Instat Climatic Guide

Roger Stern, Derk Rijks,
Ian Dale, Joan Knock

January 2006

Contents

Acknowledgements

The first version of this climatic guide was prepared in June 1989. It was translated into French by M. Kadi and colleagues of the Algerian Meteorological Service and into Spanish by P. Ramirez and M. Campas from the Costa Rican Meteorological Service.

We acknowledge the contribution of Huge Hack, a former member of the Department of Meteorology at Reading, who wrote the commands and the original chapter for calculating potential evaporation, using the Penman formula. Also the continued help, criticisms and bug-finding of James Morison and James Milford, former members of the Department of Meteorology, and Mike Dennett from the School of Plant Sciences at Reading.

We are also grateful to the UK Met Office, WMO and ACMAD for providing the finance and support that has made possible the revised Climatic Guide produced in 1997. It was written while Roger Stern was attached to ACMAD, Niamey, for three months in Spring 1997, following discussions with Joan Knock during her visit to ACMAD in March 1997.

Ric Coe, a former member of the Department of Applied Statistics of the University of Reading, now working at the World Agroforestry Centre (ICRAF), Nairobi, prepared the first version of much of the material that was incorporated into the later chapters of the 1997 guide.

We are most grateful to the Met Office in the UK, and to the Statistical Services Centre of the University of Reading, who together funded the production of the Windows upgrade to Instat, and to DFID who funded the upgrading of this guide to the initial (2001) Windows version of Instat and to the later (2005) version.

We would also like to thank the many participants who have attended the SIAC (Statistics in Agro-/Applied Climatology) courses run by the Statistical Services Centre, in Reading, by WMO abroad, and by IMTR, Nairobi and AGRHYMET, Niger. Their suggestions have led to many improvements in Instat.

Chapter 1 – About This Guide

1.1 Who is this guide for?

This guide is concerned with the analysis of climatic data. It is for four types of reader:

Those concerned with the collection of climatic data. This includes staff at meteorological services, but there are many others. For example schools and colleges, farms, agricultural institutes and many individuals who collect climatic records.

The second group is the users who need the results from an analysis of climatic data. They are in many walks of life, including agriculture, health, flood prevention, water supply, building, tourism and insurance.

The other two groups are concerned more with teaching and learning statistics. Looking at climatic data is an application of interest to most people; partly because of the effects that climate has on many other areas.

So the third group is those who teach statistics. This guide shows how simple statistical ideas are used in solving practical problems in one application area. The key concepts of sensible data handling are the same whatever the area.

The final group consists of those who have to learn statistics. Many people recognise that they need statistics skills for their work, but sometimes find their statistics courses are difficult to relate to real life. The materials here are complementary, by starting with the application and considering the statistical ideas that are needed to process the data.

These groups overlap. For example, many users of climatic data are also conscious of their need for training in statistics.

1.2 Why is it needed?

Many organisations have devoted more effort to collecting climatic data than to their subsequent analysis. This is similar to other areas where monitoring data are collected routinely. Sometimes the excuse for the lack of analysis is that the quality of the data is suspect. This is not a good reason, because one way to improve data quality is to analyse the existing data to demonstrate their importance and shortcomings.

It is also useful if those who collect data can do their own analysis. This is highly motivating for staff and an excellent way to encourage good data quality.

Some familiarity with the use of software under Windows is assumed. Knowledge of statistics is useful, but not essential for most chapters. Indeed, though this guide cannot substitute for a conventional statistics book, users do learn many general ideas through seeing where different techniques are useful.

1.3 What software is used?

The analyses use Instat, a simple general statistics package that also includes a range of special facilities to simplify the processing of climatic data.

In general Instat is used for 3 purposes.

It is designed to support the teaching of statistics. Most statistics packages are designed to process data, though they can be used in teaching. Instat is the other way round.

It is designed to support the analysis of climatic data.

It introduces what is meant by a statistics package. This is mainly for those who currently use only spreadsheets for their statistical work. Where Instat is not powerful enough, it can be a stepping-stone, to a more powerful statistics package.

Instat is similar to, but is not as powerful as the standard commercial statistics packages. It was first developed in the early 1980s as a simple general statistics package that was designed particularly to support the teaching of statistics. The facilities for climatic analysis were also added from the 1980s.

1.4 What is in this guide?

Illustrations and 'route maps' are provided in all chapters for those who just wish to study particular topics. We hope that some users will enjoy the way the ideas unfold in successive chapters, but do not assume that readers will wish to look at every chapter.

How much practice is needed, depends on users current experience in statistical computing. Those who are relatively inexperienced, or have never used a statistics package, may be surprised at how easy the ideas and the software are. However, beginners need practice, so just reading the guide will not be so effective.

Those with some experience of a statistics package, should find that Instat is similar to most other packages. The practice is not so important, because they should be able to visualise the results from just reading the text.

The material up to Chapter 7, uses only simple statistical ideas. Those who are fearful of statistics should find there is little difficulty and also be pleasantly surprised how much they have learned of processing climatic data and of the concepts of data processing generally.

Chapters 8 to 10 show how data analysis in any field is more interesting if users know about the background to their subject. As illustration there are ideas from biology in Chapter 8 and physics in Chapter 9, though neither are essential for the key ideas in either chapter.

More general ideas of statistics are used from Chapter 11 and other statistical packages are mentioned, in addition to Instat.

All the data sets used for illustration are supplied, both as Instat files, and as Excel spreadsheets. Chapter 3 describes how the data are organised, so readers can substitute their own data for the examples in later chapters.

The analysis of the climatic records is often a two-stage process. The first stage reduces the raw, often daily, data to a semi-processed form in terms of key summaries that correspond to users' needs. The second stage involves processing these summaries.

This two-stage process is typical of the processing of many types of data, and is one reason why users often find their statistics course did not seem relevant to real-world problems. Many courses use only small sets of semi-processed data that are tailored to the particular topic being taught. However, the real world starts with the raw data, and these are often quite large. Chapter 4 describes how to look at the raw data and Chapters 5 and 6, show how these data can be reduced in ways that can be "tailored" to the needs of the user.

The tools for this data reduction phase constitute most of the climatic part of Instat.

Chapter 7 considers simple ways that the summarised data (from the first stage) can be analysed. By the end of this chapter we will have covered many of the key concepts that are needed for a full analysis of the climatic data. Readers may also be surprised that although we are halfway through, we have yet to introduce ideas that are more complicated than descriptive statistics.

The examples in these first chapters are mainly of rainfall records, though the methods apply equally to other climatic variables, like wind speed and temperature data. Rainfall is however the most commonly measured and in many countries it is the single most important variable.

Chapters 8 to 10 are slightly more specialised and may be of less interest to the general statistical audience, so we first describe the remaining chapters. Chapter 11 illustrates a range of topics that include the analysis of extremes, the use and misuse of correlations and the use of regression methods to study crop-weather relations. Then Chapter 12 gives 3 case studies, the first of which looks at 5 alternative methods of analysing rainfall data. This is based on a real problem, and illustrates that, unlike in most training courses, there may be no best method of analysis. The other examples concern a simple irrigation problem and the processing of storm data, where records are available every 5 minutes.

Chapter 13 is the most ambitious. It describes a more recent method of analysing rainfall records that is much more precise than the methods in common use. Hence it could be used on quite short records, of perhaps 5 to 10 years, and give the same precision as with a 30-year record, using the simpler methods described in earlier chapters. This method is not new to statisticians. What is newer is the attempt to describe and illustrate the method to a wide set of readers, most of who are statistical sceptics.

If the claims made in this chapter are correct, then one application is to split long records into subsets and try to quantify any climate change. So, while that the ideas are usable, even within schools, their application is one of modern research.

Chapter 8 is devoted to the analysis of temperature data.

Chapter 9 describes agricultural climatology in general and then considers one aspect in detail, namely the calculation of evaporation. The evaporation is the partner to the rainfall; it goes up while the rainfall comes down. The suggested method of calculating evaporation makes use of other climatic variables, usually sunshine, temperature, wind speed and humidity. This subject is beloved of climatologists, because most have a physics background and they have here an important application of these ideas. The more general reader may omit sections of this chapter, but what is useful to see is how the quite complex ideas translate into simple Instat dialogues that anyone can use.

Rainfall and evaporation are used together in Chapter 10, which describes a simple crop index used for drought monitoring purposes. This is a useful index in its own right; also the methods of analysis are the same as for many other more complicated crop models.

1.5 What else?

Readers who are not confident in computing or statistics should recognise the three different subjects that are described here, namely climatology, statistics and computing. The material becomes easier if you separate these subjects as far as possible.

Those who are adept in computing discover that sensible use of the computer is of great help in their understanding of statistics. In contrast, beginners to computing, sometimes find it difficult to use the computer in their statistics courses. They are still trying to master the computing ideas and this becomes mixed with the statistical objectives.

Computing ideas are raised at various points in this guide, because users of Windows software sometimes limit the analyses they conduct, by not exploiting the software fully. So we show how Instat can be used in different ways, to solve problems raised by users in their needs for data analysis. These sections should be recognised as computing topics and perhaps omitted initially by those who are less experienced in computing.

Statistics and climatology have two features in common. Both are relevant to a wide range of applications and many specialists in those application areas treat both statisticians and climatologists as an unwelcome nuisance! Perhaps by working together they can be welcomed a little more.

Chapter 2 – Delivering products

2.1 Introduction

To use climatic data fully it is important to be able to deliver products. The examples in this chapter describe the steps and the end-point in this process. Data are supplied in the right form for the analysis. The objectives are specified and your task is to prepare the tables and graphs for a report and a presentation.

The problem builds on a study in Southern Zambia. This is the most drought-prone area of the country. Everyone knows that there is 'climate change'! Some farmers are emigrating North, citing climate change as their reason. However a local non-governmental organisation (NGO) called the Conservation Farming Unit, questions this reasoning. They are not convinced that any climate change has necessarily affected the farming practices. They therefore commissioned a study that used daily climatic data from a number of stations in Southern Zambia. The results were supplied as a report, and presentations of the results were also made to the NGO and to the local FAO Officers. The key conclusions were later made into a number of short plays that were broadcast on local radio, and also played at village meetings.

This chapter uses data from Bulawayo in Zimbabwe, which has a similar pattern of rainfall to Southern Zambia. We have been supplied with the daily data on rainfall and maximum and minimum temperatures, from 1951 to 2001. Here, partly for simplicity, we largely use the monthly summaries.

For the work in this chapter we draw an analogy with the preparation of a meal. The first key requirement is that you have the food, which here is the climatic data. In a real meal, the food may be supplied in a form that is ready for cooking, or it may need preparation prior to cooking. Here the data are in pre-packed form, so the analysis can proceed quickly.

You also need the right tools. In a kitchen they are the saucepans, etc, while here they are the computer, together with the required software.

You need some general cooking skills. These are the basic computing skills, plus initial skills of Instat, at least from the tutorial.

Finally your objectives must be clear. This corresponds to having a specific meal in mind, so that a recipe can be used. Of course you may have to adapt slightly as you go along. You might find some oddities in the data, just as cooks have to improvise if they suddenly find that one of the ingredients is not available.

If everything is well organised, the cook can prepares the meal very quickly. This is just what is done in the products in this chapter. This leaves time to make sure the dishes, for us the results, are presented attractively. Then users will enjoy consuming what is presented.

Section 2.2 describes the data and the production of the initial presentation graphs and tables. In Section 2.3, they are incorporated into a report and a presentation. Then we look at how further results can be produced, in Section 2.4. Finally we check what users need, so they are able to deliver products in the future, with similar ease.

2.2 The data

Monthly data are used in this chapter. Daily data are the starting point in most of this guide, because many of the objectives require daily data. But here the emphasis is on objectives for which the monthly data are suitable.

The data are already in an Instat worksheet. Hence they can be opened from the library in Instat library.

Note: lines that start with ⇨ indicate instructions that you should follow.

⇨ From the desktop **double click** on the Instat Plus icon

From the opening screen select **File ⇒ Open From Library** as shown in Fig. 2.2a. Choose the file called **bulmon.wor**, which is from ***Bul***awayo, with ***mon***thly data.

Fig. 2.2a File menu

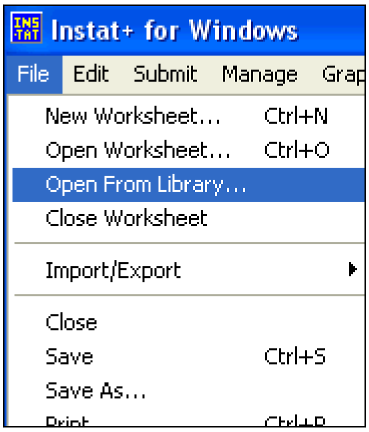
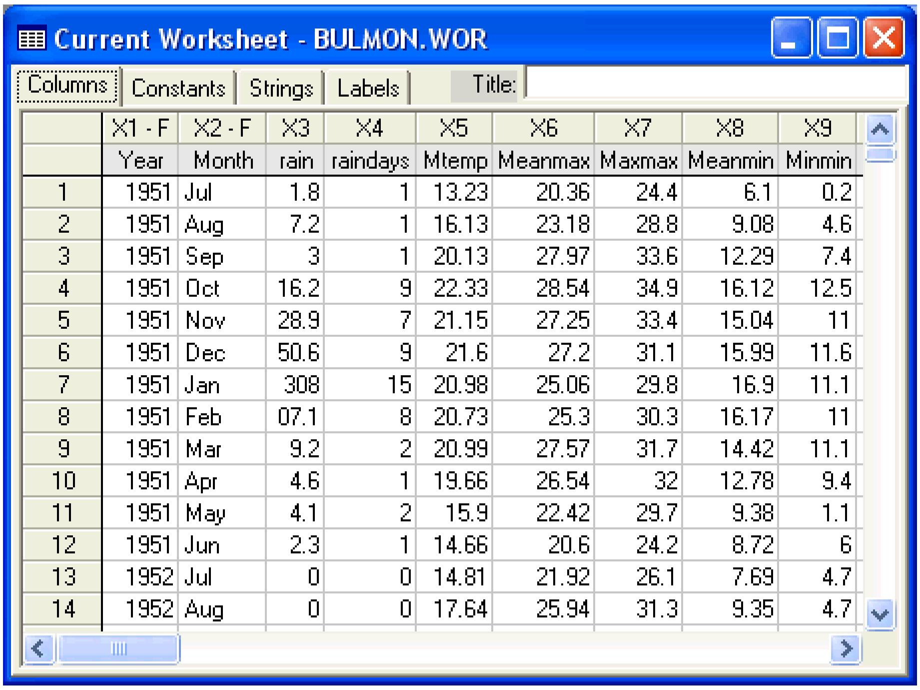


Fig. 2.2b Monthly rainfall and temperature data from Bulawayo



	X1 - F	X2 - F	X3	X4	X5	X6	X7	X8	X9
	Year	Month	rain	raindays	Mtemp	Meanmax	Maxmax	Meanmin	Minmin
1	1951	Jul	1.8	1	13.23	20.36	24.4	6.1	0.2
2	1951	Aug	7.2	1	16.13	23.18	28.8	9.08	4.6
3	1951	Sep	3	1	20.13	27.97	33.6	12.29	7.4
4	1951	Oct	16.2	9	22.33	28.54	34.9	16.12	12.5
5	1951	Nov	28.9	7	21.15	27.25	33.4	15.04	11
6	1951	Dec	50.6	9	21.6	27.2	31.1	15.99	11.6
7	1951	Jan	308	15	20.98	25.06	29.8	16.9	11.1
8	1951	Feb	07.1	8	20.73	25.3	30.3	16.17	11
9	1951	Mar	9.2	2	20.99	27.57	31.7	14.42	11.1
10	1951	Apr	4.6	1	19.66	26.54	32	12.78	9.4
11	1951	May	4.1	2	15.9	22.42	29.7	9.38	1.1
12	1951	Jun	2.3	1	14.66	20.6	24.2	8.72	6
13	1952	Jul	0	0	14.81	21.92	26.1	7.69	4.7
14	1952	Aug	0	0	17.64	25.94	31.3	9.35	4.7

⇨ Examine the worksheet (see Fig. 2.2b). The contents are as follows:

Column	Type	Description
X1-Year	Factor(F)[1]	There are 50 years of data, from July 1951 to April 2001. The 'year' 1951 is from July 1951 to June 1952
X2 - Month	Factor(F)	The levels of the factor are from July to June, to correspond to the rainy season
X3 -rain	Variate	Monthly rainfall total (mm)
X4 - raindays	Variate	Number of rain days in each month (rain > 0.85mm)
X5 - Mtemp	Variate	Mean monthly temperature (degrees C)
X6 - Meanmax	Variate	Mean of the daily maximums in the month
X7 - Maxmax	Variate	Highest of the daily maximums
X8 - Meanmin	Variate	Mean of the daily minimums
X9 - Minmin	Variate	Lowest of the daily minimums

There are 598 rows of data in this worksheet, i.e. 50 years by 12 months (less a couple of months).

The task is to write a short report that describes the patterns of rainfall and temperatures. One aim is to assess whether there is obvious evidence of climate change. This evidence might justify requesting the detailed (daily) data, to undertake a more detailed study.

The first step is to explore the data, and then consider how appropriate results could be presented. To explore the data we start with a boxplot of the rainfall totals.

[1] A 'factor' is a 'category' column, to allow summaries, etc to be given for each 'level'. For example the Month factor in Fig. 2.2b has 12 levels, going from July to June.

From the **Graphics** menu, choose **Boxplot** (Fig. 2-2c).

⇨ Complete the dialogue as shown in Fig. 2.2d.

Fig. 2-2c Graphics menu

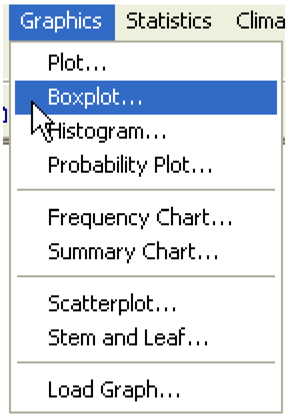
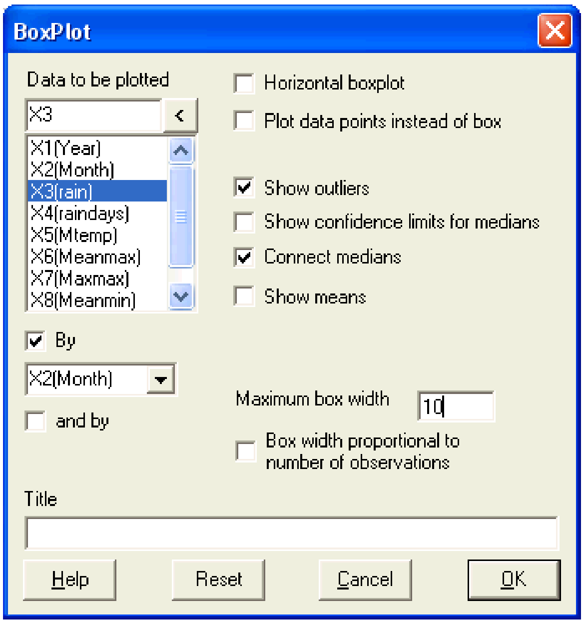


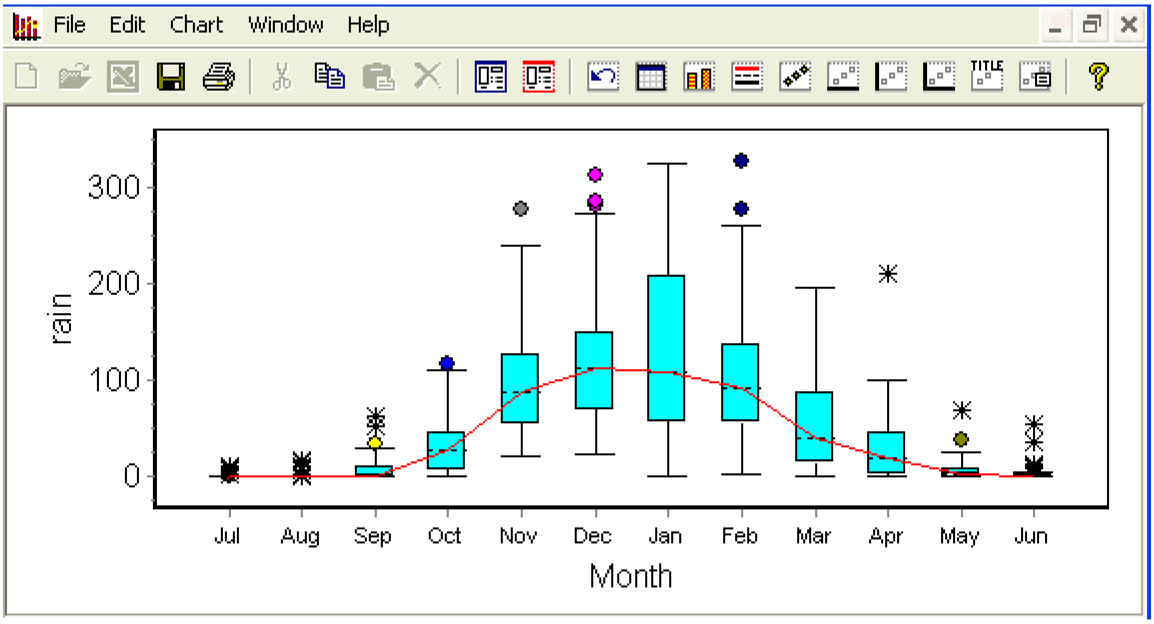
Fig. 2-2d Boxplot dialogue



The resulting plot is in Fig. 2.2e. It looks almost as though it could be tidied and put into a presentation, or report. But at this early stage we are looking for oddities in the data.

There seems little that is odd, except perhaps that the minimum value in January, which is one of the rainiest months, is zero, unlike November, December or February. As an example, we investigate further.

Fig. 2.2e Resulting boxplots, showing also the graphics menu and toolbar



As a first check, the steps below get the same information as Fig. 2.2e, but in a table.

From the **Statistics** menu, choose **Tables** and then **Summary** (Fig. 2-2f).

Complete the dialogue as shown in Fig. 2.2g.

Fig. 2.2f Statistics menu

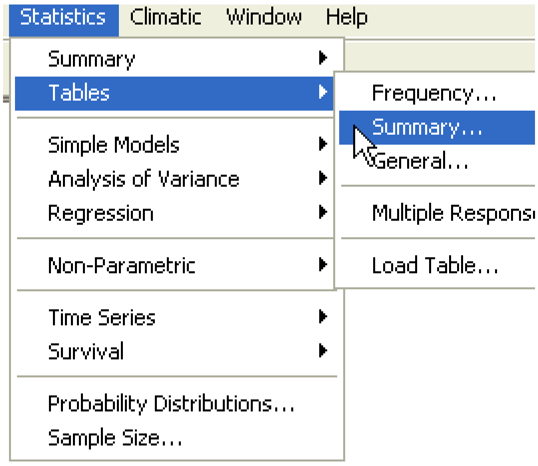
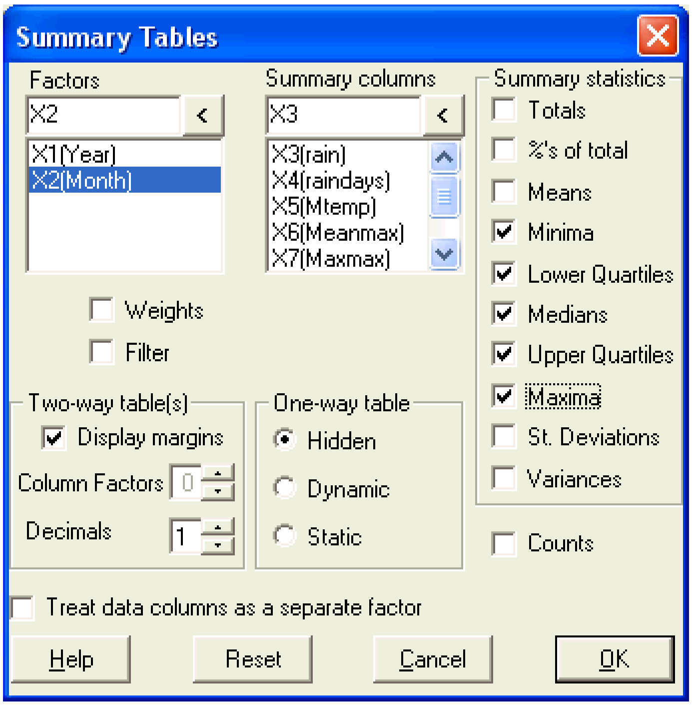
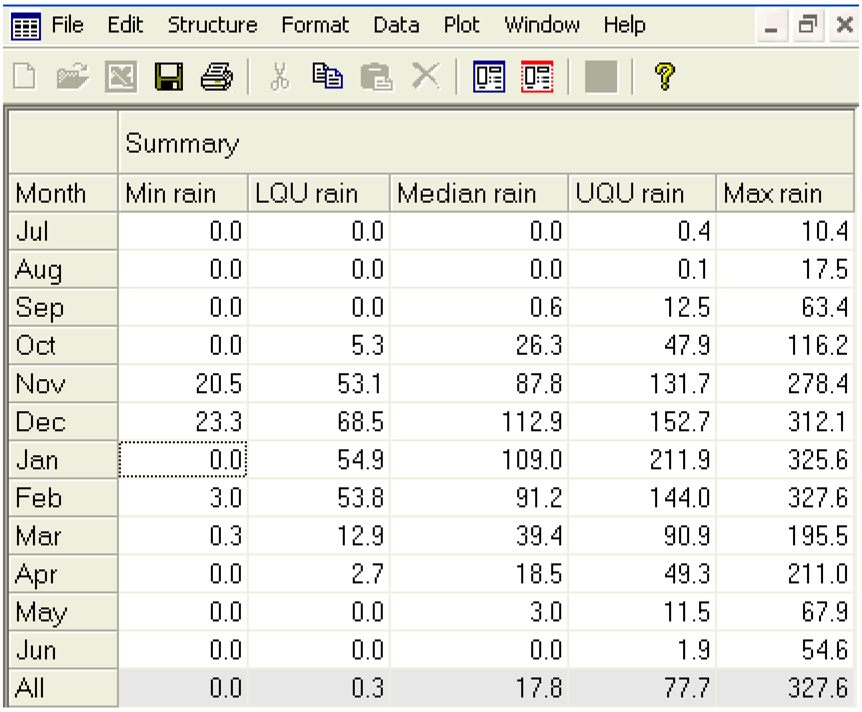


Fig. 2.2g Summary tables dialogue



The resulting table is shown in Fig. 2.2h. Look particularly at January, to see that the years do range from zero to over 300mm.

Fig. 2.2h Resulting table[2], showing also the tables menu bar



	Summary				
Month	Min rain	LQU rain	Median rain	UQU rain	Max rain
Jul	0.0	0.0	0.0	0.4	10.4
Aug	0.0	0.0	0.0	0.1	17.5
Sep	0.0	0.0	0.6	12.5	63.4
Oct	0.0	5.3	26.3	47.9	116.2
Nov	20.5	53.1	87.8	131.7	278.4
Dec	23.3	68.5	112.9	152.7	312.1
Jan	0.0	54.9	109.0	211.9	325.6
Feb	3.0	53.8	91.2	144.0	327.6
Mar	0.3	12.9	39.4	90.9	195.5
Apr	0.0	2.7	18.5	49.3	211.0
May	0.0	0.0	3.0	11.5	67.9
Jun	0.0	0.0	0.0	1.9	54.6
All	0.0	0.3	17.8	77.7	327.6

Prepare a two-way table, as shown in Fig. 2.2i, to look at the individual values.

[2] In this table the bottom row is called 'All'. This is called the 'margin' of the table, and in this case it gives the annual values. You may choose to have tables with, or without their margins. With rainfall, the margin will often be the annual total, as in Fig. 2.2j.

The results are in Fig. 2.2j, and confirm there was no rain in January in 1955. You still need to check if this was the only such year.

Fig. 2.2i two-way table

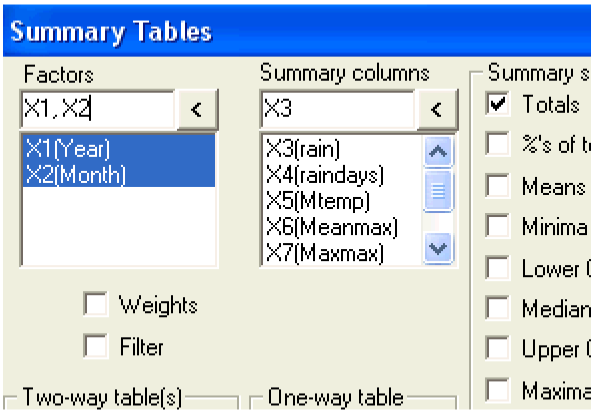


Fig. 2.2j Resulting table, with zero in January

	Month						
Year	Jan	Feb	Mar	Apr	May	Jun	All
1951	308.0	107.1	9.2	4.6	4.1	2.3	843.0
1952	240.6	111.1	70.1	37.0	21.5	0.0	930.8
1953	200.8	108.1	42.0	25.3	2.6	1.3	526.6
1954	213.7	260.9	102.0	57.1	17.3	12.2	1093.5
1955	0.0	214.0	59.7	86.9	0.3	0.0	563.2
1956	78.7	139.6	92.2	16.7	24.2	0.0	606.0
1957	216.5	91.3	29.2	31.8	0.0	0.0	590.6

Use **Structure ⇒ Sort**, Fig. 2.2k and sort on the January data, Fig. 2.2l.

Fig. 2.2k Table menu items

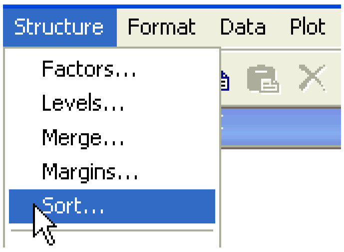
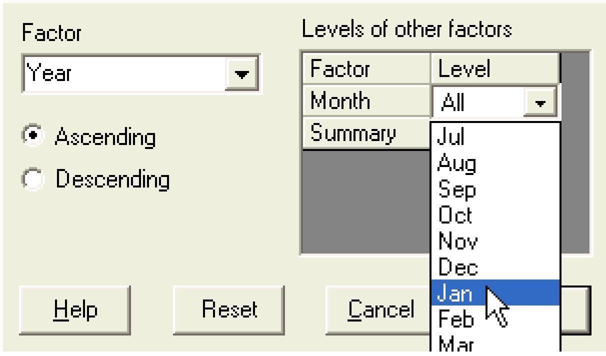


Fig. 2.2l Sorting on January rainfall



The results are in Fig. 2.2m and show there was just one year with zero rainfall in January. The daily data could be used to look more closely, but with zero in just one year, there does not seem much cause for concern, so we proceed with the analysis.

Fig. 2.2m Data sorted on the January values

	Month						
Year	Jan	Feb	Mar	Apr	May	Jun	All
1955	0.0	214.0	59.7	86.9	0.3	0.0	563.2
1969	5.8	69.4	10.4	18.3	0.0	3.5	336.0
1959	26.3	72.2	44.5	14.9	13.0	1.8	329.1
1964	26.6	5.9	2.6	1.0	4.6	0.0	400.5
1976	30.5	245.3	120.5	1.2	5.8	0.0	750.9
1982	34.3	26.9	40.6	62.4	13.1	2.7	433.6
1967	36.2	113.9	10.4	68.6	16.7	0.0	404.4

One task is to give an idea of the pattern of rainfall in good and bad years. One way is through a table that shows the two years with lowest total rainfall, the two with highest and two with average total rainfall. We are aiming for the table in Fig. 2.2n.

Fig. 2.2n Monthly rainfall from 6 'special' years

	Month							
Year	Oct	Nov	Dec	Jan	Feb	Mar	Apr	All
1972	40	43	48	108	24	11	13	288
1963	33	26	98	78	51	2	1	288
1985	23	42	119	139	16	11	211	561
1956	5	118	114	77	140	92	17	562
1954	53	104	273	214	261	102	57	1064
1977	22	96	312	298	278	120	88	1215

This is an example of a presentation table, (rather than a table for data exploration). The different rows indicate what the monthly rainfall may be like, in good, bad and 'middling' years.

There are some decisions to be made, before the table can be produced. Thus:

- Is the table to be for all 12 months, or just those in the rainy season? We decide on just the months October to April.
- We need the 'worst' and 'best' years. With just the monthly data, we use the annual totals for the definition. Will the totals be defined on the whole year, or just the 7 months of the rainy season? We decide on the seasonal total.

The table is currently as shown in Fig. 2.2m. The following steps transform it into Fig. 2.2n.

From the table menu, choose **Structure ⇒ Levels**, Fig. 2.2o. Complete it as shown in Fig. 2.2o. The option to exclude the levels is so they do not contribute to the total.

Fig. 2.2o Excluding months

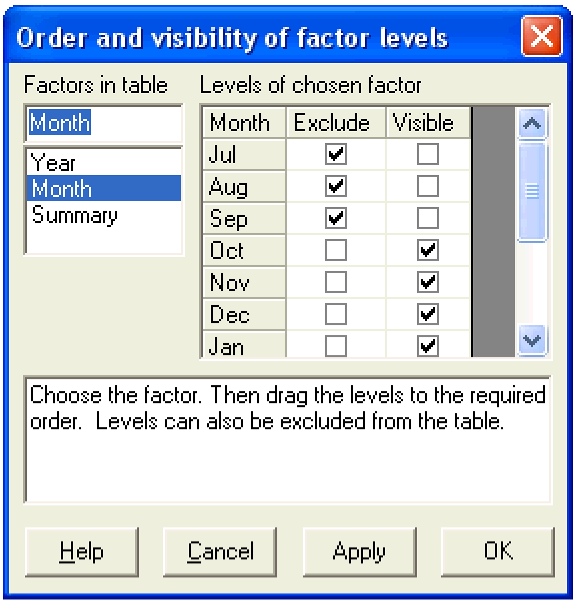
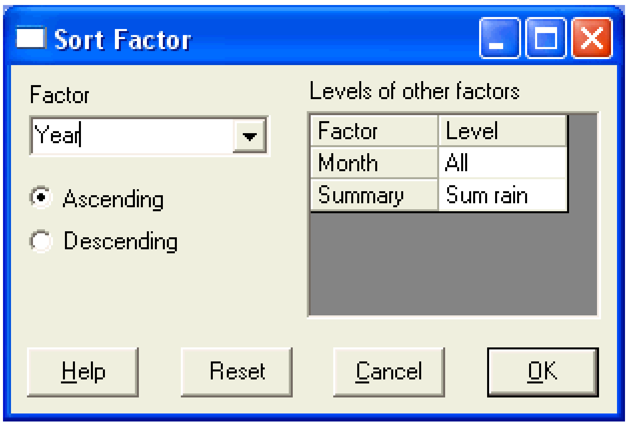


Fig. 2.2p Sorting on the seasonal total



Return to the **Structure ⇒ Sort** dialogue, Fig. 2.2p. Sort this time on the seasonal rainfall totals.

Return to the **Structure ⇒ Levels** dialogue, Fig. 2.2q, and complete it as shown. Keep the two years at the top, the two at the bottom, and 1985 and 1956 as two years round the middle.

Now use **Structure ⇒ Margins**. The final table, Fig. 2.2n, includes the side margin, but not the bottom margin. Complete the dialogue as shown in Fig. 2.2r.

Fig. 2.2q Keeping just the extreme and average years

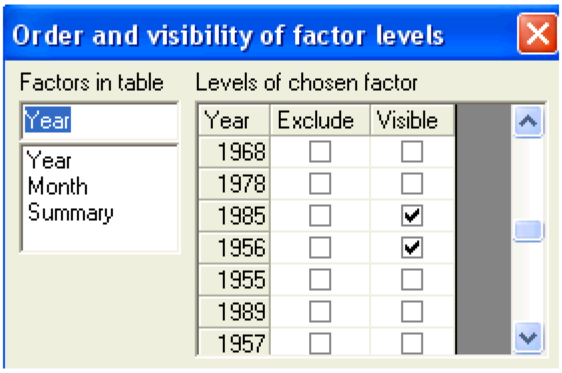
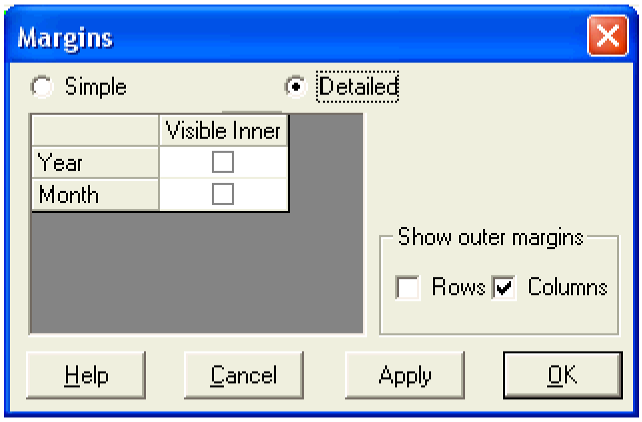


Fig. 2.2r Keeping just one margin



It remains to add the shading, which gives the impression of a series of bar charts, see Fig. 2.2n. This uses the **Format ⇒ Cells** dialogue, Fig. 2.2s.

In Fig. 2.2s set the **Apply** option to **Factor level**, and choose the **Month** factor. Then set the shading separately for each month.

We use the final column (maximum rainfall) in Fig. 2.2h to give the **To** value in Fig. 2.2s for each month. The degree of shading will indicate the value in a particular year in relation to the year with the maximum.

The example in Fig. 2.2s is for January. Choose the **To** value, and also the **Fill colour** separately for each month.

Fig. 2.2s Creating a 'bar-chart' effect

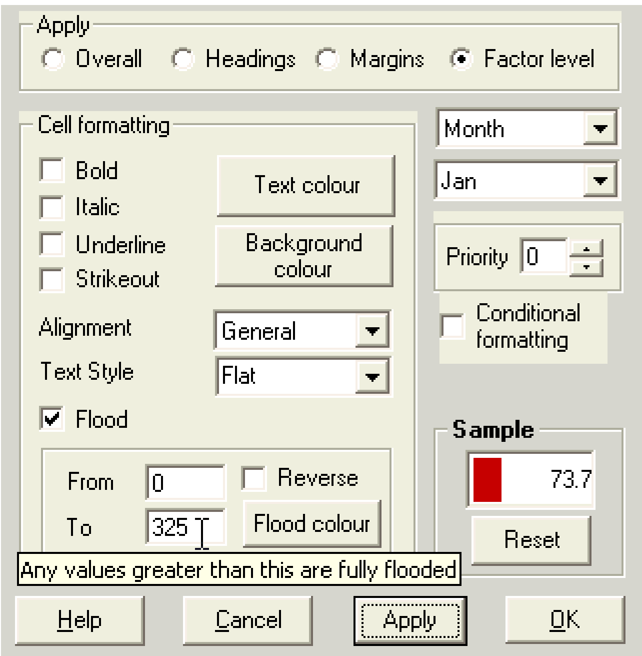
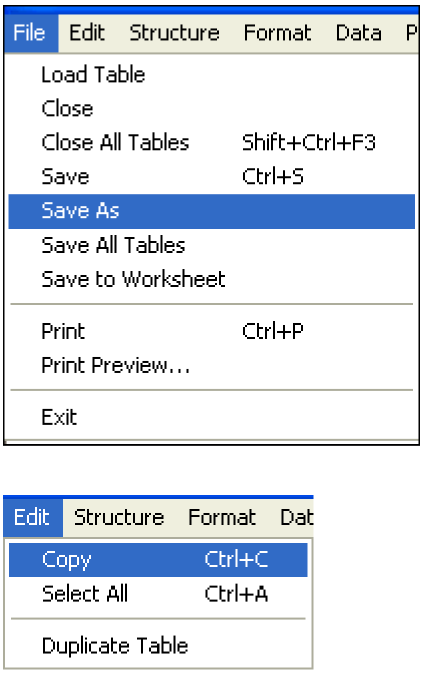


Fig. 2.2t Saving the table



The table is now as in Fig. 2.2n, and is to form part of a presentation. To save it, use **File ⇒ Save As**, Fig. 2.2t. Saving the results as a table, for Excel, or Word is easy. But this does not save the shading. Therefore save (for now) so it can load back into Instat at a later date. So choose **itb** (Instat Table Format), and give it the name **bulrain.itb**.

To copy directly into a report we will later use **Edit ⇒ Copy**. This puts a copy into the Windows clipboard, which can then be pasted (as Paste Special), into Word or Powerpoint.

We have shown how a table can be used for data exploration and also for presentation. The same applies to the graphs. Suppose the boxplot, shown in Fig. 2.2e, is suitable for a presentation graph. As with the table, check first whether there are any general improvements and one is perhaps to add the means to the display.

From the **Graphics** menu, select **Boxplot** again (Fig. 2-2u) and add the option to show and connect the means.

Fig. 2.2u Boxplot with means

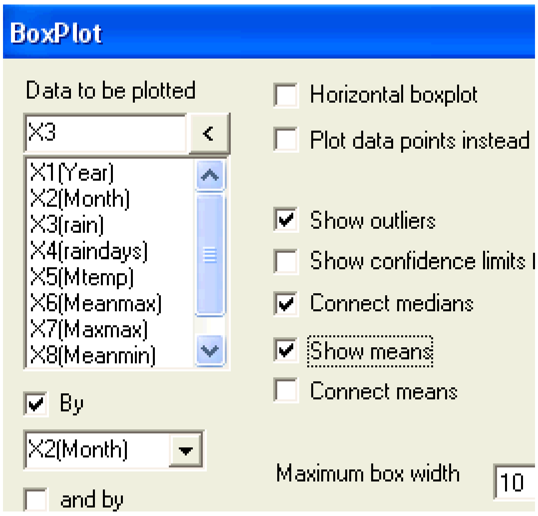
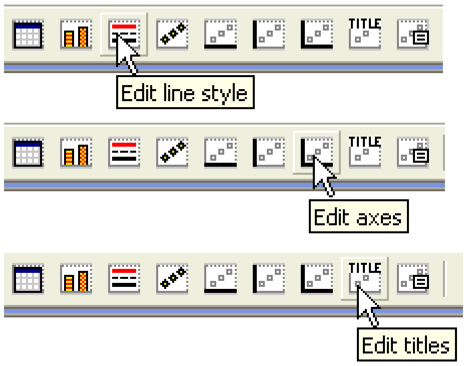


Fig. 2.2v Editing the graph



Use the toolbar to edit the graph to give something like Fig. 2.2x, as follows.

- ⇨ Edit the line style to give a thickness of 2 for the means and medians.
- ⇨ Edit the y-axis to give a more informative label, to give the tick marks at intervals of 100, and to add a grid.
- ⇨ Edit the x-axis so there is no label.
- ⇨ Edit the titles, to add a footnote that is right justified.

Fig. 2.2x The graph ready for a report

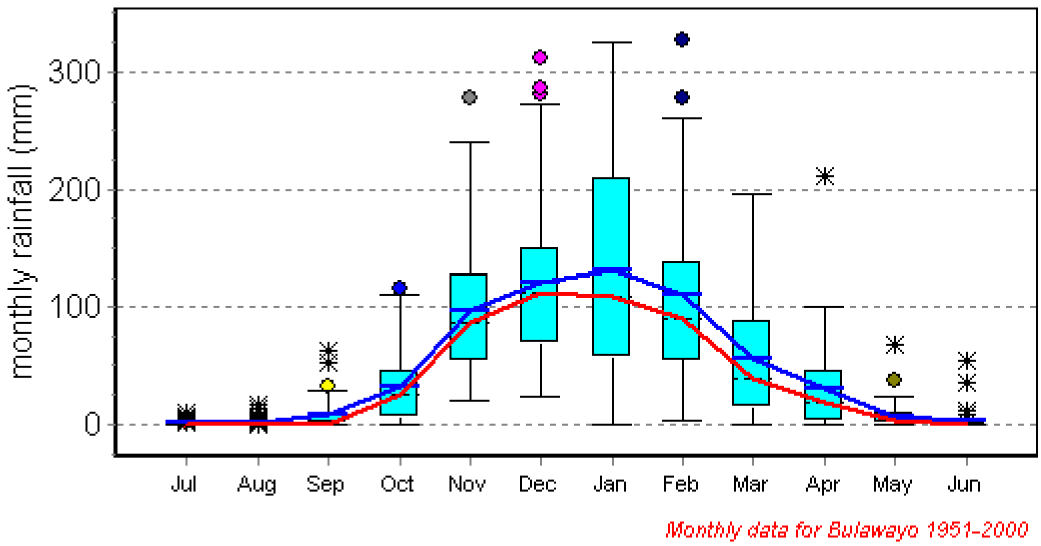


Fig. 2.2x indicates that the data are slightly skew – the mean is usually slightly higher than the median. The median rainfall is about 100mm from November to February. These are the main months so this is consistent with an annual median of 563mm and a mean of 603mm[3].

Fig. 2.2y Save or export the graph

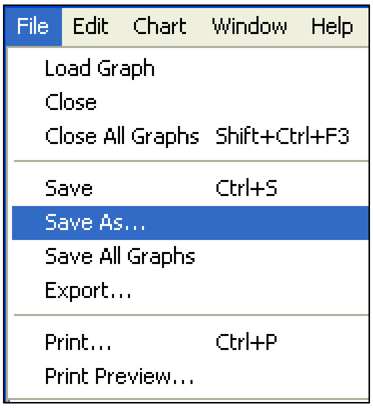
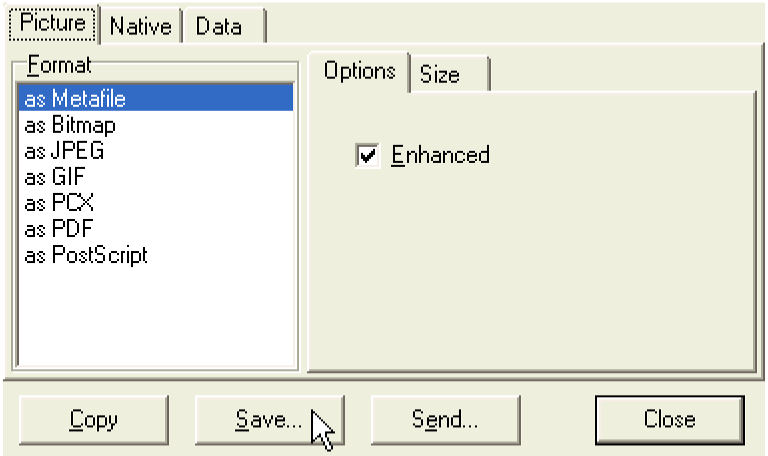


Fig. 2.2z Saving as an emf file



Save this graph using **File ⇒ Save As**, see Fig. 2.2y. Give it the name **bulmon1.igt**[4]. It can now be read back into Instat on a future occasion, for further editing if needed.

Also use **File ⇒ Export** to save a version, see Fig. 2.2z, called **bulmon1.emf**[5]. This is now ready to be inserted into a report or presentation.

In this section we have used graphs and tables, both to explore the (rainfall) data, and to prepare a graph (Fig. 2.2x) and a table (Fig. 2.2n) for a report. In the next section we include these figures in a report and a presentation.

2.3 Including results in a report and presentation

We start by showing how a graph can be included in a report. There are three possibilities:

1. The graph is prepared in Instat, for example as in Fig. 2.2x. This graph is to be exported as a picture, which is then to be displayed in a document, or a presentation.
2. The graph is prepared in Instat. Then the data behind the graph is to be exported to a preferred graphics package, for example to Excel. The graph is then constructed in Excel and put into the report.
3. The data are copied from an Instat worksheet into another package, e.g. Excel. Then the graph is constructed in that package.

For the first route, with the graph as the current Window, choose **File ⇒ Export** as shown earlier in Fig. 2.2y. This gives the dialogue shown in Fig. 2.2z. A popular format is then to save the file as an Enhanced Metafile (emf), as shown in Fig. 2.2z.

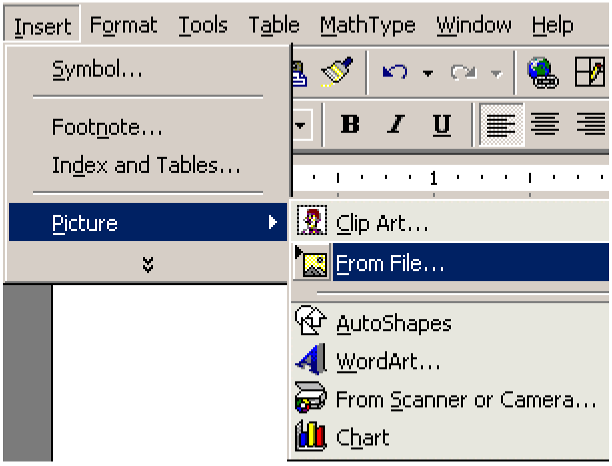
Then, if using Microsoft Office, go into either Word or PowerPoint. From within either package use **Insert ⇒Picture ⇒ From File** as shown for Word in Fig. 2.3a.

[3] One way to get the annual values is to use **Statistics ⇒ Summary ⇒ Column Statistics**. Set **X3** as the data column, and **Year** as the **By** factor. Produce the Sums, and save them, possibly into X10. This gives the annual total each year. Then use **Statistics ⇒ Summary ⇒ Describe** on X10 to give the mean of 603mm and median of 563mm.

[4] In the file extension igt, the i and g are for Instat graphics. The t is because we are using an add-in to produce the graphics called TeeChart.

[5] The extension emf stands for 'enhanced metafile'. This is a standard format for a graphics file, that can be read into many other software packages.

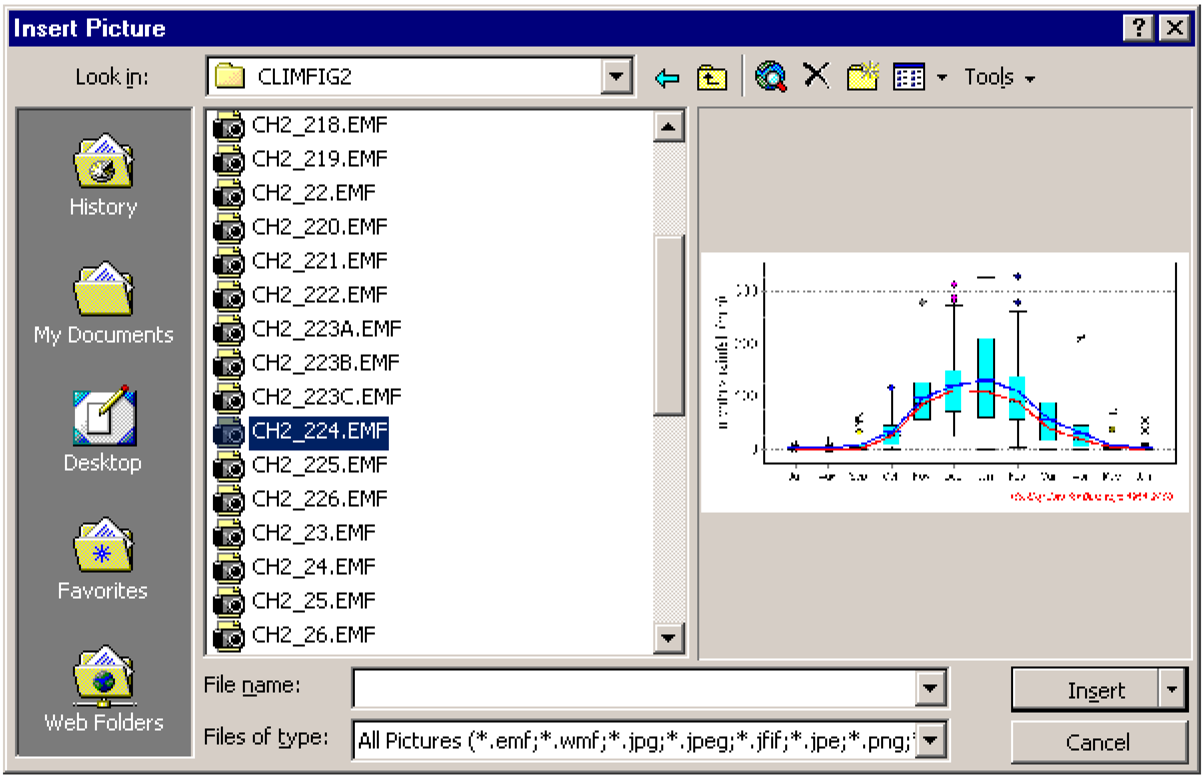
Fig. 2.3a Microsoft Word menu to insert a picture



Then browse to find the file you saved, Fig. 2.3b, and click to insert it in the document or presentation.

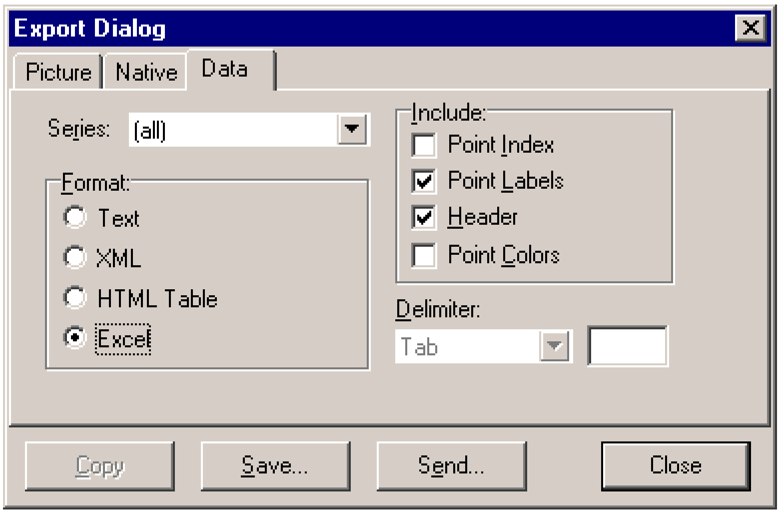
If the picture in Fig. 2.3b fails to appear, then try a different format. For example the file would be saved as an ordinary metafile (extension wmf) if you untick the option in Fig. 2.2z. This sometimes imports when emf files do not.

Fig. 2.3b Microsoft Word dialogue to insert a picture



For the second route, produce the graph in Instat, as before, but there is no reason to edit the picture in Instat. Use **File ⇒ Export** as above to give the dialogue in Fig. 2.2z. Then use the tab to export the data, see Fig. 2.3c.

Fig. 2.3c Instat's File ⇒ Export dialogue to export the data to Excel



For the third route, where the graph is constructed using a different package, you export the data from Instat, possibly by marking it in the worksheet and copying to the clipboard, and then into your graphics package, or Excel. Then you proceed with that package.

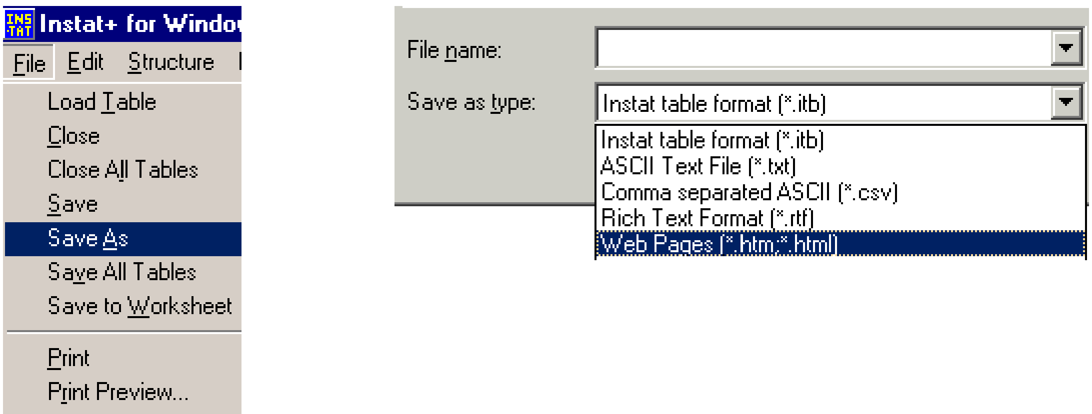
Within Word, your document will look much more consistent and professional if you use styles consistently. Perhaps the computing section in your organisation has prepared a standard template, so that all Word documents have headings, page numbering, figures and tables that are presented in a consistent manner.

Exporting an Instat table is a similar process to the export of a graph. The options are described in more detail in the Instat Introductory Guide, Section 13.11, and are outlined here. From the Instat table, you have two main alternatives:

1. Export the data in the table to Excel or to a Word table, so the numbers, and headings can be edited.
2. Export to Word or Powerpoint as a picture.

To export the data, make sure the table in Instat is the active Window, and use the **File** menu. Fig. 2.3d shows the menu (left) and some of the different export formats (right).

Fig. 2.3d Exporting a table from Instat



Choosing html (see Fig. 2.3d) exports to a format that can be read by Excel with most of the formatting intact. The data can be edited further in Excel, or transferred to Word or PowerPoint.

An alternative, is to save as an rtf file. This does not preserve colours of the cells, etc, in Word, but does transfer as a Word table, which can then be edited and formatted within Word.

Some features from an Instat table, like the flooding, have no equivalent in Word or Excel. Save the table as a picture, if these are to be transferred, or if the table is to be put straight into a Powerpoint presentation. To do this, use **Edit ⇒ Copy**, with the table as the active active window. Then go into Word or Powerpoint, and use **Edit ⇒ Paste Special**, and choose the option to copy the file as a picture.

Finally, if you are copying pictures a lot, then it is useful to have screen capture and editing software. One option is called Paintshop Pro, but there are many alternatives, for example Grabit, or Snagit. They offer options of saving pictures in a range of formats, as well as choosing parts of the screen to capture, etc. For example, if you wished to capture an Instat dialogue, or menu option, then it could be done by using the Print Screen key, and editing the screen in Word. An alternative with special software would be to capture just the dialogue, or the particular part of the screen.

2.4 Practice

The further examples used in this section are with the number of rain days and the temperature data.

Start with an exploratory plot of the mean temperatures.

Use the boxplot dialogue, with X5, as shown in Fig. 2.4a.

Fig. 2.4a Boxplot dialogue

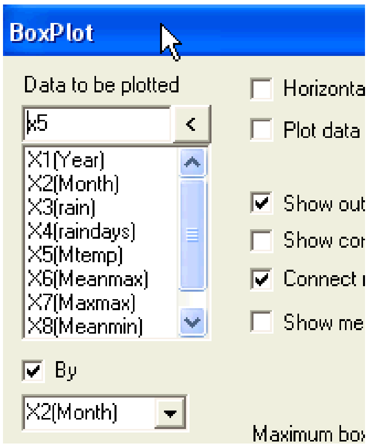
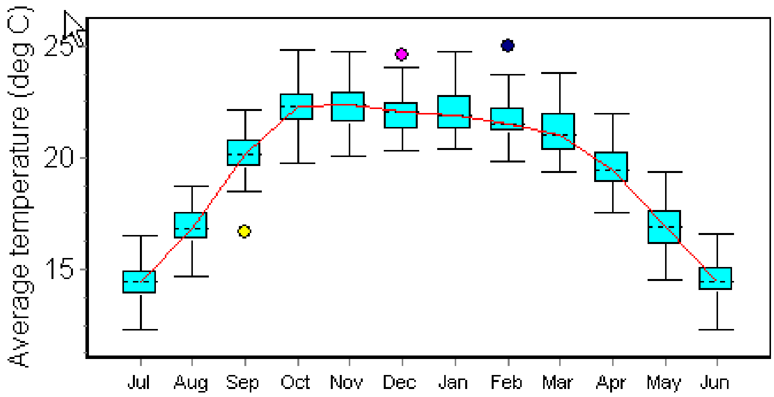


Fig. 2.4b Results for mean temperatures (Mtemp in x5)



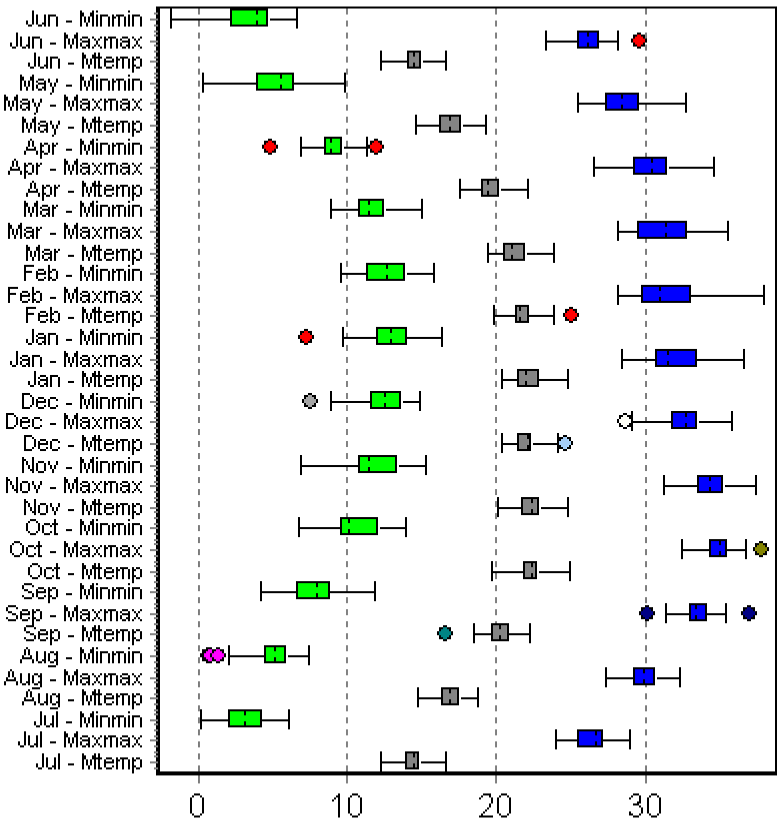
The results are in Fig. 2.4b. They show an oddity in September, where one year has a mean that is very low, compared to the others. The methods in Section 2.2 can be used to show that this oddity is in 1953. We then examined the daily data (Fig. 2.4c) to investigate further. The figure shows that there was a very cold spell from 16th to 18th September. There doesn't seem any reason to doubt the data, so we continue.

Fig. 2.4c Daily data for the 'odd' year

	X4	X5	X6	X7
	Maxtemp	Mintemp	Rain	Date
802	27.3	10.8	0	09-Sep-53
803	30.1	8.9	0	10-Sep-53
804	24.5	11.6	0	11-Sep-53
805	22.7	10.6	0	12-Sep-53
806	26.9	6.9	0	13-Sep-53
807	30.2	8.3	0	14-Sep-53
808	25.8	14.2	0	15-Sep-53
809	15.6	5.6	0	16-Sep-53
810	17.8	4.8	0	17-Sep-53
811	14.8	6.1	0	18-Sep-53
812	20.7	6.9	0	19-Sep-53

Return to the boxplot dialogue, Fig. 2.4a, and add x7 and x9. Uncheck the box to connect the medians, and check to give a horizontal plot. The result is shown in Fig. 2.4d.

Fig. 2.4d Plot of extreme and mean temperatures



Only in June have the minimum temperatures have dropped below zero. The highest maximum temperatures are between September and November. The record maximum was 38 degrees.

One objective was to assess whether there is any evidence of climate change. We therefore look at the means for each year for the mean monthly temperatures, (x5, x6 and x8) and also at the minimums for the minima, (x9) and the maximums for the maxima (x7).

Use **Statistics ⇒ Tables ⇒ General**, Fig. 2.4e. The factor column is X1(Year). First use X5, X6 and X8 as summary columns, and give the mean. Then give the maximum of X7 and the minimum of X9.

Fig. 2.4e Giving five summary statistics

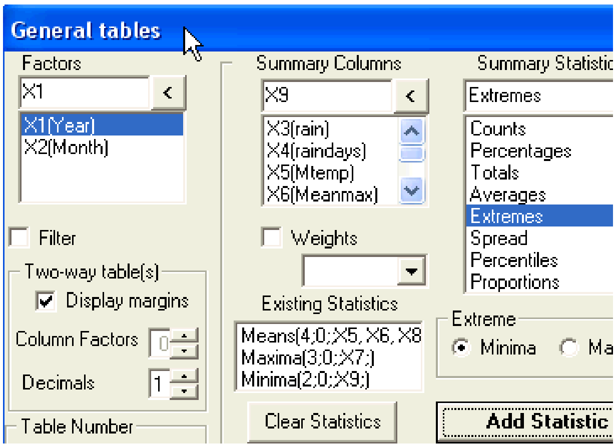


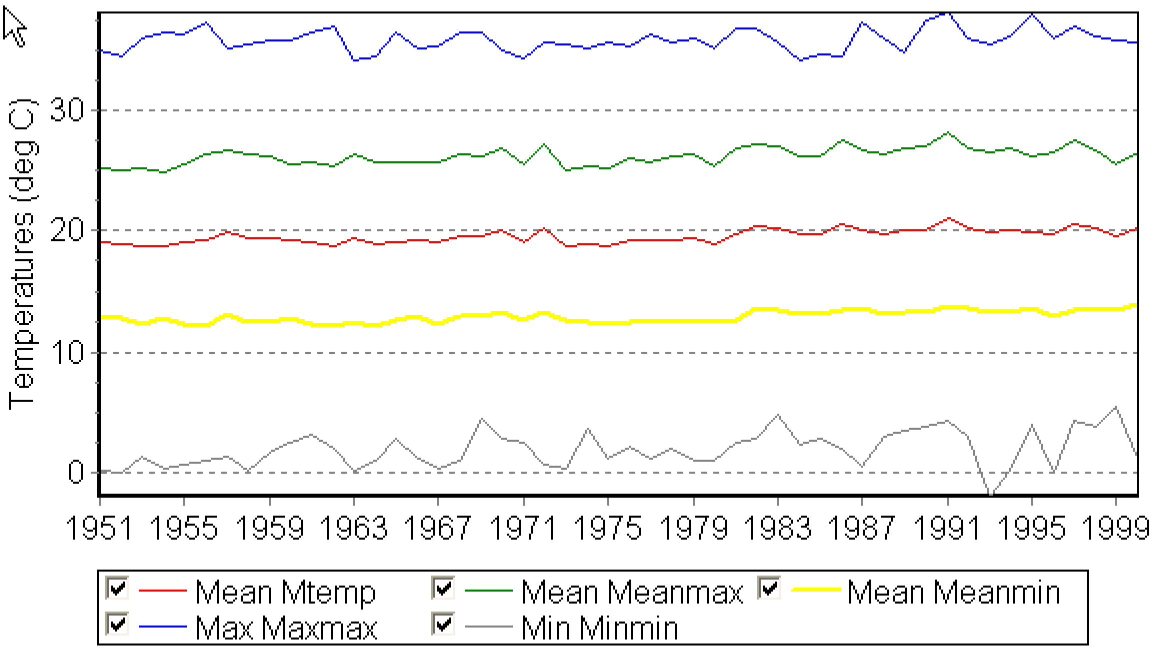
Fig. 2.4f Results for each year

	Summary				
Year	Mean M	Mean M	Mean M	Max Ma	Min Mir
1951	19.0	25.2	12.7	34.9	0.2
1952	18.9	25.0	12.8	34.4	0.1
1953	18.7	25.2	12.3	35.8	1.4
1954	18.8	24.9	12.7	36.3	0.4
1955	19.0	25.5	12.3	36.2	0.7
1956	19.2	26.3	12.1	37.1	1.1
1957	19.9	26.6	13.2	35.0	1.4
1958	19.4	26.2	12.5	35.4	0.3
1959	19.3	26.2	12.4	35.7	1.7

Use **Structure ⇒ Margins** to get rid of the margin, and then use **Plot ⇒ Create Graph from Summaries** to provide a line plot from the table, Fig. 2.4g.

There is a possible indication of climate change, in a slight rise, particularly in the mean and maximum temperatures. In a confirmatory study, this can be assessed using regression analysis.

Fig. 2.4g Time-series plot for temperatures



However, for agriculture it is the pattern of rainfall that is more important. So return to the rainfall data. The start of the season is particularly important, so look at the data for November. This is the 5th month (starting from July).

Generate a column to distinguish November from the other months, see Fig. 2.4h. The result is in Fig. 2.4i.

Fig. 2.4h Logical calculation to select November data

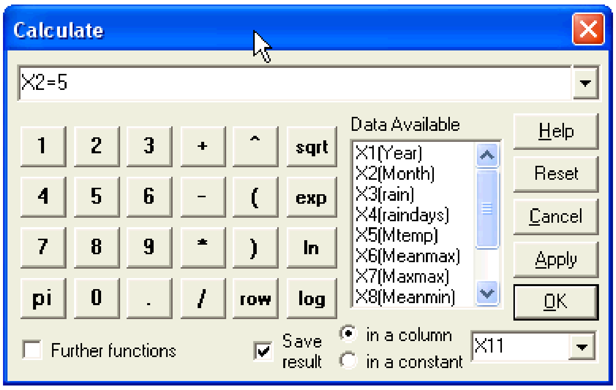


Fig. 2.4i Results

	X1 - F Year	X2 - F Month	X3 rain	X4 raindays	X11
4	1951	Oct	116.2	9	0
5	1951	Nov	128.9	7	-1
6	1951	Dec	150.6	9	0
7	1951	Jan	308	15	0
8	1951	Feb	107.1	8	0
9	1951	Mar	9.2	2	0
10	1951	Apr	4.6	1	0
11	1951	May	4.1	2	0
12	1951	Jun	2.3	1	0
13	1952	Jul	0	0	0
14	1952	Aug	0	0	0
15	1952	Sep	14	3	0
16	1952	Oct	35.1	5	0
17	1952	Nov	240.8	9	-1
18	1952	Dec	160.6	11	0

Now use **Statistics ⇒ Tables ⇒ Summary**, Fig. 2.4j, with X1 (Year) as the factor and both x3 (rain) and x4 (raindays) as the columns to summarise. Use the Totals as the summary, and uncheck the option to give the margin.

⇨ To enable the graph to show the rainfall totals and the raindays together, scale the rain column by 0.1, Fig. 2-4k, effectively giving rainfall in cms rather than mm.

Fig. 2.4j Totals and raindays in November

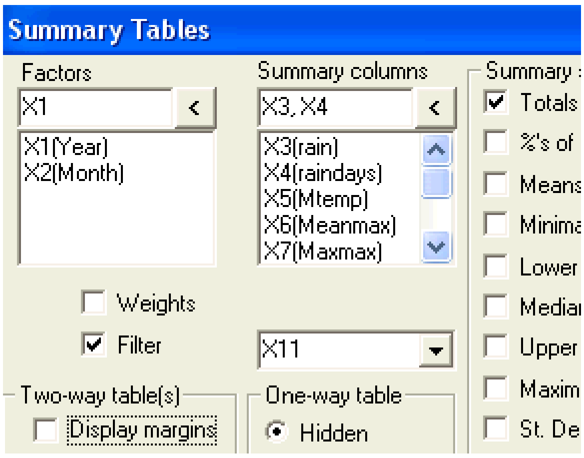
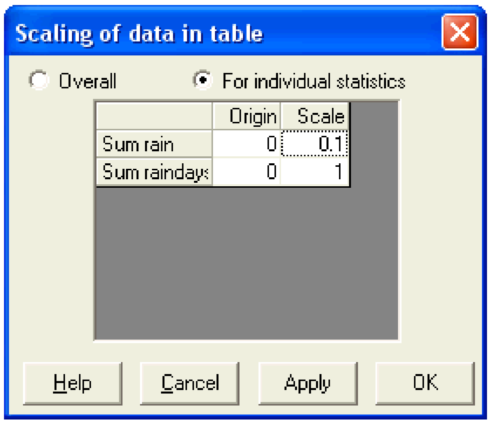
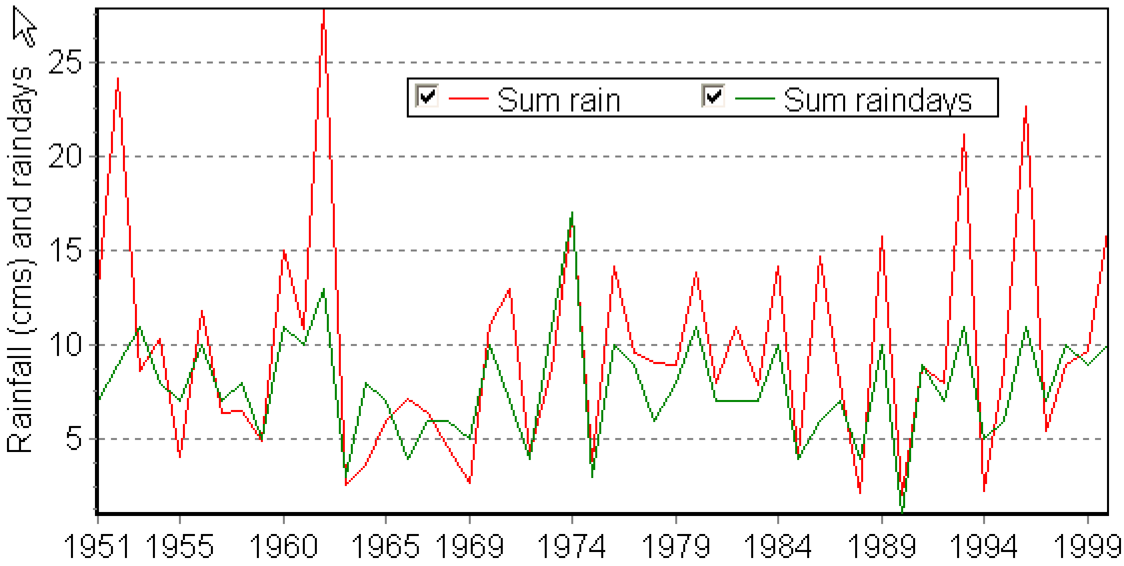


Fig. 2.4k (Tables) Data ⇒ Scale



Then use **Plot ⇒ Summary** to give the graph shown in Fig. 2.4l. We have edited it slightly, to include it in a report or a presentation. It does not indicate any climate change.

Fig. 2.4l Rain days and rainfall totals (cms) in November



The graphs in Fig. 2.4d, 2.4g and 2.4l can be saved, using **File ⇒ Export**, and added to a report or a presentation.

2.5 In conclusion

In this supposed report, one objective is to assess the case for a more comprehensive study, particularly to investigate the possibility of climate change in more detail. This would use the original (daily) data from a number of stations. We believe that this initial analysis makes the case.

In section 2.1 we stressed that the analysis is relatively easy if all components are in place. This was particularly that the data were organised so that the analyses could be performed directly. And there were no major oddities in the data.

More care is needed if the analyses were to be included in a final, rather than a draft report. For example the maximum temperatures are in December and this is ok with a file starting in July. But the minimum annual temperatures are round June/July, and this should be done with a factor column that starts in January, not one starting in July.

On the organisational side – at some stage staff may need training in good use of a word processor and presentation software. Basic training can be done by self-study. In a more advanced course, staff could learn about how to use templates that have been prepared for them. Computing staff should learn how to construct templates, and also how to make them accessible to all the other staff.

Chapter 3 – Preparing Climatic Data

3.1 Introduction

Section 3.2 lists the datasets used in this guide. Readers will also wish to analyse their own data. Sometimes these have first to be entered, an easy task if there is not too much. Ways that data can be entered are described in Sections 3.3 and 3.4.

Users may already have data in computerised form, which they wish to import into Instat, or to another statistics package. If the data are currently in a spreadsheet, they are easily imported into any statistics package. This is also described in Sections 3.3 and 3.4, because a spreadsheet is a convenient way of entering data.

For users who are inexperienced in computing there are some sections of this chapter that could be omitted at this stage. For example:

- Just follow Section 3.2, or this introduction and then omit the chapter for now.
- Initially just use the data sets provided. Using your own data adds a level of complexity in computing that is separate from the ideas of analysis described in the next few chapters.
- Consciously separate the computing, statistics and climatic ideas in this guide. Come back to this chapter when you are ready to analyse some of your own data.
- If you already have your own data files, they may be organised in ways that can be awkward to import for analysis. Sometimes data are in 'ASCII' files, which are files that can be looked at in a simple editor, such as Notepad. Alternatively Instat also has its own editor. Importing data effectively can take time and more expertise than is assumed for most of this book. Some guidelines are in Section 3.5, and readers who do not have this problem can omit this section.

Section 3.6 describes the procedure for exporting data from Instat, so the analysis can, if necessary, use other software in addition.

If you have a large amount of climatic data then it is important to recognise the data management stage and consider how the data can be organised. Allow sufficient time for the data management, which can take longer than the subsequent analysis. Entry into a spreadsheet, or directly into a statistics package is not advisable for large volumes of data and the Clicom and ClimSoft projects are described briefly in Section 3.7. Clicom is an initiative that has been coordinated by WMO (World Meteorological Organisation) for the last 20 years. This project is particularly to provide a database system for Meteorological Services in developing countries to enter their climatic data. ClimSoft is intended as a replacement for Clicom.

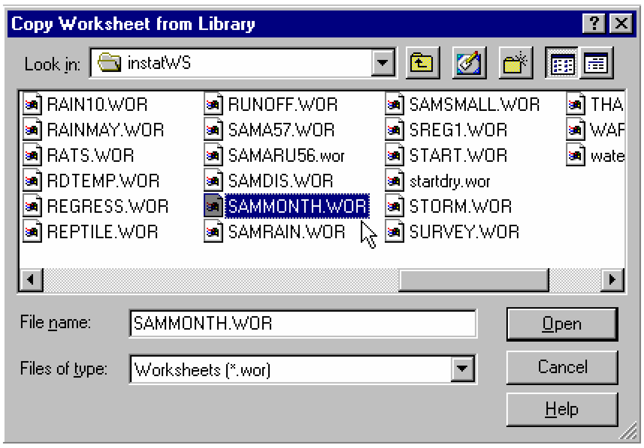
Data from the current Clicom can easily be imported into Instat.

3.2 Using existing data sets

When Instat is installed, sample datasets are also automatically installed in a 'read only' folder. These are used to illustrate all the analyses described in this guide, such as those in Chapter 2. Use **File ⇒ Open From Library**, as shown in Fig. 3.2a to open any of these worksheets. When you make changes, they will be saved in your working folder. Thus the original version can always be reopened from the library if needed.

Fig. 3.2a Choosing a worksheet from the Library

File ⇒ Open From Library
Select a file and click **Open**

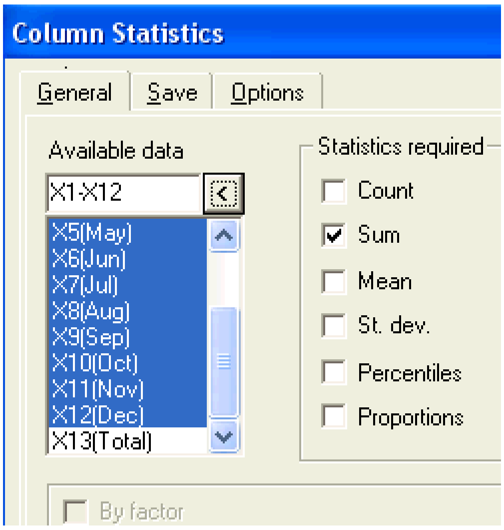


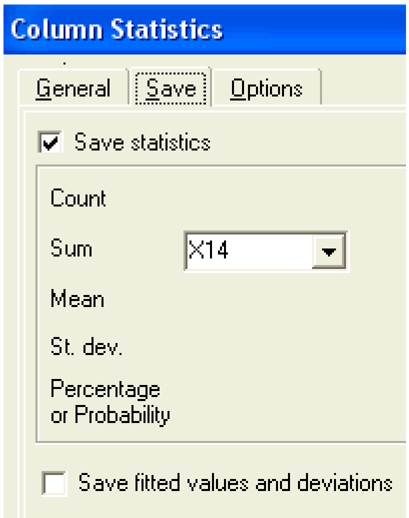
As an example, look at the file called **sammonth.wor**. This contains monthly rainfall totals for the 56 years from 1928 to 1983 for Samaru, a station in the North of Nigeria.

The aim here is primarily to show how Windows statistics packages are used, rather than to analyse data. We need some results for illustration, so use **Statistics ⇒ Summary ⇒ Column Statistics** and complete the dialogue as shown in Fig. 3.2b. Click on the Save tab (Fig. 3.2c) so the means are also saved to the worksheet. When you press **OK**, the output window, will look as shown in Fig. 3.2e. The monthly means are also saved in the worksheet, see X14 in Fig. 3.2e.

Fig. 3.2b Calculate monthly means **Fig. 3.2c Save tab**

Statistics ⇒ Summary ⇒ Column Statistics





If you maximise Instat, so it uses the whole screen, and then use **Window ⇒ Tile Vertical** as shown in Fig. 3.2d, your screen should look roughly as shown in Fig. 3.2e. There are other options that you can use from the menu in Fig. 3.2d. You can choose to tile the screen horizontally if you prefer. Because these options are used frequently both have a 'shortcut' of **<F2>** or **<Shift+F2>**.

Fig. 3.2d Windows menu

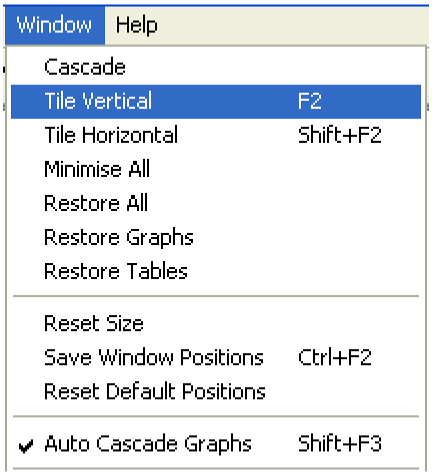
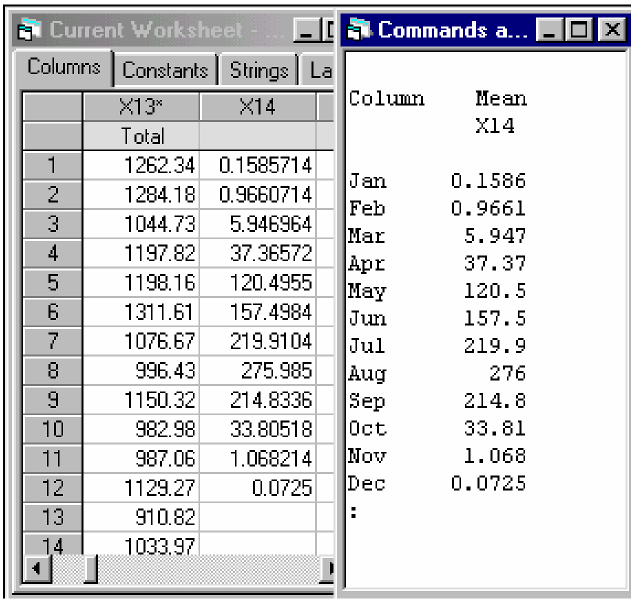


Fig. 3.2e Tiled Worksheet and Commands and Output windows



So, statistics packages have two different types of Window. They have one that looks like a spreadsheet for the data, and a second that looks more like an editor or word processor for the results.

Saving the data or results uses the **`File ⇒ Save As`** menu (Fig. 3.2f). Click somewhere in the worksheet (so it is the current Window) to save the data file, including the new contents of x13 and x14, into your working folder. You will usually keep the same name as it had in the library.

Saving the results is usually done in one of two ways. If you move to the '**Commands and Output**' Window and click anywhere in it, you can use **`File ⇒ Save As`**, shown in Fig. 3.2f to save all the output at any stage. Then use **`Edit ⇒ Select All`** (Fig. 3.2g) and press the <**Delete**> key or use **`Edit ⇒ Cut`** to clear the output window for the next stage in the analysis.

Fig. 3.2f Save the results

`File ⇒ Save As`

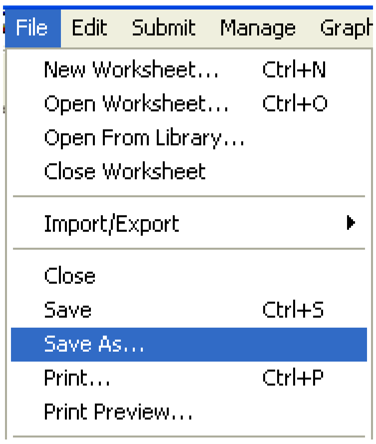
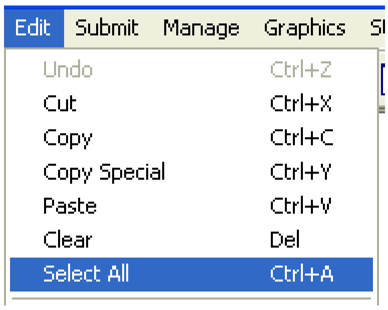


Fig. 3.2g Delete output

`Edit ⇒ Select All`



An alternative is to open a word processor at the same time and copy results over, as you produce them. To do this, mark the output to copy and then use **`Edit ⇒ Copy`** (or <**Ctrl + C**>). Then move to the Word processor and paste the results.

Statistics packages have a few more types of Window, for example for graphs, as you have already seen in Chapter 2. Copying graphs into a report is described in Chapters 2.3 and 4.2.

This feature of multiple types of window is different from spreadsheet packages.

In Fig. 3.2e the worksheet window shows some of the data in the current Instat worksheet. If you use **Manage ⇒ Worksheet Information**, as shown in Fig. 3.2h the resulting dialogue, Fig. 3.2i gives details of the size of the worksheet and how many columns, etc, already have data.

Fig. 3.2h Manage menu

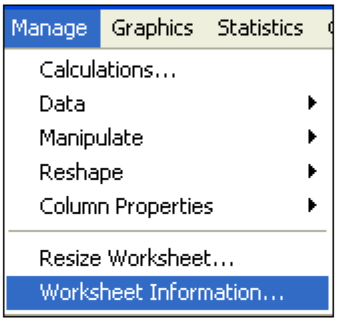
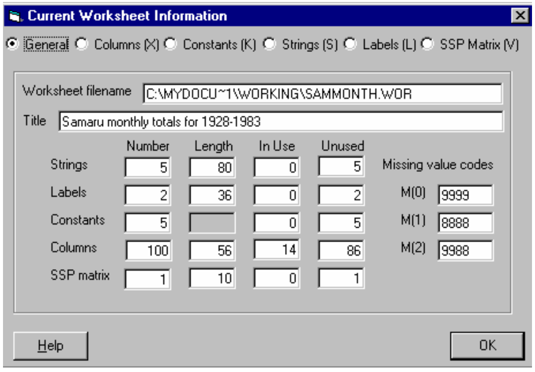


Fig. 3.2i Worksheet information dialogue

Manage ⇒ Worksheet Information



From Fig. 3.2i selecting the 'Columns' option gives more details about the number of items of data in each column, and so on (see Fig. 3.2j).

Fig. 3.2j Information about a worksheet

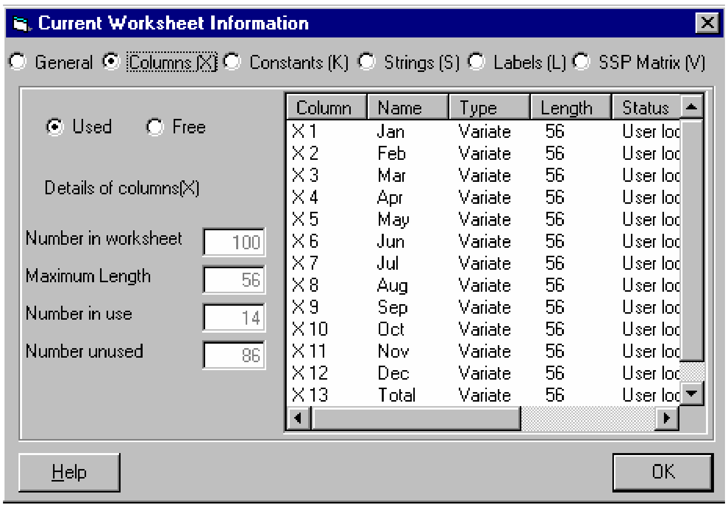
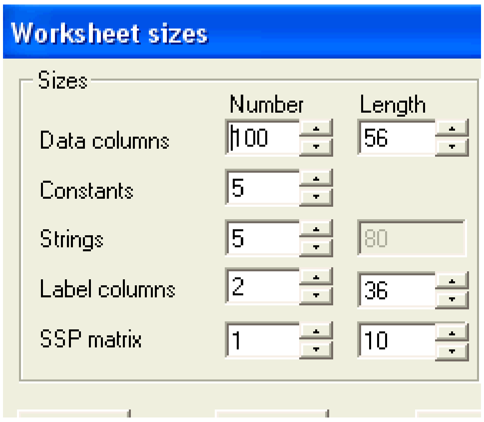


Fig. 3.2k Change worksheet size

Manage ⇒ Resize Worksheet



Unlike most statistics packages, worksheets in Instat are a fixed size determined, when first creating the file, using **File ⇒ New Worksheet**[6]. When the monthly means were calculated above, see Fig. 3.2b, we chose to save them as a further column in the worksheet. This is so they could themselves be analysed when needed. Sometimes the worksheet is not large enough for your work. Then use the **Manage ⇒ Resize Worksheet** dialogue as shown in Fig. 3.2k. Here you can change any of the dimensions.

However, there are limits. In the current Instat, a single worksheet cannot have more than 127 columns of data, i.e. variables. This is a little like a spreadsheet, where the maximum width is usually 255 columns. If you need more columns in a single worksheet then you have to move to a more powerful statistics package, where most can handle an effectively unlimited number of columns.

Currently Instat columns also have a limited length, i.e. number of cases. Many datasets in this guide have a column length of 366 and that is fine. Instat's current limit is about 32,000 rows.

[6] The **Edit ⇒ Options** dialogue may be used to change this initial size.

This compares with the limit of about 64,000 rows in a typical spreadsheet package, while most other statistics packages have no fixed limit.

In constructing Instat, we have tried to make it as consistent with other Windows software where possible. There is one difference that is illustrated by opening a new file from the library, called samsmall.wor.

Opening this new worksheet has closed the previous worksheet automatically. You can only have one worksheet open at a time. Also, you will be asked whether you want to save the changed worksheet, which now has the means added.

Instat copied the worksheet that was opened into a temporary directory. There it saves the current state of the worksheet, whenever there is a change. So the annual totals are safe. When you close that worksheet, or open a new one, Instat will ask if you want to save the changes. Alternatively you could use **File ⇒ Save** or **File ⇒ Save As** at any time, when it will suggest that the worksheet be saved into your current working directory. Then it will delete the copy from the temporary directory, once you move to a new one.

This automatic saving of the current worksheet in the temporary directory means that you should not lose data if there is ever a power cut, or even a problem in Windows or Instat. In that case it will still have a copy in the temporary directory, and will ask you if you wish to continue, with that worksheet, when you next start Instat.

Fig. 3.2l Worksheet Editor

Edit ⇒ Run Worksheet Editor

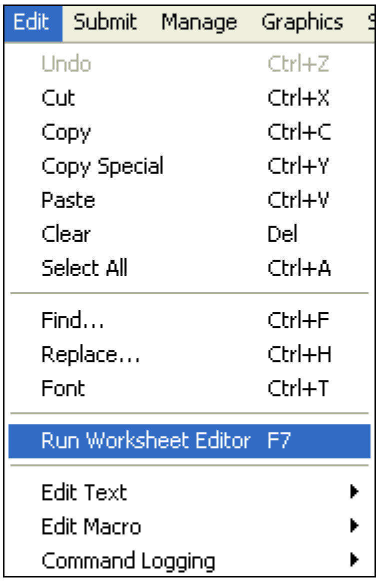
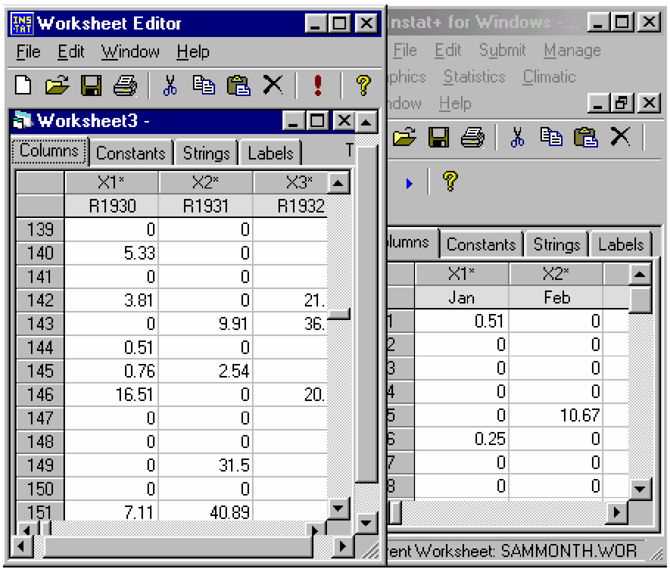


Fig. 3.2m Current worksheet, sammonth and samsmall in worksheet editor



If Instat can only have one worksheet open at a time, you may wonder how you can transfer data between different worksheets. There are two main ways. Later in this guide the **Manage ⇒ Duplicate(Copy Columns)** dialogue is used, but here we introduce the **Edit ⇒ Run Worksheet Editor** as shown in Fig. 3.2l. This is a separate editor to enter and edit data. This special worksheet editor can have more than one worksheet open and copy between them, or from one of these worksheets to the main worksheet that is being used for analysis. This worksheet editor has its own menus and just saves to the disc when you choose its **File ⇒ Save**. Fig. 3.2m shows the current worksheet **Sammonth.wor** and the worksheet **Samsmall.wor** which has been opened into the Worksheet Editor.

Finally in this section, in Fig. 3.2n lists the climatic worksheets that are used in this book. There are a total of 19 files from 10 stations.

Fig. 3.2n Climatic Worksheets

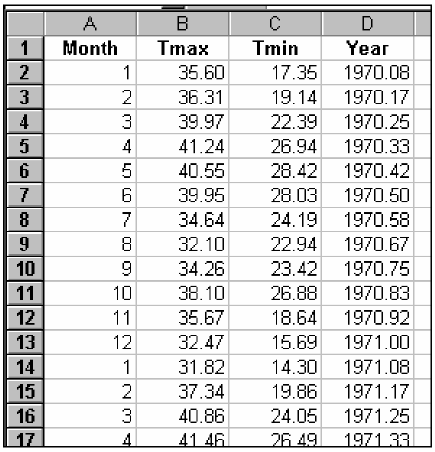
Name	Description	Chapters
Bohicon.wor	Bohicon, Benin: 1983 daily climatic data	8
Bulawayo.wor	Bulawayo, Zimbabwe: 1951-2000 daily data for rainfall and temperatures	2
Bulmon.wor	Bulawayo: monthly summaries or rainfall and temperature data	2
Kayes.wor	Monthly data from Kayes in Mali	9
Kurun7.wor	Weekly rainfall totals from Kurunegala, Sri Lanka, for 1950 to 1983	12
Kurunega.wor	Daily rainfall data from Kurunegala, Sri Lanka, for 1950 to 1983	11, 13, 15
Madaw.wor	Daily rainfall data from Madawachchiya, in the North of Sri Lanka, from 1891 to 1978	12
Malawicr.wor	10 day data for 27 years from a site in Malawi	10
Niatemp.wor	Monthly maximum and minimum temperatures from Niamey, Niger for a 7 year period	2, 3
Ntemp.wor	Daily maximum and minimum temperatures for Niamey, Niger from 1961 to 1980	8
Rain.wor	Monthly rainfall totals for the years 1950-1983 from Galle, in Sri Lanka.	
Rain10.wor	Monthly rainfall totals for Galle in Sri Lanka from 1961-70	
Rdtemp.wor	Reading, UK temperature - daily tmax and tmin are in pairs of columns for 10 years, 1976-1985.	8
Runoff.wor	Data on rainfall and runoff from Gregory (1978)	11
Samaru56.wor	56 years of daily rainfall data from Samaru, Nigeria, 1928 to 1983.	
Samdis.wor	10 day totals for 56 years from Samaru, 1928-83	
Sammonth.wor	Monthly totals for 56 years of daily rainfall data from Samaru, Nigeria, 1928 to 1983.	
Samrain.wor	56 years, 1928-83 from Samaru, Nigeria, giving events such as 'start of the rains'.	7, 11
Samsmall.wor	11 years of daily rainfall data from Samaru, Nigeria, 1930 to 1940.	5-7, 11
Storm.wor	5 minute records from 50 storms in Malawi from hydrological data .	12
Wafric2.wor	Monthly rainfall totals for 28 years from 2 neighbouring stations in Benin (West Africa)	11

3.3 Entering data

In the Instat tutorial data were entered into a new worksheet. Use **File ⇒ Open Worksheet** or **File ⇒ Open From Library** to add or to edit the data in an existing worksheet.

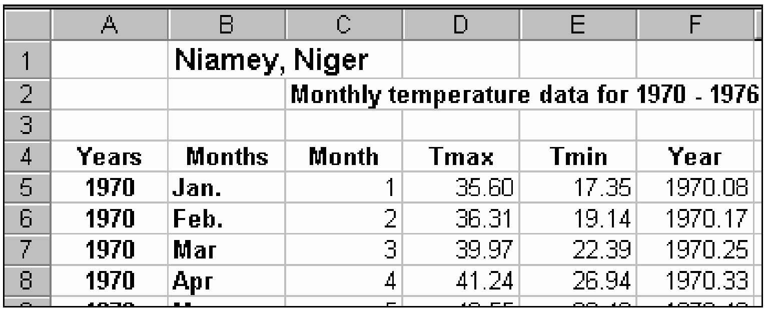
An alternative is to type the data into a spreadsheet, such as Excel and then transfer to the statistics package for some of the analysis. An example, for the Niamey temperature data, is in Fig. 3.3a. Almost every statistics package can read Excel files, so it is easy to use Excel for part of your work, and transfer to the statistics package when you need to.

Fig. 3.3a Niamey temperature data in Excel worksheet



	A	B	C	D
1	Month	Tmax	Tmin	Year
2	1	35.60	17.35	1970.08
3	2	36.31	19.14	1970.17
4	3	39.97	22.39	1970.25
5	4	41.24	26.94	1970.33
6	5	40.55	28.42	1970.42
7	6	39.95	28.03	1970.50
8	7	34.64	24.19	1970.58
9	8	32.10	22.94	1970.67
10	9	34.26	23.42	1970.75
11	10	38.10	26.88	1970.83
12	11	35.67	18.64	1970.92
13	12	32.47	15.69	1971.00
14	1	31.82	14.30	1971.08
15	2	37.34	19.86	1971.17
16	3	40.86	24.05	1971.25
17	4	41.46	26.49	1971.33

Fig. 3.3b Niamey monthly data with meta-data



	A	B	C	D	E	F
1		Niamey, Niger				
2			Monthly temperature data for 1970 - 1976			
3						
4	Years	Months	Month	Tmax	Tmin	Year
5	1970	Jan.	1	35.60	17.35	1970.08
6	1970	Feb.	2	36.31	19.14	1970.17
7	1970	Mar	3	39.97	22.39	1970.25
8	1970	Apr	4	41.24	26.94	1970.33

If you choose to start in a spreadsheet, type the data into columnsthat are headed by a name, as shown in Fig. 3.3a. Excel calls this layout a list, and has a useful description of lists in its HELP. This list format transfers easily to a statistics package.

When using a spreadsheet, we recommend you have special data sheets with just the data, and do not mix them with the results. If you wish to see the data together with the results, then paste the data to a second sheet. Then calculate summary values on this second sheet.

Sometimes you will also wish to include meta-data, i.e. a description of the data. These details may be on a separate sheet. If they are on the same sheet, such as shown in Fig. 3.3b, or if you have other information on the sheet, then the data may not be at the top left of the sheet.

Fig. 3.3c Excel help on naming

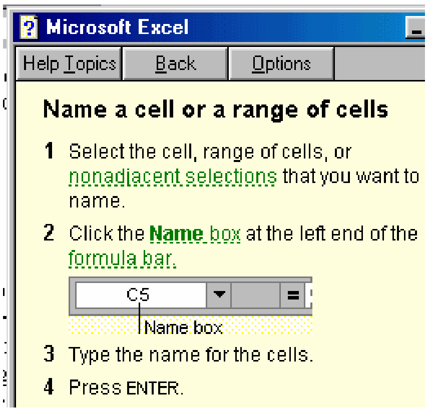
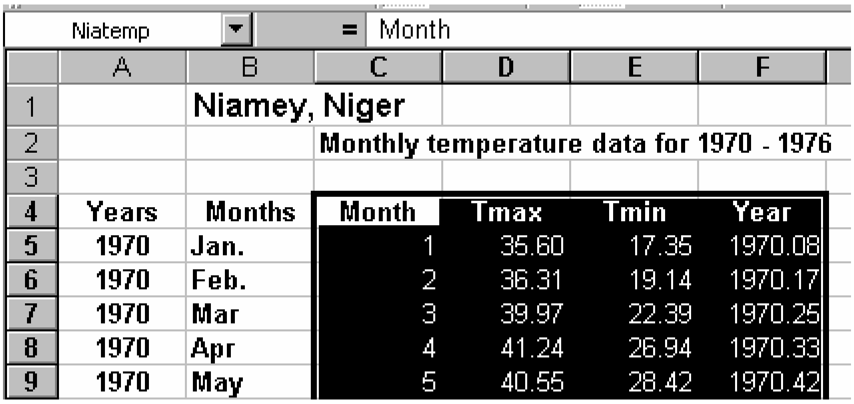


Fig. 3.3d Name a range of cells



	A	B	C	D	E	F
1		Niamey, Niger				
2			Monthly temperature data for 1970 - 1976			
3						
4	Years	Months	Month	Tmax	Tmin	Year
5	1970	Jan.	1	35.60	17.35	1970.08
6	1970	Feb.	2	36.31	19.14	1970.17
7	1970	Mar	3	39.97	22.39	1970.25
8	1970	Apr	4	41.24	26.94	1970.33
9	1970	May	5	40.55	28.42	1970.42

In this case, it is useful, in Excel, to give a name to the rectangle that contains the data to be imported to Instat, or to another statistics package. The Excel Help explains how to do this as shown in Fig. 3.3c.

There are two ways to copy the information to Instat, or to any other statistics package.

The first is simply to copy to the clipboard from Excel. Then in Instat, you must already have an open worksheet and you simply Paste into the required columns.

For the second method start by saving the Excel file in your current working directory. Then in Instat you do not need an open worksheet. Instead, use **File ⇒ Open Worksheet** and select the file type Excel2-5/95/97(*.xls) instead of the default, Worksheets(*.wor). Then select the required file from the list of Excel worksheets as shown in Fig. 3.3e.

Fig. 3.3e Open/Import dialogue

File ⇒ Open Worksheet

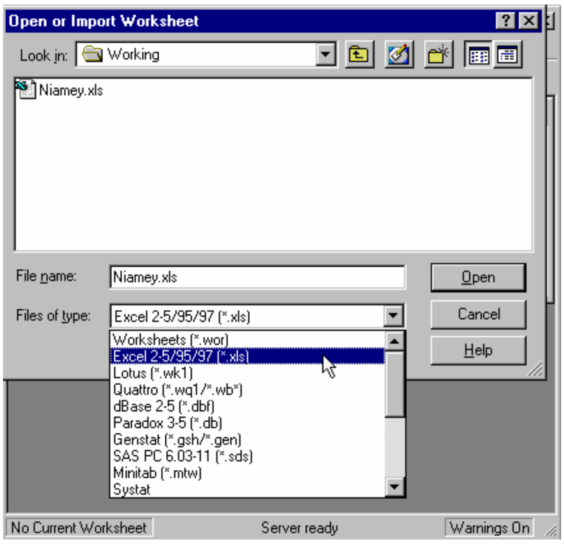
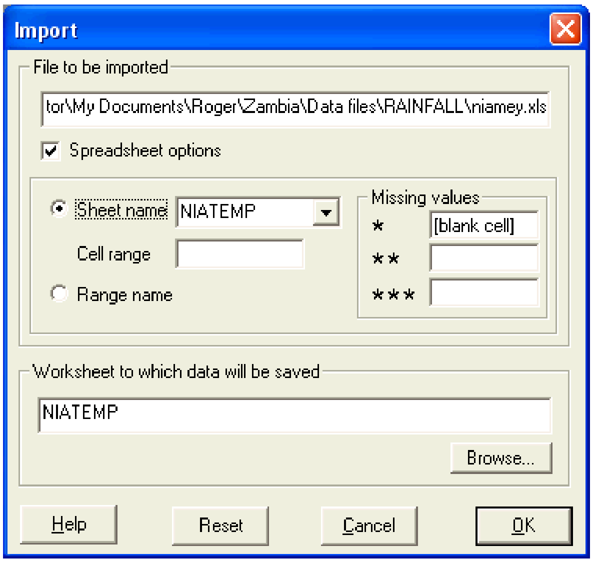


Fig. 3.3f Import from Excel workbook



You can import from many different formats and Instat recognises the format from the file extension. With a spreadsheet, unless you instruct otherwise, Instat will import from the start of the first sheet, but you can choose to import from a particular sheet, and even a specified rectangle, if you have given the name, as shown in Fig. 3.3d. Fig. 3.3f ***shows the importing*** from the workbook called *Niamey.xls*, worksheet (*Niatemp*) into an Instat worksheet called **NiameyXL**.

Fig. 3.3g Alternative way of entering data into Excel

	A	B	C	D	E	F	G	H	I	J	K	L
1	**Months**	**Tmax70**	**Tmax71**	**Tmax72**	**Tmax73**	**Tmax74**	**Tmax75**	**Tmax76**	**Tmin70**	**Tmin71**	**Tmin72**	**Tmin73**
2	**Jan.**	35.60	31.82	33.17	33.16	30.34	29.87	31.47	17.35	14.30	16.12	18.1
3	**Feb.**	36.31	37.34	36.43	36.49	34.11	34.86	35.80	19.14	19.86	19.14	20.1
4	**Mar**	39.97	40.86	39.57	38.61	38.79	38.39	37.15	22.39	24.05	23.18	23.4
5	**Apr**	41.24	41.46	40.06	41.86	41.44	41.21	39.23	26.94	26.49	26.65	27.4
6	**May**	40.55	41.24	39.23	41.66	39.95	37.95	38.30	28.42	28.02	27.65	29.0
7	**Jun**	39.95	39.06	36.97	39.61	38.40	37.36	35.52	28.03	26.12	25.32	27.2
8	**Jul**	34.64	34.19	35.82	35.39	33.41	32.31	33.41	24.19	23.72	24.91	24.5
9	**Aug**	32.10	31.40	34.65	34.16	32.04	31.79	32.45	22.94	22.37	24.41	23.1
10	**Sep**	34.26	34.62	36.75	35.66	33.89	32.92	34.43	23.42	23.64	24.65	23.6
11	**Oct**	38.10	37.43	38.17	38.96	37.51	37.12	35.37	26.88	23.15	25.26	24.5
12	**Nov**	35.67	36.47	34.91	35.34	35.37	35.97	36.29	18.64	19.40	20.20	20.0
13	**Dec**	32.47	33.15	33.23	33.54	31.18	33.26	34.14	15.69	18.25	18.47	17.0
14	**Mean**	**36.74**	**36.59**	**36.58**	**37.04**	**35.54**	**35.25**	**35.30**	**22.84**	**22.45**	**23.00**	**23.2**
15												

The idea of Excel's 'Lists' is a general concept that applies to all types of data. But what is to be done if your data were typed into Excel in a different layout, before reading this section of the guide?! Fig. 3.3a showed monthly temperature data from Niamey in Niger, in list format, but perhaps it was entered as shown in Fig. 3.3g instead.

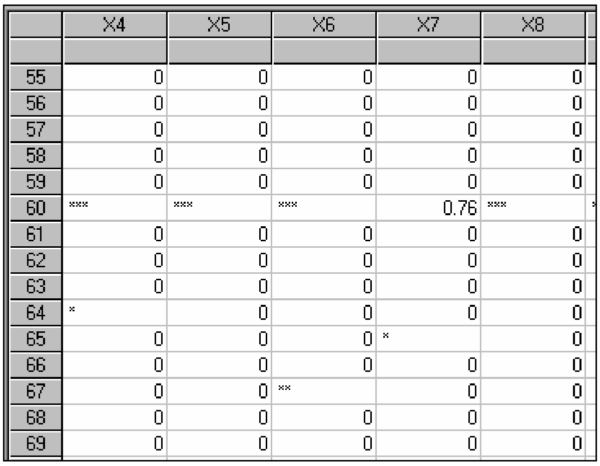
This is a common problem and methods of importing data that have been entered in a variety of ways is therefore shown in Section 3.5.

3.4 Daily data

Daily data is used in much of this guide. Figs 3.4a and 3.4b shows an Instat worksheet for 11 years of daily data. With Instat, daily data are usually stored, with one column per year in worksheets that have 366 rows. Fig. 3.4b shows that Instat has 3 different 'missing value' codes. The first is the number that will be interpreted by Instat as an ordinary missing value. The second is mainly for rainfall data where it is used to signify a 'trace' value, i.e. a very small rainfall, where the actual amount was not recorded.

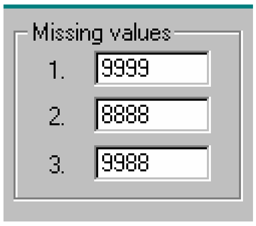
The third is a special code for 29 February in non-leap years. In Fig. 3.4a the data have been altered slightly, to show all 3 codes. They are displayed as *, ** and ***.

Fig. 3.4a Data in Instat showing missing values



	X4	X5	X6	X7	X8
55	0	0	0	0	0
56	0	0	0	0	0
57	0	0	0	0	0
58	0	0	0	0	0
59	0	0	0	0	0
60	***	***	***	0.76	***
61	0	0	0	0	0
62	0	0	0	0	0
63	0	0	0	0	0
64	*	0	0	0	0
65	0	0	0	*	0
66	0	0	0	0	0
67	0	0	**	0	0
68	0	0	0	0	0
69	0	0	0	0	0

Fig. 3.4b Missing values



Daily data can be entered directly into Instat. The first step is to create the empty worksheet, ready for the data. There is a separate dialogue that creates new worksheets with the appropriate size and missing value codes. This is under **Climatic ⇒ Manage ⇒ New Worksheet**, as shown in Fig. 3.4c and 3-4d. This is simply for convenience and only differs from the **File ⇒ New Worksheet**, in the default dimensions and missing value codes that are used.

