARATIVE PLANT ECOLOGY

Department of Animal and Plant Sciences
University of Sheffield, Sheffield S10 2TN
Tel: 0742 768555 Fax: 760159
Telex: 547216 UGSHEF G

## CENTRE FOR POPULATION BIOLOGY

Imperial College, Silwood Park
Ascot, Berks SL5 7PY
Tel: 0344 23911 Fax: 294339

## WATER RESOURCE SYSTEMS RESEARCH UNIT

Department of Civil Engineering, Newcastle University
Newcastle-upon-Tyne NE1 7RU
Tel: 091 232 8511 Fax: 091 261 1182
Telex: 53654 UNINEW G

## UNIT OF BEHAVIOURAL ECOLOGY

Department of Zoology, University of Oxford
South Parks Road, Oxford OX1 3PS
Tel: 0865 271165 Fax: 310447

Natural
Environment
Research
Council

Terrestrial and Freshwater
Sciences