Fig. 3.4c New Worksheet dialogue

Climatic ⇒ Manage ⇒ New

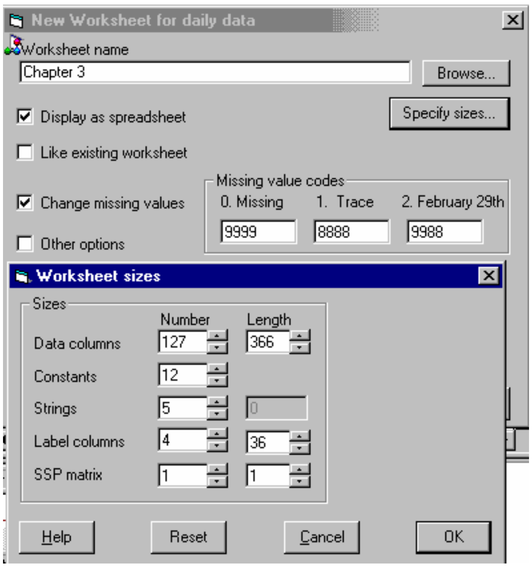
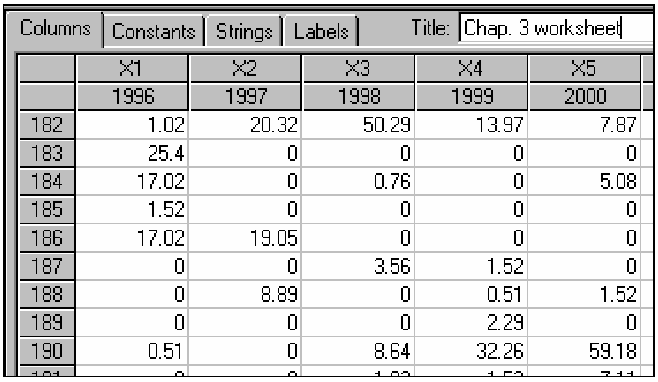


Fig. 3.4d New climatic worksheet



	X1	X2	X3	X4	X5
	1996	1997	1998	1999	2000
182	1.02	20.32	50.29	13.97	7.87
183	25.4	0	0	0	0
184	17.02	0	0.76	0	5.08
185	1.52	0	0	0	0
186	17.02	19.05	0	0	0
187	0	0	3.56	1.52	0
188	0	8.89	0	0.51	1.52
189	0	0	0	2.29	0
190	0.51	0	8.64	32.26	59.18

Then the data are typed into the worksheet, with the special codes typed either as * or as the corresponding numbers for missing, etc.

However, many users will be more comfortable typing larger volumes of data into a spreadsheet. Then missing values can be left blank, but trace and 29 February should normally be typed using the numbers that will correspond to the codes once they are in Instat. Otherwise you can change them once imported into Instat.

Fig. 3.4e shows an example of an Excel sheet containing 10 years of maximum and minimum temperatures from Reading, England.

Fig. 3.4e Temperature data in Excel from Reading (England)

	A	B	C	D	E	F	G	H	I	J
1	Mn76	Mx76	Mn77	Mx77	Mn78	Mx78	Mn79	Mx79	Mn80	Mx80
50	1.5	3.4	5.5	8.9	-4.3	1.4	-2.2	0.4	3.5	10.8
51	1	5.8	2	9.7	-3	3	-1.2	1.5	3.9	11
52	2	9.5	3.8	12.3	-3.1	2.7	0.1	2.2	1.4	10.7
53	4.1	12	4.6	9.5	-1	6.5	0	2.8	2	9.7
54	5.5	11.5	5.2	9.5	-0.5	8.4	0.6	7.5	7.6	9.1
55	7	10.5	6	10.9	6.1	13.3	1	8.5	0.8	8.4
56	8.5	13.1	5.1		6.6	12.2	-2	5.8	-0.2	6
57	6.1	12.5	4.2	7.4	7.6	10.7	-5.5	7.2	1	8.3
58	5.7	11.5	0	5.7	7.1	10.7	-4.1	3.7	-2	5.7
59	7.3	9.1	-2.9	5.9	6.9	10.9	-1.7	8.8	-0.7	6
60	5.1	10.3	9988	9988	9988	9988	9988	9988	2.4	8.5
61	2.7	14.5	0.9	9.7	6.5	10.1	-2.1	9.5	2.6	9.6
62	-1.4	10.1	6	18.4	3.8	8.5	1.6	11.5	4.6	9
63	-3.4	10	8.5	14.3	2.9		9.1	11.2	5.2	9.6
64	1.9	11.1	9.6	11.5	0.4	9.7	3.6	9.9	-0.3	6.7
65	0	12.5	4.5	11.8	-1	8.6	-0.3	10.4	-4.2	9
66	-1	10.6	6.5	11.5	-3	9.8	3.9	7.7	-2.3	7.2
67	-1.5	5.4	3.1	12.3	-1	11.7	0.8	9	3.3	10.3

To transfer these data to Instat, you have the same 2 options as described in Section 3.3. The first method is to create the worksheet, using **Climatic ⇒ Manage ⇒ New Worksheet** and then copy from Excel and paste into Instat.

The second method does not need a worksheet first. This is to use **File ⇒ Open Worksheet** and import the data directly, as described in Section 3.3. In this case make sure that the trace and missing value codes are specified correctly, see Fig. 3.4e.

Fig. 3.4f Data in list format in Excel

zimdata.xls

	A	B	C	D	E	F	G
1	Year	Mon	Day	Maxte	Mint	Rain	Date
2	1951	Jul	1	23.7	9.4	0	01-Jul-51
3	1951	Jul	2	23.9	9.7	0	02-Jul-51
4	1951	Jul	3	24.4	9.6	0	03-Jul-51
5	1951	Jul	4	23.2	8.5	0	04-Jul-51
6	1951	Jul	5	21.8	7.8	0	05-Jul-51
7	1951	Jul	6	19.9	8.4	0	06-Jul-51
8	1951	Jul	7	16.8	9.4	0	07-Jul-51
9	1951	Jul	8	18.2	4.3	0	08-Jul-51
10	1951	Jul	9	15.8	6.4	1.5	09-Jul-51
11	1951	Jul	10	16.1	7.2	0.3	10-Jul-51
12	1951	Jul	11	19	4.4	0	11-Jul-51
13	1951	Jul	12	18.3	4.3	0	12-Jul-51
14	1951	Jul	13	18.2	5.4	0	13-Jul-51
15	1951	Jul	14	21.4	4.1	0	14-Jul-51

Fig. 3.4g Data imported into Instat

	X1 - F	X2 - I	X3 - F	X4	X5	X6	X7
	Year	Mont	Day	Maxtemp	Mintemp	Rain	Date
1	1951	Jul	1	23.7	9.4	0	01-Jul-51
2	1951	Jul	2	23.9	9.7	0	02-Jul-51
3	1951	Jul	3	24.4	9.6	0	03-Jul-51
4	1951	Jul	4	23.2	8.5	0	04-Jul-51
5	1951	Jul	5	21.8	7.8	0	05-Jul-51
6	1951	Jul	6	19.9	8.4	0	06-Jul-51
7	1951	Jul	7	16.8	9.4	0	07-Jul-51
8	1951	Jul	8	18.2	4.3	0	08-Jul-51
9	1951	Jul	9	15.8	6.4	1.5	09-Jul-51
10	1951	Jul	10	16.1	7.2	0.3	10-Jul-51
11	1951	Jul	11	19	4.4	0	11-Jul-51
12	1951	Jul	12	18.3	4.3	0	12-Jul-51
13	1951	Jul	13	18.2	5.4	0	13-Jul-51
14	1951	Jul	14	21.4	4.1	0	14-Jul-51

Some readers may have noticed an anomaly in this section, compared to Section 3.3. The way daily data are held in Instat is not in the list format that was strongly encouraged in Section 3.3! Suppose there are 50 years of daily data and 10 measurements. Then list format would imply that the data should be arranged in 10 columns of length 18,300, i.e. 366 by 50.

This layout is possible with the current version of Instat, as long as the limit of 32,000 rows is not exceeded. For example Fig. 3.4f shows part of an Excel workbook with a sheet that contains 50 years of data in list form. There are 7 columns, with the first 3 containing the year, month and the day in the month. Then there are the three columns for the rainfall and the maximum and minimum temperatures. The final column is the date.

If you have your data already in the format shown in Fig. 3.4f, then you can either import the data into Instat as it stands, or reorganise the data before importing. Fig. 3.4g shows the data after importing into Instat, with the first 3 columns made into factors, and the final column in date format, as in Excel.

In Version 3 of Instat, most of the special climatic commands do not yet work with this layout of the data. The data therefore need to be 'unstacked' , to put them in the form used earlier in this section. This uses the **Climatic ⇒ Manage ⇒ Unstack Daily**, see the menu in Fig. 3.4h.

Fig. 3.4h Unstack daily data menu

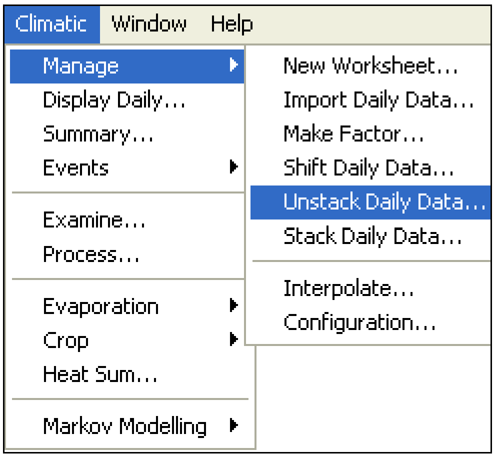
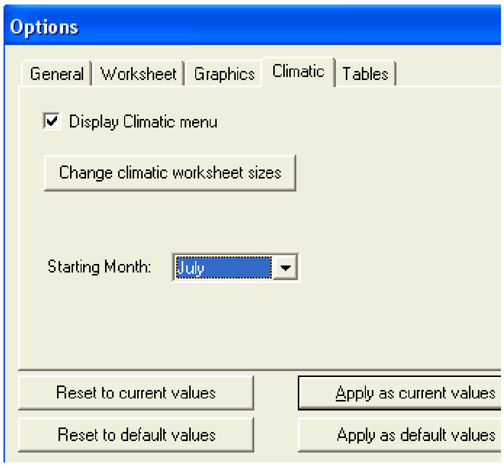


Fig. 3.4i Checking on the starting month



There is one further issue, and that is when the year is to start. In Zimbabwe the growing season is from November to April, and the data start in July. This is when we would like our 'year' to be.

Before using the Unstack dialogue, check **Climatic ⇒ Manage ⇒ Configure**, Fig. 3.4i, to see how the default month of the year is set. If it is not July, then set it to July.

Fig. 3.4j Unstacking the data

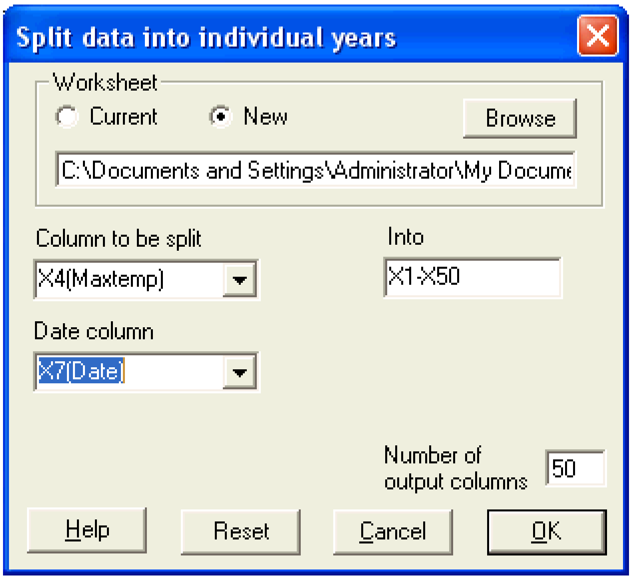
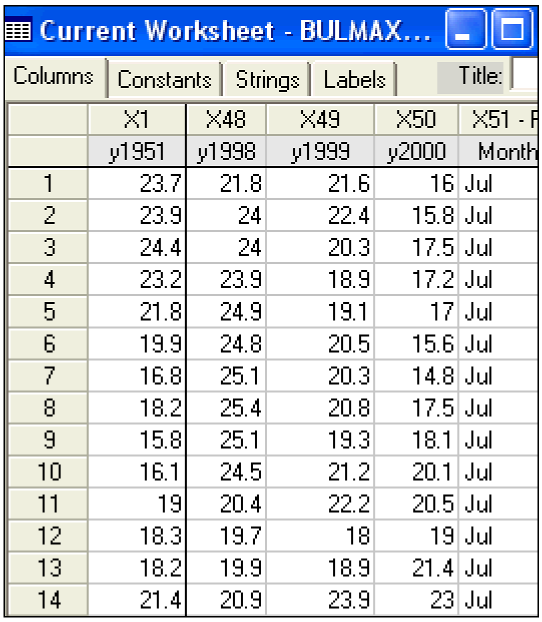


Fig. 3.4k Result, including Month factor



	X1	X48	X49	X50	X51 - F
	y1951	y1998	y1999	y2000	Month
1	23.7	21.8	21.6	16	Jul
2	23.9	24	22.4	15.8	Jul
3	24.4	24	20.3	17.5	Jul
4	23.2	23.9	18.9	17.2	Jul
5	21.8	24.9	19.1	17	Jul
6	19.9	24.8	20.5	15.6	Jul
7	16.8	25.1	20.3	14.8	Jul
8	18.2	25.4	20.8	17.5	Jul
9	15.8	25.1	19.3	18.1	Jul
10	16.1	24.5	21.2	20.1	Jul
11	19	20.4	22.2	20.5	Jul
12	18.3	19.7	18	19	Jul
13	18.2	19.9	18.9	21.4	Jul
14	21.4	20.9	23.9	23	Jul

Now use **Climatic ⇒ Manage ⇒ Unstack Daily Data,** Fig. 3.4j. The results are shown in Fig. 3.4k, after naming the columns and adding a month factor.

3.5 Importing data

Importing data from a spreadsheet, particularly from Excel, has already been described in Sections 3.3 and 3.4. If the data are organised correctly for analysis in Instat, then the importing is a simple process. Otherwise some manipulation of the data has to be done. This manipulation can either be before, or after importing to Instat.

An example with daily data was used in Section 3.4. A second example uses the monthly temperature data from Niamey in Niger that was introduced in Section 3.3, but this time we assume that the data are as was shown in Fig. 3.3g. This typifies the situation where the data are not in the right 'shape'.

The data are in the sheet called TempMean in the Excel workbook niamey.xls. They are to be imported and then reorganised, once in Instat. Go into Instat and use the importing dialogue as shown in Fig. 3.5a. The resulting Instat worksheet is shown in Fig. 3.5b.

Fig. 3.5a Import data from Excel

File ⇒ Open Worksheet

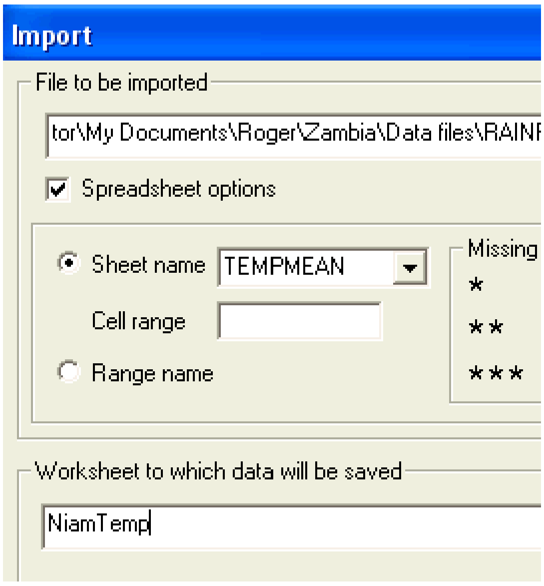
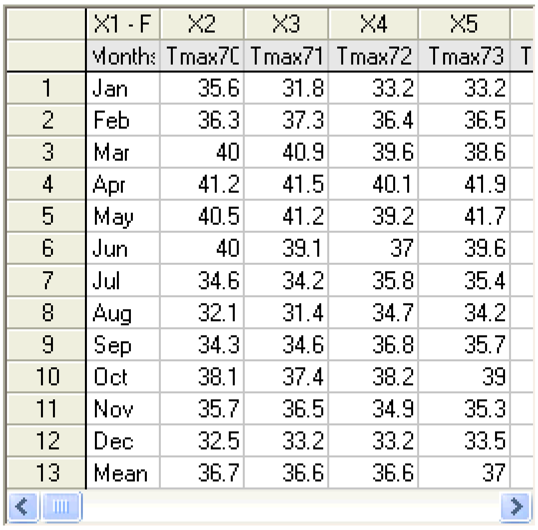


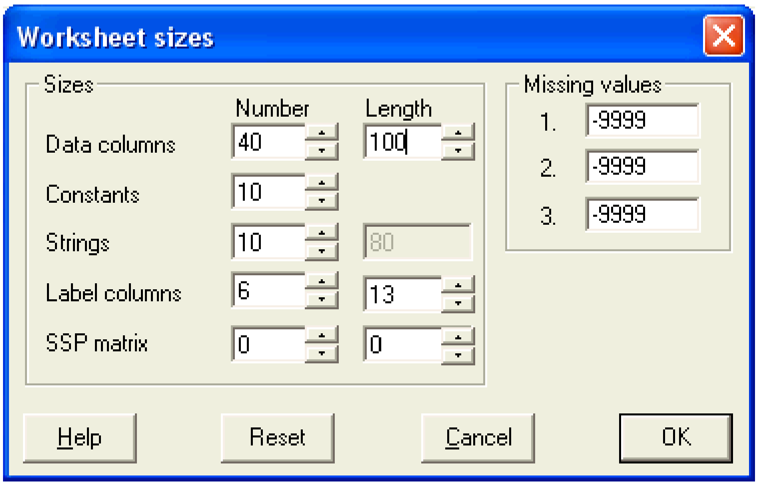
Fig. 3.5b Data in Instat



	X1 - F	X2	X3	X4	X5
	Months	Tmax70	Tmax71	Tmax72	Tmax73
1	Jan	35.6	31.8	33.2	33.2
2	Feb	36.3	37.3	36.4	36.5
3	Mar	40	40.9	39.6	38.6
4	Apr	41.2	41.5	40.1	41.9
5	May	40.5	41.2	39.2	41.7
6	Jun	40	39.1	37	39.6
7	Jul	34.6	34.2	35.8	35.4
8	Aug	32.1	31.4	34.7	34.2
9	Sep	34.3	34.6	36.8	35.7
10	Oct	38.1	37.4	38.2	39
11	Nov	35.7	36.5	34.9	35.3
12	Dec	32.5	33.2	33.2	33.5
13	Mean	36.7	36.6	36.6	37

Fig. 3.5c Changing the worksheet size

Manage ⇒ Resize Worksheet



The steps to reorganise the data use dialogues from Instat's manage menu as follows:

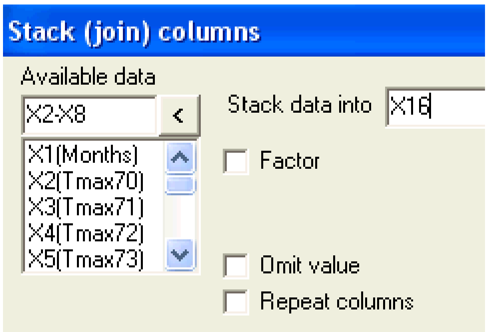
1 The Instat worksheet has only 13 rows, so use **Manage ⇒ Resize Worksheet**, Fig. 3.5c, to add more rows to the worksheet.

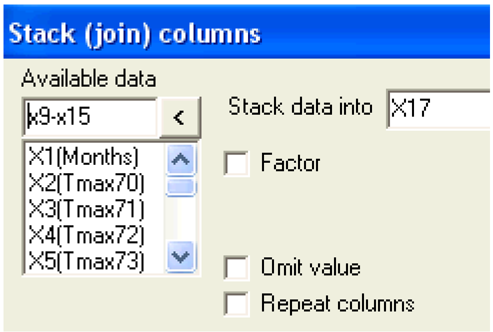
2 Row 13 contains the temperature means, so to remove this row, highlight the data in row 13 and press **<Delete>.**

3 What is needed now is the data in two long columns, rather than two columns per year. This uses the **Manage ⇒ Reshape ⇒ Stack** dialogue as shown in Fig. 3.5d.

Fig. 3.5d Stack Tmax and Tmin

Manage ⇒ Reshape ⇒ Stack





4 Now use Manage ⇒ **Data ⇒ Regular Sequence** twice to enter the data for a year and month column, as shown in Fig. 3.5e and 3.5f.

Fig. 3.5f Enter the year column

Manage ⇒ Data ⇒ Regular Seq.

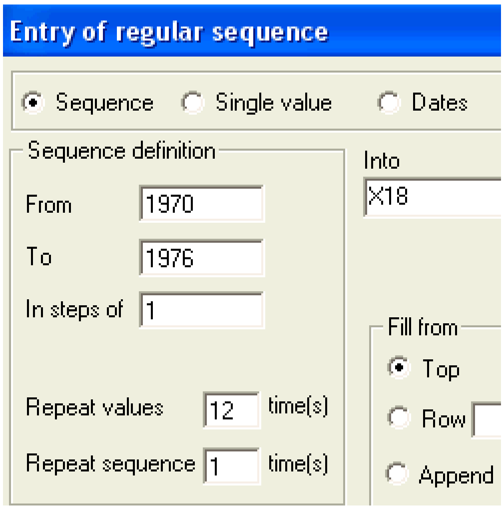
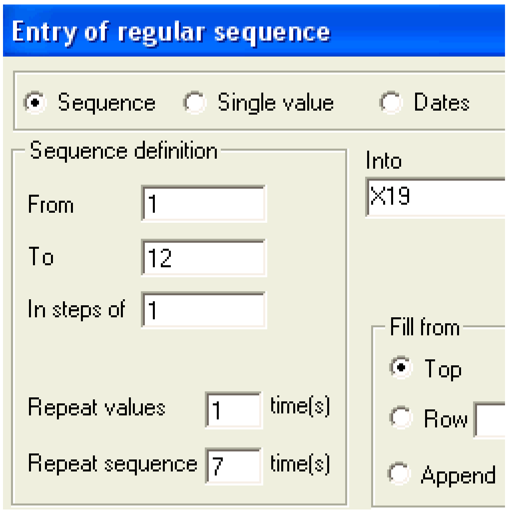


Fig. 3.5g and the month column

Manage ⇒ Data ⇒ Regular Seq.



5 We still need to calculate another year column for plotting the data as shown in Fig. 3.5h.

6 After naming the columns the data should look as shown in Fig. 3.5i.

7 Before proceeding with the next step, save the worksheet. Make sure the worksheet is the current window, and use **File ⇒ Save**.

8 Now remove the original columns and move the new ones over. Use **Manage ⇒ Data ⇒ Clear(Remove)**, as shown in Fig. 3.5j.

9 Then mark all the data (including the column names) in x16-x20, and use **Edit ⇒ Cut** (or **Ctrl<X>**). Then paste the data into x1-x5.

10 Now make x3 and x4 (previously x18 and x19, in Fig. 3.5i) into factor columns. This uses **Manage ⇒ Column Properties ⇒ Factor**, see Figs 3.5k and 3.5l.

11 The worksheet is now as shown in Figs 3.3m and 3.5n and should be saved.

Fig. 3.5h Calculating a date column for plotting

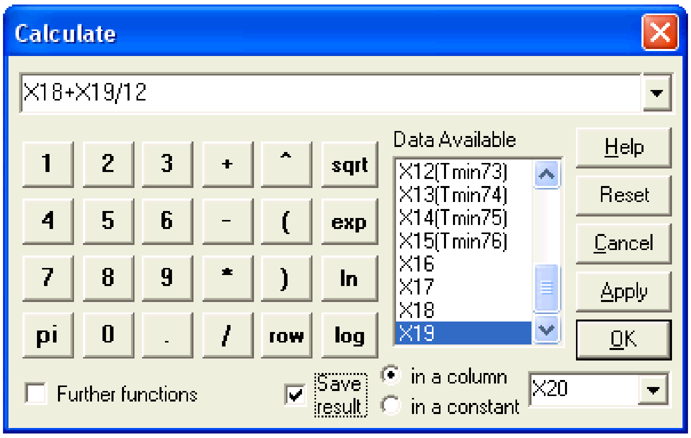


Fig. 3.5i Resulting worksheet

	X16	X17	X18	X19	X20
	Tmax	Tmin	Year	Month	Yr
1	35.6	17.4	1970	1	1970.1
2	36.3	19.1	1970	2	1970.2
3	40	22.4	1970	3	1970.3
4	41.2	26.9	1970	4	1970.3
5	40.5	28.4	1970	5	1970.4
6	40	28	1970	6	1970.5
7	34.6	24.2	1970	7	1970.6
8	32.1	22.9	1970	8	1970.7
9	34.3	23.4	1970	9	1970.8
10	38.1	26.9	1970	10	1970.8
11	35.7	18.6	1970	11	1970.9
12	32.5	15.7	1970	12	1971.0
13	31.8	14.3	1971	1	1971.1

Fig. 3.5j Removing the original data

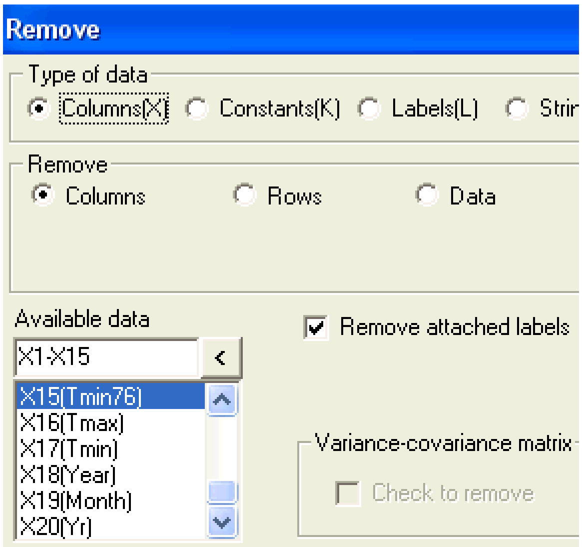


Fig. 3.5k Making Year into a factor column

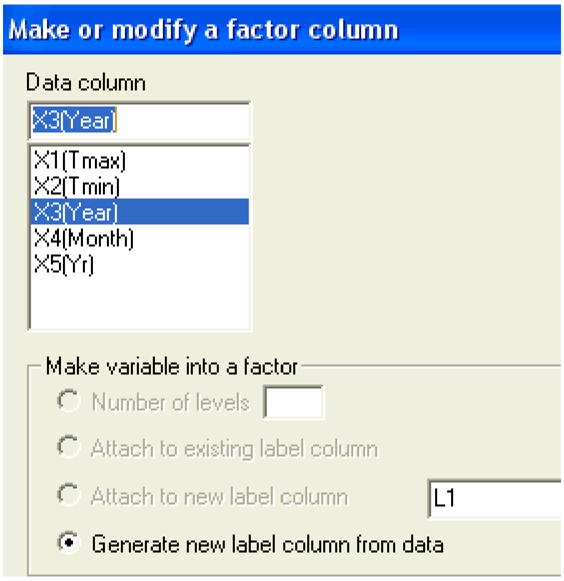


Fig. 3.5l Making Month into a factor

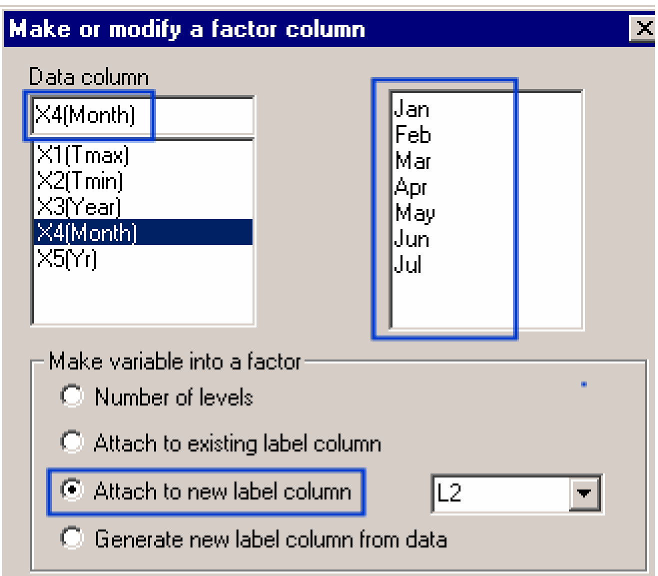


Fig. 3.5m The data in the worksheet

Columns | Constants | Strings | Labels

	X1	X2	X3 - F	X4 - F	X5
	Tmax	Tmin	Year	Month	Yr
1	35.6	17.4	1970	Jan	1970.1
2	36.3	19.1	1970	Feb	1970.2
3	40.0	22.4	1970	Mar	1970.3
4	41.2	26.9	1970	Apr	1970.3
5	40.5	28.4	1970	May	1970.4
6	40.0	28.0	1970	Jun	1970.5
7	34.6	24.2	1970	Jul	1970.6
8	32.1	22.9	1970	Aug	1970.7
9	34.3	23.4	1970	Sep	1970.8
10	38.1	26.9	1970	Oct	1970.8
11	35.7	18.6	1970	Nov	1970.9

Fig. 3.5n And the label columns

Columns | Constants | Strings | Labels

	L1*	L2*	L3
1	1970	Jan	
2	1971	Feb	
3	1972	Mar	
4	1973	Apr	
5	1974	May	
6	1975	Jun	
7	1976	Jul	
8		Aug	
9		Sep	
10		Oct	
11		Nov	
12		Dec	

The alternative, of 'reshaping' the data in Excel, is equally easy, particularly if you have the 'add-in' that has a 'Stack' dialogue that is similar to the one shown with Instat in Figs 3.5d and 3.5e.

The example above is typical in requiring several steps to prepare data in Instat or in another statistics package for the analysis. Another way to reorganise the data is now shown.

1. As an example, import the data again as shown in Fig. 3.5a, but give a different name.
2. Also resize the worksheet, so the columns are of length 100.
3. Now use **Submit ⇒ Run Macro** and look in the climatic library for the macro called **niatemp.ins**, Fig. 3.5o. When you press OK this macro runs, and the data are transformed as shown in Figs 3.5m and 3.5n.

Fig. 3.5o Running the macro to organise the niatemp data

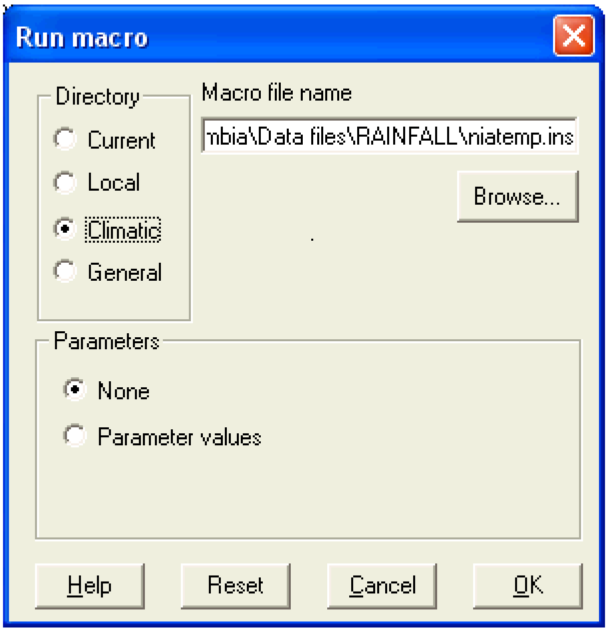


Fig. 3.5p Contents of the niatemp macro

```
DELete 13 x1-x15
JOIn X2-X8;INTo X16
JOIn x9-x15;INTo X17
ENTer X18; data 12(1970]1976)
ENTer X19; data (1]12)7
X20 = X18+X19/12
REMove X1-X15;lab
COPy X16-X20;INTo X1-X5
REMove x16-x20
name x1 'Tmax
name x2 'Tmin
name x3 'Year
name x4 'Month
name x5 'Yr
FACtor 'Year' ;levels L1
ENTer 12;data Jan Feb Mar Apr May Ju
FACtor 'Month L2
LOCk x1-x5
```

The key to this method is the macro of commands, shown in Fig. 3.5p. Most of them were copied from the Commands and Output Window, when going through the detailed steps earlier. They were then put into a macro as explained in Chapter 14 of this guide. They serve as a record of what was done.

It is useful to keep a record of these steps, which serves two purposes:

- If there are queries later about the data, it is clear what was done. (For example, in Fig. 3.5p and earlier in Fig. 3.5h the calculation of X20 = x18+x19/12 could perhaps become x20 = x18+(x19-1)/12. This change is easy to make and record in Fig. 3.5p.)
- If there are other similar datasets, then the routine becomes easier and more reliable.

Appendix 2 shows different data layouts and describes methods for importing these data into Instat. For example Figs 3.5q and 3.5r show the layout of data from Madagascar and Venezuela that have been imported into Instat using macros.

Fig. 3.5q Data from Madagascar

```
Site  Year      (Columns 14-57 contain 11 data values, each of
 |      | Month   width 4. e.g. 0004, 0006. In rows numbered 1 & 2
 |      |  | Row  the 11th value is discarded, so each month has
 ↓      ↓  ↓ ↓    31 days)
534155195201100040006001000030121014100000000002480294000  Jan. 52
534155195201200000000000450135070903970056017602000238000  row 2
534155195201300000134050000140000012601800104011502150024  row 3
534155195202101170102070902040113010501440000000000000000  Feb. 52
534155195202200000000000000116024700000000000000000250000
534155195202300200000000000000000000000000000000000000000
534155195203100000000000000000141013000000000000001900000  Mar. 52
534155195203202810140001400000240032000510000000000000000
.........................................................
.........................................................
.........................................................
534155195312102140023018000000000201300000002510000000000  Dec. 53
534155195312200260000000320000000020004000000000000500000
534155195312300230000000010005002300000000000000001200000014
```

Fig. 3.5r Data from Venezuela

```
M. A. R. N. R.  -  D. G. S. I. C. A. S. V.
DIRECCION DE HIDROLOGIA Y METEOROLOGIA
                                        PRECIPITACION 08000-0800 (mm).
ESTACION: SANTA CRUZ EDAFOLOGI  SERIAL: 0417   Zona: 02  Edo: ARAGUA   AÑO: 1971
Latitud: 10 10'00"  Longitud: 67 29'15"  Altitud: 444 m  Ins: 06/64  Elim: 00/00
***************************************************************************
DIA   ENE   FEB   MAR   ABR   MAY   JUN   JUL   AGO   SEP   OCT   NOV   DIC
***************************************************************************
  1    .0    .0    .0    .0    .0  19.6    .0  10.2    .0    .6   9.1    .7
  2    .0    .0    .3    .0    .0    .7    .6    .0   1.2    .0    .0   7.6
  3    .0    .0   6.0    .0   3.0   9.0    .0    .1    .0    .0    .0    .6
  4    .0    .0  24.7    .0   1.8    .0   1.0   5.6   3.0    .0    .0    .0
  5    .0    .0    .0    .0    .0   5.5    .0   3.1  19.0   2.9   4.4   3.9
  6    .0    .0    .0    .0    .0    .0    .0    .0  10.9    .0   3.2    .0
  7    .0    .0    .0    .0    .0    .0   5.5    .0   5.4    .0  14.0    .0
  8    .0    .0    .0  11.2    .0   8.3  38.0   8.6  11.9    .0    .0    .0
  9    .0    .0    .0    .8    .0  11.1  10.7   2.2    .0    .0    .0    .0
 10    .0   4.3    .0    .0  39.6   4.9    .3   5.5    .0   2.3    .0   8.0
 11    .1    .0    .0    .0    .0   3.9  16.0  42.3    .0    .0    .7    .0
---------------------------------------------------------------------------
---------------------------------------------------------------------------
 30    .0          .0    .0    .0    .0    .0   4.2  11.0   6.0  33.5    .0
 31    .0          .0          .8        10.6    .0          .2          .4
 TOT 10.6   4.3  32.7  17.8 149.3 115.6 260.5 228.9 105.4 113.2  75.8  22.8
MAX   9.6   4.3  24.7  11.2  39.6  23.2  38.8  42.3  19.0  25.6  33.5   8.0
DIAS  20    10     4     8    10    15    26    11     5    22    30    10
TOTAL ANUAL:  1136.9     MAXIMA:   42.3   EL    11/ 8
```

3.6 Exporting data

It is sometimes useful of be able to export data from Instat. One situation is when the first stage of the data processing is done in Instat, while a different package is more appropriate for the subsequent analysis. Or the data may be analysed, but a more powerful graphics package is then to be used to produce presentation plots.

A simple way is to copy the data into the Windows clipboard, using **Edit ⇒ Copy**, and then to paste the data into the other software that is now to be used.

Alternative ways of exporting from Instat are considered below. When moving data between packages you should also consider the importing facilities of the package that is to receive the data. Most can read Excel files, so that is a common intermediate format. The other common option is an ASCII file.

Some packages can import a wide variety of different formats, with Genstat possibly the most flexible of the major statistics packages. It can import directly from most other statistics packages, including Instat. Hence no exporting is needed, if you are transferring data from Instat to Genstat.

For export from Instat, one possibility is to use **File ⇒ Import/Export ⇒ Export As** (Fig. 3.6a). The **Export** dialogue is shown in Fig. 3.6b. This does not require the Instat worksheet to be opened first. The most common option is to export as an Excel file, which can also be read into almost any other package. This dialogue can also export to an ASCII file.

Fig. 3.6a Export dialogue browse option

File ⇒ Import/Export ⇒ Export As

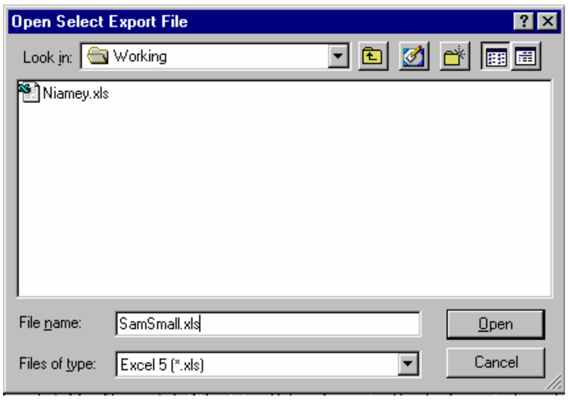


Fig. 3.6b Completed dialogue for Excel

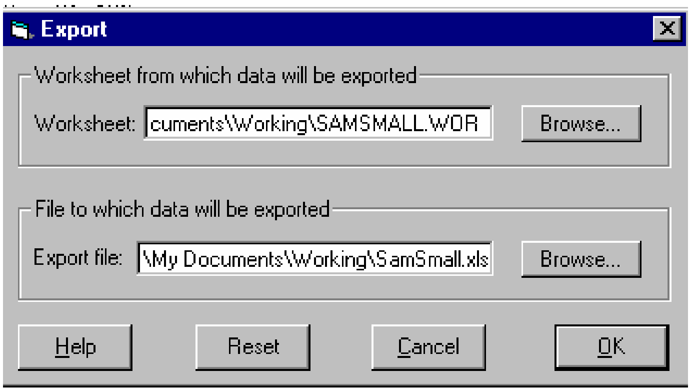
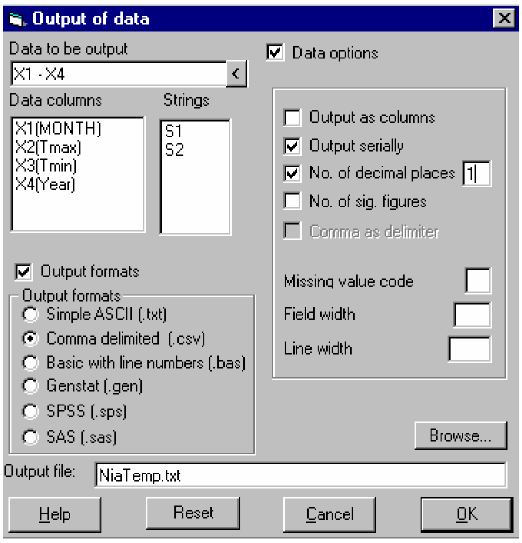


Fig. 3.6c Output dialogue

File ⇒ Import/Export ⇒ Output



The **File ⇒ Import/Export ⇒ Output** dialogue, shown in Fig. 3.6c, provides another way of exporting data from Instat. This needs a worksheet to be open, and can only export as an ASCII file. However, it gives the option to choose precisely which columns of data to export and the format of the exported data. Options include a comma separated file (.csv) that imports automatically to Excel, and formats that add the necessary commands to import easily to the major statistics packages, SPSS, SAS or Genstat.

3.7 Managing climatic data

Clicom is a project organised by the World Meteorological Organisation (WMO). It is designed particularly to improve the facilities for processing climatic data at National Meteorological Offices worldwide. The software is generally available and is also of use to any organisation that has to manage large volumes of climatic or similar data.

New Clicom-style systems have been developed, based on Microsoft Access or Oracle. One project, called ClimSoft, uses Access and is being developed in Africa; further information is available from http://www.weather.co.zw/. Another system, called CliData, is based on Oracle and has been developed by the Czech Met Service. See http://www.clidata.cz/ for details. The World Climate Programme at WMO can be consulted for details on many other developments.

Each of these systems provides facilities for the entry, quality control, management and archiving of climatic data. They are also equally useful for the entry and management of other monitoring data that are collected regularly, particularly other environmental elements, such as pollen counts, river flow, or water quality, that might usefully be processed with, and in the same way, as climatic data.

The original Clicom consists of a series of specially written programs for the data entry and quality control, plus a database with about 35 linked tables, which uses a standard database package called DataEase.

The alternative systems differ in the way they hold the climatic data, but they all have similar tables to hold the meta-data. The meta-data includes information on the following:

- The stations where data are collected, e.g. their location.
- The (climatic) elements that are recorded at each station.
- Definitions of each element.
- Once the actual climatic data have been entered the database can then also be used to give inventory information. For example a query could be of what information exists in the database that is from stations that are within 20kms of a particular location.

Some products and simple applications are distributed as part of these systems. They consist primarily of reports using the database software. They are used to produce tables and graphs of means, totals or extremes for periods that can be defined by the user.

Instat can be viewed as an example of an application. In the successors to Clicom, transfer of data to Instat will be through the **File ⇒ Import/Export ⇒ ODBC Query** dialogue. Meanwhile Instat has a special dialogue to facilitate importing data from the original Clicom and this is described now.

There are two stages to the process. The first is to export the data from Clicom as an ASCII file. The second stage is to read this file into Instat. Data from Clicom Tutorial 2 are used for illustration. The exported file is included for those who wish to understand the process that underlies this part of the work, but do not possess the Clicom package. The first stage is conducted from within Clicom.

Maximum and minimum temperatures for 2 years, 1980 and 1981, have been exported for the station Kundrip. The exported file looks something like Fig. 3.7a. It consists of 48 lines, i.e. 2 years for 2 elements with one month of data on each line. The data consist of 31 values for each month, with each value followed by the contents of the corresponding flag field. When the flag is 'null' there are 2 commas together. The end of the month is 'padded' with missing values for February, April, June, September and November, which have less than 31 days.

You must be in Instat for the second stage. The output from Clicom is an ASCII file and hence may by displayed by **Edit ⇒ Edit Text ⇒ Open from Library** and selecting **Kund.out**. (You do not need to do this for the importing process.)

Fig. 3.7a Structure of the export file to be imported into Instat

```
102,KUND0002 ,002, ,1980-01,23.3,,23.3,,23.6,,23.2,  ...  ,24.7,
102,KUND0002 ,003, ,1980-01,4.7,,5.3,,8.2,,8.5,  ...  ,4.2,
102,KUND0002 ,002, ,1980-02,23.8,,23.5,,23.3,,21, ...  ,-99999,M
102,KUND0002 ,003, ,1980-02,8.2,,10.2,,12.9,,13.5,  ...  ,-99999,M
102,KUND0002 ,002, ,1980-03,24.5,,23.5,,24.8,,26,  ...  ,25.5,
102,KUND0002 ,003, ,1980-03,12.5,,14,,13.4,,11.5,  ...  ,11.6,
  ...                 ...                 ...                 ....
102,KUND0002 ,002, ,1980-06,19.5,,20,,22.8,,22.6,  ...  ,-99999,M
102,KUND0002 ,003, ,1980-06,16,,16.2,,19.1,D,15.3,  ...  ,-99999,M
  ...                 ...                 ...                 ....
  ...                 ...                 ...                 ....
102,KUND0002 ,002, ,1981-12,22.8,,22.8,,23.3,,23.9,  ...  ,22.2,
102,KUND0002 ,003, ,1981-12,13.6,,11.7,,15,,12.8,  ...  ,14.2,
```

To import the data, first use **Climatic ⇒ Manage ⇒ New Worksheet** and create a new climatic worksheet called **Kundrip.wor**. Then use **Climatic ⇒ Manage ⇒ Import Daily Data** as shown in Fig. 3.7b. The file Kund.out, is in the Instat library.

Fig. 3.7b Import data from Clicom

Clim ⇒ Manage ⇒ New ⇒ Kundrip
Clim ⇒ Manage ⇒ Import Daily

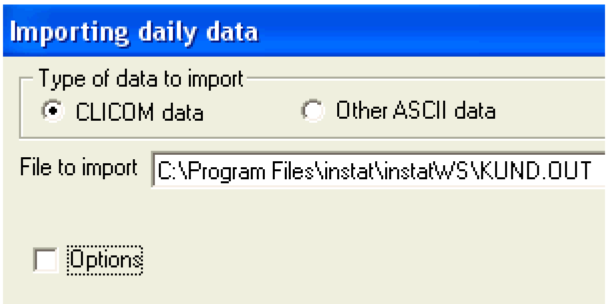
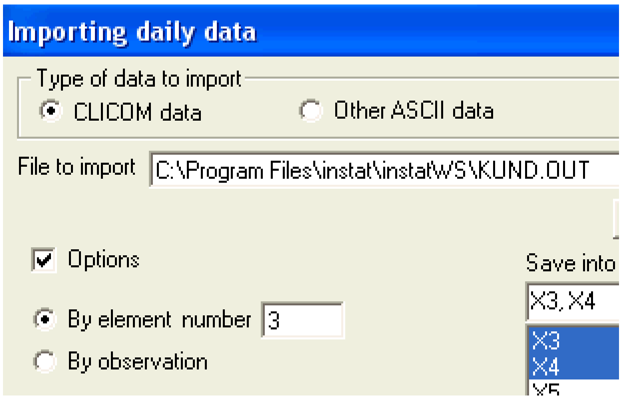


Fig. 3.7c Import minimum temperatures



This will import the data for element 002 (maximum temperature) into the first two columns, X1 and X2. This import is for one element only. To import the minimum temperatures (element 003) into X3 and X4, use the dialogue shown in Fig. 3.7c.

The resulting data in Instat is in Fig. 3.7d.

Fig. 3.7d Data imported from Clicom

	X1	X2	X3	X4
	Tmax80	Tmax81	Tmin80	Tmin81
1	23.3	23.9	4.7	14.7
2	23.3	24.2	5.3	13.3
3	23.6	25	8.2	13.9
4	23.2	23.3	8.5	13.3
5	23.5	22.2	5.6	12.2
6	22.2	25	12.1	11.1
7	24	23.1	14.2	10
8	24.2	26.4	11.5	11.1
9	24.1	22.2	9.7	9.2
10	25.2	26.4	14.9	8.6

In this import any missing values in the data are changed to the missing value code within the worksheet. Trace values are also translated into the special code. Other 'flag' fields are reported, but no special action is taken. The non-existent February 29th is given its own special code in 1981 and the 'padded' missing values at the end of months with less than 31 days are ignored. Hence each of X1-X4 has 366 elements.

This importing procedure is designed primarily for the transfer of a long record for a single element. As shown above, it is possible to transfer more than one element, but they are then read into Instat one at a time.

Clicom can also export data by observation, rather than by element. Fig. 3.7e shows the export of one year of data for 4 elements, namely rainfall, max and min temperatures and sunshine hours. This output file is in fixed format with one line for each day.

Fig. 3.7e Data output by observation

```
KUND0002,1980,01,01,      .0, ,    23.3, ,     4.7, ,     4.1,
KUND0002,1980,01,02,      .0, ,    23.3, ,     5.3, ,     9.8,
KUND0002,1980,01,03,      .0, ,    23.6, ,     8.2, ,     5.9,
KUND0002,1980,01,04,      .0, ,    23.2, ,     8.5, ,     6.0,
KUND0002,1980,01,05,      .0, ,    23.5, ,     5.6, ,     8.8,
KUND0002,1980,01,06,      .0, ,    22.2, ,    12.1, ,    10.6,
KUND0002,1980,01,07,      .0, ,    24.0, ,    14.2, ,    10.8,
  ...          ...             ...        ....
  ...          ...             ...        ....
KUND0002,1980,12,31,     1.5, ,    21.8, ,    11.3, ,     3.7,
```

Within Instat, the same dialogue is used for the import as shown in Fig. 3.7c. The **`By observation`** option indicates that the data are by observation rather than by element. With the same worksheet as before, the dialogue is in Fig. 3.7f and the resulting data in Instat are in Fig. 3.7g.

Fig. 3.7f Import data by observation

`Clim ⇒ Manage ⇒ Import Daily`

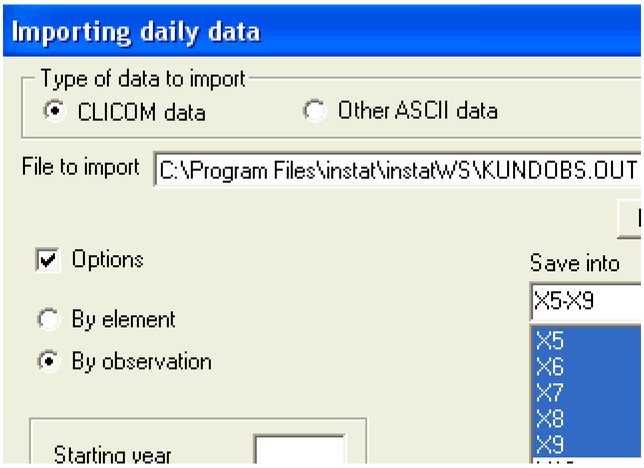


Fig. 3.7g Results

	X5	X6	X7	X8	X9
	rain	tmpmax	tmpmin	sunhrs	date
1	0.0	23.3	4.7	4.1	01-Jan-80
2	0.0	23.3	5.3	9.8	02-Jan-80
3	0.0	23.6	8.2	5.9	03-Jan-80
4	0.0	23.2	8.5	6	04-Jan-80
5	0.0	23.5	5.6	8.8	05-Jan-80
6	0.0	22.2	12.1	10.6	06-Jan-80
7	0.0	24	14.2	10.8	07-Jan-80
8	0.6	24.2	11.5	9.6	08-Jan-80
9	0.0	24.1	9.7	10.3	09-Jan-80
10	0.0	25.2	14.9	10.2	10-Jan-80
11	0.0	24.5	7.4	9	11-Jan-80
12	0.0	24.7	7.9	7.2	12-Jan-80

Five columns of data are imported, with the day number in the final column.

The data in Fig. 3.7e, when exported by observation, are already in columns and can alternatively be read into Instat with the ASCII dialogue. This process is illustrated below. It is the obvious way to transfer data that are other than daily. If data from sources other than Clicom can be exported into ASCII files with a similar structure they can be read into Instat in the same way. If a different statistics package is used, it is likely to have similar facilities for reading data from an ASCII file.

The data file in Fig. 3.7e has 12 fields, namely:

1) Station Name
2) Year
3) Month
4) Day
5) Rainfall value
6) Rainfall Flag
7) Max. Temperature
8) Max Temperature Flag
9) Min. Temperature
10) Min Temperature Flag
11) Sunshine Hours
12) Sunshine Hours Flag

The climatic element values are in fields 5, 7, 9 and 11. One way to import the data is shown in Fig. 3.7h.

Fig. 3.7h Import from an ASCII file

File ⇒ Imp/Exp ⇒ Import ASCII

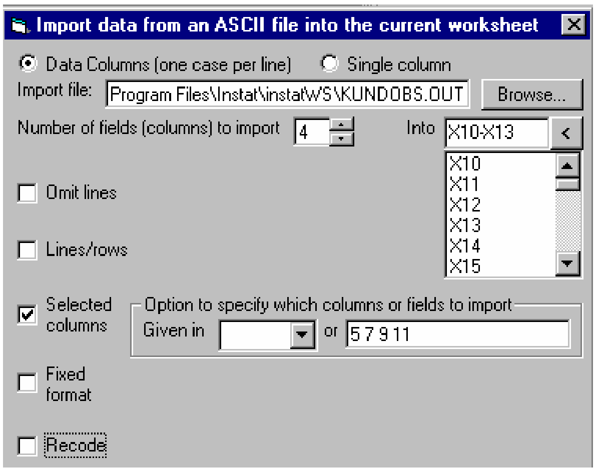
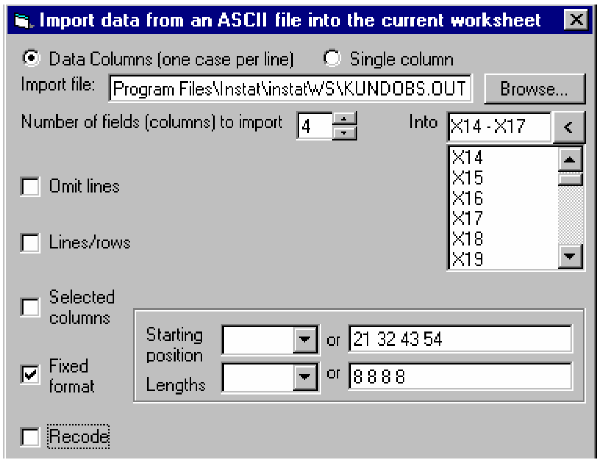


Fig. 3.7i Importing as fixed format



A second way of reading the data, takes advantage of the fact the data are in 'fixed format', i.e. each line has the numbers in the same place. Thus, counting characters across each line, shows the rainfall field starts in position 21, maximum temperature in 32, minimum temperature in 43 and sunshine hours in 54. These starting positions and field widths are entered into the dialogue as shown in Fig. 3.7i, before importing the data.

There is a price to be paid for the use of the general Import ASCII instead of the special Clicom import. In particular the flag information is ignored, so any 'trace' rainfall values (i.e. when the precipitation amount is too small to be measured accurately) are input as zero. Missing values are also not transformed to those of the worksheet. These are not serious problems, particularly if they are considered before importing into Instat. The missing value code can be set as the data are exported from Clicom and can therefore be made the same as in Instat. Alternatively, all editors have the facility to 'find and replace' repeatedly. This can be used for both the missing and the trace values. For the latter the text " **.0,T,**" has merely to be replaced by "**8888,T,**".

Chapter 4 – Presenting Climatic Data

4.1 Introduction

Climatic data are often collected daily. It is important to have simple facilities to examine the raw data, as well as to summarise the observations.

In general, displays of the raw data are interesting in their own right and are also effective as part of the quality control process for the data. Routine methods of quality control on entry are, of course, vital. However, it is easy to miss problems in the data until they are analysed. Hence it is useful if analyses start by re-examining the raw data.

In Sections 4.2 and 4.3 the tabulation and plotting of daily data is considered. This may be used to give a general overview of the records or to highlight features of particular importance. The spell and water balance dialogues are also introduced, that are described in detail in Chapter 6.

4.2 Tabular and graphical presentation of daily data

4.2.1 Display of rainfall data

Within Instat, daily records can be displayed with the **Display Daily** dialogue. The results are illustrated with daily data for 1930 from Samaru. These are in a worksheet called Samsmall. Start by opening the worksheet using **File ⇒ Open Worksheet** or **File ⇒ Open From Library** and selecting **Samsmall.wor**

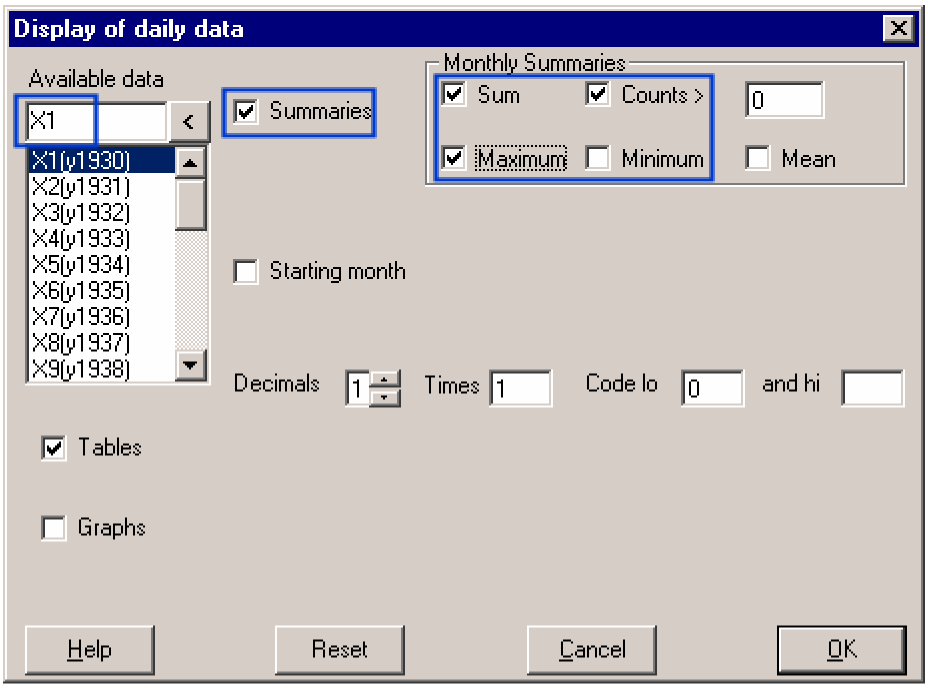
The simplest use of the dialogue **Climatic ⇒ Display Daily** is to select column **X1** and press OK.

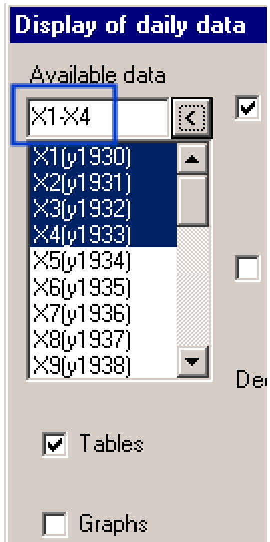
Monthly statistics and zero rainfalls displayed as '--', can be requested as shown in Fig. 4.2a.

Fig. 4.2a Monthly statistics with the tabulation of daily data

File ⇒ Open Worksheet ⇒ Samsmall.wor

Climatic ⇒ Display Daily





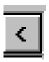
In Fig. 4.2b, only one year is displayed, i.e. X1 (1930). To display more than one year of rainfall data, either select, for example X1, X2, X3 and X4, and press the  button, or type **X1-X4** in the **Available data** field of the **Display Daily** dialogue, Fig. 4.2a.

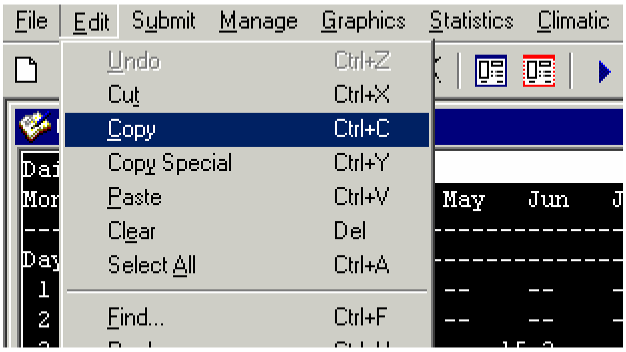
Fig. 4.2b Display of daily rainfall data (mm) from Samsmall, 1930 (dry days coded as '--')

```
Daily data for: y1930
Mon    Jan    Feb    Mar    Apr    May    Jun    Jul    Aug    Sep    Oct    Nov    Dec
----------------------------------------------------------------------------
Day.   ---------------------------------------------------------------------
 1     --     --     --     --     --     --     --    1.0   14.5    --     --     --
 2     --     --     --     --     --     --     --    3.3   26.7    --     --     --
 3     --     --     --     --     --   15.2     --    7.9    1.8    --     --     --
 4     --     --     --     --     --     --    2.3    3.3     --    --     --     --
 5     --     --     --     --     --    6.1     --    3.1    8.6    --     --     --
 6     --     --     --     --     --    3.8    8.4   12.5    9.9    --     --     --
............................................................................
28     --     --     --     --     --    4.8     --   25.9     --    --     --     --
29     --            --     --     --     --    6.9    0.5     --    --     --     --
30     --            --     --    7.1     --   93.0   12.5     --    --     --     --
31     --            --            --           0.8   28.5           --            --
Total                                                       (Overall:  1044.7)
      0.0    0.0    0.0   33.8  115.6   89.7  228.4  417.3  158.0   2.0    0.0    0.0
Maximum                                                     (Overall:  93.0)
      0.0    0.0    0.0   22.1   33.0   18.0   93.0   53.1   26.7   1.8    0.0    0.0
Number greater than 0                                       (Overall:  95)
        0      0      0      5     12     10     20     28     18     2      0      0
```

The output from most non-graphical dialogues is displayed in the **Commands and Output** window. To include the output in a report, mark the output and use **Edit ⇒ Copy** to move the information to the clipboard, see Fig. 4.2d. Then move to your document and select **Edit ⇒ Paste**, to copy the results over.

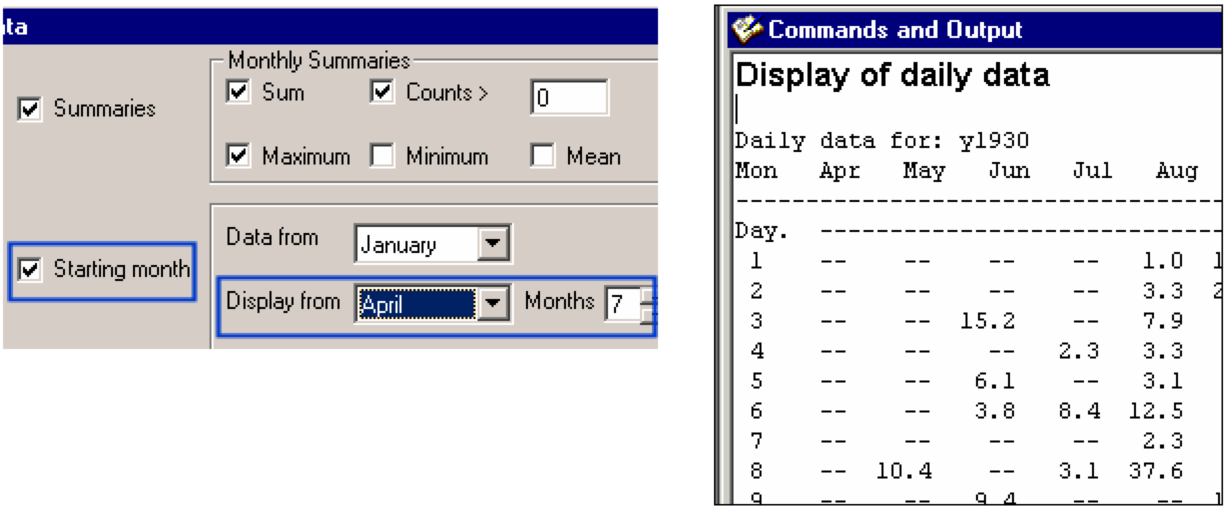
Fig. 4.2c Copying results to a report

Edit ⇒ Copy



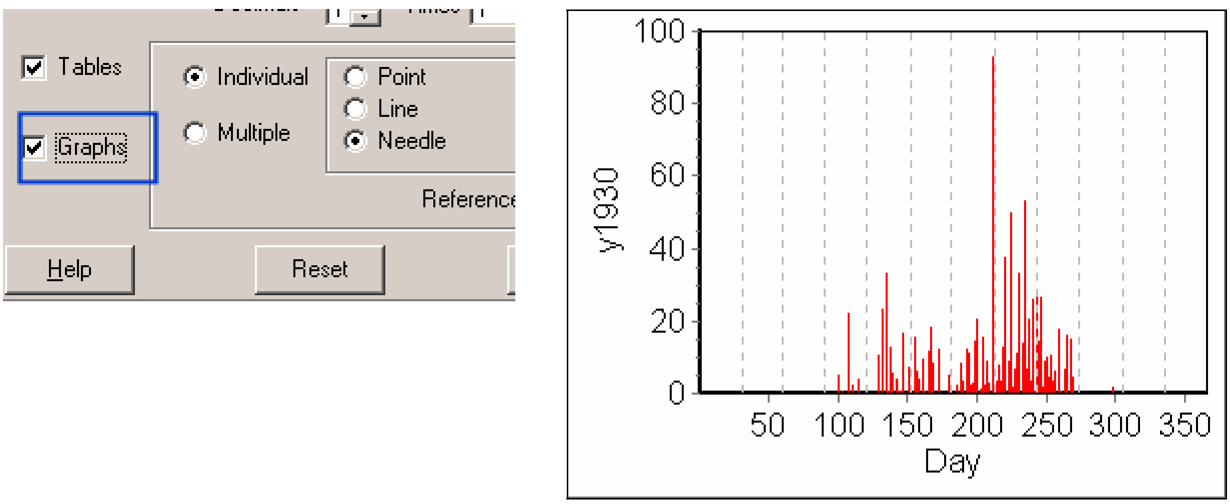
As there is no rain in January to March nor November and December, it is possible to display just the months April to October by completing the **Display Daily** dialogue as in Fig. 4.2e.

Fig. 4.2d Display specific months



The daily data can also be displayed as a graph. Return to the **Display of daily data** dialogue, tick the Graphs box and complete the dialogue as shown in Fig. 4.2e to produce the graphs, one of which is also shown in Fig. 4.2e.

Fig. 4.2e Graph of rainfall from Samaru for 1930



The columns have been given names to help in interpretation of the graph. To compare the four graphs, use **Window ⇒ Tile** or **<F3>**. An alternative is to select Multiple Graphs (see Fig. 4.2f), when all the years will be shown on one graph. However, this may not be so clear as comparing the individual graphs, especially if many years are selected.

Each of the graphs can be saved using **File ⇒ Export** and choosing the format; emf (Enhanced metafiles) were used in this guide. (Within a Word document or Powerpoint presentation, use **Insert ⇒ Picture ⇒ From File** to insert them.)

To close all the graphs use **File ⇒ Close All Graphs**. When you close the worksheet, Instat will ask if you wish to save the changes. They do not need to be saved.

4.2.2 Display of temperature data

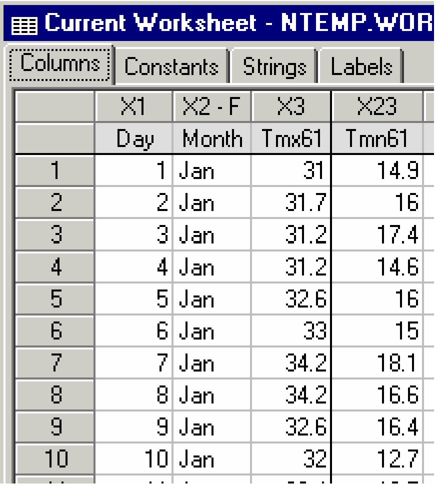
The graph in Fig. 4.2g is a needle plot, i.e. the rainfall points are joined vertically to the x-axis. For other types of climatic data, it is better to use a line plot. This is illustrated, using temperature data from Niamey in Niger.

Open the worksheet **Ntemp.wor** (Fig. 4.2i), either directly or from the library. Note that X3 contains the maximum temperature and X23 the minimum temperature for 1961.

Return to the **Display Daily** dialogue and complete the dialogue as shown in Fig. 4.2j.

Fig. 4.2f Temperature data from Niamey

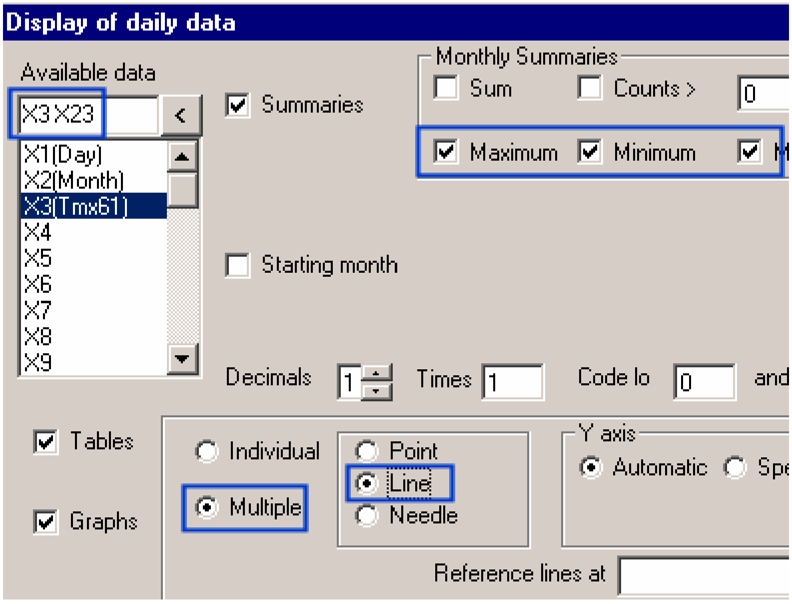
File ⇒ Open Worksheet ⇒ Ntemp.wor or (Open From Library)


Current Worksheet - NTEMP.WOR
Columns | Constants | Strings | Labels

	X1	X2 - F	X3	X23
	Day	Month	Tmx61	Tmn61
1	1	Jan	31	14.9
2	2	Jan	31.7	16
3	3	Jan	31.2	17.4
4	4	Jan	31.2	14.6
5	5	Jan	32.6	16
6	6	Jan	33	15
7	7	Jan	34.2	18.1
8	8	Jan	34.2	16.6
9	9	Jan	32.6	16.4
10	10	Jan	32	12.7

Fig. 4.2g Displaying temperature data

Climatic ⇒ Display Daily

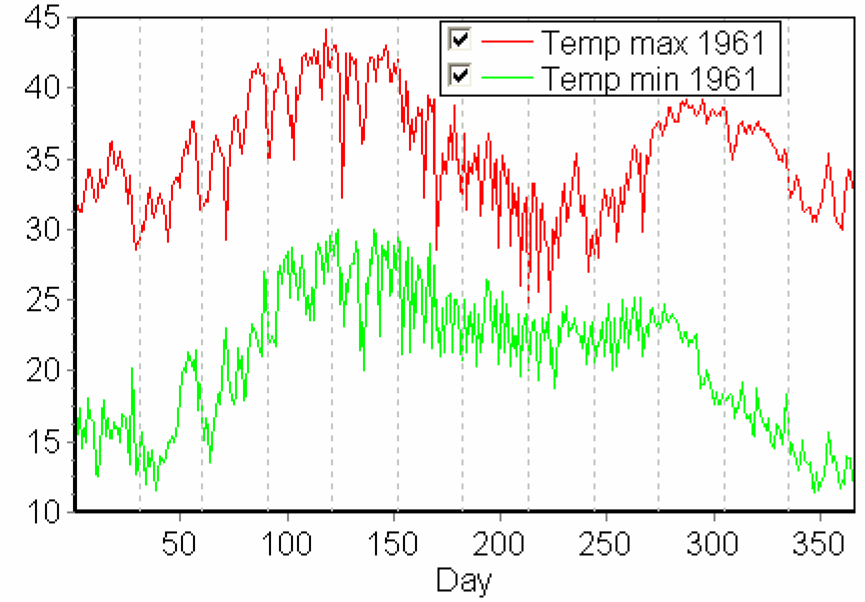


Part of the tabular output for X3 (Tmx61) is in Fig. 4.2h and the graph of Tmx61 and Tmn61 is in Fig. 4.2i.

Fig. 4.2h Daily maximum temperatures for 1961 at Niamey, Niger

```
Display of daily data

Daily data for: Tmx61
Mon    Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
--------------------------------------------------------------------------
Day.  --------------------------------------------------------------------
 1   31.0  30.2  31.7  35.4  43.0  40.8  36.8  29.2  30.0  37.5  38.4  32.2
 2   31.7  30.0  32.1  37.0  43.0  38.4  31.4  33.3  28.0  36.6  36.5  32.6
 3   31.2  31.2  32.0  39.7  42.0  36.9  34.8  33.2  31.0  37.4  36.0  33.0
 4   31.2  32.2  33.4  39.7  40.2  40.4  34.2  31.6  31.6  37.3  35.0  33.8
.........................................................................
28   29.5  31.4  40.9  41.7  40.5  35.0  31.0  27.0  37.2  38.2  35.6  34.3
29   28.5        41.0  41.5  41.7  33.0  31.4  29.0  37.4  38.0  35.3  34.1
30   29.1        39.6  42.5  41.0  32.4  32.3  29.6  37.6  38.6  33.4  33.0
31   29.3        34.7        42.0        24.6  27.8        38.6        33.3
Mean                                                    (Overall: 36.5)
     32.8  32.9  36.7  40.5  40.6  36.3  32.9  30.5  33.3  38.3  36.4  32.3
Minimum                                                 (Overall: 24.2)
     28.5  29.2  29.3  35.0  32.3  28.6  24.6  24.2  28.0  36.6  33.4  30.0
Maximum                                                 (Overall: 44.2)
     36.2  37.6  41.7  44.2  43.0  40.8  36.8  35.4  37.6  39.2  38.4  35.4
```

Fig. 4.2l Plots from the Display command of maximum and minimum temperatures in 1961



4.3 The 'Spell' and 'Water Balance' dialogues

Other dialogues can display daily climatic data in a table format to emphasise different aspects of the pattern of rainfall through the year. For illustration, return to the data from Samaru , i.e. **File** ⇒ **Open** and select **Samsmall.wor**. If you saved the changes before, then use **Manage** ⇒ **Data** ⇒ **Clear(Remove)** to delete unwanted data in X12-X60.

Fig. 4.3a The Climatic menu

File ⇒ Open ⇒ Samsmall.wor
Climatic ⇒ Events ⇒ Spells

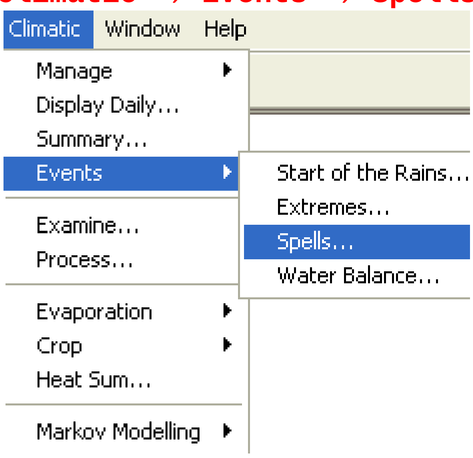
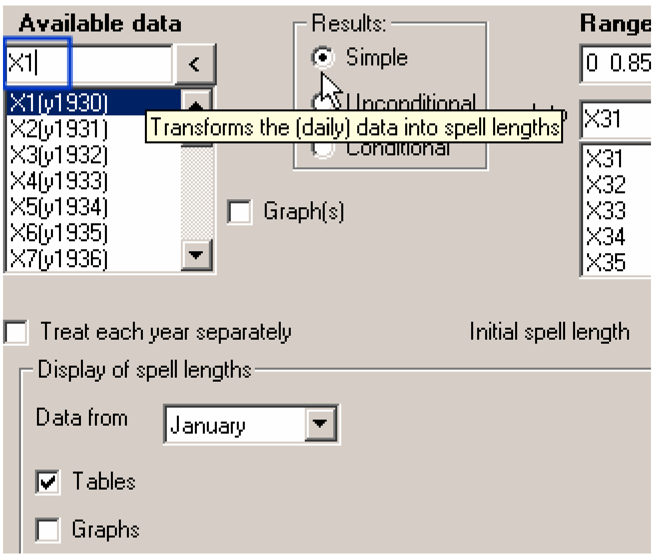


Fig. 4.3b Calculating dry spell lengths
Climatic ⇒ Events ⇒ Spells



4.3.1 Spell Lengths

To calculate dry spell lengths, select **Climatic** ⇒ **Events** ⇒ **Spells** (Fig. 4.3a). Complete the **Spell lengths** dialogue as shown in Fig. 4.3b. A dry day was defined as a day with less than 0.85mm of rain. Our reasons for choosing this seemingly curious value are considered in more detail in Chapter 6. However, one principle in constructing Instat has been to give as much choice as possible to the user. Hence other thresholds may be substituted or the results from a range of thresholds may be compared.

Fig. 4.3c displays the dry spell lengths, which were stored in X12.

Fig. 4.3c Dry spell lengths for Samaru (1930)

Spell Lengths

Mon	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Day.												
1	m	m	m	m	7	2	3	--	--	6	8	38
2	m	m	m	m	8	3	4	--	--	7	9	39
3	m	m	m	m	9	--	5	--	--	8	10	40
4	m	m	m	m	10	1	--	--	1	9	11	41
5	m	m	m	m	11	--	1	--	--	10	12	42
6	m	m	m	m	12	--	--	--	--	11	13	43
7	m	m	m	m	13	1	1	--	--	12	14	44
8	m	m	m	m	--	2	--	--	--	13	15	45
9	m	m	m	m	1	--	1	1	--	14	16	46
10	m	m	m	--	2	1	2	--	--	15	17	47
11	m	m	m	1	--	2	--	--	--	16	18	48
...												
30	m		m	6	--	2	--	--	5	6	37	67
31	m		m		1		1	--		7		68
Maximum										(Overall: 68)		
	m	m	m	6	13	7	5	1	5	28	37	68
Number of missing values										(Overall: 99)		
	31	28	31	9	0	0	0	0	0	0	0	0

The early part of the year is coded as missing. In the absence of information at the end of 1929 it is not possible to give the spell lengths for 1930 until the first wet day. As shown in Fig. 4.3c, that was on 10th April. Fig. 4.2b earlier, showed that the rainfall on 10th April was just over 5mm. This is not enough for sowing a cereal crop, and the results in Fig. 4.3c show that there was a dry spell of 13 days ending on May 8th, so the sowing early would be unlikely to be a success. The output also shows the maximum dry spell for each month. In this year there was no dry spell of more than 7 days from June to September.

4.3.2 Water Balance

Many definitions of the end of the growing season use a simple water balance equation. In this chapter the water balance dialogue in Instat is used as a further illustration of displaying daily data.

To calculate the Water balance, select **Climatic ⇒ Events ⇒ Water Balance** (Fig. 4.3a). The water balance dialogue requires the rainfall data and either a number or a column containing evaporation data. Complete the dialogue to look like Fig. 4.3d. The evaporation is taken as 5mm per day and the storage limit as 100mm.

The plot of the water balance is shown in Fig. 4.3e and the numerical results are in Fig. 4.3f. In Fig. 4.3f, data are automatically coded '--' when the water balance is zero and '++' when full.

Fig. 4.3d The water balance dialogue

Climatic ⇒ Events ⇒ Water Balance

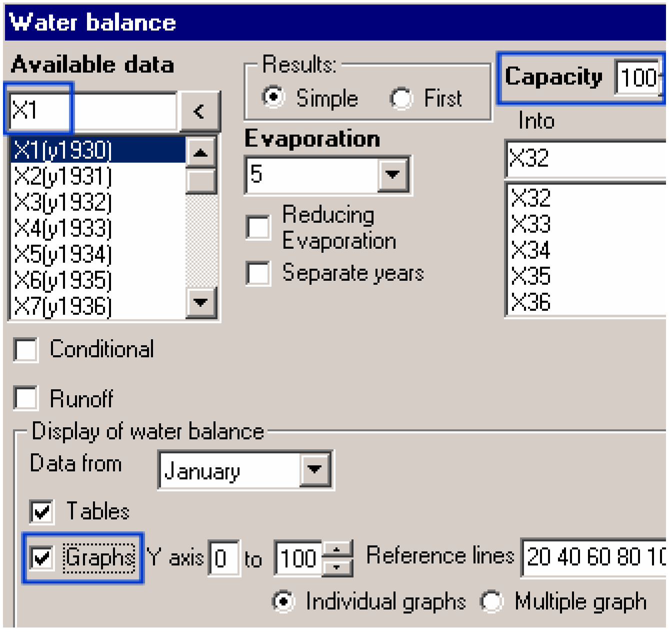


Fig. 4.3e Water balance plot for 1930

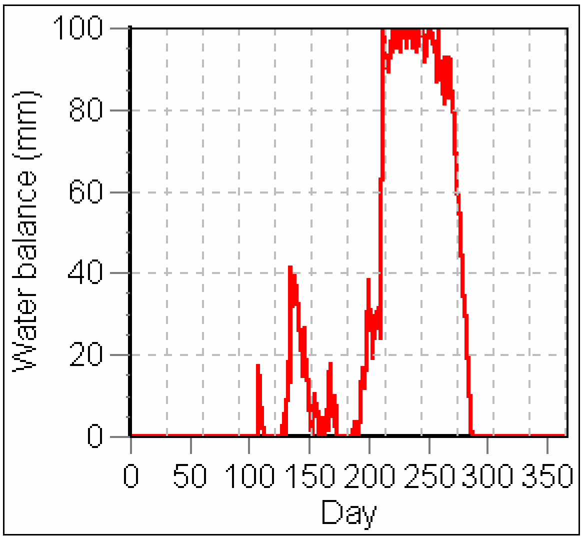


Fig. 4.3f Simple water balance for Samaru, 1930

Mon Day.	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	--	--	--	--	--	--	--	92	++	62	--	--
2	--	--	--	--	--	--	--	90	++	57	--	--
3	--	--	--	--	--	10	--	93	97	52	--	--
4	--	--	--	--	--	5	--	91	92	47	--	--
5	--	--	--	--	--	6	--	89	95	42	--	--
6	--	--	--	--	--	5	3	97	++	37	--	--
7	--	--	--	--	--	0	--	94	++	32	--	--
8	--	--	--	--	5	--	--	++	99	27	--	--
9	--	--	--	--	0	4	--	95	++	22	--	--
10	--	--	--	0	--	--	--	99	97	17	--	--
11	--	--	--	--	--	--	7	++	96	12	--	--
12	--	--	--	--	18	--	11	++	97	7	--	--
13	--	--	--	--	13	6	17	97	92	2	--	--
14	--	--	--	--	41	1	14	94	87	--	--	--
15	--	--	--	--	37	14	12	96	99	--	--	--
16	--	--	--	--	32	18	21	++	94	--	--	--
17	--	--	--	17	39	13	23	++	90	--	--	--
18	--	--	--	12	34	8	38	++	85	--	--	--
19	--	--	--	9	35	3	33	95	86	--	--	--
20	--	--	--	4	30	10	28	++	82	--	--	--
21	--	--	--	--	28	5	24	++	93	--	--	--
22	--	--	--	--	23	--	19	++	88	--	--	--
23	--	--	--	--	19	--	29	++	83	--	--	--
24	--	--	--	--	15	--	27	95	93	--	--	--
25	--	--	--	--	26	--	30	++	92	--	--	--
26	--	--	--	--	21	--	32	99	87	--	--	--
27	--	--	--	--	16	--	29	94	82	--	--	--
28	--	--	--	--	11	--	24	++	77	--	--	--
29	--		--	--	6	--	26	96	72	--	--	--
30	--		--	--	8	--	++	++	67	--	--	--
31	--		--		3		96	++		--		--

With this simple water balance, the results in Fig. 4.3f indicate that a planting on 10th April would have suffered from no available water between 21st April and 8th May. Planting on 12th May would have been better, but again there was no water by 8th June. Little rain fell until July and the profile was not full until 30th July. If the end of the season is defined as the first date the soil profile is empty after 1st September, then Fig. 4.3f also shows the end date is 14th October. Hence, with the rainfall of 1930 and this hypothetical evaporation, the length of the season, from 12 May to 14 October, was about 5 months.

4.4 Graphical presentation of the daily data

Figs 4.2g, 4.2l showed the graphs produced from the **Display Daily** dialogue. Sometimes these automatic graphs are not sufficient and the **Graphics ⇒ Plot** dialogue can then be used. One example is to plot the rainfall and the water balance together. Assume the water balance dialogue has been used as shown earlier, so the results are in X13. Then from the **Graphics** menu, select **Plot**. Complete the plot dialogue as shown in Figs 4.4a to 4.4d. Notice that using Define Lines and Symbols, rainfall is displayed as a vertical plot in red and the water balance in blue.

Fig. 4.4a Defining rain and water balance plot

Graphics ⇒ Plot

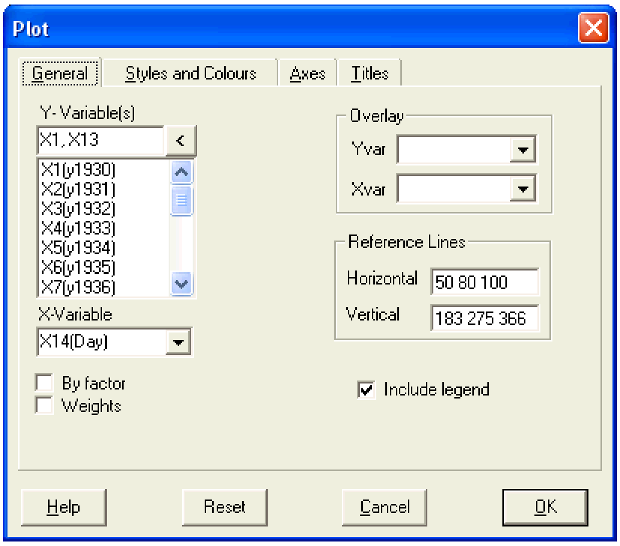


Fig. 4.4b Defining axes

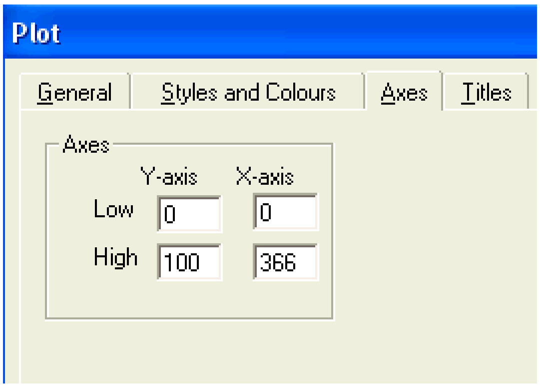


Fig. 4.4c Defining symbol and line styles

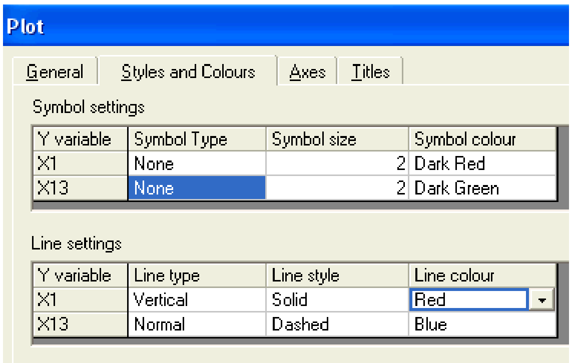
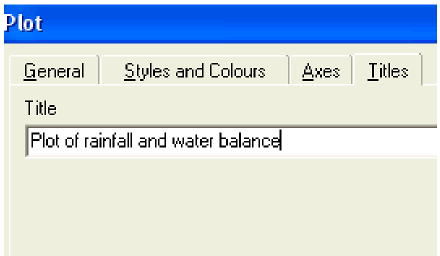


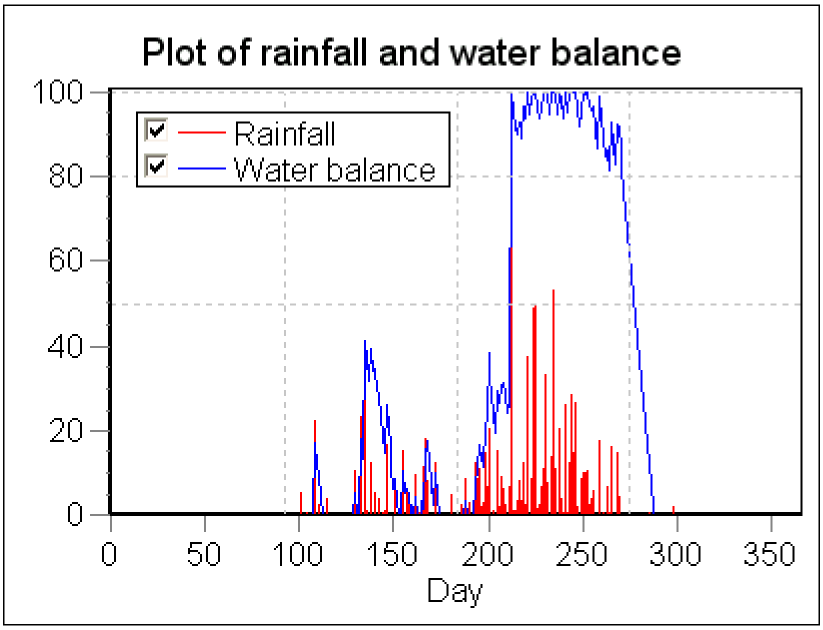
Fig. 4.4d Adding the title



Careful choice of plotting options can give clear graphs, which assist interpretation. In Fig. 4.4e, for instance, the graph with both the water balance and the rainfall, with the rainfall as a 'needle' plot, can be informative and easy to interpret.

It is also useful to add reference lines to the plot. The horizontal lines emphasise the state of the water balance, while the vertical lines indicate the different seasons.

Fig. 4.4e Plot of rain and water balance for Samaru 1930



If several years of rainfall and water balance data are to be looked at together, it may be simpler to use the Instat commands or even write a macro to do the same thing. (How to write Instat macros is explained in Chapter 14).

For example to compare the years 1930-1933 graphically, the commands shown in Fig. 4.4f could be copied into an Input window. If this looks difficult use the **Climatic ⇒ Events ⇒ Water Balance** dialogue followed by the **Graphics ⇒ Plot** dialogue instead. Then paste the commands that were generated, into an Input Window.

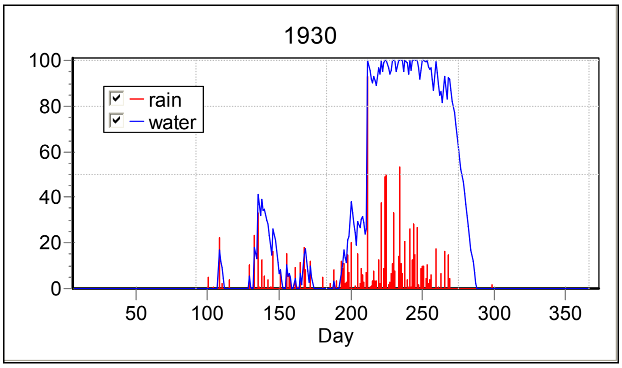
When typing commands **<Shift> <F11>** can be used to edit the last command. This makes it easy to produce the successive plot commands that are very similar to each other.

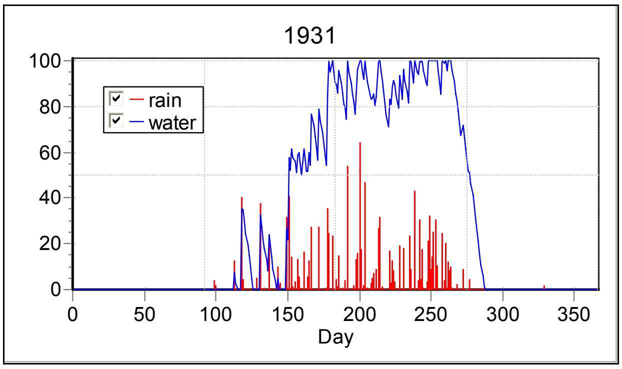
Fig. 4.4f Instat commands to produce plot in Fig. 4.4g

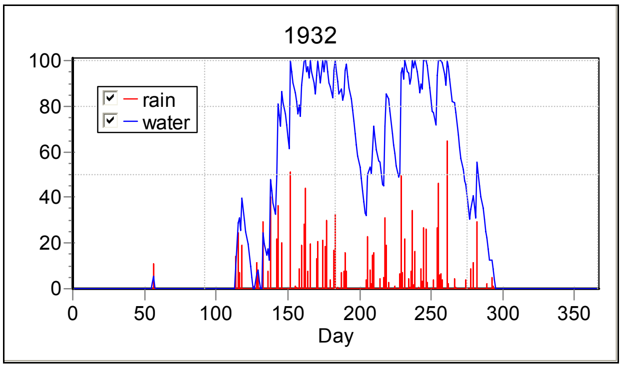
```
Water X1-X4; evaporation 5; capacity 100; into x21-x24
Line x1 2 1 1  : Line x2 2 1 1  :  Lin x3 2 1 1  : Lin x4 2 1 1
Line x21 1 1 4  :  Line x22 1 1 4  : Lin x23  1 1 4  :  Line x24  1 1 4
plot x1x21 'Day';xax 0 366;yax 0 100;href 50 80 100;vref 92 183 275 366; tit"1930";nolegend
plot x2x22 'Day';xax 0 366;yax 0 100;href 50 80 100;vref 92 183 275 366; tit"1931";nolegend
plot x3x23 'Day';xax 0 366;yax 0 100;href 50 80 100;vref 92 183 275 366; tit"1932";nolegend
plot x4x24 'Day';xax 0 366;yax 0 100;href 50 80 100;vref 92 183 275 366; tit"1933";nolegend
```

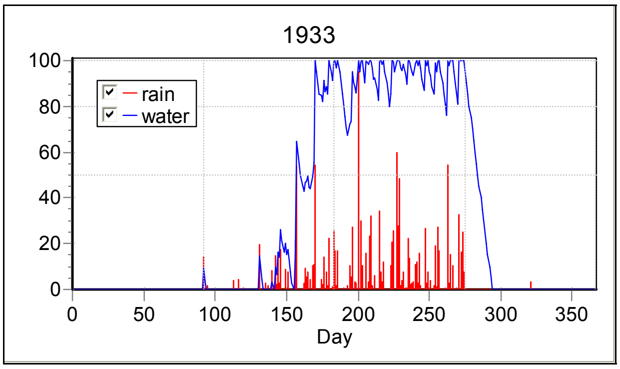
Each plot command produces a graph. To compare the four graphs, use **Window ⇒ Graphs ⇒ Tile** to produce a picture similar to Fig. 4.4g.

Fig. 4.4g Rainfall and water balance plots for years 1930-1933









Each of the graphs can be individually saved, using **File ⇒ Export** and inserted into a Word document.

4.5 Conclusions

The climatic dialogues introduced in this chapter have been used mainly for data exploration. This is a quick, but important task at the start of the analysis. The data are then summarised, as described in later chapters, and other oddities may appear. Then it is useful to return to the ideas in this chapter, to investigate the data from those particular years in more detail.

Chapter 5 – Summarizing Climatic Data

5.1 Introduction

Our main task in this chapter is to show how to summarise daily data over monthly or 10 day periods. These are only examples and results can be calculated for any period of the user's choice, e.g. 5 or 7 days.

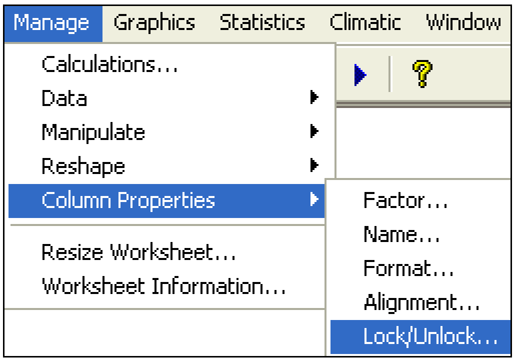
Starting from the daily records, the full analysis of the climatic data is therefore often a two-step process. The first is that of an initial summary and is described in this Chapter and in Chapter 6. Then Chapter 7 describes the second step, of processing the data that have been summarised.

5.2 Preliminary steps

The data are daily rainfall records for 11 years from Samaru, Nigeria. It is the set that was used for presentations and Graphics in Chapter 4. The first task, as in Chapter 4, is to open the worksheet. Use **File ⇒ Open** and select **Samsmall.wor**.

The 11 years of data are in columns X1-X11 and the first preliminary step is to make sure that the columns of data are locked to prevent them from accidentally being overwritten. Use **Manage ⇒ Column Properties ⇒ Lock / Unlock** as in Fig. 5.2a.

Fig. 5.2a Column Properties submenu



If the columns are already locked, the **Lock / Unlock Columns** dialogue will look as in Fig. 5.2b. If the columns are not locked, the dialogue may look as in Fig. 5.2c.

Fig. 5.2b Columns X1-X11 already locked

Manage ⇒ Data ⇒ Lock/Unlock

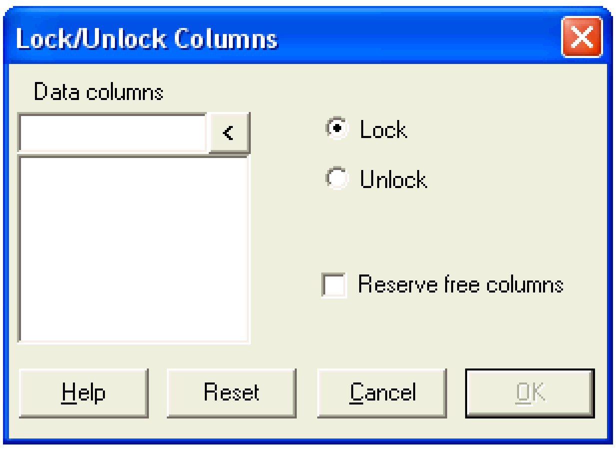
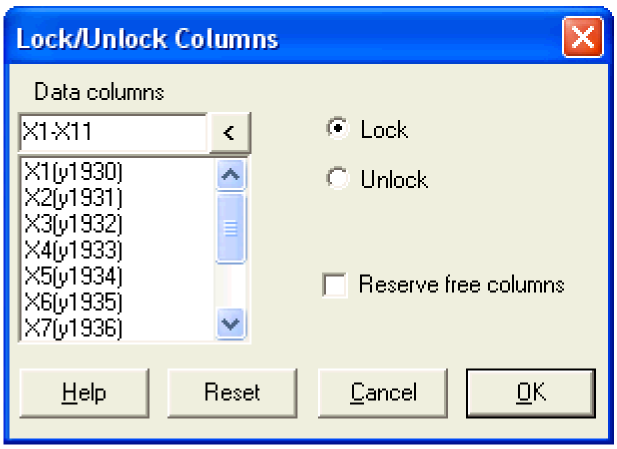


Fig. 5.2c If the columns are not locked

Manage ⇒ Data ⇒ Lock/Unlock



The status of the columns could have been seen directly from the worksheet, as shown below, where the locked columns have a * beside them, e.g. **X1***

This step only has to be done once. The columns will remain locked the next time the same set of data is used.

To see the contents of the worksheet, use **Manage ⇒ Worksheet Information** (Fig. 5.2d and 5.2e).

Fig. 5.2d Information about the worksheet

Manage ⇒ Worksheet Information

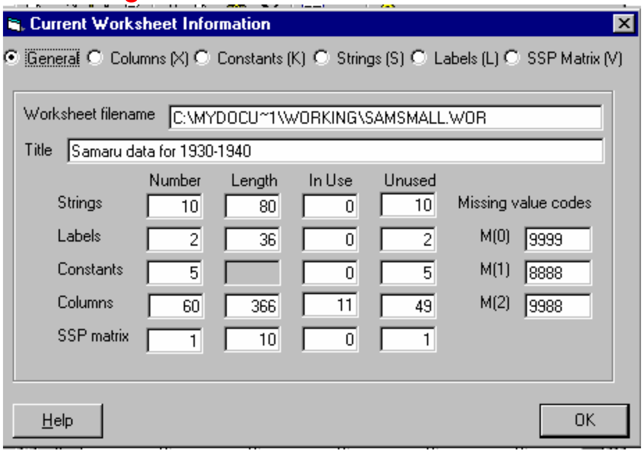
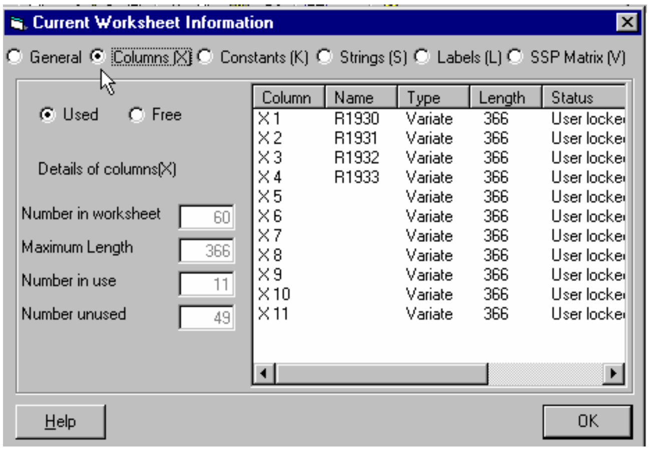


Fig. 5.2e Column details

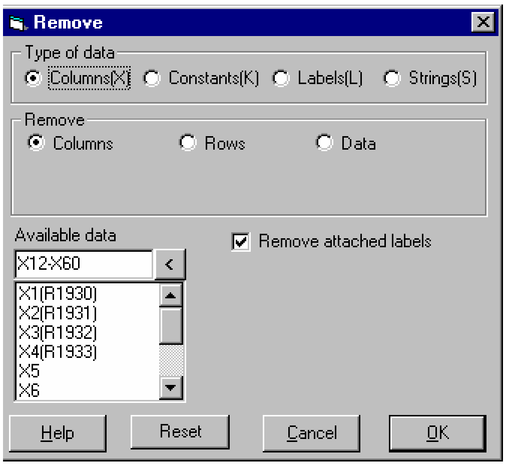


The 60 columns are each of length 366. Clicking the Columns (X) option, shows (Fig. 5.2e), that X1-X11 are 'User locked'. They contain the raw data, i.e. the daily rainfall values for 11 years. If there are extra unlocked columns in your worksheet, they may contain the results from a previous analysis.

If you need to remove the extra columns to start each analysis with a 'clean' worksheet, use **Manage ⇒ Data ⇒ Clear (Remove)**. To remove any labels at the same time as removing the columns, tick the Remove attached labels box. (Fig. 5.2f).

Fig. 5.2f Remove extra columns

Manage ⇒ Data ⇒ Clear(Remove)



Usually at this point, following good statistical practice, the raw data would be examined, using the methods described in the last chapter. These steps are omitted here to demonstrate other facilities.

The two steps above, of checking that the data columns are locked, and of clearing unwanted columns are general and apply in many chapters. The final preliminary step is particularly needed for this chapter. This is to produce a column, here X12, which uses the numbers 1 to

12, to specify the months. There is a special dialogue to perform this task, so use **Climatic ⇒ Manage ⇒ Make Factor** (Fig. 5.2g) and complete the dialogue as in Fig. 5.2h.

Fig. 5.2g Climatic Manage menu

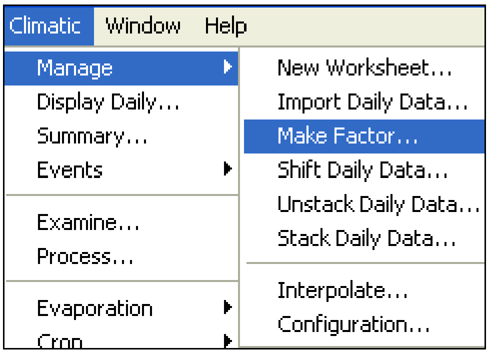
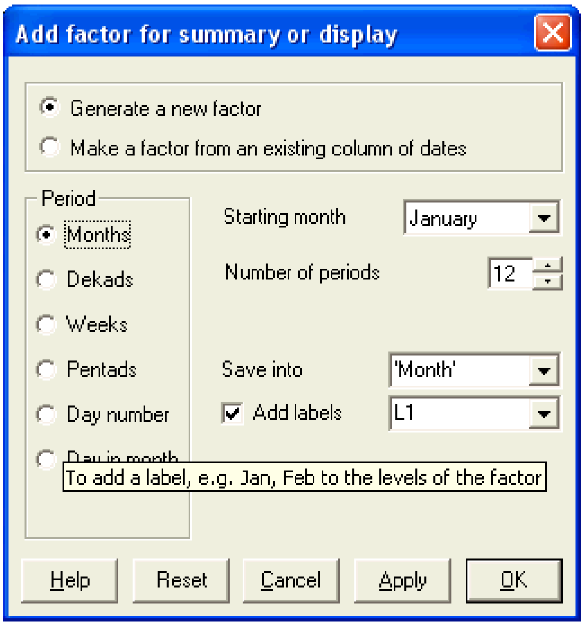


Fig. 5.2h Make factor column with monthly labels

Climatic ⇒ Manage ⇒ Make Factor



This uses the label column, L1 to label the months. The worksheet, Fig. 5.2i indicates what has been done at this stage. Notice, from the dialogue in Fig. 5.2h**,** that it would have been equally easy to produce a column that identifies the dekads, weeks or pentades in the year.

Fig. 5.2h Samsmall with display of labels

Columns | Constants | Strings | Labels

	L1*	L2
1	Jan	
2	Feb	
3	Mar	
4	Apr	
5	May	
6	Jun	
7	Jul	
8	Aug	
9	Sep	
10	Oct	
11	Nov	
12	Dec	
13		

....and data showing April/May

	X10*	X11*	X12 - F Month
117	0	0	Apr
118	0	26.67	Apr
119	0	0	Apr
120	0	0	Apr
121	0	0	Apr
122	0	0	May
123	4.57	0	May
124	0	0	May
125	0	0	May
126	0	3.05	May
127	0	17.78	May

5.3 Single Summaries

This section describes how single summary columns can be formed. The routine to give all the monthly totals is in the next section. In both cases the dialogue is **Climatic ⇒ Summary**. Complete it as in Fig. 5.3a.

Fig. 5.3a Summary dialogue for annual rainfall totals

`Climatic ⇒ Summary`

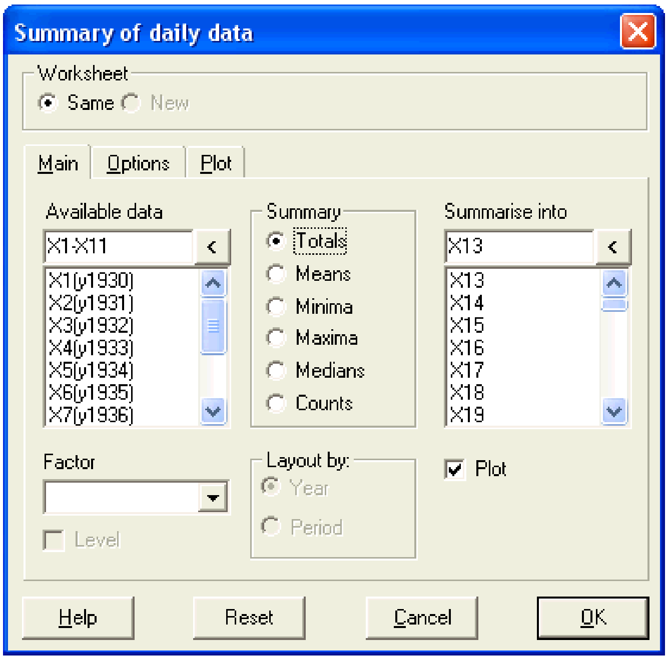


Fig. 5.3b Annual rainfall totals

	X13
1	1044.73
2	1197.82
3	1198.16
4	1311.61
5	1076.67
6	996.43
7	1150.32
8	982.98
9	987.06
10	1129.27
11	910.82

The annual totals are displayed in the **Commands and Output** window and can also be seen in the spreadsheet (Fig. 5.3b).

The dialogue in Fig. 5.3a includes an option to produce the graph shown in Fig. 5.3c. This plots the annual totals by year, together with a solid line, for the mean of the annual totals over the 11 years (1089.6m).

Fig. 5.3c Annual totals for 11 years

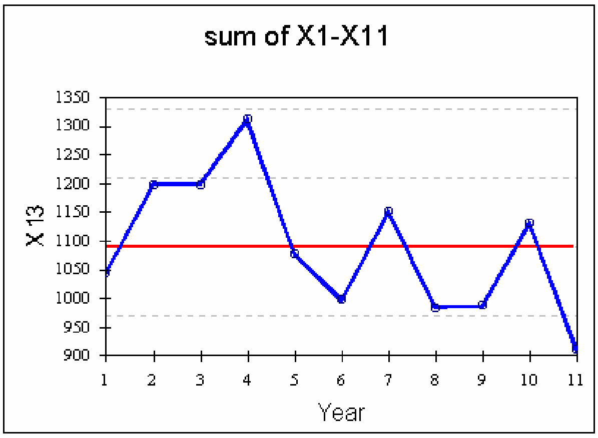
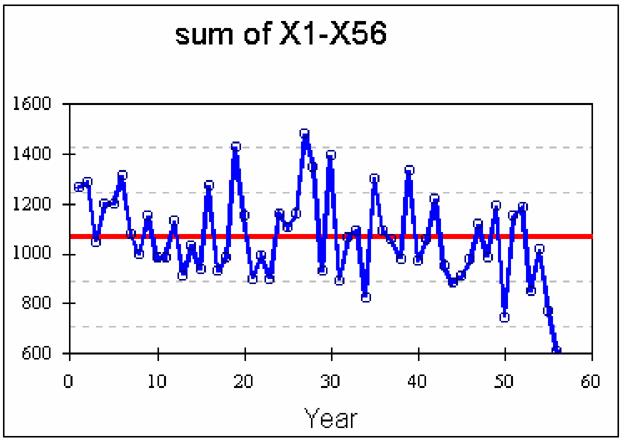


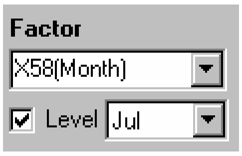
Fig. 5.3d Annual totals for a longer record

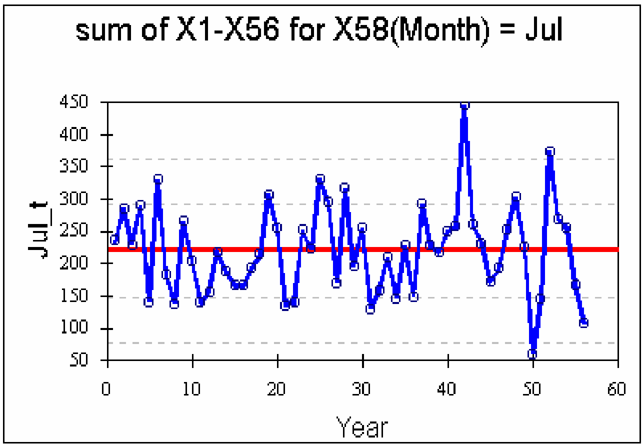


When long records are analysed, this type of graph gives the first indication of any trend or other pattern in the totals. An example is shown in Fig. 5.3d. This is the same analysis as above, but for a longer record of 56 years (1928 to 1983 in samaru56.wor) from the same station. In this graph the fainter lines are at 1 and 2 standard deviations about the overall mean of 1068.1mm. This graph indicates perhaps a tendency towards a lower annual total in the later years.

At the end of Section 5.2 a monthly (factor) column was produced. One use of this column is in Fig. 5.3e. This shows part of the same dialogue as before, but where just the data for July are analysed. The graph is shown in Fig. 5.3e with the longer record of 56 years.

Fig. 5.3e Summary to show July





This section has shown how to provide single summary columns, from the daily data. The graphs in Fig. 5.3c and e have also started the process of looking at the summary values.

5.4 Monthly summaries

The last section showed how to produce annual and individual monthly totals. Now the results are given for all the months. The same dialogue is used, completed as shown in Fig. 5.4a. The differences from Section 5.3 are as follows:

Use the factor column for the months, but without specifying a particular level. This gives the results for all the levels, i.e. all the months in this case.

The summary columns can be provided in 2 different ways. These are the options in Fig. 5.4a and 5.4b of layout by **Year** or by **Month**.

Fig. 5.4a Summary for months – by Year

`Climatic ⇒ Summary`

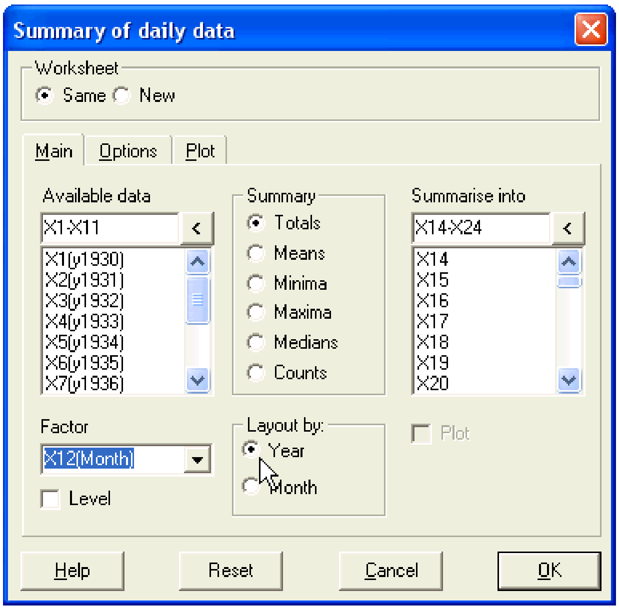
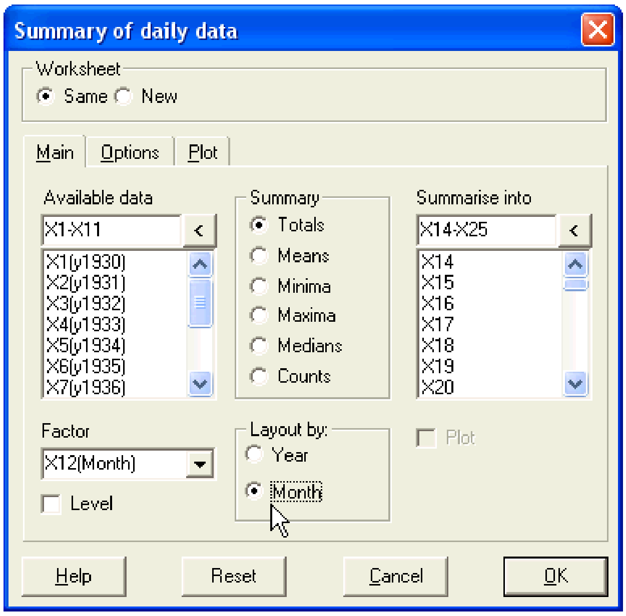


Fig. 5.4b And by Month



The layout of the monthly totals in Fig. 5.4c is by year. Thus each column is of length 12 and gives the totals for a given year. X13 gives the totals for 1930, X14 for 1931 and so on. This layout is used in further analyses where monthly values, rather than daily data are required.

Fig. 5.4c Monthly totals for the successive years

	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23	X24
1	0	0	0	0.25	0	0	0	0	0	0	0
2	0	0	10.67	0	0	0	0.76	0	0	6.1	0
3	0	0	0	0	0	8.89	0.76	0	0.76	14.99	0
4	33.79	62.99	64.77	24.64	73.4	9.4	64.51	21.85	67.55	12.19	26.67
5	115.56	148.85	226.57	99.31	180.33	106.68	151.37	110.73	135.37	159.76	73.65
6	89.67	210.29	252.23	233.15	141.49	232.16	111.24	128.28	179.57	133.84	139.95
7	228.35	289.8	138.69	329.42	181.1	136.65	265.68	203.71	138.69	156.46	216.92
8	417.32	243.59	224.05	351.01	339.35	354.58	208.02	277.37	239.55	288.28	210.82
9	158.01	236.21	223.01	270.78	106.39	145.53	315.21	240.02	213.63	265.7	203.2
10	2.03	4.57	58.17	0	54.61	2.54	28.71	1.02	11.94	91.95	39.61
11	0	1.52	0	3.05	0	0	0	0	0	0	0
12	0	0	0	0	0	0	4.06	0	0	0	0

The alternative layout is shown in Fig. 5.4d. In this case the summaries have been 'transposed' so each column now contains the summary for a given period, in this case for a given month. So, in Fig. 5.4d**,** X25 gives the totals for January. Each column is of length 11, because the record has just 11 years. This layout will be used for the further exploration and processing of the data in Chapter 7.

Fig. 5.4d Totals by month.

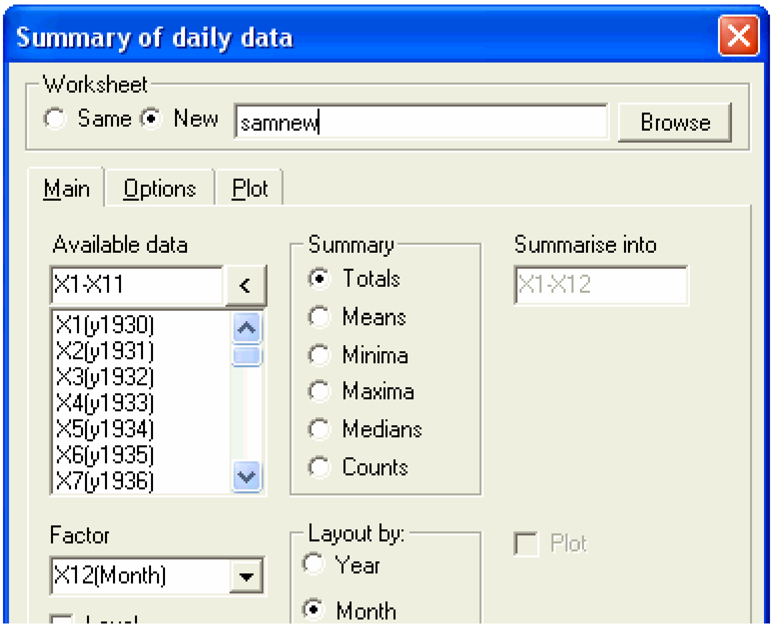
	X25	X26	X27	X28	X29	X30	X31	X32	X33	X34	X35	X36
	Jan_t	Feb_t	Mar_t	Apr_t	May_t	Jun_t	Jul_t	Aug_t	Sep_t	Oct_t	Nov_t	Dec_t
1	0	0	0	33.79	115.56	89.67	228.35	417.32	158.01	2.03	0	0
2	0	0	0	62.99	148.85	210.29	289.8	243.59	236.21	4.57	1.52	0
3	0	10.67	0	64.77	226.57	252.23	138.69	224.05	223.01	58.17	0	0
4	0.25	0	0	24.64	99.31	233.15	329.42	351.01	270.78	0	3.05	0
5	0	0	0	73.4	180.33	141.49	181.1	339.35	106.39	54.61	0	0
6	0	0	8.89	9.4	106.68	232.16	136.65	354.58	145.53	2.54	0	0
7	0	0.76	0.76	64.51	151.37	111.24	265.68	208.02	315.21	28.71	0	4.06
8	0	0	0	21.85	110.73	128.28	203.71	277.37	240.02	1.02	0	0
9	0	0	0.76	67.55	135.37	179.57	138.69	239.55	213.63	11.94	0	0
10	0	6.1	14.99	12.19	159.76	133.84	156.46	288.28	265.7	91.95	0	0
11	0	0	0	26.67	73.65	139.95	216.92	210.82	203.2	39.61	0	0

So far the monthly totals have been saved in the same Instat worksheet that holds the daily data. It is sometimes convenient for the summary values to be stored in a new worksheet. This step is essential if the daily record is very long, because the current version of Instat has an upper limit of 127 columns.

Use the same dialogue as before, but with the option to write the summary values to a new worksheet. To activate this option, first enter the factor to be used for the summaries, as shown in Fig. 5.4a. Then, as shown in Fig. 5.4e**,** choose the name of a new worksheet. The Browse option on the dialogue is to help **avoid** an existing worksheet, rather than to choose one!

Fig. 5.4e Putting the summaries into a new worksheet

`Climatic ⇒ Summary`



5.5 More summaries

Instat is designed to give as much flexibility as possible, so the analysis can be tailored to the user's requirements. As an example,consider the summary of temperature data from Niamey, Niger to review the ideas introduced in this chapter. The worksheet is called Ntemp.wor and will be used again in Chapter 8, which is on ways of processing temperature data.

Section 5.2 covered the preliminary steps, as follows:

a) After opening the worksheet, check that the data columns are locked. Here they are X3 to X22 for the daily maximum temperatures from 1961 to 1980, and X23-X42 for the minimum values. See Fig. 5.5a.

b) Remove any working columns that may remain from a previous analysis. Either use the dialogue shown in Fig. 5.5b or type **REMove X44-X100** into the 'Commands and Output' window.

c) Decide on the level that the data are to be summarised. Fig. 5.5c is by dekad.

Fig. 5.5a Lock data columns

File ⇒ Open ⇒ Ntemp.wor
Manage ⇒ Column Properties ⇒ Lock/Unlock

Fig. 5.5b Remove unwanted columns

Manage ⇒ Data ⇒ Remove(Clear)

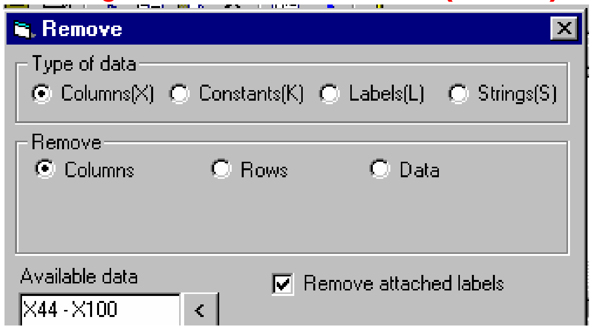


Fig. 5.5c Make a dekad factor column

Climatic ⇒ Manage ⇒ Make Factor

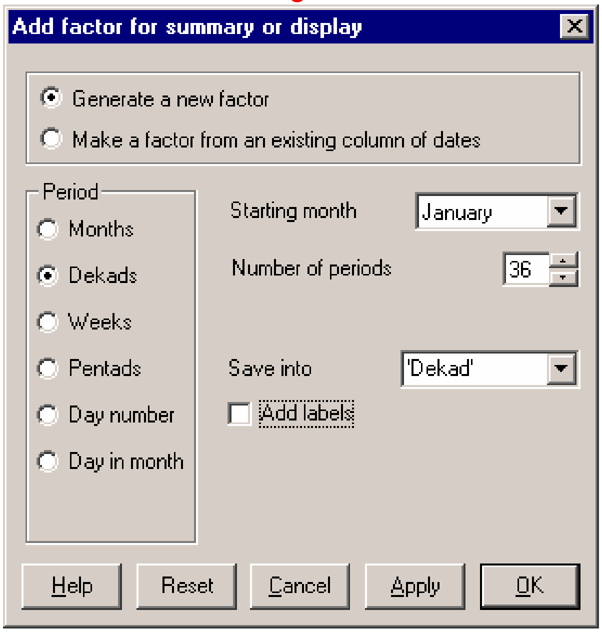
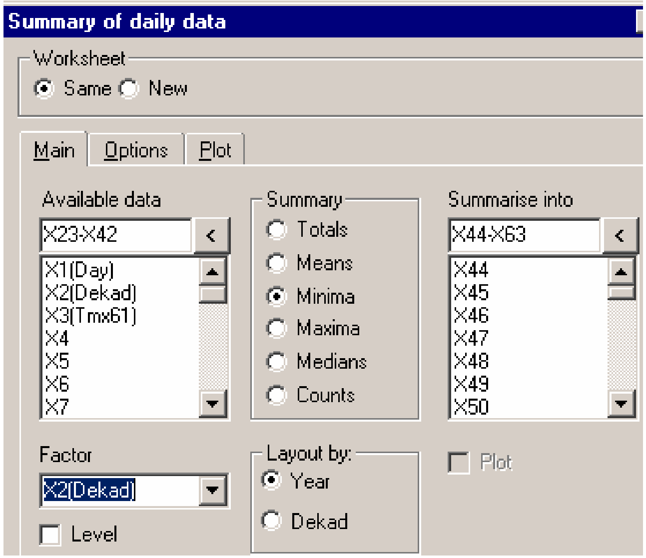


Fig. 5.5d Summarise the data

Climatic ⇒ Summary



Then use the **Climatic => Summary** dialogue as described in Section 5.4. Part of this dialogue is in Fig. 5.5d. At this stage you must decide:

a) Which columns of daily data to process. The example is for the minimum temperatures

b) Which summary to give. The example is of the dekad minima.

c) Whether the resulting columns should be by year, or by dekad. This depends on what you plan to do later in the analysis. Sometimes both layouts are needed for different analyses as shown in Fig. 5.4b and Fig. 5.4c. Then use the **Climatic => Summary** dialogue twice as shown in Section 5.4.

d) Whether to put the summary columns into the same worksheet, or into a new worksheet.

5.6 Satisfying the user

The Windows-style dialogues make it very easy to use software such as Instat. One possible problem is that a user may become helpless if a product is needed that cannot be given with the dialogues.

A simple example is used, but the strategy of how to cope with non-standard requests is more important than the particular example.

The example is to give the rainfall totals for each of the 4 seasons, Jan-March, April-June, July-Sept and Oct-Dec. This may be of interest to compare actual values with 3-month predictions from seasonal forecasts. Here, however, the main point is that the 'route' suggested in this chapter needs modification. This is because the **Climatic ⇒ Manage ⇒ Make Factor** dialogue, introduced in Section 5.2, supports a summary for Months, Dekads, Weeks or Pentades, but does not include the option of 'seasons'.

Typically, when the problem is not trivial there is more than one solution, and here three alternative approaches are considered.

Fig. 5.6a Total rainfall for July – Sept.

File ⇒ Open Worksheet ⇒ samsmall.wor
Manage ⇒ Manipulate ⇒ Row Statistics

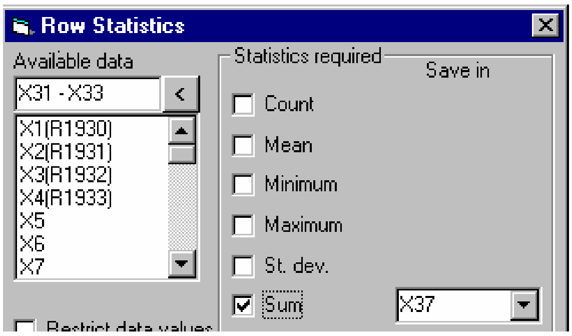


Fig. 5.6b Manage ⇒ Calculations

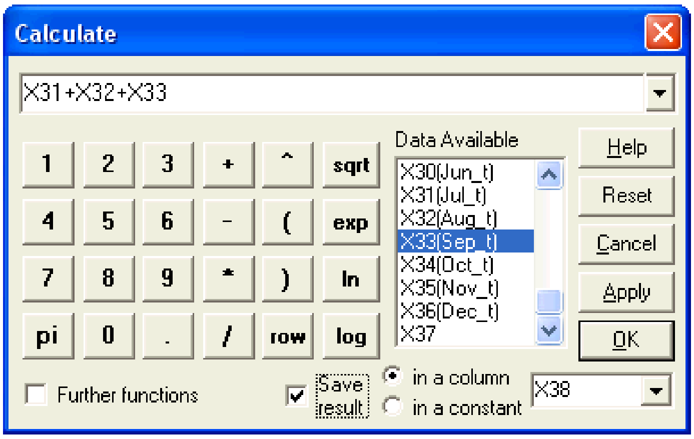


Fig. 5.6c Results

	X37	X38
1	803.68	803.68
2	769.6	769.6
3	585.75	585.75
4	951.21	951.21
5	626.84	626.84
6	636.76	636.76
7	788.91	788.91
8	721.1	721.1
9	591.87	591.87
10	710.44	710.44
11	630.94	630.94
12		

The first approach is to summarise the data over months and then to add the totals to give the seasonal values. The results were shown in Fig. 5.4a, where for example x31 to x33 gives the totals for July to September.

Then use either **Manage ⇒ Manipulate ⇒ Row Statistics** or **Manage ⇒ Calculations** as shown in Fig. 5.6a and 5.6b. The seasonal totals are stored in X36 and X37 respectively and are identical. Repeat for each season.

This approach is also useful if the user wishes to examine the more detailed monthly results en route to the calculation of the seasonal totals. However, it does require repeated use of the **Row Statistics** or **Calculations** dialogue and it was 'lucky' that the seasons were a whole number of months. If they were for 3-month periods from 15th June, etc, then the method could not be used.

The second approach is more general. The task of the **Climatic ⇒ Manage ⇒ Make Factor** dialogue is to make a column that can be used to split the data into (monthly) groups. This dialogue cannot be used, but all that is needed is to make a factor to specify the 4 seasons. Enter the numbers 1 to 4 into the next free column, as follows:

- 1 is for Jan to March. Including Feb 29 there are 31+29+31=91 days
- 2 is for April to June. There are 30+31+30=91 days

- 3 is for July to September. There are 31+31+30=92 days
- 4 is for October to December. There are 31+30+31=92 days

These could be typed directly into the worksheet, but a simpler way is to use the **Manage ⇒ Data ⇒ Enter** dialogue as shown in Fig. 5.6d. Name the column as **Season** and then define the column to be a factor as shown in Fig. 5.6e .

Fig. 5.6d Enter the seasonal code numbers

Manage ⇒ Data ⇒ Enter

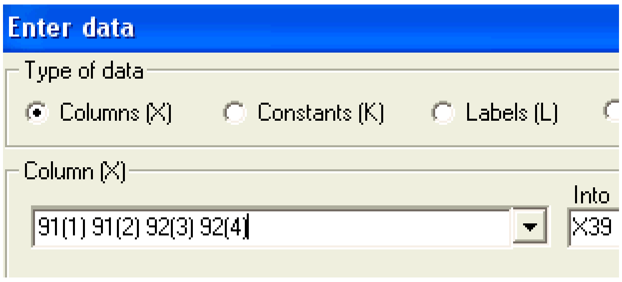
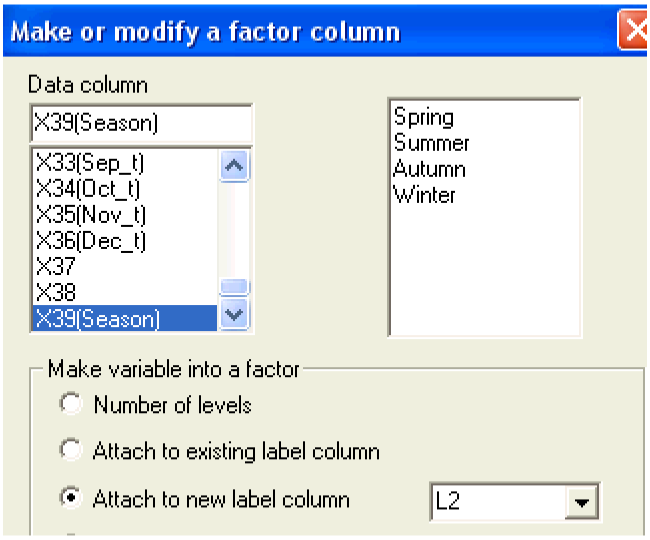


Fig. 5.6e Define a seasonal factor

Manage ⇒ Properties ⇒ Define Factor



Now summarise the data by season (Fig. 5.6f).

Fig. 5.6f Total rainfall by season

Climatic ⇒ Summary

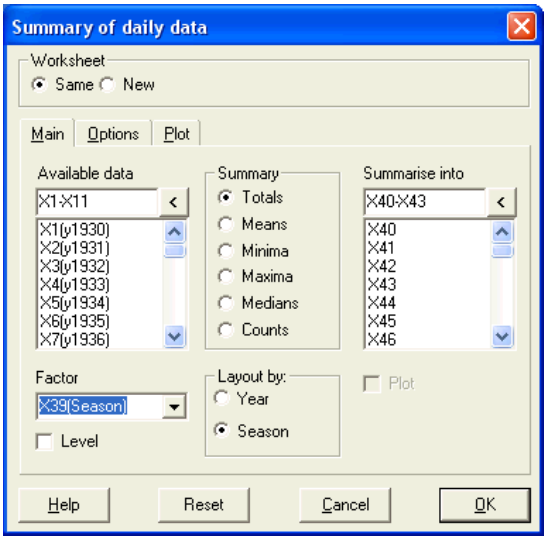


Fig. 5.6g Seasonal totals

	X40	X41	X42	X43
	Spring_t	Summer_t	Autumn_t	Winter_t
1	0	239.02	803.68	2.03
2	0	422.13	769.6	6.09
3	10.67	543.57	585.75	58.17
4	0.25	357.1	951.21	3.05
5	0	395.22	626.84	54.61
6	8.89	348.24	636.76	2.54
7	1.52	327.12	788.91	32.77
8	0	260.86	721.1	1.02
9	0.76	382.49	591.87	11.94
10	21.09	305.79	710.44	91.95
11	0	240.27	630.94	39.61
12				

This second method is flexible and can be used for periods of any number of days. It has required more knowledge of Instat, but still used the dialogues in a simple way.

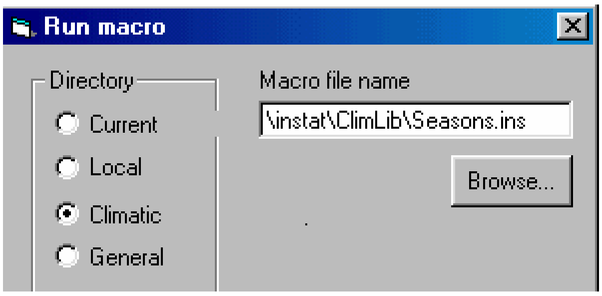
The final method builds on this second approach and is more general still. It is not needed to analyse the data from one station, but would be useful if the analysis is repeated for many sites. By automating the analysis, the steps illustrated in Fig. 5.6d to 5-6g are done more easily.

This third method is to write a macro to undertake one or more steps in the analysis. Writing a macro demands more knowledge of Instat, so here the macro, called ***Seasons.ins*** is supplied. The task is simply to use it.

Use **Submit ⇒ Run Macro** and complete the dialogue as shown in Fig. 5.6h. Use the 'Browse' button to find the macro called ***Seasons***. Then click **OK**. This macro asks what worksheet you would like to process. Then it opens the worksheet of your choice, produces the factor column and does the seasonal summary automatically. If you wish to see the Instat commands in the macro then use **Edit ⇒ View / Edit Macro ⇒ Open** dialogue and complete it in the same way as shown in Fig. 5.6h.)

Fig. 5.6h Run a macro to give a seasonal summary of the daily data

Submit ⇒ Run Macro



The general message from this section is as follows. The dialogues are used in a simple way for many summaries of the climatic data. Users can then undertake these analyses without a deep knowledge of the software package, in this case Instat. Sometimes the requirement is for a summary that cannot be handled in such a simple way. Then:

- More knowledge of the software is needed.
- There may be different approaches. Then each approach is likely to have advantages and limitations.

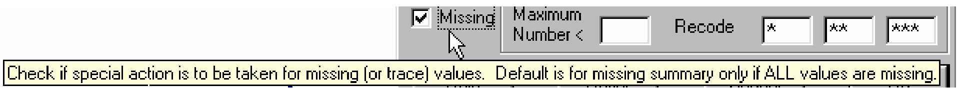
5.7 Coping with missing values

The daily data may include missing values. Their input was considered in Chapter 3. In addition Instat has two other special codes, one that is conventionally used for trace rainfalls, and the other to signify February 29th in non-leap years.

Most of Instat's dialogues and commands handle missing values in a sensible way. In the summary stage for daily records, the user can tailor the standard way of coping with missing data.

The dialogue for the summary of the daily data, for example Fig. 5.3a, includes a checkbox labelled Missing. Fig. 5.7a shows this part of the dialogue, and the different options are described in this section.

Fig. 5.7a Climatic ⇒ Summary dialogue showing the options for missing values



As shown in Fig. 5.7a Instat's three 'missing' value codes are as follows:

*	= missing data
**	= trace
***	= February 29th

For simplicity, consider a single month of rainfall data, for which decade totals are to be given. The data are in Fig. 5.7b. The data are in a worksheet called **climmiss.wor**.

First consider the default use of the **Summary** dialogue for the rainfall totals, Fig. 5.7c.

The results in Fig. 5.7d, show that the number of missing values (given in brackets), only refers to the first missing value code (*). Using the dialogue again, but requesting the Count, instead of the Sum, gives the number of days used in the calculation. This is just the non-missing values, where all the codes (*, ** and ***) have been ignored. The result is in Fig. 5.7e.

Fig. 5.7b Rainfall data (climmiss.wor) to illustrate options for missing values

Row	Decade X1	Rain X2	*Decade*	*No.*	*No.non-missing*	*	**	***
1	1	0						
2	1	0						
3	1	0						
4	1	0						
5	1	0						
6	1	**						
7	1	0						
8	1	0						
9	1	0						
10	1	0	*1*	*10*	*9*	*0*	*1*	*0*
11	2	0						
12	2	4.1						
13	2	0						
14	2	0						
15	2	*						
16	2	**						
17	2	0						
18	2	0						
19	2	0.5						
20	2	0	*2*	*10*	*8*	*1*	*1*	*0*
21	3	3.6						
22	3	2.7						
23	3	0.1						
24	3	*						
25	3	*						
26	3	*						
27	3	0						
28	3	0						
29	3	***	*3*	*9*	*5*	*3*	*0*	*1*

Fig. 5.7c The default Summary dialogue

Climatic ⇒ Summary

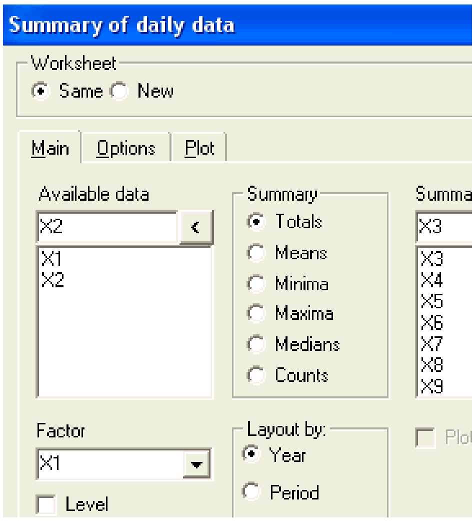


Fig. 5.7c Results

```
Summary of daily data

STAts X2;sum  X3;BY X1

      Sum  for each level of X1

Column          Sum
               X3

X2
  1              0
  2 (1)        4.6
  3 (3)        6.4

(n) - level has n missing value(s)
```

Typing the command directly into the Commands and Output Window can generate the results in Fig. 5-7d. It can be given as follows:

: **STAts X2; BY X1; SUM**

This is simpler, than the dialogue in that the results are not saved back to the worksheet.

The output in Fig. 5.7e is given with the command

: **STAts X2; BY X1; SUM; COUnt; MISs**

where the statistics for the count and the number of missing values have also been requested.

Fig. 5.7e Summary Statistics including ;COUnt and ;MISsing

: STAts X2; BY X1; SUM; COUnt; MISs

```
: STAts X2; BY X1; SUM; COUnt; MISs
Statistics for each level of X1

Column          Sum      Count     Missing


X2
  1               0          9           0
  2 (1)         4.6          8           1
  3 (3)         6.4          5           3

(n) - level has n missing value(s)
```

This shows that:

- The code * is treated as missing and the totals are calculated for the non-missing values. This is normally sensible if there are just one or two missing values. When there are many missing values, an alternative is to set the summary statistic (here the total) also to missing. This option is shown below.
- The codes ** and *** are ignored, i.e. treated as though the observation does not exist. This is correct for *** because February 29th does not exist in this year. It is not sensible for the trace values (coded as **). It would be better to recode them to zero, or a small value, before calculating the totals. It does no harm when the rainfall totals are calculated, because the total would still be the same, had ** been recoded to zero. The results would not have been the same if a different statistics had been requested, for example the mean.

One alternative uses the option in the dialogue shown in Fig. 5.7f.

Fig. 5.7f Summary where no missing values are allowed

The corresponding command and results are given in Fig. 5.7g. With this option, the summary is set to missing when there are any, i.e. one or more missing values.

Fig. 5.7g No missing values allowed

:STAts X2;BY X1; SUM; COU; NOMis 1

```
: stats x2;by x1; sum; count; nomiss 1
Statistics for each level of X1

Column          Sum      Count


X2
  1               0          9
  2 (1)           *          8
  3 (3)           *          5

(n) - level has n missing value(s)
```

Fig. 5.7h Specifying the number of missing values allowed

:STAts X2;BY X1; SUM; COU; NOMis 3

```
: stats x2;by x1; sum; count; nomiss 3
Statistics for each level of X1

Column          Sum      Count


X2
  1               0          9
  2 (1)         4.6          8
  3 (3)           *          5

(n) - level has n missing value(s)
```

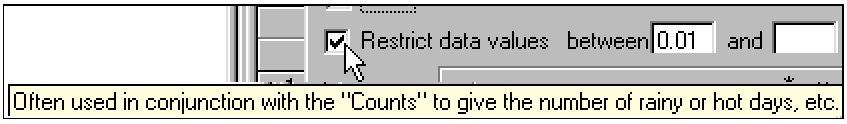
If this is too extreme, then replacing the 1, in Fig. 5.7f, by 3, or giving the command:

: **STAts X2; BY X1; SUM; COUnt ; NOMiss 3**

will give the summary statistic if there are 2 or less missing values, but set it to missing if there are 3 or more. The results are in Fig. 5.7h.

The trace values (**) become important when calculating the number of rainy days in each decade. In the **Summary** dialogue, shown in Fig. 5.7i specify the counts and also check the restrict box. Given a lower limit of 0.01, or a higher value, if small amounts are to be ignored.

Fig. 5.7i Calculating the number of rain days



The results are the same as from the command:

: **STAts X2; BY X1; COUnt ; REStrict 0.01**

and are shown in Fig. 5.7j.

Fig. 5.7j Count of rainy days, when trace values are ignored

```
: stats x2; by x1; count; restrict 0.01

    Count  for each level of X1

Column        Count


X2
  1              0
  2 (1)          2
  3 (3)          3

(n) - level has n missing value(s)
```

This default of ignoring trace values may not be appropriate, as a small amount of rain has fallen. If so, then one possibility is to recode the trace values within the summary dialogue, as shown in Fig. 5.7k.

Fig. 5.7k Temporary recoding of trace values

This recodes the data within the calculation, but leaves the data column unaltered.

The other option is to recode the trace values once and for all. This uses the Recode option in the general dialogue to transform data. It is given by **Manage ⇒ Manipulate ⇒ Recode**. Then complete the dialogue as shown in Fig. 5.7l.

Fig. 5.7l Recoding trace values

`Manage ⇒ Manipulate ⇒ Recode`

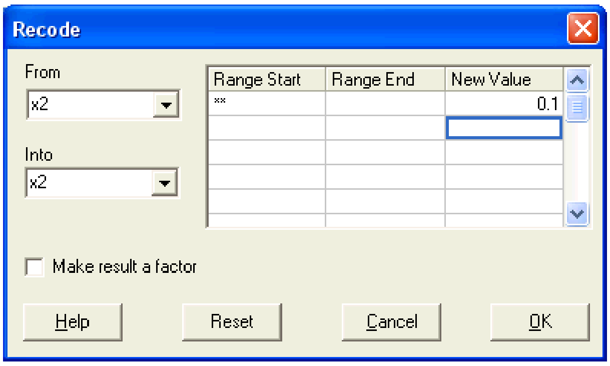


Fig. 5.7m Results following recoding trace to 0.1

```
: stats x2;by x1; sum; count
Statistics for each level of X1

Column          Sum      Count


X2
  1             0.1         10
  2 (1)         4.7          9
  3 (3)         6.4          5

(n) - level has n missing value(s)
: stats x2;by x1; sum; count; restrict 0.01
Statistics for each level of X1

Column          Sum      Count


X2
  1             0.1          1
  2 (1)         4.7          3
  3 (3)         6.4          3

(n) - level has n missing value(s)
```

In Fig. 5.7l the recoded data are saved back into the same column. If WARnings are set to ON, you will be asked if you wish to overwrite the column. If you do not wish to be asked, then change the state of warnings using the key **F6**, or by **Edit ⇒ Options ⇒ Disable Warnings**. The current status of the warnings is in the bottom right hand corner of the Instat screen. It is usually appropriate to take the option in Fig. 5.7l, of making a once-and-for-all decision on how to deal with trace values. The results, following this step, are in Fig. 5.7m.

5.8 Conclusions

This chapter has described the production of summary values from the daily data.

A monthly and a ten-day analysis have been shown. However, one of the principles in Instat has been to give the user as much flexibility as possible. It is equally straightforward to provide 7 or 5 day summaries and Section 5.6 showed how 'seasonal' summaries can be found, where the user defines each season.

Most of this chapter considered rainfall totals, but Section 5.5 looked at minimum temperatures to emphasise that other statistics, such as mean, maximum, or the proportion greater than a specified threshold can also be studied.

The *events* in this chapter, such as annual or monthly rainfall totals, will be familiar to most users. What is perhaps new is the ease with which these summaries can be produced from the daily data. Hence it makes sense to start with the full records, rather than to have separate files with different summaries. This idea is extended in the next chapter, where other summary *events*, such as the date of the start of the season are considered. Then, in Chapter 7 shows how to process both the types of event derived in this chapter and those produced in Chapter 6.

Chapter 6 – Studying Climatic 'Events'

6.1 Introduction

The general theme of this chapter is that access to the daily records enables users to conduct many analyses that are not possible with just the 10-day or monthly totals.

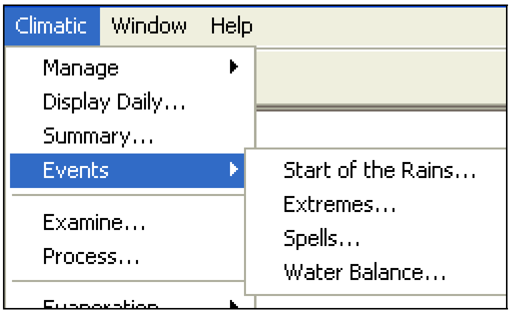
The word ***event*** is in the chapter title. An ***event*** here is just a characteristic of interest for which there is a single observed value each year. Some examples are as follows:

- **The total rainfall in January** is an event. The total rainfall of 74 mm in January 1984 is an occurrence of the event.
- **The 10-day rainfall totals** through the year are 36 events.
- **The number of rainy days in the year**, with rain defined as all non-zero values, is an event.
- **The length of dry spells** in May would not be considered as an event because there can be more than one dry spell in May in any given year. The length of the longest spell in May is an event.
- **The date of the start of the rains** for the first definition, considered in this chapter, is an event. The starting date, 17th April in 1930, is an occurrence of the event.

Given the events of interest, a simple analysis of climatic events is therefore in two stages. The first stage is to calculate the value taken by each event in each year. If there are 11 years of data, then this first stage produces columns of length 11. This first stage was considered for events such as rainfall totals in the last chapter, and here events such as the start and length of the season are found.

The dialogues used in this chapter are mainly under the **Climatic ⇒ Events** menu shown in Fig. 6.1a.

Fig. 6.1a The Events submenu



The main message from this chapter is that access to detailed (usually daily) records allows climatic data to be summarised in a wide variety of ways that can be tailored to particular user's requirements. Thus Meteorological Services, and others who process climatic data are therefore able to provide special analyses that correspond to any particular needs. This complements the provision of regular reports that are intended to be of general interest.

Section 6.2 considers the **Start of the Rains** dialogue for a variety of definitions of the beginning of the season. This is continued in Section 6.4 combined with the consideration of dry spells that are introduced in Section 6.3. Then Section 6.5 introduces a simple water-balance calculation. One use of this simple model is to provide a definition of the end of the season. Extremes are considered in Section 6.6.

All the definitions of events described in Sections 6.3 to 6.6 are easy to produce, because they correspond to options of the **Climatic ⇒ Events** dialogues. It is important that the dialogues do not limit users. Section 6.7 describes further ways of responding to users needs, even if the requirements do not correspond to the use of a single dialogue.

In this Chapter only the analysis of rainfall data is described. However the analyses are not limited in this way. For example dry spells can equally become hot spells or windy spells. Also the dialogues do not 'know' that the data are daily. The limits are therefore more in the user's imagination.

This chapter is still concerned with the first stage in the two-step process. The processing described here produces columns that are of length equal to the number of years of data. The second stage is to process these events and that is described in Chapter 7.

6.2 The start of the season

Users should still follow the preliminary steps described in Section 5.2. This ensures that the data columns are locked, and that unnecessary working columns are removed. Many of the examples below use the file called **samsmall.wor** and usually start from the **Manage ⇒ Data ⇒ Clear (Remove)** dialogue, or by typing **remove x12-x60** in the 'Commands and Output' window.

Consider two definitions of the start of the rains. For the first, define the start to be:

- the first occasion with more than 20mm in 1 or 2 days after 1st April

This is equivalent to the first time after *1st April* that the running *2-day* total is more than *20mm*.

This is an example of a type of definition. The user may vary any of the 3 criteria that are given in *italics*. For illustration, change just one of them and consider an alternative definition, namely

- the first occasion with more than 20mm in 1 or 2 days after 1st May

The 11 years of data from Samaru, Nigeria are again used for illustration. The analysis is straightforward from the daily data. The data for the first of the years, 1930, are given in Fig. 6.2a. Here 17th April is seen to be the first occasion with more than 20 mm after 1st April, while 12th May is the first occasion after 1st May.

Fig. 6.2a Data for 1930 from Samaru, with dates of the start of the rains marked

File ⇒ Open Worksheet ⇒ samsmall.wor
Manage ⇒ Data ⇒ Clear(Remove) and remove **X12-X60**
Climatic ⇒ Display Daily and display **X1**

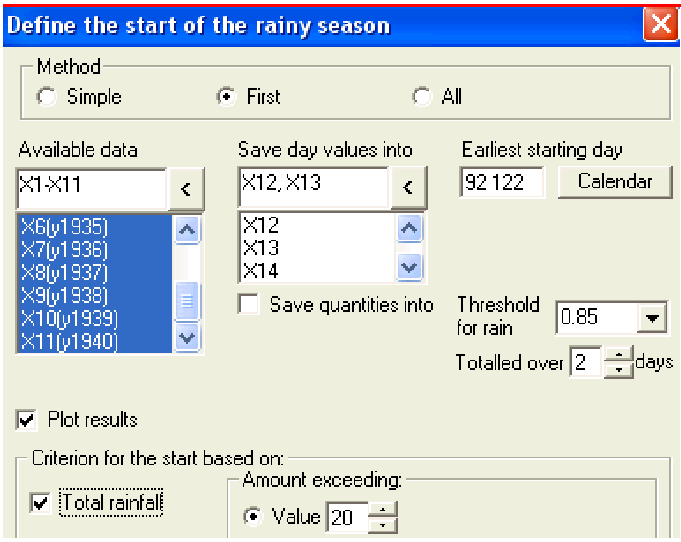
Daily data for: y1930

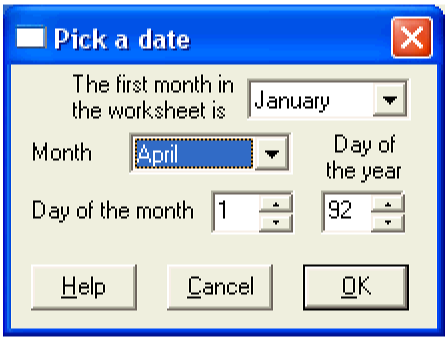
Mon Day.	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	--	--	--	--	--	--	--	1.0	14.5	--	--	--
2	--	--	--	--	--	--	--	3.3	26.7	--	--	--
3	--	--	--	--	--	15.2	--	7.9	1.8	--	--	--
4	--	--	--	--	--	--	2.3	3.3	--	--	--	--
5	--	--	--	--	--	6.1	--	3.1	8.6	--	--	--
6	--	--	--	--	--	3.8	8.4	12.5	9.9	--	--	--
7	--	--	--	--	--	--	--	2.3	9.9	--	--	--
8	--	--	--	--	10.4	--	3.1	37.6	4.3	--	--	--
9	--	--	--	--	--	9.4	--	--	10.4	--	--	--
10	--	--	--	5.1	--	--	--	8.6	2.3	--	--	--
11	--	--	--	--	1.8	0.5	12.2	49.0	3.6	0.3	--	--
12	--	--	--	--	23.1	--	8.6	49.8	5.8	--	--	--
13	--	--	--	0.5	--	11.4	10.9	1.5	--	--	--	--
14	--	--	--	--	33.0	--	2.0	2.5	--	--	--	--
15	--	--	--	--	0.8	18.0	2.8	6.6	17.5	--	--	--
16	--	--	--	--	--	8.1	14.5	10.9	--	--	--	--
17	--	--	--	22.1	12.5	--	6.9	33.3	0.5	--	--	--
18	--	--	--	--	--	--	20.3	7.6	--	--	--	--

Part of the **Climatic ⇒ Events ⇒ Start of the Rains** dialogue is shown in Fig. 6.2b, corresponding to the 2 definitions given above. This dialogue involves the use of a calendar to pick the earliest starting date as is shown in Fig. 6.2c, or type the day numbers into Fig. 6.2b.

Fig. 6.2b Start of the rains dialogue **Fig. 6.2c Calendar**

Climatic ⇒ Events ⇒ Start of the Rains





The results are in Fig. 6.2d and 6.2e in the worksheet, where the first year corresponds to the data in Fig. 6.2a i.e. April 17th and May 12th. By default the results are displayed as day numbers, as in Fig. 6.2d but the **Manage ⇒ Column Properties ⇒ Format** dialogue may be used to display dates.

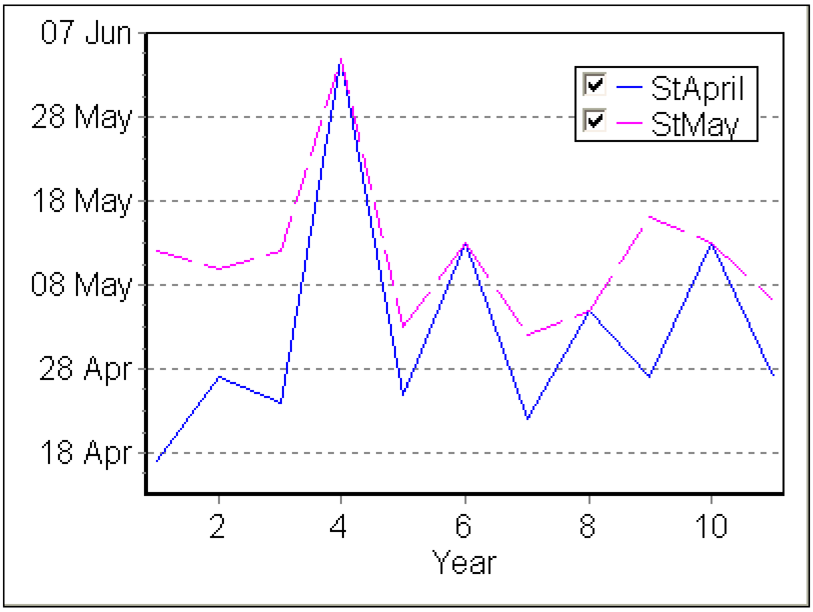
Fig. 6.2d Results for the 11 years **Fig. 6.2e 'Day of year' format**

	StApril	StMay
1	108	133
2	118	131
3	115	133
4	156	156
5	116	124
6	134	134
7	113	123
8	126	126
9	118	137
10	134	134
11	118	127

	X12	X13
	StApril	StMay
1	17 Apr	12 May
2	27 Apr	10 May
3	24 Apr	12 May
4	04 Jun	04 Jun
5	25 Apr	03 May
6	13 May	13 May
7	22 Apr	02 May
8	05 May	05 May
9	27 Apr	16 May
10	13 May	13 May
11	27 Apr	06 May

In completing the dialogue in Fig. 6.2b the option was ticked to graph the results and these are shown in Fig. 6.2f. The first definition gives an earlier starting date in 7 of the 11 years.

Fig. 6.2f Graph of start of the rains for April and May



It is useful to be familiar with the translation of day numbers and dates. To see the correspondence, use the **Manage ⇒ Data ⇒ Regular Sequence**, as shown in Fig. 6.2g to enter a column containing the numbers 1 to 366. Then the **Climatic ⇒ Display Daily** dialogue completed as shown in Fig. 6.2h gives the day numbers in the format shown in Fig. 6.2i. They are also given in Appendix 1.

Fig. 6.2g Entry of values 1 to 366

Manage ⇒ Data ⇒ Regular Sequence

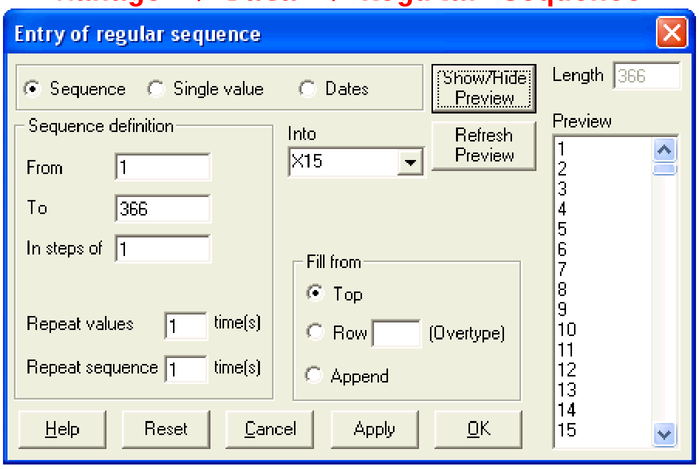
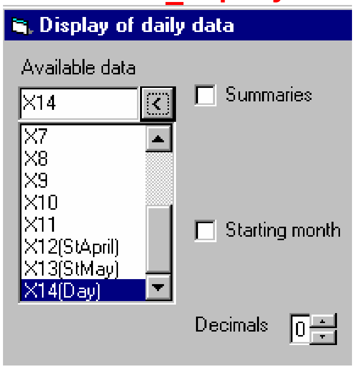


Fig. 6.2h Display of days in the year

Climatic ⇒ Display Daily



Comparison with Fig. 6.2a or ***6-2d*** confirms the starts for 1930 on day 108 (17th April) and 133 (12th May) for the 2 definitions.

Fig. 6.2i Day numbers (starting from January 1st) for the days of the year

Mon Day.	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	1	32	61	92	122	153	183	214	245	275	306	336
2	2	33	62	93	123	154	184	215	246	276	307	337
3	3	34	63	94	124	155	185	216	247	277	308	338
4	4	35	64	95	125	156	186	217	248	278	309	339
5	5	36	65	96	126	157	187	218	249	279	310	340
6	6	37	66	97	127	158	188	219	250	280	311	341
7	7	38	67	98	128	159	189	220	251	281	312	342
8	8	39	68	99	129	160	190	221	252	282	313	343
9	9	40	69	100	130	161	191	222	253	283	314	344
10	10	41	70	101	131	162	192	223	254	284	315	345
11	11	42	71	102	132	163	193	224	255	285	316	346
12	12	43	72	103	133	164	194	225	256	286	317	347
29	29	60	89	120	150	181	211	242	273	303	334	364
30	30		90	121	151	182	212	243	274	304	335	365
31	31		91		152		213	244		305		366

In training workshops we introduce these analyses by looking at these 11 years of data 'by hand'. The **Climatic ⇒ Display Daily** dialogue is used to display the daily data from X1-X11, as shown in Fig. 6.2a. Then it takes just a few minutes for the dates to be found for the alternative definitions. After the 'hand' analysis we show the use of the **Climatic ⇒ Events ⇒ Start of the Rains** dialogue.

Two different definitions of the start of the season are shown in Figs 6.2j and 6.2k.

Fig. 6.2j Definition involving evaporation

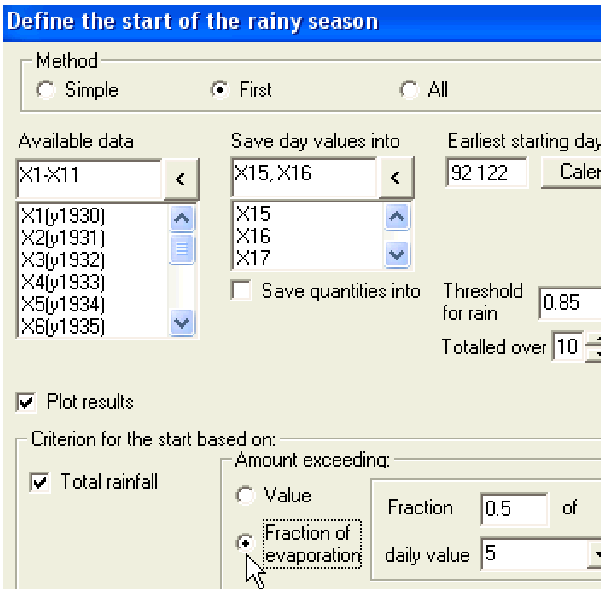
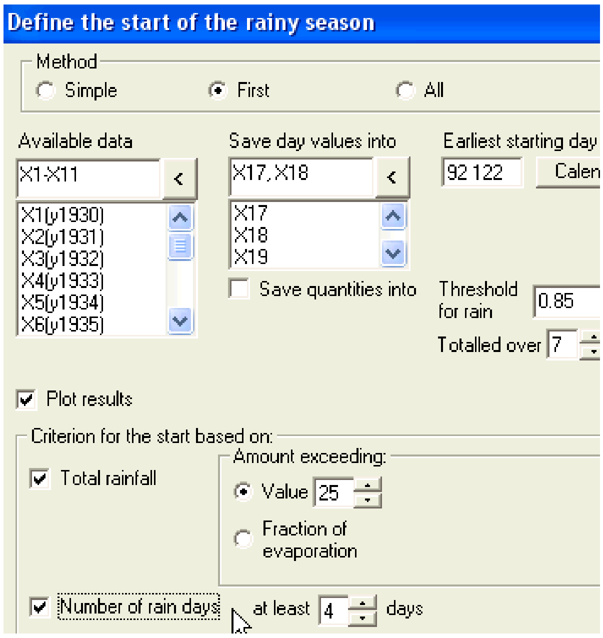


Fig. 6.2k With a minimum no. of rain days



The first defines the start as:

- the first occasion after ***1st April*** that the ***10-day*** *total* exceeds ***half*** the evaporation.

This type of definition has been used in the past, based on rainfall data that has already been totalled before being compared to the evaporation. Now, with access to the daily data, the running 10-day totals are used instead.

Daily evaporation data are needed for this definition. For simplicity here evaporation is taken to be a fixed 5 mm per day. In practice the evaporation should be estimated (e.g. using Penman/Monteith see Chapter 9).

The next definition adds a further condition, that there must be a certain number of rainy days in the period, as well as the required total. The start is defined as:

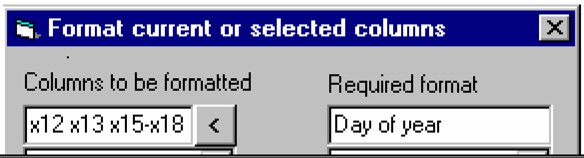
- the first occasion the ***7-day*** *total* exceeds ***25mm*** and includes at least ***4 rainy days***.

This definition is similar to one used by Raman (1974) in a study of the start of the rains in India.

The results for all six definitions are shown in Fig. 6.2l. In some years the dates are similar for most definitions, i.e. a large rainfall starts the season for any of them.

Fig. 6.2l Results from the 'Start' definitions shown above

Manage ⇒ View Data ⇒ Format



	X12	X13	X15	X16	X17	X18
	StApril	StMay				
1	17 Apr	12 May	17 Apr	12 May	14 May	14 May
2	27 Apr	10 May	27 Apr	01 May	30 May	30 May
3	24 Apr	12 May	24 Apr	01 May	27 Apr	01 May
4	04 Jun	04 Jun	19 May	19 May	19 May	19 May
5	25 Apr	03 May	25 Apr	01 May	21 May	21 May
6	13 May	13 May	13 May	13 May	13 May	13 May
7	22 Apr	02 May	22 Apr	01 May	01 May	01 May
8	05 May	05 May	05 May	05 May	06 May	06 May
9	27 Apr	16 May	22 Apr	01 May	24 Apr	01 May
10	13 May	13 May	12 May	12 May	14 May	14 May
11	27 Apr	06 May	27 Apr	01 May	27 May	27 May

One more type of definition adds the condition that there should not be a long dry spell in the period immediately following the start. An example is as follows:

- the first occasion with more than 20mm in a 2 day period after 1st April **and no dry spell of 10 days or more within the following 30 days**

This includes the idea of dry spells, so is covered in Section 6.4, after spell lengths have been considered in general.

6.3 Spell lengths

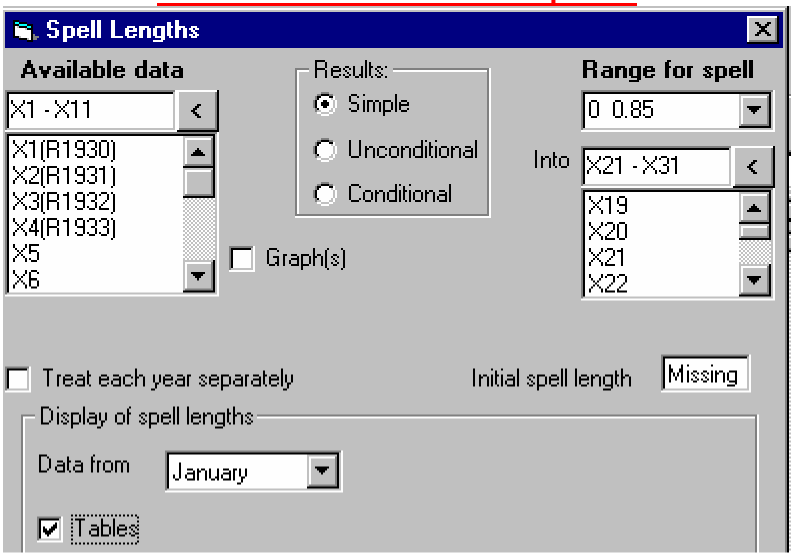
This section uses the **Spells** dialogue, the third under **Climatic ⇒ Events**. It may be applied to any variable, for example to examine hot spells in temperature data. Here dry spells in the rainfall records are considered.

A preliminary task, when looking at dry spells, is to define a dry day. The obvious definition is any day with zero rainfall. However, when comparing stations, one often finds that different observers are not equally conscientious in the recording of small rainfalls, or in the extent to which data are rounded. We usually use a value of just under 1mm and define a day to be dry if its value is less than this threshold. The value of 0.85mm avoids rounding problems in both mms and inches (it is between 0.03 and 0.04 inches). This threshold is used in all the examples considered here. For some applications a higher threshold such as 2.45 mm or even 4.95 mm, may be appropriate.

The 11 years of data from Samaru are again used to illustrate the methods. The simple use of the **Climatic ⇒ Events ⇒ Spells** dialogue was introduced in Chapter 4 and is shown again in Fig. 6.3a. Here the daily records for the 11 years are simply transformed into spell lengths.

Fig. 6.3a Spells dialogue to transform rainfall data into spell lengths

Climatic ⇒ Events ⇒ Spells



The results for 1931 are shown in Fig. 6.3b. Notice the assumption that the data follow from one year to the next. Hence, 1st January 1931 is the 69th consecutive dry day because the last rain occurred on 24th October 1930.

Fig. 6.3b Simple use of the Spells dialogue for 1931

Mon	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Day.												
1	69	100	128	159	3	--	2	--	1	3	29	7
2	70	101	129	160	4	--	--	1	2	4	30	8
3	71	102	130	161	5	--	--	--	--	--	31	9
4	72	103	131	162	6	1	--	1	--	1	32	10
5	73	104	132	163	7	--	1	2	--	2	33	11
6	74	105	133	164	8	--	2	3	--	3	34	12
7	75	106	134	165	--	1	3	4	--	4	35	13
8	76	107	135	--	1	2	--	--	--	5	36	14
9	77	108	136	--	2	--	1	--	--	6	37	15
10	78	109	137	1	--	1	--	--	--	7	38	16
28	96	127	155	--	--	1	--	--	--	25	4	34
29	97		156	1	1	--	1	--	1	26	5	35
30	98		157	2	--	1	--	--	2	27	6	36
31	99		158		1		--	--		28		37
Maximum										(Overall: 165)		
	99	127	158	165	8	5	3	4	3	28	51	37

In training workshops we start by producing these results for the 11 years and examine the longest dry spells in May, June and July by 'hand'. From Fig. 6.3b we see the results are 8, 5 and 3 days for 1931.

In Fig. 6.3c the Spells dialogue is used to find these maximum spell lengths directly. In the dialogue choose 'Unconditional' results and use the calendar to specify the periods that correspond to the months of May, June and July.

Fig. 6.3c Spells to find maximum spell lengths
Climatic ⇒ Events ⇒ Spells

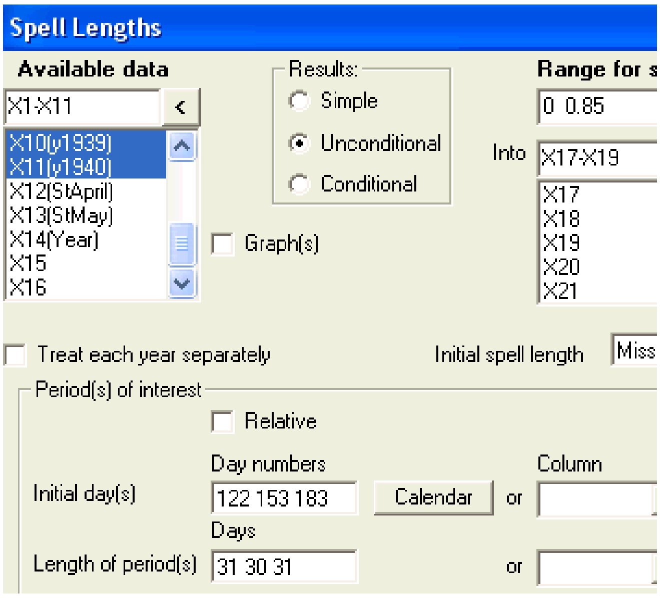


Fig. 6.3d Spell lengths for May, June and July

	X32	X33	X34
	Sp_May	Sp_Jun	Sp_Jul
1	13	7	5
2	8	5	3
3	8	3	13
4	14	4	6
5	8	3	3
6	9	3	4
7	6	6	4
8	24	6	4
9	13	4	7
10	12	3	4
11	9	5	7

The results are shown in Fig. 6.3d and the second row gives the same values, (8, 5 and 3 days), as the summary at the bottom of Fig. 6.3b.

The results indicate that long dry spells were sometimes found in May, but they were rare in June and July. Inferences from these data must, of course, be treated with caution because the record is so short.

6.4 Combining spells and the start of the rains

Sometimes it is useful to consider spell lengths for 'crop' periods, rather than fixed times in the year. An example is in Fig. 6.4a. Here the starting date each year is taken from X12. These are the dates produced in Section 6.2 using the criterion that the start is the

- first occasion after ***1st April*** with more than ***20mm*** in a ***1 or 2*** day period

This use of the SPEll command has given the dry spell lengths for 4 successive 30 day periods after sowing, i.e. the first period is days 1 to 30 of the crop, etc.

Fig. 6.4a Spells relative to starting dates

Climatic ⇒ Events ⇒ Spells

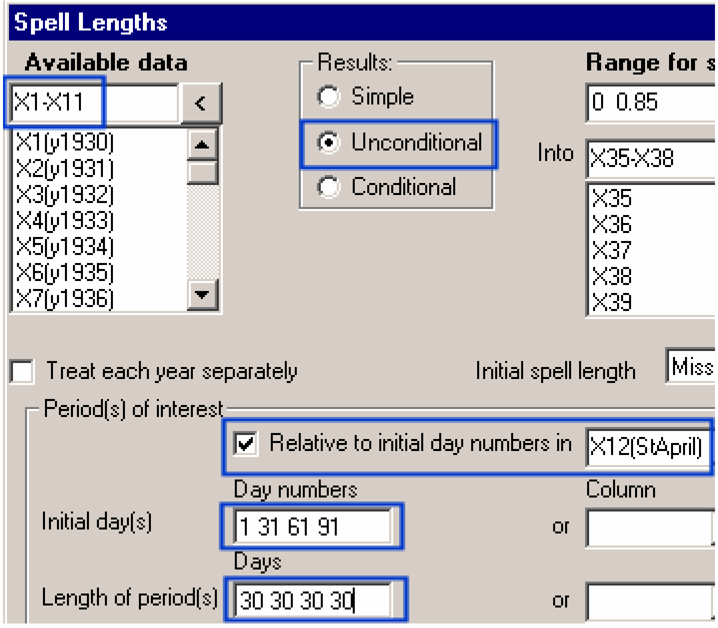


Fig. 6.4b Results

	X35	X36	X37	X38
1	13	4	7	2
2	8	5	3	4
3	8	5	13	4
4	4	6	4	5
5	8	3	3	2
6	7	4	2	2
7	6	6	4	4
8	7	6	4	2
9	13	4	4	7
10	4	3	4	5
11	9	5	7	3
12				

If a 120-day crop is being considered, the results in Fig. 6.4b indicate that the difficult time is the first 30 days after sowing. There are dry spells of 8 days or more in 6 of the 11 years. In the subsequent 90 days, there is only 1 year (1932) that has a dry spell of longer than 7 days. A long dry spell is needed for harvesting some crops and the data in the last column indicate that, in this case, there is rarely such a spell.

The results in Fig. 6.4b provide a useful reminder of the general role of the dialogues described in this chapter. The original daily data were in X1 to X11. The user specified three 'events' in the dialogue in Fig. 6.3c and four 'events' in Fig. 6.4a. With the 11 years of data, each event resulted in a column of length 11. These columns are now ready for further processing using the methods described in Chapters 7 and 11.

The next example introduces the conditional option in the Spells dialogue. The purpose is to estimate the risk of a long dry spell in the 30 days after sowing **for any potential sowing date**. Each event therefore gives the length of the longest spell in the 30 days assuming day 0 was a sowing date - and therefore had rain.

21 columns of data are produced, so some working columns may have to be removed first, to give enough space in the worksheet.

Fig. 6.4c Enter starting day numbers

Manage ⇒ Data ⇒ Remove x20-x60

Manage ⇒ Data ⇒ Regular Sequence

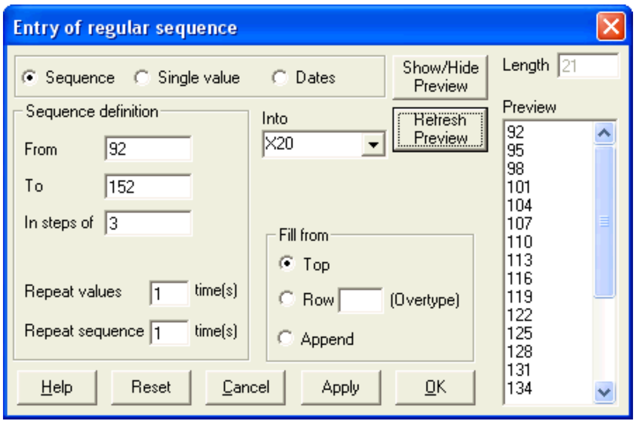
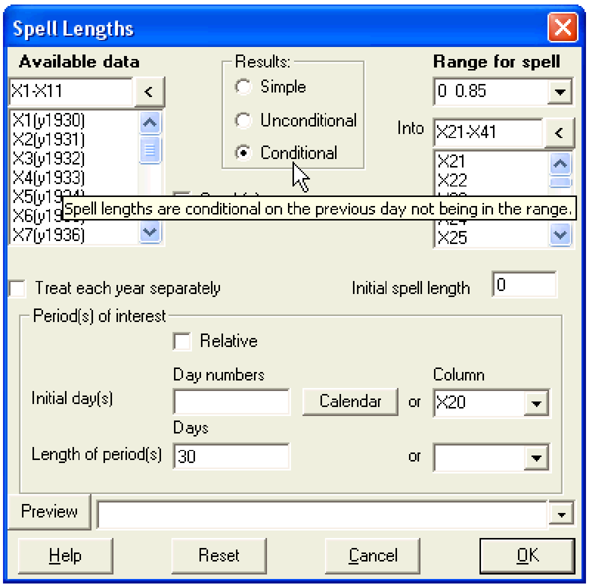


Fig. 6.4d Spells dialogue

Climatic ⇒ Events ⇒ Spells



The first step is to enter a column containing the starting dates of the periods. This is shown in Fig. 6.4c where day 92 is 1st April with successive starts at 3-day intervals. Then use the Spells dialogue as shown in Fig. 6.4d.

The results are shown in Fig. 6.4e and give the maximum spell lengths for a series of overlapping 30-day periods. The first is for April 1 to 30 and the remainder start after successive 3-day intervals. The last 30-day period is 31st May to 29th June, in X41.

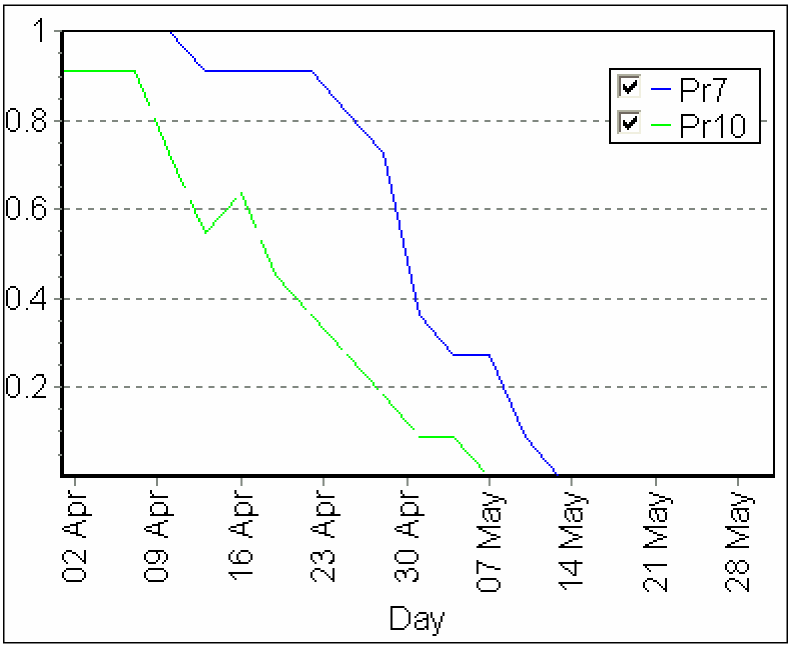
Fig. 6.4e Maximum spell lengths

	X21	X22	X23	X24	X25	X26	X36	X37	X38	X39	X40	X41
1	9	9	12	13	13	13	4	4	4	4	6	7
2	12	12	12	12	9	8	5	3	4	4	5	5
3	22	19	16	13	10	8	5	5	5	5	3	3
4	17	17	15	14	14	14	4	4	4	4	4	4
5	14	11	8	8	8	8	3	3	3	3	3	3
6	26	23	20	17	14	11	7	7	7	6	3	3
7	17	14	11	8	5	6	6	6	6	6	5	5
8	22	24	24	23	20	17	7	6	6	6	6	6
9	18	15	12	9	10	13	5	5	5	3	4	4
10	15	12	12	12	12	12	4	4	4	4	3	3
11	26	23	20	17	14	11	5	5	5	5	5	5
12												

In Fig. 6.4e the first columns correspond to sowing in early April, and show most years would have had a dry spell of more than 10 days within the following 30 days. In contrast, by the final columns the longest spell in any year is 7 days.

In **Chapter 7** these data are processed further, to produce the graph shown in Fig. 6.4f. As an example of the use of these results, suppose that in the next year, 30th April, i.e. day 121, is a potential sowing date. This corresponds roughly to X31 above. From the spell-length data or from the lower curve in Fig. 6.4f, the chance of a dry spell of more than 10 days within the next 30 days is seen to be only about 0.1 (i.e. 1 year in 11). About half the years are estimated to have a dry spell of more than 7 days within the next 30 days. Knowledge of these risks may help the user decide whether to sow and the planner to define a sowing strategy. These results must be treated with extreme caution because they are still merely a summary of just 11 years of data. They also make assumptions about the independence of the data in successive periods in the year. This latter problem is considered further in Chapter 13.

Fig. 6.4f Chance of dry spells



Now return to the **Start of the Rains** dialogue that was considered in Section 6.2. Consider a definition of the start that incorporates a dry spell condition. This definition is as follows:

- The first occasion with more than 20mm in a 2 day period after 1st April and **no dry spell of 10 days or more within the following 30 days**

Fig. 6.4g Start of the rains and dry spells

Climatic ⇒ Events ⇒ Start of the Rains

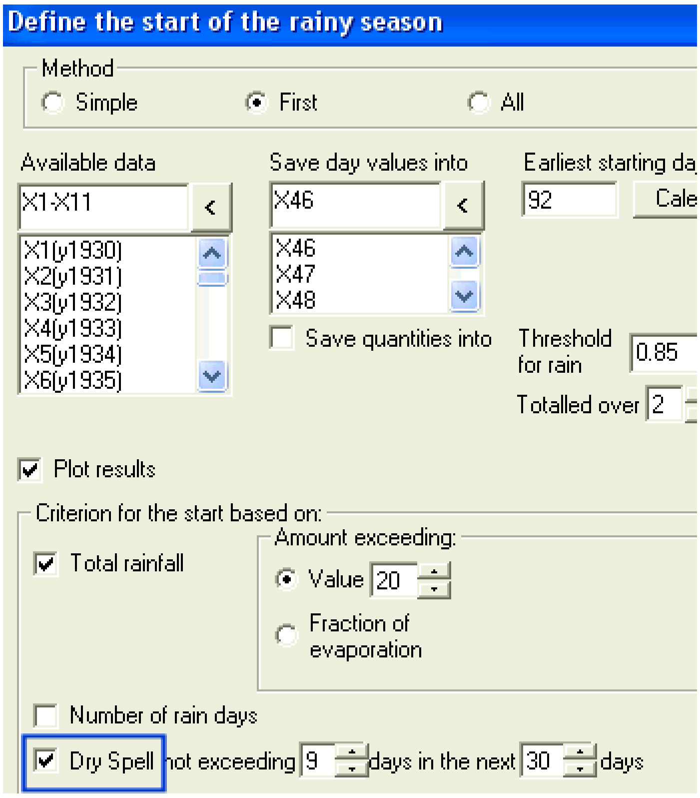


Fig. 6.4h Starting dates

	X42
	stdry1
1	12 May
2	27 Apr
3	24 Apr
4	04 Jun
5	25 Apr
6	13 May
7	22 Apr
8	05 May
9	16 May
10	13 May
11	27 Apr

The **Climatic ⇒ Events ⇒ Start of the rains** dialogue is shown in Fig. 6.4g. Tick the Dry Spell check box to add this extra condition.

The results are in Fig. 6.4h. Compared to the results without the extra condition of dry spells, shown in Fig. 6.2e, these may be considered as the **successful** starting dates. The values are the same in all but 2 years indicating that the initial sowing was successful in 9 of the 11 years. These results are considered further in Section 6.7.

6.5 Water balance

The next dialogue in the **Climatic ⇒ Events** sub-menu is for **Water Balance**. This was introduced in Chapter 4. Use the 11 years of Samaru data as the example and start, as usual, by clearing unwanted data from the worksheet. Use **Manage ⇒ Data ⇒ Clear(Remove)**, or type : **rem x12-x60** in the Commands and Output window.

In its simplest form, the **Climatic ⇒ Events ⇒ Water balance** dialogue is used similarly to those for **Spells** and the **Start of the rains**, namely as a transformation of the daily data. For example the dialogue for the 11 years of data, completed as shown in Fig. 6.5a, results in the output shown in Fig. 6.5b.

Usually we would use a column of data containing the daily evaporation. For simplicity here, assume a single constant value of 5 mm per day. Choose 100mm as the soil capacity. The results in Fig. 6.5b are the same as in Chapter 4, Fig. 4.3e. There, Fig. 4.3f also showed the graph that is available from the dialogue.

Fig. 6.5a Simple water balance

`File ⇒ Open Worksheet ⇒ samsmall.wor`
`Manage ⇒ Data ⇒ Clear(Remove) ⇒ X12-X60`
`Climatic ⇒ Events ⇒ Water balance`

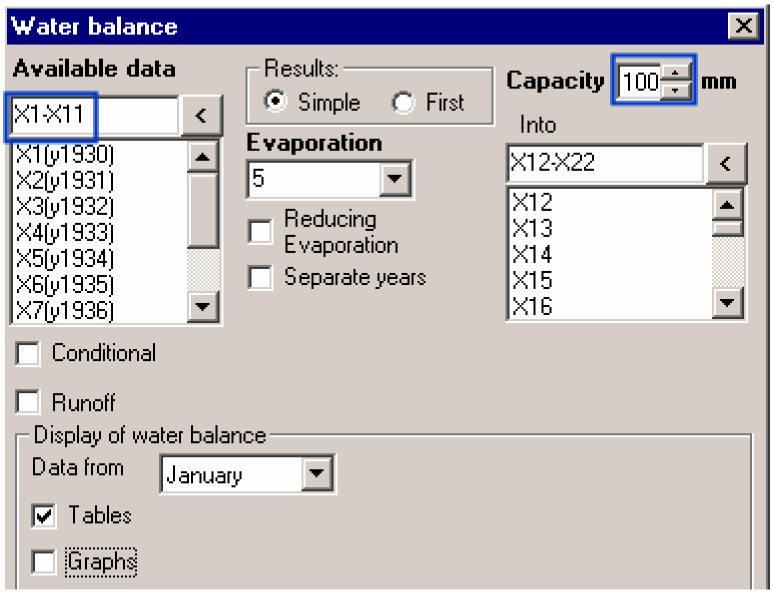


Fig. 6.5b Simple water balance for 1930

Mon	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Day.												
1	--	--	--	--	2	3	3	81	++	69	7	--
2	--	--	--	--	--	14	--	76	95	64	2	--
3	--	--	--	--	--	9	--	79	90	59	--	--
4	--	--	--	--	--	4	--	++	94	54	--	--
5	--	--	--	--	--	14	16	97	89	55	--	--
6	--	--	--	--	13	9	25	++	++	50	--	--
7	--	--	--	--	8	5	20	++	96	45	--	--
8	--	--	--	--	7	--	23	++	99	40	--	--
28	--	--	--	17	2	18	67	83	74	4	--	--
29	--	--	--	12	18	13	64	78	79	--	--	--
30	--		--	7	13	8	60	73	74	17	--	--
31	--		--		8		78	85		12		--
Minimum										(Overall: 0)		
	0	0	0	0	0	0	0	71	74	0	0	0
Maximum										(Overall: 100)		
	0	0	0	22	18	42	78	100	100	69	7	0

In training workshops we usually produce the output for the 11 years in the form shown above. Participants then study the results, particularly to identify the date, in each year, that the water balance first falls to zero, which is 14th October in Fig. 6.5b. This is a possible definition of the end of the growing season.

We digress briefly to make a statistical point. Some studies need the 'average water balance'. This is easily found from the results above. In Fig. 6.5c the **`Manage ⇒ Manipulate ⇒ Row Statistics`** dialogue is used on the water balances calculated in Fig. 6.5a to produce the 'average water balance'.

The alternative, in some studies, is to average the rainfall instead, and then calculate the water balance from the average rainfall data. This use of the **Row Statistics** dialogue for the rainfall is in Fig. 6.5d. Then the **`Water balance`** dialogue is used as in Fig. 6.5a, but with the one column of average rainfall data.

Fig. 6.5c Row statistics on water balance
`Manage ⇒ Manipulate ⇒ Row Stats`

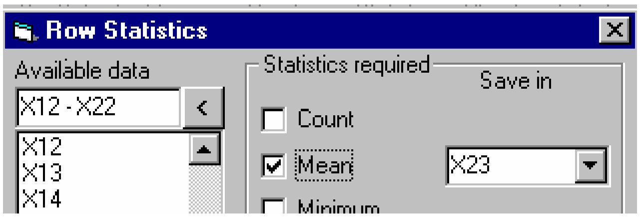
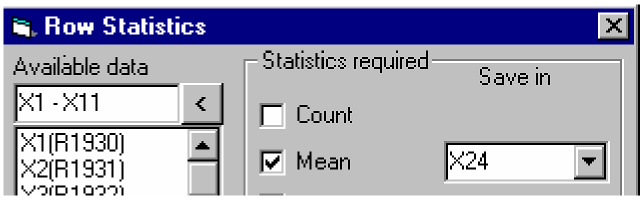
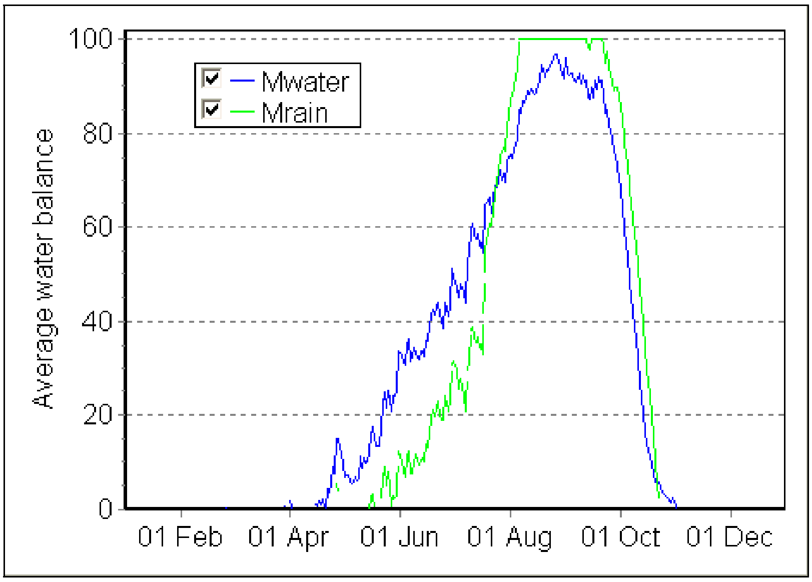


Fig. 6.5d Row statistics on the rainfall
`Manage ⇒ Manipulate ⇒ Row Stats`



The results from the 2 methods are compared graphically in Fig. 6.5e. They differ considerably (and the difference would be even more marked at sites that do not have such a clearly defined rainy season). To understand that they must be different, see for example, the beginning of the season. There the **average rainfall** per day is less than the evaporation, so the resulting water balance remains at zero until almost June. However, individual years do have more than 5 mm of rain occasionally, so the **average water balance** is greater than zero from mid-April.

Fig. 6.5e Averages are not always as simple as they may seem



In the pre-computer age, the averaging was done early in the analysis (i.e. the average rainfall was calculated first), mainly for ease of computation. Now both methods are equally easy, hence average **at the end of the analysis**, i.e. here average the water balance to give *Mwater* in Fig. 6.5e **not** the rainfall. The result is then simple to interpret, it is just the **average of the water balances**. It is also easy to move to a study of the year-to-year variability by comparing the water balance for individual years with its average. This is what is done now, to examine the end of the season.

After this digression, consider the end of the growing season. A possible definition is the first occasion, after a given date that the water balance drops below a certain level. For Samaru, define the end to be

- the first day after *1st September* that the water balance *drops to zero*

See from Fig. 6.5b that this date was 14th October in 1930. As a day number this is the 288th day of the year.

The use of the **Climatic ⇒ Events ⇒ Water balance** dialogue to produce these results is shown in Fig. 6.5f, with the results shown in Fig. 6.5g. This has again produced a column of length 11, because 11 years of data have been processed.

Fig. 6.5f Water balance for the end of the season

Climatic ⇒ Events ⇒ Water balance

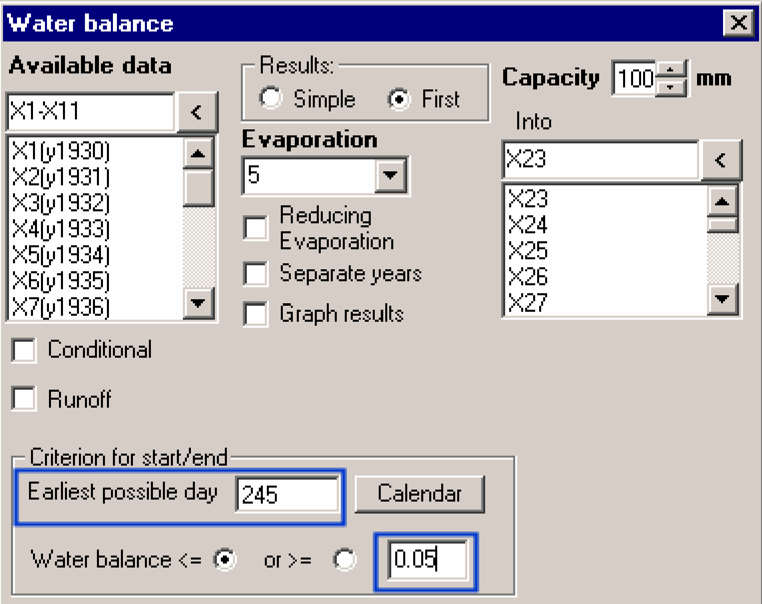


Fig. 6.5g Results

	X27
1	14 Oct
2	14 Oct
3	21 Oct
4	20 Oct
5	17 Oct
6	13 Oct
7	18 Oct
8	21 Oct
9	15 Oct
10	02 Nov
11	16 Oct
12	

To conclude this section, one further feature of the Water balance dialogue is considered. With the simple use of the command it is assumed that the evaporation proceeds at a constant rate, irrespective of the amount of water left in the soil. An extension allows the drop in the water balance to also depend on the current amount of water. As an example, take the first year of data (for 1930), for which the simple water balance was tabulated in Fig. 6.5b. The relevant part of the dialogue is shown in Fig. 6.5h.

Fig. 6.5h Decreased evaporation in dry periods

Climatic ⇒ Events ⇒ Water

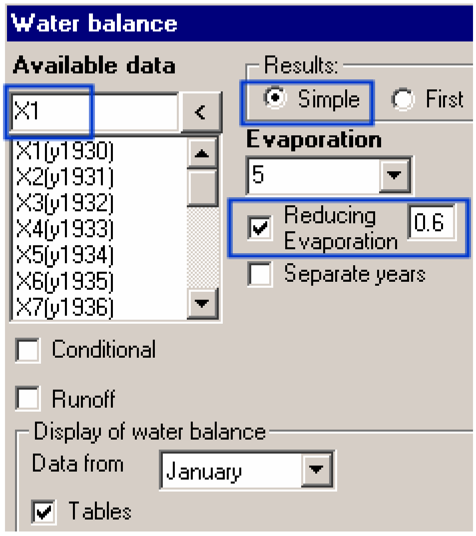
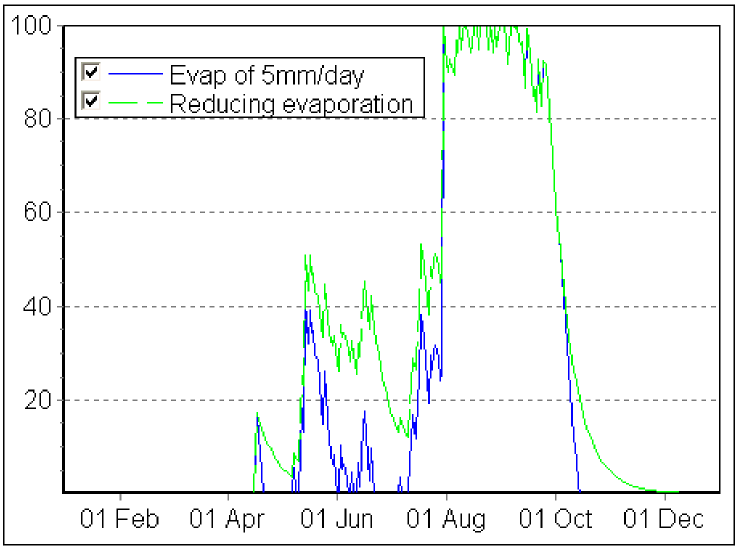


Fig. 6.5i Water balance for 1930 using alternative definitions



With the reducing evaporation set to 0.6, as given in Fig. 6.5h, the evaporation is reduced whenever the water balance is less than 60% of the soil water holding capacity, i.e. less than 60 mm in this case. So, whenever the water content is more than 60 mm, the evaporation is 5 mm as before. It is also 5 mm on rainy days with more than 5 mm of rain. If the water balance is 30 mm then the evaporation is reduced to 2.5 mm, and so on.

The results comparing the water balance with and without the adjustment for evaporation are plotted for 1930 in Fig. 6.5i. An inspection of figures such as these are to help decisions on whether the adjustment is useful for the given application.

6.6 Extremes

There is often need to know the risk of extreme events, particularly those that are damaging, such as heavy rainfall, high flood flows in rivers, high winds, or extreme temperatures, either hot or cold.

This section concentrates on rainfall extremes, in particular it starts with the daily data and shows how the extremes can be calculated. The way the extremes are then analysed is in later chapters, particularly in Section 11.4.

Fig. 6.6a Display of daily data

File ⇒ Open From Library ⇒ Kurunega
Climatic ⇒ Display Daily

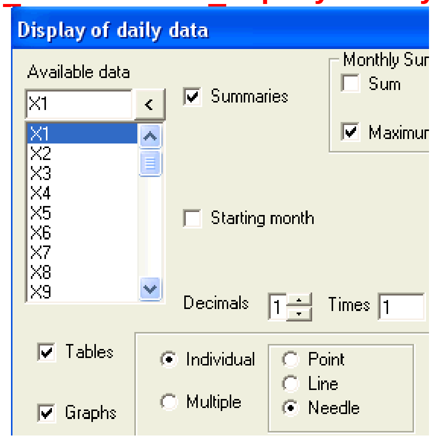
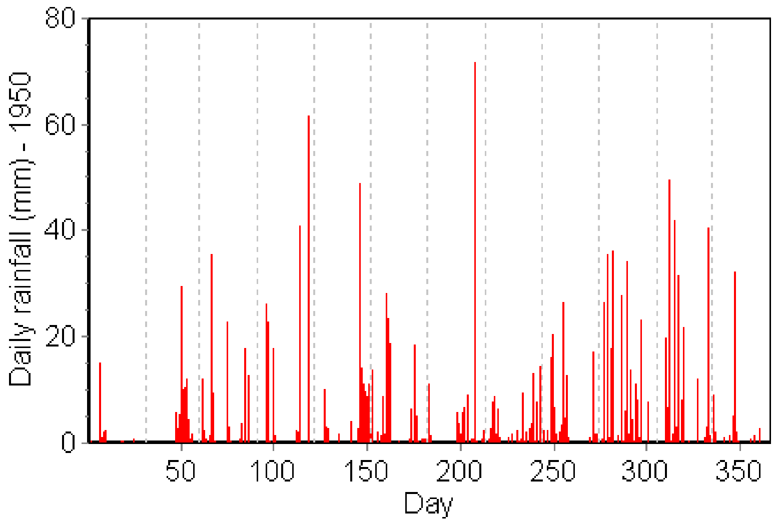


Fig. 6.6b Graph for 1950



Start by **File ⇒ Open From Library** and select the worksheet **Kurunega.wor**. Next use **Climatic ⇒ Display Daily** and select X1 to display the data for 1950 (Fig. 6.6b).

The graph in Fig. 6.6b shows the largest daily rainfall in this year was about 70mm and was on about day 210 of the year. The numerical results in Fig. 6.6c confirm the value to be 71.6mm, and in July.

Fig. 6.6c Daily data for 1950

Mon	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Day.												
1	--	--	--	--	--	13.7	11.2	--	2.3	--	--	9.1
2	0.3	--	12.2	--	--	--	1.3	0.8	--	0.5	--	2.0
...												
29	--		--	--	8.6	0.5	0.8	--	1.5	--	1.3	--
30	--		--	--	10.9	--	2.3	14.5	--	--	--	--
31	--		--		2.5		--	--		--		--
Maximum									(Overall: 71.6)			
	15.0	29.5	35.6	61.5	48.8	28.2	71.6	14.5	26.4	36.1	49.5	32.3

To find the maximum values in each year, use **Climatic ⇒ Events ⇒ Extremes**, Fig. 6.6d. This saves the maximum value each year, together with the day number in the year when it occurred. It also saves the year numbers. The results are columns x35 to x37 of the worksheet, Fig. 6.6e. As usual these columns are of length equal to the number of years of data, which is 34 in this case. They are therefore ready for the next stage in the analysis

Fig. 6.6d Extremes dialogue

Climatic ⇒ Events ⇒ Extremes

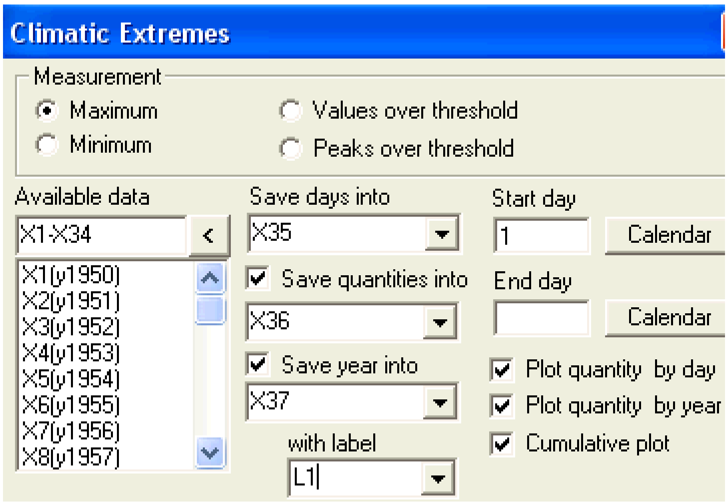


Fig. 6.6e Results saved to the worksheet

	X35	X36	X37 - F
1	26 Jul	71.63	1950
2	01 Nov	97.54	1951
3	19 Apr	87.63	1952
4	25 Oct	90.93	1953
5	13 Mar	91.69	1954
6	19 Oct	97.79	1955
7	17 Jun	122.43	1956
8	25 Dec	157.99	1957
9	23 Mar	168.4	1958
10	05 Apr	105.16	1959
11	03 Apr	71.63	1960
12	23 Oct	85.6	1961
13	09 Oct	97.79	1962

The dialogue also allows for three different graphs to be given. They are shown, after some tidying, in Fig. 6.6f, 6.6g and 6.6h.

Fig. 6.6f Time series plot

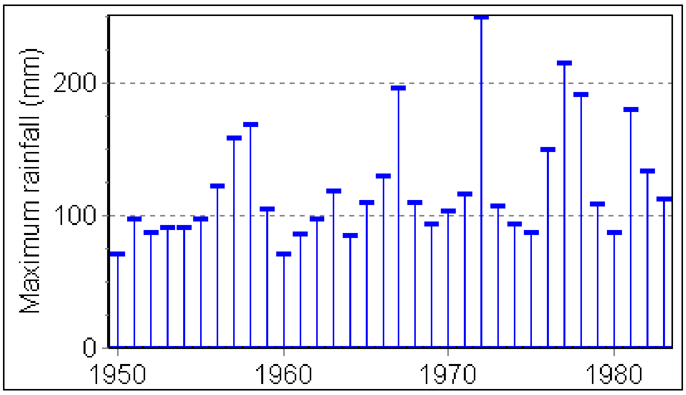
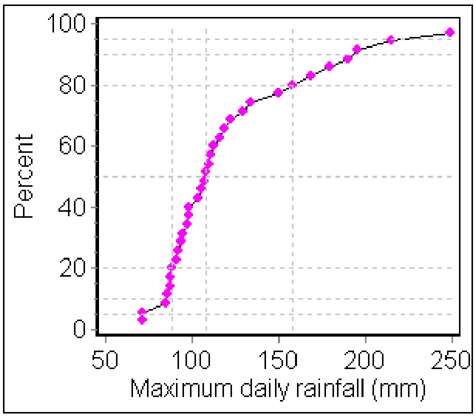


Fig. 6.6g Probability plot



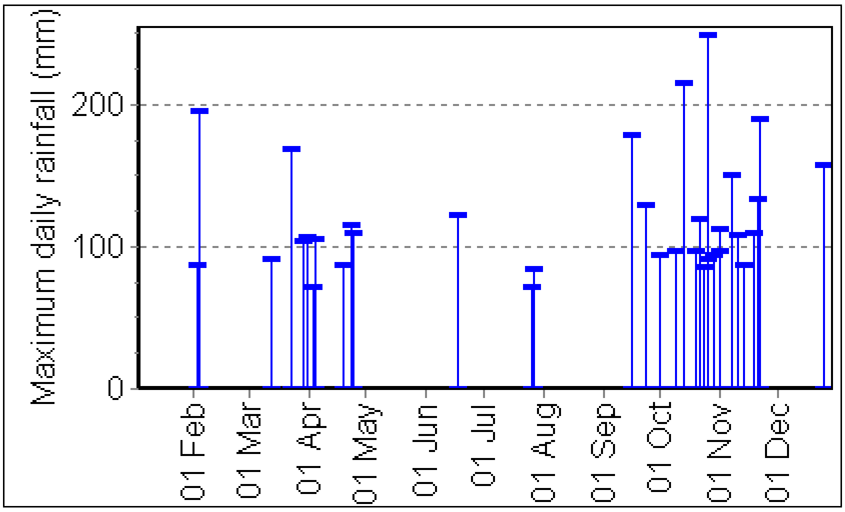
A time-series plot is given in Fig. 6.6f. The largest daily extreme was in 1971 and was about 250mm. Just over half the years had a value of more than 100mm. There does not seem to be any obvious trend to the data.

Fig. 6.6g gives a probability plot. Here you can fix a percentage (on the y-axis) and read the value from the x-axis. For example the 80% point is about 160mm. So the annual maximum exceeded 160mm on one year in five, and this is called the five-year return period.

Or you can fix an amount (on the x-axis) and read the percentage of years. For example take 100m to see that the maximum daily value was less than 100mm in about 40% of the years.

The final graph indicates when in the year the maximums occurred. Fig. 6.6h shows that more than half the extremes were between mid-September and the end of November. This is the main rainy season, so that is perhaps hardly surprising. Perhaps it is more of a surprise that the maximum daily value occurred so often outside this period.

Fig. 6.6h Time of year plot



Sometimes problems result from a sequence of consecutive wet days, rather than an extreme value on a single day. Fig. 6.6i specifies the extremes from running 7-day totals.

Fig. 6.6i 7-day extremes

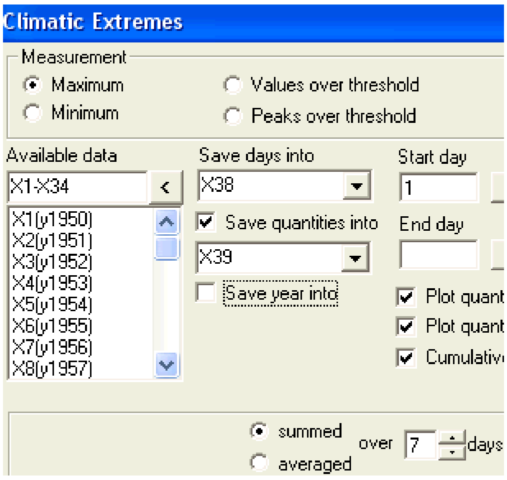
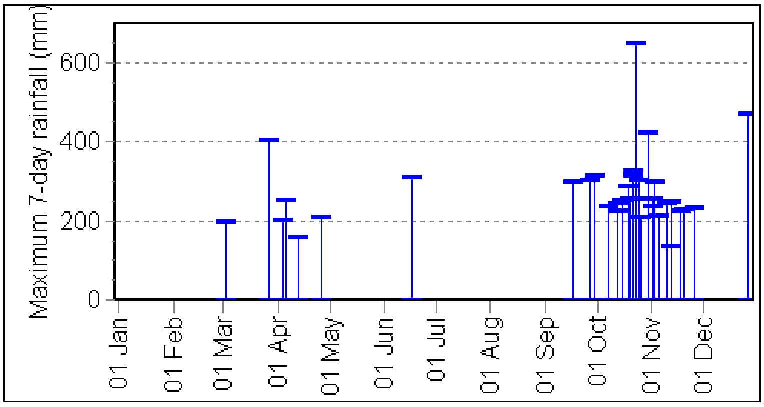


Fig. 6.6j 7-day maximums against the time of year



One of the resulting graphs is in Fig. 6.6j. The largest ever 7-day total was over 600mm. And these maximum values were usually between the end of September and the end of November.

With the series of maximum values, considered so far, the analysis ignores other large events, which in one year, may exceed the maximum value in another year.

The alternative is to break the 'rule', of looking for just a single value each year. Instead, fix a limit and ask for all values that exceed the limit. Look first at single days and, ask for all the values over 110mm, Fig. 6.6k.

The result is in Fig. 6.6l. There were 30 occasions on which there was a single day with more than 110mm, and half of these occasions were in the three years 1958, 1972 and 1977.

Fig. 6.6kValues >110mm

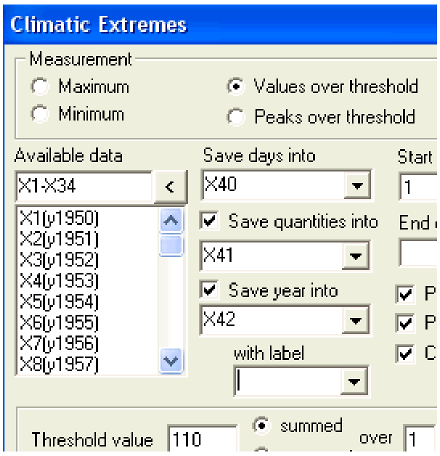
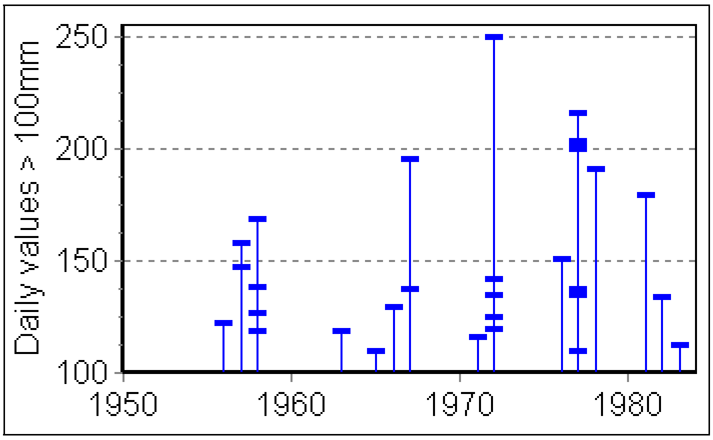


Fig. 6.6l There are 30 such events



One complication with this approach is that the same rain-bearing system may give multiple events that qualify.

To avoid this problem use the option to give just the peaks over the threshold, see Fig. 6.6m. This selects the largest within each group of consecutive days. Fig. 6.6m specifies a threshold of 250mm in 7 consecutive days. A 'group' is defined to consist of all consecutive values with more than 250mm that are separated by less than 5 days.

Fig. 6.6m

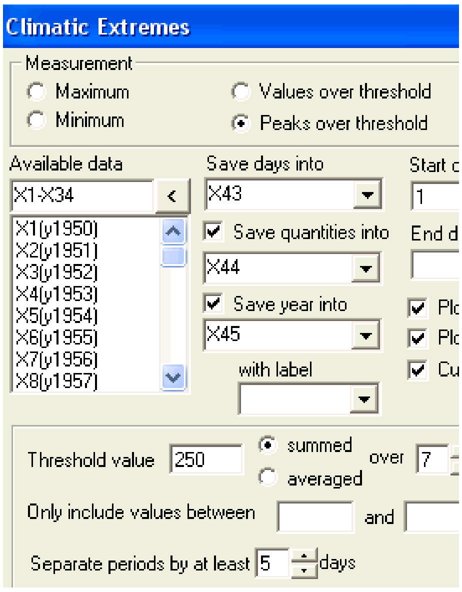
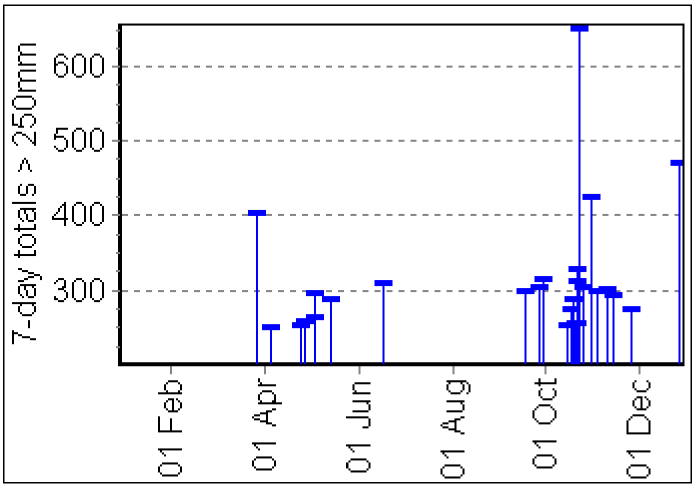


Fig. 6.6n



The extremes are processed further in Section 11.4.

6.7 More Events

This section might also be called 'responding to the users needs' and is the parallel to Section 5.6. It consists of a series of five requests, or challenges that relate to the events considered in the previous sections of this Chapter, but do not all correspond to just the use of one of the **Climatic ⇒ Events** dialogues.

As examples, use either the 11 years of data in samsmall.wor, or the longer record of 56 years in samaru56.wor. The illustrations here are with the short record.

6.7.1 Length of the season

The first request is to calculate the length of the growing season. This is straightforward. It needs a suitable definition of the start of the season, perhaps using one from Section 6.3

Fig. 6.7a shows the results using 20mm in a 2-day period after 1st April (**X12** in Fig. 6.2g) and also, from Section 6.5 where there was no long dry spell in the following 30 days (**X42** in Fig. 6.4h). The third column in Fig. 6.7a is from the definition of the end of the season, using the simple water balance, in Section 6.6 (**X27** in Fig. 6.5g).

Fig. 6.7a Data to calculate the length of the season

File ⇒ Open ⇒ samsmall.wor
Recalculate the events as shown earlier

	X12	X13	X14
	St_1	St_Dry	End_1
1	108	133	288
2	118	118	288
3	115	115	295
4	156	156	294
5	116	116	291
6	134	134	288
7	113	113	292
8	126	126	295
9	118	137	290
10	134	134	307
11	118	118	291

To calculate the lengths use the **Manage ⇒ Calculations** dialogue, Fig. 6.7b. The columns of the lengths of the seasons for the two starting definitions are shown in Fig. 6.7c. The lengths differ between zero and 25 days for the two different start definitions.

Fig. 6.7b Calculate the length of the season

Manage ⇒ Calculations

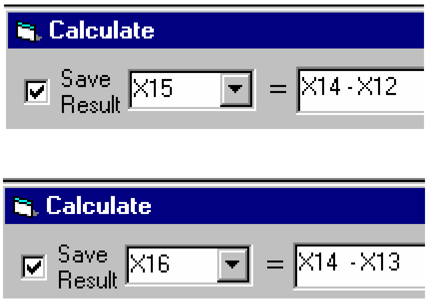


Fig. 6.7c Seasons length

	X15	X16
1	180	155
2	170	170
3	180	180
4	138	138
5	175	175
6	154	154
7	179	179
8	169	169
9	172	153
10	173	173
11	173	173

6.7.2 The risk from planting

Use the two definitions of the start of the season in Fig. 6.7a. Assume that the sowing failed if there was a 10-day dry spell, so this is the risk that is to be found. In this case **X13** gives the successful sowing date and hence the initial sowing failed each time **X13** is different from **X12**. So use the **Manage ⇒ Calculations** dialogue again, as shown in Fig. 6.7d. This adds column X17, as shown in Fig. 6.7e. The sowing failed in 2 of the 11 years. Hence the risk is estimated as 2/11 or 18%. In those two years the farmers had to wait either 19 or 25 days for a successful sowing.

Fig. 6.7d Calculate risk of planting too early

Manage ⇒ Calculations

Fig. 6.7e Years >0 at risk

	X17
1	25
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	19
10	0
11	0

6.7.3 The rainfall amount at the start of the rains

The third challenge starts with the same definition of the start of the season that was considered in Section 6.3, namely:

- the first occasion with more than **20mm** in **1 or 2 days** after ***1st April***

but it applies equally to any of the definitions that were used. Assume the user would like the dates, as in Section 6.2, but would also like to know the actual total that triggered the start. Thus they would like the results shown in Fig. 6.7g. These totals used to require a step-by-step approach, but they are now an option from the **Start of the Rains** dialogue, Fig. 6.7f.

Fig. 6.7f Start of the Rains

Climatic ⇒ Events ⇒ Start of the Rains

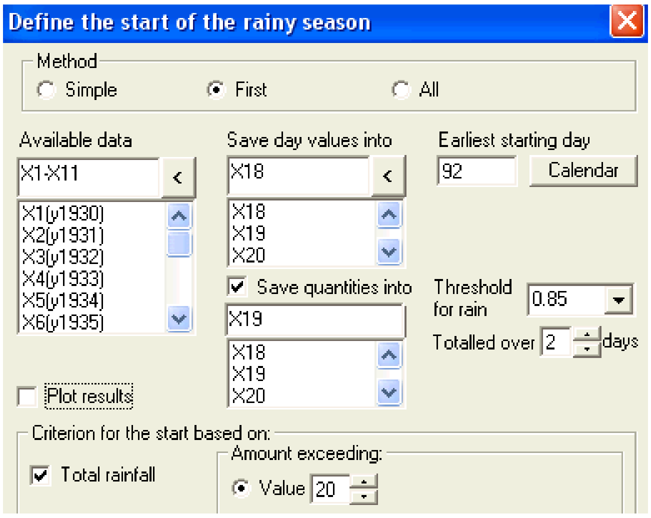


Fig. 6.7g Results

	X18	X19
1	17 Apr	22.1
2	27 Apr	40.13
3	24 Apr	38.61
4	04 Jun	21.08
5	25 Apr	55.63
6	13 May	24.63
7	22 Apr	30.48
8	05 May	23.62
9	27 Apr	20.83
10	13 May	20.83
11	27 Apr	26.67
12		

The results are in Fig. 6.7g. This is an example of a challenge that was needed sufficiently that it is now in a dialogue.

6.7.4 Total rainfall during the growing season

The fourth challenge is to calculate the total rainfall in the growing season. Fig. 6.7a gives the first and last days we need to consider, so all that is needed is to calculate the total rainfall between these dates.

This is another challenge that was required sufficiently that it is now an option within an Instat dialogue. Use the **Climatic ⇒ Summary** dialogue, as shown in Fig. 6.7h, with the option in Fig. 6.7i.

Fig. 6.7h Sum rainfall amounts

Climatic ⇒ Summary

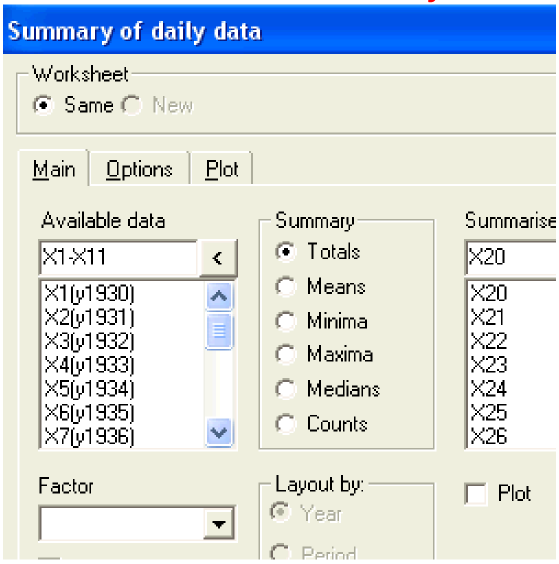
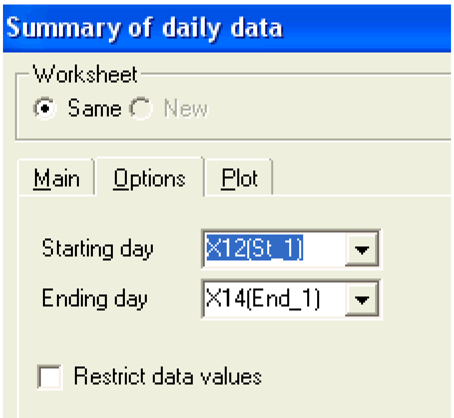


Fig. 6.7i Total rainfall in season



The results are in Figs 6.7j and 6.7k, and show the seasonal totals vary from 877mm to 1184mm.

Fig. 6.7j Results in the output window

```
Summary of daily data

STAts X1-X11;sum  X20;FROm X12;TO X14

Column          Sum      From      To
                X20      St_1    End_1

y1930          1037       108      288
y1931          1178       118      288
y1932          1174       115      295
y1933          1184       156      294
y1934          1043       116      291
y1935         956.5       134      288
y1936          1139       113      292
y1937         947.9       126      295
y1938         939.8       118      290
y1939          1056       134      307
y1940         877.1       118      291
:
```

Fig. 6.7k Results in the worksheet

	X12	X14	X20
	St_1	End_1	
1	108	288	1037.4
2	118	288	1178.0
3	115	295	1173.5
4	156	294	1184.4
5	116	291	1043.1
6	134	288	956.5
7	113	292	1139.4
8	126	295	947.9
9	118	290	939.8
10	134	307	1056.1
11	118	291	877.1
12			

6.7.5 No dialogue satisfies the client's request!

For the final challenge consider a user who asks for a definition of the start of the season that is not in any dialogue. The request starts innocently with the definition of the start as:

- the first occasion that the water balance exceeds 20mm after 1st April

This clearly involves the **Climatic => Events => Water Balance** dialogue, rather than the one called Start of the Rains. This is in Fig. 6.7m and shows the dialogue allows for this option. The results for a particular evaporation and water capacity are in Fig. 6.7n.

Fig. 6.7m Water balance dialogue

Climatic ⇒ Events ⇒ Water Balance

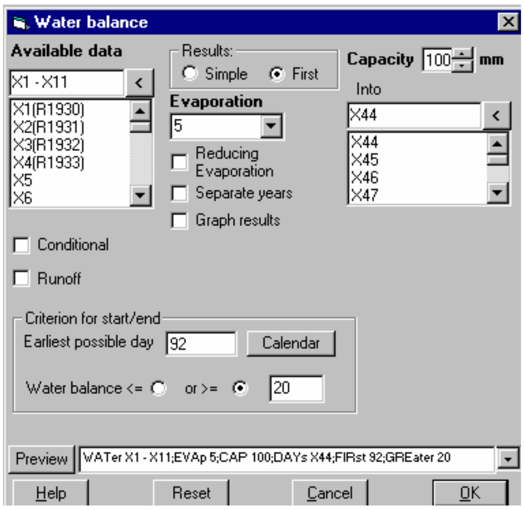
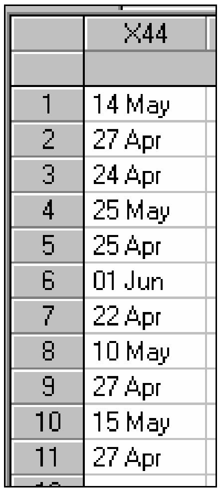


Fig. 6.7n Results



	X44
1	14 May
2	27 Apr
3	24 Apr
4	25 May
5	25 Apr
6	01 Jun
7	22 Apr
8	10 May
9	27 Apr
10	15 May
11	27 Apr

The client then modifies the definition and asks for one that corresponds to a successful start as follows:

- the first occasion that the water balance exceeds 20mm after 1st April and does not drop to zero in the following 30 days.

This typifies a challenge where there is no dialogue. There is also currently no macro written, and writing a macro would not be a trivial task. So what is to be done?

If you enjoy computing, perhaps you could write a macro. If not does someone in your organisation have more experience of the software?

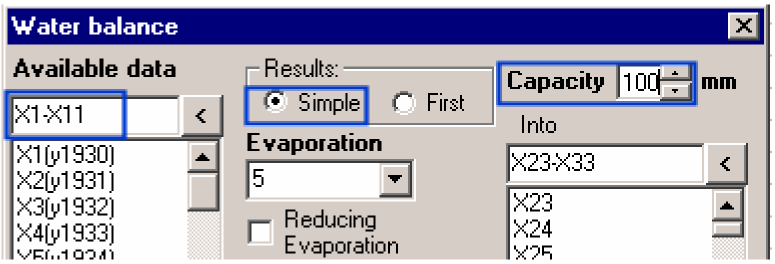
But this takes time and meanwhile your client is waiting.

In such a situation, a simple method can often be used to start with. Even if an 'automatic' method is produced later these results will be useful to check the method.

A solution is in Fig. 6.7o. Use the **Water balance** dialogue and produce just the simple results.

Fig. 6.7o Simple water balance

Climatic => Events => Water Balance



The data for 1931 are shown in Fig. 6.7p. There were 2 unsuccessful plantings on 27th April and 10th May, with the successful planting on 28th May. It is a quick procedure to examine the data each year and then enter the resulting values.

Many users find this 'hand analysis' to be useful and this examination of the data can be done jointly with the client. They will learn a lot in this way, and may modify their definition to something even more appropriate for their work as a result!

Fig. 6.7p Water balance table for 1931

Mon	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Day.												
1	--	--	--	--	20	62	90	++	96	56	--	--
2	--	--	--	--	15	58	90	95	91	51	--	--
3	--	--	--	--	10	56	86	91	89	51	--	--
4	--	--	--	--	5	51	95	86	++	46	--	--
5	--	--	--	--	--	59	90	81	++	41	--	--
6	--	--	--	--	--	59	86	76	++	36	--	--
7	--	--	--	--	0	55	81	71	++	31	--	--
8	--	--	--	--	--	50	80	83	++	26	--	--
9	--	--	--	--	--	61	75	81	++	21	--	--
10	--	--	--	--	32	56	++	88	++	16	--	--
11	--	--	--	--	27	51	95	91	95	11	--	--
12	--	--	--	--	22	52	90	89	90	6	--	--
13	--	--	--	--	18	60	85	84	85	1	--	--
14	--	--	--	--	13	55	82	79	++	--	--	--
15	--	--	--	--	8	76	77	93	99	--	--	--
16	--	--	--	--	24	71	85	88	++	--	--	--
17	--	--	--	--	19	67	96	83	96	--	--	--
18	--	--	--	--	14	62	++	96	++	--	--	--
19	--	--	--	--	9	57	++	91	++	--	--	--
20	--	--	--	--	4	79	95	86	++	--	--	--
21	--	--	--	--	--	74	92	81	96	--	--	--
22	--	--	--	8	5	69	++	99	91	--	--	--
23	--	--	--	3	--	64	95	++	86	--	--	--
24	--	--	--	--	--	59	91	95	83	--	--	--
25	--	--	--	--	--	54	86	90	78	--	--	--
26	--	--	--	--	--	85	83	++	73	--	--	--
27	--	--	--	35	--	++	83	95	68	--	--	--
28	--	--	--	35	27	95	85	94	71	--	--	--
29	--		--	30	22	++	80	++	66	--	--	--
30	--		--	25	57	95	84	99	61	--	--	--
31	--		--		52		++	++		--		--

This final challenge was the only one where there is currently no neat solution. The fact that there still was a solution, even though it involves a little 'hand analysis', should give users the confidence that they can use the software to respond to the real needs of the client, rather than limiting the client to just the analyses that correspond to the current dialogues.

6.8 Conclusions

Chapters 5 and 6 have shown how easily a flexible analysis of rainfall 'events' can be conducted when daily records are available.

Instat provides general tools, so the user can choose the precise definition. For example, an agronomist may have different definitions of the start of the season for alternative crops or soil types. It is relatively easy to provide the results in a 'tailor made' report for chosen definitions, and to assess in addition, to which aspects of the definitions the results are sensitive.

In general the 'events' may be monthly totals (Chapter 5), a date of the start of the rains (Sections 6.2 and 6.4), maximum spell lengths (Sections 6.3 and 6.4) and so on. For each event, a year of data provides one number. These chapters, have concentrated on the first stage of the analysis. In most of the examples the dialogues produced columns of length 11, because there were just 11 years of data.

These columns of 'events' may then be processed by the dialogues in Instat's **Statistics** menu or using the **Climatic ⇒ Examine** and **Climatic ⇒ Process** dialogues. This is described in the next chapter.

Chapter 7 – Processing the Events

7.1 Introduction

In Chapters 5 and 6 we showed how the daily data can be summarised to give 'events' of interest. This chapter covers the second step, which is to process the events.

The examples in Figs 7.1a to 7.1c have already been through the first step.

Sammonth.wor is in Fig. 7.1a. It contains the monthly totals in x1-x12 for the data from Samaru in Nigeria from 1928 to 1983. The annual totals are in x13. They were generated from the daily file samaru56.wor using the methods described in Chapter 5.

Fig. 7.1a Monthly totals from Samaru in sammonth.wor

File ⇒ Open From Library ⇒ sammonth.wor

Current Worksheet - SAMMONTH.WOR

Columns | Constants | Strings | Labels — Title: Samaru monthly totals for 1928-1983

	X1*	X2*	X3*	X4*	X5*	X11*	X12*	X13*
	Jan	Feb	Mar	Apr	May	Nov	Dec	Total
1	0.5	0.0	52.6	56.1	182.1	0.0	0.0	1262.3
2	0.0	0.0	12.7	1.5	210.8	0.0	0.0	1284.2
3	0.0	0.0	0.0	33.8	115.6	0.0	0.0	1044.7
4	0.0	0.0	0.0	63.0	148.9	1.5	0.0	1197.8
5	0.0	10.7	0.0	64.8	226.6	0.0	0.0	1198.2
6	0.3	0.0	0.0	24.6	99.3	3.0	0.0	1311.6
7	0.0	0.0	0.0	73.4	180.3	0.0	0.0	1076.7
8	0.0	0.0	8.9	9.4	106.7	0.0	0.0	996.4

Kurun7.wor is in Fig. 7.1b. It contains the weekly totals in x1-x52 for the 34 years from 1950 to 1983 from Kurunegala in Sri Lanka. This has a bimodal rainfall pattern. The summaries were generated from the daily file **kurunega.wor** using the methods in Chapter 5.

Fig. 7.1b Weekly totals from Kurunegala, Sri Lanka

File ⇒ Open From Library ⇒ kurun7.wor

Current Worksheet - KURUN7.WOR

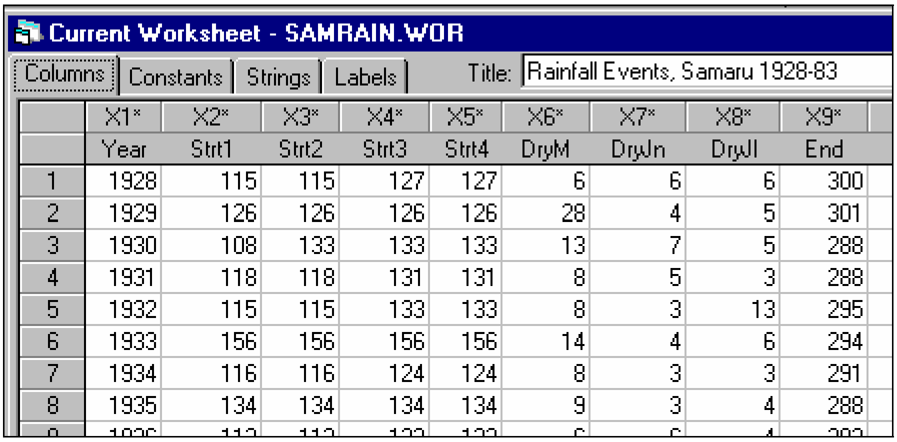
Columns | Constants | Strings | Labels — Title: 7 day rainfall totals Kurunegala 1950-83

	X1*	X2*	X3*	X4*	X5*	X6*	X7*	X50*	X51*	X52*
1	15.2	5.3	0.5	0.5	0.0	0.0	14.0	39.6	2.0	2.8
2	72.9	90.4	120.4	36.6	0.0	52.8	0.0	0.0	8.4	43.7
3	31.8	127.0	0.5	2.8	0.5	3.6	0.0	43.4	0.0	0.3
4	0.0	188.5	156.7	14.2	0.5	18.5	0.0	8.6	2.0	35.6
5	122.7	109.0	37.6	0.5	9.4	56.4	0.0	36.6	32.8	33.5
6	61.5	90.7	52.8	54.6	46.7	30.0	7.4	0.3	0.0	6.6
7	11.9	19.8	114.8	18.0	28.2	51.0	0.0	0.3	28.5	126.5
8	57.1	29.7	11.4	54.4	35.3	74.2	2.0	32.0	132.1	410.5
9	2.3	10.4	48.0	1.3	221.0	3.8	15.5	0.3	45.0	47.2
10	39.6	0.0	43.4	8.6	36.6	0.3	0.3	34.8	2.8	3.6

Samrain.wor is in Fig. 7.1c. It was generated from samaru56.wor using the methods described in Chapter 6. The contents are:

Column	Name	Contents
X1	Year	Year number from 1928 to 1983
X2	Strt1	First date after 1 April with >20mm in a 2 day period
X3	Strt2	As X2 with no 10 day dry spell in next 30 days
X4	Strt3	First date after 1 May with >20mm in a 2 day period
X5	Strt4	As X4 with no 10 day dry spell in next 30 days
X6	DryMay	Longest dry spell in May (dry = < 0.85mm)
X7	DryJun	Longest dry spell in June
X8	DryJul	Longest dry spell in July
X9	End	First date after 1 September with empty water balance. Capacity is 100mm and evaporation is 5mm per day

Fig. 7.1c samrain.wor


Current Worksheet - SAMRAIN.WOR

Columns | Constants | Strings | Labels | Title: Rainfall Events, Samaru 1928-83

	X1* Year	X2* Strt1	X3* Strt2	X4* Strt3	X5* Strt4	X6* DryM	X7* DryJn	X8* DryJl	X9* End
1	1928	115	115	127	127	6	6	6	300
2	1929	126	126	126	126	28	4	5	301
3	1930	108	133	133	133	13	7	5	288
4	1931	118	118	131	131	8	5	3	288
5	1932	115	115	133	133	8	3	13	295
6	1933	156	156	156	156	14	4	6	294
7	1934	116	116	124	124	8	3	3	291
8	1935	134	134	134	134	9	3	4	288

Section 7.2 uses dialogues from the **Statistics ⇒ Summary** menu in Instat, shown in Fig. 7.1d. Later chapters, for example Chapter 11, use other options, particularly from the **Statistics ⇒ Simple Models** menu.

Fig. 7.1d Statistics menu

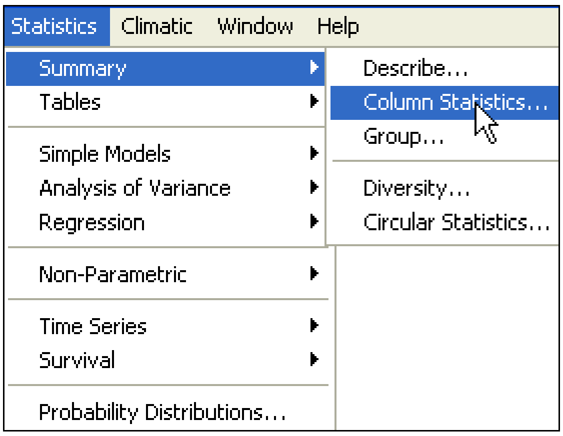
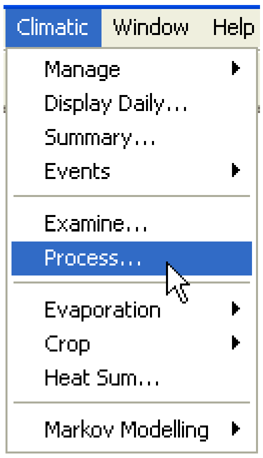


Fig. 7.1e Climatic menu



Using Instat's **Statistics** menu in Section 7.2 emphasises that this chapter describes general statistical summaries that can be done with any statistics package. Hence users who prefer a different statistics package, or who require methods of analysis that exceeds Instat's capabilities, can instead use Instat for just the initial summary stage described in Chapters 5 and 6 and then export the summary columns. Instat's **File ⇒ Import/Export** menu includes options to export as ASCII files or in a variety of other formats including Excel. The data in Fig. 7.1a to 7-1c are also available as Excel files.

Instat's **Climatic** menu, shown in Fig. 7.1e, includes two dialogues used in Sections 7.3 and 7.4. Good statistical practice requires users to continually explore their data in a critical way, and the **Climatic ⇒ Examine** dialogue, described in Section 7.3, provides options to make exploration of the 'events' into an easy procedure.

A common requirement is to process the summary data to give return periods (percentage points) or risks (probabilities). These are available using the statistics dialogues, as shown in Section 7.2, but the **Climatic ⇒ Process** dialogue makes their calculation into a one-step process. The simpler options of this dialogue are covered in Section 7.4, while options that involve the gamma distribution and 'smoothing' are in Chapters 11 and 12.

7.2 Summarising events

Fig. 7.2a shows the **Statistics ⇒ Summary ⇒ Describe** dialogue to process the annual totals from the **sammonth.wor** worksheet. The results are in Fig. 7.2b.

Fig. 7.2a Describe dialogue

File ⇒ Open ⇒ sammonth.wor
Statistics ⇒ Summary ⇒ Describe

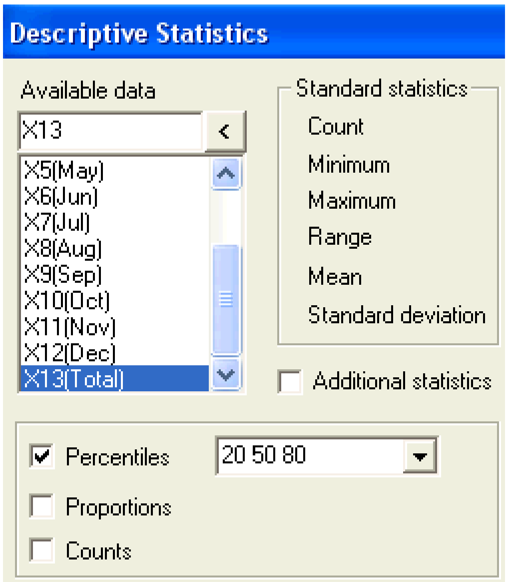
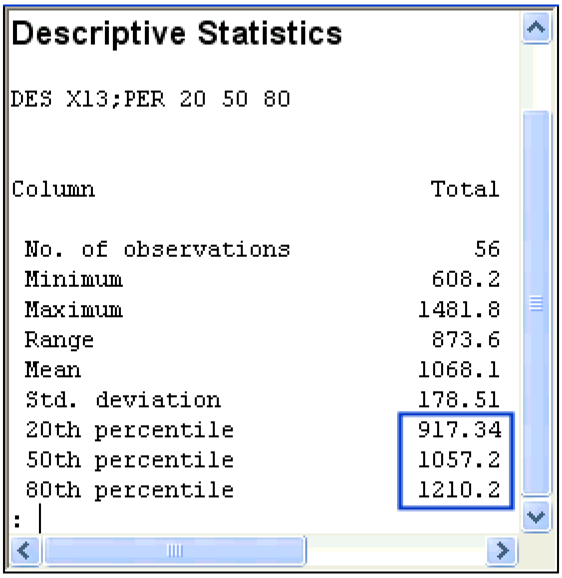


Fig. 7.2b Annual totals

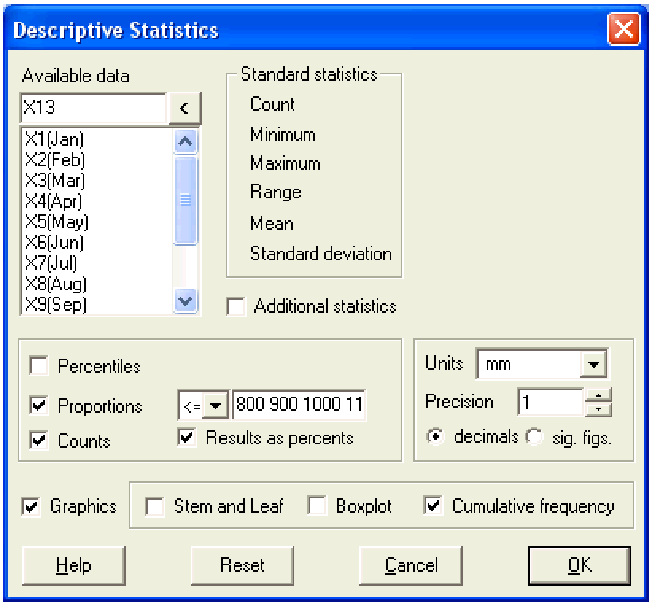


In addition to the standard summaries, the option has been set to give the 20% 50% and 80% points of the data. The median, or 50% point, is seen to be 1057mm.

The 80% point is 1210mm. One expects more than this total in just 20% of the years, i.e. on 1 year in 5. This is therefore sometimes called the 5-year return period. Similarly the 20% point is 915mm. One expects more than this total on 4 years in 5. Of course one expects **less** than 915mm on 1 year in 5, so it can also be thought of as a 5-year return period for small rainfall totals!

The converse is to fix on particular amounts and find the percentage of years that were less than them. The dialogue is in Fig. 7.2c. It also provides a graph.

Fig. 7.2c Calculating proportions

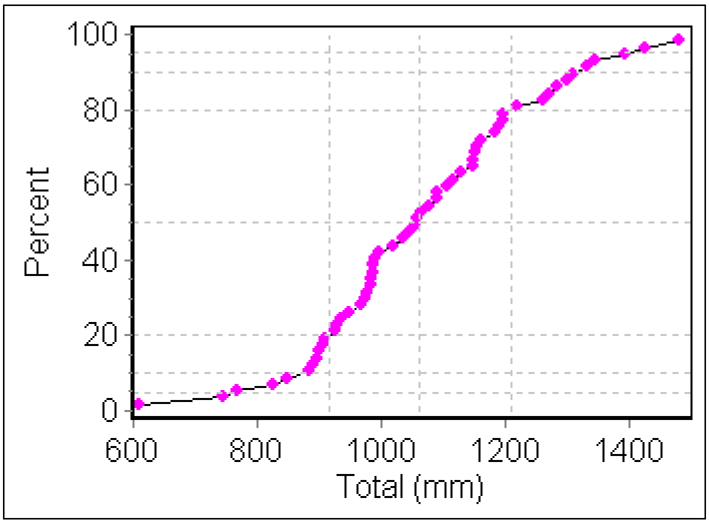


The results are in Fig. 7.2d. They show, for example, that 5% of the values were less than 800mm, and this was 3 out of the 56 values.

Fig. 7.2d Results

Column	Total
No. of observations	56
Minimum	608.2 mm
Maximum	1481.8 mm
Range	873.6 mm
Mean	1068.1 mm
Std. deviation	178.5 mm
Count <= 800 mm	3
Count <= 900 mm	9
Count <= 1000 mm	24
Count <= 1100 mm	33
Count <= 1200 mm	45
% of data <= 800 mm	5.4
% of data <= 900 mm	16.1
% of data <= 1000 mm	42.9
% of data <= 1100 mm	58.9
% of data <= 1200 mm	80.4

Fig. 7.2e Cumulative frequency plot



The graph in Fig. 7.2e can be read in either direction. Either fix on a percentage, say 80% (on the y-axis) and find that the corresponding value, which is just over 1200mm. Or fix on a value, say 800mm (on the x-axis), and find the percentage is about 5%.

The **Describe** dialogue (Fig. 7.2b) can be used on multiple columns and is useful to give quick summaries. The results are displayed in the output window.

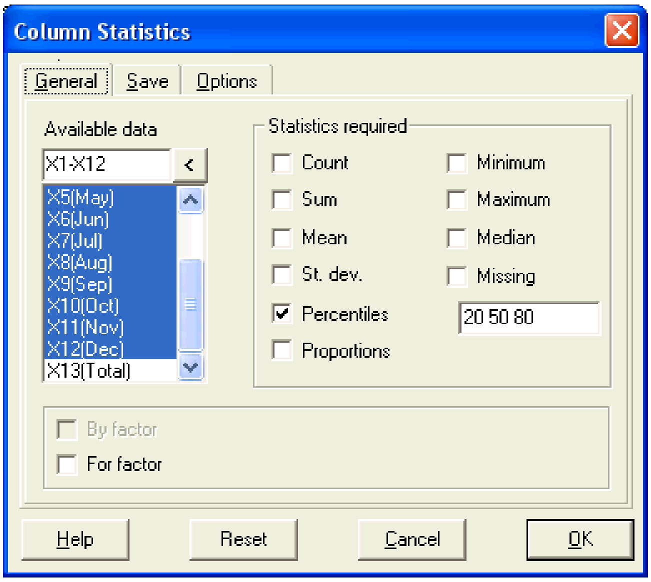
The **Statistics** ⇒ **Summary** ⇒ **Column Statistics** provides similar summaries, but the user chooses exactly which summary statistics to give. Fig. 7.2f shows this dialogue for the 20%, 50% and 80% points of the monthly data in X1-X12 of the **sammonth** worksheet.

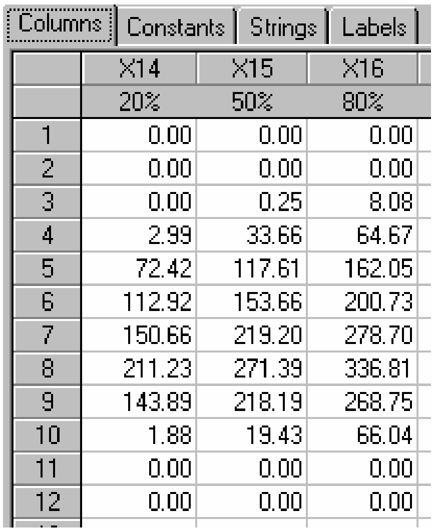
The results are again in the output window, but they can also be stored in further columns in the worksheet, as shown in Fig. 7.2g.

Fig. 7.2f Column Statistics dialogue

Fig. 7.2g Percentage points

Stats ⇒ Summary ⇒ Column Stats





	X14	X15	X16
	20%	50%	80%
1	0.00	0.00	0.00
2	0.00	0.00	0.00
3	0.00	0.25	8.08
4	2.99	33.66	64.67
5	72.42	117.61	162.05
6	112.92	153.66	200.73
7	150.66	219.20	278.70
8	211.23	271.39	336.81
9	143.89	218.19	268.75
10	1.88	19.43	66.04
11	0.00	0.00	0.00
12	0.00	0.00	0.00

To graph the percentage points for the different months, in Fig. 7.2g, add a column giving the values 1 to 12. This can be typed into the worksheet, or use the **Manage ⇒ Data ⇒ Regular Sequence** dialogue, as shown in Fig. 7.2h.

Fig. 7.2h Enter month numbers

Manage ⇒ Data ⇒ Regular Sequence

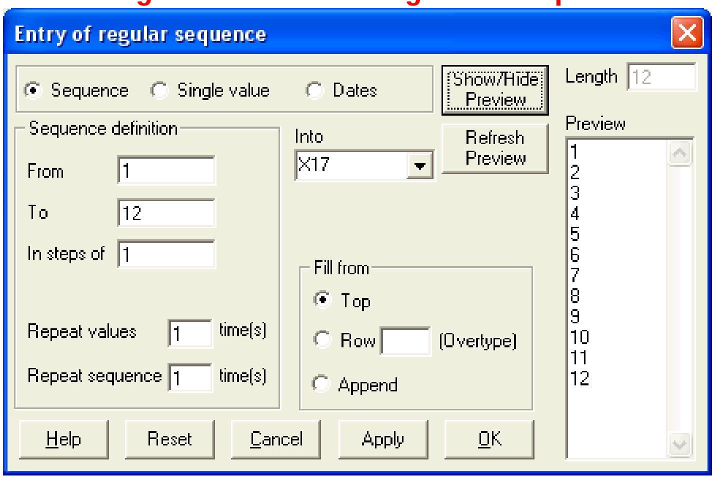
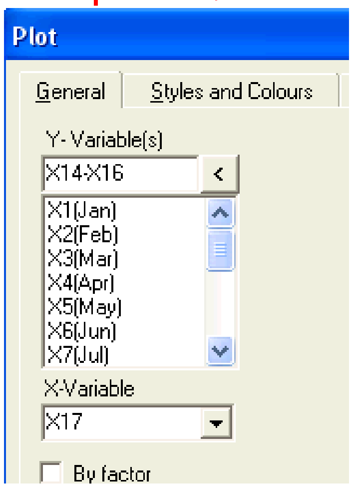


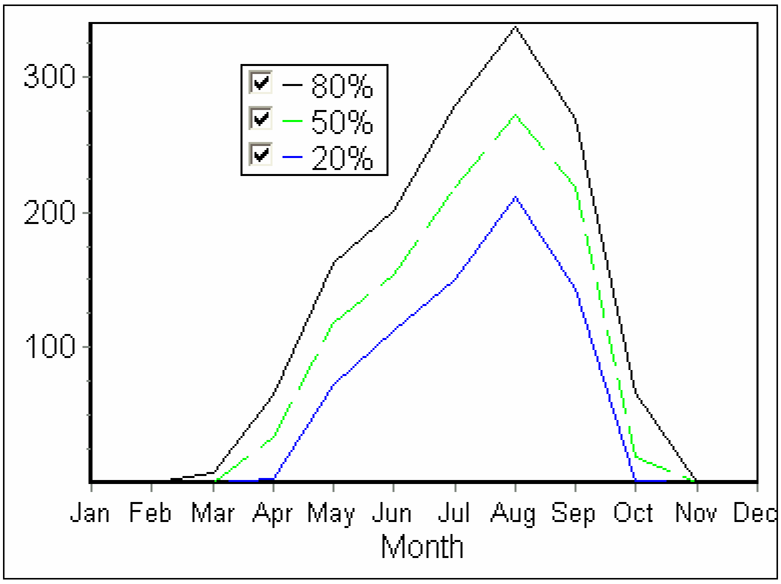
Fig. 7.2i Plot dialogue

Graphics ⇒ Plot



Then use the **Graphics ⇒ Plot** dialogue, Fig. 7.2i, with the lines option, to give the graph in Fig. 7.2j.

Fig. 7.2j Percentage points for monthly data

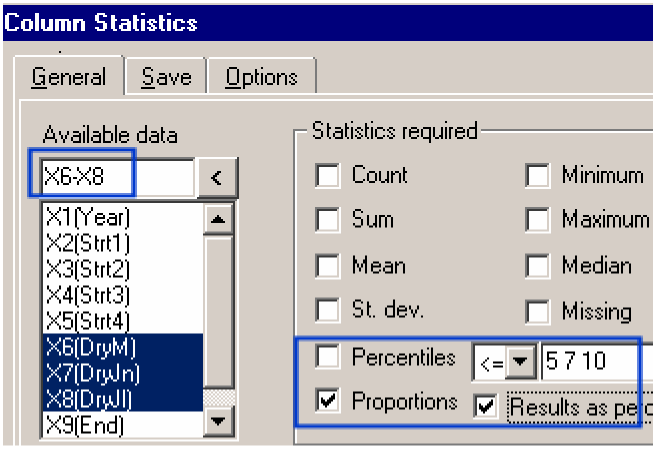


The **samrain.wor** worksheet, see Fig. 7.1c, is used for a second example of this type of summary. Here the 3 columns of data that give the longest dry spells each year in May, June and July are processed.

To assess the risk (or probability) of a long dry spell in each month, define 'long' alternatively as more than 5, 7 or 10 days. Use the **Statistics ⇒ Summary ⇒ Column Statistics** dialogue. The results are in Fig. 7.2k.

Fig. 7.2k Risk of a long dry spell

File ⇒ Open Worksheet ⇒ samrain.wor
Statistics ⇒ Summary ⇒ Column Statistics



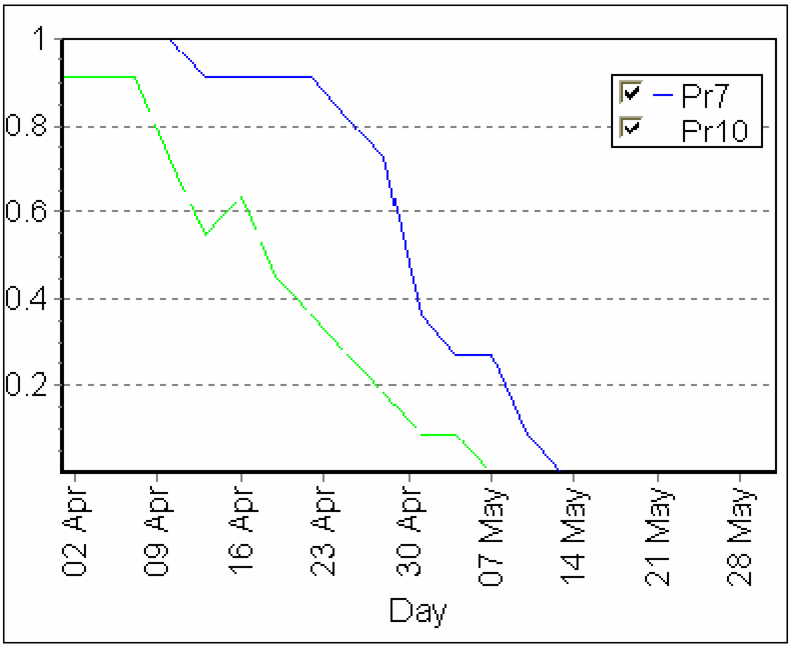
Column Statistics

Column	%<=5	%<=7	%<=10
DryM	5.4	35.7	58.9
DryJn	62.5	92.9	96.4
DryJl	62.5	89.3	98.2

The longest dry spell in May in most years is shown to be more than 5 days, but about 60% of years have the longest dry spell in May of 10 days or less. In contrast more than 60% of the years have no dry spell as long as 5 days in June or in July.

In Chapter 6, Fig. 6-4f showed the **conditional** risk of a 7 or 10-day dry spell in 30 days, for a range of dates from the beginning of April until mid June, for the 11 years of data from Samaru. The steps are outlined below, to produce the graph, in Fig. 7.2l, of the risk of a dry spell of 7 or 10 days or more from **samaru56.wor**.

Fig. 7.2l Chance of dry spells



The steps are as follows:

1) With the samaru56.wor data, first clear extra columns, then enter a column of the initial day numbers, from 92, 95, to 152 into **X57**, using **Manage ⇒ Data ⇒ Regular Sequence**, as shown in Section 6.5.
2) Use the **Climatic ⇒ Events ⇒ Spells** dialogue to give the 21 columns of **conditional** spell lengths in **X58-X78**.
3) Use the **Statistics ⇒ Data Summary ⇒ Column Statistics** dialogue roughly as in Fig. 7.2k to give 2 columns of spell lengths of 7 and 10 days into **X79** and **X80**.
4) Finally use the **Graphics ⇒ Plot** dialogue to give the graph in Fig. 7.2l.

From Fig. 7.2l**,** the chance of a dry spell of more than 10 days in the 30 days starting with a rainy day at the end of April (day 121) is just over 20%, i.e. 1 year in 5, while the chance of a seven-day spell is about 50%, i.e. in half of the years. By a month later, the chance of a 7-day dry spell is reduced to about 1 year in 20.

7.3 Exploring the events data

Chapter 4 looked at the daily data, both in tables and graphs. Data exploration is important at any stage and this section considers different ways of looking at the 'events'.

The example uses the monthly totals from the sammonth.wor worksheet.

Use the **Climatic ⇒ Examine** dialogue, shown in Fig. 7.3a.

Fig. 7.3a Examine dialogue

File ⇒ Open ⇒ sammonth.wor
Climatic ⇒ Examine

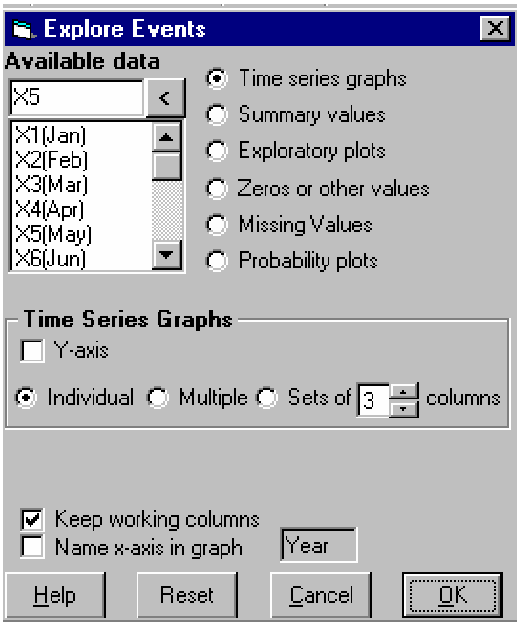
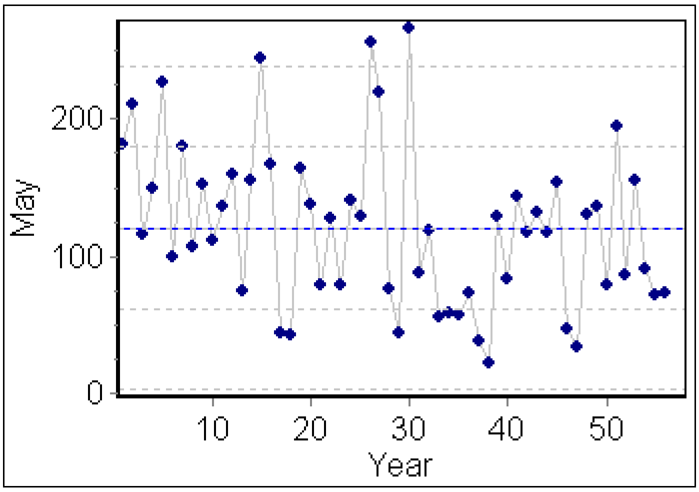


Fig. 7.3b Monthly totals for May



The first option is a time-series plot. Begin by plotting a single column, using the data for May. The dialogue is completed as shown in Fig. 7.3a. The plot is shown in Fig. 7.3b, where the solid line indicates the mean for May over the 56 years and the fainter lines are at one and two standard deviations.

In the graph, a column called 'Year' was generated for the x-axis. If a column with this name, and the correct length already exists, it will be used instead. Hence to clarify the graph use the **Manage ⇒ Data ⇒ Regular sequence** dialogue, as shown in Fig. 7.3c, to add a column that gives the actual years for these data. Now the **Climatic ⇒ Examine** dialogue is used again, as shown in Fig. 7.3a, and gives the graph shown in Fig. 7.3d, i.e. with a more appropriately labelled x axis.

Fig. 7.3c Enter years 1928 to 1983

Manage ⇒ Data ⇒ Regular sequence

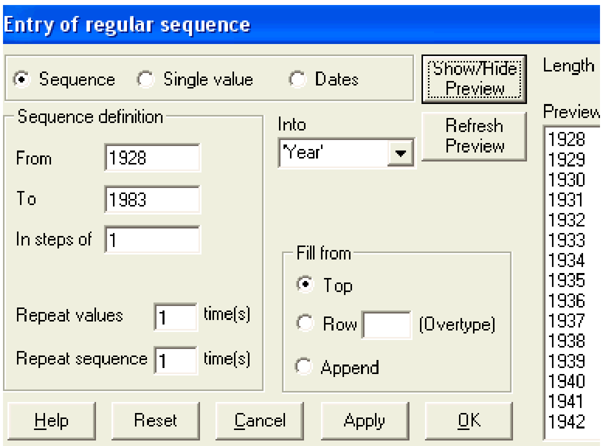


Fig. 7.3d Graph after replot

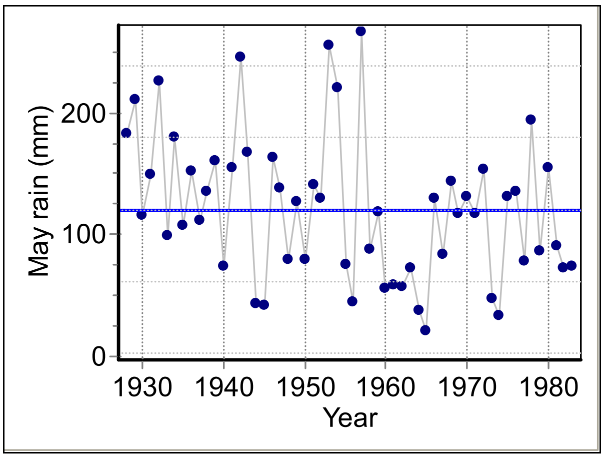
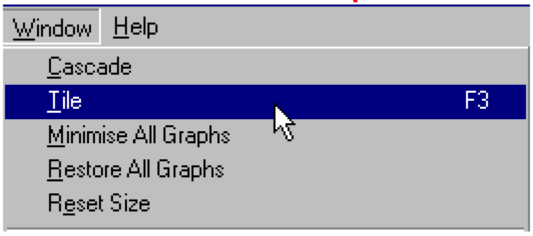


Fig. 7.3d includes one further feature. The box was checked to keep the working columns, see Fig. 7.3a. The appropriate vertical reference lines were added. If, as here, the reference lines coincide with the tick marks, then this can be done once the plot is displayed. Otherwise use the **Graphics ⇒ Plot** dialogue, or type: **replot ;vref 1930 1940 1950 1960 1970 1980** as a command.

The graph for a single month, given above, could equally have been generated from the **Climatic ⇒ Summary** dialogue, as described in Chapter 5. This dialogue also generates multiple graphs. When generating many graphs it is useful to develop a strategy for managing the graphics windows. Fig. 7.3e shows the options.

Fig. 7.3e Options for the graphics windows

Window ⇒ Graphs

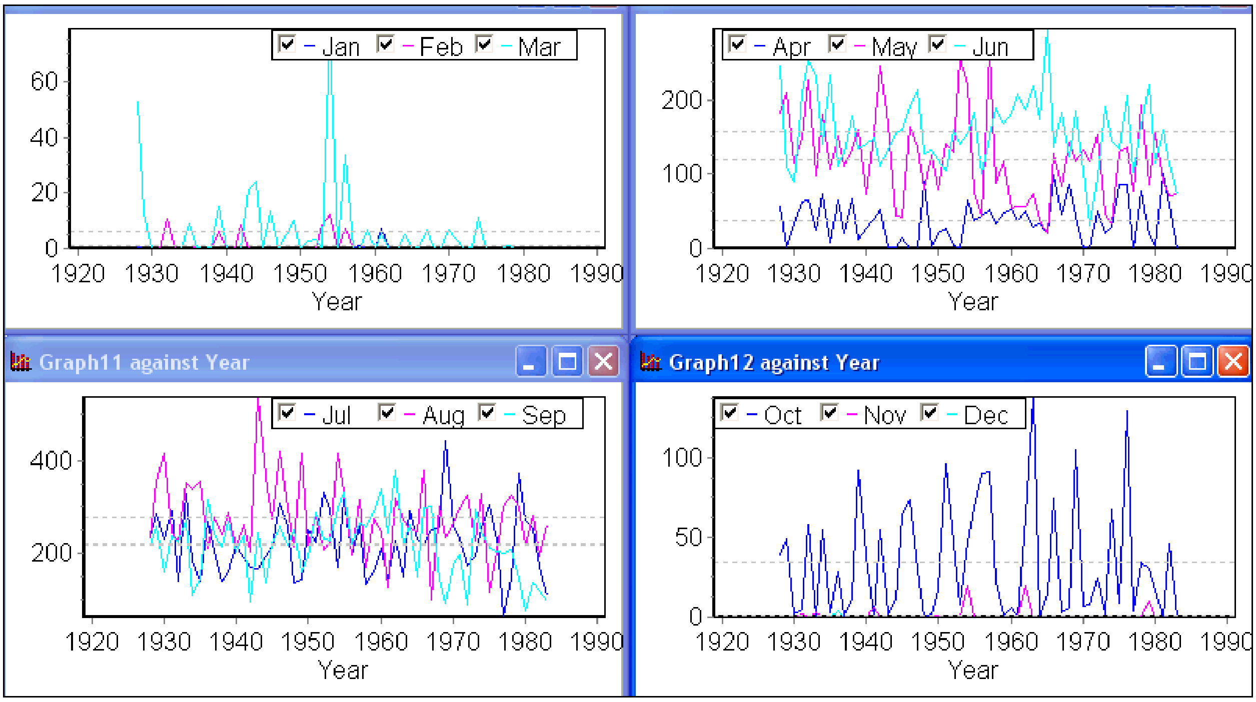


Pressing **<F3>** or using **Window ⇒ Tile**, tiles the graphic windows. Use **File ⇒ Remove all graphs** to remove all the old graphs, before starting a new stage.

The default is for successive graphs to cascade down the screen. Then the final ones can disappear! They can all be seen, for example, by using **<F3>** to tile, but alternatively use **Window ⇒ Auto Cascade Graphs** to switch off this feature.

Fig. 7.3f Tiled graphs of monthly rainfall for 1928-1983

Climatic ⇒ Examine



Remove all the graphics windows, then use the **Climatic ⇒ Examine** dialogue again, but for all 12 columns, X1 to X12. Choose the option to plot in sets of 3 columns to give the 4 graphs that can then be tiled as shown in Fig. 7.3f.

Data exploration is designed partly to assist in searching the data for oddities. In the graphs above there is one obvious possible oddity in the March data in 1954. The total in this year was 79mm, which is far higher than in any other year. The statistical literature includes many methods to assess whether a given value might be an **outlier**, but none are as useful as having access to the raw data. So look back at the daily records in samaru56.wor and use the **Climatic ⇒ Display Daily** dialogue.

Fig. 7.3g shows that most of this total of 79 mm was a rainfall of 74 mm on 17th March. This is in the dry season and was the highest single-day value in the year. This must have been a momentous event for that time of the year. The same year had 8.4mm on the same day in

February, and this should prompt an examination of the paper record as well as the record from neighbouring stations.

Fig. 7.3g Daily data for 1954

`Climatic ⇒ Display Daily (with samaru56.wor for X27)`

Mon	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Day.												
1	--	--	--	6.4	--	--	7.4	--	--	14.5	--	--
2	--	--	--	--	--	--	--	--	19.3	--	--	--
...												
14	--	--	--	7.1	1.0	--	--	--	--	--	17.3	--
15	--	--	--	--	--	--	--	31.5	0.5	--	--	--
16	--	3.8	--	--	21.1	21.1	0.3	1.8	32.5	--	--	--
17	--	8.4	74.2	--	--	13.5	38.1	0.8	--	--	--	--
18	--	--	--	--	--	2.5	9.4	--	16.8	0.5	--	--
19	--	--	--	5.6	--	9.6	2.3	4.3	1.3	1.3	--	--
...												
28	--	--	--	--	--	--	--	17.5	--	--	--	--
29	--		--	--	13.5	27.2	--	--	--	--	--	--
30	--		--	--	--	--	--	16.3	--	--	--	--
31	--		--		--		--	10.4		--		--
Total										(Overall: 1481.8)		
	0.0	12.2	79.0	64.8	219.7	156.5	167.9	417.1	298.4	47.0	19.3	0.0
Maximum										(Overall: 74.2)		
	0.0	8.4	74.2	34.3	52.6	27.2	38.9	64.0	51.1	24.1	17.3	0.0

Now return to the monthly data and use the second option from the **Climatic ⇒ Examine** dialogue. This gives summary values for the set of events, in this case, for the monthly totals. The dialogue is completed as shown in Fig. 7.3h. Ask for a line graph, include the quartiles and also rename the x-axis as Month, rather than the default of 'Period'. The results are in Fig. 7.3i and a table of the results is also given in the output window.

Fig. 7.3h Examine the monthly totals

`Climatic ⇒ Examine`

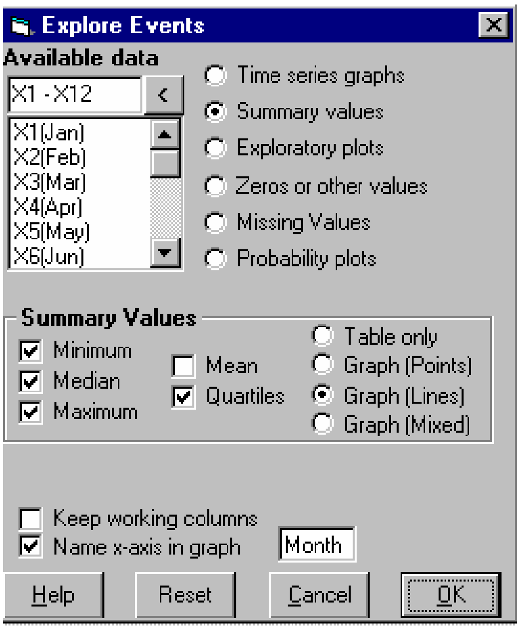
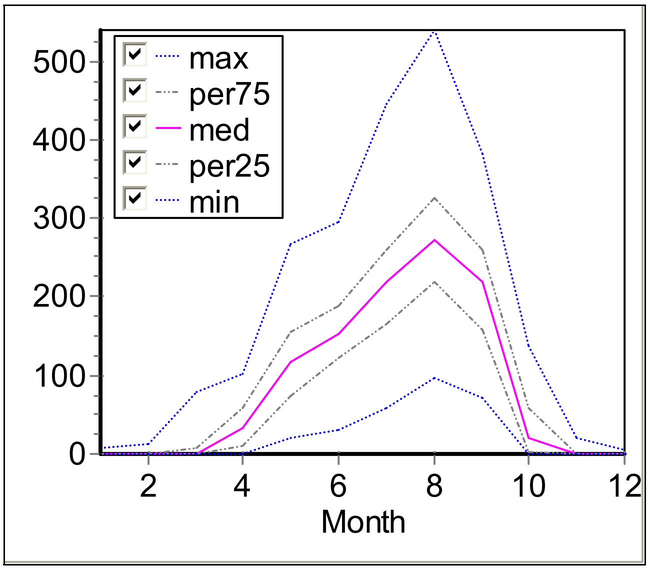


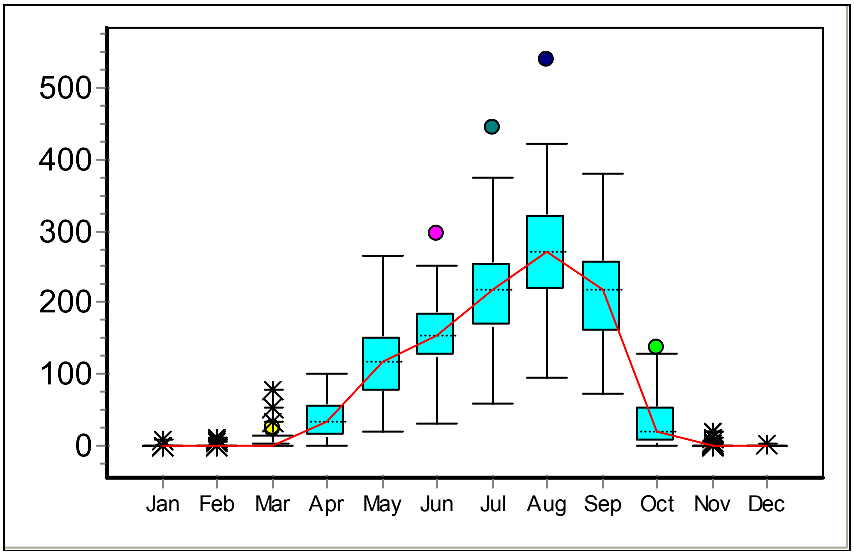
Fig. 7.3i Plot of monthly totals



The third option in the **Climatic ⇒ Examine** dialogue gives exploratory plots, and boxplots are particularly useful. Fig. 7.3j shows the plots for the 12 months and is similar in shape to Fig. 7.3i. These plots indicate that the distribution of the totals in each of these months is roughly symmetrically, but June, July and August have one surprisingly high value. For example, in August, a closer examination of the data shows the lowest total is about 100 mm, and the median is 271 mm. There are 3 years with high totals of between 400 and 420 mm and above this there is 1943, with 540 mm.

Fig. 7.3j Boxplots for Samaru monthly data

`Climatic ⇒ Examine`



Sometimes the monthly total may be zero. When processing the events, the presence of zeros is often an important feature of the data and hence there is a special option in the **`Climatic ⇒ Examine`** dialogue as shown in Fig. 7.3k. With the default settings this option gives the tabular output shown in Fig. 7.3l and the graph shown in Fig. 7.3m.

Fig. 7.3k Dialogue for zero rainfall

`Climatic ⇒ Examine`

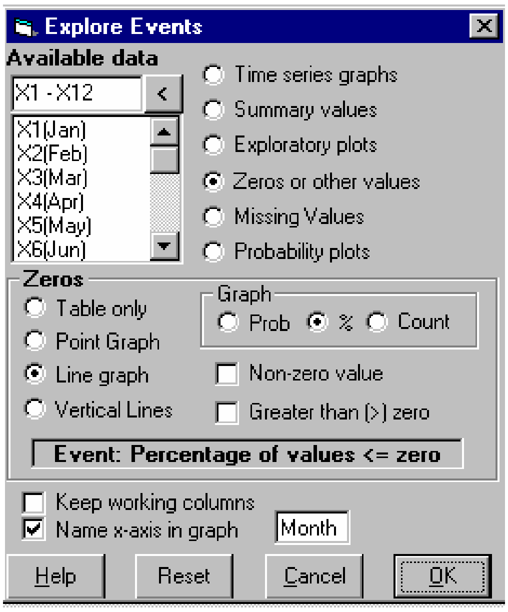
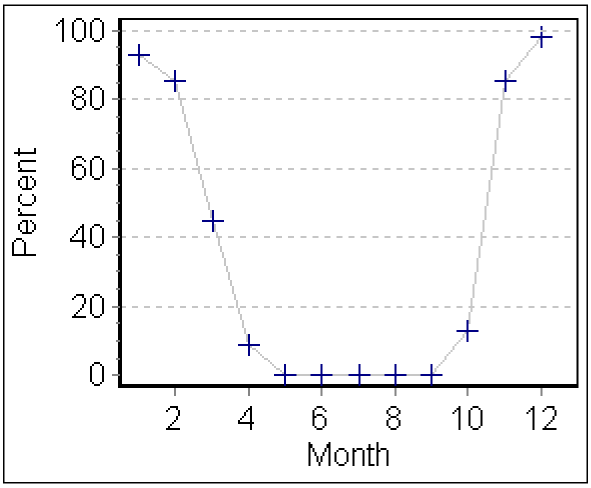


Fig. 7.3l Tabular output

Row	prob	Percent	Count
1	0.929	92.9	52
2	0.857	85.7	48
3	0.446	44.6	25
4	0.089	8.9	5
5	0.000	0.0	0
6	0.000	0.0	0
7	0.000	0.0	0
8	0.000	0.0	0
9	0.000	0.0	0
10	0.125	12.5	7
11	0.857	85.7	48
12	0.982	98.2	55

The graph and table show there was rain sometimes in every month, but each of the months of November to February was totally dry in more than 80% of the years.

Fig. 7.3m Graph of % of zero monthly totals



A challenge is to find in how many 'seasons' the 4-month period from November to February was totally dry. The slight 'catch' in the analysis is that the relevant January and February are in the year following the November and December data. Hence the data have to be 'shifted' before the seasonal totals are calculated. The result is that 38, i.e. 68% of the 55 years were dry for this 4-month period.

The fifth option in the **Climatic ⇒ Examine** dialogue is for missing values, not illustrated here. The final option, is to give a probability plot of the data. Try the normal model where the graph should be roughly a straight line if the data are normally distributed. Fig. 7.3n and 7.3o show the dialogue and graph for March, where a Normal model would be foolish, because many of the years are zero. The graphs for the 4 months from June to September are in Fig. 7.3p. They show a normal model is reasonable, though the largest observations, that were evident when constructing the boxplots, are still a 'feature'.

Fig. 7.3n A Normal model

Climatic ⇒ Examine

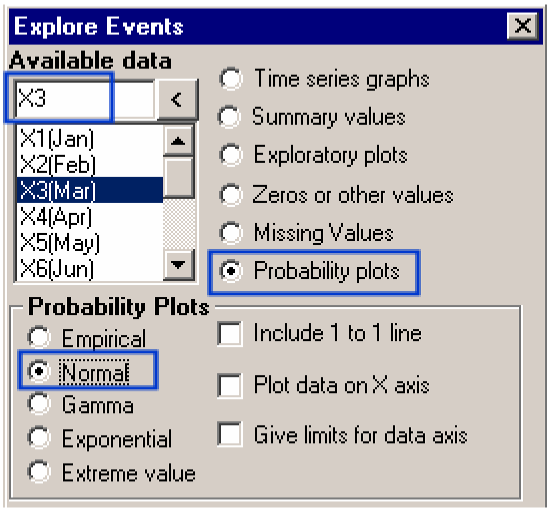


Fig. 7.3o Normal model for March

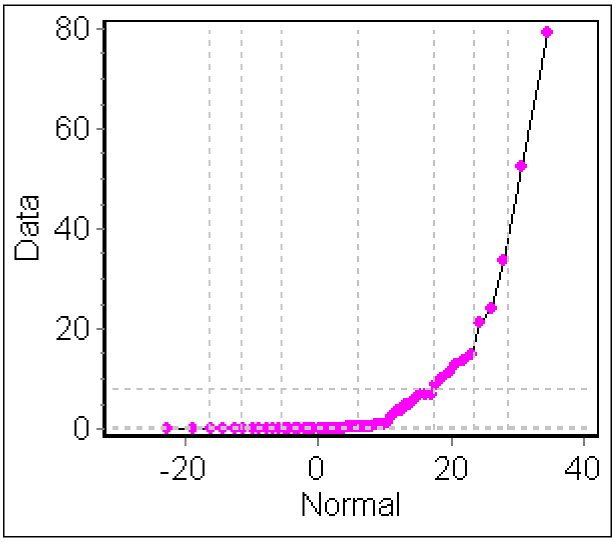
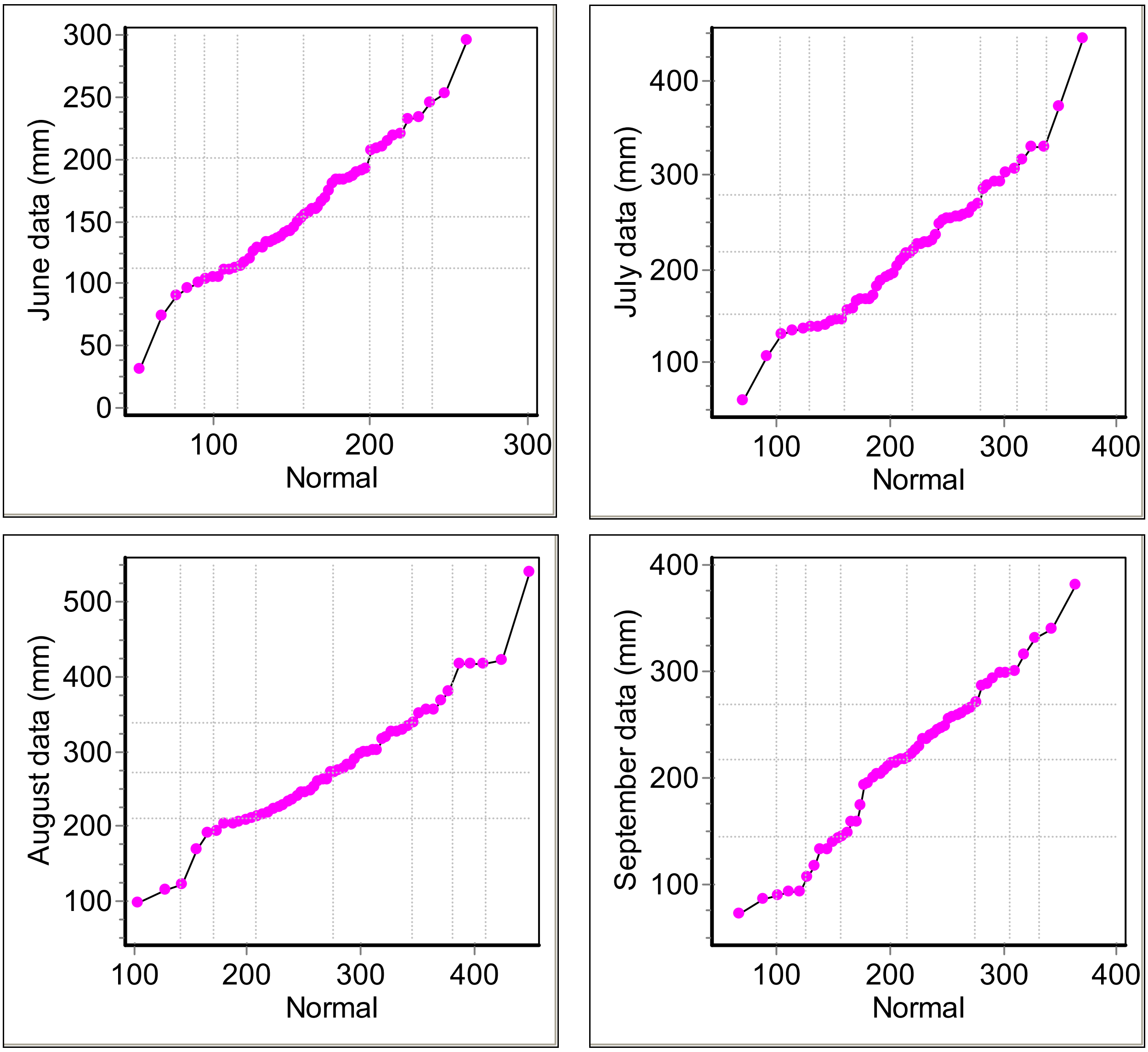


Fig. 7.3p Normal plots for June - September



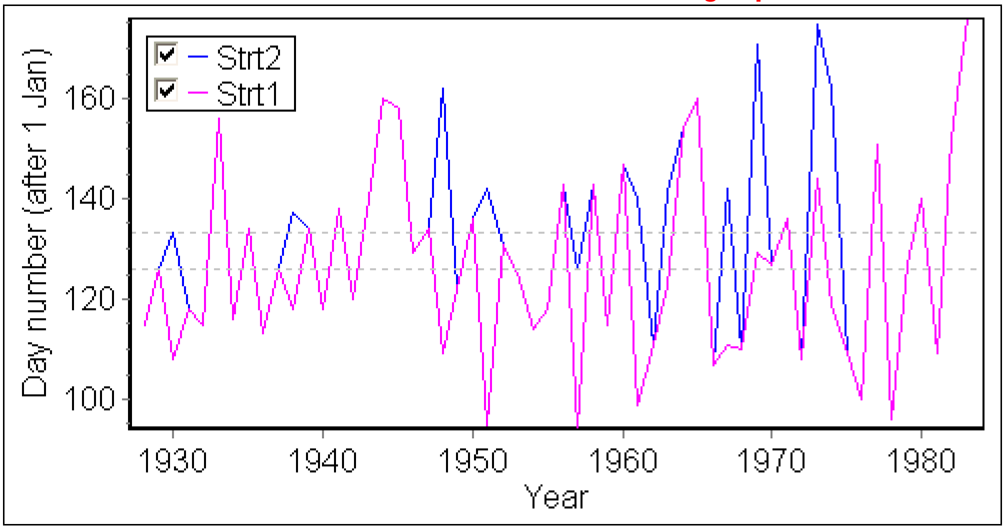
Now consider the **samrain** worksheet and examine the columns on the start of the rains. Here some results are shown, but without pictures of the dialogue.

Fig. 7.3q shows an example of a time series plot of the data on the start of the season. Plotting 2 definitions together is convenient to compare them. For example, Fig. 7.3q shows the added dry-spell condition gives a later starting date in about 10 of the years.

Fig. 7.3q Start of the rains in April – 2 definitions

File ⇒ Open Worksheet ⇒ samrain.wor

Climatic ⇒ Examine with **Time Series graph** of **X2 X3**



The second option in the dialogue is to produce summary values. In this case the graph is not particularly informative, but the 'Table only' option is useful, and can be used on all the columns together. This gives similar information to the boxplots, shown in Fig. 7.3r. These plots indicate the difference in spread of the starting dates, for the two sets of definitions. They also indicate how exceptionally late the final year of 1983 was, when the starting date was day 176, (24th June), for all definitions.

Fig. 7.3r Boxplots of Start of the rains in April and May

Climatic ⇒ Examine with **boxplots for X2-X5**

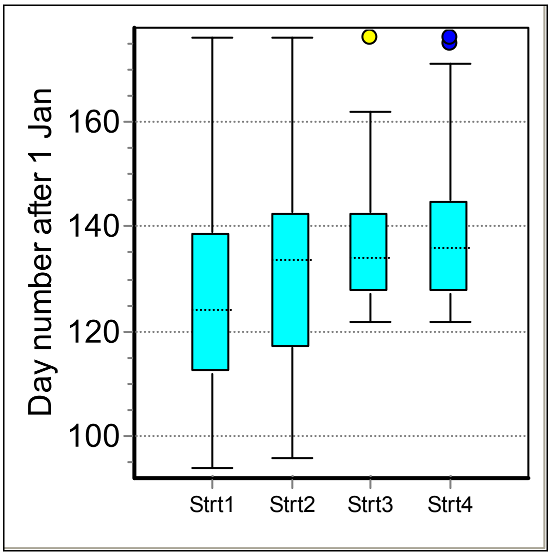
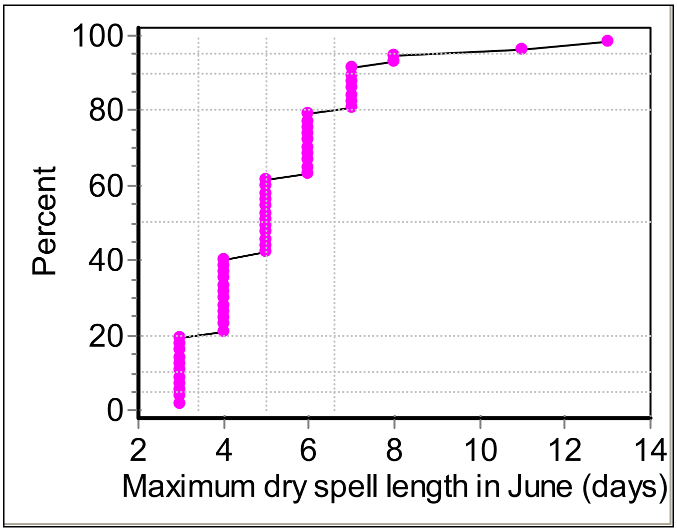


Fig. 7.3s Cumulative dry spells in June

Climatic ⇒ Examine with **Empirical Probability plot** of **X7; data plotted on X axis**



The boxplots also indicate the distribution is slightly skew, at least for the later definitions. This is confirmed by the probability plots (not shown). They indicate a normal model would be reasonable for the first definition, but slightly suspect for the others.

The final example is the dry spell lengths, again from the 56 years in Samaru. Fig. 7.3s shows the cumulative distribution, for the maximum spell lengths in June, and with the data plotted on the x-axis. The median is 5 days and less than 1 year in 5 had a dry spell length of 7 days or more.

Everything that has been described in this section can also be done using the dialogues in the **Statistics** and **Graphics** menus. The advantage of the **Examine** dialogue is that it is a one-step process. Hence we hope users will explore and not simply summarise their data.

7.4 Automating the process of calculating risks and return periods

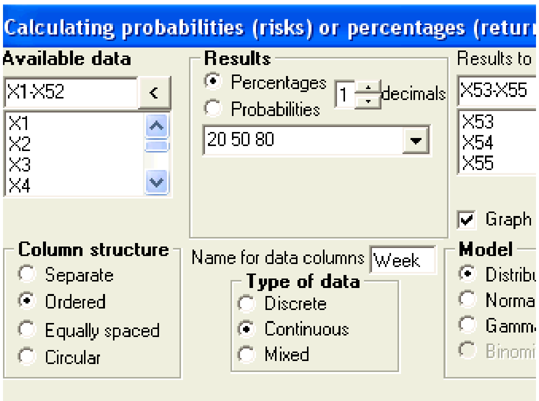
The **Climatic** ⇒ **Process** dialogue is first illustrated for the same examples considered in Section 7.2.

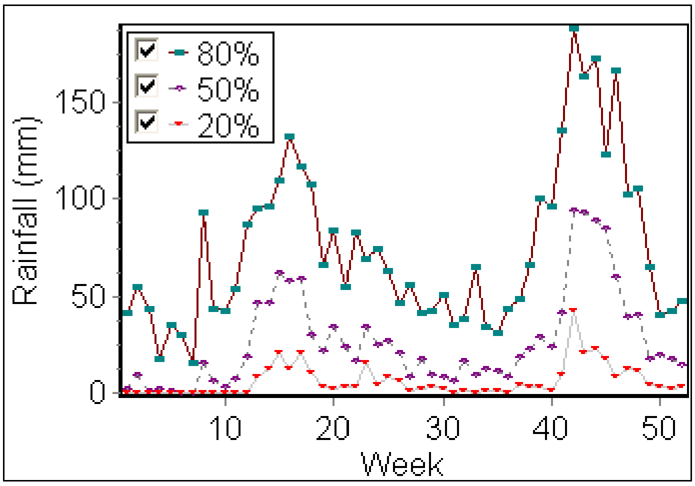
The first example gives the 20%, 50% and 80% of the rainfall totals, but this time with the weekly totals from Kurunega in Sri Lanka. Use the file called **kurun7.wor** that has these totals in x1-x52, as shown in Fig. 7.1b.

The **Climatic** ⇒ **Process** dialogue is shown in Fig. 7.4a, together with the resulting graph.

Fig. 7.4a Calculate percentage points

File ⇒ Open ⇒ Kurun7.wor
Climatic ⇒ Process



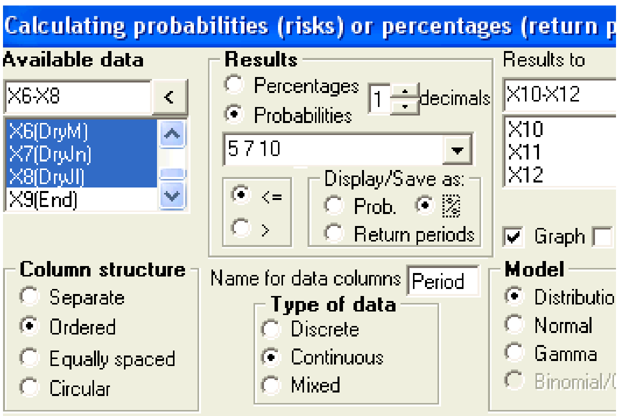


The steps described in Section 7.2, are all handled automatically, from this dialogue. It is designed to produce tables and/or graphs of either percentage points (return periods) as above, or probabilities (risks).

There is less need for this special dialogue for the second example, on the chance of a long dry spells in May, June and July from the worksheet **samrain.wor**. As in Section 7.2 define 'long' as greater than 5, 7 or 10 days. The dialogue and results are in Fig. 7.4b.

Fig. 7.4b Risk of a long dry spell

File ⇒ Open ⇒ samrain.wor
Climatic ⇒ Process

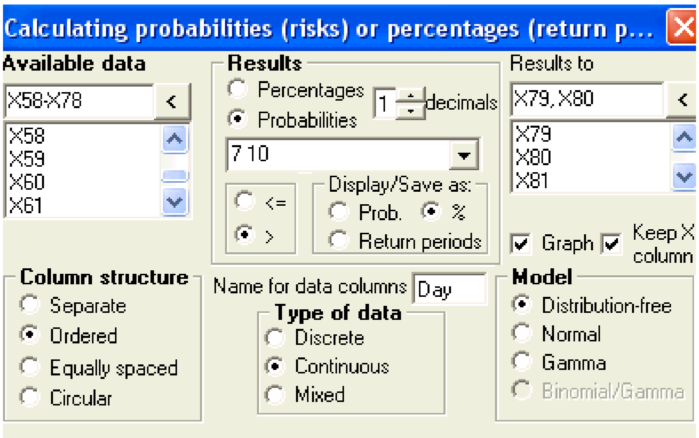


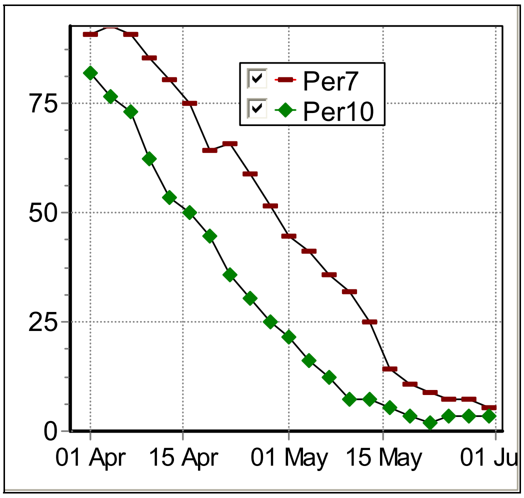
Period	Per5	Per7	Per10
1	5.4	35.7	58.9
2	62.5	92.9	96.4
3	62.5	89.3	98.2

The final example is that of the conditional spell lengths. Section 7.2 described the four stages to produce the plot shown in Fig. 7.2i. The **Process** dialogue still needs the first stage of producing the 'events', as described in Chapter 6, Section 6.4. Then the **Climatic ⇒ Process** dialogue is used as shown in Fig. 7.4c to produce the plot directly.

Fig. 7.4c The conditional risk of dry spells

File ⇒ Open Worksheet ⇒ samaru56
Climatic ⇒ Process





The **Climatic ⇒ Process** dialogue is designed to simplify the process of analysing the 'events' that have been generated using the methods described in Chapters 5 and 6. Often the need is to estimate percentage points (return periods) or probabilities (risks) and the dialogue provides a structure for the alternative analyses that are possible.

Consider first the structure of the data within each column, i.e. for each event. They might be spell lengths, or rainfall totals, or dates for the start and end of the season.

Section 7.2 and all the examples in this section so far have used a *distribution-free approach* to the analyses. This approach is applicable for all types of data. An alternative would be to assume that the data for a given event follow a *normal distribution*. For example annual rainfall totals and even monthly rainfall totals from the middle of the rainy season seem to be approximately normally distributed. The use of the normal model is described in Chapter 11, Section 11.2.

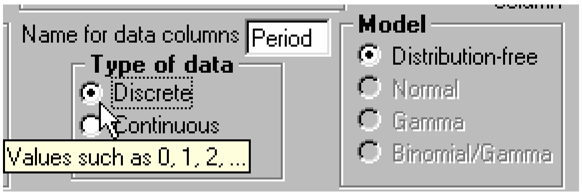
Chapter 11, Section 11.3, introduces the *gamma distributiong*, which is often a suitable model for rainfall totals if shorter periods than a month are considered. However both the gamma and normal distributions are models for continuous data and not for situations where the whole period may be dry, i.e. when there are zeros in the data. When the total may be zero, one possibility is to use the distribution-free approach. The alternative is to separate the problem into 2 parts, which we have called the *binomial/gamma model*. The first part of the model considers the chance of rain (binomial). Then the second part examines the distribution of the rainfall amounts in those years in which there was rain.

Fig. 7.4d shows the types of model that can be fitted for the different events that have been generated in Chapters 5 and 6. If the type of data is **discrete** then the dialogue only allows a distribution-free approach. Spell lengths, of the sort found in Chapter 6, Sections 6.4 and 6.5 are an example of discrete data. In contrast rainfall totals are **continuous**, i.e. they can take any value. Then a normal, a gamma model, or the distribution-free approach is possible.

We have termed **mixed**, the type of data that is sometimes continuous and sometimes zero. In this case either use the binomial/gamma model or, as usual, adopt the distribution-free approach.

Fig. 7.4d Process dialogue: 3 types of model and data

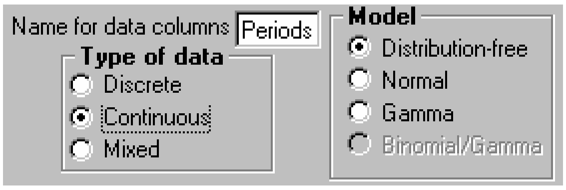
Discrete



Samaru spell lengths

	X32	X33	X34
	Sp_May	Sp_Jun	Sp_Jul
1	13	7	5
2	8	5	3
3	8	3	13
4	14	4	6
5	8	3	3
6	9	3	4
7	6	6	4
8	24	6	4
9	13	4	7
10	12	3	4
11	9	5	7

Continuous

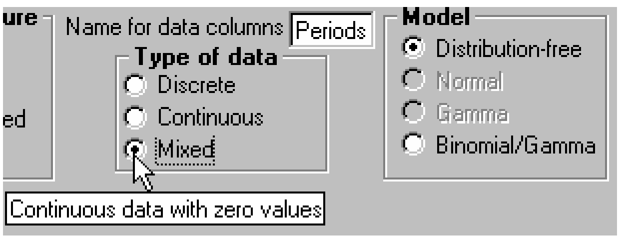


Samaru monthly totals

Samaru monthly totals for 1928-1983

X5*	X6*	X7*
May	Jun	Jul
182.09	245.36	235.71
210.82	110.98	285.24
115.56	89.67	228.35
148.85	210.29	289.8
226.57	252.23	138.69

Mixed



Rainfall totals

X25	X26	X27
Jan_T	Feb_T	Mar_T
0	0	0
0	0	0
0	10.67	0
0.25	0	0
0	0	0
0	0	8.89
0	0.76	0.76
0	0	0
0	0	0.76
0	6.1	14.99

In some cases the type of the data is not so clear-cut. For example the normal model may sometimes be used for events such as the start of the rains. It is often of little concern whether these data are taken as discrete (whole dates) or continuous (large rainfalls can be at any time of the day). But when a proportion of the years has the same value, usually the earliest possible date, then this model is not appropriate.

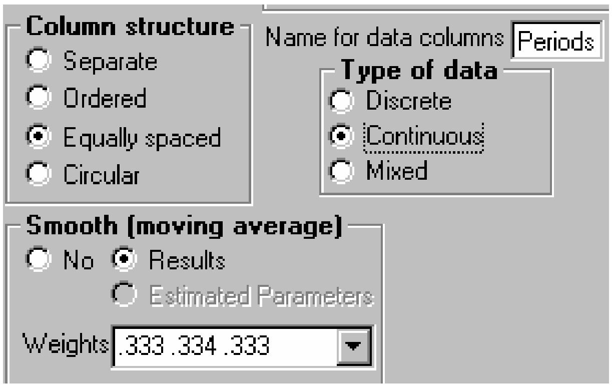
The options in this dialogue are models we have found to be useful, but they are not the only possibilities. For example Chapter 11 considers a different model for extreme values and a wide range of distributions are available to model both discrete and continuous data. For example the geometric, or negative binomial distributions are popular in ecology to model discrete data.

As well as considering the structure of data within each column, or event, it is also useful to consider the relationship between successive columns, if they are to be analysed together. In **Climatic ⇒ Process**, this is called this the **Column structure**. The column structure dictates the way graphs may be drawn.

The simplest is for the events to be **Separate**, like alternative definitions of the start of the rains. In this case any graphs will just be point plots. Sometimes the events have a natural ordering, like definitions of the start that are progressively more demanding. Then graphs have both points and lines (Fig. 7.3b).

With the yearly rainfall totals for different months in Fig. 7.3f the data are also (roughly) equally spaced, at monthly intervals. Then the line graphs permit the analysis to include an element of smoothing, see Fig. 7.4e. Smoothing is discussed further in Chapter 12, Section 12.2.

Fig. 7.4e Process dialogue for smoothing



The final possibility is called **circular** and simply affects the way the smoothing is done. It allows for situations, like the weekly data analysed in Fig. 7.2b from Kurunegela. In this case the 52 columns are **circular** in that X52 is the last week of December, and is also just before x1, which is the first week in January.

7.5 Conducting an effective analysis

Earlier, users' lack of confidence in statistics meant that they sometimes did not exploit climatic records fully. Now, with statistical software in Windows, it has become much easier to analyse the data. The **Climatic ⇒ Examine** and **Process** dialogues are designed to make the analysis as simple as possible. Both these dialogues simply collect appropriate methods that are available individually in the **Statistics, Graphics** and **Manage** menus.

The danger now is that inexperienced users may be limited to the analyses that use these dialogues. Our aim in this section is similar to that in Sections 5.6 and 6.7. Users should always consider the analysis that is appropriate, given the objectives of the study. If that coincides with an easy-to-use dialogue, then so much the better. But if not, then they should have the confidence to undertake the appropriate analyses.

If the analysis using the **Climatic ⇒ Examine**, or **Climatic ⇒ Process** dialogues is not appropriate, you have at least 3 alternatives:

1) Use these dialogues, with the option to keep the working columns. Then modify the presentation according to your needs.

2) Use the individual dialogues instead. For example, use the methods described in Section 7.2, rather than the **Climatic ⇒ Process** dialogue, shown in Section 7.4.

3) Use the commands that are generated by the dialogues. If you have many stations to analyse, then consider constructing a macro, tailored to the analysis you require. Writing macros is described in Chapter 14.

As an example of the three methods use the first option in the **Climate ⇒ Examine** dialogue in Fig. 7.5a, to give the graph in Fig. 7.5b. Alternatively, a similar graph is given using the individual dialogues, principally the **Graphics ⇒ Plot** dialogue shown in Fig. 7.5e or through the use of the commands shown in Fig. 7.5f.

Fig. 7.5a Examine Dialogue

File ⇒ Open ⇒ sammonth.wor
Climatic ⇒ Examine

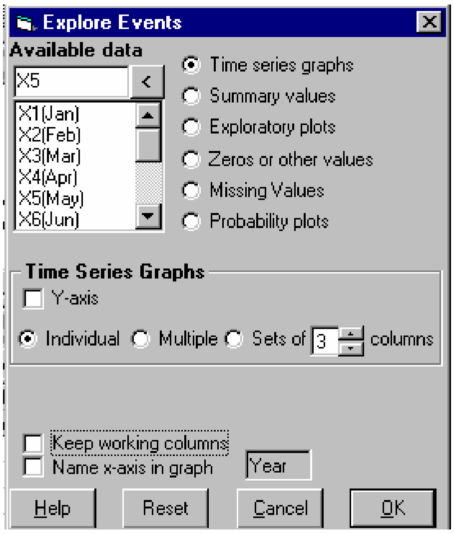


Fig. 7.5b Result

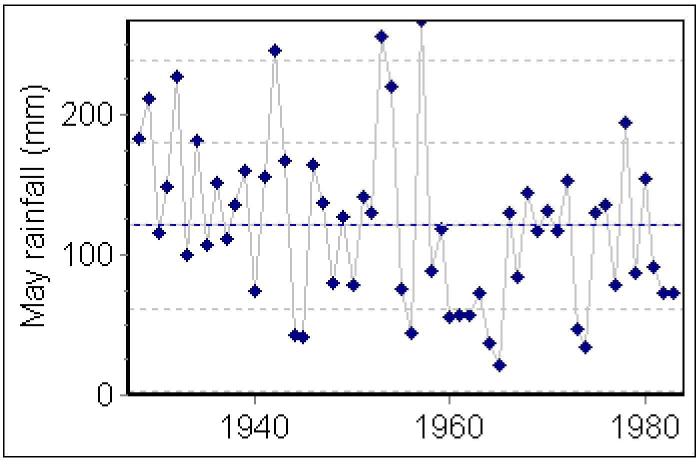


Fig. 7.5c Step by step approach

Manage ⇒ Data ⇒ Regular Sequence

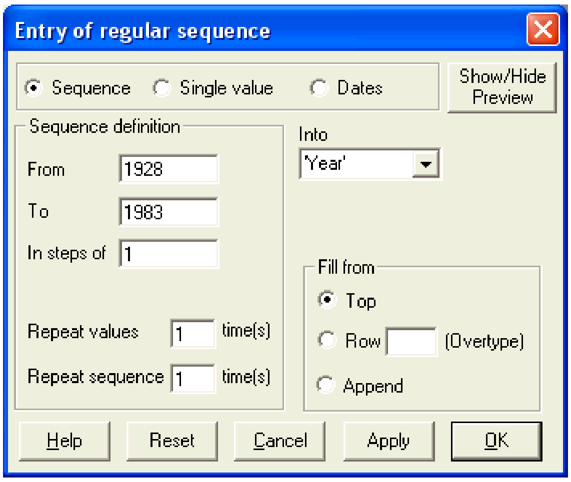


Fig. 7.5d Step 2

Manage ⇒ Calculations

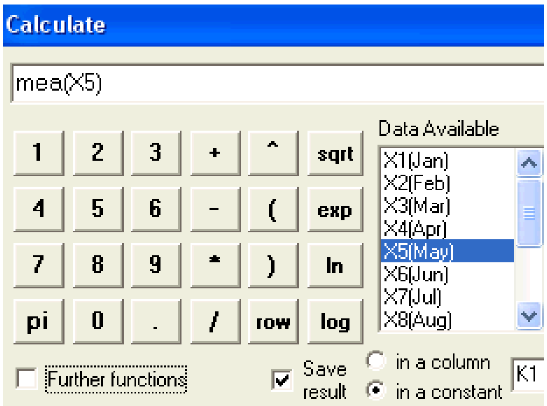


Fig. 7.5e Step 3
Graphics ⇒ Plot

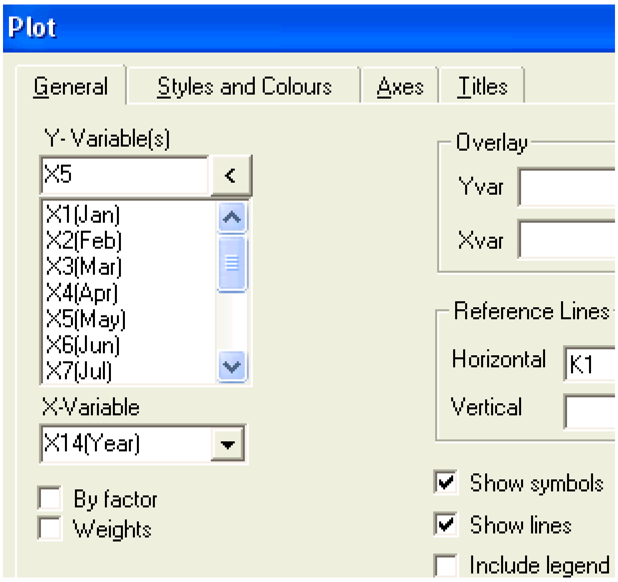


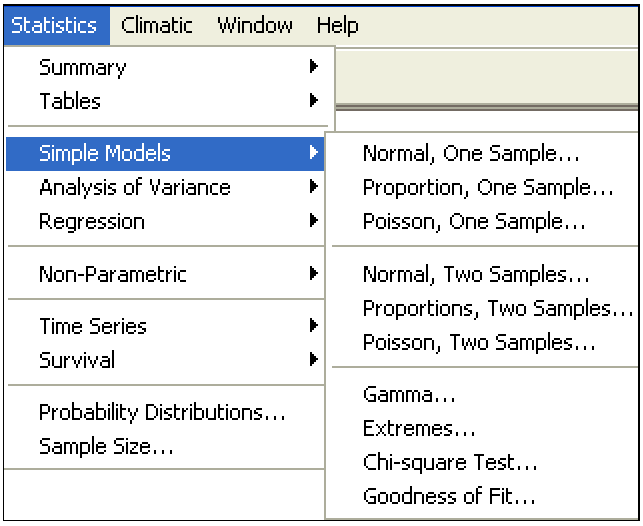
Fig. 7.5f Commands for macro

```
: enter 'Year; data (1928]1983)
: K1=mea(X5)
: line X5 1 1 67
: symbol X5 1 1 67
: plot X5 'Year; href K1;nolegend
```

In general, if the individual steps are used, the **Statistics ⇒ Summary ⇒ Column Statistics** dialogue that from Section 7.2 is the key to many analyses. That is for most of the options in the **Climatic ⇒ Examine** dialogue and also for the **Climatic ⇒ Process** dialogue if a distribution-free analysis is adopted.

If a parametric model is used in the **Climatic ⇒ Process** dialogue, then an alternative is the **Statistics ⇒ Simple Models** submenu, as shown in Fig. 7.5g. This includes dialogues for the Normal, the Gamma and the binomial (Proportion, One Sample) models.

Fig. 7.5g Simple models submenu



Smoothing, Fig. 7.5h, can improve the precision of estimates of the probabilities and percentage points. This option applies equally for the different models. Smoothing is a large topic, but the current version of Instat includes just one simple method, under the **Statistics ⇒ Time Series ⇒ Moving Average** dialogue, shown in Fig. 7.5i.

Fig. 7.5h Part of process dialogue

Climatic ⇒ Process

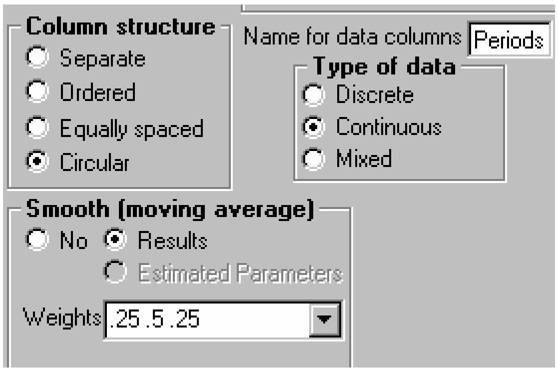
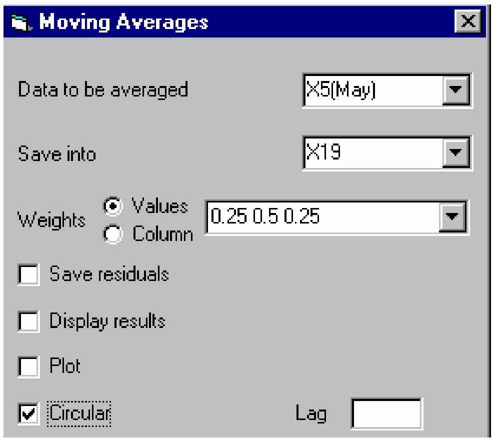


Fig. 7.5i Part of Moving-average dialogue

Stats ⇒ Time Series ⇒ Moving Average



7.6 Teaching statistical methods

In this guide we have so far used only simple statistical methods, though some ideas of modelling were mentioned in the last section. More knowledge of statistics is needed for some of the analyses in later chapters. This guide is not designed to teach statistics, but illustrations are given here on how the computer can change the way that statistical ideas can be introduced.

For illustration consider two ideas, both concerned with the normal distribution. The first concerns the start, end and length of the rainy season and the second is on the estimation of percentage points for the start of the rains. The worksheet **samrain.wor** is used. There the **Statistics ⇒ Data Summary ⇒ Describe** dialogue can be used to show that for the column 'Strt1 the mean starting date was day 125, i.e. 4th May, and the standard deviation was 16.8 days. Similarly the mean of the ending dates was day 294 (i.e. 19th October), with standard deviation of 8.1 days.

To help understand the properties of the normal distribution use the **Statistics ⇒ Probability Distributions** dialogue. Use the final option, called **Plot,** and enter the mean and standard deviation of the end of the rains as shown in Fig. 7.6a.

Fig. 7.6a Distribution for start of the rains

File ⇒ Open ⇒ samrain.wor
Statistics ⇒ Prob. Distributions

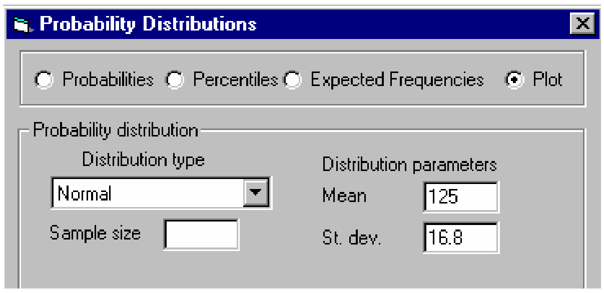
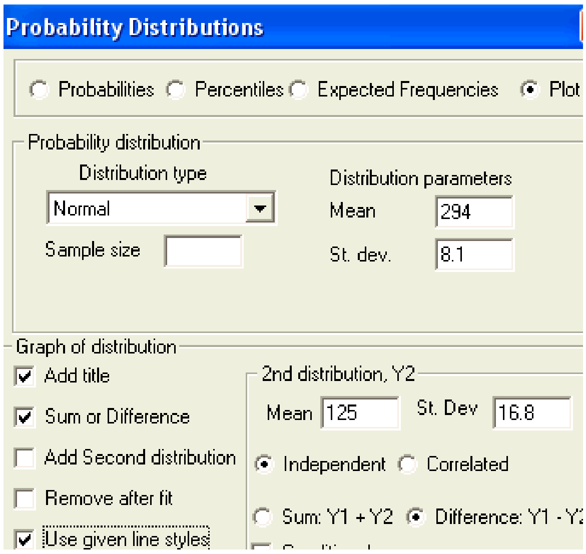
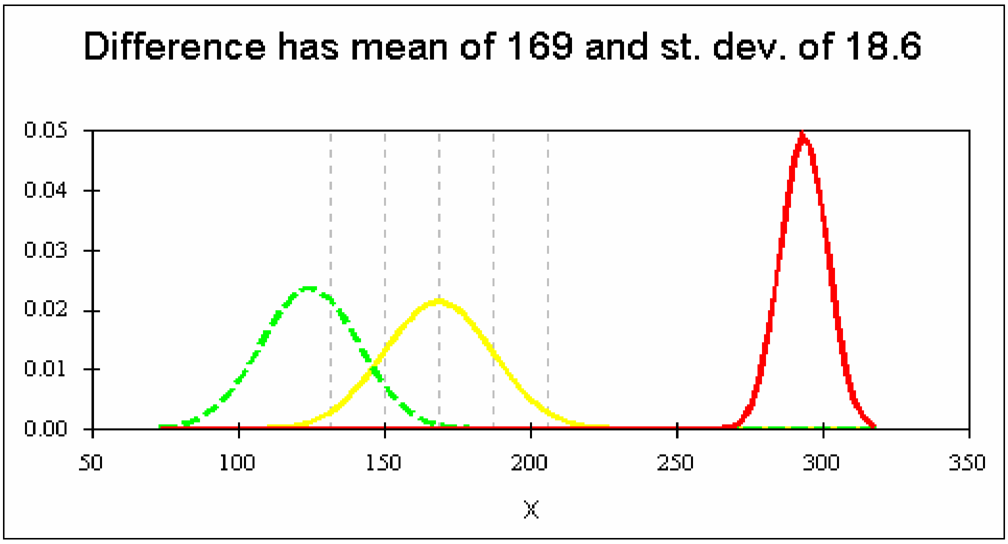


Fig. 7.6b The start, end and difference



This can be plotted as it stands, but more interesting is to plot this together with the distribution of the start. Also calculate the difference, which is the length of the season. The completed dialogue is in Fig. 7.6b and the resulting graph is in Fig. 7.6c.

Fig. 7.6c Normal plot for the start, end and difference



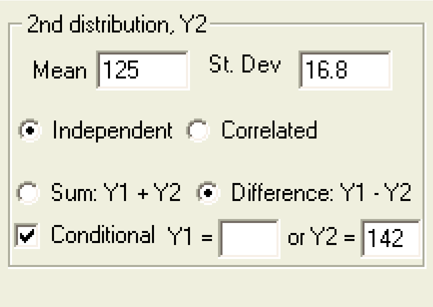
With the assumptions of a normal distribution for the start and end of the season, the difference, i.e. the lengths, are also normally distributed, with the mean and standard deviation given in the graph. This result assumes there is no correlation between the start and end. The dialogue in Fig. 7.6b allows exploration of the way the distribution of the lengths would be different if there were a correlation.

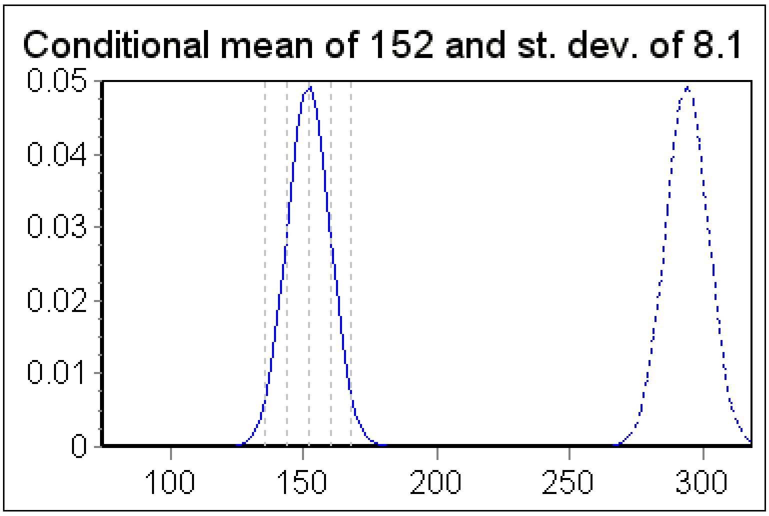
These graphs can form the basis of a discussion of the role of the normal distribution in the modelling of climatic data. In pre-computer days it was used routinely, because the analysis was computationally simple. The **Climatic ⇒ Process** dialogue shows that alternatives include a distribution-free approach, or a gamma model. In our teaching we stress that, when just a single station is to be analysed, the value of a 'parametric model', such as the normal or gamma distribution, is often modest, compared to the distribution-free approach. When data from a set of stations is available, one objective is often to compare the pattern of rainfall, etc, at the different stations. This is usually much easier, by comparing parameters (from the fitted models) than with the distribution-free approach.

A number of the dialogues, like spells, have considered the idea of conditional lengths. Here an option is to look at the conditional distribution of the length of the season, given, i.e. conditional on, a specific start date. An example is in Fig. 7.6d for a later starting date on 20th May, day 142, with the result in Fig. 7.6e. This is a clear way to show that once you know the date of planting, the variability in the length of the season is less. This is because one of the two causes of variability has been eliminated.

Fig. 7.6d Conditional length of season given the planting date

Stats ⇒ Prob. Dist





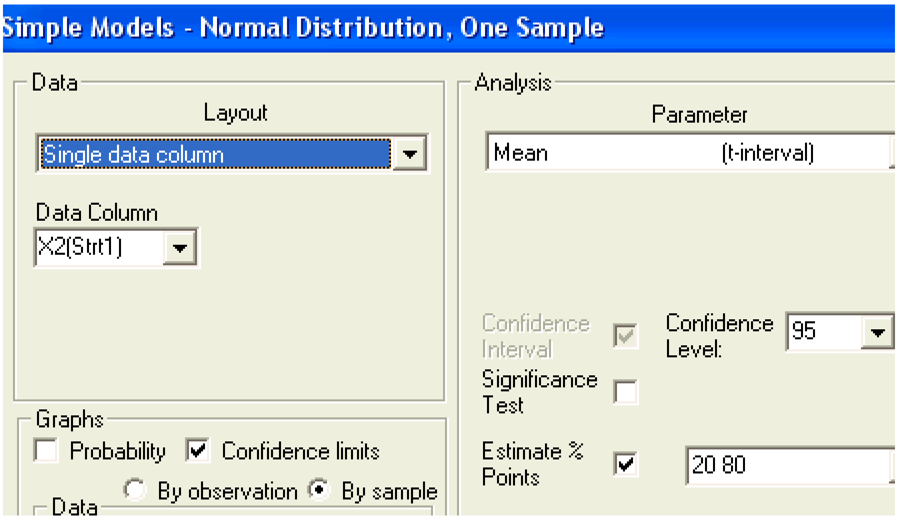
Further explanation of the ideas above is left to a statistics course, or book. Here the main point is that access to computers and suitable software can be used to support the teaching of this idea.

The **Statistics ⇒ Simple Models ⇒ Normal, One Sample** dialogue, completed as shown in Fig. 7.6e, is used for the second example. This is for one definition of the start of the rains, and estimates the same percentage points (return periods) as considered in Section 7.2.

Fig. 7.6e Normal One Sample dialogue

File ⇒ Open ⇒ Samrain.wor

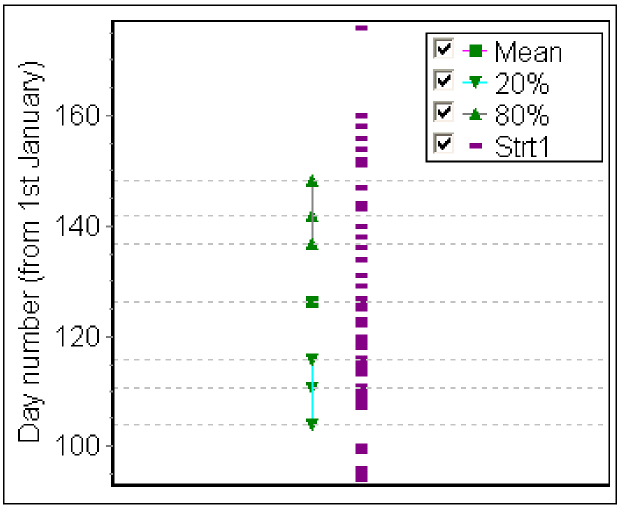
Statistics ⇒ Simple Models ⇒ Normal, One Sample



The numerical and graphical results are in Fig. 7.6f. They show the estimation of percentage points using a Normal model and also the idea of confidence limits. Those who understand what is meant by the 95% confidence limits for the 80% point (or 5 year return period), which here is day 137 to 148, have understood some key concepts in statistics!

Fig. 7.6f Percentage points with confidence intervals

```
95% confidence interval for mean
              121.18    to    131.18

         Estimated        Approximate 95%
%        value   s.e.    confidence limits
20       110.5 2.909        104    115.7
80       141.9 2.909      136.7    148.3
Note: The estimates of % points are
  sensitive to the assumption of normality.
```



Instat has always been designed with the dual aims of supporting the teaching of statistics and the analysis of climatic data. The Windows style interface permits statistical ideas to be taught more visually, just as it has helped make the analysis of climatic data into a simpler process.

This teaching support is particularly easy for those who are already familiar with any statistics package. Under Windows style systems most statistics packages are similar to use, hence mixing them is not an onerous task. So users of a different statistics package could use Instat, to help with the teaching of basic statistical ideas. This ease of support, applies particularly to readers of this guide, who are already familiar with Instat.

More examples of the use of Instat to support the teaching of statistics are in the Instat Introductory Guide, where the ideas in this section are considered in more detail in chapters 14 and 15. There are also further examples in Chapter 11 of this climatic guide.

7.7 Conclusions

By this stage users should have found that the analysis of (daily) climatic data is straightforward. They have seen that access to the daily records permits a wide variety of events to be studied and these events can be tailored to the needs of different clients.

This guide also mentions a wide range of statistical ideas, but it is not intended as a text on statistics. Statistical methods and climatic analysis have one point in common, namely that the ideas are needed by users who work in many application areas. These include agriculture, health, water and many industries, such as power and building.

Hence perhaps this guide can be used in two complementary ways. The first is to help users to incorporate climatic analyses for their area of application. The second is to support the teaching of statistical ideas for a wide audience.

Chapter 8 – Analysing Temperature Data

8.1 Introduction

Many of the techniques used in earlier chapters to analyse rainfall data can be applied to other climatic variables. This chapter shows ways of analysing temperature data, that parallel those shown in Chapter 5, 6 and 7 for rainfall data. Temperature extremes and degree days are also considered.

One difference between rainfall and temperature analyses is that there are often multiple columns of temperatures to process. For example you may wish to consider maximum and minimum temperatures together.

A second issue is the summary statistic that is used. In rainfall data it is the total, while here the maximum, mean or minimum (of say the daily maximums) might be of interest.

The wide range of ways of looking at temperature data also gives rise to an issue concerning the use of Instat or other statistical software. Chapters 1 to 7 of this guide have concentrated largely on the simple use of Instats' dialogues. Keeping the computing as simple as possible allows the user to concentrate on the climatological and statistical issues. From this chapter onwards, it will be useful if you are also prepared to learn a little about how to run Instat through giving commands and also that you are able to run "macros". These alternative ways of using Instat apply equally to the use of other statistics packages. Section 8.3 introduces the idea of running a "macro", and further macros are used in later sections of this chapter. Chapter 14 considers the use of macros in more detail.

We hope you will be pleasantly surprised by the ease of these other ways of running Instat. The alternative is to use the dialogues repetitively in a way that becomes very time-consuming and boring. It is also illogical to do repetitive tasks yourself, because computers were designed to handle repetitive calculations!

8.2 Temperature analysis

Temperature is a critical determinant of plant and animal growth and survival. For all organisms, we can determine upper and lower lethal temperatures, in addition to upper and lower critical temperatures. These critical temperatures are those limits that cause distinct, though sub-lethal responses - for example anther sterility in wheat at temperatures above 28°C, tomato fruit shedding at temperatures below 5°C, or cessation of feeding in sheep above 38°C. Between these limits, temperature conditions are often viewed as 'favourable' and little consideration is given to temperature alone, particularly in tropical areas where temperature rarely changes markedly. However, temperature analysis can be important in many situations where crops, livestock, stored products, pests and diseases are all affected by the temperature conditions.

In cool, temperate areas, agrometeorologists are most concerned with low temperatures and the determination of the length of the growing season by spring and autumn frosts. In warm, temperate and subtropical conditions, there can be concern over occasional damaging low temperatures, as well as with excessive temperatures, usually occurring at the same time as lack of water and high evaporative demand. In tropical areas, temperatures can be too low for certain agricultural practices, and excessive temperatures are always a concern.

The most common form of climatic data, available for many locations, are summaries of monthly averages of temperature. While they are useful for comparing one location with another and to show seasonal patterns, these summaries are too long and typically out of phase with critical stages of crop development and essential field operations. Plant growth and development are more closely related to critical values in the rise and fall of seasonal temperature, precipitation and daylength patterns and their interactive effects. They are also related to accumulations of temperature between phenological stages and to conditions when essential operations, such as planting and harvest must be performed. Daily temperatures are therefore of much more value.

While daily screen minima and maxima are generally readily available, grass minima or soil temperatures at various depths may be the more appropriate measure in certain conditions.

Not only are the daily values of temperature and extremes important, so is the timing relative to agricultural activities. Accumulated temperature (or 'growing degree-days') are commonly used to analyse temperature patterns, as the development processes of many organisms are approximately linearly related to the accumulated temperature above a threshold or critical temperature. Given the long-term records for a site, or for an area, it is possible to calculate mean growing degree-days above a certain threshold temperature over the whole season, or for any particular period (e.g. where perhaps growth is determined by water supply). Also, either by examining the distribution of the observed accumulated temperature or assuming that degree-days are distributed normally, probable values can be usefully combined with weather forecasts for certain agricultural activities.

8.3 Preliminary analysis

A worksheet of Niamey daily temperature data was used briefly in Chapter 4. Use **File ⇒ Open Worksheet** or **File ⇒ Open from Library** and select Ntemp.wor. The worksheet contains daily Tmin and Tmax data in pairs of columns for 20 years, from 1961-1980.

This layout of the data, shown in Fig. 8.3a, is the same as for the rainfall, but it is less convenient for temperatures, because there are two elements. An alternative is also shown in Fig. 8.3a and Table 8.1. Both these layouts will be used.

Fig. 8.3a Alternative layouts for temperature data

Current Worksheet - NTEMP.WOR

Columns | Constants | Strings | Labels | Title: Ni

	X1	X2 - F	X3	X4	X23	X24
	Day	Month	Tmx61	Tmx62	Tmn61	Tmn62
1	1	Jan	31.0	32.8	14.9	11.7
2	2	Jan	31.7	31.2	16.0	14.0
3	3	Jan	31.2	30.4	17.4	13.6
4	4	Jan	31.2	30.5	14.6	12.0
5	5	Jan	32.6	31.0	16.0	11.0
6	6	Jan	33.0	32.0	15.0	11.5
7	7	Jan	34.2	33.0	18.1	11.6
8	8	Jan	34.2	34.2	16.6	12.1
360	360	Dec	30.0	35.1	12.2	15.4
361	361	Dec	32.2	35.7	12.0	16.6
362	362	Dec	33.3	36.4	14.0	16.4
363	363	Dec	34.3	36.6	13.8	17.4
364	364	Dec	34.1	36.0	13.8	16.9
365	365	Dec	33.0	33.6	12.2	18.2
366	366	Dec	33.3	32.2	13.3	14.6

Current Worksheet - NTEMPL.WOR

Columns | Constants | Strings | Labels

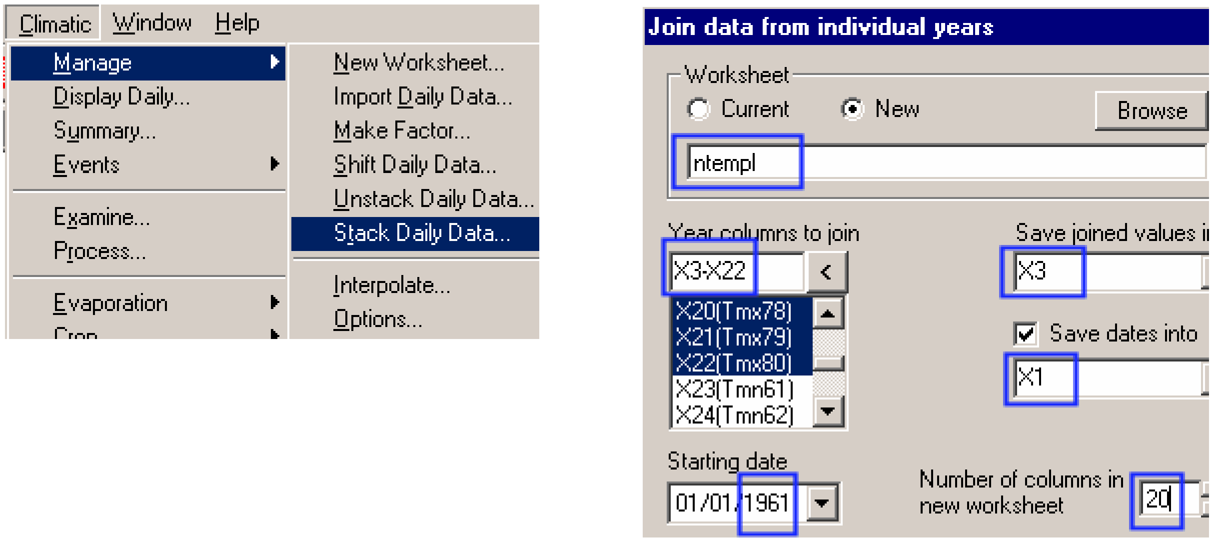
	X1	X2 - F	X3	X4
	Date	Month	Tmax	Tmin
1	01-Jan-61	Jan	31.0	14.9
2	02-Jan-61	Jan	31.7	16.0
3	03-Jan-61	Jan	31.2	17.4
4	04-Jan-61	Jan	31.2	14.6
5	05-Jan-61	Jan	32.6	16.0
6	06-Jan-61	Jan	33.0	15.0
7	07-Jan-61	Jan	34.2	18.1
8	08-Jan-61	Jan	34.2	16.6
7300	26-Dec-80	Dec	25.7	11.0
7301	27-Dec-80	Dec	26.2	12.0
7302	28-Dec-80	Dec	29.4	13.2
7303	29-Dec-80	Dec	32.4	11.5
7304	30-Dec-80	Dec	33.0	13.2
7305	31-Dec-80	Dec	33.7	15.0
7306				

Table 8.1 Alternative data layouts

"Short" – one column per year Worksheet: ntemp.wor, (length 366)		**"Long" – one column per element** Worksheet: ntempl.wor (length 7305)	
Column	**Contents**	**Column**	**Contents**
X1	day number (1 - 366)	X1	Date (I Jan 1961 - 31 Dec 1980)
X2	Factor for the months	X2	Factor for the months
X3 - X22	Tmx for 1961-1980	X3	Tmax for 1961 - 1980
X23 - X42	Tmn for 1961-1980	X3	Tmin for 1961 - 1980

Both layouts are useful, so the process of generating one from the other is described. The relevant menu is **Climatic ⇒ Manage**, shown in Fig. 8.3b. Starting from the **ntemp.wor** file, the **Climatic ⇒ Manage ⇒ Stack Daily Data** dialogue is used, as shown also in Fig. 8.3b.

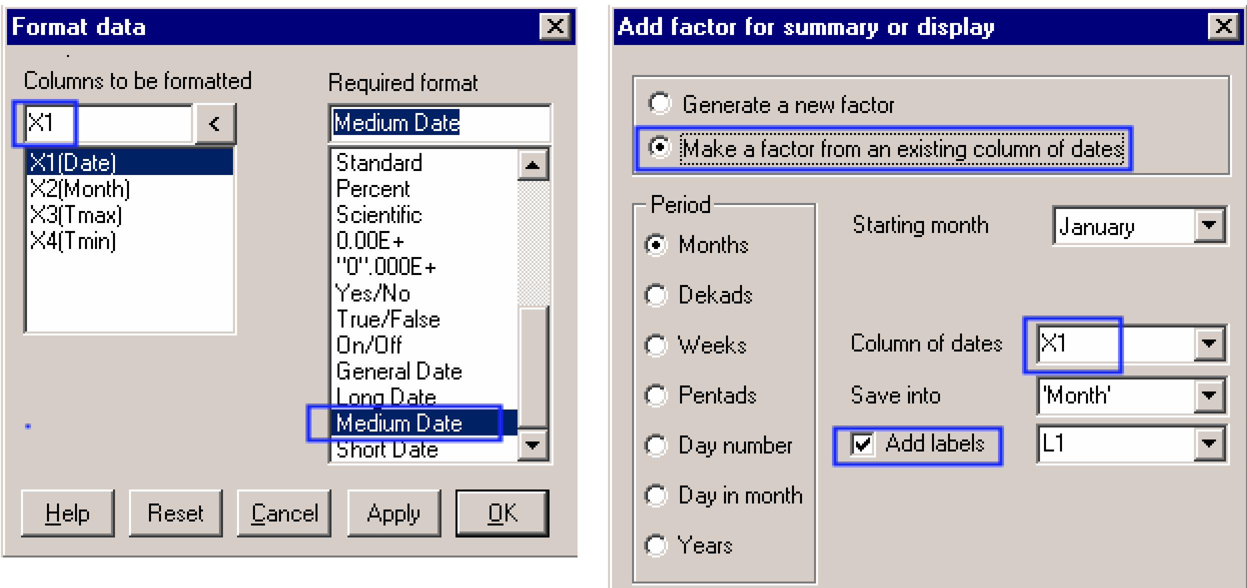
Fig. 8.3b Changing the layout of the daily data



A second use of the same dialogue puts the minimum temperatures into X4. The **Manage ⇒ Column Properties ⇒ Format** dialogue, Fig. 8.3c can then be used to format the date column.

The **Climatic ⇒ Manage ⇒ Make Factor** dialogue can be used with the new layout as shown in Fig. 8.3c. It has been used to generate a month factor, but could alternatively make dekads, etc.

Fig. 8.3c Displaying in date format and adding a factor column



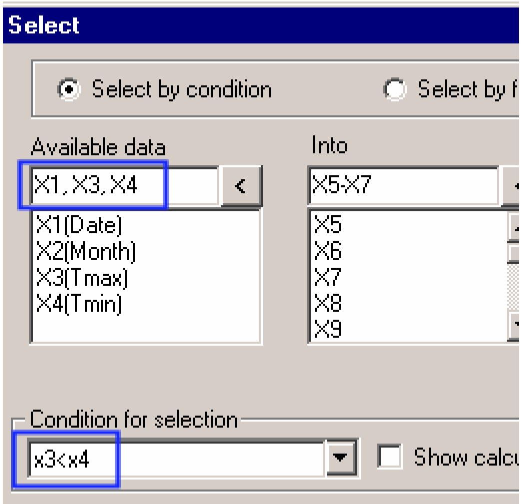
The converse is to start with data in the long format, and transform into columns of length 366. This uses the **Climatic ⇒ Manage ⇒ Unstack Daily Data** dialogue.

Quality control checks could be performed, for example to check whether Tmax is ever less than Tmin. This can be done on by **Manage ⇒ Reshape ⇒ Select** as shown in Fig. 8.3d.

Fig. 8.3d Checking temperature data are sensible

File ⇒ Open and select **ntempl.wor**

Manage ⇒ Reshape ⇒ Select



With condition for selection x3 < x4+4

X5	X6	X7
10-Aug-61	24.2	20.5
19-Jul-64	24.4	21.4
09-Aug-64	26.0	24.0
11-Aug-65	26.2	23.0
01-Sep-65	25.2	22.4
12-Sep-67	23.0	19.8
03-Aug-74	23.2	20.9
30-Apr-76	34.9	31.0
08-Aug-77	26.5	23.0
27-Jul-78	25.6	22.1

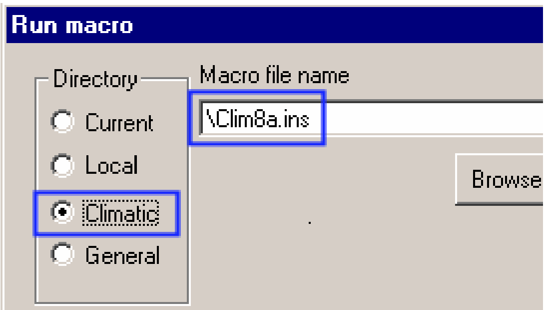
This copies the day numbers in the year for any days for which Tmin > Tmax and then shows in the **Commands and Output** window the number of rows that were selected. It is desirable that no data are selected and this was the case.

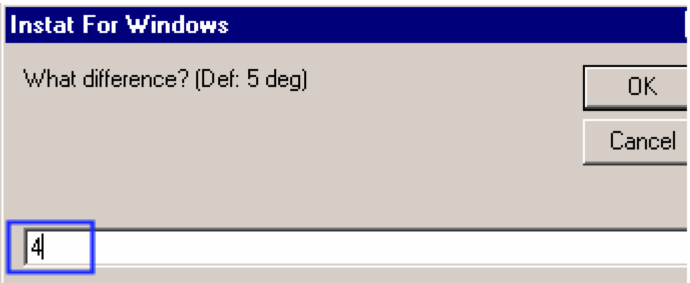
This basic quality control check would normally be performed when entering the data, for example using Clicom or ClimSoft. In a statistics package the checks can be taken further. Instead of just checking whether Tmin is greater than Tmax, the temperature range can be examined each day. For example, relaxing the condition in Fig. 8.3d to (Tmax-Tmin)<4 gave 10 occasions. The results are also in Fig. 8.3d.

A macro has been written to automate this check and to graph the data for any of the years that fail the check. It uses the data in the "short" layout, (ntemp.wor) but could equally have been written for ntempl.wor. To run it, use **Submit => Run Macro**, select the file **Clim8a.ins** from the Instat Climatic Library and click **OK**. (Fig. 8.3e)

Fig. 8.3e Running an Instat macro

Submit ⇒ Run Macro ⇒ Clim8a.ins

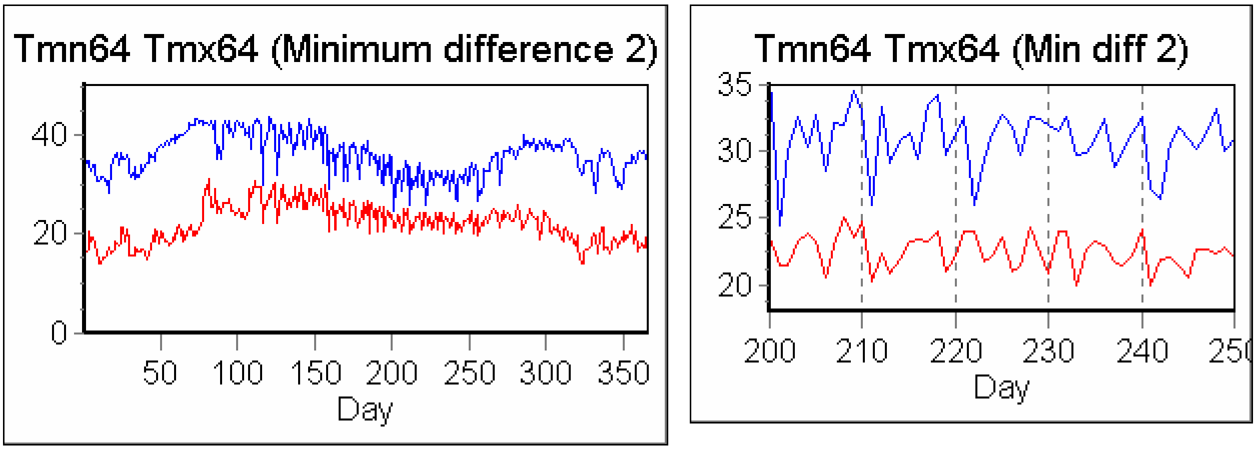




The macro asks what difference to investigate and 4 degrees have been given in Fig. 8.3e.

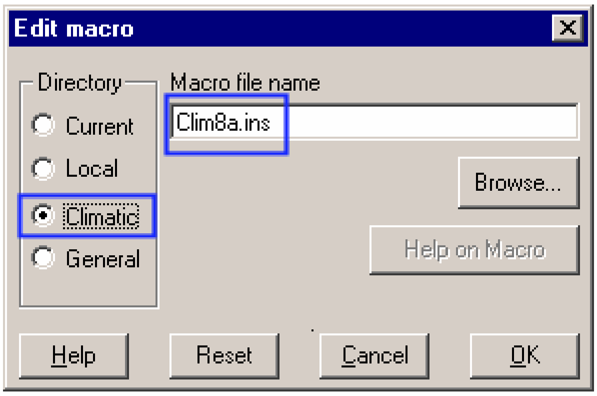
The graph for the data in 1964 is given in Fig. 8.3f, when the difference, at 2.0 degrees, was the smallest in the 20 years. It is then possible to "home in", as shown also in Fig. 8.3f, to look in more detail at the particular time of year when this occurred.

Fig. 8.3f Part of the output from running the macro



Those who wish to see the Instat commands that were used for this analysis can use **Edit** ⇒ **View / Edit Macro** ⇒ **Open** and select **Clim8a.ins,** Fig. 8.3g.

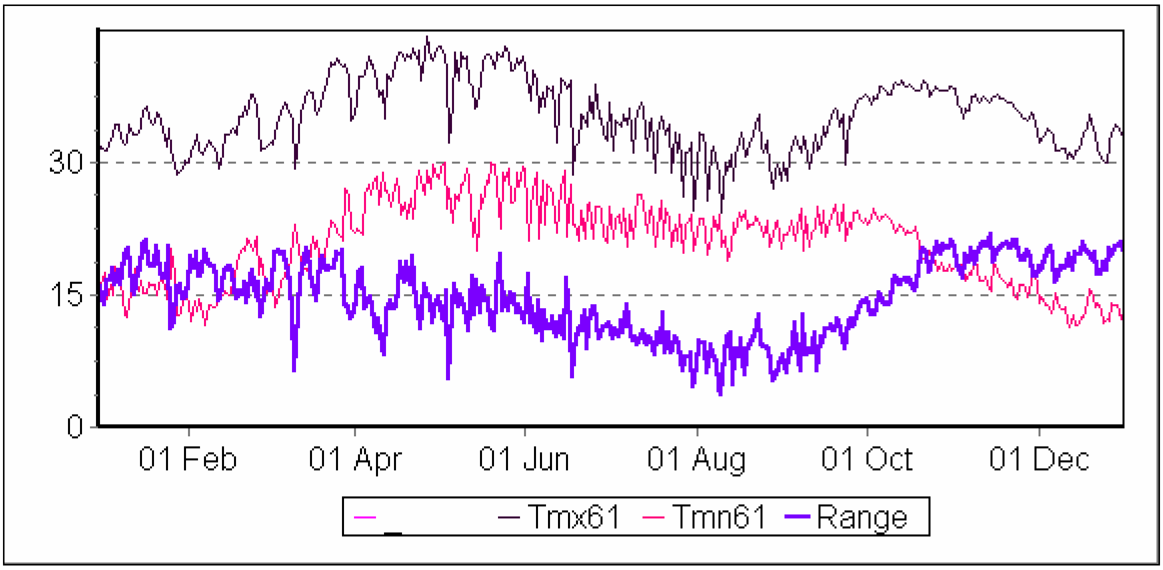
Fig. 8.3g Viewing the Instat commands in the macro
Edit ⇒ View/Edit Macro ⇒ Open and select Clim8a.ins



The first lines of the macro:

```
Note Clim_8a.ins
REStore : WARn OFF
open @ntemp
loop 1;if (narg=0)
  exit;com "par 1 5;inp";mess "What dif
loop 1
K1=%p1
%1=43 : %2=23 : %3=3
%4=num(3)
%5=%4-1:%6=%4-2
LOOp 1; REPeat 20
```

Fig. 8.3h Plot of temperature maximum, minimum and range
File ⇒ **Open** and select **ntemp.wor**
Manage ⇒ **Calculate** put **x3-x23** into **x43** (call it **'Range**)
Graphics ⇒ **Plot**



For some analyses, it might be useful to calculate the temperature range (i.e. max-min) and plot the min, max and range on the same plot. Fig. 8.3f gives an example for 1961 of the Instat commands that could be typed and run from the Commands and Output window.

8.4 Temperature events

Daily temperature records can be examined for the occurrence of average, low (or high) temperature conditions. Thus the analyses from Chapters 5 and 6 can be repeated to derive temperature events. They can then be processed to give the, probabilities of occurrence and specified return periods as shown in Chapter 7. Examples include assessing the risk of frosts to high value fruit production in temperate or high altitude tropical localities, or planning cropping calendars for cold-sensitive crops such as rice, where temperatures below 15°C affect development and below 12°C cause permanent physical damage.

8.4.1 Monthly summaries

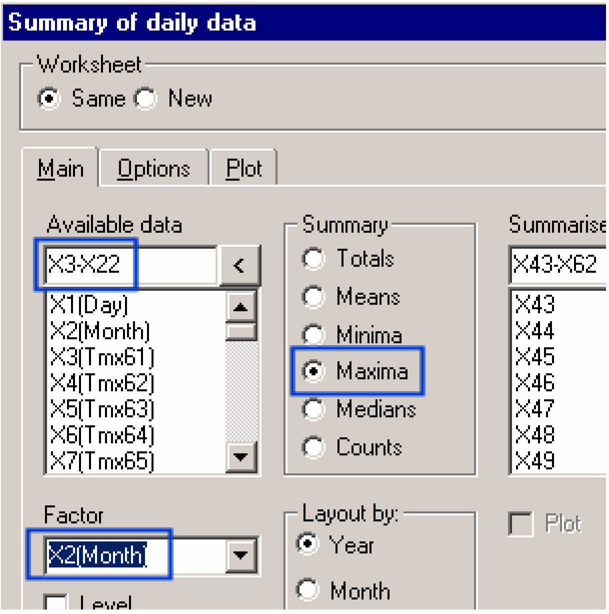
The first calculations might be monthly summary statistics from the maximum temperatures. If the maximums of the maximum temperatures, within each month are required, use the **Climatic ⇒ Summary** and complete the dialogue as shown in Fig. 8.4a.

Fig. 8.4a Maximum temperatures for each month

File ⇒ Open and open the worksheet **ntemp.wor**

Manage ⇒ Data ⇒ Clear(Remove) and remove **x43-x80**

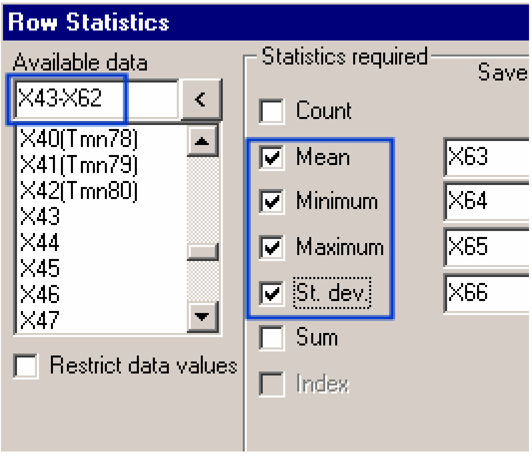
Climatic ⇒ Summary



	X43	X44	X45	X46
1	36.2	35.5	39	36.4
2	37.6	40.5	41.6	39.8
3	41.7	41.2	41.4	43
4	44.2	42.8	43.2	43.4
5	43	42.6	42.4	43
6	40.8	42.2	43.2	41.5
7	36.8	39.6	40	38.6
8	35.4	35.6	37.4	34.2
9	37.6	36.3	38.5	36.8
10	39.2	39.7	40.2	39.9
11	38.4	39.4	39.5	39
12	35.4	37	37.7	37
13				

For example, from Fig. 8.4a the monthly January maximum was 36.2 degrees in 1961 and 35.5 degrees in 1962. The Row dialogue, i.e. **Manage ⇒ Manipulate ⇒ Row Statistics** (Fig. 8.4b) can be used to summarise these monthly maximums.

Fig. 8.4b Summarising monthly maxima data
Manage ⇒ Manipulate ⇒ Row Statistics



X63	X64	X65	X66	X67 - F
MeanTmax	MinTmax	MaxTmax	SdeTmax	Mth
36.7	33.6	39.3	1.30	Jan
39.8	37.2	41.6	1.33	Feb
41.9	39.4	43.6	1.21	Mar
43.4	42.0	45.0	0.85	Apr
43.2	42.2	45.2	0.86	May
41.4	38.3	43.5	1.38	Jun
38.1	36.0	40.8	1.41	Jul
36.3	34.2	38.5	1.19	Aug
38.2	34.8	40.8	1.52	Sep
39.9	38.0	41.6	0.91	Oct
38.8	37.6	40.5	0.77	Nov
36.9	35.0	39.8	1.15	Dec

An alternative way to get these summaries is to use the **Layout** by **Month** in Fig. 8.4a, and then use the **Column Statistics** dialogue. This was shown in Chapter 5 for the rainfall data.

Fig. 8.4c Graph of the monthly summaries

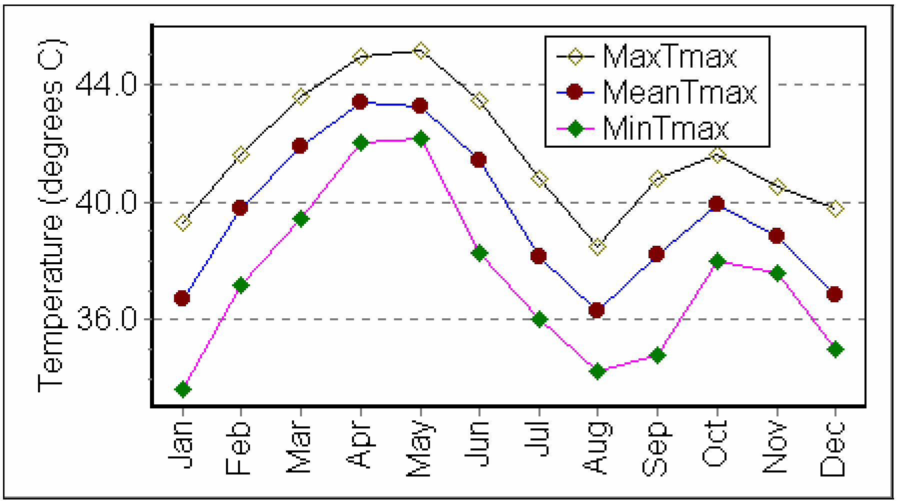


Fig. 8.4b and Fig. 8.4c show, for example, that April and May were the hottest months, when all the years had at least one day with a maximum temperature of 42°C or more (X65). The overall record high in the 20 years was 45.2°C (X66). August, December and January were the only months in the 20 years that never had a day with a maximum temperature above 40°C.

Many other summary statistics could be produced, for example for the daily minima or the daily mean temperatures. One issue, when producing reports is to be clear, exactly what summary statistics have been produced. The complication is the fact that results such as Fig. 8.4c have included 3 levels of summary, as follows:

The daily values are themselves a summary of the instantaneous temperatures

The monthly, or 10-day summary is then the second level, see Fig. 8.4a for an example.

The third level is the summary of these values over the different years of the record, such as was done in Fig. 8.4b, leading to the results in Figs 8.4b and 8.4c.

For example, from Fig. 8.4b, the value of 39.3°C in X66 was the instantaneous maximum in January over all the years. Similarly 33.6°C was the lowest daily maximum in any January in the 20 years.

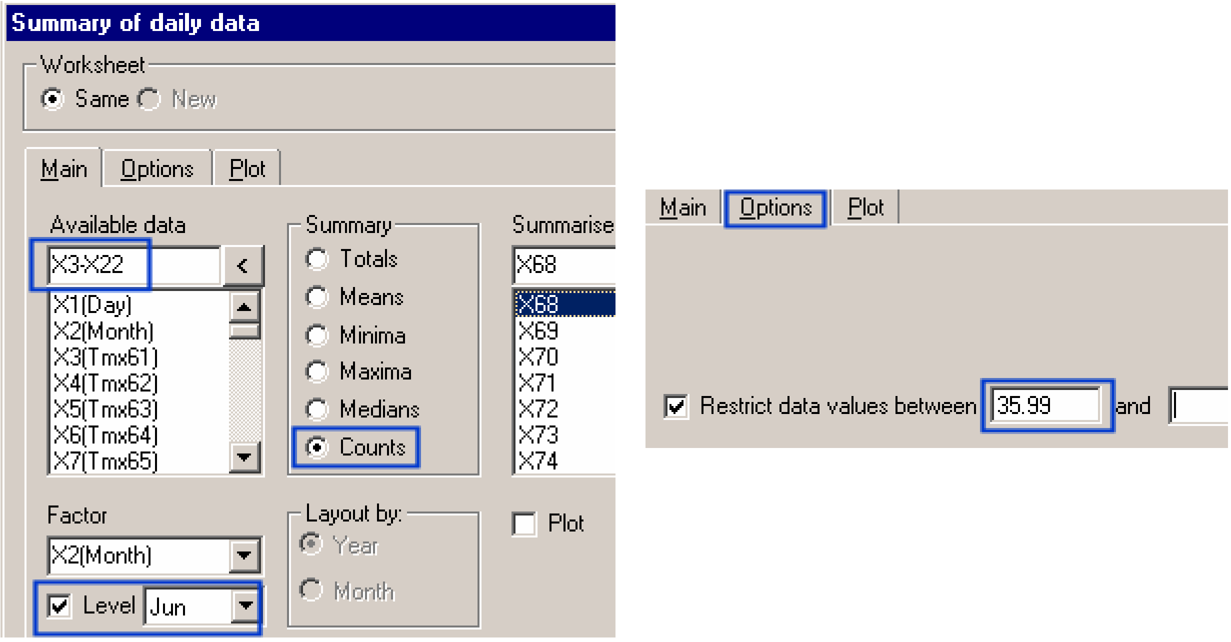
8.4.2 Frequencies at different thresholds

In Niger, high temperatures can be a problem, particularly to small seedlings in June and July. Applying mulch can reduce the soil temperature by 2°C or more. To show how this type of problem can be studied, consider how many days in June have maximum temperatures greater than 36, 38, 40, 42 and 44°C. To find the number of days greater than or equal to 36°C, complete the dialogue as in Fig. 8.4d. The option is used to restrict attention to just those values of 36 degrees or more.

The resulting column is then renamed, and the dialogue repeated for 38, 40, 42 and 44°C. The results are in Fig. 8.4e.

Fig. 8.4d Number of days ≥ 36°C

Climatic ⇒ Summary



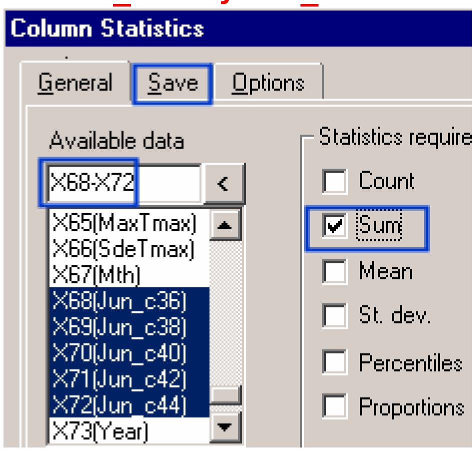
These calculations have produced the five "events" of interest, as for the rainfall data in Chapters 5 and 6. For example, in Fig. 8.4e, in 1961 17 of the 20 years had at least one day with the maximum temperature ≥ 36°C, while only 2 years had a day ≥ 40°C

We now proceed to the analysis, as in Chapter 7 for the rainfall data. To calculate the total number of days over the 20 years, use the **Statistics ⇒ Summary ⇒ Column Statistics** dialogue as shown in Fig. 8.4e. Also click on the Save tab to save the summary into X74.

Fig. 8.4e Number of days with max temp ≥ 36, 38, 40, 42, 44°C

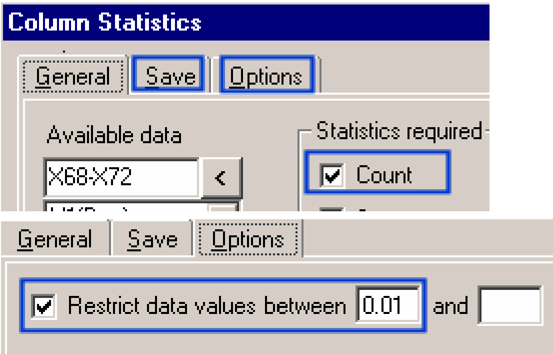
X68	X69	X70	X71	X72	X73
Jun_c36	Jun_c38	Jun_c40	Jun_c42	Jun_c44	Year
17	12	2	0	0	1961
25	12	4	1	0	1962
20	13	8	1	0	1963
20	9	4	0	0	1964
16	6	5	0	0	1965
15	7	2	0	0	1966
14	8	6	2	0	1967
12	4	1	0	0	1968
17	11	5	2	0	1969
28	28	17	3	0	1970
27	19	14	3	0	1971

Stats ⇒ Summary ⇒ Column Stats



Now use the **Manage ⇒ Calculations** dialogue, o calculate the percentage of days in June over the 20 years, that have temperatures equal or greater than 36, 38, 40, 42 and 44°C. Then the Column Statistics dialogue is used again, Fig. 8.4f, to find the number of years that have a day above the given threshold.

Fig. 8.4f Calculating frequencies at different thresholds
Manage ⇒ Calculations with **100*X74/(20*30)** into **X75**
Stats ⇒ Summary ⇒ Column Stats



	X74	X75	X76	X77
	Ndays	%Day	Nyears	Thold
1	391	65	20	36
2	231	39	20	38
3	102	17	19	40
4	20	3	8	42
5	0	0	0	44

The results are also shown in Fig. 8.4f. They indicate that the crops have to withstand a day with a maximum June temperature of more than 40°C on 17% of days and in 19 of the 20 years. There were only 20 occasions with a maximum temperature of 42°C or more and only 8 years had any days above this threshold.

To show how these results can be used, assume first a variety of millet that can withstand a soil temperature that corresponds to an air temperature of 40°C, but has problems with higher temperatures. For this variety a comparison of the results between 42°C and 40°C shows that the application of mulch would considerably reduce the number of days of stress for the plant.

However, mulching would be of less value for crops that could not withstand a maximum temperature of more than 38°C, because even reducing the temperature by 2°C would still leave almost all years with temperatures that are too high.

The idea of looking at thresholds has not been considered before in this guide, but it applies equally to other climatic elements. For example for rainfall data the chance of a rain day with more than 50mm could be studied. It is also useful to study extremes of wind speeds.

8.5 Examples of other temperature events

This section considers other temperature events, using data from Reading, England for illustration.

8.5.1 Occurrence of frost

In a temperate environment, such as Reading, the incidence of late frosts in the spring determines planting dates for tender crops. With a daily temperature record the probability of frosts can be assessed. First, define a frost as a day when Tmin ≤ 0°C. Obviously, the definition could be changed to use grass minimum temperatures (which more closely reflect the temperatures that plants are exposed to), or mean daily temperatures, or any other value of temperature of interest.

The worksheet, Rdtemp.wor, contains the data. The daily Tmin and Tmax are in pairs of columns for 10 years, 1976-1985.

X1	is the day number (1 - 366)
X3, X5, X7 … X21	are Tmin for 1976-1985
X4, X6, X8 … X22	are Tmax for 1976-1985
X2	is free, but will be used as a factor column to denote months or decades

The **select** command has been used in the macro in Fig. 8.5a to produce graphs for the cumulative probability of the date of the last Spring frost (or the first Autumn frost) occurring before (or after) June 1st - day number 153. A frost has been defined as occurring when the screen minimum air temperature ≤ 0°C, (mild frost), or ≤ -2°C (brr!).

Fig. 8.5a Part of macro to calculate cumulative probabilities of first and last frosts

```
Note Clim8b.ins
restore:warn off
remove x24- x33
%l=3
loop l; rep 10
  select xl; into x24; if x%l<=0 and row <153; last; append
  select xl; into x25; if x%l<=-2 and row <153; last; append
  select xl; into x26; if x%l<=0 and row >153; first; append
  select xl; into x27; if x%l<=-2 and row >153; first; append
   %l=%l+2
loop l
sort x24 x24
name x24 'Days
```

The results from running the macro are shown in Fig. 8.5b.

Fig. 8.5b Plots of last spring and first autumn frosts at Reading

File ⇒ Open From Library ⇒ RdTemp.wor

Submit ⇒ Run Macro ⇒ Clim8b.ins from Climatic library

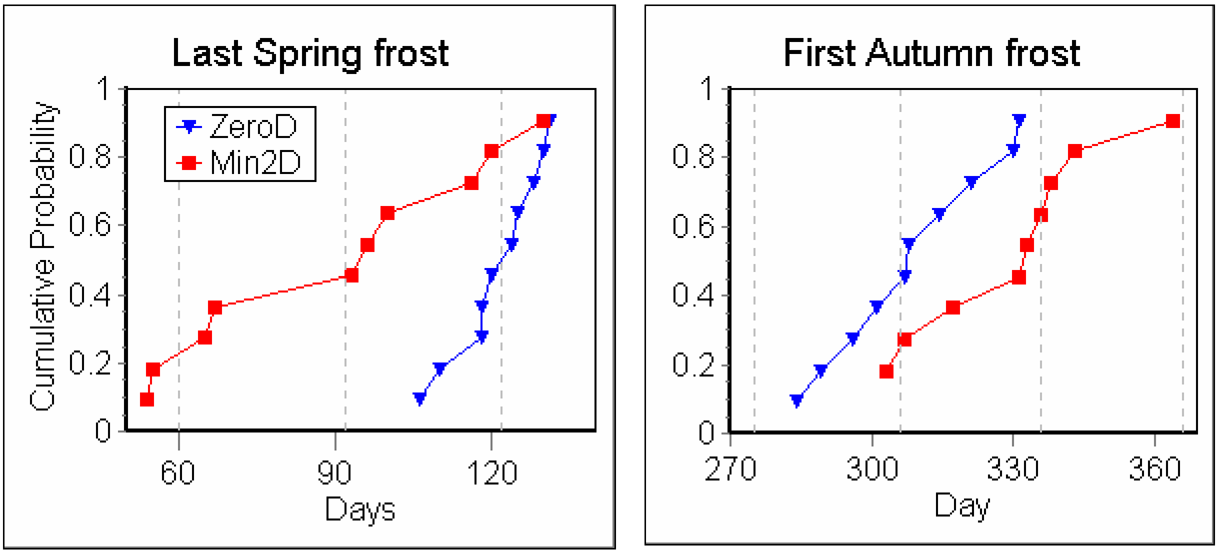


Fig. 8.5b shows that every year had a mild frost (of ≤ 0°C) as late as mid-April or May, but some years did not have a day with minimum temperatures less than ≤ -2°C after February.

After the summer, some years had frost by mid October and even temperatures of ≤ -2°C were possible by the end of October. All years had at least one mild frost by the end of November.

8.5.2 Accumulated temperatures

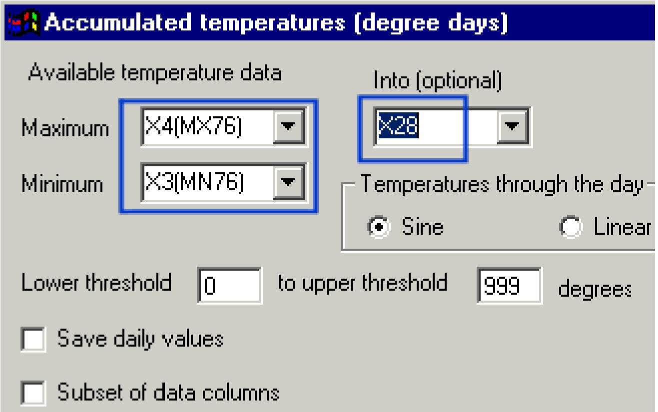
Often it is not just specific events or extremes (frosts, heat waves etc.) that are of concern, but the assessment of the overall temperature that crops or animals (livestock, pest populations, even humans) have been exposed to. The development of many crops can often be regarded as being controlled by the integrated temperature to which the plants (and some diseases) have been exposed. There is usually some upper or lower threshold temperature, outside which development through the various stages (e.g. germination, flowering, senescence) will not proceed. Other environmental factors also influence these relationships (such as day length and vernalisation requirement) but generally the relationship between accumulated temperature (sometimes called heat sum) is a useful predictor of crop developmental stage.

The same data set, RdTemp.wor, is used to analyse the variability in the accumulated temperatures in the Spring. The agricultural industry in Britain has adopted the practice that first originated in the Netherlands. This uses the occurrence of an accumulated temperature of 200 degree-days (sometimes referred to as 'Tsum 200') above a threshold of 0°C to time the optimum application of nitrogen fertilisers to grass crops in the Spring.

Instat allows the calculation of the accumulated temperatures from daily Tmax and Tmin records, using any upper or lower thresholds. An example of the dialogue is shown in Fig. 8.5c. This dialogue assumes the time course of temperature over the day follows a sine wave (Snyder, 1985) though the alternative of linear interpolation can also be specified.

Fig. 8.5c Degree day dialogue

File ⇒ Open Worksheet ⇒ RdTemp.wor
Climatic ⇒ Heat Sum

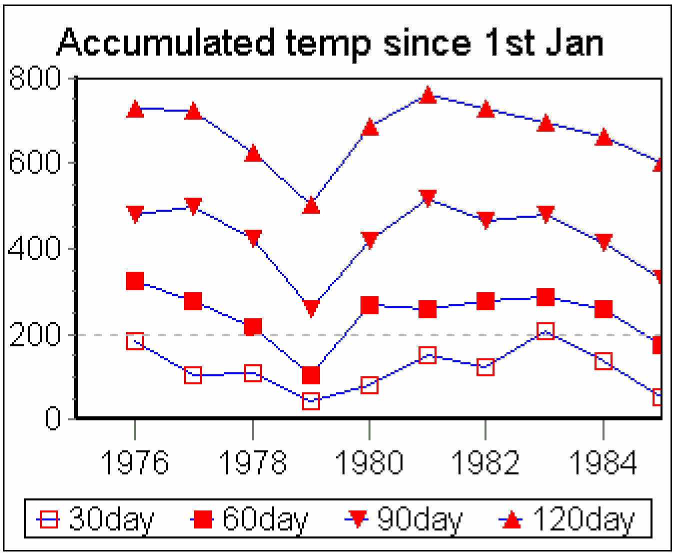


	X1	X2	X3	X4	X28
	Dayno		MN76	MX76	heatsum
1	01 Jan		2.5	11.9	7.2
2	02 Jan		3	13.4	15.4
3	03 Jan		5.2	7	21.5
4	04 Jan		-2	10.6	26.2
5	05 Jan		1.7	10.6	32.3
6	06 Jan		8.3	11.5	42.2
7	07 Jan		6.9	9.2	50.3
8	08 Jan		7.6	10.5	59.3
9	09 Jan		5.4	12	68.0
10	10 Jan		7.3	11.4	77.4
11	11 Jan		9.3	11.7	87.9
12	12 Jan		7.8	12.1	97.8

The dialogue in Fig. 8.5c uses the lower threshold of 0 and the upper of 999 (i.e. no upper restriction) and saves the sum of the temperatures in x28.

Fig. 8.5d Plots of accumulated temperatures

Submit ⇒ Run Macro ⇒ Clim8c.ins **Submit ⇒ Run Macro ⇒ Clim8d**



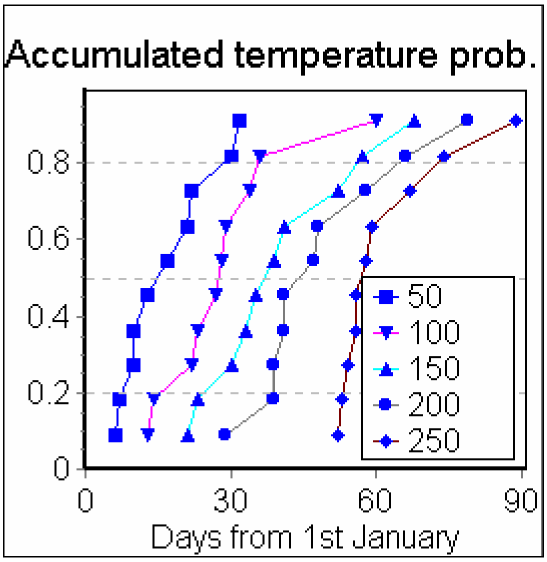


Fig. 8.5d shows results from the calculation of degree-days using a macro. The number of days has been fixed, at 30, 60, 90 and 120, to investigate the differences in degree-days on specific

dates. The results show that 1979 was a particularly cold year. The horizontal line at 200 also shows that 1983 had 200 degree-days already by the end of January, and 1976 was close to this value.

The alternative is to fix the required number of degree days and look at the distribution of different dates. These can conveniently be plotted by calculating the cumulative probabilities as was used earlier in Fig. 8.5b. The resulting plot is shown in Fig. 8.5d, for 'Tsum 50' up to 'Tsum 250'. This shows that the 'Tsum 200' mentioned above, varied from the end of January (this was in 1983) until mid-March (in 1979).

8.6 Using temperature data to model pests and diseases

Pest and disease modelling has at least three purposes. First, to help make decisions about how crops are grown and managed. Second, to estimate yields to help in government and corporate decisions about buying and selling produce. Third, to improve understanding and so suggest new or improved management methods.

Two basic techniques are available:

- To search for patterns in past records of pests and diseases and weather;
- To mimic the individual biological process and build a simulation that describes the system.

These two methods are not mutually exclusive and each may inform the other.

This section considers one example of the type of calculations that may be done to predict the optimal time to spray a crop. It still uses the idea of accumulated temperatures, but without any of the special climatic facilities in Instat. Hence these calculations can be done using any statistics package.

8.6.1 Basic biology

Crop disease may be caused by viruses, fungi, bacteria or higher plants. These either destroy plant tissues or divert photosynthate to themselves, reducing the growth of the crop. Crop pests are animals which eat parts of the crop. These include insects, nematodes and mites.

Animals mostly have ways of protecting themselves from desiccation, so temperature variables, which control how fast they grow and reproduce, tend to be more important than wetness variables.

Fungi are the most important cause of diseases of plants. Most fungi reproduce by spores, single cells adapted to survive alone for a time. These spores may be moved to new, healthy plant tissue by rain or by wind, or may lie dormant in soil until a suitable host plant grows close enough.

Fungi are mostly very vulnerable to drying out, so rain or humidity variables are often closely related to the development of fungi disease. The same is true for bacteria. When humidity or rain is suitable, temperature may have a very strong influence on fungal growth.

Viruses have to be spread by some other organism, usually an animal such as an insect or a nematode. Relationships of viruses with weather are therefore often more like those of pests, or related very closely to plant growth patterns.

8.6.2 Predict timing of insect activity

In Britain the pea moth is a pest of the legume crop, peas. The moths pass the winter as dormant pupae and emerge as adults in late May or early June. The timing of this emergence can be determined fairly easily by trapping adult males in traps baited with the sex-attracting chemical of the female. Females mate as soon as they can and begin to lay eggs within 2-3 days of emergence. The eggs mature at a rate that depends on temperature, and hatch into young larvae. The larvae find their way to the developing pods and burrow inside; they then grow inside the pea pods, destroying the crop. Larvae that develop into larvae before harvest or in wild plants, form the founders of next year's population.

Insecticide does not kill eggs, adults or larvae inside pods. It also gradually disappears once applied. It has therefore to be applied at exactly the right time, so the larvae receive a maximum dose after they hatch and before they penetrate the pods. The traps indicate to the advisors and farmers when the adults have emerged. The question is, how long after this should insecticide be applied?

The following experimental data are for pea moth eggs laid in the laboratory:

Mean temperatures:	14.3	18	20.3	21.5	25.3
Days to 50% hatch:	14.3	8.9	7.3	6.3	4.8

The problem is

a) to calculate the rate of development (i.e. the inverse of the time taken)

b) plot the rate against temperature

c) find the base temperature below which development does not proceed

d) using the slope of the regression of rate of development on temperature, find a 'temperature sum' (or accumulated temperature) above the base temperature, which corresponds to 50% hatch

e) formulate spray advice to farmers based on this temperature sum

To solve this problem the concept of accumulated temperatures, which was introduced in Section 8.5.2, is applied

If the spread of a process is directly proportional to temperature, then the time taken to complete the process depends inversely on temperature. This means that a graph of time taken to complete the process will be a curve.

Fig. 8.6a Completion time v temp

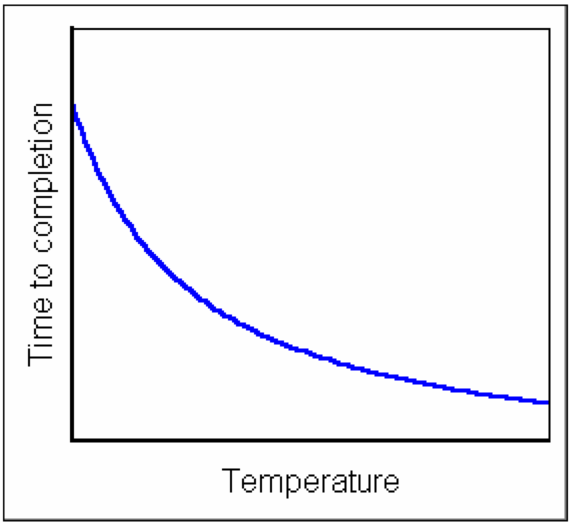
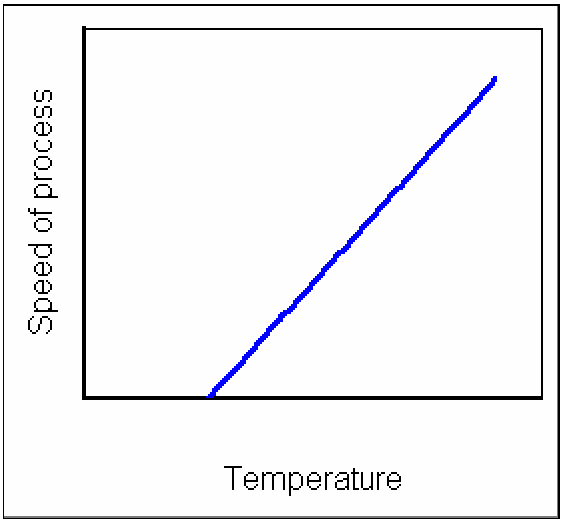


Fig. 8.6b Process speed (rate) v temp



Things are simplest to understand, if the speed of the process depends linearly on temperature above 0°C, so that the speed of process is a straight line passing through 0°C. Then, if at 10°C the process takes 1 day, it will only take half a day at 20°C. The product of temperature * time is the same in both cases, i.e 1 °C-day. This is said to be the *accumulated temperature*, or the *temperature sum*, required for the process. Things are a little more difficult if the speed of the process depends linearly on temperature, but stops at a temperature different from 0°C. In this case, the speed of the process, r, will obey

$$r = c + mT$$

where c is a constant (the intercept on the r axis), m is the slope of the fitted line and T is the temperature. This can be rewritten as

$$r = m(T + c/m) = m(T-b)$$

where *b* is a new constant, the intercept on the *T* axis, called the *base temperature*. The 'distance' the process has gone can be found by multiplying *r* by the time elapsed, *t*. (If *r* varies, integration will be needed). Therefore the distance gone is

$$\int_t m(T-b)t = m\int_t (T-b)t$$

The distance scale is arbitrary, so we can just say that when the process is complete we have travelled 1 unit. Then when the process is just complete,

$$\frac{1}{m} = \int_t (T-b)t$$

All that is known is the time to complete the process at a given temperature, which we will call *t(m)*. But the rate at which it was completed, *r*, must be 1/*t(m)*. So, plotting this rate against temperature, gives *b* and *m*. Then, given a temperature series against time, the right hand side of the last equation can be calculated. If it is greater than 1/*m*, the process is complete.

Fig. 8.6c Pea moth data

File ⇒ New Worksheet

Enter data

File ⇒ Save As ⇒ Peamoth.wor

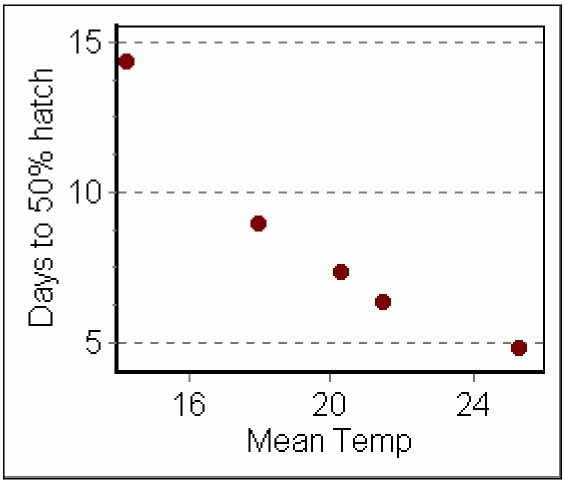
Current Worksheet - peamoth.wor

Columns | Constants | Strings | Labels

	X1	X2	X3
	Temp	Days	
1	14.3	14.3	
2	18	8.9	
3	20.3	7.3	
4	21.5	6.3	
5	25.3	4.8	
6			

Fig. 8.6d Plot of Days vs Temperature

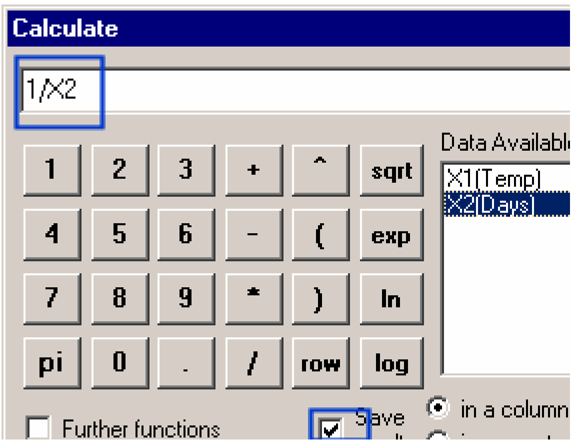
Graphics ⇒ Plot (X2 v X1)



These concepts are now applied to the pea moth problem. First use **File ⇒ New Worksheet** to create a new worksheet. Enter the data as shown in Fig. 8.6c, then save as **Peamoth.wor**.

Fig. 8.6e Plot of Rate vs Temperature

Manage ⇒ Calculations



Graphics ⇒ Plot (X3 v X1)

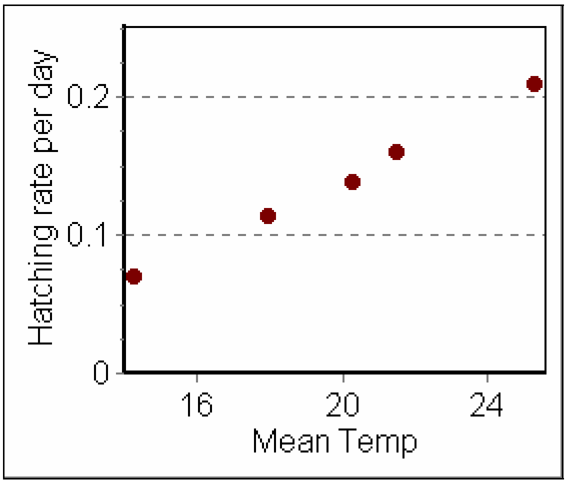
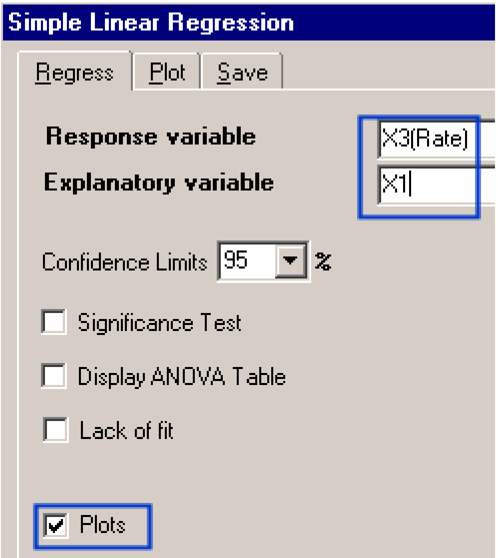


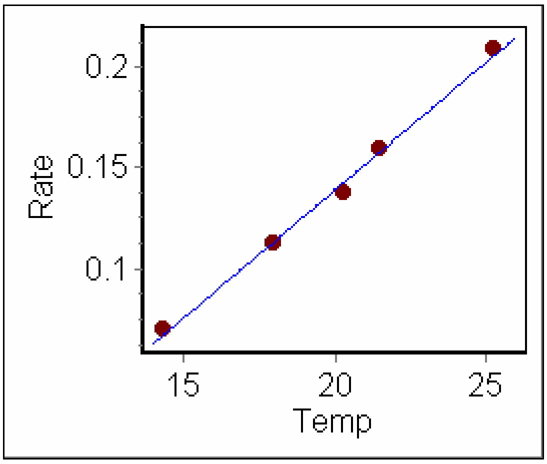
Fig. 8.6d shows the graph of the days to 50% hatch v mean temperature and Fig. 8.6e the plot of its inverse (Rate) against mean temperature. These are examples of the graphs shown in Figs 8.6a and 8.6b.

A regression line is now fitted to these data (Fig. 8.6f), from which, also in Fig. 8.6f, we calculate the baseline temperature (*Tb*) of about 9°C and the required accumulated temperature as about 80°C.

Fig. 8.6f Fitting a regression line

`Statistics ⇒ Regression ⇒ Simple`





Simple Linear Regression

```
----------------- DETAILS OF THE FITTED LINE ------------------

Fitted equation      : Rate = -0.1133   +   0.0126 * Temp

Standard error of slope : 0.0005  with 3 d.f.
95% confidence interval for slope  0.01101    to    0.01419

R-squared              : 0.9953

: ?1/0.1133    : Note this gives the base temperature
  8.8261
: ? 1/0.0126   : Note this gives the required accumulated temperature
  79.365
```

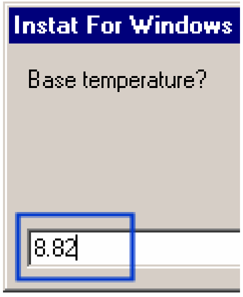
This provides the type of information needed to give advice to the farmers. From the traps, we know the day to start the calculations. Then suppose that the mean temperatures for the following days are 7, 10.4, 13.6 ... 19.9. The macro, **Clim8e.ins**, which is run in Fig. 8.6g calculates the accumulated temperature, *TAcc*, for day *n* and temperature *Tn*, using

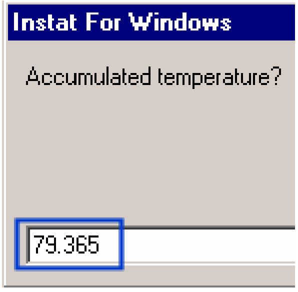
$$TAcc = TAcc + (Tn - Tb)$$

When run, it asks for the base temperature, and the required accumulated value, Fig. 8.6g. It then calculates that after 10 days, they will have exceeded the required threshold. This is when the farmers are advised to spray.

Fig. 8.6g Calculating accumulated temperatures to identify when to spray

`Submit` ⇒ `Run Macro` `and select` **`Clim8e.ins`**





	X1	X2	X3
	Tn	Day	Tacc
1	7	1	0
2	10.4	2	1.58
3	13.6	3	6.36
4	18.8	4	16.34
5	20.5	5	28.02
6	18.6	6	37.8
7	20.3	7	49.28
8	19.4	8	59.86
9	20.8	9	71.84
10	18.6	10	81.62
11	19.9	11	
12			

Chapter 9 – Agricultural Climatology

9.1 Introduction

The first section of this chapter describes the scope of agricultural climatology. The remainder of the chapter is devoted to one of the key building blocks, namely the calculation of evapotranspiration. In these sections the case for the use of the Penman-Monteith equation is given. The use of Instat dialogues for the calculations are described. For reference we also give the formulae and a hand calculated example of their use.

Subsequent chapters consider other topics in agricultural climatology. Chapter 10 combines the calculation of evapotranspiration with the use of crop coefficients in a simple crop index. Chapter 8 is devoted to the analysis of temperature data, including a study of different methods for studying the impact of climate on problems of pests and diseases. At a simpler level, calculations of water balance, described in Chapter 6, can use the calculated evapotranspiration from this chapter.

9.2 The scope of agricultural climatology

Agricultural meteorology is the study of all aspects of the links between the atmosphere and agriculture. It is obviously concerned with the effects of weather on agriculture, but also covers the influences of agriculture (and forestry) on the atmosphere. Agricultural climatology is concerned with the variation in time and space of all aspects of agricultural meteorology.

Agricultural meteorology is a relatively new subject, developing rapidly. Until recently it has mainly been practised by people who have been involved in one of the established sciences, often meteorology, but there is now sufficient theory and established methodology for it to be regarded as a subject in its own right. Its boundaries are diffuse since it merges into other established disciplines such as soil science and crop physiology as well as agriculture and meteorology.

Operational agricultural climatology is usually carried out by staff within the national Met. Service. Sometimes this is in association with agricultural advisory staff, and in the most successful cases there are interdisciplinary teams working together. Meteorologists alone rarely have the structure or information to reach the farming communities. Internationally, the WMO agricultural climatology effort is organised through the Agricultural Meteorology Division of the World Climate Programme.

The range of topics, which may be considered as part of agricultural climatology, is summarised under the following interrelated headings:

a) **Distribution of crops, animals and farming systems**
Crops, animals and farming systems can only succeed at a locality if they are adapted to that climate, including its variation from year to year.

b) **Effects of weather on crop yields**
Weather is often the major factor that affects year-to-year variation in crop growth, development and yields. Effects may be direct (physiological) or indirect (pests, diseases or crop management).

c) **Pests and diseases**
Development of pests and diseases depend upon temperature and humidity, their usual spatial and temporal distribution is often climatically determined, and their dispersion and spread is often controlled by the weather on all scales.

d) **Exchanges of mass and energy**
Crop growth depends upon the supply and use of solar radiation and carbon dioxide. All living organisms exchange heat, moisture and other substances with their environment. The use of water by evapo-transpiration is a major interest.

e) **Microclimates**
All organisms live within their own microclimates and usually show physiological responses to these environments.

f) **Modification of microclimate**
Examples include mulches, shelter, drainage, irrigation, shade, glasshouses and animal housing.

g) **Agricultural management**
Many aspects are weather dependent, e.g. sowing crops, applying herbicides and other agrochemicals, animal nutrition, grazing periods, crop harvesting, processing and transport.

h) **Climate change and agriculture**
Most farming systems will need to change and adapt if (or when) climates change. All the topics mentioned above are relevant in this context.

The above topics cover a range of spatial scales (micro, meso and macro). For example crop distribution includes macro-scale features such as the distribution of rice in the tropics and micro-scale features such as planting at the bottom edges of ridges to maximise water availability on slopes in dry areas.

Decision-making in agriculture may be tactical or strategic and agricultural climatology plays an important part in both types of decision. Tactical decisions are immediate decisions, i.e. what to do now. Examples are: how much to irrigate a crop that is currently growing; whether to harvest a crop tomorrow or next week. Tactical decisions are therefore often based on climatic information plus current and forecast weather information.

Farming strategy is concerned with longer term planning, e.g. what crops to grow; how much irrigation capacity is needed. This normally requires an evaluation of needs and risks over a sequence of years.

A careful climatic analysis can also underpin strategic decision-making that incorporates seasonal forecasting and climatic change scenarios.

9.3 Working with users

9.3.1 Who are the users of agrometeorological information?

Users belong to different major groups, though to provide an effective service, each application must be considered unique. The agrometeorologist must understand the overall objectives of each project to work effectively in a team:

a) Are the interests in benefits in the technical, economic, social, security, sustainability, environment, leisure, or other domains?

b) Are the users policy makers, monitoring agents, or practising producers?

c) Can users reap the benefits from the product individually, or will remuneration come collectively, e.g. through a Chamber of Agriculture, a commodity agency, a marketing unit?

The initial user may be a colleague within the Met service, asking for an analysis of data. "Outside" users include government, NGOs, or an information dissemination unit, such as a local radio. They may be farmers or groups of farmers or farmer’s organizations, plant or animal health protection, forestry or livestock services, fertilizer companies or soil conservation groups.

Other users in the Ministry of Agriculture, the Ministry of Planning or agricultural research institutes, may aim for the development of sustainable agriculture, for warnings on alarm situations, bush and forest fires, locust control, for drought alleviation measures, flood control, the planning of the movement of stocks of food or seeds, or a crop monitoring activity. They may be Embassies, marketing or post-harvest crop management services, entities engaged in the conservation of the environment, or any other kinds of "customer".

To identify users in agricultural, livestock husbandry or forestry activities, it is useful to contact the relevant Ministries, the "Chambers of Agriculture" or equivalent units, the commodity institutes, or the district agricultural and livestock services. This identification of the user and of the product(s) required, can be through listening to the requirements of the work of persons in other disciplines, and of the aspects that could make their work safer, easier, more efficient, more reliable, etc.

9.3.2 How to begin?

The process can start with a dialogue on the users' work, discussing each step in the agricultural production process and the effect of weather factors on these steps. An assessment may also be made of when and how agrometeorological information products can realistically increase the efficiency of the steps. Such a discussion can deal with the factors that determine the highest possible rate of production, factors that limit the production or those that reduce production below levels already established. This could than lead to a joint definition of the agrometeorological product that is needed. Such a dialogue requires that the agrometeorologist knows about the products that (s)he can offer or develop. They may require the joint provision of input data, hence joint data collection , analysis and formulation of the message in a user-adapted language.

Before a decision is made to furnish a product, consideration should be given to how the user can effectively manage the product. The technology must be of realistic service.

9.3.3 What does the user require?

Often, a combined use of climatological inputs, both observed and forecast, is needed, and the relative reliability of each component should be identified. Sometimes the users only vaguely realize how meteorological information can help them; by going, together, step by step, through the work or production process can one identify areas where meteorology might be of use and where the benefits of provision of information could be examined. One may have to make a "user sensitivity analysis" and to assess (and communicate) the value of the information product supplied.

User-tailored weather information for planning, adaptation of the system, and day-to-day operations in agriculture involving the dosage and timing of application of inputs, is one of the major factors that can increase the efficiency of these measures and help to reduce the risks on the investments made. This aspect defines one group of clients. Another group is involved in general matters or in activities that precede or follow after agricultural production has been achieved: marketing, processing, consumer orientation, legal and administrative matters and environmental issues.

Users may wish to obtain:

a) a description of the basic factors determining the atmospheric environment for agriculture (solar radiation, temperatures, water availability in all its forms, humidity, the wind regime and other characteristics, such as weather "hazards");

b) a description of the requirements for each application;

c) a quantitative formulation of the relationships between weather and vegetation, soil, open water and animals and the reciprocal effects of these "surfaces" on their nearby atmospheric environment;

d) a process to "match" the requirements of the users to the meteorological conditions that may exist, to optimize the use of all the resources provided by the weather and the other inputs and to minimize the influence of adverse conditions (Rijks, 1986).

9.4 Products

Possible products fall into different groups:

a) basic data;

b) basic data together with an analysis and/or an advisory message for specific applications, possibly combined with non-meteorological data, such as those derived from remote sensing;

c) methods, techniques, software packages for specific applications.

Examples are given of products in the second group, dealing with the planning of agriculture, day-to-day operations and monitoring and associated measures respectively.

9.4.1 Products to help planning

In **planning**, products are based on values (maximum, optimum, minimum and probabilities) of the relevant parameters for different crops. For example:

Types of factor	Elements
factors that define maximum production:	temperature, solar radiation;
factors that may limit production:	water balance, conditions for nutrient uptake and weeding;
factors that reduce production:	pests and diseases.

The products include:

a) **the agro-meteorological characterisation** of regions (FAO, 1978; Rijks, 1994) for the application of results of research;

b) **the probability of rainfall** for crop water balance calculations to plan the agricultural system, to assess the possible length of the rainfed cropping season (Manning, 1956; Rijks, 1976) and to provide information for longer-term, infrastructural, measures like land-layout for erosion control and soil conservation, intercropping systems, and contour-ridging for the conservation and use of water (Rijks, 1977). Similar information is needed for the study of relations in catchment management and for planning of irrigation system layout and related studies;

c) **information about (extreme) low or high temperature** regimes and their duration and localization, that affect the development and growth of crops and animals, and in some cases the state of the infrastructure serving agriculture, the frequency of the risk of occurrence of frost, or of heat stress for crops and livestock;

d) **information about solar radiation and sunshine hours**, for the calculation of photosynthesis, crop growth, evapotranspiration, crop drying, and for applications in the sphere of the agricultural infrastructure and operations, like the construction of animal shelters, animal health care or for on-farm (renewable) energy generation and conservation;

e) **information required in the livestock industry** including the assessment of the potential pasture productivity, of the seasonal food and water supply and quality, assessment of the risk of overgrazing or of bush fires, hay making, housing, animal health and productivity, the introduction of highly-productive species and the drying with solar energy of meat and fish;

f) **information on weather factors** (water balance, temperature and humidity regimes, day length etc.) that help with the year-to-year selection of varieties adapted to the variability of the length of season;

g) **parameters dealing with the choice and use of farm machinery**, fertilizer applications, pest and disease management;

h) **quantitative values** (maximum, optimum, minimum, probabilities) of the relevant parameters of models of development of pests and diseases (Franquin and Rijks, 1983) and of migrant pests (Rainey *et al*, 1990);

i) **information to implement measures of microclimate manipulation and modification** (Stigter, 1994) for the establishment and management of windbreaks and for planning agroforestry plantations;

j) **information on the probability of certain conditions of solar radiation**, temperature and water availability for the development of intercropping and multiple cropping systems, so that natural inputs are exploited optimally;

k) **information for planning the feasibility and efficiency of on-farm water storage facilities** (Baradas and Sutrisno, 1981);

l) **conditions for the selection of different forest species**, their establishment, the risk and incidence of forest pests and diseases, information on the risk and for the forecasting of bush and forest fires and for forest fire management practices;

m) **assessment of the solar and wind energy potential**;

n) **information on humidity**, a major element in the assessment of the risk of occurrence of crop diseases and some crop pests. Also, low humidity may inhibit fertilization during flowering;

o) **wind regimes** may influence lodging, and thus perhaps the need for ridging, and the movement of crop and animal pests and their control. Extreme winds may cause significant damage to fruit trees (Mellaart *et al*, 1999). Information on wind regimes is essential for the construction of windbreaks and the establishment of fire-breaks in bush- and forest-fire control.

9.4.2 Products for day-to-day operations

For **day-to-day agricultural operations, products** may be:

a) **the probability of rainfall** for crop water balance calculations to decide on agricultural activities or processes for different crops, such as accessibility of the fields, land preparation, sowing, germination, weeding, thinning, ridging to prevent lodging, supplemental irrigation, fertilizer application, crop protection measures, ripening, harvesting and post-harvesting operations such as drying and storage (e.g. Traore *et al*, 1992; Direction Nationale de la Météorologie, 1998);

b) **climatic, probability and forecast information** for the planning of irrigation systems, risks of water shortages, optimization of the water use efficiency (the ratio of yield per unit water), information for day-to-day scheduling irrigation scheduling models, using real-time data and forecasts (e.g. Rijks and Gbeckor Kove, 1990; Friesland *et al*., 1998; Smith, 1999, this volume);

c) **information on meteorological factors that affect the efficiency of energy**, one of the most expensive recurrent inputs in agriculture, whether the energy be of fossil, human, animal, mechanical, thermal, solar (electrical) or chemical nature, through a choice of optimum timing and amount of such inputs;

d) **information of the risk of weather hazards** for crops and animals, hail, frost, hot dry winds that may cause sterilization of pollen, floods, droughts etc.;

e) **information to foresee the optimum time for harvesting** (e.g. of vine-grapes, Gerbier and Remois, 1977; Strydom, 1999);

f) **information for the improvement of storage conditions** (e.g. of groundnuts in Gambia, Rijks, 1987);

g) **the practice of agricultural aviation**, for sowing, fertilizer application and surveying in addition to crop protection, requires (agro)meteorological information for the assessment of the needs and the potential benefit of an intervention, as well as for the application operations.

Other clients may, in relation to studies of the effects of climate variability, or in crop monitoring and yield forecasting procedures, need the outlook, the "scenarios" that could occur (and their probability), following different (agro)meteorological or general weather events.

9.4.3 Monitoring and evaluating products

For monitoring agricultural production and the social, economic and legal aspects of agriculture, (agro-)meteorological information is used in:

a) crop monitoring activities in food security programmes;

b) modelling of potential production of natural grazing areas by livestock services, which in turn may have an effect on transhumance;

c) promoting the efficiency of the use of water and energy, the reduction of pollution and the conservation of the environment;

d) monitoring, and forecasting floods and droughts and the alleviation of their effects;

e) monitoring of desertification, avoidance of overgrazing, salinization, wind- and water-erosion;

f) information for wildlife conservation and management

More details and references are given by Rijks (2000).

A product has value only to the extent that it is being used to the clients' satisfaction. Therefore, feedback on the benefits that are derived from the decisions based on their use, be they technical, social, economic, environmental or other, must be obtained regularly.

Evaluations may consider whether the value of the product is to minimize damage, risk or costs, to maximize output, net return, or to enhance the value of, or the possibility of exploiting, other resources. Feedback on the indirect benefits from post-harvesting operations and agro-industrial procedures are also important.

9.5 Introduction to evapotranspiration

Penman (1948) produced the first physically based equation for evapotranspiration of water from land and water surfaces. His first equation was for evaporation from an open water surface (e.g. a lake), but it was soon modified to deal with evapotranspiration from vegetation.

In this form, Penman's equation estimated potential evapotranspiration, i.e. the rate of water loss from a short green crop, fully covering the ground and fully supplied with water.

Under these conditions (full crop cover and soil near field capacity) the rate of evapotranspiration is a consequence of atmospheric, rather than surface, conditions.

The appeal of Penman's equation was that it dealt with the physical basis of evapotranspiration. There are two physical aspects to evapotranspiration:

a) The need for energy to provide the latent heat to convert liquid water to water vapour. This is normally provided by solar radiation, through the energy balance equation.

b) The need to transport the water vapour away from the evaporating surface into the atmosphere. This is a function of turbulent transfer, which depends upon the interaction of wind with the surface and of atmospheric water vapour content.

When Penman derived his equation, the physics for (a) was well understood. The physics for (b) was less clear and Penman used an empirical relationship between evapotranspiration and windspeed and vapour pressure deficit. This was based on experimental work at Rothamsted Experimental Station and provided an excellent basis for his equation.

Gradually, scientists realised that evapotranspiration from crops differed slightly from the value predicted from the Penman equation and various empirical factors were introduced to correct for this. Physically, these arose from (b) above, in particular that turbulent transfer depends upon the nature of the surface and that all crops could not be treated as short grass.

Monteith (1965) completed the scientific theory by introducing the concept of resistances to describe the turbulent transfer of water vapour to the atmosphere. Combining these concepts with the Penman equation gave the Penman-Monteith equation.

Prior to 1990, the standard method of estimating crop water was based on the original Penman equation, with various modifications, as outlined by Doorenbos and Pruitt (1976) and Gommes (1983).

In 1990 FAO reviewed their methods and reached the following conclusions.

a) The Penman equation should be replaced by the Penman-Monteith equation. This requires a more exact definition of potential evapotranspiration and hence they introduced reference crop evapotranspiration. This is defined as "the rate of evapotranspiration from a hypothetic crop with an assumed crop height (12 cm) and a fixed canopy resistance (70 sm$^{(-1)}$) and reflection coefficient (0.23) which would closely resemble evapotranspiration from an extensive surface of green grass cover of uniform height, actively growing, completely shading the ground and not short of water."

   This is really just a more precise definition of what Penman (1956) called potential evapotranspiration: "The rate of evapotranspiration from a short green crop, fully covering the ground and fully supplied with water".

b) For the time being, crop water use could be estimated by multiplying the reference crop evapotranspiration by a crop coefficient ($k_{(c)}$) to allow for differences in crop type and crop cover. This would follow the guideline of Doorenbos and Pruitt (1976)

c) In the longer term (10 years), information on the canopy and aerodynamic resistances of crops would be collated to allow the direct use of the full Penman-Monteith equation.

9.6 Dialogues to calculate evapotranspiration

The Instat dialogues under the **Climatic ⇒ Evaporation** menu are used for the Penman-Monteith calculations as shown in Fig. 9.5a.

Fig. 9.6a Evaporation sub-menu

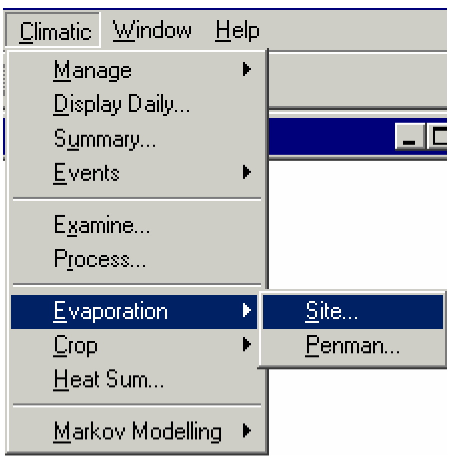


Fig. 9.6b Data from Kayes, Mali

File ⇒ Open From Library ⇒ Kayes.wor

	X1*	X2*	X3*	X4*
	Tmean	RHmean	suh	msc
1	25.4	22	8.16	2.8
2	28	19	8.32	2.9
3	31.2	17	9.71	2.8
4	34	19	9.53	2.7
5	35.5	30	8.19	2.8
6	32.1	55	7.3	2.8
7	28.8	72	6.35	2.6
8	27.4	79	6.03	2.6
9	27.9	79	6.97	2.3
10	29.5	69	7.77	2
11	28.8	43	8.27	2.4
12	25.8	28	6.48	2.6
13				

The four climatic variables to measure are radiation, wind-speed, temperature and humidity and should first be entered into a worksheet. An example uses monthly data for Kayes in Mali, shown in Fig. 9.5b. The data that were measured are:

a) mean temperature in °C X1
b) mean % relative humidity X2
c) wind-speed at 2m in m/s X4
d) daily hours of bright sunshine X3

The first dialogue is **Climatic ⇒ Evaporation ⇒ Site**, Fig. 9.5c, and the four parts to this dialogue need to be completed. The first gives the location of the site. The latitude of Kayes is 14° 26' N and the altitude is 47m.

Fig. 9.6c Site information for Kayes, Mali

Climatic ⇒ Evaporation ⇒ Site

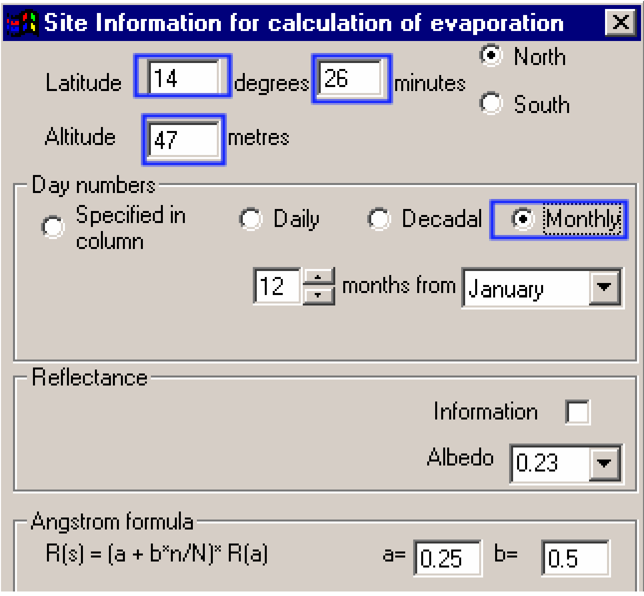


Fig. 9.6d Results

	X5	X6
	const	Date
1	14.43333	16
2	47	46
3	0.23	76
4	0.25	106.5
5	0.5	137
6	0.0670267	167.5
7		198
8		229
9		259.5
10		290
11		320.5
12		351

A column of day numbers from 1st January is need and Instat calls this column 'Date. This column can be generated if the data are daily, decades or monthly. Otherwise you need to provide a column of day numbers.

The third part of the dialogue allows the reflection coefficient to be specified as described in Section 9.8. The default value is 0.23 for the reference crop and various other values are suggested if you tick the "Information" check box shown in Fig. 9.6c.

The final section specifies the intercept and slope of the Angstrom formula (see Section 9.8). This is need if radiation is measured as hours of sunshine. These constants are normally determined locally.

The result of running the dialogue is to store the constants in a further column called 'Const and to give the column with the day numbers. The 'Const and 'Date columns generated for Kayes are shown in Fig. 9.6d.

One complication in the calculation of evapotranspiration is the variety of measurements and units that are in common use. Instat caters for this in the **Climatic ⇒ Evaporation ⇒ Penman** dialogue, by transforming the measured data into a fixed set of variables.

The transformed columns are also given fixed names that are used in the Penman-Monteith calculations. The 4 types of measurements are considered in turn.

1) The ***radiation*** data can be given as sunshine hours or measured radiation. If sunshine hours, the column is called **'suh**. If radiation, then it is transformed if necessary to MJm^{-2} and called **'mjo**.

2) The ***wind-speed*** data is transformed to metres/sec and called **'msc**

3) The ***temperature*** data may be given as the mean. If the maximum and minimum values are given, then the mean temperature is calculated and called **'Tmean**.

4) The ***humidity*** can be given as a variety of ways as shown in the dialogue. It is transformed to vapour pressure and called **'ed**.

If the Penman-Monteith checkbox in Fig. 9.6e is left unchecked, then the dialogue simply calculates the columns that are required for the calculation. For illustration this was done in Fig. 9.6e with the calculated column 'ed in Fig. 9.6f.

Fig. 9.6e Data for Penman-Monteith calculation

`Climatic ⇒ Evaporation ⇒ Penman`

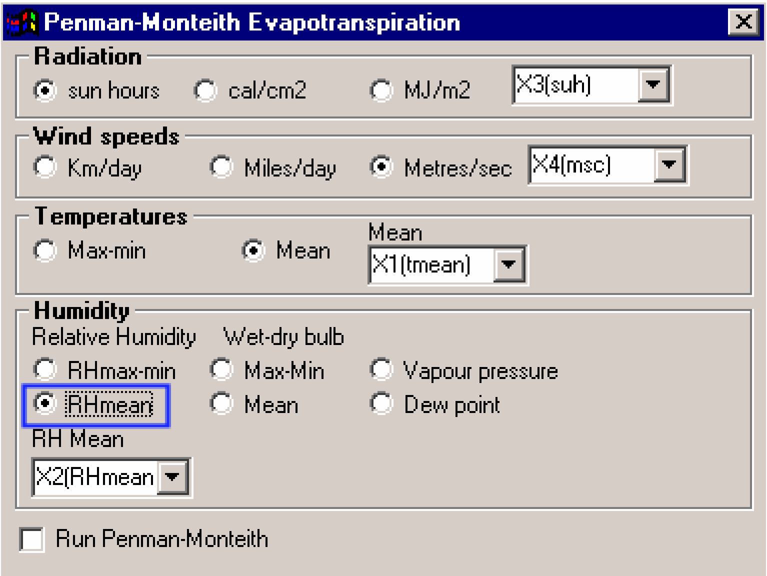
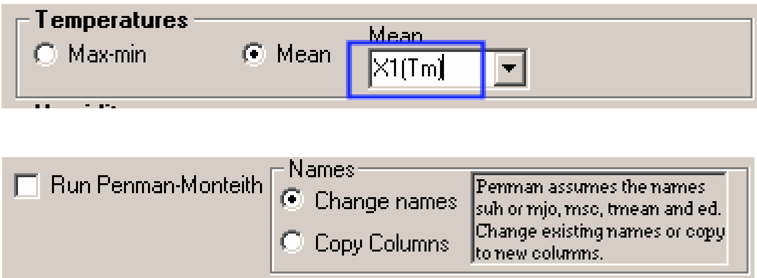


Fig. 9.6f Calculate 'ed

	X7
	ed
1	0.714
2	0.718
3	0.773
4	1.011
5	1.735
6	2.631
7	2.852
8	2.884
9	2.970
10	2.846
11	1.703
12	0.930

Fig. 9.6g Rename 'Tm to 'Tmean in Penman dialogue

	X1*
	Tm
1	25.4
2	28
3	31.2
4	34
5	35.5
6	32.1
7	28.8



In the **`Climatic ⇒ Evaporation ⇒ Penman`** dialogue, columns that already have the names that are used later, are picked up automatically by the dialogue as can be seen in Fig. 9.6e. The data in Kayes worksheet were given the correct names. Fig. 9.6g shows what happens if this is not done. The temperature, X1, has been renamed to 'Tm. Now the dialogue has to be informed which is the temperature column. Instat then gives the option either to produce a copy with the correct name or to change the name of the existing column.

With the correct names in the **`Penman`** dialogue all the information is complete and the **`OK`** button is therefore enabled. Tick the two boxes at the bottom of the dialogue (Fig. 9.6h).

The graph produced after clicking **`OK`** is shown in Fig. 9.6i. It shows that the evapotranspiration (ET_0) is the sum of the aerodynamic and radiation terms and illustrates the different patterns of variation in the two terms throughout the year.

The Penman dialogue also produces the columns listed in Fig. 9.6j.

Fig. 9.6h Penman calculations:
`Climatic ⇒ Evaporation ⇒ Penman`

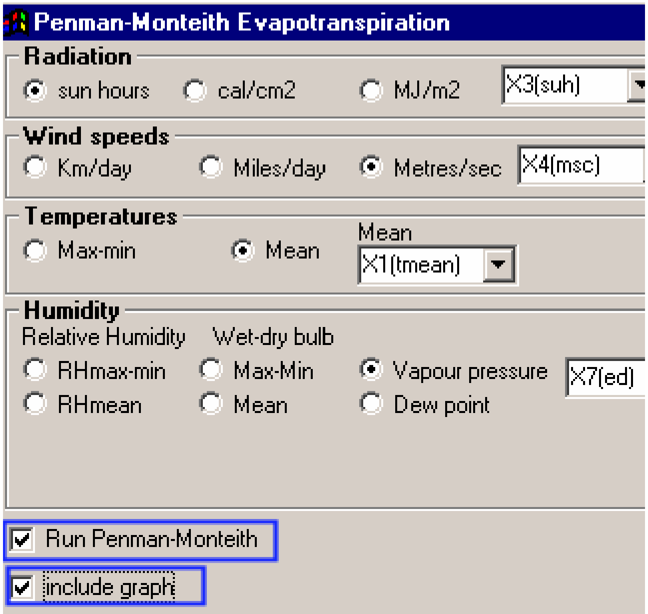


Fig. 9.6i Evapotranspiration

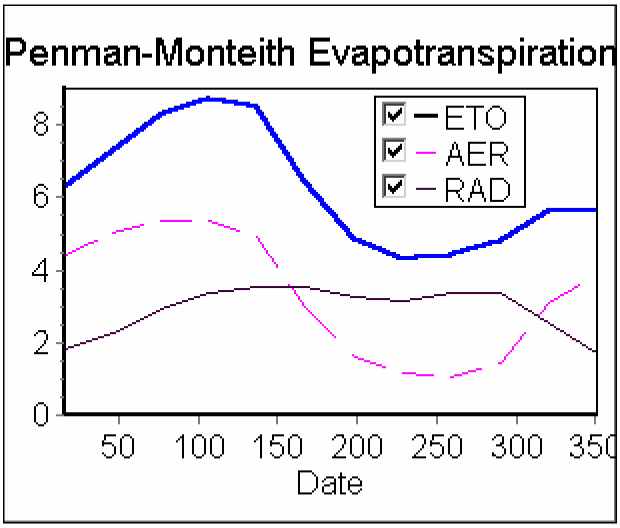


Fig. 9.6j Columns calculated for Penman-Monteith

Name	Element	Equation (Section 9.8)
'VPD	Vapour pressure deficit	69
'Rs	Solar radiation	55
'Rns	Net short-wave radiation	55
'Rnl	Net long-wave loss	63
'Rn	Net radiation	50
'AER	Aerodynamic term	69
'RAD	Radiation term	69
'ETO	Crop evapotranspiration	69

	X17	X18	X19	X23	X24	X25	X26
	AER	Rs	Rns	Rnl	Rn	RAD	ETO
1	4.4	18.3	14.1	6.5	7.6	1.8	6.3
2	5.0	20.1	15.5	6.7	8.8	2.2	7.3
3	5.4	23.8	18.3	7.6	10.7	2.9	8.3
4	5.4	24.2	18.7	6.9	11.8	3.4	8.7
5	4.9	22.0	17.0	4.7	12.3	3.5	8.5
6	3.0	20.4	15.7	2.9	12.8	3.5	6.5
7	1.6	19.0	14.7	2.3	12.4	3.3	4.9
8	1.2	18.7	14.4	2.2	12.2	3.1	4.3
9	1.1	19.7	15.2	2.5	12.7	3.4	4.4
10	1.5	19.6	15.1	3.0	12.1	3.3	4.8
11	3.1	18.6	14.3	4.8	9.5	2.5	5.6
12	3.9	15.5	11.9	5.0	6.9	1.7	5.6

9.7 Calculating evapotranspiration for daily data

Data from Bohicon in Benin for 1983 are used to illustrate the analysis on a daily basis. Open the worksheet **Bohicon.wor**. Fig. 9.7a shows the names given to the columns and the first few days of the data.

Fig. 9.7a Daily data from Bohicon, Benin

`File ⇒ Open From Library ⇒ bohicon.wor`

Col.	Name	Contents
X1	Day	Day Number
X2	Wind	Wind speed (m/s)
X3	Sunhrs	Sunshine hours
X4	Vap	Vap pressure (mb)
X5	Minhum	Min RH (%)
X6	Maxhum	Max RH (%)
X7	Mintemp	Min temp. (deg C)
X8	Maxtemp	Max temp. (deg C)
X9	Rain	Rainfall

X1	X2	X3	X4	X5	X6	X7	X8	X9
Day	Wind	Sunhrs	Vap	Minhum	Maxhum	Mintemp	Maxtemp	Rain
1	0.3	4.5	19.7	24	98	22	23.4	0
2	0	7.6	23	30	96	21	34.8	0
3	0	0.8	21	36	93	20.4	33.7	0
4	0.7	6.9	16.5	27	97	18.5	33.8	0
5	1	5	11.4	23	58	19.6	33	0
6	2	0	7.7	17	32	20	29.3	0
7	1.6	0	8.8	20	35	20	30.2	0
8	0.7	6.8	10.5	19	62	21	32.5	0
9	0.7	7.2	13.5	22	68	18.2	33.4	0
10	0	6.8	12.5	19	66	19	33	0
11	0.6	5.6	12.1	16	63	20	33.7	0
12	1.4	7.1	11.8	16	59	21.7	33.8	0
13	2	9.4	10.7	18	41	21.5	34.3	0
14	0.6	8	10.4	18	51	21.3	34.5	0

Bohicon is 07 degrees 10 minutes N at an altitude of 167m, so the Site dialogue is completed as shown in Fig. 9.7b.

Fig. 9.7b Site data for Bohicon

`Climatic ⇒ Evap ⇒ Site`

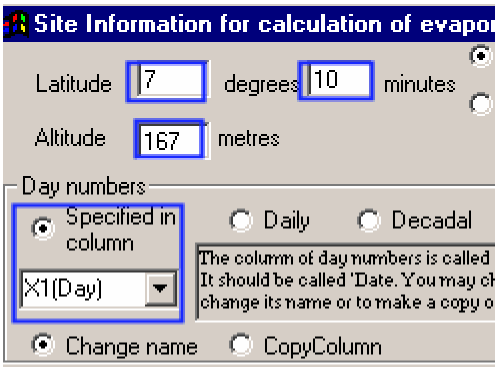
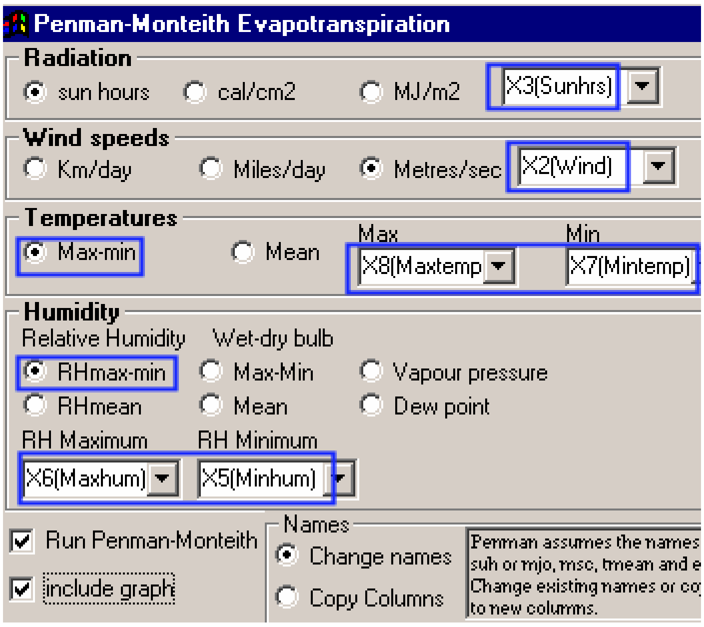


Fig. 9.7c Penman dialogue for Bohicon

`Climatic ⇒ Evaporation ⇒ Penman`



The Penman dialogue is shown in Fig. 9.7c.

The evapotranspiration is stored in ETO, the radiation term in RAD and the aerodynamic term in AER. The values of ETO calculated using Penman-Monteith are shown in Fig. 9.7d together with a graph of ETO, RAD and AER.

Fig. 9.7d Evapotranspiration from daily data from Bohicon

Daily data for: ETO

Mon	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Day.												
1	2.7	5.0	3.5	5.1	5.4	3.7	3.0	3.5	3.7	4.0	4.7	3.7
2	3.5	3.9	3.7	5.4	5.1	5.0	3.5	3.8	3.7	4.7	5.6	4.3
3	2.0	3.4	4.9	3.2	5.5	2.0	3.8	3.0	3.2	4.2	5.2	4.3
.	...	...	...	...	...	...	...	...	...	...	...	...
27	5.7	4.8	5.9	5.0	5.4	3.6	2.7	2.9	4.0	4.4	4.6	3.9
28	7.2	4.0	6.2	5.7	2.4	2.7	4.5	2.6	3.6	4.5	4.1	3.7
29	5.8		6.2	5.1	4.8	3.5	3.4	2.8	3.8	4.7	4.5	2.9
30	5.6		5.3	4.1	5.1	4.8	2.7	3.2	2.8	4.1	2.7	3.1
31	4.5		4.6		5.0		2.8	3.2		5.1		4.3
Mean										(Overall: 4.4)		
	4.1	5.5	5.4	5.2	4.7	3.7	3.5	3.2	3.5	4.2	4.7	4.2

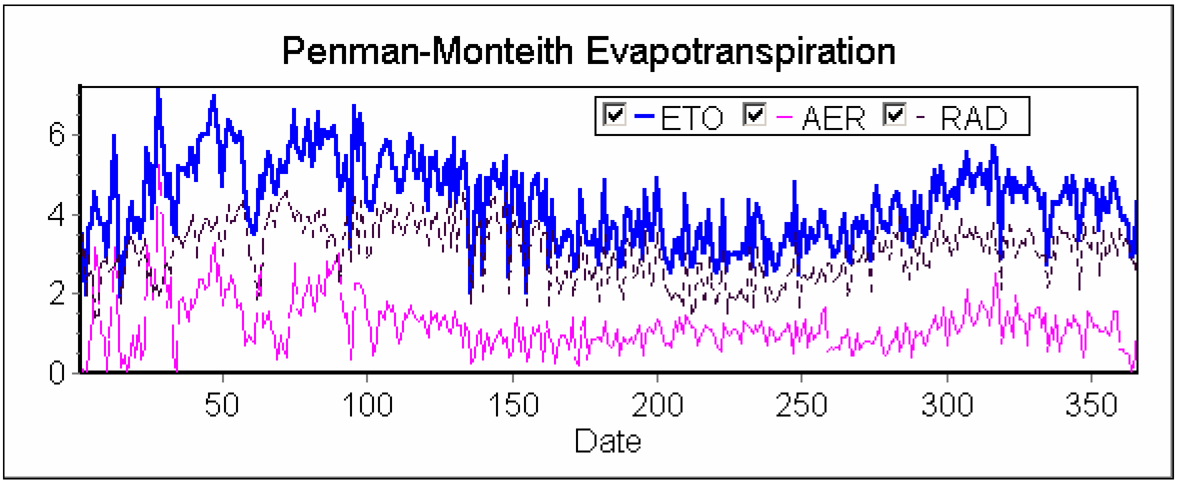


Fig. 9.5d shows that peak values of ET0 are found in February and March with monthly means of about 6 mm per day. Mean values decrease through April and May to 4 mm, a level which is maintained, apart from a small peak (5 mm) in November for the remainder of the year. Monthly means are also given in Fig. 9.7d. The graph shows that the radiation term is the dominant component and is relatively constant through the year. The aerodynamic term is about 2 mm per day in January to March and about 1 mm per day until the secondary peak in November.

In some studies the average value of ET0 may be required. This is illustrated by calculating the mean on a decade basis. Use **Climatic ⇒ Manage ⇒ Make Factor** as shown in Fig. 9.7e.

Fig. 9.7e Make a column of decade numbers

Climatic ⇒ Manage ⇒ Make Factor

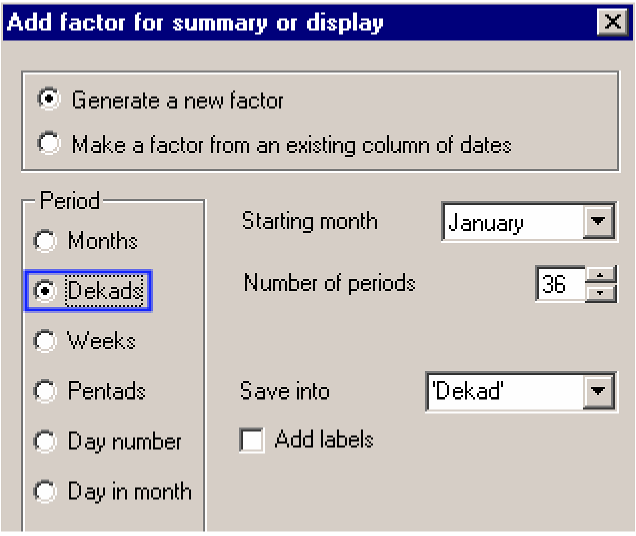
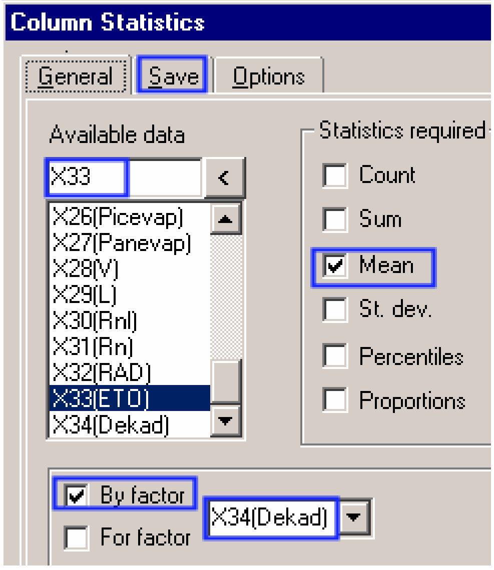


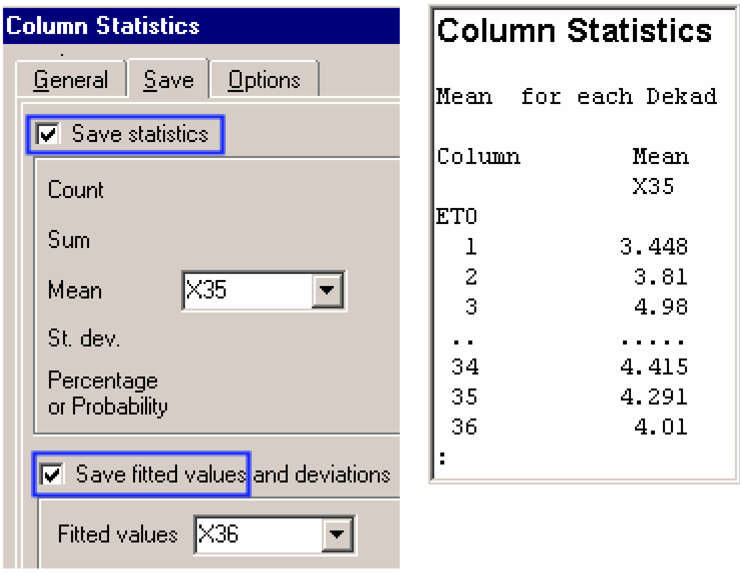
Fig. 9.7f Calculate 10-day means

Stats ⇒ Summary ⇒ Column Stats



Use **Statistics ⇒ Summary ⇒ Column Statistics** as shown in Fig. 9.7f to calculate the 10-day means. Click on the **Save** tab and save both the means and the fitted values.

Fig. 9.7f Saving the 10-day means and the fitted values



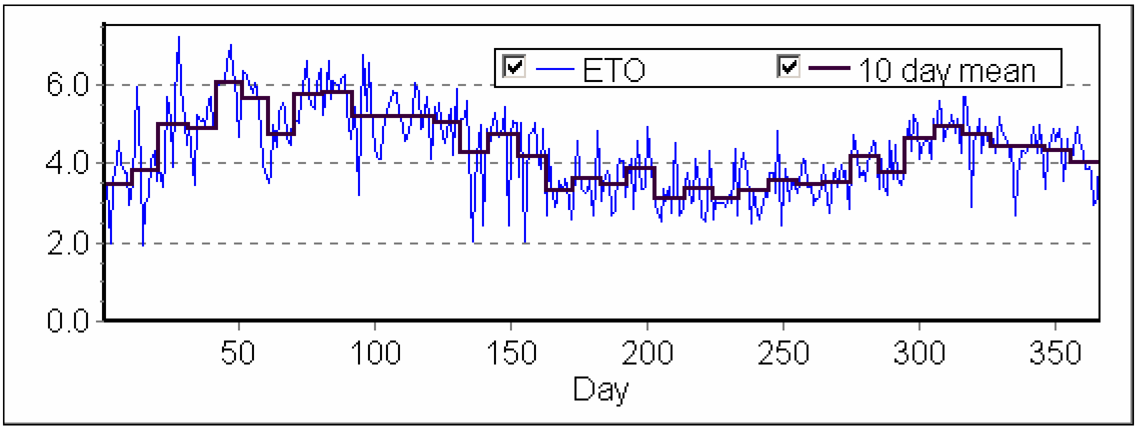
	X35	X36
1	3.4	3.4
2	3.8	3.4
3	5.0	3.4
4	4.9	3.4
5	6.0	3.4
6	5.6	3.4
7	4.7	3.4
8	5.7	3.4
9	5.8	3.4
10	5.2	3.4
11	5.2	3.8
12	5.2	3.8
13	5.0	3.8

In Fig. 9.5f, X35 contains the means and is a column with 36 values. X36, while containing the same numbers as X35, is of length 366. It is displayed using the "**Display of daily data**" dialogue in Fig. 9.7g and also plotted, together with the daily values.

Fig. 9.7g Table and plot of 10-day means

Daily data for: X36

Mon / Day.	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	3.4	4.9	4.7	5.2	5.0	4.1	3.5	3.3	3.5	4.1	4.9	4.4
2	3.4	4.9	4.7	5.2	5.0	4.1	3.5	3.3	3.5	4.1	4.9	4.4
3	3.4	4.9	4.7	5.2	5.0	4.1	3.5	3.3	3.5	4.1	4.9	4.4
4	3.4	4.9	4.7	5.2	5.0	4.1	3.5	3.3	3.5	4.1	4.9	4.4
.	...	...	...	...	...	...	...	...	...	...	...	...
28	5.0	5.6	5.8	5.2	4.7	3.6	3.1	3.3	3.5	4.6	4.4	4.0
29	5.0	5.6	5.8	5.2	4.7	3.6	3.1	3.3	3.5	4.6	4.4	4.0
30	5.0		5.8	5.2	4.7	3.6	3.1	3.3	3.5	4.6	4.4	4.0
31	5.0		5.8		4.7		3.1	3.3		4.6		4.0



The same point about averaging can be made here, as was discussed for the water balance dialogue in Chapter 6. If the average evaporation is required on a 10-day basis, then should the basic climatic data be averaged before using Penman-Monteith? Or should the ET0 values be averaged, as was done here? Computationally either method is possible, (the Penman dialogues merely require columns of data of the same length). As usual, the averaging should be at the end (as was done here to produce Fig. 9.7g), rather than at the beginning of the calculations.

Similarly, where data are available for more than one year, the calculations should be done for each year and the results then averaged. This can be done by having a worksheet with the data end-to-end for the successive years.

Partly to test Instat, ten years of daily data were used in a worksheet with 3660 rows and 40 columns. After calculating Penman-Monteith's evaporation for each day of the ten years, the **Statistics ⇒ Summary ⇒ Column Statistics** dialogue was used to evaluate both the mean and the standard deviation for each 10-day period.

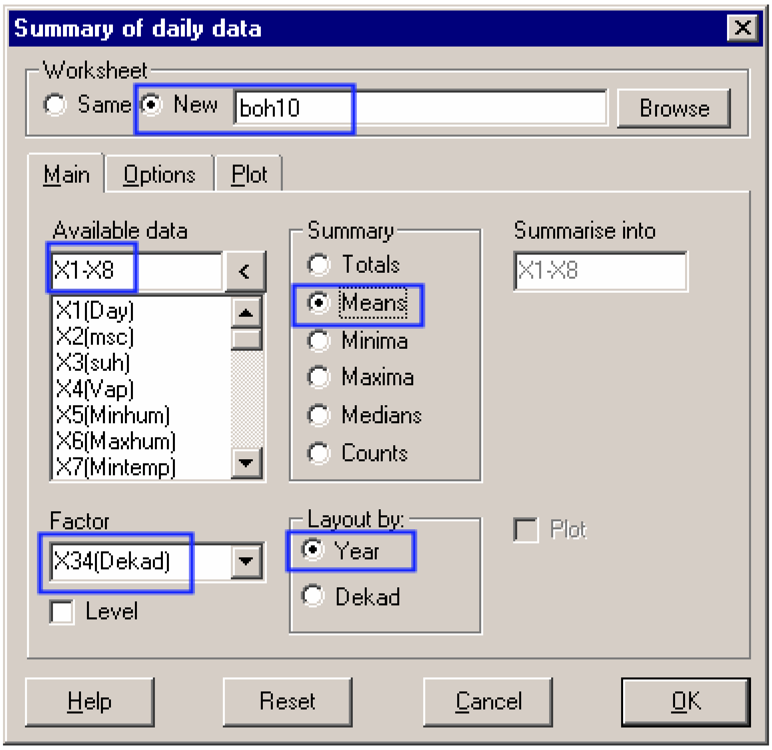
In some countries it may be considered impractical (because of lack of data or of time) to do the analysis for all sites on a daily basis. In this case a preliminary piece of "research" should be to evaluate the consequences of a 10-day or monthly analysis for sites for which daily data are available.

This “research” is straightforward with Instat. The calculation of the decades was shown in Fig. 9.7f. (An alternative is to calculate monthly summaries.) The dialogue to summarise the input data is shown in Fig. 9.7j.

To transform the data in X2 to X8 to 10 day means, use **Climatic ⇒ Summary** and complete the dialogue as shown below. The decade means are put into X2 to X8 in a new worksheet

Fig. 9.7j Calculate decade means for Bohicon data

`Climatic` ⇒ `Summary`



Run the Penman dialogues again on a 10-day basis, **Climatic ⇒ Evaporation ⇒ Site and Penman**.

Compare the results with those given in Fig. 9.7h. In this case the results were similar, with a maximum discrepancy of 0.3 mm per day

9.8 Reference crop evapotranspiration

Introduction

In summary, the key FAO 1990 conclusions were:

a) The Penman-Monteith equation was the best method and should be used for estimating evapotranspiration. This conclusion was based on recent comparative studies. If this is not possible, the older Penman equation can still be used.

b) Information on canopy and aerodynamic resistances would be collated with the aim of introducing the direct use of the Penman-Monteith equation.

c) The previous FAO method of estimating reference crop evapotranspiration and then using crop coefficients to allow for different crops and growth stages can still be used, until new software and information on resistances has been developed and distributed.

Recommended combination formula for reference evapotranspiration ($ET_{(0)}$)

From: FAO (1992). Report on the Expert Consultation of FAO Methodologies for Crop Water Requirements, held at FAO, Rome, May 1990.

Defining reference evapotranspiration ($ET_{(0)}$) as "the rate of evapotranspiration from a hypothetical crop with an assumed crop height of 12cm, a fixed canopy resistance of 70 $sm^{(-1)}$ and an albedo of 0.23, closely resembling the evapotranspiration from an extensive surface of green grass of uniform height, actively growing, completely shading the ground and not short of water", the estimation of the $ET_{(0)}$ can be determined with the combination formula based on the Penman-Monteith approach. For the estimation of daily values the equation becomes:

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma\frac{900}{T+273}U_2(e_a - e_d)}{\Delta + \gamma(1 + 0.34U_2)} \qquad (69)$$

where: $ET_{(0)}$: reference crop evapotranspiration [mm $d^{(-1)}$]
$R_{(n)}$: net radiation at crop surface [MJ $m^{(-2)}$ $d^{(-1)}$]
G : soil heat flux [MJ $m^{(-2)}$ $d^{(-1)}$]
T : average temperature [°C]
$U_{(2)}$: wind-speed measured in 2m height [m $s^{(-1)}$]
$(e_{(a)}-e_{(d)})$ vapour pressure deficit [kPa]
Δ : slope vapour pressure curve [kPa °$C^{(-1)}$]
γ : psychrometric constant [kPa °$C^{(-1)}$]
900 : conversion factor

Use of the Penman-Monteith equation for reference crop evapotranspiration

The FAO Penman-Monteith reference equation has assumptions about the resistances:

a) $r_{(c)}$ is taken as 70 $sm^{(-1)}$, for grass 0.12m high

This is based on the relationship $r_{(c)} = 2*r_{(s)}$ / LAI where $r_{(s)}$ is the average 24 hour stomatal resistance of a single leaf and LAI is leaf area index which is related to height (LAI = 24 * height in m).

b) $r_{(a)}$ is taken as $208/U_{(2)}$ where $U_{(2)}$ is the 24 hour mean windspeed at a height of 2m. Note that $r_{(c)}/r_{(a)}$ then becomes $70/(208/U_{(2)})$ or $0.34U_{(2)}$.

The equation incorporates some precise physical parameters:

c) 0.408 is the reciprocal of the latent heat of water (2.45 MJ $kg^{(-1)}$) and changes the units of $(R_{(n)}$-G) from MJ $day^{(-1)}$ to mm $day^{(-1)}$.

d) 900/(T +273) includes ρ (density of air) and $C_{(p)}$ specific heat.

e) γ is the psychrometric constant, normally taken as 0.066 kPa $C^{(-1)}$ but this does in theory depend upon atmospheric pressure (P in kPa): γ= 0.000665*P. Normal pressure is about 100 kPa (equal to 1000 mb). This correction can be important at high altitudes. FAO recommend that P is estimated from station altitude (z metres) by:

$$P = 101.3\left(\frac{293 - 0.0065z}{293}\right)^{5.26}$$

Estimation of Net Radiation

The greatest practical difficulties and apparent complexities of using the Penman type of equations results from the fact that usually measurements of net radiation are not available and must be estimated.

Net radiation $(R_{(n)})$ is:

$R_{(n)} = (1 - \alpha)R_{(s)} + L_{(d)} - L_{(u)}$

$L_{(d)}$ = downward long-wave radiation; $L_{(u)}$ = upward long-wave radiation

which in FAO terminology becomes:

$R_{(n)} = R_{(ns)} + R_{(nl)}$ (50)

$R_{(ns)}$ = net short-wave radiation; $R_{(nl)}$ = net long-wave radiation

If measurements of solar radiation are available they should be used. If not, solar radiation $R_{(s)}$ is estimated, using the Ångstrom equation, from hours of bright sunshine (n), day length (N) and solar radiation at the top of the atmosphere $R_{(a)}$. Both $R_{(a)}$ and N depend only on the latitude and day of the year:

$R_{(s)} = (a + b*n/N)* R_{(a)}$ where a and b are constants to be determined locally. Note that a represents the fraction of radiation reaching the earth's surface on a completely overcast day and (a+b) that on a completely clear day.

Note that the 0.77 in FAO equation (55) is $(1 - \alpha)$, where α is taken as 0.23 for the reference crop. (α = reflection coefficient)

$$R_{ns} = 0.77(0.25 + 0.50\frac{n}{N})R_a \qquad (55)$$

Net longwave loss ($R_{(nl)}$) is estimated by FAO equation (63). This is an old established equation that longwave loss depends basically on Stefan's law (σT^4) but increases with clear skies (n/N) and decreases with water vapour in the atmosphere ($e_{(d)}$). The modified formula was proposed by Boltzmann. The expression used is usually called the Stefan-Boltzmann equation.

Note that the 2.45* $10^{(-9)}$ is actually half Stefan's constant because maximum and minimum temperatures have been included separately in the equation. Often the data available include only mean temperature. In this case the mean temperature should be included twice in equation 63.

$$R_{nl} = 2.45.10^{-9}\left(0.9\frac{n}{N} + 0.1\right)\left(0.34 - 0.14\sqrt{e_d}\right)\left(T_{kx}^4 + T_{kn}^4\right) \qquad (63)$$

Equation (68) gives a rough estimate of the soil heat flux (G), but G is normally taken as zero, when the time span of calculation permits.

$$G = 0.14(T_{month_n} - T_{month_{n-1}}) = 0 \qquad (68)$$

List of symbols:

$R_{(n)}$	:	net radiation [MJ $m^{(-2)}$ $d^{(-1)}$]	
$R_{(ns)}$	:	net shortwave radiation [MJ $m^{(-2)}$ $d^{(-1)}$]	
$R_{(nl)}$	:	net longwave radiation [MJ $m^{(-2)}$ $d^{(-1)}$]	
$R_{(a)}$	:	extraterrestrial radiation [MJ $m^{(-2)}$ $d^{(-1)}$]	
n/N	:	relative sunshine fraction	
$T_{(kx)}$	:	maximum temperature [K]	(Tmax °C + 273)
$T_{(kn)}$	:	minimum temperature [K]	(Tmin °C + 273)
$e_{(d)}$	:	actual vapour pressure [kPa]	
G	:	soil heat flux [MJ $m^{(-2)}$ $d^{(-1)}$]	

Data requirements for the estimation of reference crop evapotranspiration

In principle, four climatic variables are required to estimate reference crop evapotranspiration.

a) temperature - preferably maximum and minimum;
b) relative humidity or vapour pressure;
c) windspeed at a height of 2 m;
d) solar radiation or hours of bright sunshine.

Various complexities arise in practice. Usually solar radiation is estimated from hours of bright sunshine. Estimation of net long-wave radiation- involves temperature, vapour pressure deficit, hours of bright sunshine (as a surrogate for cloud cover).

If there is one meteorological observation provided, this is usually at 0900 local time. Relative humidity at this time may not provide a good estimate of the daily mean. It is usually preferable to calculate vapour pressure at this time and to estimate the mean vapour pressure deficit by combining the actual vapour pressure with the saturation vapour pressure at maximum and minimum temperatures.

A calculation of reference crop evapotranspiration, using the FAO Penman-Monteith equation

A typical data set is that for Cairo (latitude 30°N, altitude 75 m) in June:

Mean temperature	28.5°C	Relative humidity	55%
Daily wind run at 2m	232 km	Hours of bright sun	11.5 h

1. Calculate vapour pressure deficit ($e_{(a)} - e_{(d)}$)

From tables, $e_{(a)}$ (saturated vapour pressure) at 28.5 C is 3.89 kPa
$e_{(d)}$ (actual vapour pressure) is RH*$e_{(a)}$ / 100 = 55*3.89/100 = 2.14 kPa
($e_{(a)} - e_{(d)}$) = (3.89-2.14) = 1.75 kPa

2. Calculate the available energy (R_n - G)

For this we need values of solar radiation at the top of the atmosphere ($R_{(a)}$) and daylength (N). These depend only upon the latitude and the day of the year. From tables (or by calculation) these for mid June are:

$R_{(a)} = 41.2$ MJm$^{(-2)}$ and N = 13.9 h

Equation (55) estimates solar radiation, $R_{(ns)}$, retained at the surface as:

$R_{(ns)}$=0.77 * (0.25 + 0.50*11.5/13.9) * 41.2 = 0.77 * 27.3 = 21.0 MJ m$^{(-2)}$

(Note that the incoming solar radiation is 27.3 MJ m$^{(-2)}$)

Equation (63) estimates the net longwave loss ($R_{(nl)}$) from Stefan's constant, temperature, n/N (which represents the amount of cloud cover) and the actual vapour pressure of the air ($e_{(d)}$)

$R_{(nl)}$=2.45*10$^{(-9)}$ * (0.9*11.5/13.9+0.1) * (0.34- 0.14*√(2.14)) * 2*(273+28.5)^4

Note that the temperature in the last bracket is in K (Kelvin) not °C because 273 has been added to the temperature in °C. Also as we do not have separate maximum and minimum temperatures as in equation (63), we use the mean temperature twice.

This becomes

$R_{(nl)}$ = 2.45*10$^{(-9)}$ * 0.845 * 0.135 * 2* (301.5)^4

= 2.45 * 0.845 * 0.135 * 2 * 82.63 * 10$^{(-1)}$

= 4.6 MJ m$^{(-2)}$

Assume G = 0 and then:

$R_{(n)}$ - G = $R_{(ns)}$ - R_{nl} - G = 21.0 - 4.6 - 0 = 16.4 MJ m$^{(-2)}$

3. Calculate other values required

We also need to know Δ (slope of saturated vapour pressure curve at this temperature) - from tables Δ = 0.226 kPa C$^{(-1)}$

γ (the psychrometric constant) = 0.066 kPa C$^{(-1)}$

Wind-speed $U_{(2)}$ in equation (69) is the mean 24 hour wind-speed in ms$^{(-1)}$. We have the daily wind-run in km per day (232 km). Therefore to get the mean wind-speed $U_{(2)}$ in ms$^{(-1)}$, divide the wind-run by the number of seconds in 24 hours and multiply by 1000 (km to m).

$U_{(2)}$ = 232 * 1000 / (24 * 3600) = 2.685 ms$^{(-1)}$

4. Substitute these values in equation (69)

Therefore $ET_{(0)}$ =

((0.408*0.226*16.4) + (0.066*(900/301.5)*2.685*1 .75))

/ (0.226 + 0.066*(I +0.34*2.685))

= 1.512 / 0.352 + 0.925 / 0.352

= 4.295 + 2.628 = 6.923 mm

The first term in the line above is often called the radiation term and the second the aerodynamic term. So ET_0 (reference crop evapotranspiration) is 6.9 mm per day.

9.9 Comparing Penman-Monteith with other methods

Comparison of the Penman-Monteith and original Penman equations

The original Penman equation is present in many agricultural textbooks and it is worthwhile comparing the two equations. Essentially Penman assumed that the canopy resistance ($r_{(c)}$) was zero and that the aerodynamic resistance ($r_{(a)}$) was one particular function of the daily wind run. This can be seen by comparing the two equations, though any numerical comparison must take account that Penman's equation has units of mm of water whereas the Penman-Monteith as presented here is in MJm$^{(-2)}$.

The function chosen by Penman has proved to be very robust and the Penman equation still gives good estimates of potential (i.e. reference crop) evapotranspiration. In many environments differences between the equations are virtually zero, but the Penman-Monteith equation has been adopted because of its complete theoretical treatment and more general applicability.

Penman-Monteith

$$\lambda E = \frac{\Delta(R_n - G) + \rho C_p(e_s - e)/r_a}{[\Delta + \gamma(1 + r_c/r_a)]}$$

Original Penman

$$E = \frac{\Delta(R_n - G) + \gamma 0.26(1 + 0.0062U)(e_s - e)}{[\Delta + \gamma]}$$

i.e. the equations are the same, allowing for constants and changes of units if

$r_{(c)} = 0$ and $1/r_{(a)}$ is proportional to $(1 + 0.0062U)$.

Pan Evaporation

The best estimates of reference crop evapotranspiration will be obtained from the above equations. Sometimes the necessary climatic data are not available but measurements of pan evaporation are. These do not give good estimates of crop evapotranspiration because their energy balance and resistances differ from crops. The best procedure is to correct the pan reading to an estimate of the reference crop value by multiplying by a pan coefficient. Pan coefficients depend on the type of pan, its exposure and the environment. Some values from FAO-24 (Doorenbos and Pruitt, 1976) are given below.

		Relative	Humidity	
	Class A Pan		Sunken Pan	
Wind run (km day$^{(-1)}$)	<40%	>70%	<40%	>70%
<175	0.65	0.85	1.00	1.00
175-425	0.60	0.75	0.85	0.90
>425	0.55	0.65	0.75	0.75

9.10 Conclusions

This chapter has mainly been concerned with the calculation of evapotranspiration. This has changed in two ways from the calculation used in early versions of Instat. The first is that Penman-Monteith is now used, rather than the FAO modified Penman method. The second is that there are dialogues, rather than commands, for the calculations.

Some readers may be curious why **only** use the Penman-Monteith equations are provided and there are no simpler formulae, particularly for stations that do not have all the measurements that are necessary. Our view is that it is probably preferable to use a nearby station for all or part of the measurements. However, the actual calculations are now in the macro called **pencalc.ins** so readers who would like to provide alternatives could adapt the macro. Macro writing is described in Chapter 14.

Chapter 10 – A Crop Performance Index

10.1 Introduction

Frère and Popov (1979) introduced a crop monitoring index based on a water balance calculated using actual precipitation and estimated evapotranspiration (Er) using Penman's method. The index is used for Early Warning Systems in a number of countries.

This monitoring scheme is introduced in Sections 10.2 and 10.3 as a case study of the use of climatic data. Extensions to the basic monitoring scheme include its use with many years of data, described in Sections 10.4.

Section 10.4 also describes preliminary analyses of the climatic data, and suggests ways of summarising the results from the index. These topics apply equally when more sophisticated crop models are used. These crop growth simulation models usually need daily data, and the use of daily records with this monitoring scheme is described in Section 10.5.

10.2 The crop monitoring scheme

The crop performance index is calculated from the difference between precipitation and the crop water requirements. The latter are calculated using crop water coefficients and Er. Allen *et al.* (1998)[7] provide guidelines for calculating crop water requirements for many crops, and this guide is one of the most popular pages on the FAO website (www.fao.org/documents).

The index has a maximum value of 100 and is reduced through the season if the crop is under stress. Stress, which results in a decrement in the index, can be caused by a water deficit or surplus.

Rainfall, potential evapotranspiration and crop coefficients must be supplied. together with a sowing date (or a criterion for sowing) and the soil water holding capacity. In Frère and Popov (1979) all calculations use 10-day data with a planting period assumed to be the first period with rainfall greater than 30 mm. The form of the calculations can be seen in Fig. 10.2a, taken from Frère and Popov, (1979) page 9 and using the format of FAO (1986).

The steps of the analysis leading to the index, I, given in Fig. 10.2a, are as follows:-

- The sowing date is the first decade in June. This is the first decade with precipitation greater than the 30 mm.
- The water requirement, WR, in each dekad is given by multiplying the PET by the crop coefficient Kc. For example, in the third decade in June
  WR = 44 * 0.4 = 18 mm
- The total water requirement for the season can then be calculated i.e.
  WR =Σ(PET * Kc)
  WR = 16 + 14 + 18 + ... + 20 = 328 mm in this example
- The difference between the actual rainfall, Pa, and the water required, WR, is calculated and added to the existing reserves (Ra) in the soil. These reserves cannot go above the maximum water holding capacity, 60 mm (in this case), nor can they become negative. A surplus or deficit is recorded under the S/D heading.
- For example, in the third dekad in June, P - WR = 115 - 18 = 97 mm. The reserves, Ra, were 51 mm, hence they reach the maximum value of 60 mm and there is 88 mm to be recorded under the S/D category.
- The index begins at 100 and is reduced in 2 ways. First, if there is a surplus of greater than 100 mm then the index is reduced by 3 units. Second, if there is a deficit, the index is reduced by the percentage of this deficit in relation to the total water requirements for the season.

[7] R.G. Allen, L.S. Pereira, D. Raes and M. Smith (1998) *Crop Evapotranspiration – Guidelines for Computing Crop Water Requirements.* FAO Irrigation and Drainage Paper 56.

Fig. 10.2a Cumulative water balance and crop performance index for Ziquinchor (Sahel)

FAO Agrometeorological Rainfed Crops Monitoring - SHEET 1
Station **Ziquinchor** Country **Senegal** Season **1978**
Lat. **12.33N** Long. **16.16** Alt. **23** m Crop/Cultivator Sorghum Lgs(no.of days) **130**
Soil Water Retention Capacity : **60** mm Total Water Requirements : **328** mm

No.	Dekad/Month	P(N)	P(a)	d(a)	PET	K(c)	WR	P-WR	R(a)	S/D	I%
	1/5	1	**0**	0	**64**						
	2/5	3	**0**	0	**61**						
	3/5	6	**2**	1	**62**						
1	1/6	23	**46**	3	**52**	**0.3**	16	30	30	0	100
2	2/6	40	**35**	4	**48**	**0.3**	14	21	51	0	100
3	3/6	62	**115**	6	**44**	**0.4**	18	97	60	88	100
4	1/7	118	**104**	6	**41**	**0.4**	17	87	60	87	100
5	2/7	121	**100**	5	**38**	**0.5**	19	81	60	81	100
6	3/7	124	**202**	7	**41**	**0.7**	29	173	60	173	97
7	1/8	176	**218**	9	**35**	**0.8**	28	190	60	190	94
8	2/8	180	**56**	5	**34**	**1.0**	34	22	60	22	94
9	3/8	176	**149**	7	**38**	**1.0**	38	111	60	111	91
10	1/9	130	**81**	5	**37**	**1.0**	37	44	60	44	91
11	2/9	120	**87**	7	**39**	**0.9**	35	52	60	52	91
12	3/9	111	**74**	6	**40**	**0.6**	24	50	60	50	91
13	1/10	85	**34**	3	**41**	**0.5**	20	14	60	14	91
	2/10	35	**72**	5	**41**						
	3/10	26			**42**						

P(N) = Normal precipitation / period
P(a) = Actual precipitation / period
d(a) = Number of days of rainfall
PET = Potential Evapotranspiration
K(c) = Crop Coefficient
WR = Water requirement of the crop = PET*Kc
R(a) = Water reserve in the soil
S/D = Surplus or deficit of water
I = Index

In this example there is a surplus of more than 100 mm on 3 occasions, which is why the index finishes at 91. There was never a deficit. Had there been a deficit of say 33 mm this, with the total water requirement of 328 mm, would have caused a drop of 10 units in the index.

10.3 The water satisfaction index dialogue

The data in Fig. 10.2a are in the worksheet called **FAOEx1.wor**, Fig. 10.3a, which has a title "**FAO Crop Monitoring Model – Ziquinchor May-Oct 1978**". Only three columns from Fig. 10.2a are needed, giving the rainfall, the evaporation and the crop coefficients. The length of the crop coefficient column determines the growing period, here it is a 130 day crop.

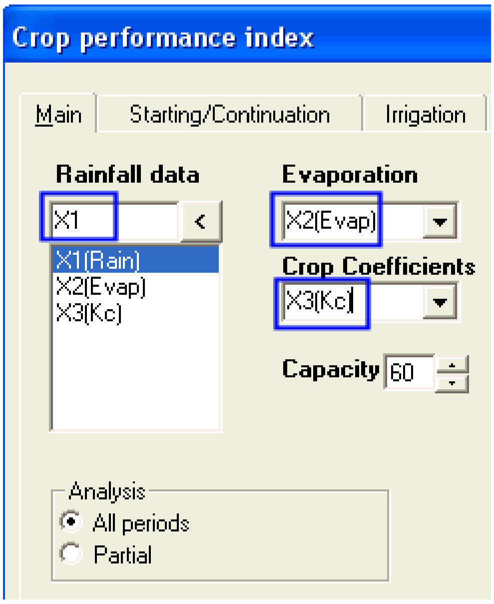
Fig. 10.3a Data for crop modelling in Instat
File => Open from library => FAOEx1.wor

	X1	X2	X3
	Rain	Evap	Kc
1	0	64	0.3
2	0	61	0.3
3	2	62	0.4
4	46	52	0.4
5	35	48	0.5
6	115	44	0.7
7	104	41	0.8
8	100	38	1
9	202	41	1
10	218	35	1
11	56	34	0.9

The simplest use of the Crop index dialogue is **Climatic => Crop Index** and complete the dialogue as shown in Fig. 10.3b.

Fig. 10.3b Crop Index dialogue and output

Climatic ⇒ Crop ⇒ Water Satisfaction Index



Crop performance index

Analysis for rainfall data in Rain
Total water requirement = 328mm.

Period Date	Crop	WR mm	Ra mm	Sur/Def mm	Index
4	1	15.6	30.4	0.0	100
5	2	14.4	51.0	0.0	100
6	3	17.6	60.0	88.4	100
7	4	16.4	60.0	87.6	100
8	5	19.0	60.0	81.0	100
9	6	28.7	60.0	173.3	97
10	7	28.0	60.0	190.0	94
11	8	34.0	60.0	22.0	94
12	9	38.0	60.0	111.0	91
13	10	37.0	60.0	44.0	91
14	11	35.1	60.0	51.9	91
15	12	24.0	60.0	50.0	91
16	13	20.5	60.0	13.5	91

(Loss from surplus = 9.0 in 3 periods)
:

The first column in Fig. 10.3b shows the crop was planted in period 4. This was the first decade in June. It was a 130-day crop as indicated by the crop periods given in the second column. The remaining columns give the water requirement, WR; the soil reserves, Ra; the surplus or deficit and the value of the index. The final line of results indicates the drop in the index from 100 to 91 is solely the result of waterlogging, rather than water shortage.

Fig. 10.3c Cumulative water balance for Dori, 1978

FAO Agrometeorological Rainfed Crops Monitoring - **SHEET 1**
Station **Dori** Country **Burkina Faso** Season **1978**
Lat. Long. Alt. m Crop/Cultivator Lgs(no.of days) **90**
Soil Water Retention Capacity : mm Total Water Requirements : mm

No.	Dekad/Month	P(N)	P(a)	d(a)	PET	K(c)	WR	P-WR	R(a)	S/D	I%
	1/5		4		75						
	2/5		11		78						
	3/5		14		80						
	1/6		23		68						
	2/6		18		63						
1	3/6		58		59	0.3	18	40	40	0	100
2	1/7		51		59	0.4	24	27	60	5	100
3	2/7		48		57	0.5	29	19	60	19	100
4	3/7		134		59	0.8	47	87	60	87	100
5	1/8		9		48	1.0	48	-39	21	0	100
6	2/8		29		47	1.0	47	-18	3	0	100
7	**3/8**		**28**		**50**	**1.0**	**50**	**-22**	**0**	**-19**	**94**
8	1/9		15		47	0.6	28	-13	0	-13	90
9	2/9		32		50	0.5	25	7	7	0	90
	3/9		6		52						
	1/10		0		55						
	2/10		3		59						
	3/10		0		59						

P(a) = Actual precipitation / period K(c) = Crop Coefficient
PET = Potential Evapotranspiration I = Index

Fig. 10.3c gives a second example, also from Frère and Popov (1979). This is for Dori, also in the Sahel. Here a 90-day crop is grown and the index decreases because of a water shortage in August. There is, for example, a deficit of 19 mm in the third decade of August. With a total water requirement of 316 mm, this is 6% of that value. Hence the index drops by 6 units to 94.

Repeating the steps described for Ziquinchor gives the results shown in Fig. 10.3d.

Fig. 10.3d Crop index for data from Dori

File => Open from library => dori.wor

Climatic ⇒ Crop ⇒ Water Satisfaction Index

	X1	X2	X3
	Rain	Evap	Coeff
1	4	75	0.3
2	11	78	0.4
3	14	80	0.5
4	23	68	0.8
5	18	63	1
6	58	59	1
7	51	59	1
8	48	57	0.6
9	134	59	0.5
10	9	48	
11	29	47	
12	28	50	
13	15	47	
14	32	50	
15	6	52	
16	0	55	
17	3	59	
18	0	59	

```
Crop performance index

Analysis for rainfall data in Rain
Total water requirement = 315mm.
Period      WR        Ra     Sur/Def   Index
Date Crop   mm        mm       mm
6    1     17.7      40.3     0.0      100
7    2     23.6      60.0     7.7      100
8    3     28.5      60.0    19.5      100
9    4     47.2      60.0    86.8      100
10   5     48.0      21.0     0.0      100
11   6     47.0       3.0     0.0      100
12   7     50.0       0.0   -19.0       94
13   8     28.2       0.0   -13.2       90
14   9     25.0       7.0     0.0       90
:
```

When the crop dialogue is used in its simplest form, other aspects of the analysis have default values which are consistent with those used by Frère and Popov, (1979). Thus

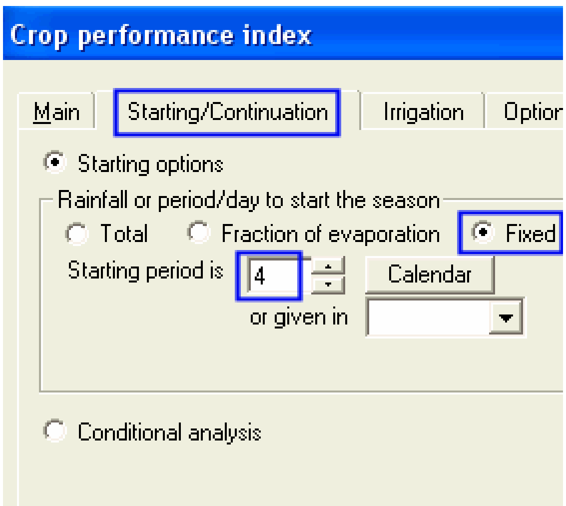
- the starting period is the first with rainfall greater than 30 mm
- the soil capacity is 60 mm
- 3 units are lost from the index if there is more than 100 mm surplus.

These default values can be changed with appropriate subcommands

Take the sowing period as an example. This can be fixed as decade 4, when there was 23mm of rain, as shown in Fig. 10.3e

Fig. 10.3e Specifying the start dekad

Climatic ⇒ Crop ⇒ Water Satisfaction Index ⇒ Starting/Continuation

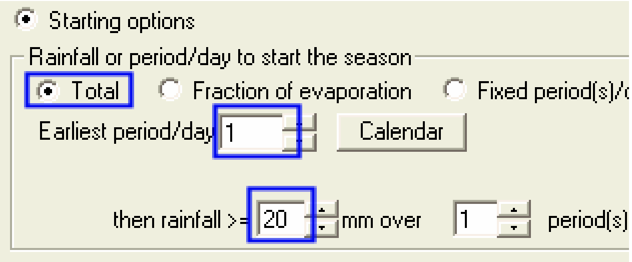



```
Crop performance index

Analysis for rainfall data in Rain
Total water requirement = 340mm.
Period      WR        Ra     Sur/Def   Index
Date Crop   mm        mm       mm
4    1     20.4       2.6     0.0      100
5    2     25.2       0.0    -4.6       99
6    3     29.5      28.5     0.0       99
7    4     47.2      32.3     0.0       99
8    5     57.0      23.3     0.0       99
9    6     59.0      60.0    38.3       99
10   7     48.0      21.0     0.0       99
11   8     28.2      21.8     0.0       99
12   9     25.0      24.8     0.0       99
:
```

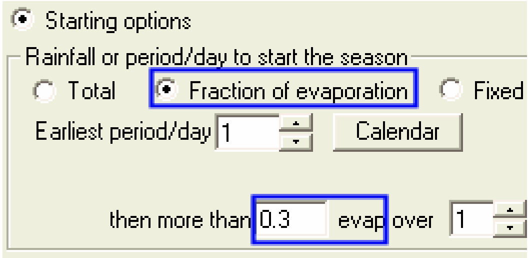
To specify a starting period to be rainfall is greater than 20mm, rather than the default of 30mm, the option in the dialogue is completed as shown in Fig. 10.3f.

Fig. 10.3f Starting period when first rainfall > 20 mm

first rainfall > 20 mm



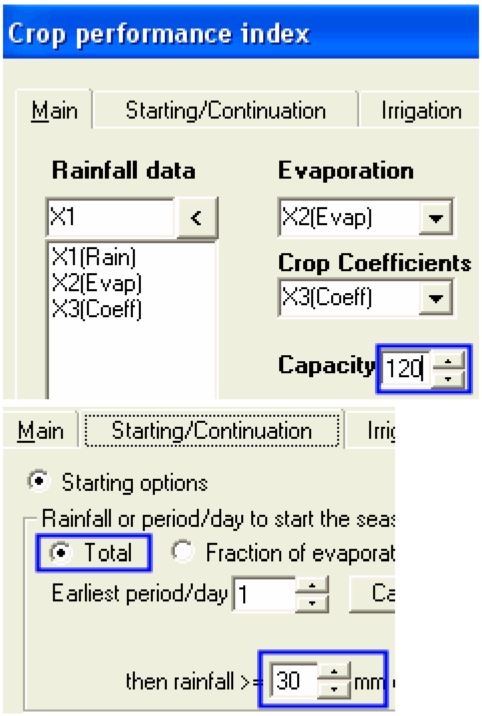
first rainfall > 30% of evaporation



Finally, with a crop coefficient of 0.3, the sowing period could be the first with rainfall greater than 30% of the evaporation, Fig. 10.3f. This is the assumption made for sowing in FAO (1986). For Dori, each of these three dialogues produces the same result, which is shown in Fig. 10.3e, namely that planting is in period 4 and the index at the end of the season is then 99, rather than 90.

Fig. 10.3g considers the original sowing criterion but changes the capacity to 120 mm, instead of the default value of 60 mm. In this case the index remains at 100.

Fig. 10.3g Change the capacity from 60mm to 120mm



Crop performance index

Analysis for rainfall data in Rain
Total water requirement = 315mm.

Period Date	Crop	WR mm	Ra mm	Sur/Def mm	Index
6	1	17.7	40.3	0.0	100
7	2	23.6	67.7	0.0	100
8	3	28.5	87.2	0.0	100
9	4	47.2	120.0	54.0	100
10	5	48.0	81.0	0.0	100
11	6	47.0	63.0	0.0	100
12	7	50.0	41.0	0.0	100
13	8	28.2	27.8	0.0	100
14	9	25.0	34.8	0.0	100

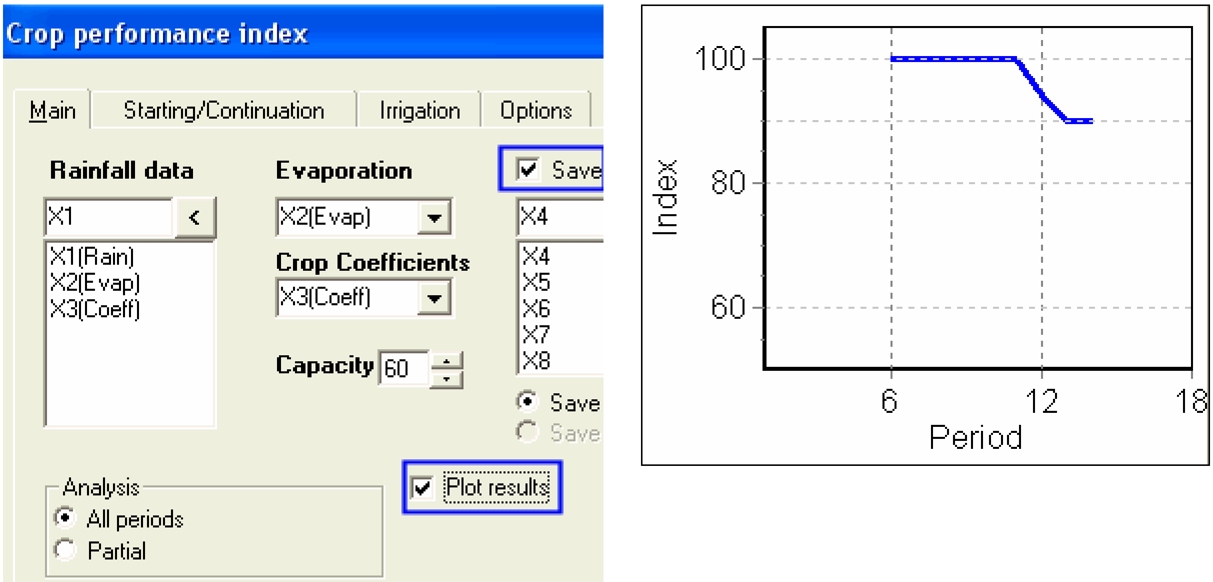
:

The successive values of the index can be saved and plotted, as shown in Fig. 10.3h, for the initial analysis for Dori.

The analyses so far have all used dekads, as in the FAO publications. However, the dialogue can be used with groups of any size. All that is required is that the climatic and crop variables are consistent with each other (i.e. each rainfall period should be grouped over the same number of days as the corresponding evaporation period, and each crop coefficient should correspond to these climatic periods).

Choosing the crop coefficients for a given run of the index, can draw on the work by Allen et al (1998) for FAO. This provides the coefficients for many crops, for each of the main growth stages, and also tabulates the length of these stages.

Fig. 10.3h Save and possibly graph the index



Once the values and the lengths are available, the Climatic ⇒ Crop ⇒ Crop Coefficients dialogue may be used to generate values for each period. An example is in Fig. 10.3i for the data in FAOex1.wor. This was a 130-day crop, with coefficients shown in Fig. 10.3i. This is roughly as follows:

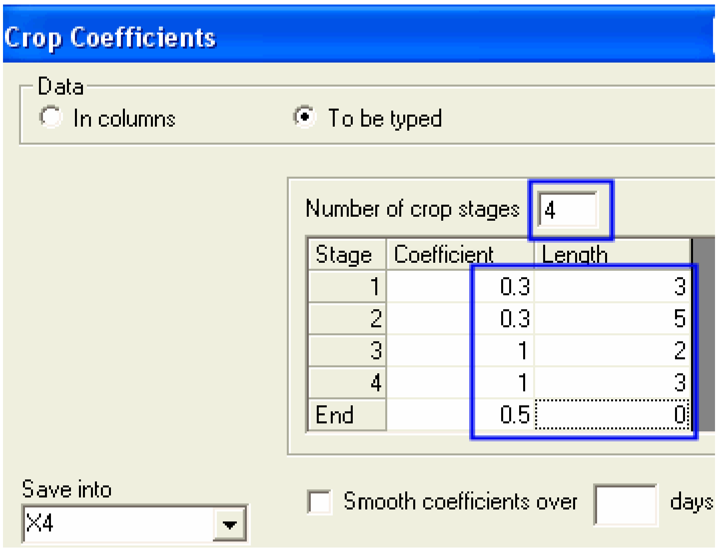
- Initial 30 days Index at 0.3
- Growth 50 days Index increases to 1.0
- Mid 20 days Index remains at 1.0
- Final 30 days Index drops to 0.5

The way this information is input to Instat is shown in Fig. 10.3i, together with the results, in X4.

Fig. 10.3i Calculating coefficients from information at the crop stages

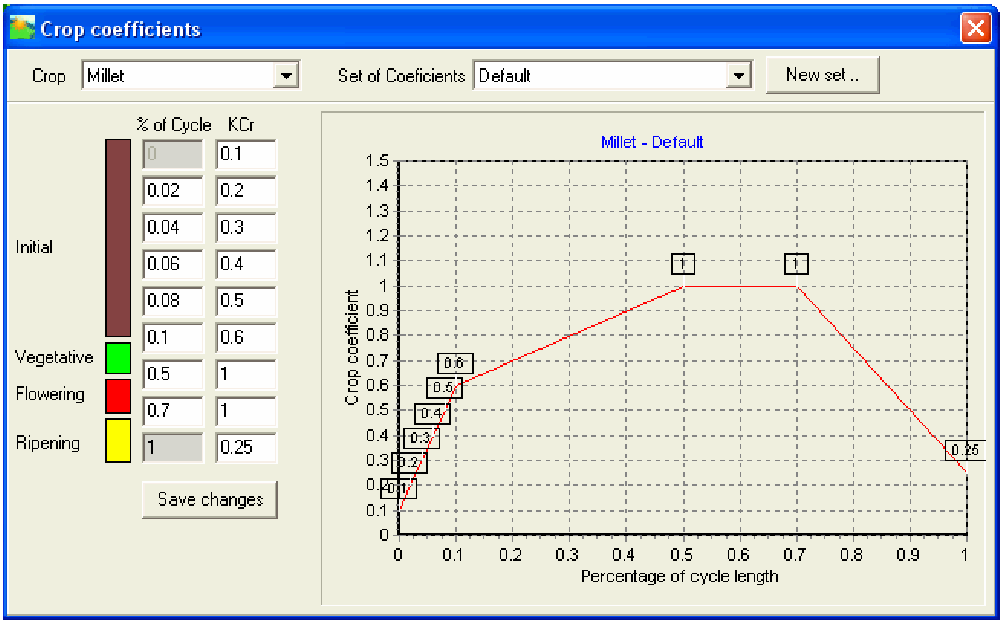
`Climatic ⇒ Crop ⇒ Crop Coefficients`

	X3	X4
	Kc	
1	0.3	0.3
2	0.3	0.3
3	0.4	0.3
4	0.4	0.44
5	0.5	0.58
6	0.7	0.72
7	0.8	0.86
8	1	1
9	1	1
10	1	1
11	0.9	0.833
12	0.6	0.667
13	0.5	0.5



FAO have also upgraded their own software, now called AgroMetShell, that calculates and maps the water satisfaction index. This includes a flexible method of choosing the crop coefficients, shown for millet in Fig. 10.3j.

Fig. 10.3j FAO AgroMetShell calculation of crop coefficients



10.4 The water satisfaction index with many years of data

Fig. 10.4a Malawicr data

File ⇒ Open From Library ⇒ malawicr.wor

X1 - X27	Rainfall totals for 24 dekads (from 1st Oct.) for 1954/5 to 1982/3 (omitting 1962/3 & 1981/2)
X28	Evaporation totals for dekads (from 1st October)
X29	Crop coefficients for groundnut (11 periods)
X30	Crop coefficients for cotton (17 periods)
X31	Numbers 1,2,... 17 (for plotting)
X32	Numbers 1,2,... 24 (for plotting)

	X1	X2	X3	X4	X5	X6	X7	X8	X9
1	0	0	0	0	0	18.3	0	0	0
2	0	0	18.8	0	0	0	5.8	4.3	8.9
3	0	0	0	0	0	9.1	2.5	0	0
4	18.8	0	0	0	0	4.8	3.6	0	123.2
5	59.2	9.9	52.1	0	17	3	0	30.7	25.4
6	42.2	65.3	45	0	24.9	3.6	14	13.2	61.7
7	43.4	71.1	58.7	8.6	37.6	15	2	102.4	39.4
8	16.8	112.5	79.5	74.4	32.3	128.5	73.2	128.8	94
9	67.1	72.9	83.6	51.3	16.3	19.8	67.8	13.4	118.1
10	47.5	10.4	95.5	224	84.6	12.7	9.4	91.2	49.5
11	152.6	131.3	132.3	53	49	1.5	89.9	43.1	170.2
12	105.4	44.2	30.5	208.3	26.2	62.7	31.2	27.4	25.4
13	191.3	182.6	63.8	46.2	66	20.6	86.9	82.3	59.7
14	193	80.8	56.6	135.1	82.6	62.5	132.8	194.3	48.8
15	111.3	8.2	3.8	19.6	41.9	35	8.9	5.1	41.9
16	101.6	71.6	134.1	44.7	94.5	82.2	156.5	0	24.1
17	0	102.4	5.8	12.2	23.6	67.6	20.8	88.6	58.4
18	0	78.2	23.6	20.8	1.3	0	18	75.9	0
19	0	107.2	26.9	5.8	0	6.4	46.7	16.5	0
20	0	0	57.7	0	0	0	30.7	0	0
21	8.3	32.5	0	0	0	0	4.3	0	0
22	66.8	0	3.6	0	0	0	5.1	0	0
23	0	6.6	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0

When many years of rainfall data are available, the year-to-year variation of the index can be evaluated. This can put the values for the current year into perspective. For example, if the index is 88, halfway through the current season, this value can be compared with the historical record at the same point in the season.

A further option is to continue from this point, to give the distribution of final values of the index in past years. This is called a **conditional analysis**, since all years are processed **conditional** on the starting values from the current season. An example is at the end of this section.

This section uses ten-day data for 27 years from a site in Malawi. Part of the data from this file, called **Malawicr.wor**, are shown in Fig. 10.4a.

10.4.1 Preliminary examination of the data

The first step in such a study should be an examination of the rainfall data. Here two topics are outlined, where the statistics are often misused:

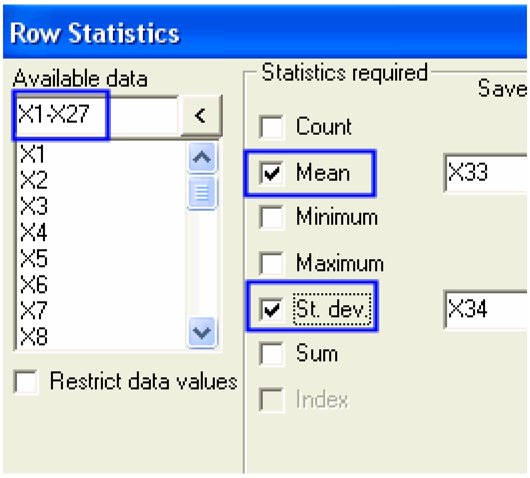
- The use of rainfall and evaporation to examine season lengths
- The use of the coefficient of variation (c.v.)

In a full study, the rainfall records would be transposed, so each column holds the data for a single decade within the year, rather than for each year. This idea was described in Chapter 7. With the data in its current form, the **Manage ⇒ Manipulate ⇒ Row Statistics** dialogue can be used to look at summary values (Fig. 10.4b).

Fig. 10.4b Calculating summary statistics for Malawi data

Manage ⇒ Manipulate ⇒ Row Statistics

Manage ⇒ Calculate (x35 = x28/2)



	X32	X33	X34	X35
	Dekad	Mean	Stdev	halfevap
1	1	2.1	5.0	28.8
2	2	2.7	7.2	29.5
3	3	5.0	8.8	30.3
4	4	12.5	24.4	30.8
5	5	15.4	19.0	31.3
6	6	34.5	31.9	34.3
7	7	59.1	45.1	27.5
8	8	62.6	39.0	25.5
9	9	54.6	48.9	24.3
10	10	64.4	48.8	23.3
11	11	64.9	50.8	22.5
12	12	61.3	47.0	22.5

The calculation of the mean and standard deviation for each dekad is shown in Fig. 10.4b. The mean is to be compared with half the evaporation, so that is also calculated, Fig. 10.4b.

Comparing x33 with x35 (Fig. 10.4b) shows that dekad 6 is where the mean rainfall first exceeds half the evaporation. The results though the year can be plotted, see Fig. 10.4c.

This plot is sometimes used to give the "average" length of the season, i.e. the season is the period during which the average rainfall (X33) is greater than half the evaporation (X35). Here it is periods 6 to 19. It does give a first indication, though the difficulties in interpreting this type of result have been discussed in earlier chapters, for example Section 6.5.

It is also useful to have an idea of the variability of the season length from year to year. The start is easy. For example, defining the start of the season as the first decade with rainfall of more than half the evaporation then the Start of the Rains dialogue, described in Chapter 6, can be used. It is designed for daily data, but can equally be used with dekads. It is shown in Fig. 10.4d, together with the results generated.

Fig. 10.4c Plot of mean rainfall with evaporation/2

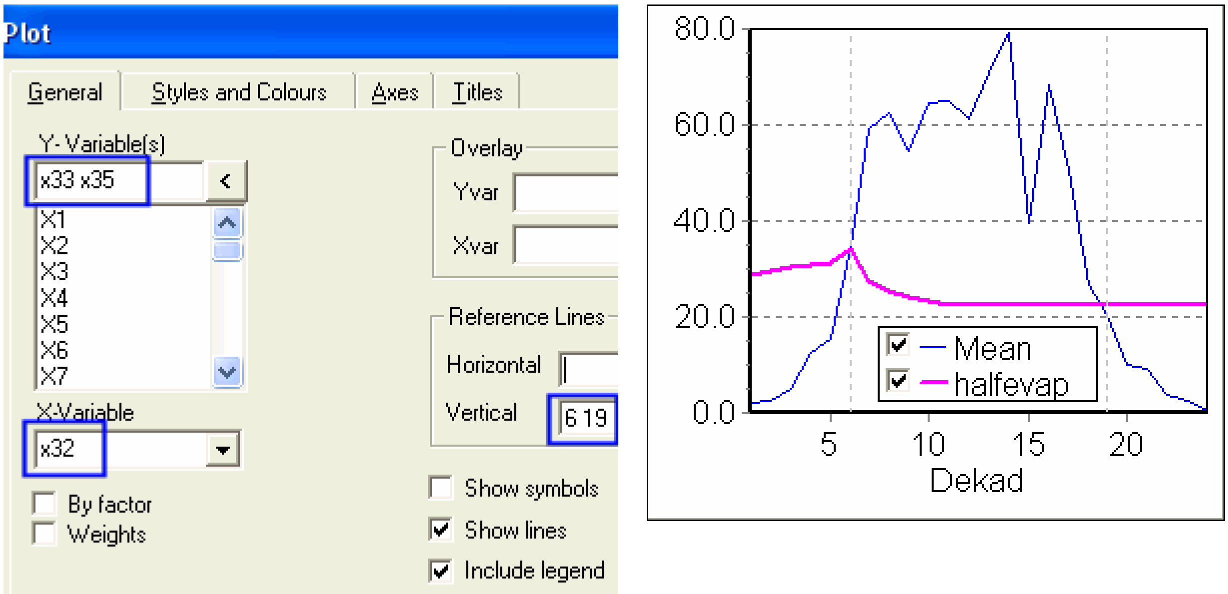
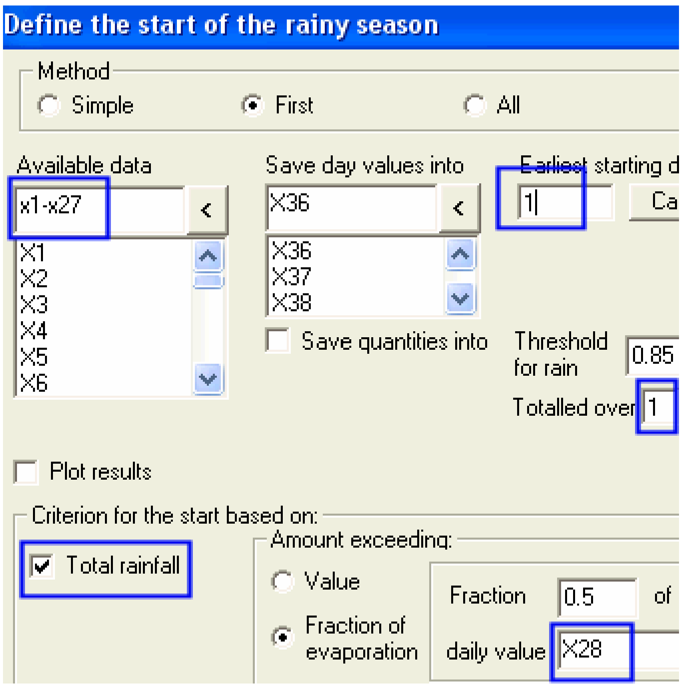
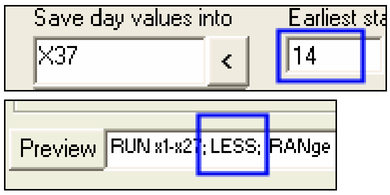


Fig. 10.4d Calculating the start and end of the season

`Climatic ⇒ Events ⇒ Start of the Rains`



Changes for the end of the season

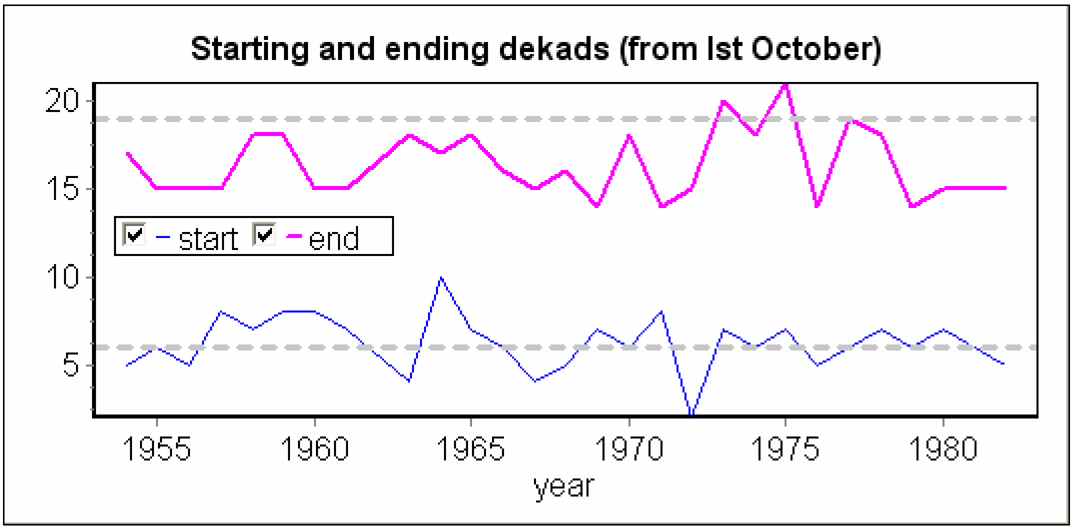


	X36	X37	X38
	start	end	year
1	5	17	1954
2	6	15	1955
3	5	15	1956
4	8	15	1957
5	7	18	1958
6	8	18	1959
7	8	15	1960
8	7	15	1961
9	4	18	1963
10	10	17	1964
11	7	18	1965

The same dialogue may also be used for the end of the season, as indicated in Fig. 10.4d. The starting dekad is changed, here to dekad 14, because that was the one with the highest mean rainfall in the season, see Fig. 10.4c. Then we want to find the first dekad from 14, where the rainfall is **less** than half the evapotranspiration. The option to look for less is not on the dialogue, but it can be added to the command that generates the analysis, as shown in Fig. 10.4d.

A summary of the results in Fig. 10.4d show the mean start is 6.2, very close to the value in Fig. 10.4c. The mean of the ending dekads is 16.4, a long way from the value of dekad 19 in Fig. 10.4c. The values each year are plotted in Fig. 10.4e, with the supposed average dekads from Fig. 10.4c marked.

Fig. 10.4e The start and end of the season



Comparing Fig. 10.4c and 10.4e shows that, if the value each year is available, then the methods shown in Fig. 10.4d should be used. This calculates the starting and ending dekad each year, and then averages these values.

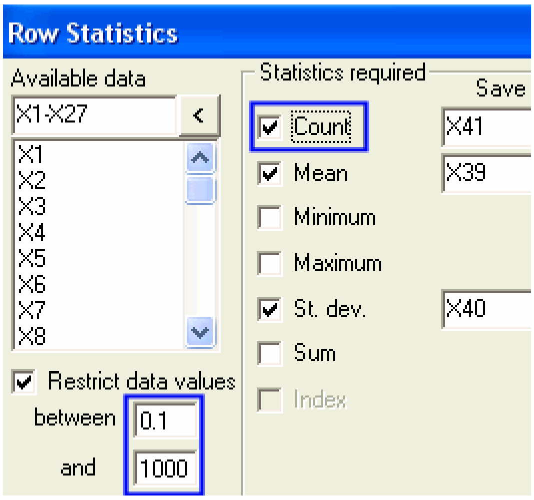
The second topic involves examining the rainfall data through the season. Fig. 10.4a shows the decade totals are often zero at the start and end of the "year". When analysing data with zeros it is usually useful to split the analysis into two parts. The number of zeros is considered first and then the non-zero values are analysed further.

If the data in Fig. 10.4a were transposed to be by decade, then the **Climatic ⇒ Examine** dialogue, could be used. It includes an option to tabulate and graph the zeros. With the data as in Fig. 10.4a, use the **Manage ⇒ Manipulate ⇒ Row Statistics** dialogue, as shown in Fig. 10.4f. In the dialogue "restrict" the data to omit the zeros and include the count of the number of observations.

Fig. 10.4f Dealing with zeros in data

Manage ⇒ Manipulate ⇒ Row Statistics

Manage ⇒ Calculate ⇒ x42 = 100*(27-x41)/27

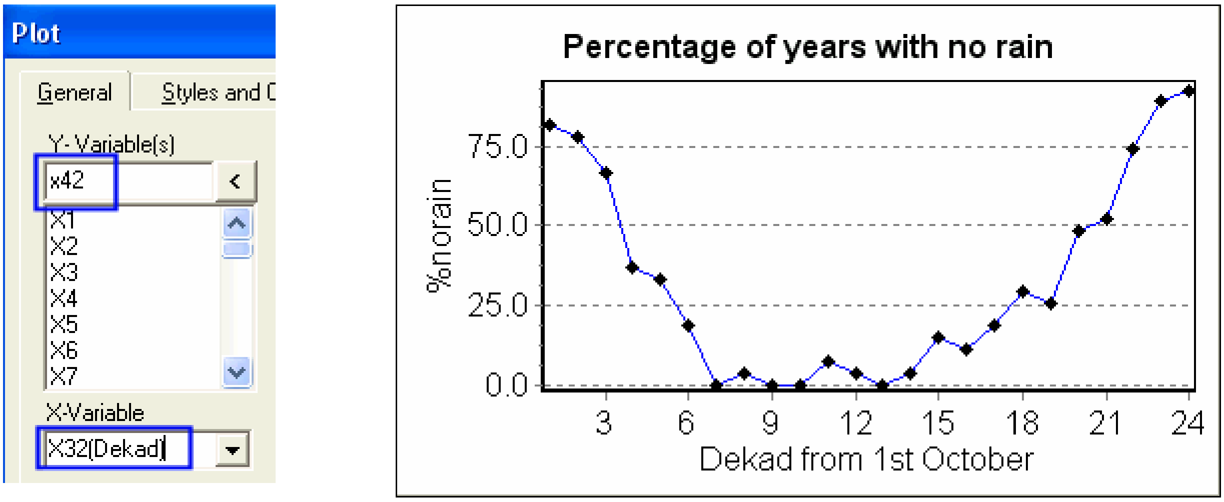


	X39	X40	X41	X42
	Mea	Std	rain	%norain
1	11.2	5.7	5	81.5
2	12.1	11.5	6	77.8
3	14.9	9.3	9	66.7
4	19.9	28.5	17	37.0
5	23.1	19.0	18	33.3
6	42.3	30.3	22	18.5
7	59.1	45.1	27	0.0
8	65.0	37.7	26	3.7
9	54.6	48.9	27	0.0
10	64.4	48.8	27	0.0
11	70.1	49.2	25	7.4

For example, Fig. 10.4f shows that the 3rd dekad in October had rain in 9 of the 27 years (X41). Hence the calculation shows that 2/3 of these years were dry, (X42). In the 9 years with rain the mean was 14.9mm and the standard deviation 9.3mm.

The percentage of years can then be plotted, as shown in Fig. 10.4g.

Fig. 10.4g Plot of the percentage of zeros



Now look at the non-zero values. The coefficient of variation is often calculated in climatic studies. This is defined as

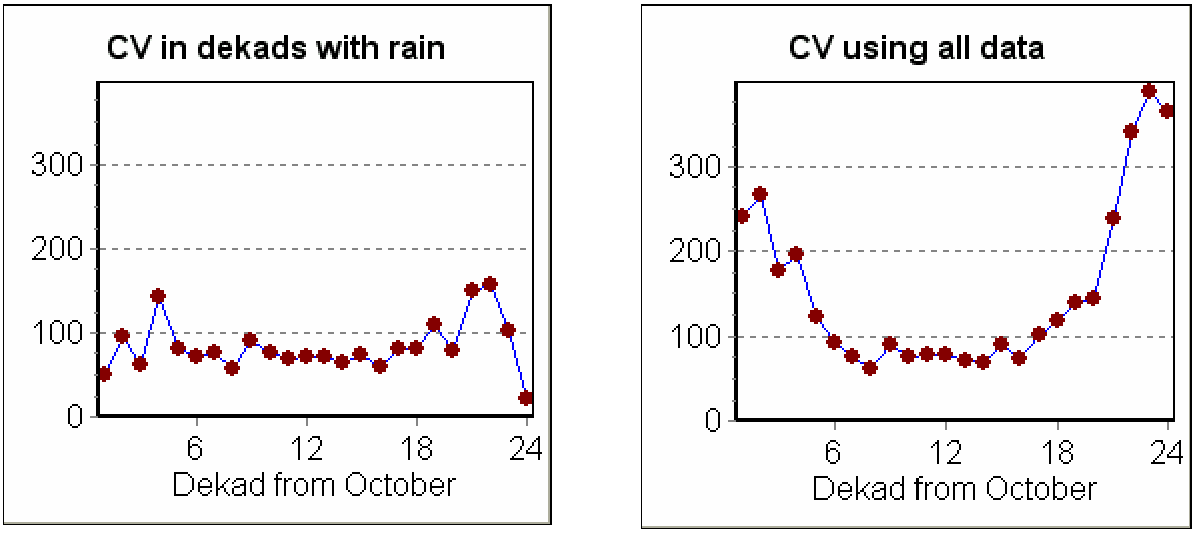
c.v. = 100 * (Standard deviation)/Mean

i.e. here x44 = 100 * x40 / x39

Fig. 10.4h Coefficient of variation of 10 day rainfall data through the season

`Manage ⇒ Calc ⇒ x43=100*x40/x39`
`Graphics ⇒ Plot x43 by dekad`

`Manage ⇒ Calc ⇒ x44=100*x34/x33`
`Graphics ⇒ Plot x44 by dekad`



Use the **`Manage => Calculations`** dialogue to give the c.v. for each dekad, using the non-zero values, then graph the c.v. as shown in the left side of Fig. 10.4h. The coefficient of variation seems roughly constant through the year, at about 80%. Its value is, of course, much less "stable" at the start and end of the year, because there are then fewer non-zero values.

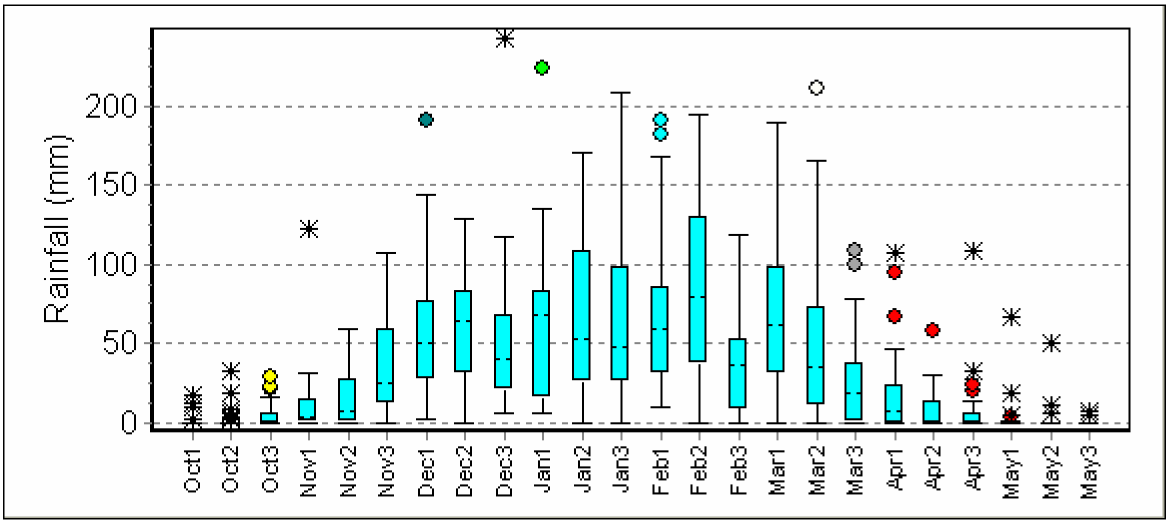
The c.v. is often reported in the summary of rainfall totals, but usually without omitting the zero values first. This uses the results from our initial calculation of summary statistics, shown in Fig. 10.4b. Repeating the calculation and plot gives the results in the right of Fig. 10.4h. Although the shape looks impressive, a comparison with Fig. 10.4g shows that this is effectively a very complicated way of giving the same information, namely that non-zero values are less likely in the middle of the season!

The general message is that when there is obvious structure in the data, and here that structure includes "dry decades", then take great care with any analysis that ignores the structure.

Once the zero values have been plotted, as in Fig. 10.4g, check that the c.v. provides a useful summary of the data. Often it is easier to interpret the mean, and perhaps the standard deviation separately. Here, the dekad totals are very skew, hence using the standard deviation is suspect. Often percentiles are more appropriate summaries. As an example, the data in **malawicr.wor** were transposed, and then plotted as boxplots, Fig. 10.4i.

Fig. 10.4i Boxplots of the 10.day data

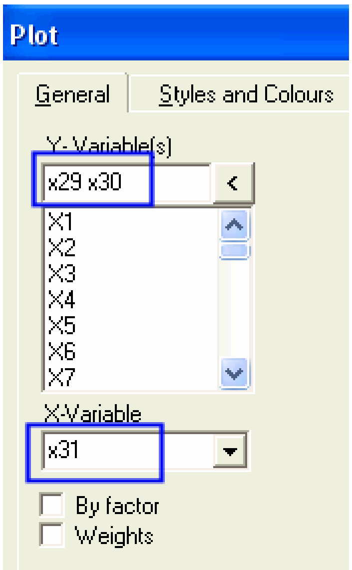
`File ⇒ New, Manage ⇒ Data ⇒ Copy x1-x27 from malawicr to x1-x24, then Graphics ⇒ Boxplot`

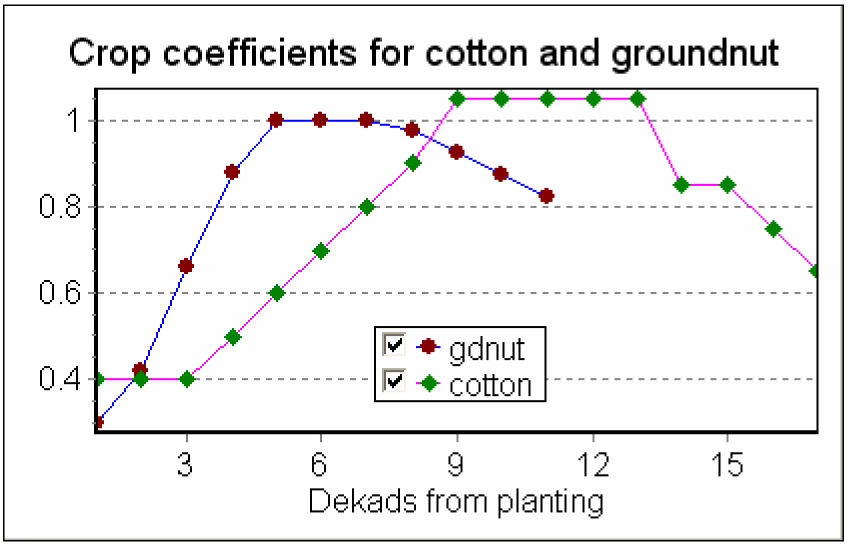


10.4.2 Year to year variation of the crop indices

After the preliminary analysis, the main task is to look at the crop index dialogue for each year. Fig. 10.4j plots the crop coefficients for cotton and groundnuts. We start by looking at the cotton data and then compare the water satisfaction index for the two crops.

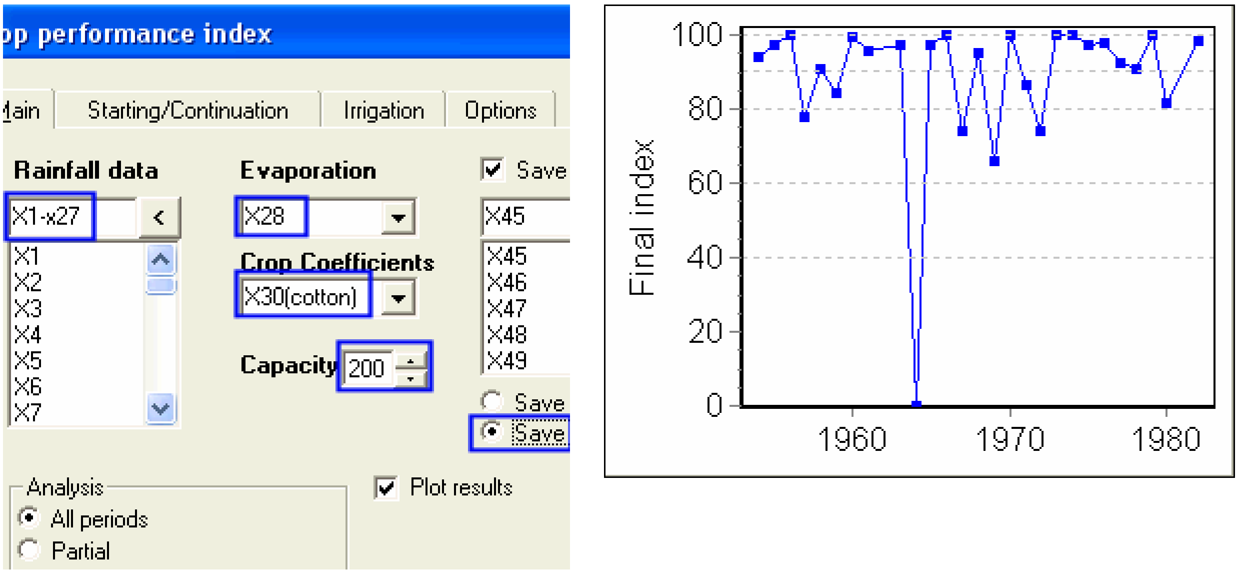
Fig. 10.4j The crop coefficients for cotton and groundnut





In the analysis for cotton the soil water capacity is taken as 200mm. Then the dialogue is as Fig. 10.4k.

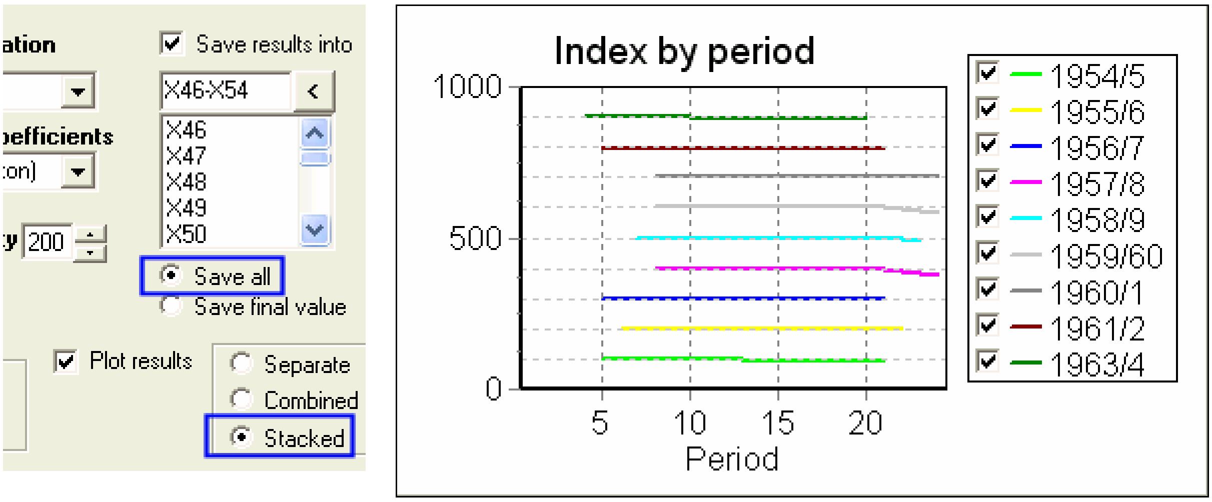
Fig. 10.4k The crop index for 27 years of data – 170-day cotton crop



The results for the first year are given in Fig. 10.4k along with a graph for the final index for each year. The dip to zero in year 10 is because it was not possible to sow in that year (1964/65) and Instat gives a warning message in the output for X10.

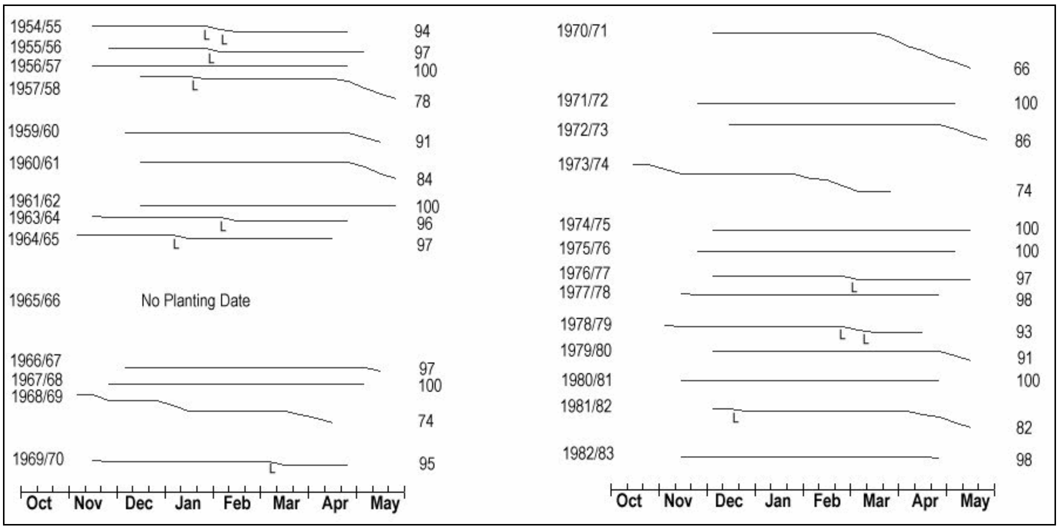
Fig. 10.4l shows part of the **Crop Index** dialogue, with the option to save and plot all the values for the index in each year. A single "stacked-graph" is chosen as the presentation and just the first 9 years are analysed.

Fig. 10.4l Stacked graph of crop index for cotton



The curious y-axis in Fig. 10.4l indicates that Instat's current graphics capabilities have been "stretched". The presentation is useful as an exploratory tool. An alternative is to export the numerical results, so that a more flexible graphics package can be used. Fig. 10.4m shows the result, using Excel, which has all the information in Fig. 10.4l and has also differentiated between the losses resulting from water logging and water shortage.

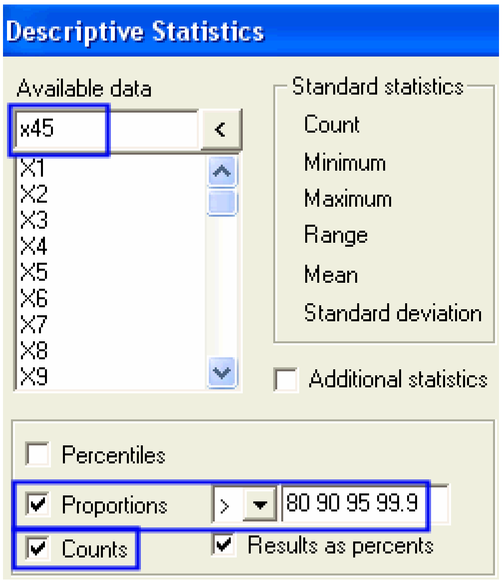
Fig. 10.4m Variation of the crop index for 27 years of data from a site in Malawi. L indicates a reduction due to waterlogging.



With the crop performance index dialogue used as in Fig. 10.4k, the value of the final index each year has been saved in X45. This column can now be summarised. Sowing was not possible in one of the years (1964/5). Once this fact is noted, we put the value as "missing", by typing * in **row 10** of **X45** in the spreadsheet window. (An alternative is to type **X45(10) = *** in the Commands and Output window.) The results for the remaining years can then be summarised, Fig. 10.4n.

Fig. 10.4n Summary statistics of the final index for cotton

`Statistics ⇒ Summary ⇒ Describe`



Descriptive Statistics

Column	X45
No. of observations	27
No. not missing	26
Minimum	65.634
Maximum	100
Range	34.366
Mean	91.746
Std. deviation	9.7423
Count > 80	22
Count > 90	19
Count > 95	14
Count > 99.9	6
% of data > 80	84.6
% of data > 90	73.1
% of data > 95	53.8
% of data > 99.9	23.1

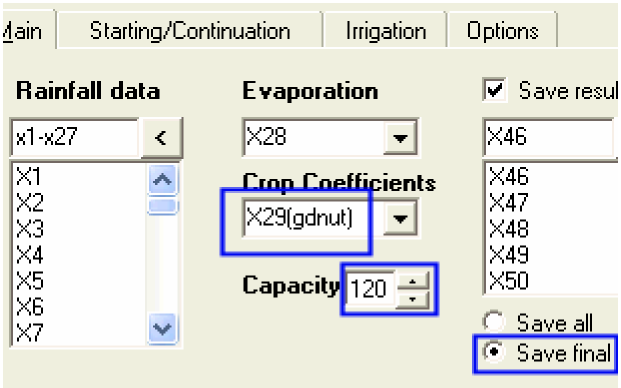
Fig. 10.4n shows that the index remained at 100 for 6 of the years, more than 20%. It was more than 90 in 19, i.e. almost ¾ of the years

In a number of the years the index dropped at the end of the season. Perhaps a 170-day crop is too long for this site. A plot of the crop coefficients for cotton and groundnut was in Figs 10.4i and 10.4o shows the use of the **Crop Index** dialogue for the groundnuts. Here the soil capacity within the range of the groundnut roots was taken as 120 mm. We have also assumed that up to an extra 180 mm (rather than the default of 100 mm) does not cause any harm to the crop.

The numerical results for the two crops are compared in Figs 10.4o and 10.4p. There were 2 very low values for the groundnuts, namely 1967/8 and 1972/3. In these years the index for cotton was low, but that for the groundnuts was even lower. In the other years the index for groundnuts was never lower than 90. The most dramatic difference was in 1969/70, row 15, when the index for cotton was its lowest (at 66), while the groundnuts were almost over before the problems, and therefore had an index of 90. Note also that the index for groundnuts was 100 in 1964/5 when there was no sowing date for cotton.

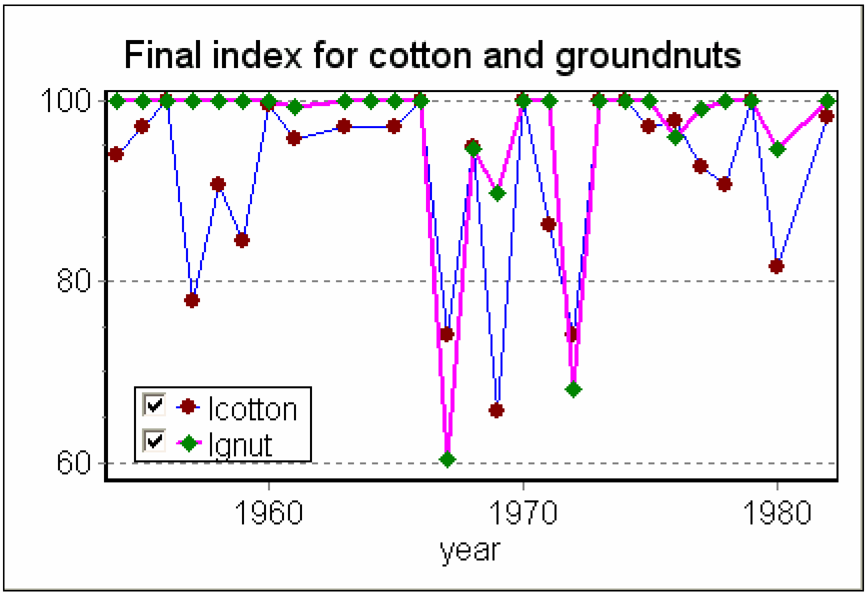
Fig. 10.4o Comparison of the final index for cotton and groundnut

`Manage ⇒ Remove(Clear) ⇒ X44-X75`
`Climatic ⇒ Crop Index`



`Manage ⇒ Calc`
`x47=x46-x45`

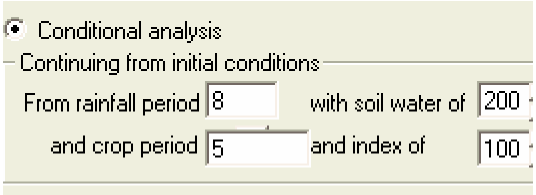
	X38	X45	X46	X47
	year	cottor	lgnut	diff
1	1954	94	100	6
2	1955	97	100	3
3	1956	100	100	0
4	1957	78	100	22
5	1958	91	100	9
6	1959	84	100	16
7	1960	100	100	0
8	1961	96	99	4
9	1963	97	100	3
10	1964	*	100	*
11	1965	97	100	3
12	1966	100	100	0
13	1967	74	60	-14
14	1968	95	95	0
15	1969	66	90	24
16	1970	100	100	0
17	1971	86	100	14
18	1972	74	68	-6
19	1973	100	100	0
20	1974	100	100	0
21	1975	97	100	3
22	1976	98	96	-2
23	1977	93	99	7
24	1978	91	100	9
25	1979	100	100	0
26	1980	82	95	13
27	1982	98	100	2



This example indicates how climatic data can be used to compare alternative cropping strategies. Even with this simple index, it might now be instructive to look in more detail at the years when one crop fared worse than the other. The main concepts of using a modelling approach to indicate differences in cropping strategies are also unchanged if a more realistic crop simulation model, is substituted for the simple FAO index.

Finally in this section a **conditional** question is considered. Suppose that in the current year cotton was sown in period 4, and after 40 days the index was still 100 with the soil full. What would have happened to the index thereafter in the past years, **if the season had started in this way**? The main part of the dialogue is the same as shown in Fig. 10.4k dialogue to investigate this is shown in. The **conditional analysis** tab was used and completed as shown in Fig. 10.4p. The option was also used to display results only in periods when the index changed. The results for one year (1980/81) are shown in Fig. 10.4q when the index dropped to 91.

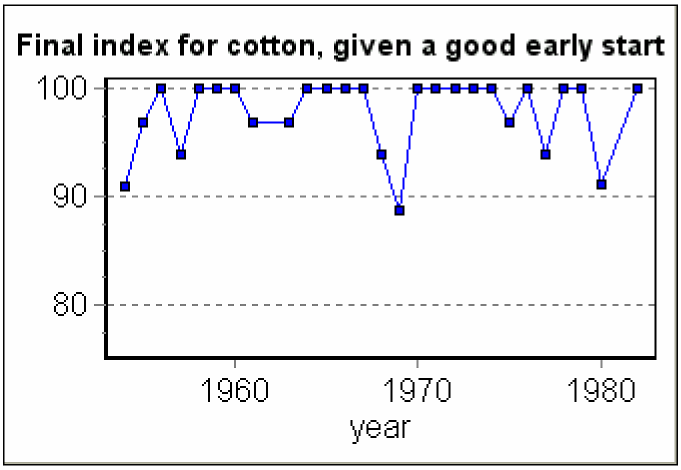
Fig. 10.4p A conditional analysis with the water satisfaction index




```
Analysis for rainfall data in X26
Total water requirement = 623mm.
Period      WR        Ra     Sur/Def  Index
Date Crop   mm        mm       mm
8     5    30.6     192.3     0.0     100
9     6    34.0     200.0    201.2    97
19    16   33.8      0.0     -8.6     96
20    17   29.2      0.0    -27.5     91
(Loss from surplus = 3.0 in 1 periods)
```

Some of the results are given in Fig. 10.4q. With the good start to the season, the minimum value of the final index is 89. There were only 2 years in which there was a water shortage. The other years in which the index fell below 100 were problems of water surplus. The results can also be compared with the unconditional analysis, given in Fig. 10.4n. With this good start, the results in Fig. 10.4q show that the index remained at 100 in 17, i.e. in 2/3 of the years, compared to only 6 years from the unconditional analysis.

Fig. 10.4q Results from a conditional analysis




```
Descriptive Statistics

Column                        X48

 No. of observations           27
 Minimum                   88.795
 Maximum                      100
 Range                     11.205
 Mean                      97.814
 Std. deviation            3.3919
 Count > 90                    26
 Count > 95                    21
 Count > 99.9                  17
 % of data > 90              96.3
 % of data > 95              77.8
 % of data > 99.9            63.0
```

10.5 Using daily data

If this index is used for a given site and crop for many years of data then the rainfall is the only element that varies from year to year in this model. Hence the variability in the index is simply a reflection of the variability in the rainfall. So one interpretation of the index is that it is a summary of the rainfall, i.e. like the annual total or any other summary value.

However, unlike the annual total it is designed to be a relevant summary for a given crop and soil capacity. It seems intuitively to be a useful summary of the rainfall and, when calculated over many years of data, can give a rough comparison of the suitability of different crops and soils.

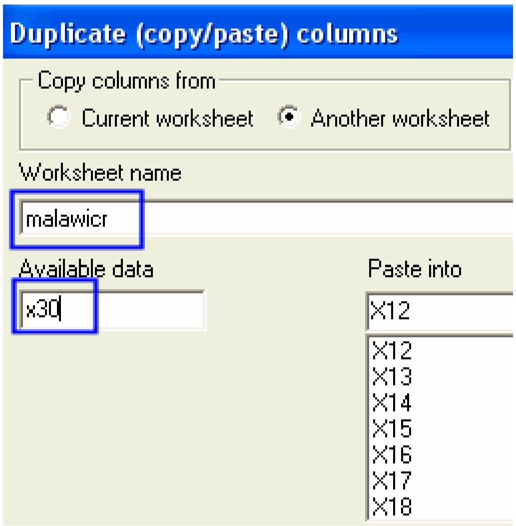
The index is thus just another **event**. It may be of value by itself or in conjunction with other events, such as the sowing date each year or the longest dry spell in the growing season.

If the index is another "event", then there is no reason for summarising the data, into decade totals, before calculating the index. Frere and Popov (1979) was written before many countries had computers, and the hand analysis would be tedious with daily data. Now that the index is no longer calculated by hand, the daily data should be used.

The example here is with the same data from Samaru, that was used in Chapters 5 to 7. The cotton crop coefficients are copied from Malawicr worksheet using **Manage ⇒Data ⇒ Duplicate (Copy columns)**, Fig. 10.5a. Alternatively they can be copied to the clipboard from Malawicr and then pasted into the samsmall worksheet.

Fig. 10.5a Preparing to use the index on a daily basis

File ⇒ Open and open **samsmall.wor**
Manage ⇒ Data ⇒ Clear(Remove) x12-x60
Manage ⇒ Data ⇒ Duplicate(Copy columns)

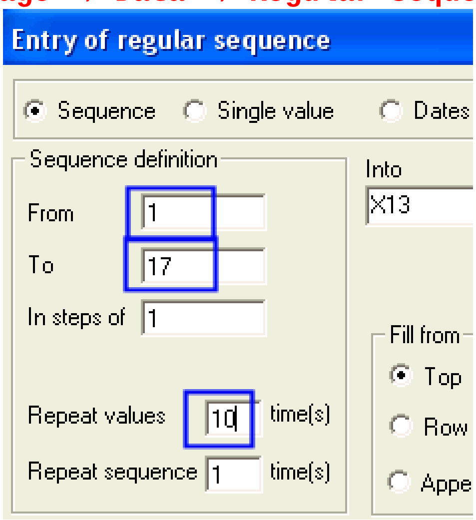


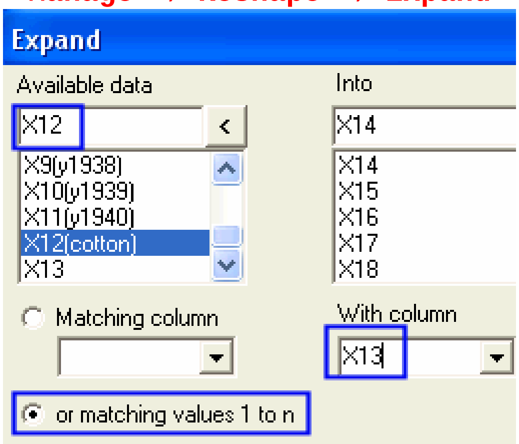
	X10	X11	X12	X13	X14
	y1939	y1940	cotton		
1	0	0	0.4	1	0.4
2	0	0	0.4	1	0.4
3	0	0	0.4	1	0.4
4	0	0	0.5	1	0.4
5	0	0	0.6	1	0.4
6	0	0	0.7	1	0.4
7	0	0	0.8	1	0.4
8	0	0	0.9	1	0.4
9	0	0	1.05	1	0.4
10	0	0	1.05	1	0.4
11	0	0	1.05	2	0.4
12	0	0	1.05	2	0.4
13	0	0	1.05	2	0.4
14	0	0	0.85	2	0.4
15	0	0	0.85	2	0.4
16	0	0	0.75	2	0.4
17	0	0	0.65	2	0.4
18	0	0		2	0.4

Part of the Samsmall worksheet is shown in Fig. 10.5a. One minor complication, that the crop values are on a 10.day basis and daily values are needed. This is a common general problem, for example evaporation data may be on a decadal or monthly basis and there are daily rainfall records. Here, for cotton, there are 17 values, corresponding to a 170-day crop.

Fig. 10.5b Expand the cotton coefficient

Manage ⇒ Data ⇒ Regular Sequence **Manage ⇒ Reshape ⇒ Expand**





One approach is in Fig. 10.5b. Frst enter a regular sequence as shown and then use the **Manage ⇒ Reshape ⇒ Expand** dialogue. The resulting columns are also shown in Fig. 10.5a.

Evaporation data are also needed. For the illustration here a constant column with 5mm per day for the 366 days is entered, using the **Manage ⇒ Data ⇒ Regular Sequence** dialogue again.

Then use the **Climatic ⇒ Crop ⇒ Water Satisfaction Index**, as shown in Fig. 10.5c. A soil depth of 60cm was used and the same definition of the start of the season adopted as in Chapter 6 with 1st May as the earliest planting date

Fig. 10.5c Crop index for daily data

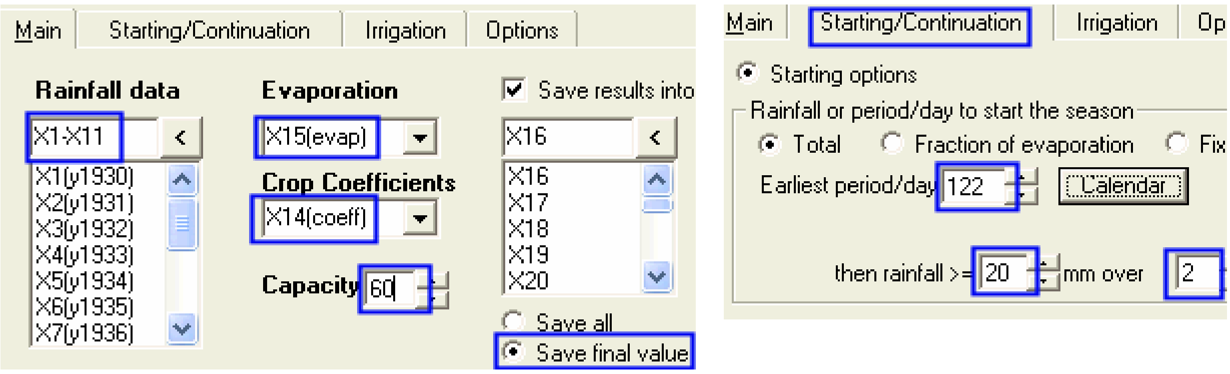


Fig. 10.5d Final index for daily data from Samaru for 3 different soil depths

X15	X16	X17	X18	X19	X20
evap	Ind60		Year	Ind100	Ind150
5	90	1	1930	96	100
5	90	2	1931	97	100
5	95	3	1932	100	100
5	77	4	1933	83	91
5	98	5	1934	100	100
5	88	6	1935	94	100
5	99	7	1936	100	100
5	98	8	1937	100	100
5	88	9	1938	94	100
5	100	10	1939	100	100
5	95	11	1940	99	99
5					
5					

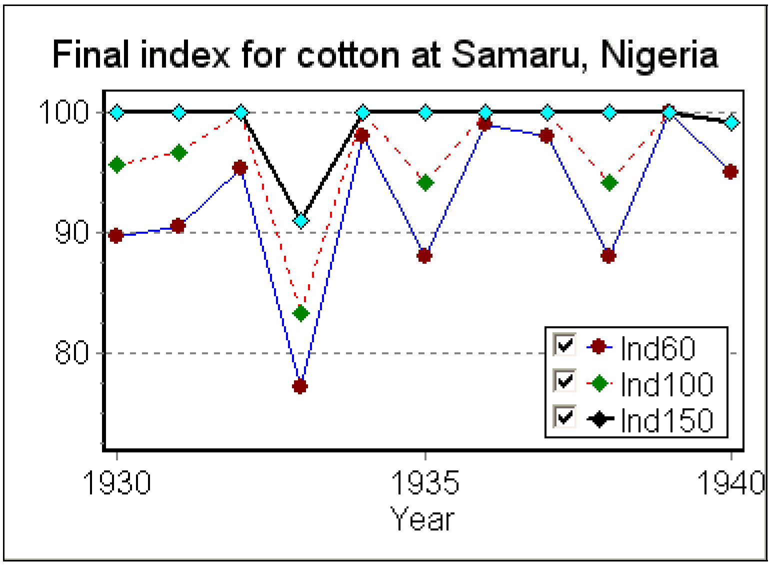


Fig. 10.5d shows the results for 3 different soil depth, 60, 100 and 150cms. the results are seen to be sensitive to the depth. For example, in 1930 the final index is 90, 96 and 100 at the three depths. With the deep soil the index is 100 in all but one year. The exceptional year was 1933, when planting was very late.

10.6 Conclusions

This chapter has combined a discussion of the FAO index, with a review of some statistical ideas, most of which were introduced in Chapter 7. The index is primarily a way of generating further events, and seems of considerable value. While remaining simple, the indices provide a summary of the rainfall data that can be interpreted for the growing season, in relation to one or more cropping strategies.

This index is too simple to be viewed as a crop model in expecting it to closely mimic crop yields, though it might be a useful variable (perhaps in conjunction with others) in a regression study relating yield to climatic data. The use of regression to study crop-weather relationships is in Chapter 11.

One reason for considering how to process the results from using this index with many years of data is because the methods of analysis are identical if a more complex crop model is used. With more complex models, it remains important to have facilities to look at the climatic and other inputs to the model. There also still needs to be ways of running the model in a flexible way, using climatic data from many years.

Chapter 11 – Further Topics

11.1 Introduction

This chapter examines further topics that are important in the analysis of climatic data. Section 11.2 considers the estimation of risks and percentage points. This covers both the calculation of the estimates and the evaluation of confidence intervals.

The use of the gamma distribution in the modelling of rainfall data is in Section 11.3. This distribution is used here for modelling 10 day totals. In Chapter 12 it is used for weekly totals and in Chapter 13 for daily values. In both these sections the estimation of the 20% point is considered. With annual data the 20% point corresponds to a 5-year return period. Sometimes estimates of higher return periods are required. This is usually to evaluate extreme events and the analysis of extremes is described in Section 11.4.

The final two sections of this chapter deal with the important subjects of correlation and regression. In Section 11.5 looks at the use and also the misuse of correlations in climatology. Section 11.6 gives a brief review of how regression methods can be used to study crop-weather relationships.

11.2 Estimating risks and percentage points

11.2.1 Overview

It is often important to estimate the proportion of occasions on which a certain event happens. With climatic records, one example would be the risk of frost. There are a variety of ways in which this risk might be estimated. If the question refers to a particular day of the year and the minimum temperature has been modelled with a normal distribution, then the estimate is given by the probability that this normal distribution is less than zero. An alternative is to look at the data and record, for each year, whether the temperature on that day was less than zero. This reduces the data for that day to a series of '0's and '1's, where we might code '0' for temperatures less than zero and '1' for those greater than zero. The proportion of zeros is then taken as the chance of frost.

A more complicated example is to estimate the chance that there is ever frost in the year. This is important for crops such as coffee, which cannot withstand any frost. It is difficult if we have a model of daily temperatures with, for example, a normal distribution and wish to use the model. We consider one way this can be done in the section on extremes, later in this chapter. A simple alternative, when there are many years of data, is to scan the data and record '0' for any year in which the temperature ever drops to zero and '1' for the other years. If, from 40 years, there are four years when the temperature dropped to zero, the estimate is 4/40=0.1, i.e. 10% of the years.

This calculation has, of course, 'thrown away' a lot of information. Although there is a temperature reading every day of the year, these are all replaced by either a '0' or a '1' for the year. Hence, the resulting estimate (10% in this case) is not very precise. However, it is useful to have such a simple method.

Two situations can be distinguished. The first is where the original data are recorded as one of two alternatives. For example, in a survey there might be questions on whether or not farmers:

- i) were tenants;
- ii) had crop damage;
- iii) applied fertilizer in the last year;
- iv) grew cotton.

Again, simple recordings of climatic data might just note whether or not there:

- i) was rain;
- ii) or frost;
- iii) or snow;
- iv) or sufficient wind to generate power

without recording the exact amounts.

The second situation is where more detail has been recorded, such as the actual temperature. Users can then, if they wish, reduce such data to a simple '0-1' form for analysis, as described above.

In Chapter 7 there have been numerous examples of the estimates of probability (or risks). For example: What is the probability of planting by the end of April? What is the probability of more than 20mm of rain in a given 10 day period?

There have also been examples where percentages have been estimated, e.g. what is the 20% point of the planting date? For annual records, this can be considered as a return period, i.e. the 20% corresponds to the 100/20=5 years return period.

In this section we revisit the estimation of risks (probabilities) or return periods (percentage points) and also how to give standard errors, or confidence limits, for these estimates.

11.2.2 Estimating risks

The dataset **Samrain.wor** contains data from Samaru, Nigeria, for 56 years, from 1928 to 1975, for which the first stage in the analysis have been run, as described in Chapter 6. Part of the resulting dataset is in Fig. 11.2a.

Fig. 11.2a Samrain.wor with 'events' from Samaru

`File ⇒ Open From Library ⇒ samrain.wor`

Current Worksheet - SAMRAIN.WOR

Columns | Constants | Strings | Labels   Title: Rainfall Events, Samaru 1928-83

	X1*	X2*	X3*	X4*	X5*	X6*	X7*	X8*	X9*
	Year	Strt1	Strt2	Strt3	Strt4	DryM	DryJn	DryJl	End
1	1928	115	115	127	127	6	6	6	300
2	1929	126	126	126	126	28	4	5	301
3	1930	108	133	133	133	13	7	5	288
4	1931	118	118	131	131	8	5	3	288
5	1932	115	115	133	133	8	3	13	295
6	1933	156	156	156	156	14	4	6	294
7	1934	116	116	124	124	8	3	3	291
8	1935	134	134	134	134	9	3	4	288

Here, the dates of the start of the rains in X2 and X3 are considered. X2 is the date of the start of the rains using the definition described in Chapter 6, i.e. the first occasion after 1st April with more than 20mm in 1 or 2 days. X3 uses the same definition as X2, but with the added condition that there should not be a dry spell of 10 or more days in the 30 days after planting.

Thus, X2 can be interpreted as the *first* planting date and X3 as the *successful* planting date. Now look at the differences between the two columns with **`Manage ⇒ Calculations`** and calculate X10 (Fig. 11.2b) and then **`Manage ⇒ View Data ⇒ Display`** to look at X10 (Fig. 11.2c). This was one of the tasks considered in Section 7.6.

Fig. 11.2b Calculate planting date differences

Manage ⇒ Calculations

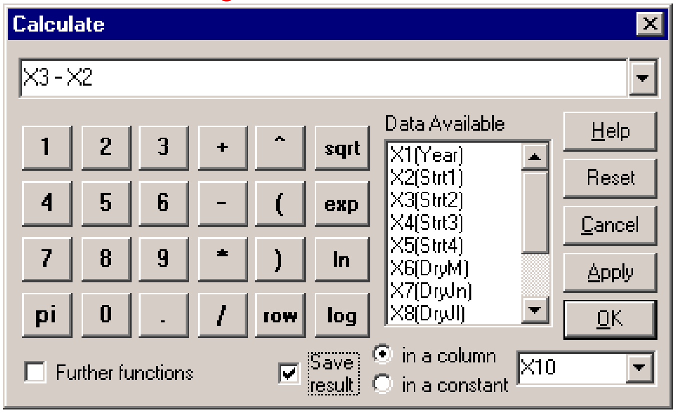


Fig. 11.2c Display the differences

Manage ⇒ Data ⇒ Display

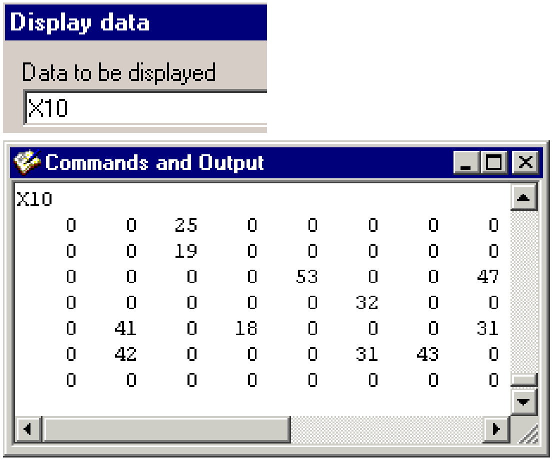


Fig. 11.2d shows there are 11 observations where the difference is non-zero, i.e. the first planting date was not successful. Instead of manually calculating the non-zero cases, use **Manage ⇒ Reshape ⇒ Select**, (Fig. 11.2d) followed by **Statistics ⇒ Summary ⇒ Describe** and choose **X11,** (Fig. 11.2e).

Fig. 11.2d Select non-zero cases

Manage ⇒ Reshape ⇒ Select

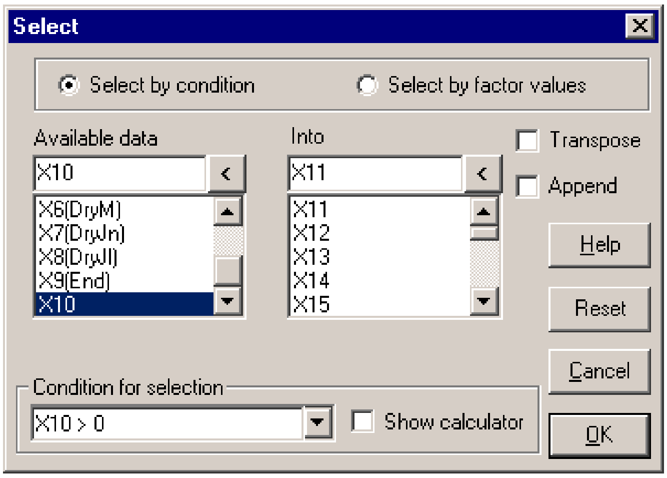
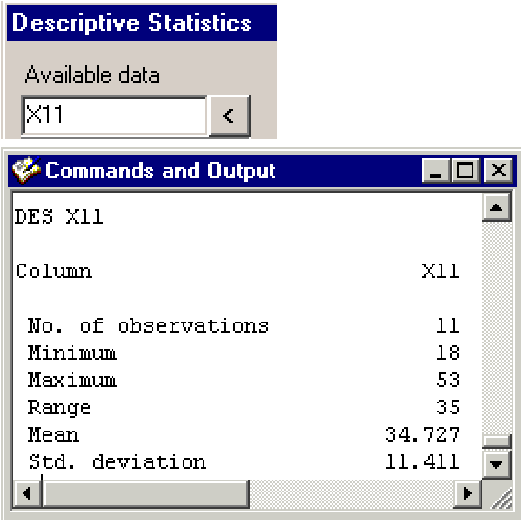


Fig. 11.2e Results

Stats ⇒ Summary ⇒ Describe



There are many ways that the analysis can proceed. For example Fig. 11.2e shows that the first planting date was not successful in 11 out of the 56 years and in those years, the extra time before a successful planting had a mean of 35 days with a standard deviation of 11 days.

Here, of particular interest is that the first planting failed in 11 out of the 56 years, i.e. the risk of failure is 11/56=0.20 or 20%. This sort of information can help in comparing alternative planting strategies. For example, X4 and X5 give the same definitions, but with 1st May as the earliest planting date. The risk can be evaluated similarly as 4/56=0.07 or 7%.

This does not necessarily mean that the 1st May strategy is better. It is safer, but has lost the chance of planting in April and hence of having a longer season. Further analysis of 'Strt2' (X3) in Fig. 11.2f shows that in 18 years, i.e. 32% of the years, there was a successful planting before the end of April (day number 121).

Fig. 11.2f Proportions for X3

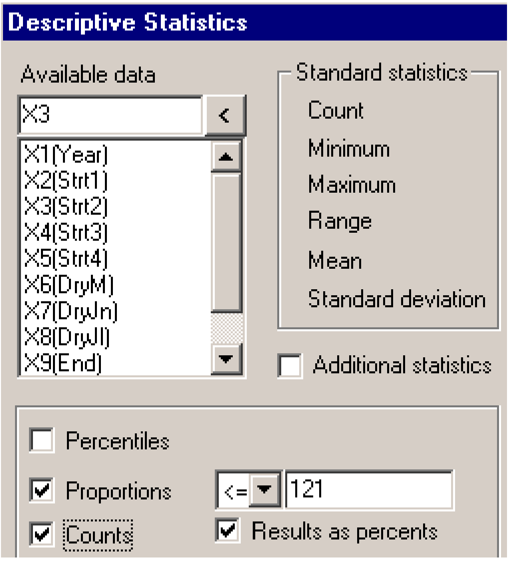
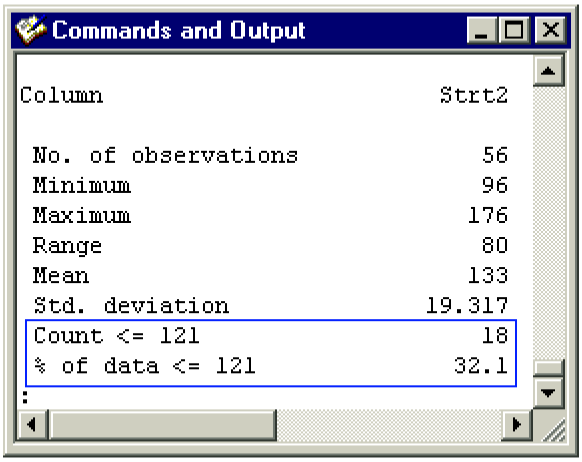


Fig. 11.2g Results

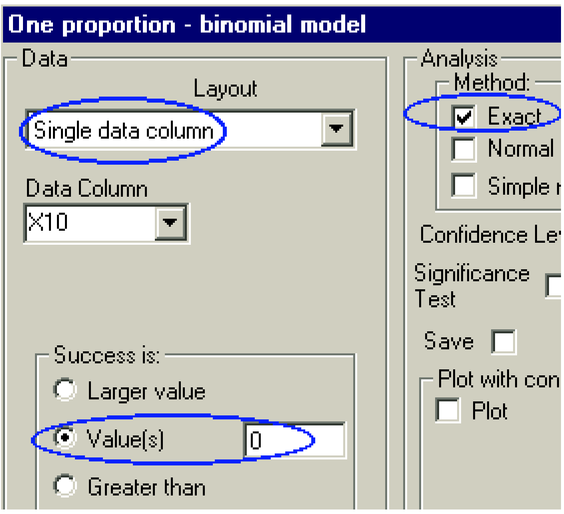


Once risks have been evaluated, e.g. 11/56=20%, it is important to calculate the precision of the estimates. Use the **Statistics ⇒ Simple Models ⇒ Proportion, One Sample** as shown in Fig. 11.2h. This dialogue can be used in two ways and both are shown. on the first is to use the data in X10 directly, defining the value 0 (i.e. no difference between Strt1 and Strt2) as a success. Alternatively, if the summary values are already available, here 45 out of the 56 years then this can be used as shown in Fig. 11.2h. Similarly, confidence limits for the risk, which is 11 out of the 56 years are given by subtracting 1 from the proportions, or by rerunning the dialogue with 11 out of 56, rather than 45 out of 56.

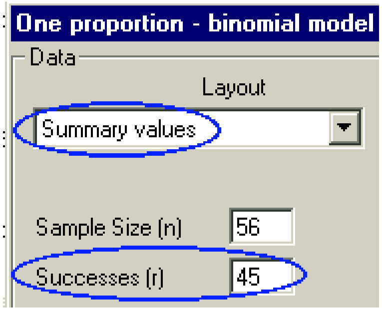
The results are shown in Fig. 11.2i.

Fig. 11.2g Proportion, One sample dialogues

Statistics ⇒ Simple Models ⇒ Proportion, One Sample



Successes



Failures

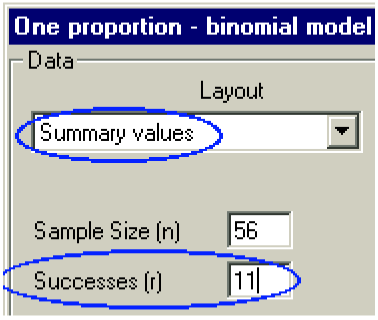
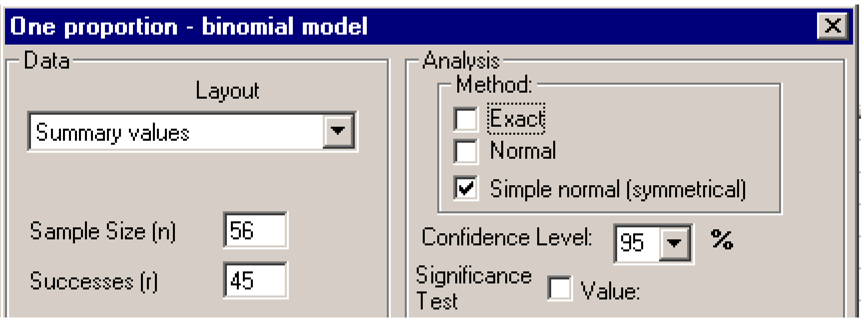


Fig. 11.2i 'Success' and 'Failure' of planting

```
Success
Binomial model, single sample
 Sample size             56
 Successes               45
 Proportion              0.804

Exact results:
95% confidence interval for prop.   0.676    to    0.898

Failure
Binomial model, single sample
 Sample size             56
 Successes               11
 Proportion              0.196

Exact results:
95% confidence interval for prop.   0.102    to    0.324
```

The dialogue in Fig. 11.2h, used the default option and calculated exact confidence limits for the proportion. Fig. 11.2j shows same dialogue, but with the **Simple normal (approximate)** confidence limits. This is what was described in some textbooks and can be calculated 'by hand'.

Fig. 11.2j Normal approximation to confidence interval




```
One proportion - binomial model

BINomial;stats 56 45;SUCcess 0;SIMple
Binomial model, single sample

 Sample size           56
 Successes             45
 Proportion            0.804

Approx s.e. of proportion = 0.053

Simple normal approximation:
95% confidence interval for prop.   0.700    to    0.908
```

The calculations in Fig. 11.2j uses the fact that the standard error of a binomial probability is √[(p(1-p)/n)], here √ [(0.20*0.80/56)]=0.053 or 5%. Hence, with the normal approximation to the binomial distribution, the approximate 95% confidence limits are

0.20 ± 1.96 * 0.05, i.e. 0.1 to 0.3

Thus the true risk, which is estimated as 20%, is likely to be between 10% and 30%. This uncertainty in the estimates of the risks has to be born in mind when interpreting the results.

11.2.3 Estimating percentage points (return periods)

Chapter 7 showed it is straightforward to estimate percentage points or return periods. Use either the **Statistics** ⇒ **Summary** ⇒ **Column Statistics** or **Describe** dialogues, Section 7.2 or **Climatic => Process** dialogue, Section 7.4. For the data in Strt2 (X3), with the 20% point, the result is day 115, i.e. 20% of years have a successful planting date by 24th April.

The analysis can be extended to give confidence limits for this estimate. With Instat, this uses a macro called **quantile.ins**. One way to run the macro is to type : **@quantile x3 20** in the Output and Command window. The results are shown in Fig. 11.2k.

Fig. 11.2k Confidence limits for a percentage point

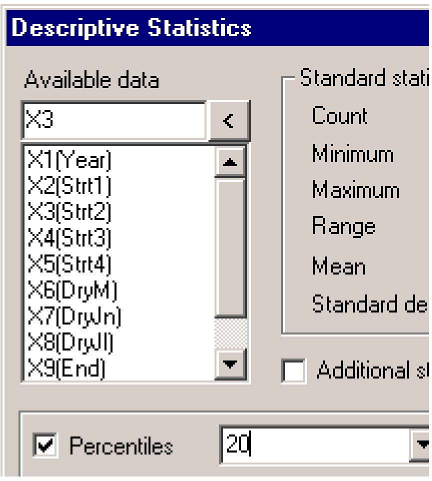
```
: @quantile x3 20

Row        Lo       Hi      %lo      %up        %      Dlo      Dhi

  1         6       16      5.1     95.7     90.6      110      118
  2         5       16      2.1     95.7     93.6      109      118
  3         6       17      5.1     97.8     92.7      110      120
  4         5       17      2.1     97.8     95.7      109      120

Estimate of 20% point
          115
```

Fig. 11.2j, shows that the estimate is day 115, i.e. 24th April. Because of the discreteness of the binomial distribution an exact 95% confidence interval is not possible and the macro has given 4 possibilities. From the column in Fig. 11.2j called %, the last line is seen to be the closest at 95.4% and this corresponds to the 5th and 16th observations in the ordered data, with corresponding observations shown to be days 111 and 122. Hence the approximate 95% confidence intervalis from 20th April to 1st May.

For those who wish to understand how these calculations work, the method is given from first principles in Fig. 11.2k and described below.

Fig. 11.2l The 20% day for a successful start of the rains is day 115

Statistics ⇒ Summary ⇒ Describe




```
DES X3;PER 20

Column                          Strt2

 No. of observations               56
 Minimum                           96
 Maximum                          176
 Range                             80
 Mean                             133
 Std. deviation                19.317
 20th percentile                  115
:
```

Fig. 11.2m Sorting the data

Manage ⇒ Manipulate ⇒ Sort

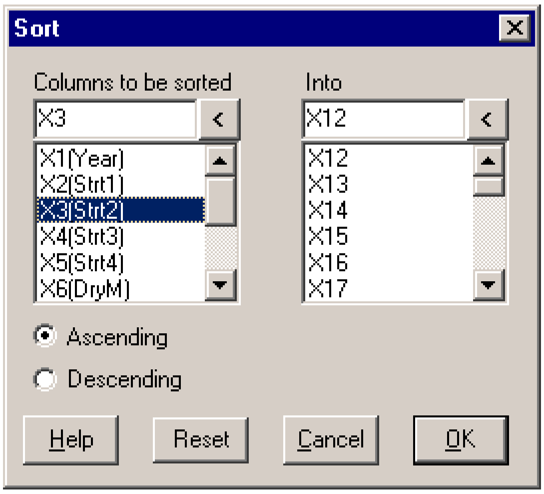


Fig. 11.2n Values for the probabilities

Manage ⇒ Data ⇒ Regular Sequence

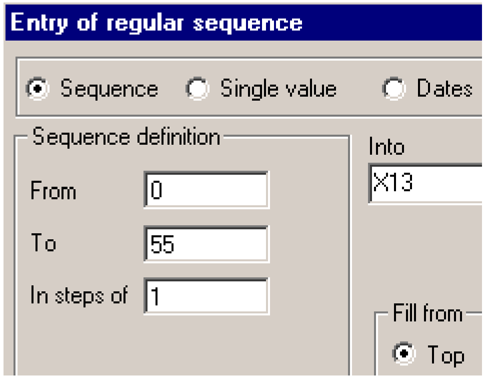


Fig. 11.2o Probabilities

Stats ⇒ Prob. Distribution

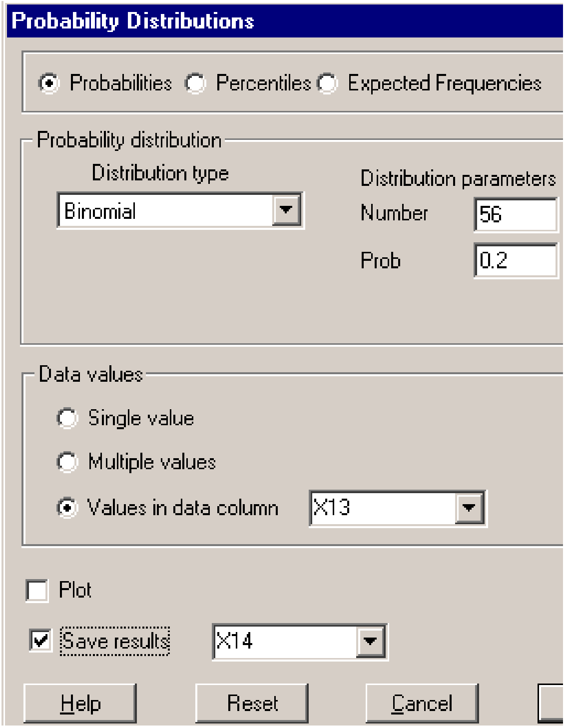


Fig. 11.2p Results

(after X15=cusum(x14))

	X12	X13	X14	X15
1	96	0	0.000	0.000
2	100	1	0.000	0.000
3	107	2	0.000	0.000
4	108	3	0.002	0.002
5	109	4	0.005	0.007
6	109	5	0.014	0.021
7	110	6	0.030	0.051
8	111	7	0.053	0.104
9	113	8	0.081	0.185
10	114	9	0.108	0.293
11	115	10	0.127	0.420
12	115	11	0.133	0.553
13	115	12	0.125	0.678
14	116	13	0.105	0.783
15	118	14	0.081	0.864
16	118	15	0.057	0.921
17	118	16	0.036	0.957
18	120	17	0.021	0.978
19	122	18	0.012	0.990

The results in Fig. 11.2p show that

a) Prob(less than or equal to 5 successes) = 0.021

b) Prob(less than or equal to 17 successes) = 0.978

We are looking for values that are close to 2.5% and 97.5% and these are close. The approximate 95% limits are given by the corresponding observations, which can be found from the sorted data in X12(5)= 109 and X12(17)= 120. Thus, the 95% interval is given by days 109 to 120 or 18th April to 29th April.

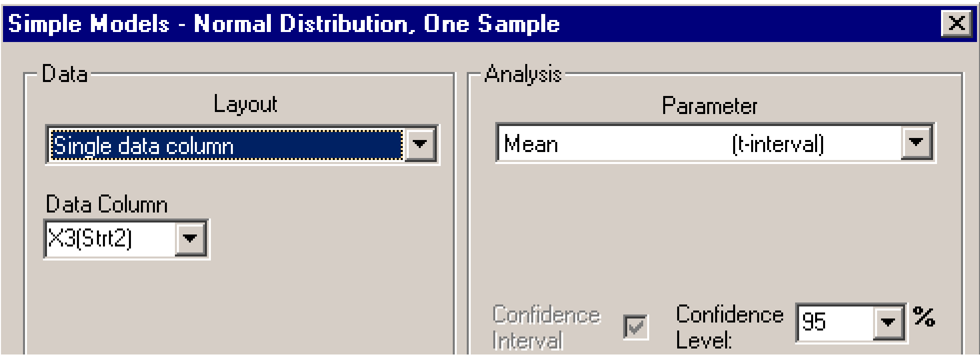
11.2.4 Percentages using a normal distribution

The analysis in the previous section was of the type that is called 'non-parametric' or 'distribution-free'. This is because the analysis used just the observed observations and assumed no particular distribution for the starting dates. The analysis above does make the assumptions that the years are independent of each other and are from the same distribution.

One alternative is to fit and use a normal model to the data on starting dates given in X3. The **Climatic => Examine** dialogue, introduced in Section 7.3 includes the option to give normal probability plots. Here the analysis is taken with the **Statistics** ⇒ **Simple Models** ⇒ **Normal, One Sample** dialogue for **X3** (Fig. 11.2l).

Fig. 11.2q Normal, one sample dialogue (Tinterval)

Statistics ⇒ Simple Models ⇒ Normal, One Sample




```
Simple Models - Normal Distribution, One Sample

TINt 'Strt2'
Normal model, one sample
Column                     Strt2
 Sample size                56
 Minimum                    96
 Maximum                    176
 Range                      80
 Mean                       133
 Std. deviation             19.317

Standard error of mean    = 2.5813    with 55 d.f.

95% confidence interval for mean    127.83     to     138.17
:
```

The results in Fig. 11.2q show that the mean date of the start of the rains is estimated as day 133, 12th May, with a standard error of about 2.6 days. With the assumption of a normal model, the mean equals the median. The median is the 50% point or 2 year return period.

Next, the 20% and 80% points (i.e. the 5 year return periods) are estimated. This uses an option in the same dialogue, Fig. 11.2r. The additional results, with a plot, are in Fig. 11.2s.

Fig. 11.2q Estimating percentage points

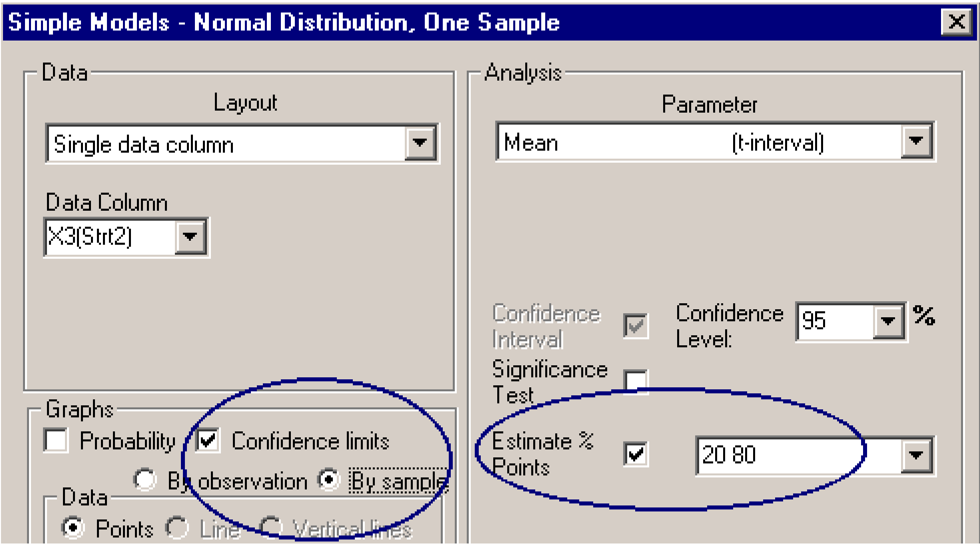
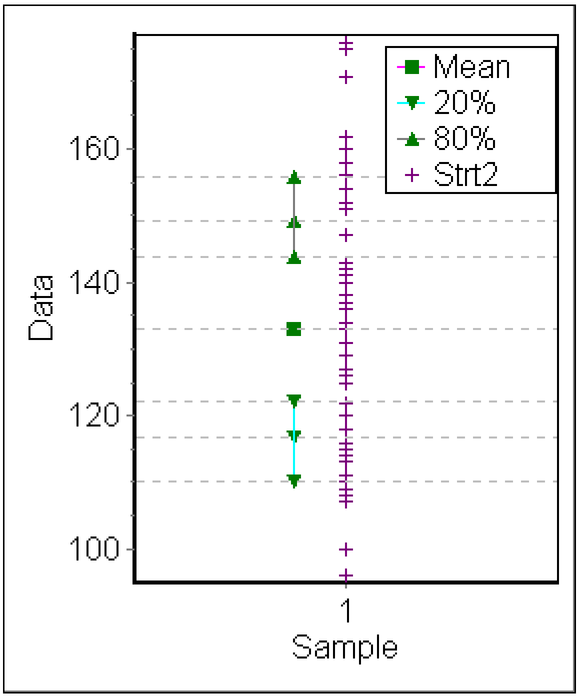


Fig. 11.2r The estimated 20% and 80% points with confidence limits

```
 Range                       80
 Mean                        133
 Std. deviation              19.317

Standard error of mean     = 2.5813    with 5

95% confidence interval for mean    127.83

                Estimated          Approximate 95%
%               value   s.e.      confidence limits
20              116.7 3.011        110.1     122.1
80              149.3 3.011        143.9     155.9
```



The results show the estimate of the 20% point is day 117 (i.e. 26th April) with 95% confidence interval from day 110 (19th April) to day 122 (i.e. 1st May). The results from the distribution-free and normal distribution approaches are therefore seen to be similar.

> These results can also be calculated from first principles. The Statistics ⇒ Probability. Distributiondialogue is used in Fig. 11.2s to estimate the 20% and 80% points, using the normal distribution with the same mean and variance as the data. The results, in Fig. 11.2t are the same as in Fig. 11.2r

Fig. 11.2s Percentiles from the normal distribution

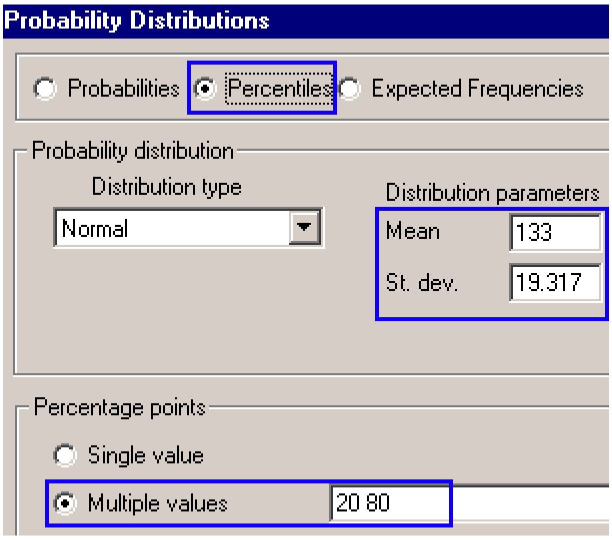


Fig. 11.2t Results

Probability Distributions

```
ENTer X40;DATa 20 80
: PER X40 X16;NORmal 133 19.317
Normal dist.  Mean 133 and s.d. 19.32
 Percentage          Value
       20%          116.74
       80%          149.26
```

The formula for this 20% point is

133 - 0.8416 * 19.317 = 116.74

because the 20% point of the standard normal distribution is -0.8416.

The standard error of the estimate is more complicated than when estimating the mean. It is given approximately by:

$$\sqrt{[s^2/n + (-0.8416)^2 * (s^2/(2(n-1)))]}$$

i.e. : ? SQR(19.317^2/52+(-0.8416*19.317)^2/(2*55)) = 3.011 or about 3 days

and the 95% confidence interval is approximately twice the standard error, i.e.

117 - 2*3 days (day 111 or 20th April)
117 + 2*3 days (day 123 or 2nd May)

Similarly the 95% confidence limits for the 80% point is roughly day 149 – 6 to 149 + 6, i.e. 22nd May to 3rd June.

The results are usually be close to, but not exactly the same as those from the Normal, One Sample dialogue, Fig. 11.2r, because Instat uses a better formula, based on the non-central t-distribution.

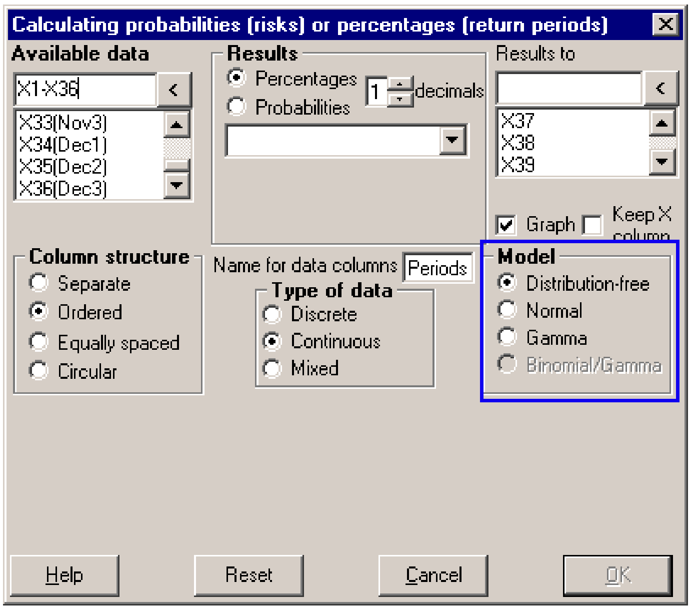
11.3 The gamma distribution

The gamma distribution is commonly used to model rainfall amounts. The **Climatic => Process** dialogue, used in Section 7.4, showed that 4 models were possible, see Fig. 11.3a. In Chapter 7 the distribution free analysis was used. Section 11.2, showed how the normal model is sometimes applied. The other two options involve the gamma model, which is described here.

Fig. 11.3a Climatic ⇒ Process dialogue

Fig. 11.3b Dekad data

`File ⇒ Open From Library ⇒ Samdis56.wor`



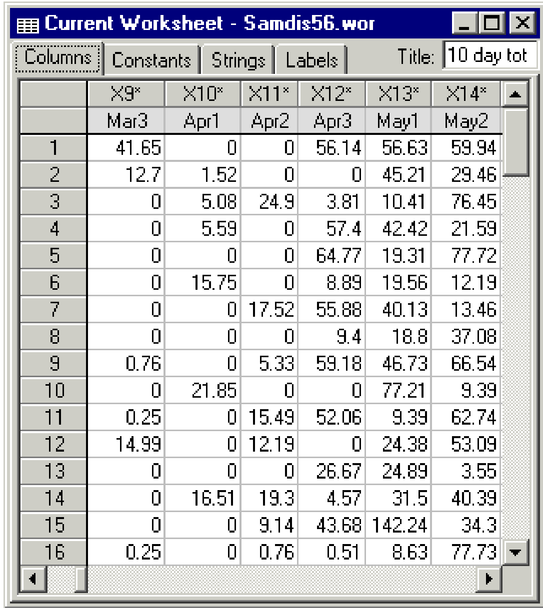


Fig. 11.3b shows the data that will be used in the subsequent examples. They consist of the 36 rainfall decade totals for the 56 years of the record at Samaru, Nigeria.

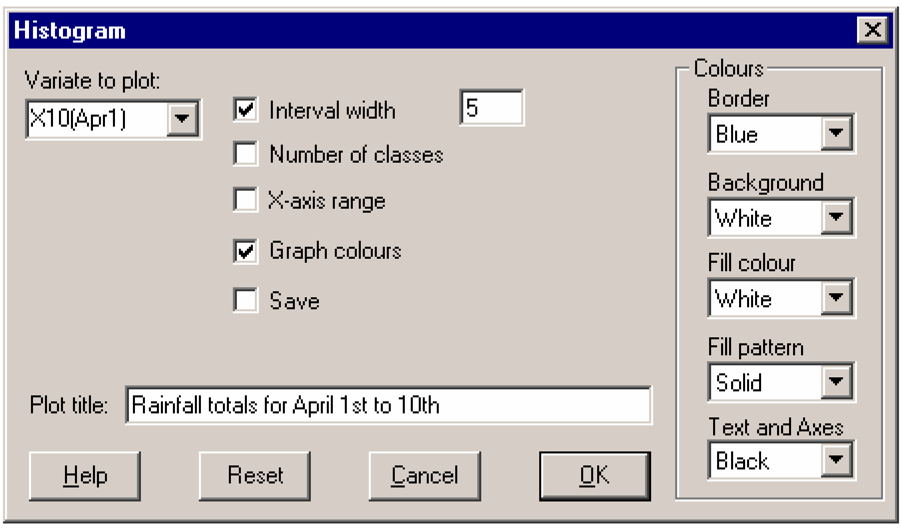
The normal distribution is often used to model rainfall totals. The normal distribution is symmetric, so it is not suitable for modelling data that has a skew distribution. This is commonly the case, as is indicated by the first histogram in Fig. 11.3c. Skew data may be transformed in an effort to fit a normal distribution, but this is not always satisfactory, partly because a different transformation may be required for successive periods.

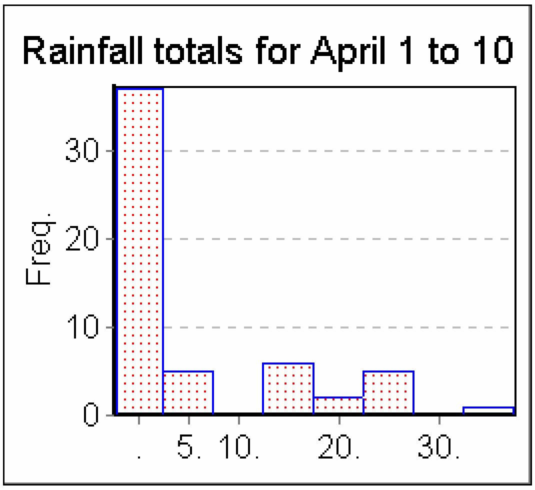
In the pre-computer age it was sometimes useful to transform because the normal distribution is easier to handle. Now, with computers, other distributions are equally tractable. The gamma distribution is popular, partly because its flexibility (see Fig. 11.3f) means it can be used to model rainfall amounts on time scales ranging from less than a day to a year.

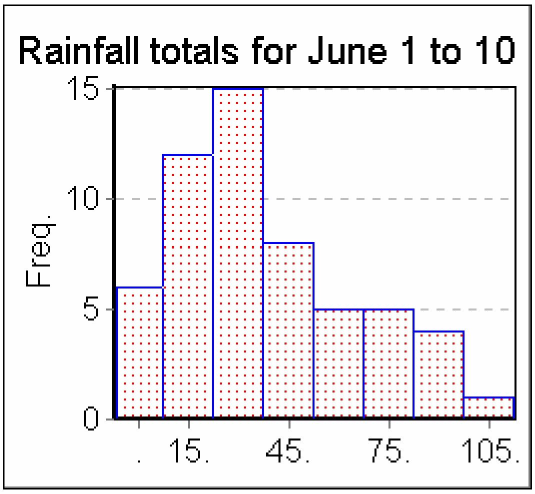
Use **`Graphics => Histogram`** dialogue (for example Fig. 11.3c for April 1st to 10th) to produce the graphs in Fig. 11.3d

Fig. 11.3c Histogram dialogue

`Graphics ⇒ Histogram`







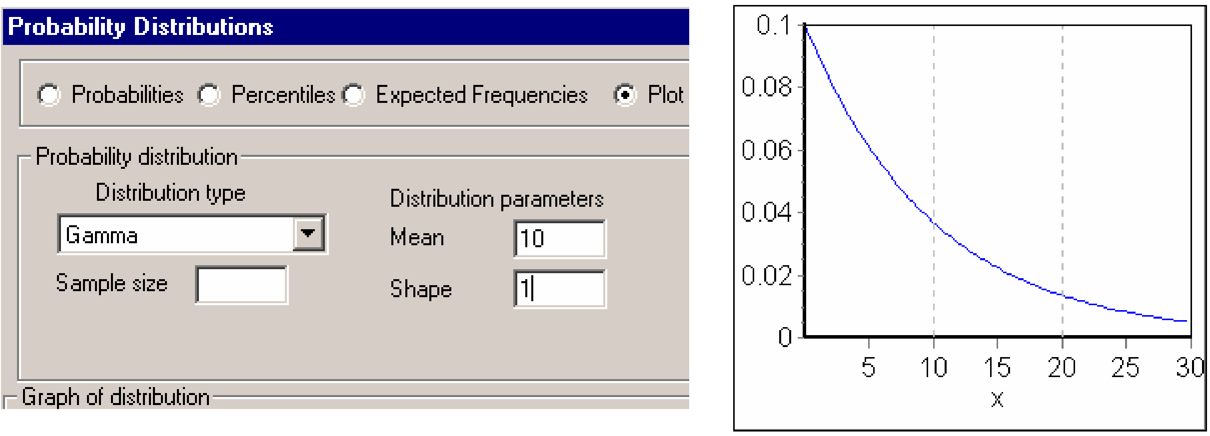
Section 11.3.2 introduces the gamma distribution and shows how probability calculations can be performed. Section 11.3.3 then demonstrates how it can be used as a model for the data in Fig. 11.3d. The June data are used first because they do not have the added complication of zeros. The problem of zeros is tackled in Section 11.3.4, when the April data are analysed. The period (1-10) April was dry in 32 of the 56 years.

11.3.2 Properties of the gamma distribution

The gamma distribution has 2 parameters, μ and k. The parameter μ is the mean of the distribution and the value of k dictates the shape. The **Statistics => Probability Distribution** dialogue may be used to look at a range of distributions, including the gamma. It is shown in Fig. 11.3e for μ=10 and k=1.

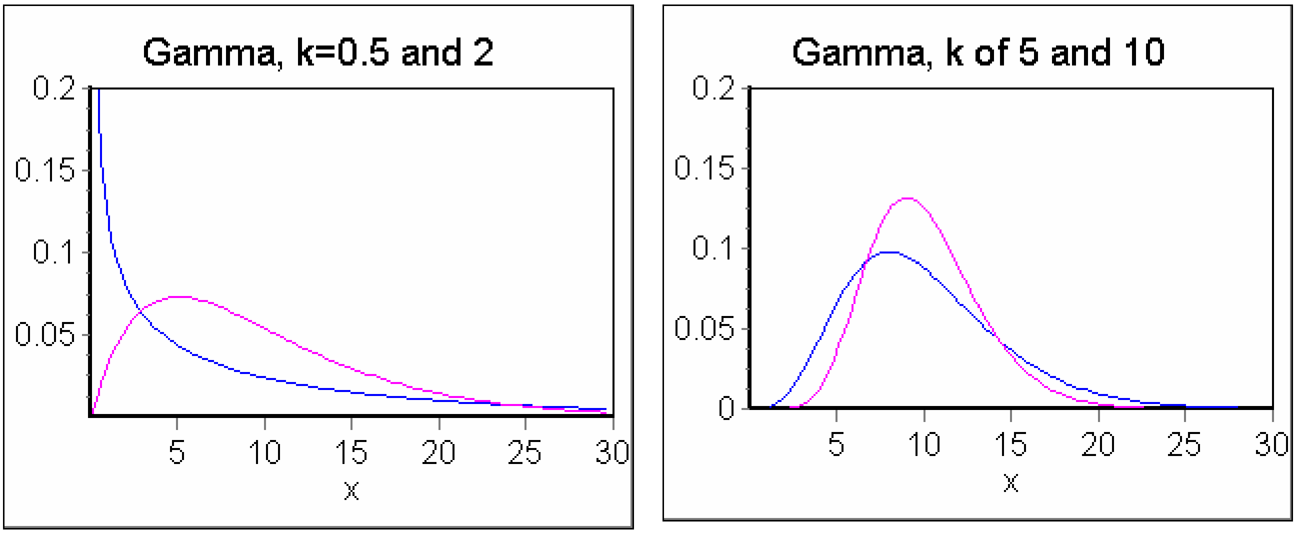
Fig. 11.3e Plot of gamma probability distribution for μ=10 and k=1

Statistics ⇒ Probability Distribution



More examples are shown in Fig. 11.3f. Low values of the parameter k are the shape that corresponds to daily data, while higher values give the shape for monthly or even annual totals. When k=10, the shape is similar to a normal distribution.

Fig. 11.3f Plots of gamma distributions for μ=10 and k=0.5, 2, 5 and 10



For reference the gamma density function is defined for x>0 as

$$P(x) = \left(\frac{k}{\mu}\right)^k x^{k-1} \frac{\exp\left(\frac{kx}{\mu}\right)}{\Gamma(k)} \qquad \mu > 0,\ k > 0$$

Thus the gamma distribution may be used to model variables that are continuous and positive. The gamma function is $\Gamma(k) = (k\text{-}1)!$ (i.e. '(k-1) factorial'), when k is an integer.

As with the normal distribution, probabilities and percentage points cannot be calculated algebraically. Probabilities and percentage points for gamma distributions are available in statistics packages, including Instat, where the **Statistics => Probability Distribution** dialogue is used.

The mean of the gamma distribution is μ and the variance is μ^2/k. Thus the coefficient of variation is $100/\sqrt{(k)}$. The coefficient of variation is therefore greater than 100% when k is less than 1, which is often the case when modelling daily data. It is sometimes thought that there is something wrong with data if the c.v. is greater than 100%, but this is not necessarily the case. The c.v. is not a useful summary for skew data and the gamma distribution is very skew when k is less than 1. The skewness of the gamma distribution is $2/\sqrt{(k)}$ and is always positive. For large values of k, the skewness is low and the gamma distribution is similar to the normal distribution.

Two special cases of the gamma distribution are important in their own right. When k is 1, the distribution has a particularly simple form, namely

$$p(x) = (1/\mu)^* e^{(-x/\mu)} \qquad \mu > 0\ ,\ x > 0$$

This is the exponential distribution and the example plotted in Fig. 11.3e. The mean is μ and the variance is μ^2, i.e. the coefficient of variation is 100%.

The second special case is when k is either a whole number or a half integer and the mean is $2k$. This is the chi-square distribution with $2k$ degrees of freedom.

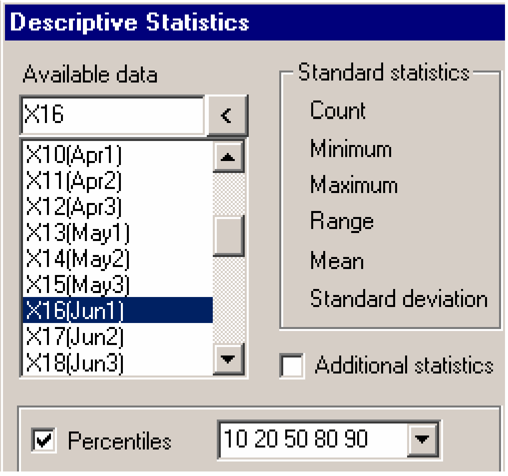
11.3.3 Fitting a gamma model to 10 day rainfall totals

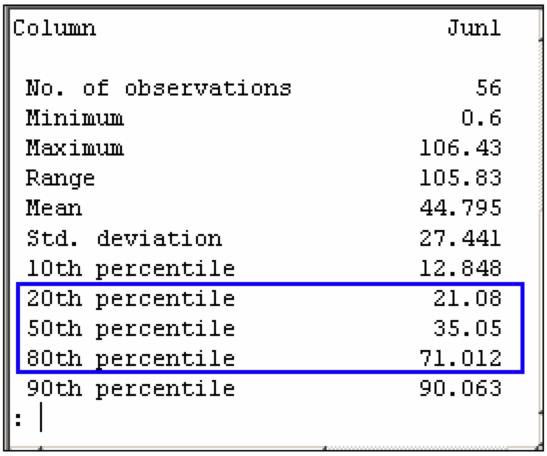
The example uses the 56 rainfall totals for the period 1-10 June. The histogram in Fig. 11.3d shows that the data are skew, though not markedly so. To provide a basis for comparison, Fig. 11.3g shows some results from the distribution-free approach that has been used in earlier chapters. The results include different percentage points, for example

20% = 21.1mm 50% = 35.5mm 80% = 71.0mm

Fig. 11.3g Summary of rainfall totals for 1-10 June

`Statistics ⇒ Summary ⇒ Describe`



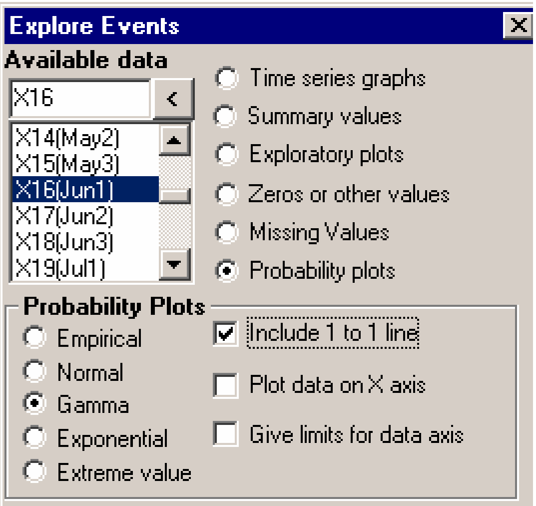


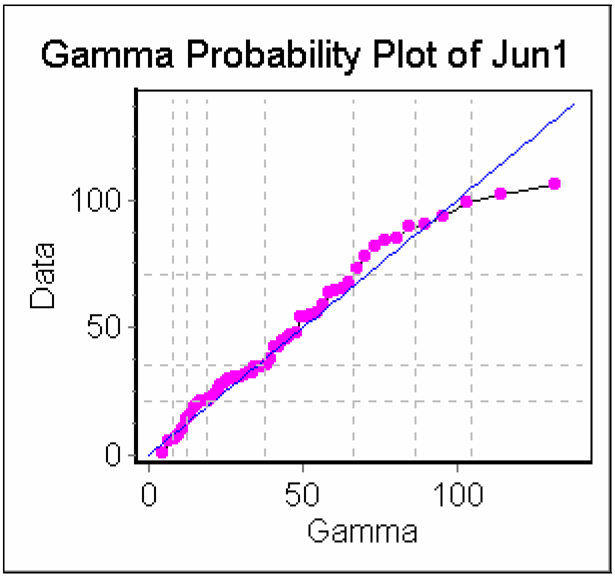
Column	Jun1
No. of observations	56
Minimum	0.6
Maximum	106.43
Range	105.83
Mean	44.795
Std. deviation	27.441
10th percentile	12.848
20th percentile	21.08
50th percentile	35.05
80th percentile	71.012
90th percentile	90.063

These observed values will be compared with the estimates from the gamma model. The first step in this modelling process is to check whether the gamma model is suitable. Either the **`Climatic ⇒ Examine`**, Fig. 11.3h, or the **`Graphics ⇒ Probability plot`** dialogues may be used to give a probability plot for the gamma distribution as shown in. If the model is appropriate, the graph should be roughly linear.

Fig. 11.3h Probability plot of gamma

`Climatic ⇒ Examine`



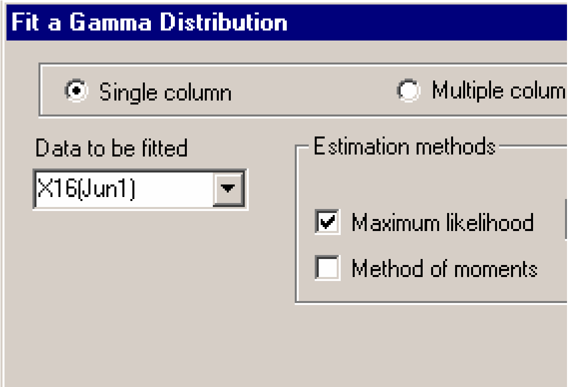


The result in Fig. 11.3h looks reasonably linear, except possibly for large values. To show the modelling process we now fit the gamma model, but return to the possible problem later.

The process of fitting involves estimating the unknown parameters, μ and k of the gamma distribution. This is shown in Fig. 11.3i.

Fig. 11.3i Fit a gamma model to rainfall totals for 1-10 June

`Statistics ⇒ Simple Models ⇒ Gamma`



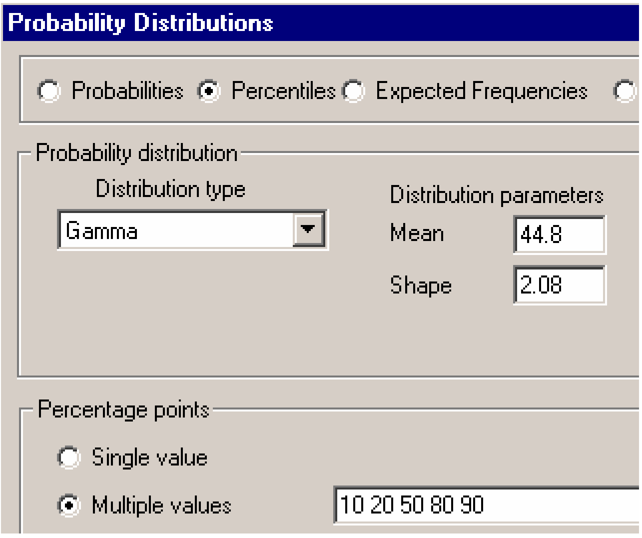

```
Fit a Gamma Distribution

GAMma X16;MAX

Column                  Jun1
No. of observations     56
No. greater than 0      56
Percentage 0            0.0%

Method of Maximum Likelihood
 muhat = 44.7954    with s.e. 4.14842
  khat = 2.08215    with s.e. 0.36635
:
```

The results from Fig. 11.3i show that the estimate of μ, $\hat{\mu}$ = 44.8mm and $\hat{k}$ =2.08.

The **Statistics** ⇒ **Probability Distributions** dialogue is used to give probability and percentage points for a gamma model as shown in Fig. 11.3j.

Fig. 11.3j Percentage points for fitted gamma model

`Statistics ⇒ Probability Distributions`




```
Probability Distributions

Gamma dist. Mean 44.8 & k of 2.08
 Percentage              Value
        10%              12.355
        20%              18.921
        50%              37.861
        80%              66.762
        90%              86.333
:
```

The results in Fig. 11.3j are seen to be very close to those in Fig. 11.3g observed from the data.

	Observed from data	Gamma model
10%	12.8 mm	12.0 mm
20%	21.1 mm	18.6 mm
50%	35.1 mm	37.7 mm
80%	71.0 mm	67.0 mm
90%	90.1 mm	86.9 mm

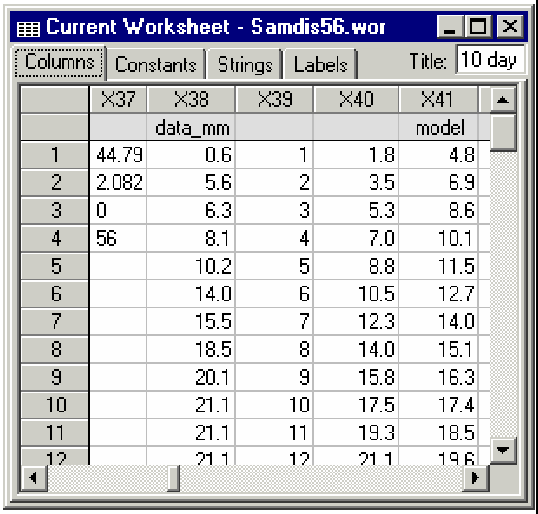
The process above shows it is now easy to fit and use the gamma distribution. The extra complexity of such a model, compared with the normal distribution, is no longer important because computer programs are available.

The use of probability plots to check the data, and the method of fitting the gamma model, are now described in more detail.

The steps to get a gamma probability plot are shown in Fig. 11.3k. In Fig. 11.3k the parameters of the fitted gamma model are saved into x37. Then the sorted data are in x38, with the fitted values from the gamma model in x41. The resulting plot is essentially the same as in Fig. 11.3h.

Fig. 11.3k Gamma probability plots from first principles

```
warn off
gamma x16; max x37
sort x16 x38
enter x39; data (1]56)
x40 = 100*x39/57

percent x40 into x41;gamma x37
name x38 'data_mm':name x41 'model
enter s1; data x
line x38 l l 64: symbol x38 l l 64
plot x38 s1 x41;nol; tit"Probability plot"
```


Current Worksheet - Samdis56.wor

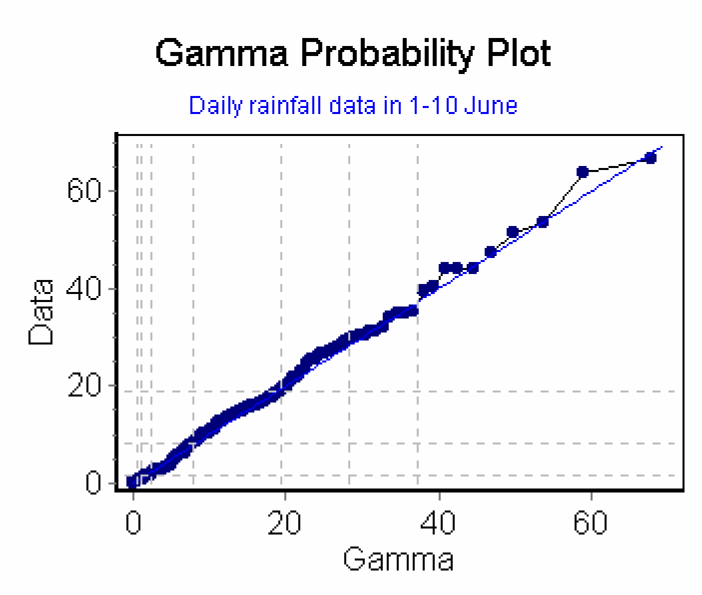
Columns | Constants | Strings | Labels | Title: 10 day

	X37	X38	X39	X40	X41
		data_mm			model
1	44.79	0.6	1	1.8	4.8
2	2.082	5.6	2	3.5	6.9
3	0	6.3	3	5.3	8.6
4	56	8.1	4	7.0	10.1
5		10.2	5	8.8	11.5
6		14.0	6	10.5	12.7
7		15.5	7	12.3	14.0
8		18.5	8	14.0	15.1
9		20.1	9	15.8	16.3
10		21.1	10	17.5	17.4
11		21.1	11	19.3	18.5
12		21.1	12	21.1	19.6

The curvature in Fig. 11.3h is now considered in more detail. One obvious step is to see whether the shape is similar for other decades. This is easy to do with the **Climatic ⇒ Examine** dialogue, Fig. 11.3h, which allows multiple columns to be specified. The result is that the periods round June, have a similar curvature, but those at the end of the season do not. In a more substantial study it would be useful also to look multiple sites.

The gamma model usually fits better than this, so a further query could be whether the daily rainfalls follow a gamma model. Fig. 11.3l plots the daily values on the 209 rainy days in the June 1-10 period during the 56 years and they seem to fit a gamma model well. So it seems to be the action of totalling that might cause the lack of fit in the 10-day data.

Fig. 11.3l Probability plot of daily data



On the fitting process, two general methods of estimating parameters from any distribution are

a) the method of Moments
b) the method of Maximum Likelihood

The method of moments is usually straightforward. It consists of equating the sample 'moments' - these are the mean and the variance - with the population values.

For the gamma model equate

$$\bar{x} = \tilde{\mu},\ \ s^2 = \mu^2 / k \qquad \text{i.e.}\ \ \tilde{\mu} = \bar{x} \qquad \tilde{k} = \bar{x}^2 / s^2$$

Fig. 11.3d shows that $\bar{x}$ =44.8mm and s=27.4mm. Hence $\tilde{\mu}$=44.8 and $\tilde{k}$=2.67. To use the gamma dialogue to give these results directly, use **Statistics => Simple Models => Gamma** (Fig. 11.3h), with the methods of moments option. The maximum likelihood estimate was shown to be $\hat{k}$=2.08.

With the gamma model, the maximum likelihood estimator of μ turns out to also be the sample mean, i.e. $\hat{\mu} = \bar{x}$.

The method of maximum likelihood is often better than the method of moments giving more precise estimates of unknown parameters, i.e. they have a lower standard error. However, the calculations are sometimes more complicated.

For the gamma model, the maximum likelihood estimator of k cannot be written down explicitly, but an approximate value is

$\hat{k}$ =*(1+√(1+4y/3)) / 4y* where y=log($\bar{x}$) - Σ[(log(x)/n]

This translates into Instat commands as

: **x37=ln(x16)**
: **k1=ln(mea(x16))-mea(x37)**
: **k2=(1+sqr(1+4*k1/3))/(4*k1)**
: **? k2**

2.085 $(\hat{k})$

The exact value of the maximum likelihood estimate may be found by solving the equation

log($\hat{k}$ *)-ψ(* $\hat{k}$ *) = log(* $\bar{x}$ *)-Σ[log(x)/n]*

where $\psi(\hat{k})$ (pronounced 'psi') is the digamma function. It has to be solved iteratively, i.e. start with an approximate value (the method of moments can provide this), then 'home-in' on the value of k for which this equation holds.

Once a statistics package has been programmed to give the estimates of the parameters, the complication in calculating the maximum likelihood estimate, $\hat{k}$, compared with the method of moments, $\tilde{k}$ become irrelevant to the user.

11.3.4 Analysing data with zeros

The analysis is slightly more complicated when there is a chance that the whole period is dry (i.e. the 10-day total is zero). This problem is considered with the data for April 1-10 from Samaru, see Fig. 11.3b. There was no rain in that period in over half the years. One way round the problem is to add a small (arbitrary) amount to each zero. This is a common method, but it should not be done. It hides the problem, whereas the fact that there are zeros is important information. A better alternative is to analyse the data in two parts.

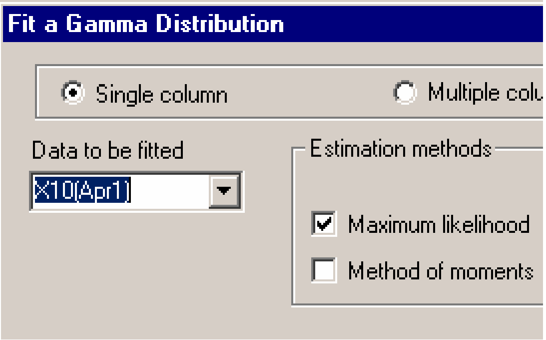
(i) Assess the chance of zero. In this example the 56 years have 32 'failures' (i.e. no rain in the period) and 24 'successes'. If years are independent and there is no trend, this is an example of the binomial distribution.

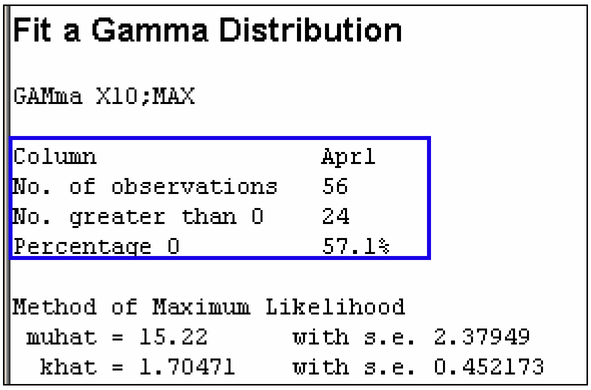
(ii) Fit a gamma distribution to the non-zero values

The analysis for the April data is shown in Fig. 11.3m. It is simple in Instat, because the gamma dialogue (for the fitting) and the probability and percentage dialogue (for the estimates) have been extended to cope with the two components of the model.

Fig. 11.3m Gamma model for rainfall totals for 1-10 April

`Statistics ⇒ Simple Models ⇒ Gamma`




Fit a Gamma Distribution

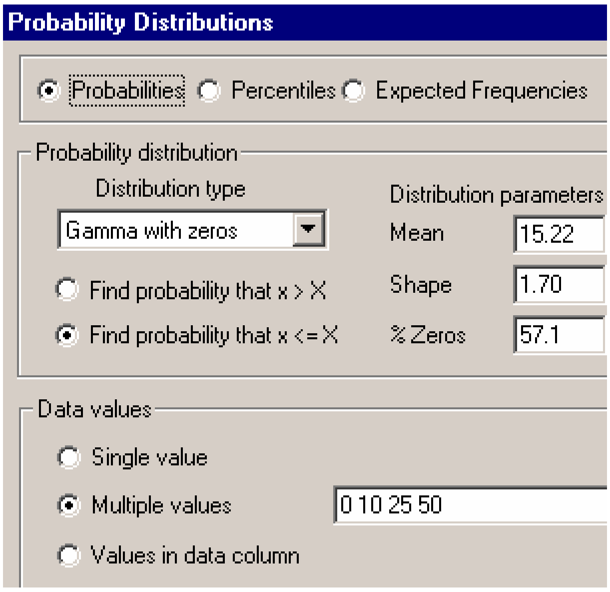
```
GAMma X10;MAX

Column                    Aprl
No. of observations       56
No. greater than 0        24
Percentage 0              57.1%

Method of Maximum Likelihood
 muhat = 15.22        with s.e. 2.37949
  khat = 1.70471      with s.e. 0.452173
```

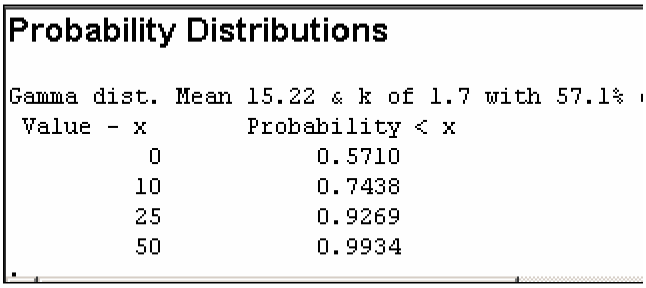
Thus the output from the Gamma dialogue in Fig. 11.3m includes the fact that 57% of the observations are zero and there are just 24 non-zero observations. The gamma distribution, fitted to these 24 values has a mean of 15.2mm and shape parameter of 1.7.

To use this model, first consider how to estimate a probability and take, for example, the chance of less than 10mm.

Fig. 11.3n Probabilities from the gamma model

`Statistics ⇒ Probability Distributions`

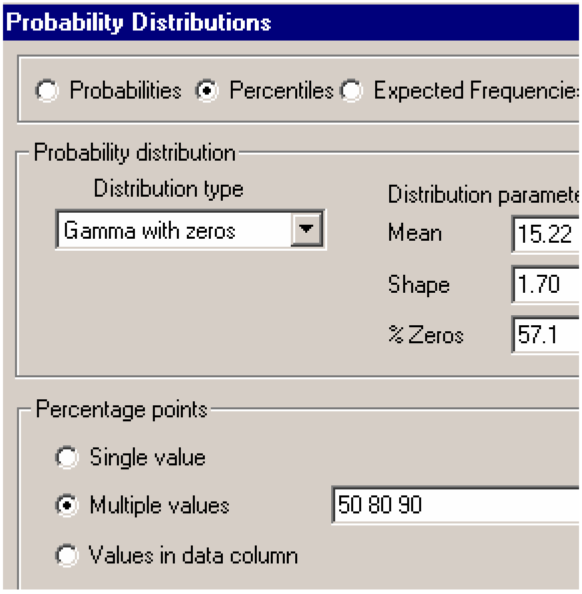


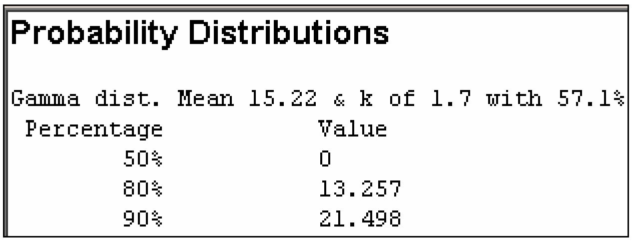

Probability Distributions

```
Gamma dist. Mean 15.22 & k of 1.7 with 57.1%
 Value - x          Probability < x
         0               0.5710
        10               0.7438
        25               0.9269
        50               0.9934
```

Fig. 11.3n shows that 74% of years are estimated to have 10mm of rainfall or less, in this decade. The estimated percentages of years above, may be compared with the descriptive summary. This shows that 42, 51 and 56 of the years had less than 10mm, 25mm and 50mm in that 10-day period, i.e. 75%, 91% and 100%.

Fig. 11.3o Percentage points





Calculating percentage points is equally straightforward and Fig. 11.3o shows the estimated 50%, 80% and 90% points of the distribution which gives 0, 13mm and 21mm. (The observed percentage points, without assuming a gamma model, are 0, 17mm and 25mm).

For those who wish to understand the calculations when there are zeros in the data, Fig. 11.3p gives the same results, but without using the special facility in Instat to cope with zero values. The steps are as follows:

- Select the non-zero values
- Count how many there are
- Fit a gamma model to these data

Now to estimate the overall chance of less than 10mm (or other values)

- Find the probability of less than 10mm from the gamma model
- Adjust the probability for the chance of zero

And to estimate the 80% (or other percentage points)

- Find what the overall 80% point corresponds to, in the non-zero years
- Find that percentage point from the gamma model

They are supplied in a macro called fig11_3p.ins for those who wish to try.

Fig. 11.3p Calculating gamma probabilities and percentiles by selecting totals >0

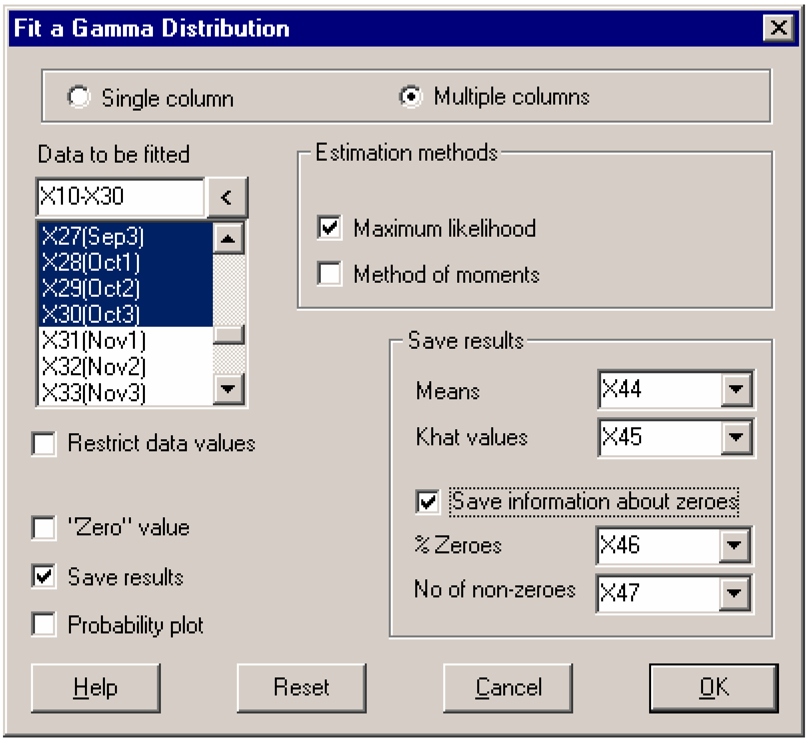
```
: select x10;into x42; if x10 > 0

Number of cases = 24
:
: k1 = count(x42)/count(x10)
: k2 = 1 - k1
: Note k1 is the prop of years with rain
: Note k2 is the prop with the 10 days dry
: gamma x42;max x43

Column                  X42
No. of observations     24
No. greater than 0      24
Percentage 0            0.0%

Method of Maximum Likelihood
 muhat = 15.22       with s.e. 2.37949
  khat = 1.70471     with s.e. 0.452173
```

```
: prob 10 k3;gamma x43;less
Gamma dist. Mean 15.22 & k of 1.705
 Probability < 10  = 0.4024
: k4 = k2 + k1*k3
: dis "The chance of <10mm = " k4

    The chance of <10mm = 0.74
:
: Note Calculate the 80% point
: k5 = 100*(0.8 - k2)/k1
: Note The 80% point is the k5'th point
: Note                 of the gamma
: percent k5 k6; gamma x43
Gamma dist. Mean 15.22 & k of 1.705
 53.333 % point is 13.251
: dis "The overall 80% point = " k6

 The overall 80% point = 13.3
```

11.3.5 Analysis of multiple periods

The gamma dialogue can be used to model many periods simultaneously. In **samdis56.wor**, the 10-day decade totals for each decade of the year are stored in X1-X36.

Fig. 11.3q Fitted gamma distributions for the data from April to October

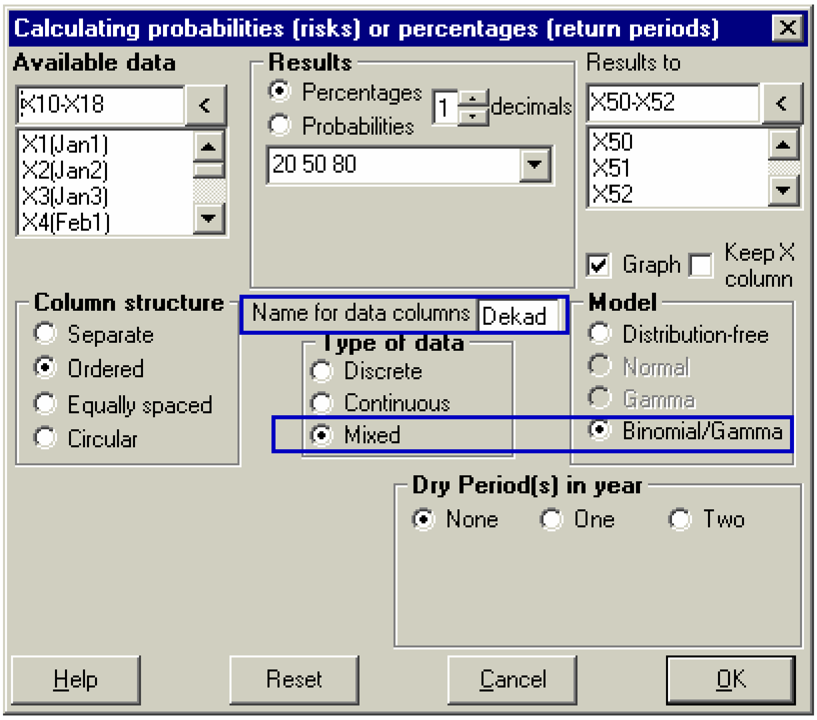


	X44	X45	X46	X47	X48 - l
	muhat	khat	%zero	nonzero	
1	15.2	1.7	57.1	24	Apr1
2	18.2	1.0	42.9	32	Apr2
3	26.6	1.3	23.2	43	Apr3
4	34.3	1.5	8.9	51	May1
5	35.4	1.3	5.4	53	May2
6	55.8	2.2	0.0	56	May3
7	44.8	2.1	0.0	56	Jun1
8	56.5	2.2	0.0	56	Jun2
9	56.2	2.9	0.0	56	Jun3
10	57.5	3.9	0.0	56	Jul1
11	72.1	4.1	1.8	55	Jul2
12	91.6	3.4	0.0	56	Jul3
13	82.1	3.1	0.0	56	Aug1
14	87.3	3.7	0.0	56	Aug2
15	106.7	5.4	0.0	56	Aug3
16	93.5	5.0	0.0	56	Sep1
17	74.4	3.0	0.0	56	Sep2
18	46.9	1.6	0.0	56	Sep3
19	27.5	0.9	26.8	41	Oct1
20	15.6	0.6	33.9	37	Oct2
21	14.6	1.0	76.8	13	Oct3

The dialogue in Fig. 11.3q shows the fitting of gamma models to the separate dekads from April to October. The results show that the end of August is the peak of the rainy season, with a mean of over 100mm in the dekad. The shape parameters (khat in Fig. 11.3q) reach about 5 at that time, compared to about 1, at the start and end of the season.

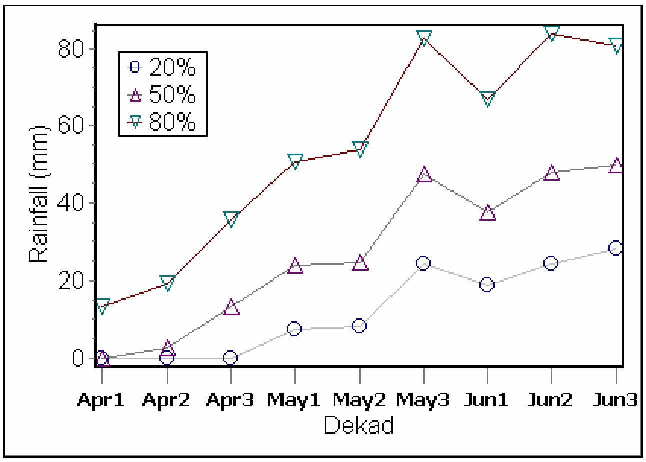
The **Climatic** ⇒ **Process** dialogue provides a one-step way of fitting the gamma models, described in this section and then estimating probabilities or percentage points. This is illustrated by showing the analysis of the 10 day totals in the **samdis56.wor** for April to June. The dialogue is in Fig. 11.3r. The 20%, 50% and 80% points of these totals are to be estimated.

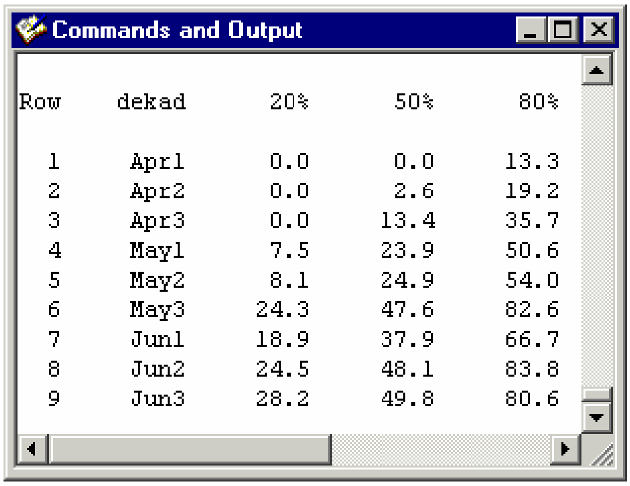
Fig. 11.3r The Climatic process dialogue



Both the numerical and graphical display of the results are shown in Fig. 11.3s. For example, in the first dekad in June the 20% point was estimated to be 18.9mm, the median, 37.9mm and the 80% point was 66.7mm.

Fig. 11.3s Graphical and tabular results




Commands and Output

Row	dekad	20%	50%	80%
1	Apr1	0.0	0.0	13.3
2	Apr2	0.0	2.6	19.2
3	Apr3	0.0	13.4	35.7
4	May1	7.5	23.9	50.6
5	May2	8.1	24.9	54.0
6	May3	24.3	47.6	82.6
7	Jun1	18.9	37.9	66.7
8	Jun2	24.5	48.1	83.8
9	Jun3	28.2	49.8	80.6

With this dialogue the data for the full year can be analysed, though there is sometimes a further complication concerning the zero values. This complication can be seen from the zero option in the **Climatic** ⇒ **Examine** dialogue shown in Fig. 11.3t.

Fig. 11.3t Percentage of dekads with no rainfall

`Climatic` ⇒ `Examine`

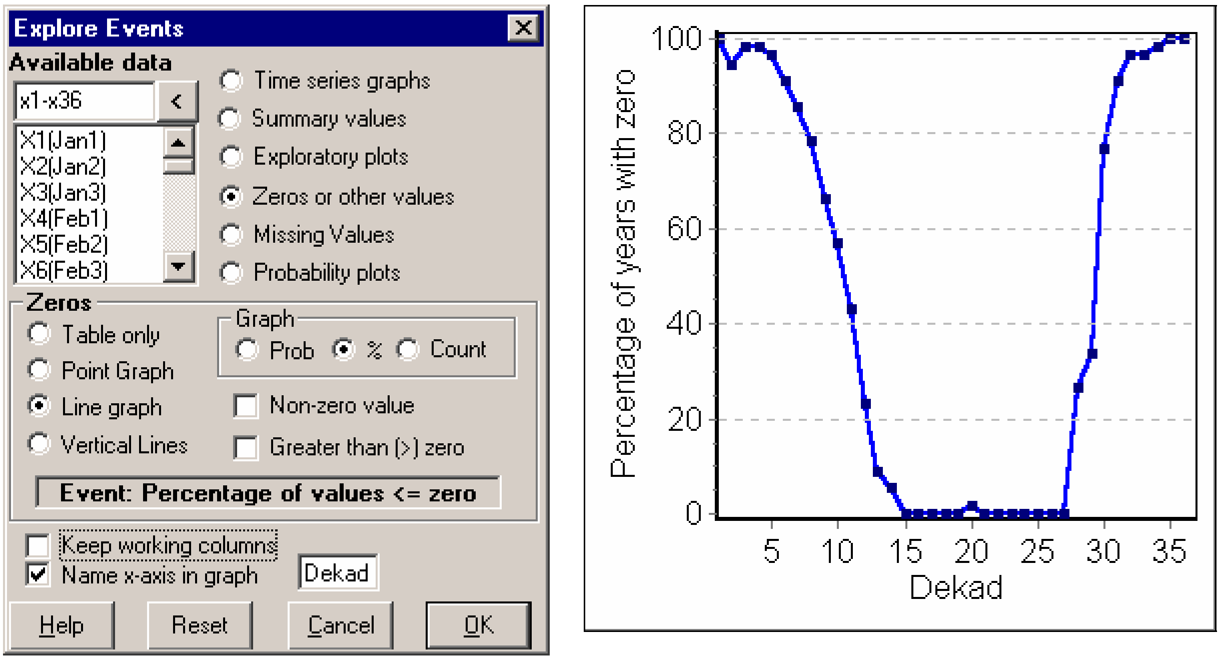


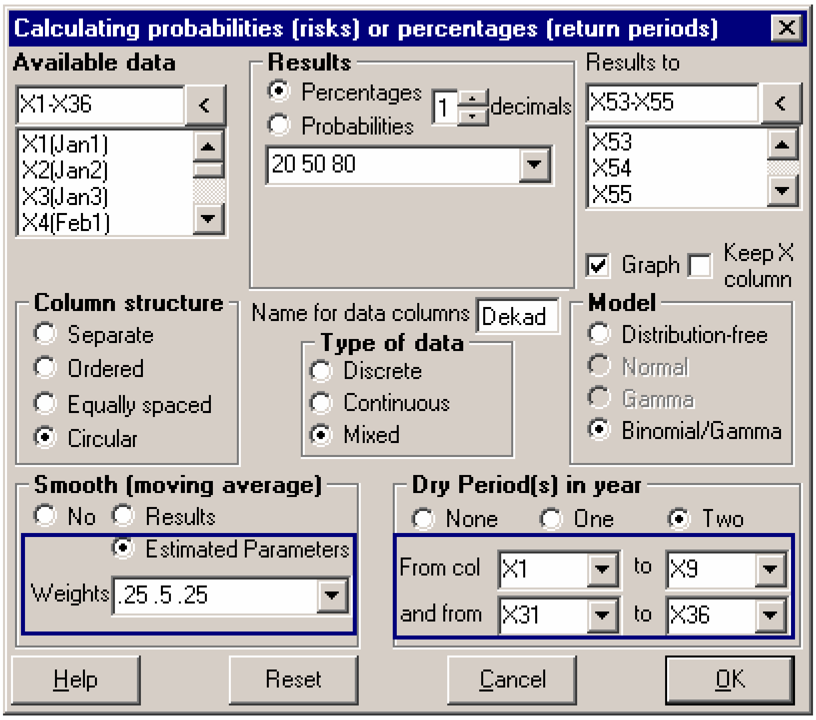
Fig. 11.3r shows there are some 10-day periods (in December and January) when there was never rain, hence there is no data to estimate the rainfall amounts. This does not matter when the **Climatic⇒ Process** dialogue is used in a simple way. By **simple** we mean that each 10-day period is analysed separately. But, it is sometimes useful to consider **smoothing** the results using neighbouring periods and then we need to know about the distribution of rainfall amounts there might be on the rare occasions that there is rain.

The option provided for this eventuality is to specify the part of the year that might be completely dry, which for this site is roughly for the months November to March. Then all the columns in that period are analysed together and a single average gamma distribution is fitted. Then the fitted value is used for any decade that is completely dry. This is just for the mean and a shape parameter of 1, i.e. an exponential model is assumed for that period.

An example is shown in Fig. 11.3u. Two dry periods are given in the dialogue, because January to March (X1-X9) is specified separately from November to December (X31-X36) as the possible very dry periods. We have also chosen one of the options for smoothing described in Chapter 12.

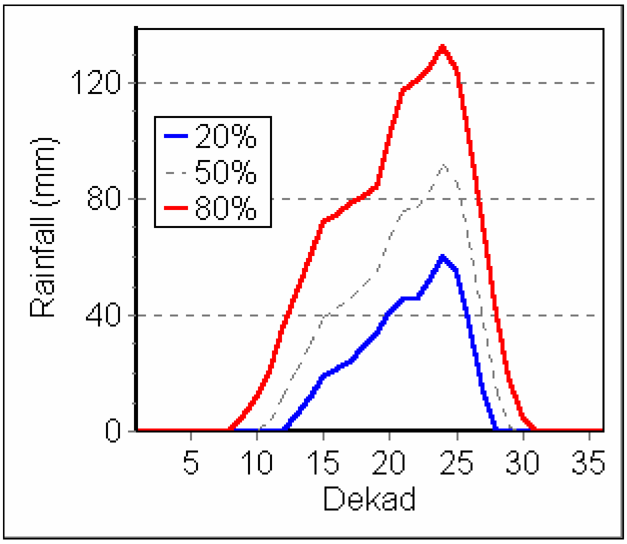
Fig. 11.3u calculating percentages for the full year

`Climatic ⇒ Process`



The graphical and some of the numerical results are in Fig. 11.3v. The slight differences from the results in Fig. 11.3s are due to the smoothing.

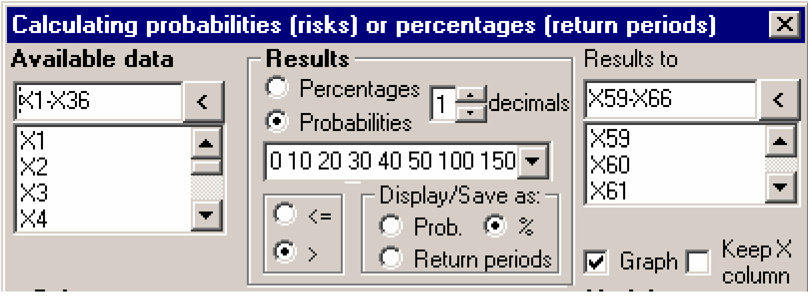
Fig. 11.3v Percentage points through the year



10	Apr1	0.0	0.0	11.5
11	Apr2	0.0	4.2	21.5
12	Apr3	0.0	12.7	35.1
13	May1	5.4	21.4	47.6
14	May2	11.5	30.2	60.4
15	May3	18.6	39.5	71.6
16	Jun1	21.6	42.8	75.0
17	Jun2	24.1	46.1	78.7
18	Jun3	28.7	50.3	80.9
19	Jul1	33.6	55.3	84.6
20	Jul2	41.0	66.7	101.1
21	Jul3	45.7	76.2	117.9
22	Aug1	45.8	77.3	120.9
23	Aug2	52.0	83.3	125.3
24	Aug3	60.4	91.9	132.8
25	Sep1	55.4	85.4	124.9
26	Sep2	37.8	64.8	102.5
27	Sep3	14.8	37.3	71.8

The **Climatic => Process** dialogue can equally be used to calculate probabilities. Fig. 11.3w Shows all that need be changed, compared to Fig. 11.3u where percentage points were calculated.

Fig. 11.3w Calculating probabilities



There is also a macro called **rainprob.ins**. This does the complete analysis, starting with the daily rainfall data. It enables a large number of stations to be processed routinely. It may be used by typing

: **@RAINPROB**

and responding to the questions. Partial results from the analysis of the daily data from the same site, in the file samaru56.wor, are in Fig. 11.3x.

Fig. 11.3x Results from the @rainprob macro

Estimated probability of decade total exceeding specified limits

Decade	0mm	10mm	20mm	30mm	50mm	100mm	150mm
Jan_1	3	0	0	0	0	0	0
Jan_2	4	0	0	0	0	0	0
Jan_3	5	0	0	0	0	0	0
May_1	89	71	52	38	19	3	1
May_2	96	88	74	59	34	7	1
May_3	98	95	84	68	38	6	1
Jun_1	100	99	94	83	50	6	0
Jun_2	100	100	95	83	50	6	0
Jun_3	100	100	96	86	56	8	1
Dec_1	3	1	0	0	0	0	0
Dec_2	1	0	0	0	0	0	0
Dec_3	0	0	0	0	0	0	0

11.3.6 Conclusions

In this topic the gamma distribution has been used to model 10-day rainfall totals. This was for illustration only. Gamma models can be used both for (non-zero) daily amounts and seasonal totals. Daily rainfalls are very skew and the value of the shape parameter, k, is usually estimated to be between 0.5 and 1.

One practical problem when modelling daily rainfalls arises from the fact that small amounts are often not recorded accurately. The estimate of the shape parameter, k, can be sensitive to these inaccuracies. There is a way round this problem, by using what is called the 'censored' data, but this is more advanced. What we normally suggest is that small amounts should be ignored, i.e. treated as zero.

With monthly or annual totals, in places where rain is frequent, the fitted gamma distribution will give quite a large value of k. With large k, the gamma distribution becomes very similar to the normal distribution. Thus, the use of gamma distributions for rainfall totals is consistent with the standard methods of analyses that assume a normal distribution provides an appropriate model for monthly or annual totals.

11.4 Extreme events

11.4.1 Overview

The probability (risk) of extreme events is important for many purposes, particularly when they are damaging, such as heavy rainfall, high flood flows in rivers, high winds, extreme temperatures (hot and cold) and even low flows in a river at a time of drought. Extreme event probabilities are required in engineering works - dam construction, roads (for example in determining the correct size of bridges or culverts), or irrigation projects. They are also needed in hydrology to determine maximum rainfalls, or maximum flood levels.

Two approaches exist in both rainfall and river flow analysis to estimate probable maximums. The first is mechanistic - where for example the probable maximum precipitation (PMP) is determined from knowledge of atmospheric conditions, such as humidity, inflow of air and precipitable fraction, or the maximum river level from a knowledge of river basin morphology, flood routing and channel hydraulics. The second approach, which is considered here, is statistical. In this case the estimate of the frequency with which a given magnitude of rainfall or streamflow may be exceeded in the future is based on the frequency with which it has been exceeded in the past. Inevitably, records are short so techniques have been developed to model the distribution of the parent population from which the sample is drawn.

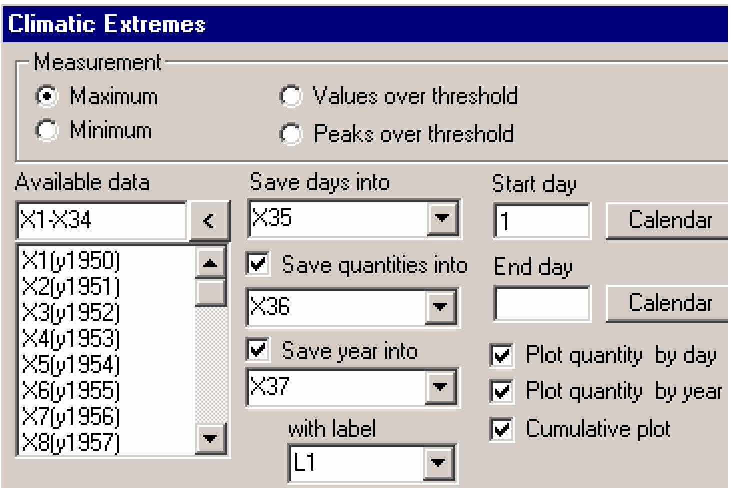
Often, for extreme events, the maximum value of the rainfall, river flow, etc. each year are selected to form the *Annual Maximum Series*. This ignores other large events that, in any one year, may exceed the largest in another year. An alternative is to include all values above a particular threshold. Extreme value frequency analysis entails the estimation of, for example, the peak flow Q(T), which is likely to be equalled or exceeded once *on average* in a specified period of T years. This peak value then has a *return period* of T years. T is the **long-term** average of the intervals between successive occurrences of a flow of size Q(T). For example, 25-year return periods may occur at intervals considerably greater, or less, than 25 years, but there will be an average of 4 occurrences in every 100 years.

11.4.2 Calculating the extremes

In text-books, the starting point for the analysis is usually a column of annual extremes. In general however, as with the analyses earlier in this guide, it is better to start with the raw data. As an example, consider the 34 years of daily rainfall data from Kurunegala in Sri Lanka (1950-83) that were processed in Section 6.6. This used the dataset **Kurunega.wor**. The **Climatic** ⇒ **Events** ⇒ **Extremes** dialogue was used in Section 6.6 to give the summary shown in Fig. 11.4a to give the summary values, that are ready for the second stage in the analysis.

Fig. 11.4a Extremes dialogue

Climatic ⇒ Events ⇒ Extremes



	X35	X36	X37 - F
1	26 Jul	71.6	1950
2	01 Nov	97.5	1951
3	19 Apr	87.6	1952
4	25 Oct	90.9	1953
5	13 Mar	91.7	1954
6	19 Oct	97.8	1955
7	17 Jun	122.4	1956
8	25 Dec	158.0	1957
9	23 Mar	168.4	1958
10	05 Apr	105.2	1959
11	03 Apr	71.6	1960
12	23 Oct	85.6	1961
13	09 Oct	97.8	1962
14	21 Oct	119.1	1963
15	27 Jul	[illegible]	1964

We emphasise the value of starting with the raw data. The first point is that these data have been checked and there is therefore confidence that the values in X36 are really the maximum values. The alternative would be to calculate the maxima separately and then input these values. This is both wasteful in time and can result in typing errors.

The second point is that initial analyses often lead to further questions. For example:

a) The four-month period from September to December is particularly important. Could an analysis of the maximums for this period be undertaken?

b) When in the year do the heavy rainfalls or annual maximums occur?

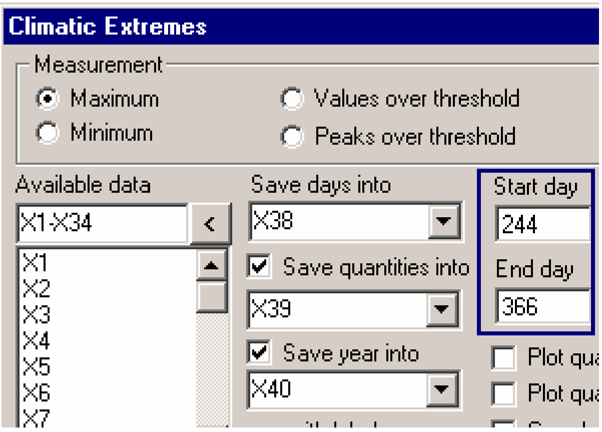
c) How many observations in the whole record are >110mm and in which years do they occur?

These are reasonable questions and all can easily be considered if there is access to the raw data. They illustrate that most analyses involve an initial data management phase, followed by the analysis. This point has been made before and studies of extremes are no different.

These three problems are considered in turn.

Fig. 11.4b shows that the analysis in Fig. 11.4a can easily be adapted to give the maximums in a specified period of the year.

Fig. 11.4b Maximum rainfall during September to December

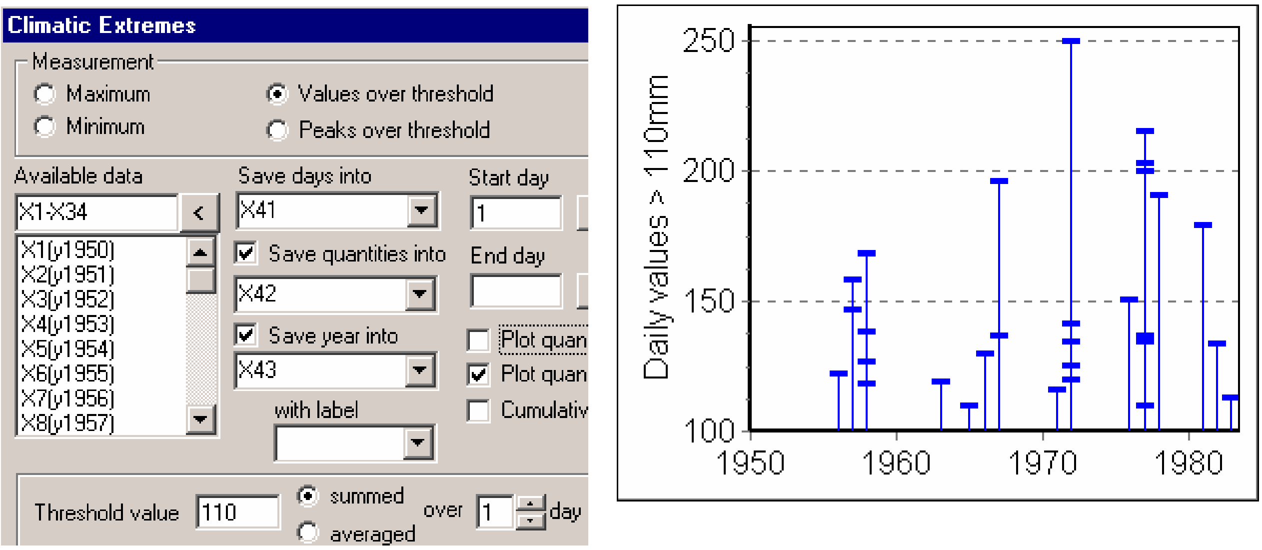


	X38	X39	X40 - F
1	07 Nov	49.5	1950
2	01 Nov	97.5	1951
3	03 Oct	72.9	1952
4	25 Oct	90.9	1953
5	15 Oct	75.9	1954
6	19 Oct	97.8	1955
7	26 Dec	82.3	1956
8	25 Dec	158.0	1957
9	24 Oct	126.8	1958
10	30 Nov	78.2	1959
11	17 Nov	69.6	1960
12	23 Oct	85.6	1961
13	09 Oct	97.8	1962
14	21 Oct	119.1	1963

The second task is to consider the annual maximums, as shown in Fig. 11.4b, but also to record which day in the year that the maximum occurred. As shown in Fig. 11.4a and ***11-4b*** this is an option on the same dialogue, and the results are in x35 and x38.

The third task is to consider all days with more than 110mm of rain. This is called the partial duration series. It is another option of the **Climatic** ⇒ **Events** ⇒ **Extremes** dialogue, Section 6.6, and Fig. 11.4c.

Fig. 11.4c Generating a partial duration series



The results from the analysis in Fig. 11.4c show that there were a total of 30 observations, in 15 of the years. What also stands out is that X28 (corresponding to 1977) had 7 occurrences. This deserves a more careful study and the **Climatic ⇒ Display Daily** dialogue (Fig. 11.4f) reveals that October 1997 was an extraordinary month. It had a total of 1735mm, with 8 days having more than 100mm of rain and 11 with more than 75mm. November was also very wet, with a total of 839mm. The total from these two months was more than the annual rainfall in most of the other years.

Fig. 11.4f Rainfall for 1977

Climatic ⇒ Display Daily for X28

Daily data for: X28

Mon Day.	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	--	--	20.6	0.5	0.3	46.5	7.1	--	6.6	62.7	42.7	--
2	--	--	--	0.5	14.5	14.0	--	--	1.0	--	41.7	--
3	--	--	--	--	12.7	5.1	0.5	--	21.1	--	46.0	--
4	--	--	--	7.4	9.4	23.9	--	--	--	22.9	1.3	--
5	--	--	5.6	24.9	2.0	2.5	10.2	--	1.5	38.4	--	--
6	--	--	--	24.6	19.0	--	0.8	--	0.8	202.4	--	--
7	--	--	0.8	--	41.9	2.5	--	--	--	8.9	92.0	--
8	--	--	--	--	7.9	--	1.0	--	--	8.1	199.6	--
9	1.8	--	--	14.7	33.5	--	3.6	--	--	16.3	--	--
10	--	--	--	15.2	17.3	--	18.3	74.9	0.3	30.7	--	4.1
11	--	--	15.0	15.2	--	--	1.8	--	2.0	86.9	26.2	--
12	--	--	--	--	--	9.1	1.0	--	--	27.2	20.3	5.6
13	--	--	--	16.0	36.3	--	51.8	5.1	--	215.4	35.1	--
14	--	--	--	--	17.5	0.5	0.3	--	1.8	94.0	79.5	--
15	--	--	--	--	9.6	1.8	--	--	--	--	62.7	--
16	--	--	--	3.3	0.5	11.9	--	0.8	--	17.0	--	19.8
17	--	1.3	--	3.6	2.5	--	--	45.7	--	110.2	--	23.1
18	--	0.5	69.1	10.7	2.5	--	1.3	36.8	--	7.6	4.1	9.4
19	--	0.3	12.2	3.6	18.5	21.8	4.3	49.0	--	106.7	--	28.5
20	--	--	31.5	3.3	0.8	13.0	0.3	31.5	--	10.4	--	1.3
21	--	--	--	4.8	27.9	5.6	--	31.5	--	76.7	--	--
22	--	--	--	--	--	--	--	2.5	--	134.6	--	--
23	--	1.3	--	40.4	--	--	--	--	30.7	203.0	52.8	--
24	--	23.6	--	4.1	--	--	--	20.6	--	109.2	--	0.8
25	--	--	18.8	42.9	--	0.5	1.3	--	--	--	11.9	--
26	--	--	--	12.2	1.5	--	--	--	1.3	--	1.5	--
27	--	0.5	5.6	9.4	0.5	0.3	--	--	--	5.6	108.2	--
28	--	--	0.3	5.3	2.8	1.0	--	--	3.8	2.5	13.2	--
29	--		3.3	31.8	11.2	0.5	--	0.8	11.9	137.2	--	--
30	--		4.6	--	14.2	--	--	1.3	27.4	--	--	--
31	--		15.2		38.4		--	5.8		--		4.1
Tot	1.8	27.4	202.4	294.4	343.4	160.5	103.4	306.3	110.2	1734.6	838.7	96.5

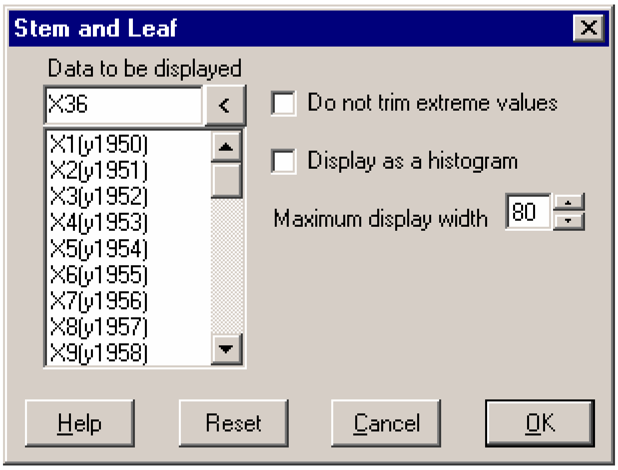
What has been done in this section parallels the first stage of the analyses considered in Chapters 5 and 6. We have looked at what might be termed ‘extreme events’.

11.4.3 Analysing the extremes

The extremes in Fig. 11.4a are used. A stem-and-leaf plot presents the data like a histogram, Fig. 11.4d.

Fig. 11.4d Stem-and-leaf plot of annual maxima rainfall

Graphics ⇒ Stem and Leaf

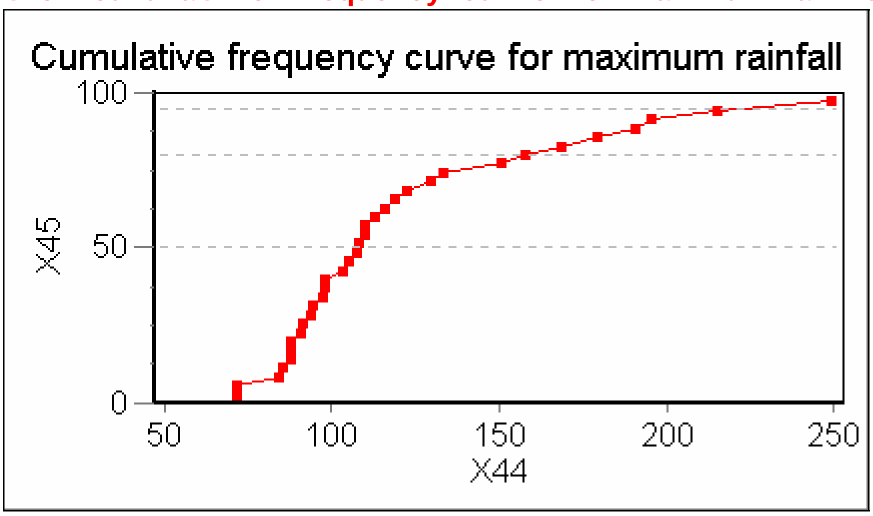



```
STEM & LEAF DISPLAY FOR X36
 Min=71   Max=249   No. obs.=34
 195 is represented by 19 5

    2     7  11
    7     8  45778
   14     9  0134777
  (5)    10  35789
   15    11  0269
   11    12  29
    9    13  3
    8    14
    8    15  07
    6    16  8
    5    17  9
    4    18
    4    19  05

        HI: 215, 249
```

The plot indicates the positive skewness of the data and that the median value is just over 100mm.

Fig. 11.4e Cumulative frequency curve of maximum data

```
warn off
sort x36 x44
ent x45;data (1]34)
x45=100 * x45/35
line x45 1 1 1
symbol x45 8 0 1
plot x45 x44;href 50 80 95;nolegend;yaxis 0 100;
  title ‘Cumulative frequency curve for maximum rainfall’
```



A cumulative frequency curve is a good way to show different percentage points (or return periods). This was shown earlier (Section 6.6), when the extremes were generated. It can be given again with the **Climatic** ⇒ **Examine** dialogue. An alternative uses the commands and the resulting plot is shown in Fig. 11.4e.

On this plot the reference lines correspond to the

50% point	(2 year return period)
80% point	(5 year return period)
95% point	(20 year return period)

The **Statistics** ⇒ **Summary** ⇒ **Describe** dialogue or the command

: **DEScribe X39; PERc 50 80 95**

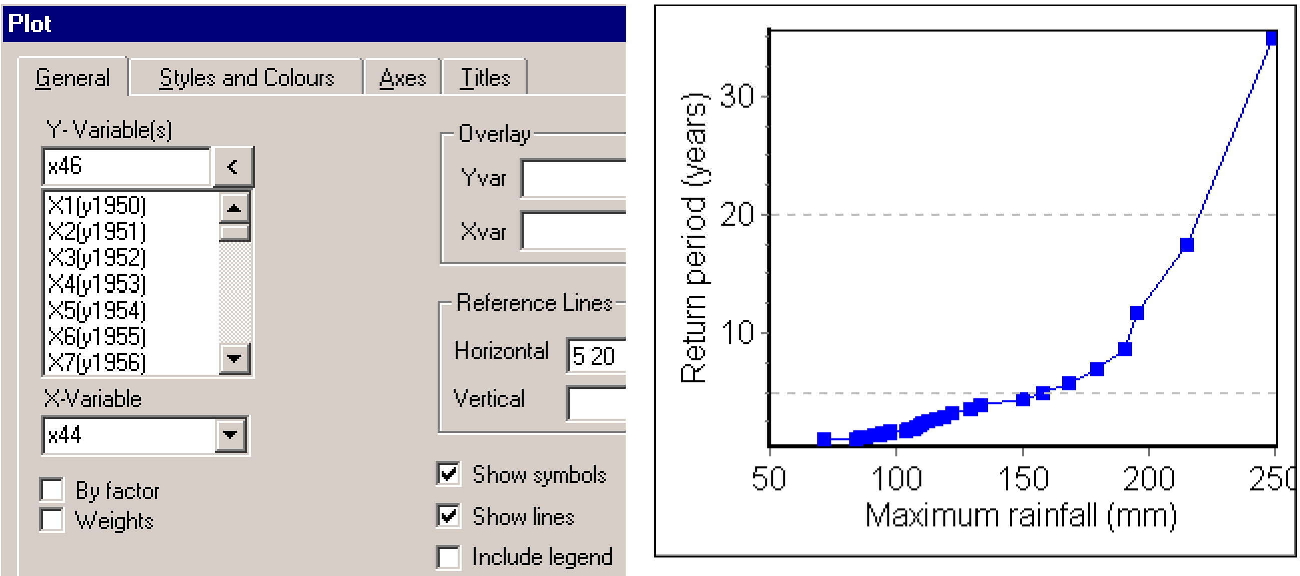
give the exact return periods for the maximum rainfall calculated in Fig. 11.4d as 108mm, 158mm and 224mm.

The cumulative probabilities, F, can be transformed into return periods, T, using T=1/(1-F) or T = 100/(100-P) for percentages and plotted as in Fig. 11.4e.

Fig. 11.4f Plot of return periods

Manage ⇒ Calculations ⇒ x46=100/(100-x44)

Graphics ⇒ Plot



Sometimes return periods are required that are larger than the number of observations. This is clearly a dangerous venture as it involves extrapolating outside the range of the data. For example in the **Statistics** ⇒ **Summary** ⇒ **Describe** dialogue or the command

: **DEScribe X39; PERc 50 80 95 99 99.9**

simply gives the largest observations for the 99% and 99.9% points.

A distribution-free approach cannot be used If extrapolation is required. Instead, a model is fitted to the data and the percentage points are estimated from the model. This remains a dubious exercise because one does not have any data with which to test whether the model is appropriate at these large return periods. Within the range of the data we can investigate whether an extreme value (a Gumbel distribution) fits the data. This is sometimes called the double exponential distribution, because the distribution function is

$$F(x) = Prob(X<x) = \exp(-\exp((x-u)/a))$$

Hence, calculate

$$L(x) = -\log(-\log(Fx)) \quad \text{then}$$

$$L(x) = (x-u)/a$$

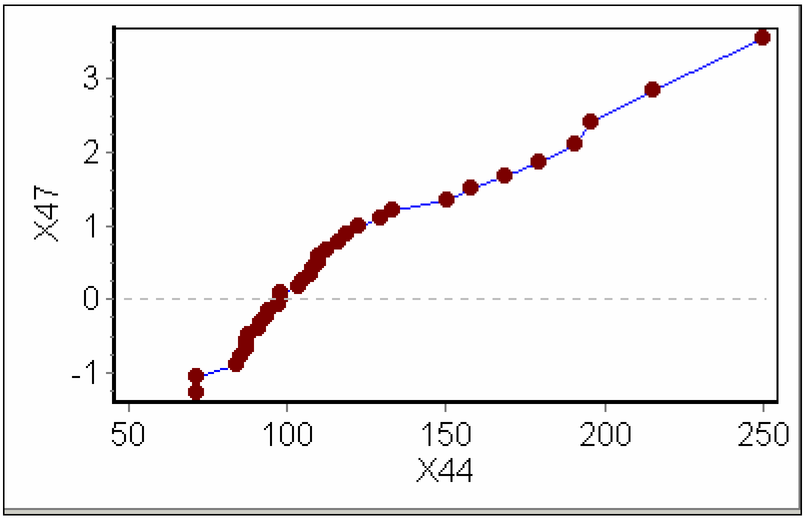
Thus plotting L(x) against x should give a straight line, if the model is a good fit.

The **Climatic** ⇒ **Examine** dialogue can be used to provide the plot. It is also not difficult the plot from first principles, see Fig. 11.4g.

Fig. 11.4g Plot of the Gumbel distribution

Manage ⇒ Calculations and calculate **X47 = -LN(-LN(X45/100))**

Graphics ⇒ Plot for **X47 v X44**



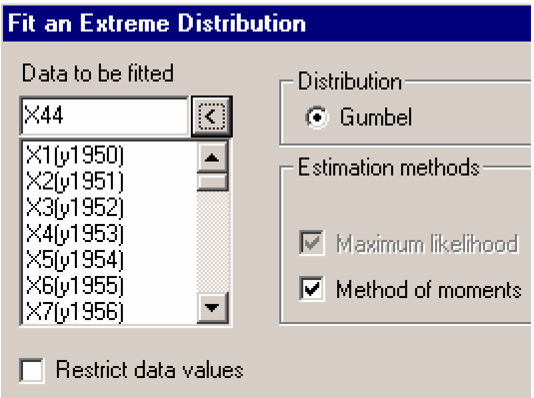
We proceed, even though the graph seems to be curved at lower values. The parameters, u and a of the extreme value distribution are related to the mean and variance by

$u = \bar{x} - 0.5772 * a$ where $a = \sqrt{(6s^2 / \pi^2)}$

Hence they can be estimated by the method of moments as in Fig. 11.4m. The estimates are $\tilde{u}$ =102 and $\tilde{a}$ =33.5.

Fig. 11.4h Using the method of moments

Statistics ⇒ Simple Models ⇒ Extremes

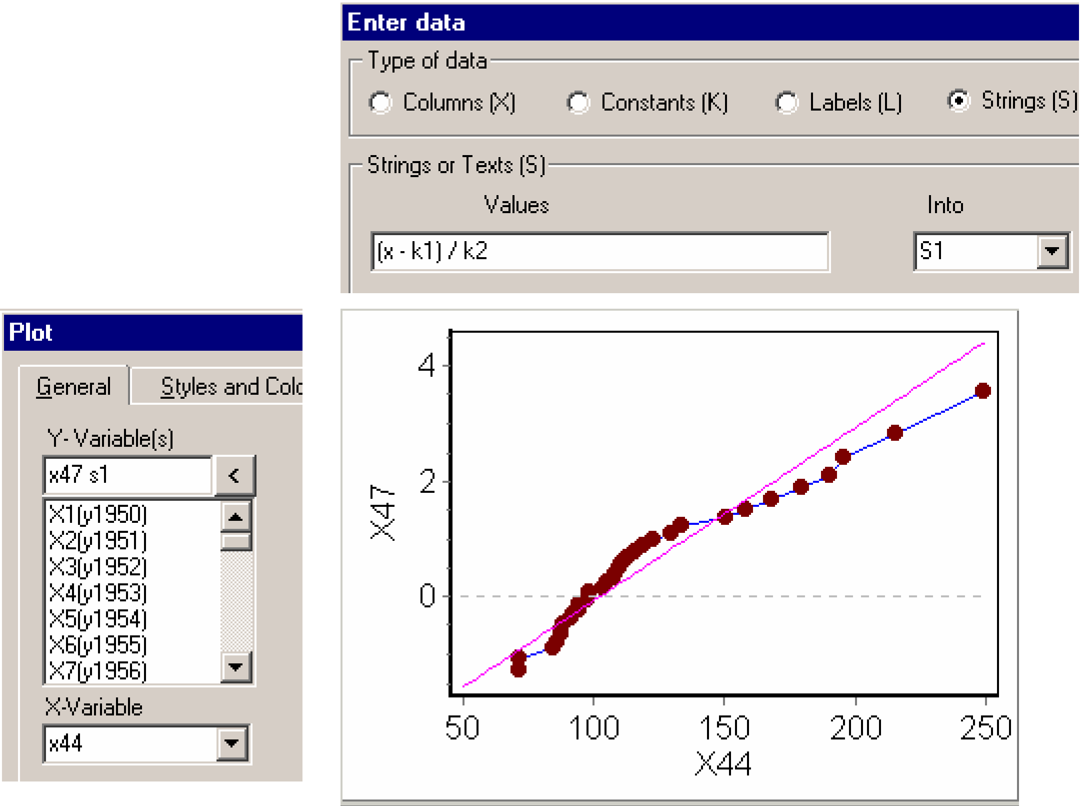


Fit an Extreme Distribution

```
Column                 X44
No. of observations    34
Mean                   121.3
Standard deviation     42.92

Method of Moments
Mode,  uhat = 102
Scale, ahat = 33.47
```

To see how this model will estimate extreme values, put K1 = uhat = 102 and K2 = ahat = 33.47, and add the standardised variable to the display, see Fig. 11.4i.

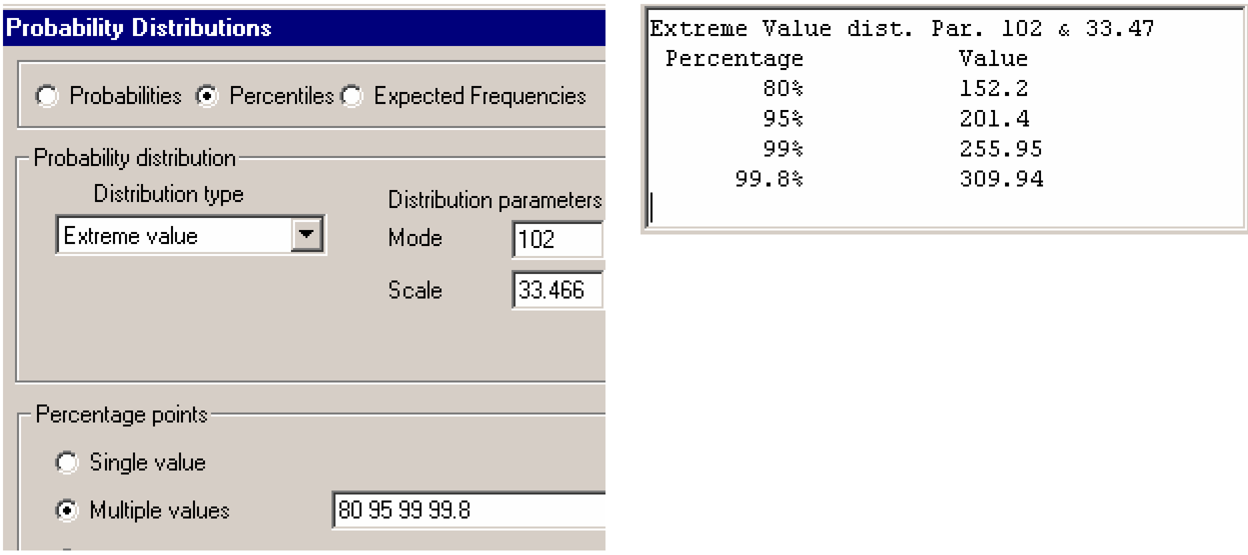
Fig. 11.4i Adding the fitted line to the probability plot



Estimates of return periods can then be found from the extreme value model, using the **Statistics ⇒ Probability Distributions** dialogue as shown in Fig. 11.4m. The estimates of the 5, 20, 100 and 500 year return periods are given as 152mm, 201mm, 256mm and 310mm.

Fig. 11.4j Calculating percentiles (return periods) from the fitted model

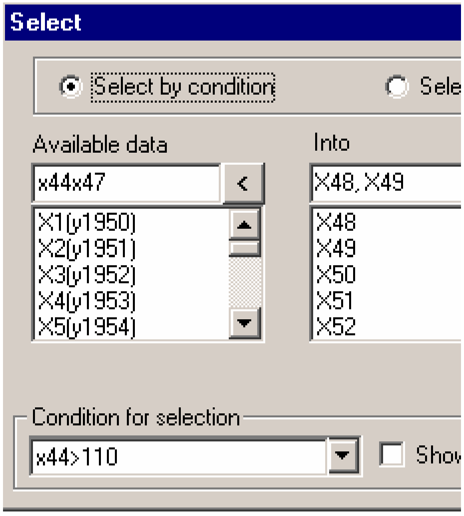
Statistics ⇒ Probability Distributions

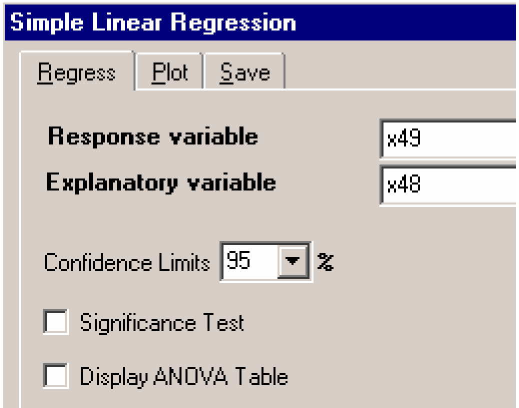


However, these estimates obviously have to be treated with caution, because the plot in Fig. 11.4i shows that the model is a poor fit at the higher return periods. There are various ways that the model could be improved. The large values look reasonably linear, so one possibility is to use the partial duration series, with a lower limit of about 110mm. A simpler possibility, which is illustrated here, is to estimate the parameters of the extreme value distribution by linear regression. This follows the observation earlier that the extreme value plot is of L(x)=(x-u)/a, i.e. a straight line, with intercept -u/a and slope 1/a. The steps are shown in Fig. 11.4k.

Fig. 11.4k Estimating the parameters for extremes > 110mm

Manage ⇒ Reshape ⇒ Select **Statistics ⇒ Regression ⇒ Simple**





The fitted regression model is **X49 = -1.593 + 0.0201 * X48**

Hence calculate k2 = 1/0.0201 = 49.112

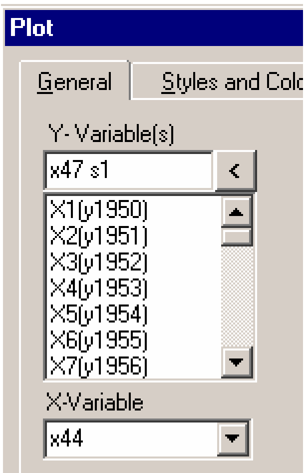
and k1 = 1.593 * k2 = 79.112

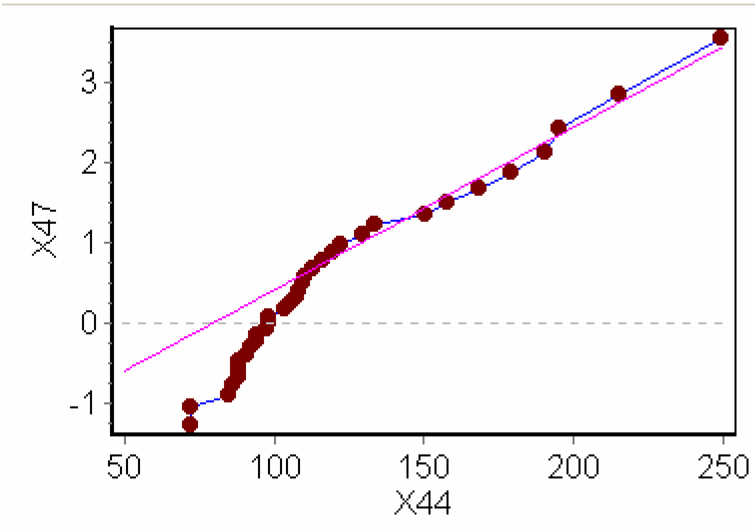
The estimates of the parameters of the extreme value distribution by this method are now

$k1$ = $\tilde{u}$ = 79.1, $k2$ = $\tilde{a}$ = 49.7

and the line is seen to be a much better fit at high return periods, Fig. 11.4l.

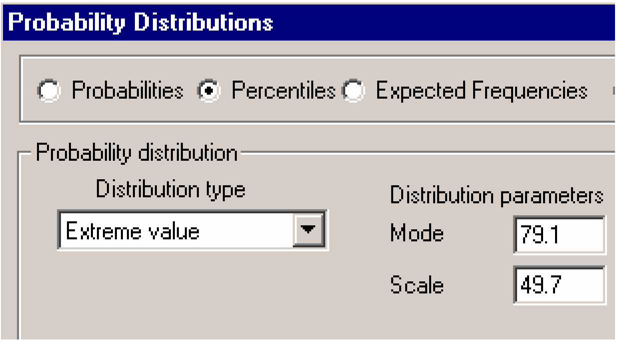
Fig. 11.4l Plotting the new model





The **Statistics => Probability Distributions** dialogue also gives estimates that seem more reasonable, such as a 100-year return period of 308mm, though it must again be emphasised that these are inevitably an extrapolation from the data.

Fig. 11.4m Percentiles (return periods) from the new model




```
Extreme Value dist. Par. 79.1 & 49.7
 Percentage            Value
       80%             153.65
       95%             226.72
       99%             307.73
     99.8%             387.92
```

11.4.4 In conclusion

We conclude this section with a general point. Earlier, in Fig. 11.4a, both the maxima and the dates on which they occurred are calculated. A simple examination of the dates shows that their distribution is bimodal, see Fig. 11.4n.

Fig. 11.4n Stem and leaf plot of the dates of the year of each extreme
Graphics => Stem and Leaf for X35

```
STEM & LEAF DISPLAY FOR X35
  Min=30   Max=360   No. obs.=34
 360 is represented by 3:6 6

    2     0:0-4 33
    8     0:5-9 788999
   11     1:0-4 111
   12     1:5-9 6
   14     2:0-4 00
  (10)    2:5-9 6678899999
   10     3:0-4 000111222
    1     3:5-9 6
```

This is reasonable and corresponds to the bimodal pattern of rainfall at this site. This aspect of bimodality has been ignored in the analysis. In general, analyses should not ignore obvious structure in the data. In this case it would probably be better to consider the distribution of extreme rainfalls separately for each season. Fig. 11.4c shows how the extremes can be calculated for the season from September to December and the analysis could be repeated for this season. The high return periods are very similar to those for the full year.

In the next chapter, this aspect of allowing for the structure of the data is discussed further in the case study in **Section 12.3**.

11.5 Correlations in climatology

The next two sections move from looking at variables one at a time, to a study of how two or more variables are related. The opportunities are endless, for example to examine the relationships between:

July and August rainfall

Dates of the start and the end of the rains

Rainfall and runoff

Sunshine hours and radiation

Accumulated temperatures (degree days) and crop yield

A simple way of studying the relationship between two variables is to calculate the correlation between them. Correlations are used, and also misused extensively and both aspects are covered here.

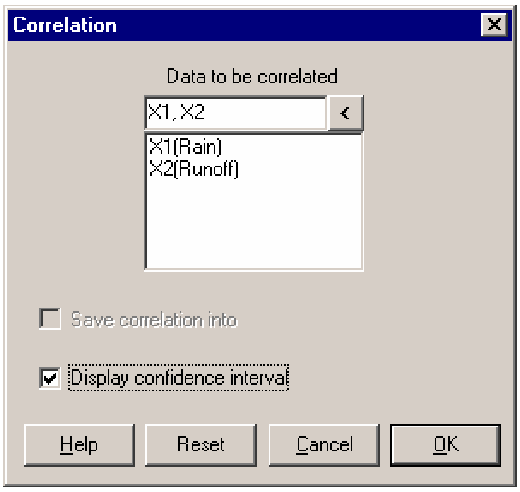
11.5.1 Example 1: Data from Gregory (1978)

Fig. 11.5a shows a set of data on runoff and rainfall, from Gregory (1978).

Fig. 11.5a Runoff Data
File ⇒ Open From Library ⇒ Runoff.wor

	X1*	X2*
	Rain	Runoff
1	46.4	31.9
2	63	46.8
3	48.8	34.2
4	60.1	47.5
5	50.6	35.2
6	57.5	40.5
7	55.5	41.3
8	57	43.5
9	60.8	44.8
10	48.3	38.5
11	59	39.1
12	41	26.5
13	66.7	46.5
14	56.4	43.4
15	58.3	40.9
16	55.7	41.3
17		

Fig. 11.5b Correlation for Runoff and Rain
Statistics ⇒ Regression ⇒ Correlation



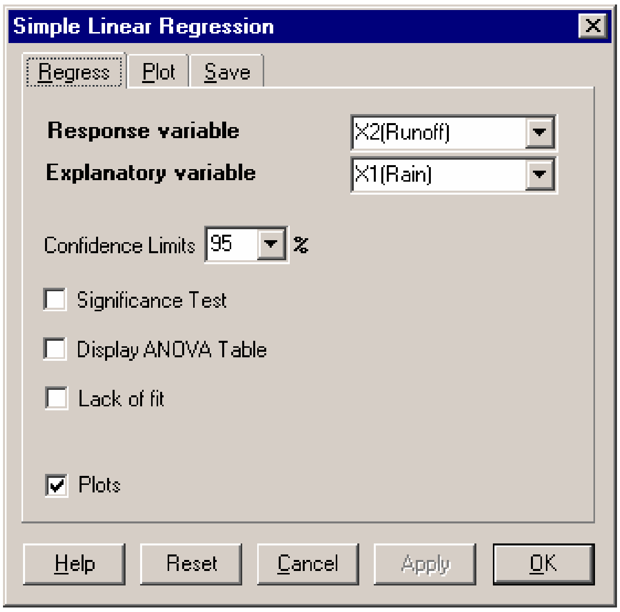

```
Correlation Rain, Runoff = 0.9214

Testing hypothesis that rho = 0
t value corresponding to correlation is 8.871 with 14 d.f.
Significance level (2 sided) is 0.00%

Approximate 95% confidence interval for rho is 0.76 to 0.98
```

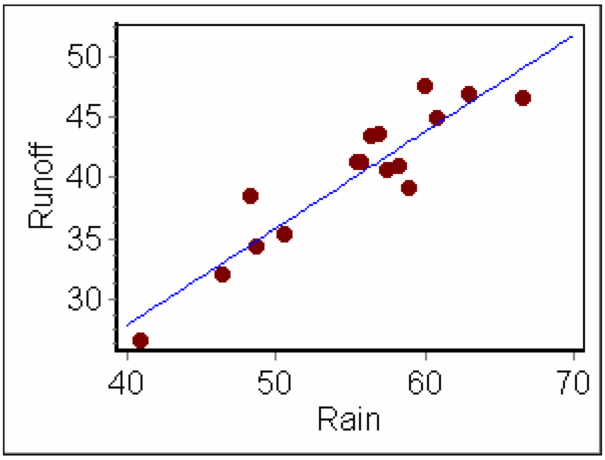
The **Statistics => Regression => Correlation** dialogue shows a high correlation of 0.92 between rainfall and runoff with the 95% confidence interval for the correlation to be 0.76 to 0.98.(Fig. 11.5b).

No misuse here, except that when a correlation is non-zero, it is often more useful to examine the equation, relating the two variables, than just to give the correlation. The **Statistics => Regression => Simple** dialogue in Fig. 11.4c produces a graph that confirms the clear relationship between rainfall and runoff.

Fig. 11.5c Simple regression analysis

Statistics ⇒ Regression ⇒ Simple





Simple Linear Regression

```
----------------- DETAILS OF THE FITTED LINE ------------------

Fitted equation      : Runoff = -4.016  +  0.7978 * Rain

Standard error of slope : 0.0899  with 14 d.f.
95% confidence interval for slope  0.60492    to    0.99073
R-squared              : 0.849
```

Fig. 11.5c shows that the equation is **runoff = -4.02 + 0.80 * rainfall** i.e. in the units measured, an increase of one in the rainfall gives an average increase of 0.8 in the runoff.

11.5.2 Example 2: Data from Benin, West Africa

The data in Fig. 11.5d are monthly rainfall totals from two stations, Porto Novo (6.30N, 2.47E) and Save (8.04N, 2.79E) in Benin, West Africa. In the worksheet **wafric2.wor**,

X1 gives the year number
X2-X5 give the totals from April to July at Porto Novo
X6-X9 give the totals from April to July at Save.

Fig. 11.5d Monthly rainfall totals

File ⇒ Open From Library ⇒ Wafric2.wor

	X1*	X2*	X3*	X4*	X5*	X6*	X7*	X8*	X9*
	Year	Apr_PN	May_PN	Jun_PN	Jul_PN	Apr_S	May_S	Jun_S	Jul_S
1	1938	263.4	104.6	155.4	97.3	110.2	106.3	213.7	27.5
2	1939	183.2	112.5	272.9	313.5	111.5	141.5	172.4	121.5
3	1940	97.1	375.2	337.8	75.3	135.5	181.8	113.9	69.0
4	1941	192.3	149.8	164.3	165.3	106.6	222.3	64.0	184.6
5	1942	37.7	338.3	423.0	24.6	55.7	162.3	130.1	51.1
6	1943	122.0	197.0	643.3	85.5	131.0	203.2	141.1	45.1
7	1944	40.2	112.3	285.8	245.9	124.5	144.1	152.8	126.5
8	1945	43.3	81.6	145.9	132.9	72.2	31.8	169.1	131.8
9	1946	139.8	244.9	210.9	0.0	66.1	101.8	110.2	61.4
10	1947	85.3	187.1	296.7	348.4	86.7	120.1	162.8	141.9
11	1948	162.0	381.8	241.0	6.6	253.1	116.5	86.6	69.6

Various correlations may be of interest, depending on the application. For example:

1) If there are correlations with the year (X1), there may be a trend in the data.

2) A positive correlation between successive months at the same site would indicate that rainy Aprils tend to be followed by rainy Mays, etc. If this is the case, then rainfall in previous months could be used as part of a long-term forecasting scheme.

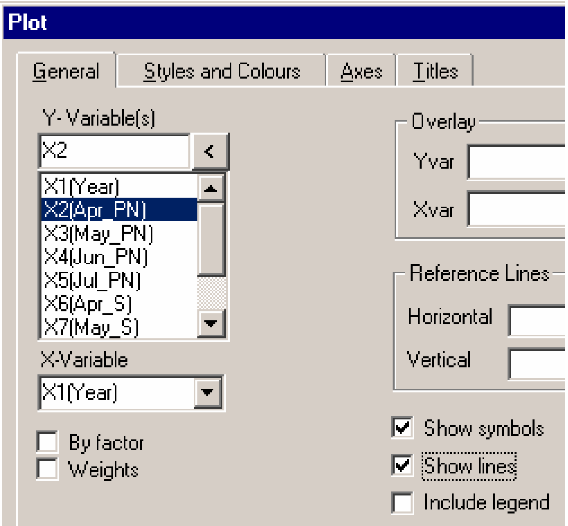
3) Correlate X2, or any other column against the same column for the previous year. These are correlations between successive years, i.e. serial correlations.

4) Correlate X2 against X6 etc. This corresponds to the spatial correlation between the two stations for the given month.

Fig. 11.5e examines a possible trend in the data by plotting, for example X2 against the year.

Fig. 11.5e Plot of monthly totals

`Graphics ⇒ Plot`



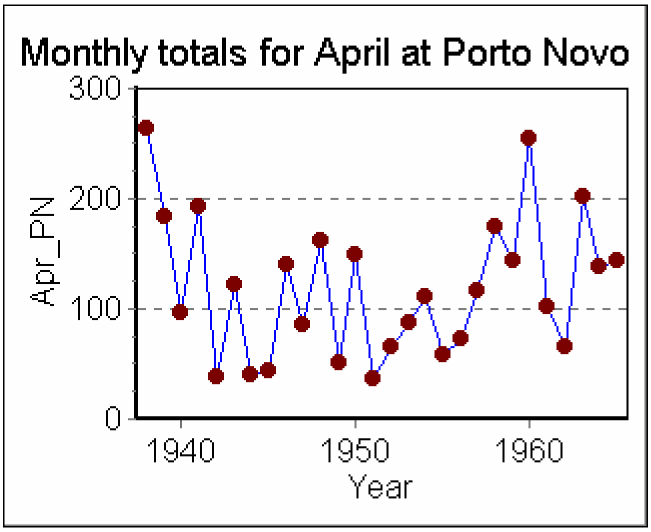
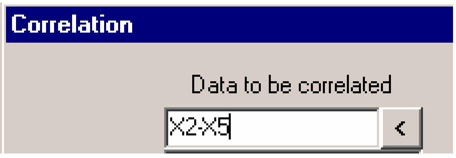
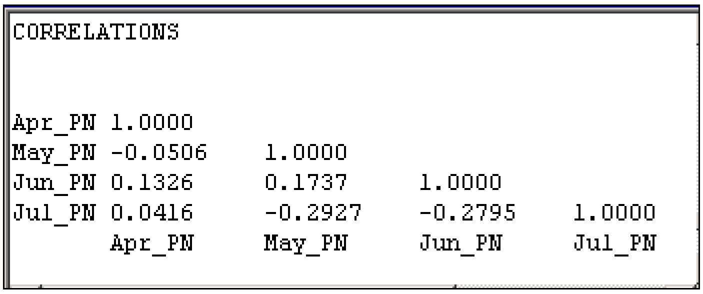


Fig. 11.5f checks for correlations between successive months in Porto Novo.

Fig. 11.5f Correlations between April-July

`Statistics ⇒ Regression ⇒ Correlation`





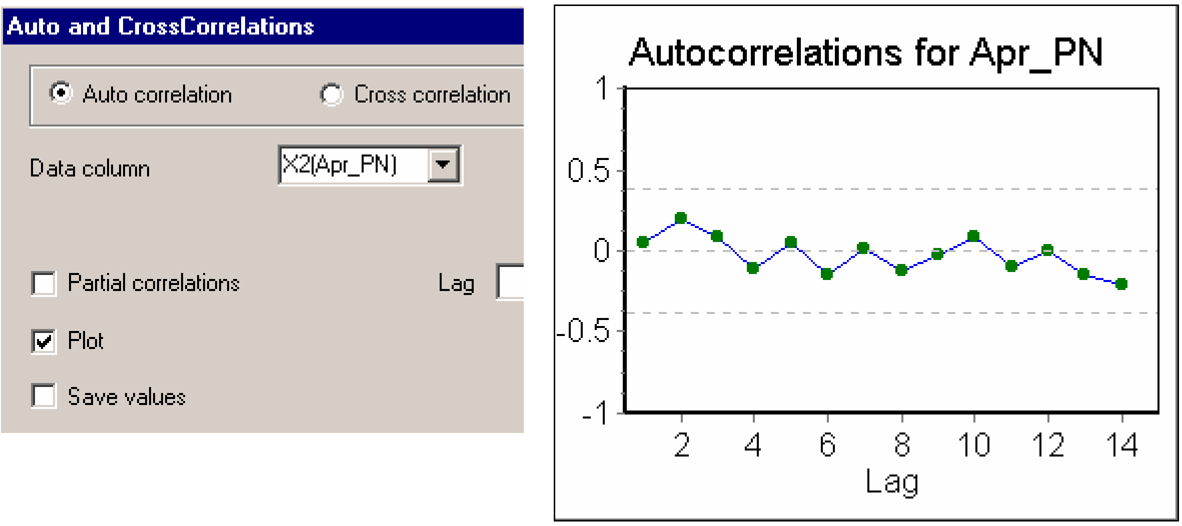
```
CORRELATIONS


Apr_PN 1.0000
May_PN -0.0506     1.0000
Jun_PN 0.1326      0.1737     1.0000
Jul_PN 0.0416      -0.2927    -0.2795    1.0000
       Apr_PN      May_PN     Jun_PN     Jul_PN
```

The results in Fig. 11.5f show no evidence of correlations between successive months at the same site.

Nxt use the **`Statistics => Time Series => Correlations`** dialogue in Fig. 11.5g to see if there are any correlations between successive years.

Fig. 11.5g Auto Correlation for X2

`Statistics ⇒ Time Series ⇒ Correlations`



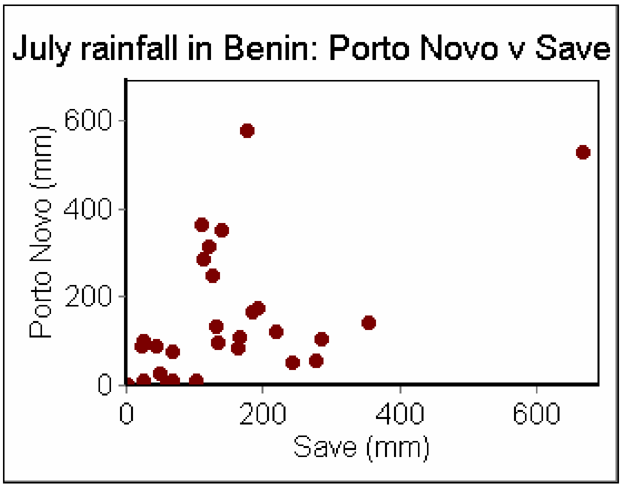
There is also no evidence of serial correlations.

> In the plot in ***Fig. 11.5g*** the horizontal lines are at $\pm 2/\sqrt{n} = \pm 2/\sqrt{28} = \pm 0.38$. These are approximate 95% confidence limits for each sample serial correlation, if the true serial correlations are zero. Hence this provides a guide to check on the magnitude of the observed correlations.

Next plot the July rainfall from Porto Novo against the July rainfall for Save and calculate the correlation coefficient for July (Fig. 11.5h).

Fig. 11.5h Plot and correlation for July at two sites

`Graphics ⇒ Plot (X5 against X9)` **`Stats ⇒ Regression ⇒ Correlation`**



Correlation

```
Correlation Jul_PN, Jul_S = 0.4783

Testing hypothesis that rho = 0
t value corresponding to correlation is
                       2.777 with 26 d.f.
Significance level (2 sided) is 1.00%

Approximate 95% confidence interval
          for rho is 0.11 to 0.73
```

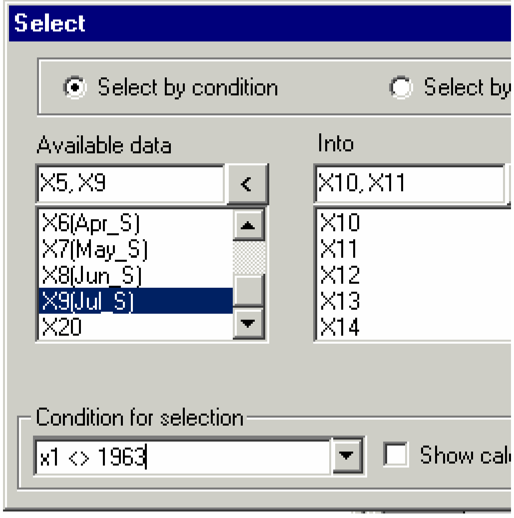
From the value of the correlation of 0.48 between the July rainfalls in x5 and x9, there seems initially to be evidence of a spatial relationship between the two sites. However, the plot in Fig. 11.5h, implies that care must be taken in the interpretation, because the high correlation is primarily due to a single point. We see later, in Section 11.5.4, how such values can be misleading in the use of correlations.

When this one point is omitted, the correlation drops to a low value, Fig. 11.5i. If the spatial correlation is real, it might also be evident in more than one month, but April (X2 with X6), May

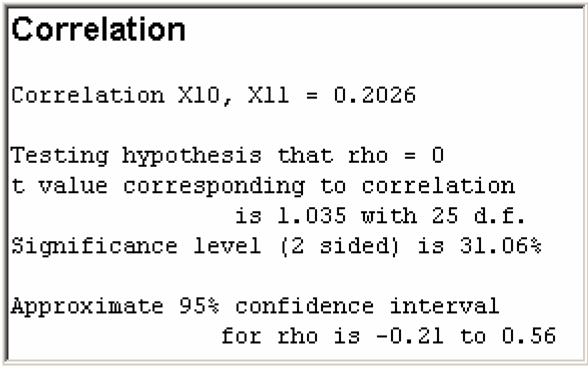
(X3 with X7) and June (X4 with X8), show nothing. Hence, perhaps surprisingly, there seems to be no evidence for a spatial correlation between the monthly rainfall totals at these two sites.

Fig. 11.5i Correlations in July without year 1963

Manage ⇒ Reshape ⇒ Select



Stats ⇒ Reg ⇒ Corr (X10 and X11)



```
Correlation

Correlation X10, X11 = 0.2026

Testing hypothesis that rho = 0
t value corresponding to correlation
                is 1.035 with 25 d.f.
Significance level (2 sided) is 31.06%

Approximate 95% confidence interval
               for rho is -0.21 to 0.56
```

Negative results are sometimes treated as a failure, but this is not always the case and negative evidence about correlations is often useful. For example, in previous chapters, data have often analysed as though successive years were independent observations from the same distribution. This assumes no trend or serial correlations in the data. Thus, negative evidence about correlations often permits a simple model to be adopted.

11.5.3 Example 3: Data from Samaru, Nigeria

The third example uses data of the type generated in Chapters 6 and 7. The data in **samrain.wor** consist of 56 years of data from Samaru, Nigeria 1928-83. Here

X1 is the year number

X2-X5 gives the date of the start of the season for 4 different definitions

X6-X8 gives the length of the longest dry spell in May, June and July

X9 gives the date of the end of the season, using a simple water balance definition

There are a number of useful correlations that can be calculated. For example, similarly to Example 2, Fig. 11.5b, any correlation with X1 would examine whether there is evidence of a trend in the data.

Here, for illustration, the second definition of the start of the season is considered. This is in X3 and is the first occasion after 1st April with more than 20mm in a two-day period and no 10-day dry spell in the following 30 days. With this definition of the start, the length of the season is also calculated and put into X10. This uses **Manage ⇒ Calculations**, or the command may be typed as

: X10 = X9 – X3

The correlation between X3 and X9, shows no evidence of any correlation between the start and end of the season. This is a useful result as is shown below.

Fig. 11.5j Boxplot of season length

`Graphics ⇒ Boxplot of X10`

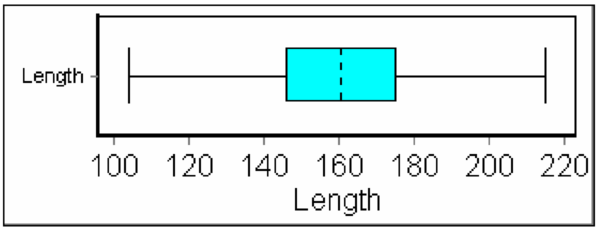
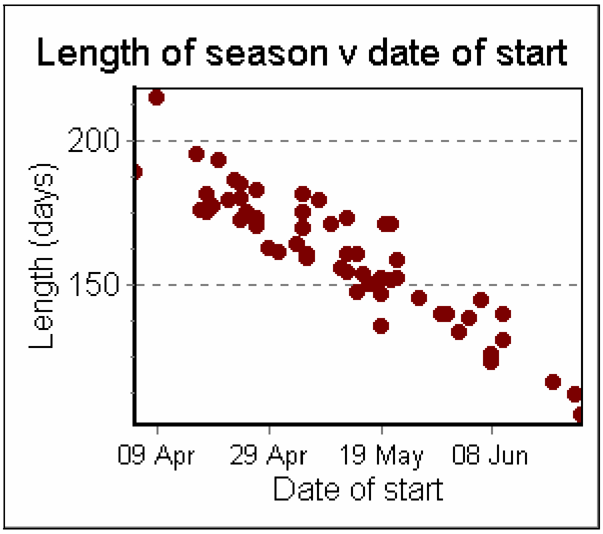


Fig. 11.5k Plot of length against start

`Graphics ⇒ Plot of X10 v X3`

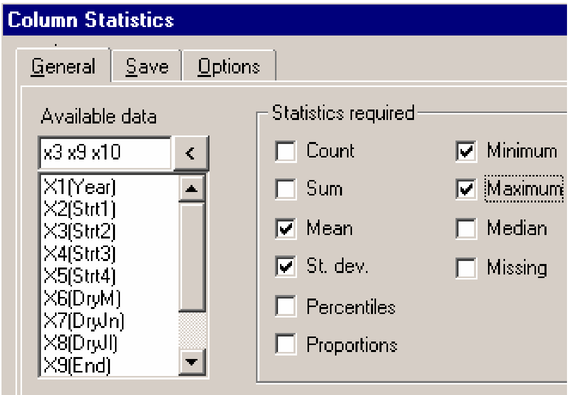


The results in Fig. 11.5j show that the length varied from less than 4 months to 7 months. In Fig. 11.5k, the length is plotted against the date of the start. This shows a clear negative trend and the correlation is -0.92.

This last correlation, between the date of the start and the length of the season, is not useful. It can also be harmful, because it may be used instead for analyses that are of value. We first explain why the correlation is not useful here and then suggest an alternative analysis.

Fig. 11.5k Statistics for start and end of rains and length of season

`Statistics ⇒ Summary ⇒ Column Statistics`



Column Statistics

Column	Mean	Min.	Max.	SDE
Strt2	133	96	176	19.32
End	292.9	275	315	8.87
Length	159.9	104	215	22.5

:

Fig. 11.5k shows the mean and standard deviation for each of the three columns, X3 (start), X9 (end) and X10 (length). There will obviously be a relationship between X3 and X10, because X10 has been calculated from X3 and X9, i.e. X10 = X9 - X3

We assume X3 and X9 to be independent (there was no evidence of a correlation between them, Fig. 11.5j.) With the assumption of independence, between X3 and X9, it is easy to calculate the correlation between X3 and X10, which is given by

$$\mathbf{-19.32 / \sqrt{(19.32^2 + 8.87^2)} = -0.91}$$

Thus, the negative correlation of –0.9is just a very complicated way of saying that the standard deviation of the start of the rains, which is just over 2 weeks, is approximately double that of the end.

If the correlation is not a useful way of summarising the relationship between the dates of the start and length of the season, then the relationship needs to be summarised in a different way. One possibility is as follows.

The general problem that farmers have in planning their cropping strategy is shown by the large variability in the length of the growing season from year to year. As an example, consider the situation when, in a given year, the farmer has planted successfully on 29th April, (i.e. day 120). What is then needed is the length of the season, **given** this planting date. The graph in Fig. 11.5k, indicates what to expectIt shows the season length is likely to be between 160 and 185 days, when the starting date was day 120.

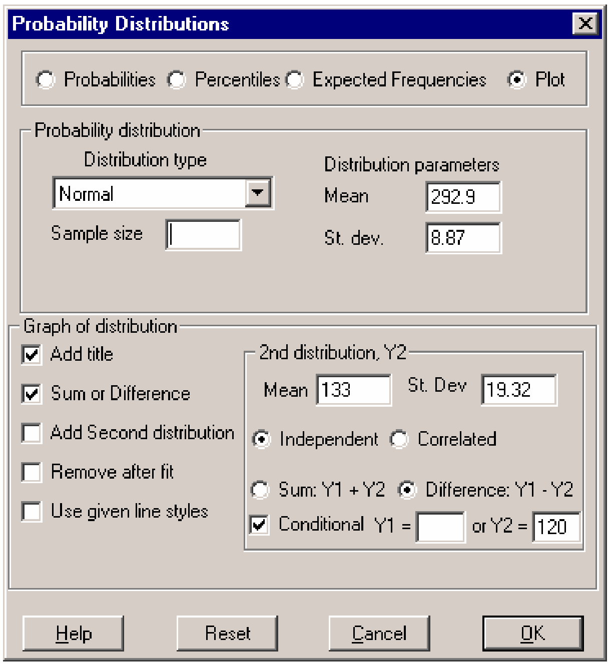
A conditional analysis is now needed , i.e. what is the distribution of the length of the season **given** the start is on a particular date. The only uncertainty now is of the date of the end of the season and the simplifying assumption is made that the start and end of the season are independent. This is possible, because there was no evidence of any relationship between X3 and X9. With this assumption, the **conditional length** of the season has a mean of (293-120)=173 days and standard deviation of 8.9 days.

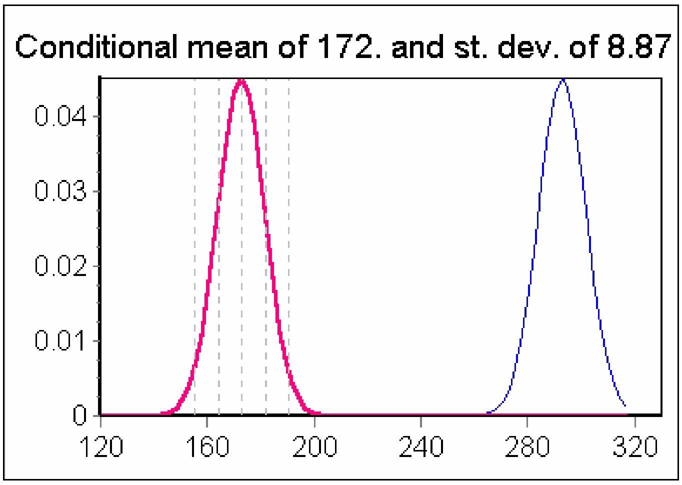
The analysis can now proceed in a distribution-free way or, as in Section 11.2, a distribution may be assumed. The distribution of the dates of the end of the season is approximately normally distributed. Adding this assumption, the properties of the normal distribution imply we can say that the farmer, planting on this date, has approximately a 2/3 chance of a season length of between 166 and 182 days and most years (95%) will have a season length of between 158 and 190 days. Hence, once the planting date is known, the uncertainty in season length is dramatically reduced and this can help in the choice of appropriate varieties, etc.

The idea of a 'conditional analysis' has been used at various points in this guide. It is important, so we conclude with a graph to reinforce the concept. Use the **Statistics ⇒ Probability Distribution** as shown in Fig. 11.5m.

Fig. 11.5m Conditional length given day 120 as the planting date

Statistics ⇒ Probability Distributions





A practical problem is how best to display this type of result, so it can easily be used. One possibility is shown in Fig. 11.5n for Niamey in Niger, taken from Sivakumar et al. (1993). This shows that if planting in Niamey can be in mid-May, then one could plan with fair confidence (88%) on a growing season length of more than 110 days, while if it is a month later, then one would hope for 90 days.

Fig. 11.5n Probabilities of growing season length at Niamey, Niger, exceeding specified durations for different dates of onset of the rains, (from Sivakumar et al. 1993)

Date of onset	Length of growing season (days) exceeding			
	90	*110*	*130*	*150*
19 May	99	88	49	11
29 May	97	70	29	2
8 June	88	49	11	0
18 June	70	29	2	0
28 June	49	11	0	0

11.5.4 Example 4: Simulated model of crop growth and yield

The final example shows another common use of correlations. Suppose a model of crop growth and yield has been constructed. To test this model, the crop is grown at a number of sites. At each site the observed yield is to be compared with the predicted yield from the model. The correlation between the observed and predicted yields is then sometimes used as a measure of the effectiveness of the model.

This evaluation process is simulated using a macro called **nomodel.ins**. It is called nomodel because the actual model is of no value, though it permits some seemingly impressive correlations to be found. In the evaluation of the model assume that the crop is grown at three types of site. The first is termed **desert**, and is on the limit of possibility for the crop. The model always predicts 0.05 tons/ha at such sites. Then there are **on-farm** sites, which typically have a yield of about 1 ton/ha. These are assumed to be the important sites and you will have to read further to see how these yields are modelled. Finally there are research institute sites. Here there are normally much higher yields than on farmers fields and the model always predicts 4 tons/ha.

Fig. 11.5o shows a run of the evaluation process. There is a high correlation between the **observed data** and the **predictions**. If correlations are used uncritically, then this model might be thought useful. However, the graph shows that this is clearly a misuse of correlations. With the anchoring points at the desert and research sites the fact that the model is useless at the important on-farm sites is effectively hidden.

If you study the macro you will see that at the on-farm sites the model simply chooses a random number from a normal distribution with mean 1 ton/ha and standard deviation of 1/4 ton/ha.

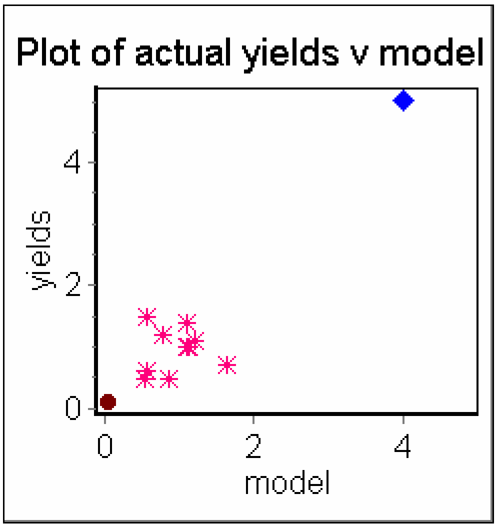
This type of misuse of correlations is common. The key to the misuse is that there is clear structure within the data. The **structure** here is that the data come from 3 distinct groups (desert, on-farm and research) and this structure has been totally ignored in the correlation analysis. This is a simple rule, namely whenever there is structure in the data, it is dangerous to ignore this structure in the analysis. In this case a much better analysis would be to consider the three groups separately. Then the fact that the model is useless at the on-farm sites would become clear.

Fig. 11.5o Test of crop model

Submit ⇒ Run Macro and select **nomodel.ins** from the **climatic library**

```
Input yields at desert site(s), default 0.1
(accept default)
Input yields at on-farm site(s), default 0.5 0
(accept default)
Input yields at research site(s), default 5
(accept default)
Row          X1       yields       model

  1       desert         0.1        0.05
  2      on_farm         0.5        0.86
  3      on_farm         0.6        0.56
  4      on_farm         1.0        1.12
  5      on_farm         0.5        0.56
  6      on_farm         1.1        1.22
  7      on_farm         0.7        1.63
  8      on_farm         1.0        1.11
  9      on_farm         1.2        0.78
 10      on_farm         1.4        1.11
 11      on_farm         1.5        0.56
 12     research         5.0        4.00


Correlation yields, model = 0.9157

Correlation > 0.8 - clearly good model,
                    worth publishing
```

Graphics ⇒ Plot ⇒ 'Yields' v 'model'



11.5.5 Conclusions

In summary, correlations are useful, but they are also overused and misused. The misuse is often either that a different summary of the data would be more informative (Section 11.5.3 on the length of the season), or that a single correlation is insufficient, because the data have structure, that needs more than one number to summarise (Section 11.5.4 on crop models).

11.6 Regression methods to study crop-weather relations

This section examines briefly the use of regression methods to construct models relating crop yield to weather variables. This is a large subject and here just two simple examples from the literature are used. This is followed by suggestions on the way a regression study to evaluate crop-weather relationships could proceed.

11.6.1 Example 1: Relationship between yield & weather variables

The first study is by Huda *et al.* (1975) on the relationships between rice yields and various weather variables. The 12 years of yield data are shown in Fig. 11.6a, together with a simple regression model, fitting the trend of yield against year number.

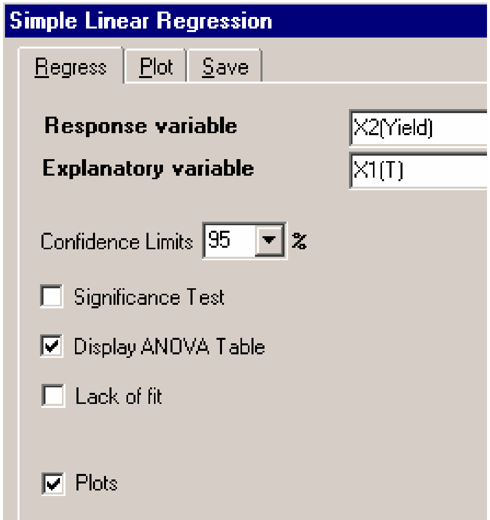
The article does not consider the trend alone, but looks, in turn, at the 5 types of climatic variable; rainfall, max and min temperatures and max and min relative humidity. The analysis is the same for each variable and is illustrated for the rainfall data. There are 24 weekly values each year, which correspond to the growing period, and these are reduced to three columns, which are as follows:

$Y0 = \Sigma y$		seasonal rainfall total
$Y1 = \Sigma iy$	(y1+2*y2+.... + 24*y24)	a weighted rainfall total
$Y2 = \Sigma i^2y$	(y1+4*y2+9*y3+... + 24²*y24)	another weighted total

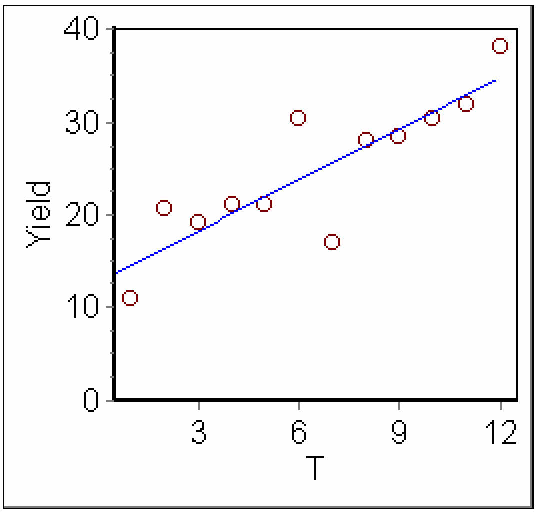
Fig. 11.6a Regression analysis of Huda data

File ⇒ Open Worksheet ⇒ Huda.wor

	X1*	X2*
	T	Yield
1	1	11.05
2	2	20.55
3	3	19.2
4	4	21.17
5	5	21.1
6	6	30.25
7	7	16.95
8	8	27.92
9	9	28.42
10	10	30.37
11	11	31.87
12	12	38.15
13		




```
---------- DETAILS OF THE FITTED LINE --------------

Fitted equation        : Yield = 12.9  +  1.824 * T
Standard error of slope : 0.3377  with 10 d.f.

95% confidence interval for slope  1.0712 to 2.5762

ANOVA for regression of Yield on T
---------------------------------------------------
Source       df        SS         MS     F value    Prob>F
---------------------------------------------------
Regression    1   475.604     475.6      29.16     0.0003
Residual     10   163.114     16.311
---------------------------------------------------
Total        11   638.718
---------------------------------------------------
R-squared = 0.7446
```



This phase of initial processing of the weather variables illustrates a general complication in studies of crop weather relationships, namely that there are many columns of possible independent variables (weather data) compared to the single column of yield data. Here there were 120 columns, consisting of 24 weekly values for each of the five elements. Here the initial processing consisted of the decision to look at the five climatic elements in turn and then reducing the 24 values to the three that are given above.

Then the regression equation given in the paper for rainfall is

yield =19.41 + 1.80T + 0.00332*Y0 - 0.000725*Y1 + 0.00000199*Y2

It is stated that 'the coefficient of determination (R^2) obtained was 0.7952 which was found to be significant at 2.5% level.'

However, the key point, omitted in the paper, is that the simpler equation in Fig. 11.6a of

yield=12.9 + 1.82T

has an R^2=0.74. Hence the important question is whether there is a case to consider the **more complicated equation**, given in the paper. To assess this, we construct the ANOVA table for the model given in the paper. The R^2=0.7952, means that 79.52% of the total sum of squares was explained by this model. The total sum of squares is 638.7, see Fig. 11.6b, hence the regression sum of squares is 507.9 and the residual sum of squares is the remaining 130.8. Further, the simple equation has a regression sum of squares of 475.6, so the extra sum of squares associated with the three rainfall variables is 32.3. These results are summarised in the ANOVA table given in Fig. 11.6b.

Fig. 11.6b Regression table from Huda *et al.* paper

```
Terms         d.f.      ssq         msq       F-ratio

Regression     4        507.9
  Trend        1        475.6      475.6      25.4
  Rainfall     3         32.3       10.8       0.6

Residual       7        130.8      18.68
------------------------------------------------------
Total         11        638.7
------------------------------------------------------
```

This result shows that the F value for the rainfall variables is even less than 1 and clearly non-significant. Hence there is no evidence of any relationship between the rainfall and the yield data. The other climatic variables have similar R^2. Thus the overall conclusion is that the data show **no evidence of any relationship between the yield and climatic variables**.

The conclusions in the paper are simply a result of poor use of regression methods - and presumably also of poor refereeing of the paper. A further disappointment is that the authors and also the journal were sufficiently impressed by the approach that they published a similar paper on maize in the following year. This type of misunderstanding, and misuse, of regression methods has given sceptics ample ammunition to discredit the approach.

11.6.2 Example 2: Rice yields in Brazil

The second study is again on rice yields, Mota and da Silva (1980) in a worksheet called Brazil.wor. The data are given in Fig. 11.6c.

Fig. 11.6c Analysis from Mota and da Silva paper

File ⇒ Open Worksheet ⇒ Brazil.wor

	X1*	X2*	X3*	X4*	X5*	X6*	X7*	X8*
	Year	Yield	Area	FebTmp	FebSun	MaxTmp	MinTmp	ObsNo
1	1957	1.9	95	13	235	27.3	15.6	1
2	1958	2.7	21	6	229	25.3	18.3	2
3	1959	2.9	100	5	234	26	17.3	3
4	1960	2.7	55	4	211	26.3	16.7	4
5	1962	2.9	57	14	211	26.3	17.3	5
6	1963	3.1	95	3	220	26.7	16.7	6
7	1964	2.3	7	4	216	25	17	7
8	1965	2.6	10	8	213	26.3	15.3	8
9	1966	3.1	90	10	229	25	17.7	9
10	1967	3.4	84	6	198	25.3	17	10
11	1968	3.5	86	7	227	26.7	17.3	11
12	1969	3.3	100	6	247	26.3	18	12
13	1970	4.1	100	3	216	25.7	17	13
14	1971	3.1	100	4	229	25.7	16	14
15	1972	3.8	100	4	191	27	16.3	15

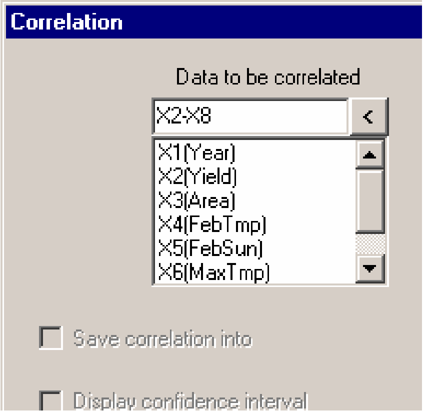
In Fig. 11.6c the variables are as follows:

X1 Harvest year
X2 Rice yield (t/ha)
X3 Percentage of area sown before 30 November
X4 Days with min temp below 15°C in February
X5 Mean number of sunshine hours in February and March
X6 Mean max temp November to January (°C)
X7 Mean min temp November to January (°C)
X8 Technological trend

The first step was to look at the correlations. There is a trend, as is seen by the large correlation of the yield against the observation number. Hence this is fitted first, as in the previous section.

Fig. 11.6d Correlations

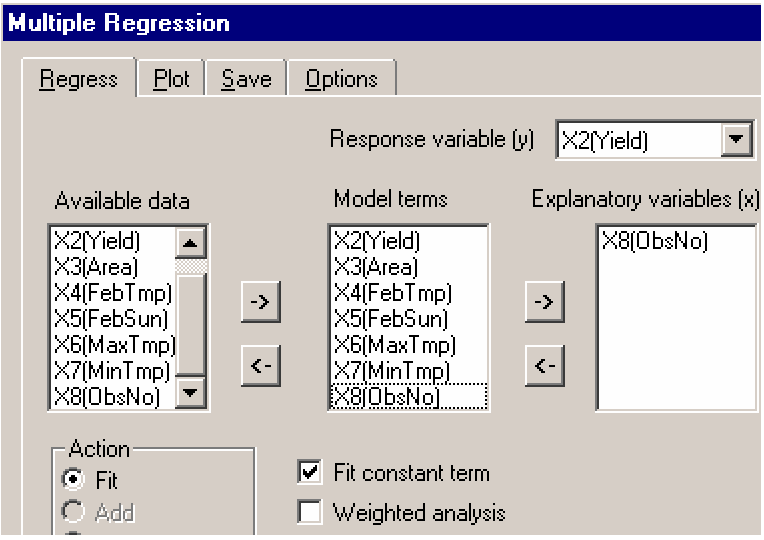
Statistics ⇒ Regression ⇒ Correlation




```
CORRELATIONS

Yield   1.0000
Area    0.5274   1.0000
FebTmp -0.4387  -0.0811  1.0000
FebSun -0.2994   0.1874  0.1745  1.0000
MaxTmp -0.0381   0.3474  0.2088 -0.0120  1.0000
MinTmp  0.2623   0.0202 -0.0595  0.3148 -0.4719  1.0000
ObsNo   0.7752   0.3859 -0.3992 -0.2583 -0.0626 -0.0744   1.0000
        Yield    Area    FebTmp  FebSun  MaxTmp   MinTmp  ObsNo
```

Fig. 11.6e Fitting the trend

Statistics ⇒ Regression ⇒ Multiple ⇒ Apply

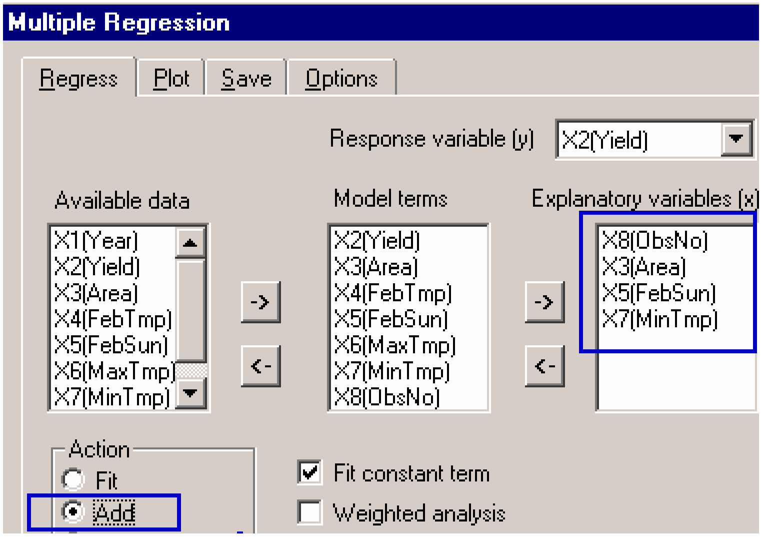


Multiple Regression

```
ANOVA for regression of Yield on ObsNo
----------------------------------------------------------------------
Source        df          SS            MS        F value      Prob>F
----------------------------------------------------------------------
Regression     1      2.66175        2.6618        19.58       0.0007
Residual      13      1.76758       0.13597
----------------------------------------------------------------------
Total         14      4.42933
----------------------------------------------------------------------
R-squared = 0.6009  (adjusted = 0.5702)
```

Now the climatic variables are added. After some experimentation the variables X3, X5 and X7 are added, as shown in Fig. 11.6f. This is the model that is fitted in the paper. It is seen that the effect of adding the three variables, X3, X5 and X7 does improve the fit of the model.

Fig. 11.6f Adding climatic variables



Multiple Regression

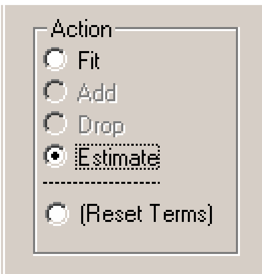
ADD 'Area' 'FebSun' 'MinTmp'

Change	df	SS	MS	F value	Prob>F
Original	1	2.66175	2.6618	39.55	0.0001
Added	**3**	**1.09461**	**0.36487**	**5.42**	**0.0179**
Residual	10	0.672976	0.0673		
Total	14	4.42933			

The parameters of the fitted model are then estimated, as shown in Fig. 11.6g. This gives the same fitted equation as in the paper, i.e.

yield = 0.3274 + 0.0059*X3 - 0.0133*X5 + 0.2733*X7 + 0.0725*X8

Fig. 11.6g Estimating the parameters of the fitted model



REGRESSION COEFFICIENTS
Y-variate: Yield

Param.	Estimate	SE	t	Prob>|t|	95% CI	
Const	0.32737	1.602	0.20	0.8422	-3.242	3.897
Area	0.00587	0.0023	2.56	0.0284	0.0008	0.011
FebSun	-0.01329	0.0054	-2.45	0.0344	-0.0254	-0.0012
MinTmp	0.2733	0.0873	3.13	0.0107	0.0787	0.4679
ObsNo	0.0725	0.0181	4.02	0.0025	0.0323	0.1127

This is a better study, within which there is just one major point of concern. This is that the variables given in Fig. 11.6c appear to be ones that achieved ‘statistical significance’ in an initial phase of the analysis. There is no indication in the paper on how many variables were included initially, but this implies an element of ‘data dredging’. The idea of ‘data dredging’ here is that with just 15 values of the yield, if enough potential independent variables are used, then by chance the ‘best’ are likely to fit well with the yields.

The effects of this problem can be assessed by checking the final model against new years of data, that were not included in the fitting. In the paper, this was done by testing the model against the observed data for 4 further years, 1973 to 1976. The concept is good, but the extra record is very short.

11.6.3 Conclusions and suggestions

Researchers sometimes voice scepticism about the potential of regression methods in studies of crop-weather relations. Some caution is valid, because it is easier to find poorly executed studies than good ones. Our view is that regression methods are useful tools in the study of crop-weather relations if (and only if) they are well used. It is difficult and time-consuming to conduct a good regression study and some suggestions are given below.

An alternative method for studying crop-weather relations is to use weather data as one type of input to a physiologically based crop growth model. This process-based modelling is sometimes viewed as being in competition with the construction of a more empirically based model. We view the two approaches as being complementary, with each being able to support the other. The process-based models are usually based on a model for individual plants, rather than large areas. Hence, when yield data are available for an area, it is useful to test whether the same variables that drive the process-based model, have an equally clear effect on these larger scale yields. Thus the process-based model can indicate the type of climatic variables that are likely to be related to the yields. An example is given later.

We now consider briefly the necessary components for an effective use of regression methods to study crop weather relations. A major problem in many studies is the lack of sufficient yield data. Unless there is a long record, data from a single site will not be sufficient. One initial step is therefore the organisation and management of the yield and ancillary data. The variables that are normally reported in a yield trial will often include:

date of sowing	soil type	total dry matter	insect, disease or
date of flowering	soil Ph	head weight	pest damage
date of harvesting	level of Phosphorus	grain weight	
	level of Nitrogen		

Sometimes variables are omitted because they are not of direct concern to a particular study. This is incorrect, because all major variables, such as insect damage, that are related to yields, need to be included if we are to unravel the contribution due to climatic variables.

The value of any regression study should be questioned, if there is so little yield data that this step of data management for the yields, is not needed. The study by Huda et al, (1975) described in Section 11.1, falls clearly into this category.

The second component of the initial phase is the management and initial processing of the climatic data. There will still be an imbalance in the volumes of the climatic and yield data and hence some method must be found to reduce the number of potential climatic variables. This guide has given a wide variety of ways in which the climatic data can be processed. For example, from the initial daily data, the following variables could be calculated:

a) total rainfall between sowing and flowering
b) length of the longest dry spell in the 30 days after sowing
c) minimum value of the water balance in the 20 day period after flowering
d) maximum humidity in the 15 days before harvest
e) maximum of max temperature in the 15 days before flowering
f) final value of the FAO crop index
g) maximum wind speed in the 15 days after sowing
h) number of heat units (degree days) in the growing period

Knowledge of the pattern of growth of the crop can be used to indicate the type of variables that are likely to be related to the yield.

Growth and physiologically based yield models now also exist for most crops. They require climatic and other inputs. The way these models use climatic data can also be a useful indication of the ways in which climatic variables could be transformed to be of use. For example, if a model is very sensitive to temperature data, then it may be that the use of heat units in a regression model would be useful. If the process-based model shows yield to be closely related to soil Ph, then a regression study with yields from different sites, that did not have access to this information, would probably be of little value.

The regression study begins after the initial data management phase. From the yield data, there may be various dependent variables that can be used. For example the head weight may give clearer results than grain weight, because it is less affected by pest damage. Hartmond, Williams and Lenz (1996) propose a transformation of the yield variables into crop growth rate and partitioning. They show that partitioning is primarily genetically controlled, while the crop growth rate depends primarily on environmental variables. Hence the crop growth rate could be a useful dependent variable.

When pest damage has been recorded, then this could also be a useful dependent variable in an ancillary study. Pests and diseases may be related to date of planting and relative humidity, etc.

We now turn to the independent variables, particularly the climatic, variables. Here knowledge of the results from research will sometimes suggest that some important independent variables may not, however, be linearly related to the yield. For example, results from Payne (1997) indicate that millet yields in the Sahel are more related to lack of nutrients than lack of water, in all but the driest years. If this is the case, then seasonal rainfall of 300mm at a particular site might give a higher yield than 200mm, but above 350mm there is no relation. Including rainfall in the regression might then not indicate a relationship, while transforming the rainfall data (for example, make all values above 350mm into 350mm) could clarify the situation.

The importance of including the structure of the data in the model has been mentioned in Sections 11.4.4, 11.5.4 and elsewhere, and regression studies are no exception. Here the structure of the yields usually includes the fact that the data are from different sites. Ideally the site factor would not remain in the final equation, but would be replaced by real explanatory variables, such as soil fertility and rainfall.

Once the possible dependent variable has been selected, together with a set of potential independent variables, different regressions can be fitted. The regression package that is used for the fitting, should be able to handle a mixture of factors (like site) and variates (like annual rainfall). Instat has reasonable facilities for this, but some other packages (Minitab, SAS, Genstat) are better.

Many packages have facilities for automatic stepwise regression or best subset regression. These can be useful, but it is important that scientists, not the computer software, remain in control. Often the problem of choosing the best subset of the independent variables is less important than that of checking that the set of independent variables includes all that are important. For example, if temperature is important, then check whether the effect is linear, or whether a second variable, such as sqrt(temp) or $temp^2$, would also be useful.

Care must also be taken that the regression equation reflects the data as a whole and not simply isolated points. The problems in Section 11.5 with the spatial correlations, Fig. 11.5b and with @nomodel in Fig. 11.5n apply equally here. The possibility of such problems can usually be checked using the many regression diagnostics that have been developed, see the Instat Introductory Guide, **Section 17.13**.

Finally we consider the issue of which software package to recommend for this type of analysis. What is needed is a package that is easy to use, and has excellent facilities for data management, for processing climatic data and for regression studies. For teaching purposes we are happy to recommend Instat, but it is not powerful enough for some of the large studies that are needed. If we had to recommend currently, then the package SAS is perhaps the best for data management, Instat is still the only package with special facilities for climatic analyses and Genstat would be a good package for regression studies.

Perhaps the lack of a single appropriate package is one reason that it is difficult to find the exemplary case studies that we seek. However, with Windows based software, it is easy for students to be use more than one package if this is needed. So perhaps soon the regression approach to the study of crop-weather relations might redeem its good name.

Chapter 12 – Case Studies

Three case studies are considered that use and extend the features of Instat that have been introduced in the previous chapters.

The first example is based on a real example of supplying water for irrigation needs. It reduces to one of estimating rainfall probabilities and percentage points through the year. This is ostensibly simple, but is used to compare five different methods. They range from a simple summary of observed data to the use of gamma distributions with smoothed parameters. The results show the value of the more complex analyses, and Instat dialogues make the fitting process into a simple routine.

The second case study is that of modelling a simple reservoir and irrigation scheme. This is used to consider the problem of the appropriate area to irrigate given the vagaries of rainfall. It uses an extended version of the crop performance index, introduced in **Chapter 10**, and illustrates many of the steps involved in stochastic modelling.

The third example is that of analysing some storm event data. These are available for 50 storms on a 5 minute basis. This case study illustrates some ways in which these types of data may be analysed. It also shows the care that must be taken in such an analysis and introduces the analysis of 'circular data', because one variable is the time of day at which storms begin.

12.1 Case Study 1 - Estimating probabilities and percentage points

In Sri Lanka rice is often grown using irrigation to supplement the rainfall. The irrigation schemes vary greatly in size. Many use reservoirs or tanks of great antiquity.

At one tank near Kurunegala the management scheme was as follows:

At the start of a week the water requirement of the crop for that week was estimated. The water currently in the field was measured and hence the amount to be added was found.

If this entire amount were released from the tank to the field, then the water requirement of the crop would always be fully met. However, if it rained during the week, the rainwater might be lost; the water level in the field would become too high and water would flow over the bunds to drainage. Clearly it would be more efficient to release less than the crop requirement, from the reservoir, allowing it to be made up by rainfall. This would also help conserve water in the tank, during the rainy season, for use during the subsequent dry season.

Of course, at the beginning of each week, it is not known how much rain would fall during the week. However, rainfall records could be examined to estimate the amount to 'expect' to get. For example if every year of the last fifty had more than 20mm rain during the corresponding week, it would be sensible to allow for at least 20mm of rain when calculating how much to release. We could actually allow for rather more than 20mm, say 40mm. If less than 40mm fell, the crop water requirement would not have been met, but the stress would be slight and the deficit could be corrected the following week.

Some work, using a simple model to describe the water in the fields and the effect of water shortage, arrived at the suggestion that the amount to allow for should be the 30% point of the distribution of total rainfall for that week. Hence, the problem that is considered here is to estimate the 30% point of the total rainfall for each week.

The data available are daily rainfall records for 34 years, 1950-1983. These are in the worksheet Kurunega.wor, i.e. X1 - 1950, X2 - 1951 ... X34 - 1983. Weekly totals have already been calculated, using the methods described in Chapter 5, and stored in the worksheet Kurun7.wor. These data are stored with one column for each week of the year, i.e. X1 gives the rainfall totals from 1 to 7 January, X2 - 8 to 14 January .... X52 - 24 to 31 December. The 34 observations in each column correspond to the 34 years of data.

12.1.2 Analyses

Five different methods of estimating the 30% point of the data through the year are considered. The first method is the simplest and merely involves calculating the observed 30% point from the data, as described for monthly and decade data in Chapter 7. This analysis was originally proposed by the consultants and criticised by the customer. Alternative (improved?) methods, which are described here, included 'smoothing' the original estimates and/or fitting gamma distributions as described in **Chapter 11**.

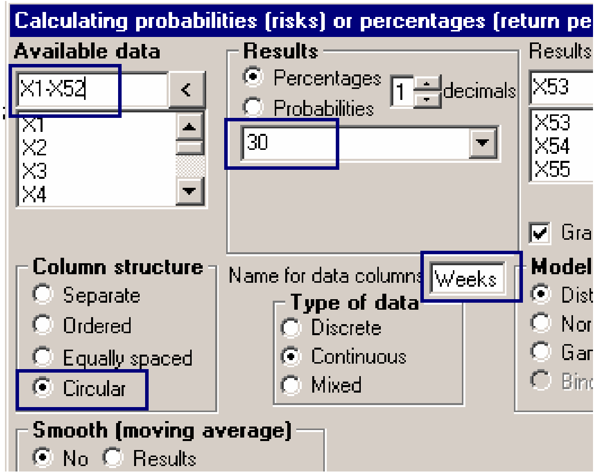
The five methods are all available using the **Climatic ⇒ Process** dialogue and we consider them in turn.

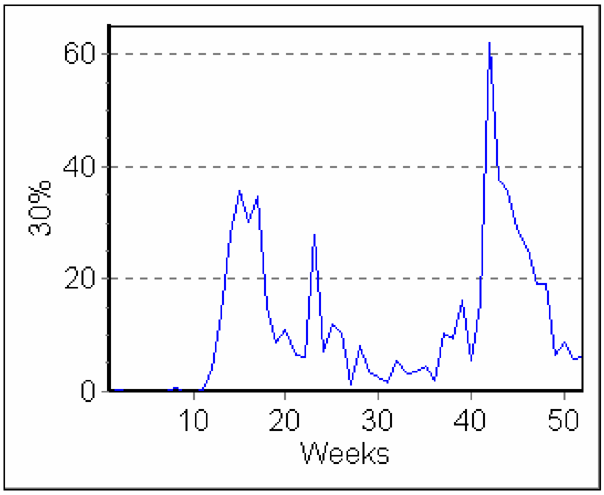
Method 1 - The simplest estimate of the 30% point is just the observed value. Start by opening the worksheet Kurun7.wor. In the **Climatic ⇒ Process** dialogue the data columns are X1-X52 and this set of columns is for the whole year, so use the circular option, as shown in Fig. 12.1a.

Fig. 12.1a Plot of 30% points for 7 day rainfall totals

File ⇒ Open From Library ⇒ Kurun7

Climatic ⇒ Process





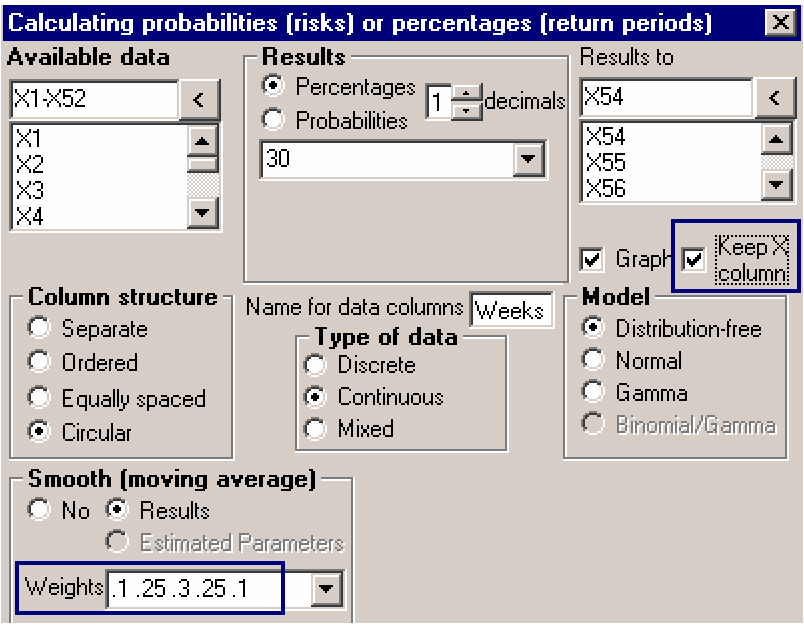
The plot shows that the main (Maha) wet season starts around week 38 and continues to week 50, with the small (Yaha) wet season from week 12 to week 20 or 25.

Two problems arise. The first is that the 'estimates' are jagged, (i.e. not smooth) through the year. This is not likely to be a 'real' feature of the climate, but due to limited data and the estimation procedure. For example the reservoir managers did not feel it was sensible to allow for 16mm in week 39, but only 5 mm in week 40. Similarly, it seems unlikely that the real value in week 42 (of 62mm) is 4 times the value in week 41 and twice that in week 43 - but that is what the estimates indicate.

The second problem with these estimates of the 30% point is that they are not very precise. The raggedness is a result of this. As an example, consider week 45, where the estimate is 28.9mm. An approximate 95% confidence interval, calculated by the methods shown in Chapter 11, is 17mm to 60mm, which is rather wide. (The result follows from the use of @quantile x45 30, **Section 11.2.3**.)

Method 2 involves smoothing the estimates from Method 1. There are various ways of smoothing data. In the **Climatic ⇒ Process** dialogue a simple smoother is used, taking five point moving averages. Because the columns are declared “circular”, the estimates at the end of the year (December) use the values for the first weeks of January to make the moving average.

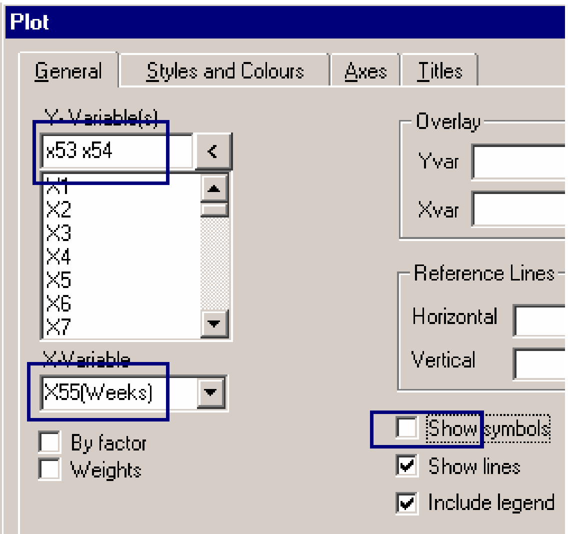
Fig. 12.1b Calculate moving averages

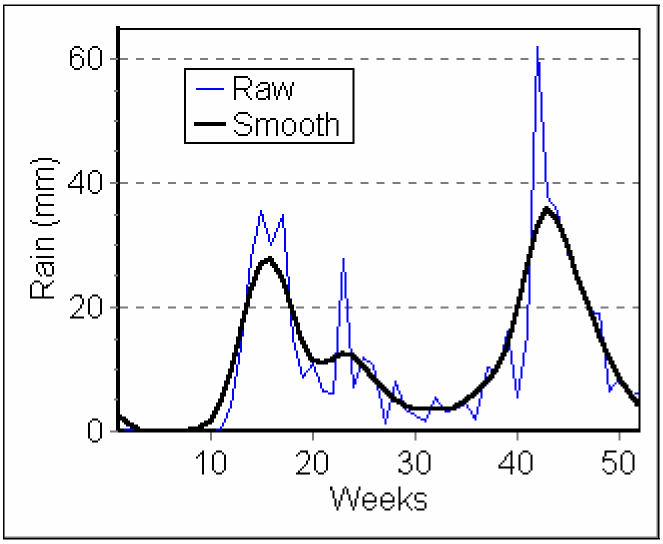


Use **Graphics ⇒ Plot** to plot the raw and smoothed 30% points as shown in Fig. 12.1c. The results from the smoothed estimates seem more reasonable than those from Method 1.

Fig. 12.1c Raw and smoothed 30% Points

Graphics ⇒ Plot





When using moving averages, the choice of 5 terms, rather than 3, and also the weights are relatively arbitrary. Some alternatives are:

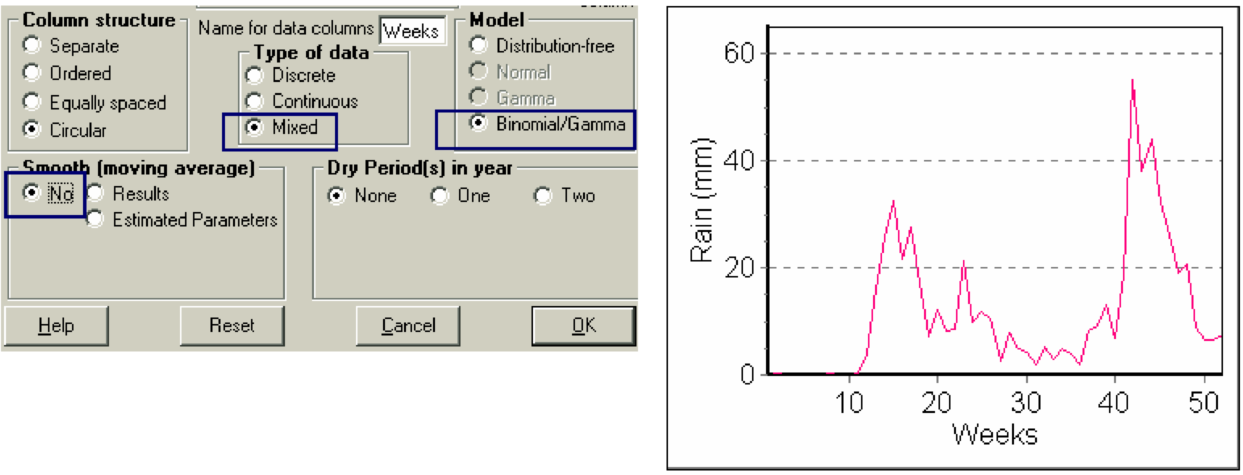
3 point	0.333	0.334	0.333		
	0.25	0.5	0.25		
5 point	0.2	0.2	0.2	0.2	0.2
	0.111	0.222	0.334	0.222	0.111
	0.1	0.2	0.4	0.2	0.1

For weekly data, we often find that the last two alternatives, the 5 point moving averages, with the most weight for the central points, gives what look like more sensible results. With a more powerful statistics package, other forms of smoothing, such as spline fitting may be preferable to the use of moving averages.

Calculating the precision of the estimates of the 30% from Method 2 depends on which moving average is used and assumptions of the extent of dependence between successive weeks. However, even conservatively, it is clear that the precision is more than doubled. This simple smoothing is therefore equivalent to using Method 1 on a record that is more than twice as long.

Method 3 involves fitting a distribution to the weekly data and then estimating the 30% point of that distribution. In the last chapter, Section 11.3, used the gamma distribution (plus a proportion of zero observations) as a model. This is called "Mixed" as the type of data in the **Climatic ⇒ Process** dialogue and uses the Binomial/Gamma model. There is no totally dry part of the year so keep the **Dry Period(s) in year** option as **None**. The dialogue and plot of the gamma model are in Fig. 12.1d.

Fig. 12.1d Gamma model with no smoothing

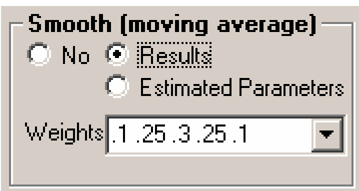


Throughout the year the two estimates from the gamma model are similar to Method 1. This suggests the model is sensible, but of course, the assumption of a gamma model should be checked, as described in Chapter 11.

Calculating confidence intervals for the estimates from the gamma model is hard, particularly when the chance that the whole period is dry is not small. However, approximate confidence intervals have been calculated and are ***slightly narrower*** (estimates more precise).

Fig. 12.1e Methods 4 and 5

Method 4



Method 5

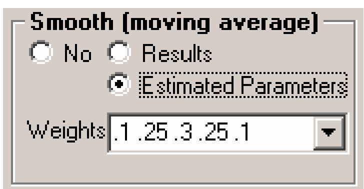
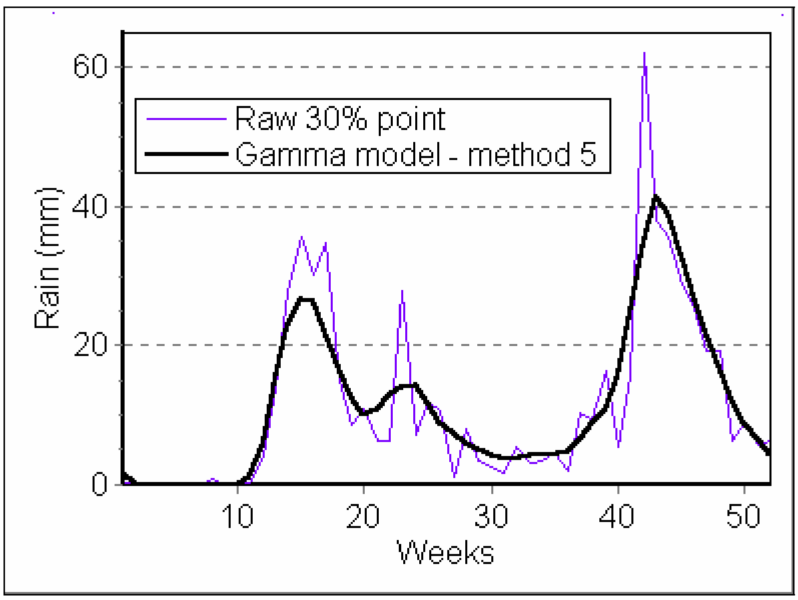


Fig. 12.1f Comparing raw 30% points and Method 5



Method 4 takes the estimates of the 30% points from the gamma model (Method 3) and smooths them using simple moving averages as shown in Fig. 12.1e.

However, given that smoothing is useful, an alternative is to smooth the estimates of the gamma parameters, rather than the final result. The 30% points are then estimated from the smoothed gamma parameters (**Method 5**). The raw and smoothed 30% points are plotted in Fig. 12.1f.

12.1.3 Choice of method

The five methods are:

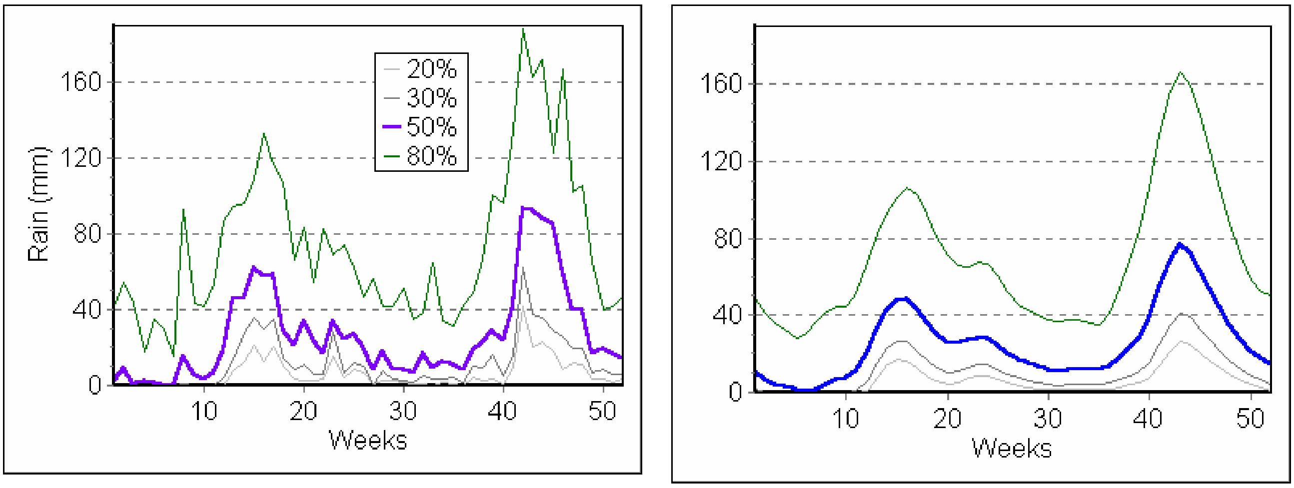
Method 1	Empirical 30% point
Method 2	Smoothed values from Method 1
Method 3	Estimates from separate gamma distributions
Method 4	Smoothed 30% points from Method 3
Method 5	Smoothed model parameters in Method 3

The precision of the estimates increases from Method 1 to Method 2 and Methods 4 and 5 are more precise still. However, the assumptions also increase. The preferred method is that which is most precise, yet does not rely on any unrealistic assumptions. Here, we believe this is Method 5, though the extra precision may be small, compared to Method 2.

This case study, has concentrated on the estimations of a particular percentage point. Fig. 12.1g shows the results using Methods 1 and 5 for the 20%, 30%, 50% and 80% points.

Fig. 12.1g 20%, 30%, 50% and 80%

Method 1 **Method 5**



A similar problem is to estimate probabilities (or risks) through the year. For example, what is the probability of more than 20mm in each week (or decade) of the year. Here again Method 5 will often be suitable and is an option of the **Climatic ⇒ Process** dialogue. The results, together with a brief description of the method, are in various reports, for example Sivakumar et al. (1993), and an example is in Fig. 11.3x.

Finally, a sixth method is more spectacular, but could become standard in the future. If smoothing is valuable in increasing precision, then perhaps one should smooth as early as possible in the analysis. Method 5 consists of the following steps:

(a) The daily data are totalled on a weekly basis
(b) Gamma distributions are fitted to the weekly totals
(c) The gamma estimates are smoothed
(d) The 30% point is calculated from the smoothed gamma

It is feasible to smooth even earlier, namely at (a) with the daily data, rather than (c). Various papers, for example Stern *et al.* (1982) and Stern and Coe (1984) explore these methods and this case study was based on a University of Reading MSc project, Sooriyarachchi (1989), which found that the sixth method doubled precision yet again.

This sixth method is considered further in **Chapter 13**. The importance of the results concerning the precision is as follows: Consider a station with a 60-year record, and suppose that Method 1 (which is standard in many publications) is used. Now Method 2, and hence certainly Method 5, is at least twice as precise, and Method 6 is twice as precise as Method 5. Hence, we should have at least the same precision from less than 15 years of data, using Method 6, as from 60 years using Method 1.

This should provide hope for the many stations in Africa and elsewhere, where short records are common. Even when long records exist, the extra precision from Methods 2 and 5, and particularly from Method 6, permit just the recent years to be used for the analysis. This is useful if there is thought to be a climatic change during the period that corresponds to the full record.

12.2 Case Study 2 - Modelling a simple reservoir & irrigation scheme

When managing an irrigation system, there are many decisions to be taken to make most efficient use of water. This case study considers a simplified scheme to show how one of these decisions might be taken.

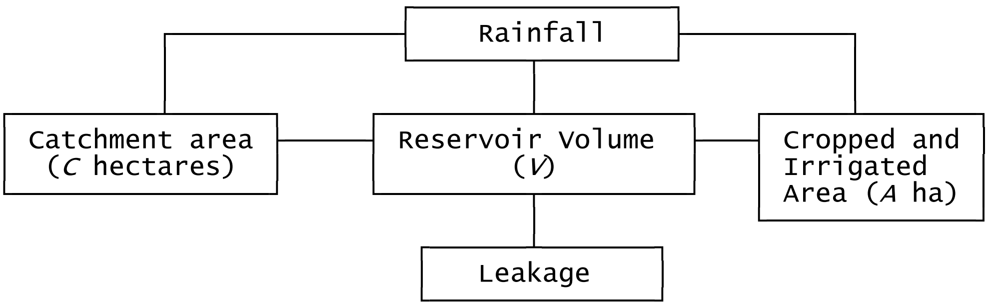
The problem is an obvious one. A reservoir contains water at the beginning of the growing season. This water can be used to irrigate a large area. However, if the whole area is planted and irrigated, we might run out of water before the end of the season. Then the crop will fail. Alternatively just a small area could be planted and irrigated. There would then be sufficient water, and hence a crop to harvest, but the yield would be limited by the small area. The question is "What area should be planted and irrigated?".

If the reservoir was the only source of water for the crop the question could be answered quite easily. However, it rains during the growing season, and this rain falls on the crop and refills the reservoir. The decision on the area to plant has to be taken at the start of the season, but the total rainfall is not known until the end of the season. The answer to the question must therefore involve probabilities, risks, or averages of the rainfall.

12.2.2 A simple reservoir

The system consists of a Catchment area, *C*. Assume that all the rain falling on the catchment runs into the reservoir. The Volume (maximum water content) is *V*m^3. Water is lost from the reservoir by:

- Leakage and evaporation. This is a proportion, p, of reservoir contents, per period.
- Irrigation. Water is used to irrigate an area A



Notes: 1 hectare = 10,000 m² 1mm rain on 1 hectare = 10m³

We first show that the calculation can be done by hand, and then show the use of Instat. In the following examples, all accounting of water inputs and outputs is done on a 10-day basis. The initial example models a situation as follows:

Catchment, *C*	3 hectares
Cropped area, *A*	1 hectare
Volume, *V*	5000m^3
Leakage rate, *r*	0 (no leakage or evaporation)
Crop water requirement	50mm per 10 day period

This initial example just considers the reservoir. Assume it is 4/5th full at the start. Table 12.1 shows the formulae used and the contents of the reservoir for 6 successive periods. The rainfall was 0, 20mm, 60mm, 0, 10mm, 0 in these 6 periods. The budget show that the irrigation need was satisfied and the reservoir had 3900m^3 at the end of the period.

Table 12.1 Irrigation budget, when 1 ha is irrigated

Period	10 days	t	**1**	**2**	**3**	**4**	**5**	**6**
Previous contents	m^3	x_{t-1}	[(a)] 4000	3500	3800	5000	4500	4400
Leakage rate	m^3	$L_t=rx_{t-1}$	0	0	0	0	0	0
Rainfall	mm	P_t	0	20	60	0	10	0
Rainfall volume	m^3	$I_t=10CP_t$	0	600	1800	0	300	0
Total available (5000 max)	m^3	$T_t=x_{t-1}+I_t-L_t$	4000	4100	5000	5000	4800	4400
Water required	mm	W_t	50	50	50	50	50	50
Irrigation required	mm	W_t-P_t	50	30	[(b)] 0	50	40	50
Irrigation volume required	m^3	$R_t=10A(W_t-P_t)$	500	300	0	500	400	500
Subsequent contents	m^3	$x_t=T_t-R_t$	3500	3800	[(c)] 5000	4500	4400	3900

Notes:

(a) Initially the reservoir is 4/5th full
(b) Period 3: W_t-P_t=-10, but negative irrigation not possible, set to 0
(c) Period 3: Final contents, x_t – the maximum reservoir size is 5000

These results could be compared with Table 12.2, where A=5, i.e. 5 hectares are cropped. Here only the rainfall value and the equivalent last two lines of Table 12.1 are given.

Table 12.2 Calculations as for Table 12.1, but with 5 hectares irrigated

Period	10 days	t	**1**	**2**	**3**	**4**	**5**	**6**
Rainfall volume	m^3	$I_t=10CP_t$	0	600	1800	0	300	0
Irrigation volume required	m^3	$R_t=10A(W_t-P_t)$	2500	1500	0	2500	2000	2500
Subsequent contents	m^3	$x_t=T_t-R_t$	1500	600	2400	0	0	0

In Table 12.2 the reservoir is empty in the fourth period and the crop therefore received no water from the reservoir after this date.

So far the model describes the reservoir, but not the crop. The crop water requirement needs to be more realistic than just 50mm per decade and the model must also describe what happens to the crop when the water requirement cannot be satisfied by rainfall or irrigation. To illustrate one way that includes the crop, the FAO model (described in Chapter 10) is used to determine the crop water requirements: the resulting index summarises the state of the crop at each stage.

Table 12.3 gives an example and shows the calculations as they can be done by hand, or with a spreadsheet. The same rainfall data are used. Assume that PEt is 50mm per decade and use the crop coefficients for an 100 day crop to determine the water requirements. Assume, as before, C=3ha, irrigation area A=1ha, maximum volume V=5000m^3 and no leakage, i.e. r=0.

The example starts with an initial volume x_0=500m^3. The budget in Table 12.3 proceeds roughly as before.

- In period 1 the water requirement by the crop is 50*0.3=15mm. This gives a total water volume requirement of 150m^3, which can be supplied from the reservoir, leaving 350m^3 at the start of the next decade. The crop requirement is satisfied, so the index remains at 100.
- In the next decade, 50*0.7=35mm is required by the crop. The rainfall provides 20mm leaving 15mm, or a total of 150mm to be met by the reservoir. In this period the rainfall has also replenished the reservoir by 600m^3, leaving 800m^3 at the start of the third decade.

Table 12.3 Reservoir calculations and crop water satisfaction index

Reservoir calculations

Period	1	2	3	4	5	6	7	8	9	10
Previous volume	500	350	800	2600	2200	2000	1400	800	300	0
Leakage	0	0	0	0	0	0	0	0	0	0
Rainfall (mm)	0	20	60	0	10	0	0	0	0	0
Rain volume	0	600	1800	0	300	0	0	0	0	0
Total	500	950	2600	2600	2500	2000	1400	800	300	0
Water required	15	35	50	50	60	60	60	50	45	30
Irrigation required	15	15	0	40	50	60	60	50	45	30
Irrig. volume required	150	150	0	400	500	600	600	500	450	300
Reservoir contents	**350**	**800**	**2600**	**2200**	**2000**	**1400**	**800**	**300**	**0**	**0**

Index calculations

Period	1	2	3	4	5	6	7	8	9	10
Rainfall	0	20	60	0	10	0	0	0	0	0
PEt	50	50	50	50	50	50	50	50	50	50
Crop Coefficient	0.3	0.7	1.0	1.0	1.2	1.2	1.2	1.0	0.9	0.6
Water required	15	35	50	50	60	60	60	50	45	30
Surplus or deficit	-15	-15	+10	-50	-50	-60	-60	-50	-45	-30
Irrigation	15	15	0	50	50	60	60	50	30	0
Final deficit	0	0	0	0	0	0	0	0	-15	-30
Index	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**97**	**90**

Total water required = 455mm

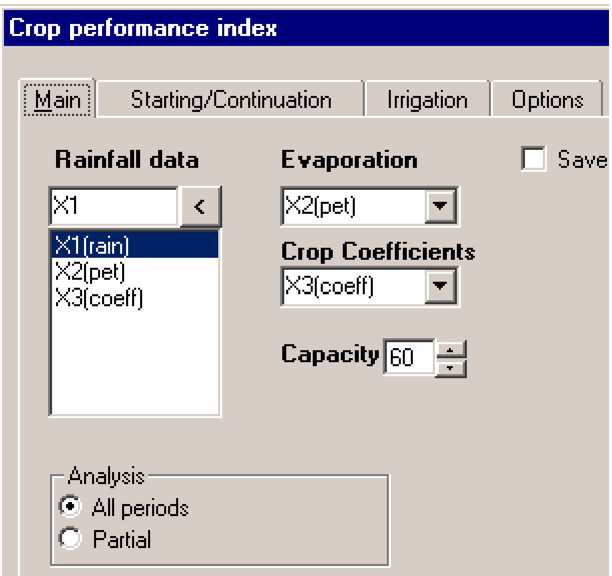
- The third decade has 60mm of rainfall. This is more than the crop needs, hence no irrigation is used and 10mm remains in the soil for the next period. The rainfall has also helped the reservoir to attain 2600m^3.
- In the fourth decade there is no rain. The crop needs 50mm, of which 10mm is given from the soil, leaving 40mm to be met from the reservoir.

Fig. 12.2a Data for reservoir example

	X1	X2	X3
	rain	pet	coeff
1	0	50	0.3
2	20	50	0.7
3	60	50	1
4	0	50	1
5	10	50	1.2
6	0	50	1.2
7	0	50	1.2
8	0	50	1
9	0	50	0.9
10	0	50	0.6
11			

Fig. 12.2b Crop performance index dialogue

`Climatic ⇒ Crop ⇒ Water Satisfaction Index`



The calculations then proceed to the penultimate period when the crop lacks 15mm of the water that it requires. This is about 3% of its total requirement, hence the index drops by the 3% to 97. In the final decade it drops a further 7% to its final value of 90.

This example can also be done in Instat, because the crop performance dialogue, used in Chapter 10, contains a reservoir option. To analyse the data in Instat use **File ⇒ New Worksheet**. Enter the data shown in Fig. 12.2a and name the sheet **Reserve.wor**. Or use this worksheet from the Instat library.

Click on the Irrigation tab in the dialogue, to give the screen shown in Fig. 12.2c. Change values as necessary to correspond to those in this figure.

Click on the Starting/Continuation tab, to give the screen shown in Fig. 12.2d. Choose a fixed starting period, and specify that it is period 1.

Fig. 12.2c Irrigation tab on dialogue

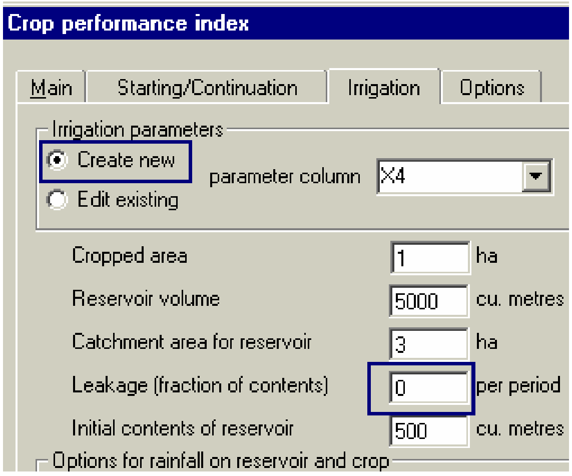
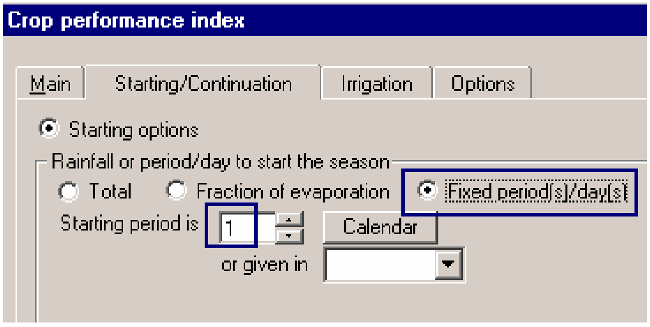


Fig. 12.2d Starting tab on dialogue



When this is run, the results are as shown in Fig. 12.2e for the reservoir and 12-2f for the crop. They are the same as shown earlier in Table 12.3.

Fig. 12.2e Results for the reservoir

```
Crop performance index

Analysis for rainfall data in rain
Catchment area      = 3          km. sq
Reservoir capacity = 5000        cu. m.
Initial vol.        = 500        cu. m.
Area irrigated      = 1          km. sq
Rate of loss        = 0.000 per period

Period  Rainfall  Reservoir (cu. m.)
          (mm.)     Needed Remaining
  1        0        150        349
  2        20       150        799
  3        60       0          2600
  4        0        400        2200
  5        10       500        2000
  6        0        600        1399
  7        0        600        799
  8        0        500        299
  9        0        450        0
  10       0        300        0
```

Fig. 12.2f Results for the crop index

```
Total water requirement = 455mm.
Period      WR       Ra     Sur/Def   Index
Date Crop   mm       mm       mm
1     1    15.0     0.0      0.0      100
2     1    35.0     0.0      0.0      100
3     2    50.0     10.0     0.0      100
4     3    50.0     0.0      0.0      100
5     4    60.0     0.0      0.0      100
6     5    60.0     0.0      0.0      100
7     6    60.0     0.0      0.0      100
8     7    50.0     0.0      0.0      100
9     8    45.0     0.0     -15.0     97
10    9    30.0     0.0     -30.0     90
:
```

Any of the options from the crop dialogue can be used. Hence the soil capacity (etc.) may be altered. Similarly, the effect of changing the cropped area, *A*, the catchment area, *C*, or the initial contents of the reservoir may be investigated by altering the corresponding values in the dialogue.

12.2.3 Application to Madawachchiya, Sri Lanka

Madawachchiya is in the northern region of Sri Lanka, which receives two wet seasons each year. The main wet season from October to December is used to grow a rainfed crop of rice. The rains during the second wet season (March to May) are too unreliable to allow a second crop, unless supplemented by irrigation. This area of Sri Lanka has many reservoirs (or tanks) which are used to supply irrigation water. The managers of every system have to decide what area can be irrigated.

Consider a scheme as follows:

Cropped Area, *A*	1 ha
Reservoir Volume, *V*	200,000 m^3
Catchment area, *C*	200 ha
Leakage coefficient, λ	0.05

It is assumed the reservoir is always full at the start of the second wet season and that planting takes place in the first decade of February (period 4 of the year). Rainfall data are available for 85 years from 1891 to 1978 (excluding 1914 and 1921). This has been summarised into 10-day totals. The Instat worksheet Madaw.wor has a column for each year.

As an example, when 80ha are irrigated, the values for analysing the first year are as shown in Fig. 12.2g.

Fig. 12.2g Crop analysis for 1891 from Madawachchiya

File ⇒ Open From Library ⇒ Madaw.wor
Climatic ⇒ Crop ⇒ Water Satisfaction Index

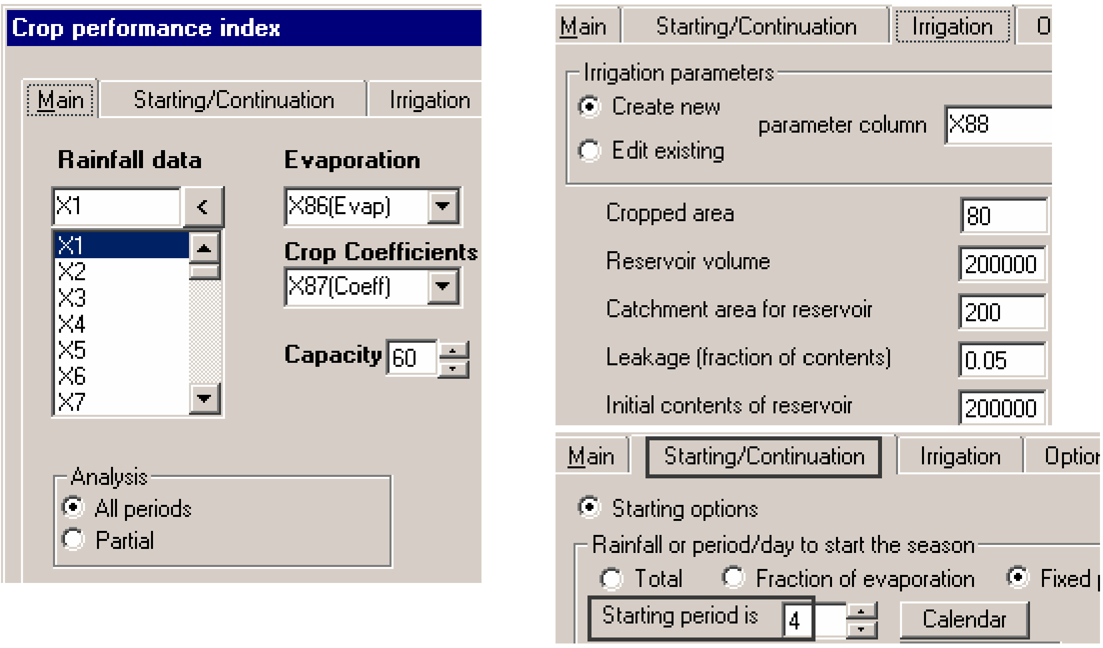


Fig. 12.2h shows the results from the first year.

Fig. 12.2h Results for the reservoir and crop for year 1

```
Analysis for rainfall data in X1
Catchment area     = 200        km. sq
Reservoir capacity = 200000     cu. m.
Initial vol.       = 200000     cu. m.
Area irrigated     = 80         km. sq
Rate of loss       = 0.050 per period
```

Period	Rainfall (mm.)	Reservoir (cu. m.) Needed	Remaining
1	2	-----	195080
2	20	-----	200000
3	1	-----	193040
4	9	36888	164280
5	8	37496	134830
6	0	44000	84088
7	18	31776	83668
8	27	24664	108160
9	0	46000	56752
10	27	28664	78591
11	75	0	200000
12	39	8759	200000
13	94	0	200000
14	86	0	200000
15	163	0	200000
16	0	0	190000
17	0	32000	148500
18	0	40000	101075

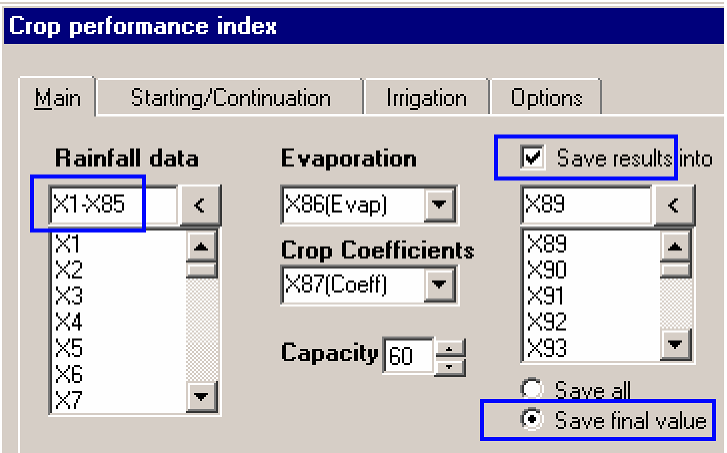
Total water requirement = 850mm.

Period Date	Crop	WR mm	Ra mm	Sur/Def mm	Index
4	1	55.0	0.0	0.0	100
5	1	55.0	0.0	0.0	100
6	2	55.0	0.0	0.0	100
7	3	57.5	0.0	0.0	100
8	4	57.5	0.0	0.0	100
9	5	57.5	0.0	0.0	100
10	6	62.5	0.0	0.0	100
11	7	62.5	12.2	0.0	100
12	8	62.5	0.0	0.0	100
13	9	62.5	31.5	0.0	100
14	10	62.5	55.1	0.0	100
15	11	50.0	60.0	108.4	97
16	12	50.0	10.0	0.0	97
17	13	50.0	0.0	0.0	97
18	14	50.0	0.0	0.0	97

(Loss from surplus = 3.0 in 1 periods)

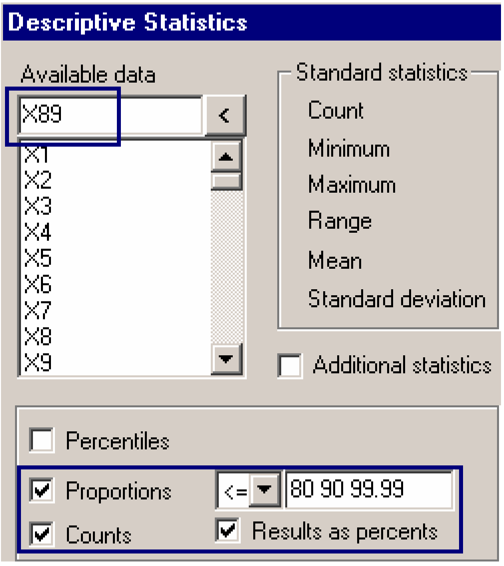
This can be repeated for each year with the dialogue shown in Fig. 12.2i.

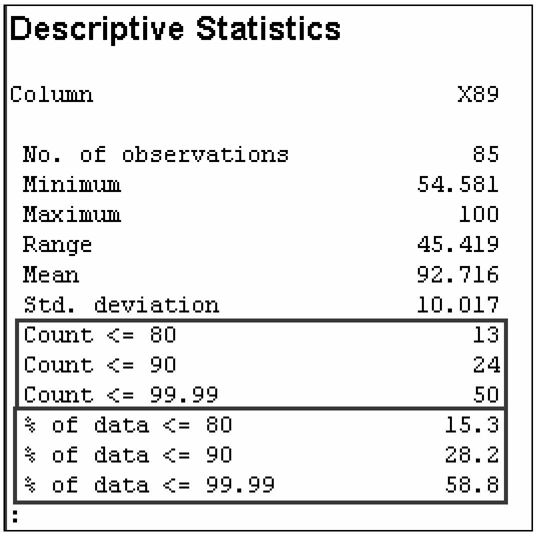
Fig. 12.2i Crop indices for each year



The data in X89 can now be analysed to assess the risk of a low crop index when 80ha are irrigated. This can be done by using **Statistics ⇒ Summary ⇒ Describe** for **X89** and requesting "**Proportions**" for **80**, **90 and 99.99**.

Fig. 12.2i Crop indices for each year





Descriptive Statistics

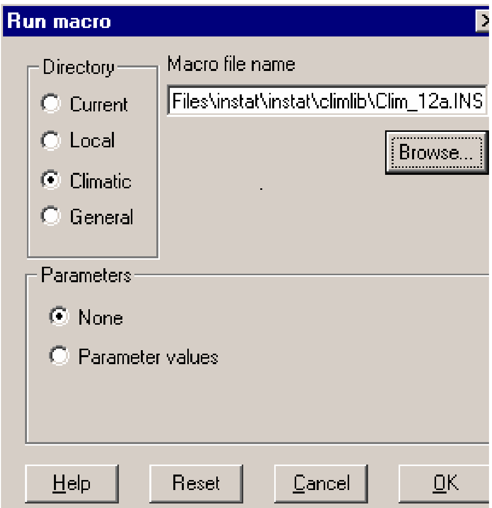
Column	X89
No. of observations	85
Minimum	54.581
Maximum	100
Range	45.419
Mean	92.716
Std. deviation	10.017
Count <= 80	13
Count <= 90	24
Count <= 99.99	50
% of data <= 80	15.3
% of data <= 90	28.2
% of data <= 99.99	58.8

This shows for example that there are 24 years = 100* 24/85 = 28% of the years when the final crop index was less than or equal to 90.

Fig. 12.2k shows the use of a macro that plots the risk of a low crop index for cropped areas (*A*) of 20 to 160 ha.

Fig. 12.2k Calculating the risk for areas *A* = 20 to 160 ha

Submit ⇒ Run Macro select **Clim_12a.ins** from **Climatic library**



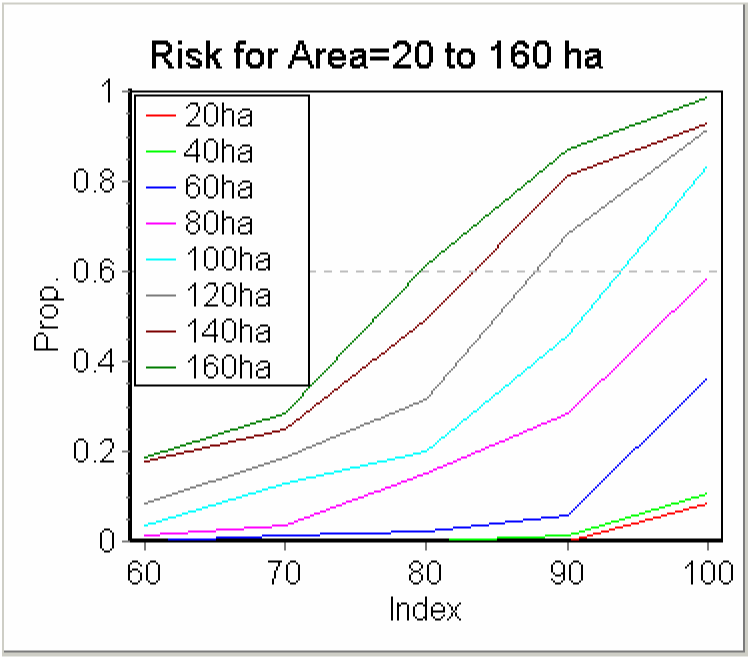


Fig. 12.2k shows the graph from a sequence of analyses with areas from 20 to 160ha. It shows a low risk of crop failure when 20 or 40ha are irrigated (bottom 2 lines in Fig. 12.2k). With 60ha irrigated, the risk that the index, *I*, will end up being less than 90 is very low, but there is about a 40% chance that *I*<100, i.e. that the water requirement is not completely satisfied. At the other extreme, with 160ha irrigated, there is a 50% chance that the final index will be less than 75.

12.2.4 Costs, profits and decisions

The analyses above **may** be sufficient to estimate what area to irrigate. For example, 40ha is only slightly more risky than 20ha and will give twice the yield, unless the crop fails. Fig. 12.2k shows that the chance of crop failure is low. There seems relatively little risk with 60ha, but above 60ha the risk begins to rise in terms of the value of the final index. Perhaps an area of 80ha is a reasonable value.

To go further needs information on how the index is related to yield and how much the crop is worth.

To illustrate one method we make some crude assumptions, that the maximum yield is 6 tonnes /ha and is related to the index, I, as shown in Fig. 12.2l. This models the yields as 6ha if the index, I, is greater than 95. The yield decreases linearly to zero as the index drops to 70.

Fig. 12.2l Yield per hectare v Crop Index

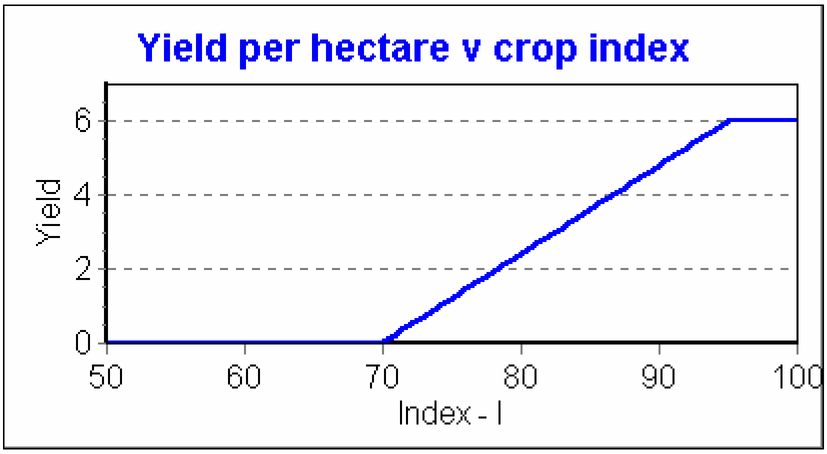


Fig. 12.2m shows the way this relationship allows the indices, saved in x91 to x98, to be translated into yields and income. If the price is 1 unit per tonne then the **income=Area*yield**. Hence the **income** each year and the risk levels can be calculated for given areas.

Fig. 12.2m Calculating "income"

Submit ⇒ Run Macro select **Clim_12b.ins** from **Climatic library**

Column	Pr.<=200	Pr.<=400	Pr.<=600	Mean	Min.	Max.
20	1	1	1	120	120	120
40	0.0118	1	1	238.9	148.7	240
60	0.0353	1	1	341.7	0	360
80	0.1529	0.2824	1	394.8	0	480
100	0.2	0.3294	1	416.9	0	600
120	0.2588	0.4353	0.7059	405	0	720
140	0.3412	0.5529	0.7294	366.3	0	840
160	0.4824	0.6235	0.7882	320.5	0	960

:

Suppose 400 units is the minimum income required. Then this can be achieved with lowest risk, 28%, by irrigating 80ha. However, the maximum income with this area is 480 units.

An alternative would be to maximise the expected (or average) income, and this is at 100ha. However, this may not satisfy the subsistence farmer, who has to survive each year, not just maximise the long-term income. In 1 year in 5, irrigating 100ha returns an income of less than 200 units.

The cost of the land preparation and planting can also be included. If this is assumed to be equal to the value of 1 tonne per ha, then the expected profit (i.e. income - cost) is very similar from irrigating 80ha as 100ha.

12.2.5 Further options

Currently the model is merely a simple water-budgeting exercise. There are many ways it could be made more realistic. However, care should be taken in building complex models, unless data are available to validate the models. Ways in which the model could be extended include the following:

- **More realistic rainfall data.** One extra option in the reservoir model is to allow a different distribution of rainfall to the reservoir and the irrigated site.
- **More realistic modelling of the reservoir.** In particular, the run-off process of the catchment could be included. The run-off will also be related to rainfall.
- **More realistic crop model.** This could include multi-layer soil models, plus a physiologically based growth model. The possible effects of pests and diseases could be included.
- **More flexible strategies for the use of water.** In practice, reservoir managers would have options, each period, of providing less than the total water requirement, or cutting supplies to part of the irrigated area.

12.3 Case Study 3 - The analysis of storm data

This is a summary of data collected as part of a detailed hydrological study in Malawi. The data are from a continuously recording rain gauge installed in Bvumbwe catchment. The gauge recorded cumulative rainfall on a paper chart that was changed every 24 hours. The data are therefore recorded in analogue form, i.e. as a line on a chart, and were digitised before analysis. For this study the digitising was done by recording the date and time that each rainfall event (storm) started. Throughout the storm the rainfall falling in each 5 minute period was read from the chart and recorded as an intensity in mm per hour.

The main objective was to recommend ways in which such data could usefully be summarised and analysed. Data for the first 50 storms of the 1983/4 rainy season are in the Instat worksheet called **Storm.wor** as follows:

X1 – X50	contains the 5-minute data for storms 1 to 50. These columns are of various lengths, depending on the duration of the storm
X51 – X54	are each of length 50, where
X51	day of each storm
X52	month
X53	year
X54	starting time, in minutes after midnight

The analysis is in four stages. The first is an exploratory look at the data. The second stage is to consider some of the standard summaries for this type of data. Then we look at different units on which results may be required and finally try further analyses of some of the summary data.

12.3.2 The exploration phase

The exploration stage includes a simple display of the data and a graph of the profile of the storms through time. Some results are shown in Fig. 12.3a. The display of the data uncovers some possible problems immediately. All columns begin with a zero, and many, such as X1, have many zeros at the end. Some have zeros in the middle, for example X8 has more than 3 hours of zero rainfall in the middle of the storm. Given that the data includes multiple storms on the same day, why were these data all considered as part of the same storm? A possible answer is that some studies recommend that rainfall events be considered as different, only when there is more than 6 dry hours between them.

Fig. 12.3a An initial look at the storm data

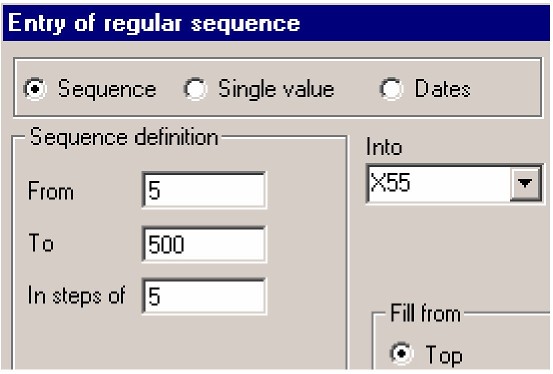
File ⇒ Open From Library ⇒ storm.wor

	X1*	X2*	X3*	X4*	X5*	X6*
	stm1	stm2	stm3	stm4	stm5	stm6
1	0.00	0.00	0.00	0.00	0.00	0.00
2	10.63	2.30	0.00	0.00	5.42	0.17
3	33.76	2.93	3.62	2.36	14.88	0.17
4	20.57	2.83	1.98	37.90	18.23	0.17
5	5.38	2.52	0.71	59.51	9.14	0.17
6	5.38	2.78	0.00	33.82	11.79	0.17
7	2.81	4.75	0.00	10.56	15.02	0.17
8	3.96	4.30	1.12	30.98	9.73	0.17
9	4.61	0.61	1.28	43.52	5.95	0.17
10	11.15	0.23	6.83	18.01	6.11	0.17
11	3.41	0.13		6.14	6.75	0.17
12	3.94	0.14		3.95	6.76	0.17
13	2.26	0.21		6.54	2.78	0.17
14	0.56	0.27		6.79	1.78	0.13
15	0.56			4.64	1.78	0.12
16	0.48			5.93	1.78	0.12
17	0.42			13.64	1.78	0.12
18	0.42			16.81	0.64	0.12
19	0.00			16.37	0.64	0.12
20	0.00			6.28		0.12

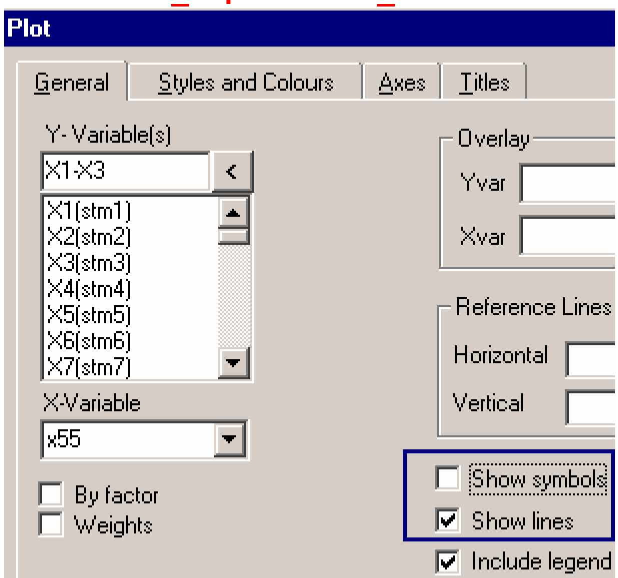
	X51*	X52*	X53*	X54*
	day	month	year	start
1	14	10	83	793
2	15	10	83	77
3	15	10	83	939
4	15	10	83	1125
5	16	10	83	867
6	27	10	83	66
7	27	10	83	1142
8	28	10	83	115
9	31	10	83	953
10	2	11	83	800
11	3	11	83	700
12	13	11	83	947
13	15	11	83	1029
14	16	11	83	141
15	23	11	83	1097
16	27	11	83	1182
17	28	11	83	763
18	29	11	83	831
19	1	12	83	685
20	1	12	83	874

Fig. 12.3b Plotting the data from individual storms

Manage ⇒ Data ⇒ Regular Sequence



Graphics ⇒ Plot

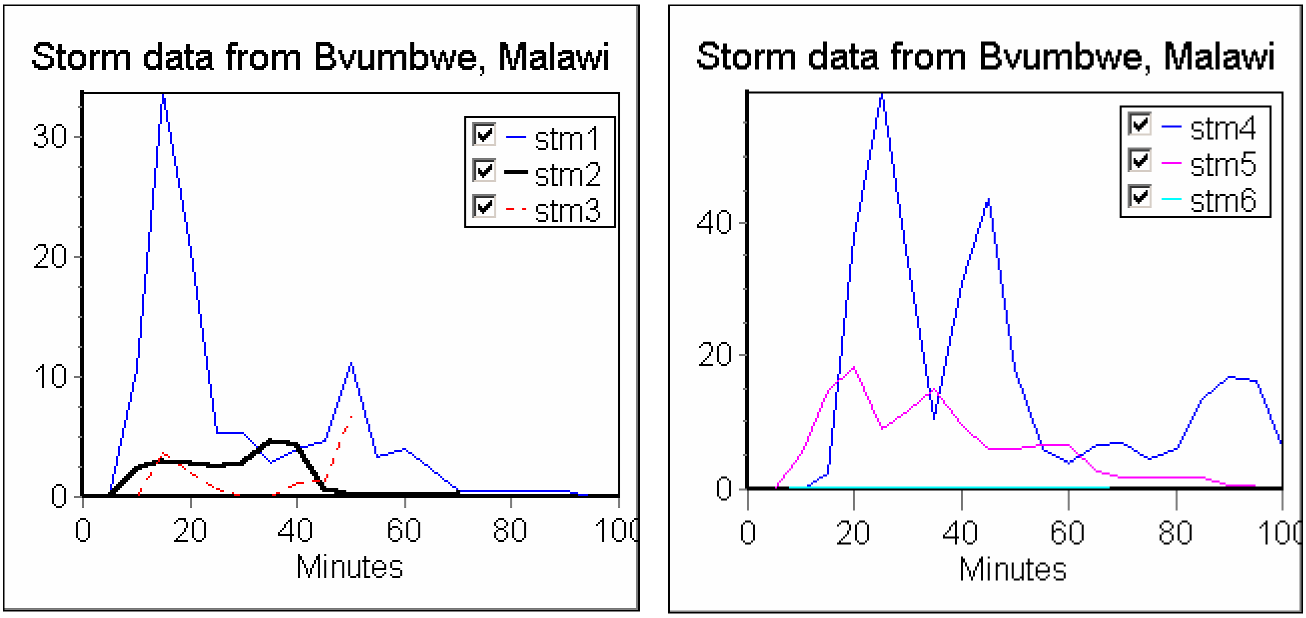


The graphs in Fig. 12.3c seem useful. Sets of graphs can be tiled and studied in turn. If necessary a graph can be scaled differently to look at part of the storm profile.

To repeat for all columns would be tedious, so a small macro was written as shown in Fig. 12.3d. It plots five storms at a time. The graphs give a good visual impression of the data and also uncover a second type of problem. Some storms seem to stop in "mid-air", as in X3(stm3). It is not clear if this is genuine, or a breakdown of the equipment.

Two general points can be made at this stage. The first concerns the availability of information **about** the data. The queries above, concerning the data, are typical that an exploratory analysis will usually give rise to some questions. These can be answered by reference to the raw data, in this case the charts, or by good documentation of the data, or by the availability of someone who understands the data well.

Fig. 12.3c Resulting graphs



The first possibility, namely inspection of the charts, is usually only feasible if the data are analysed where they are archived. This is desirable and becomes more feasible as countries get good local computing facilities. The second possibility, normally the availability of good documentation is, unfortunately, rare and is often only possible if the initial analyses were conducted on-site. The third possibility, an expert on the data is more common. However, this amounts effectively to leaving some of the database information in the heads (or filing cabinets) of the staff who are responsible for the data. This is fine while staff remain, but any transfer of these staff takes away part of the database!

The second general point, illustrated by the macro in Fig. 12.3d, is that analyses of non-standard data sets usually involves more time on organising and manipulating the data than on the actual analyses. This data management phase requires a good knowledge of the software being used.

Fig. 12.3d Macro to plot 50 storms

Editing FIG12_3D.INS

```
Note FIG12_3d.INS to calculate kinetic energy
restore : warn off : echo off
%1=1 : %2=51
par 1
loop 1; repeat 50
  select x%1; into x%2; if x%1>1
  par 2 x%2; inf 2
  loop 2; if %p2>0
    par 1 %p1x%2
    x%2=x%2/12*(11.9 + 20.1*ln(x%2))
  loop 2
  %1=%1+1 : %2=%2+1
loop 1
stats %p1; sum x109
describe x109; median
```

Within Instat the macro used in Fig. 12.3d was written just for this problem. It is therefore the type of macro that would normally become part of a personal macro library, but it is not general enough to be useful in the local site library.

Once you have finished with the graphs, they can be removed in turn, or all at once using **Window ⇒ Graphs ⇒ Close All**.

12.3.3 Summary statistics

The second stage is to consider standard summary statistics from the storm profiles. The following characteristics of storms are often calculated and their distribution found:

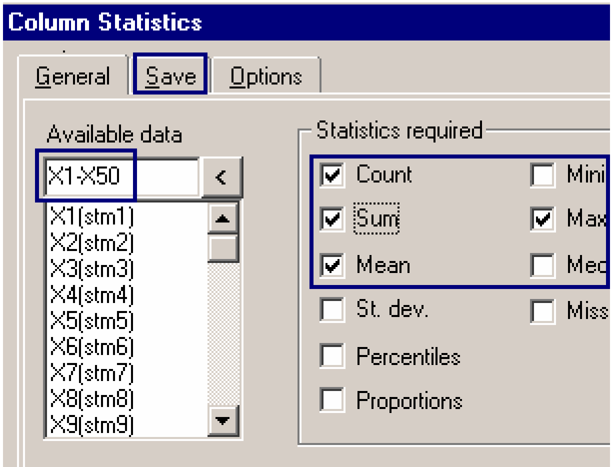
- total rainfall
- duration (minutes)
- maximum intensity
- average intensity

These statistics can be calculated with the **Statistics ⇒ Summary ⇒ Column Statistics** dialogue as shown in Fig. 12.3e. The resulting columns are each of length 50, hence are summaries at the 'storm level'. There is however a difficulty, which is typical of analyses done 'automatically'. The exploratory analysis showed that some columns, such as X1 (the first storm), have many zeros. This does not affect the total rainfall (X59), nor the maximum intensity (X58), but it does affect the duration (X56) and therefore the mean intensity (X57).

Fig. 12.3e Summary statistics for 50 storms

Statistics ⇒ Summary ⇒ Column Statistics

(with Save tab to save summaries in **X56-x59)**



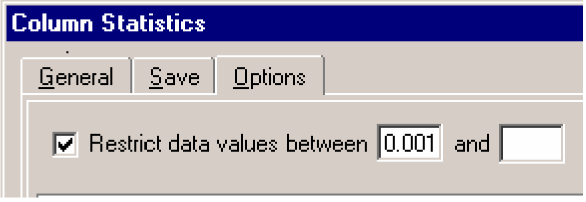
Column Statistics

Column	Count X56	Mean X57	Max. X58	Sum X59
stm1	82	1.345	33.76	110.3
stm2	14	1.715	4.752	24.01
stm3	10	1.555	6.834	15.55
stm4	79	4.652	59.51	367.5
stm5	19	6.367	18.23	121
.....	..	.....	.....	.....
.....	..	.....	.....	.....
stm48	14	10.78	63.15	150.9
stm49	14	4.24	27.81	59.37
stm50	15	0.1322	1.89	1.983

:

Fig. 12.3f Storm statistics after excluding zero periods

`Manage ⇒ Data ⇒ Clear(Remove) ⇒ X56-X59`
`Statistics ⇒ Summary ⇒ Column Statistics`




```
Column Statistics

Column      Count     Mean      Max.      Sum
             X56       X57       X58       X59

stm1          17     6.488     33.76     110.3
stm2          13     1.847     4.752     24.01
stm3           6     2.592     6.834     15.55
stm4          77     4.773     59.51     367.5
stm5          18     6.721     18.23       121
.....         ..     .....     .....     .....
.....         ..     .....     .....     .....
stm48         13      11.6     63.15     150.9
stm49          8     7.421     27.81     59.37
stm50          7    0.2833      1.89     1.983
:
```

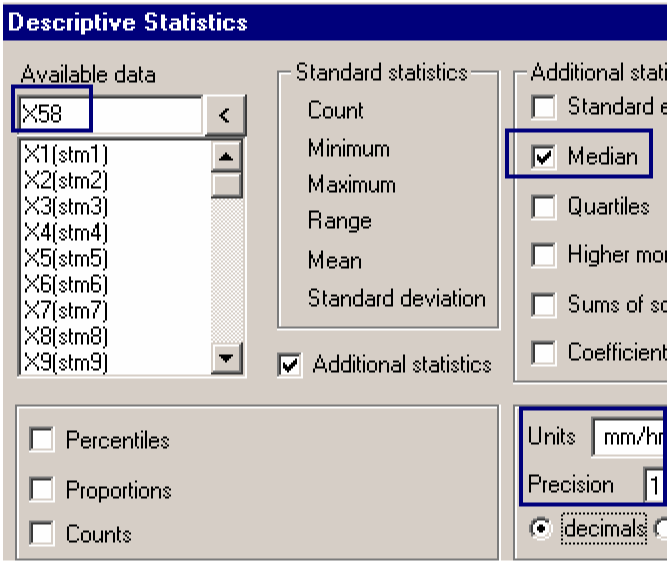
Here, as the problems have not been resolved, we have chosen to restrict the calculations to the non-zero values. Hence, of the Column Statistics dialogue is used again with the **`Restrict data values`** option as shown in Fig. 12.3f, to ignore all zero values.

The count in X56 is for every 5 minute period, so is multiplied by 5 to find the duration of the storms in minutes.

: X60 = X56 * 5

Fig. 12.3g Summary statistics of the 50 storms

`Statistics ⇒ Summary ⇒ Describe`
`(used as here, and then for` **`x60`** `with units changed to "minutes")`




```
Descriptive Statistics
Column                        X58

 No. of observations          50
 Minimum                   0.2 mm/hr
 Maximum                  84.4 mm/hr
 Range                    84.2 mm/hr
 Mean                     20.0 mm/hr
 Std. deviation           23.2 mm/hr
 Median                    8.1 mm/hr
:
```

```
Descriptive Statistics
Column                       X60

 No. of observations          50
 Minimum                      5 minutes
 Maximum                    885 minutes
 Range                      880 minutes
 Mean                       135 minutes
 Std. deviation             161 minutes
 Median                      80 minutes
:
```

A summary of the resulting data (Fig. 12.3g) shows that the maximum intensities have a median of 8 mm/hr and a maximum of 84 mm per hour. The durations have a minimum of 5 minutes, mean of 135 minutes and maximum of 885 minutes.

Many other summaries are possible and one more is considered here.

The **kinetic energy** is sometimes required. This is difficult to measure and depends on drop size. Various approximate formulae have been proposed, one of which is

$$KE=11.9 + 20.1 * \ln(i)$$

where KE is in joules per m^2 per mm of rainfall, i is the intensity in mm per hour and $\ln(i)$ is the natural log of i. Here, with 5 minute data, any value of i in 5 minutes corresponds to a total rainfall of I/12mm per hour, hence the associated KE, for that period, is given by

$$KE=i/12*(11.9+20.1*\ln(i))$$

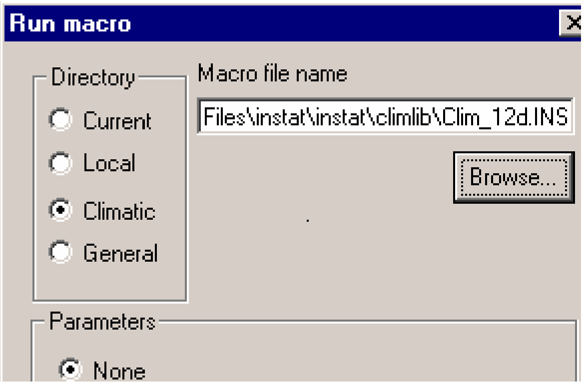
For each period of the first storm, X1, the corresponding kinetic energy could therefore be calculated, and stored in X71, by

: **X71=X1/12*(11.9+20.1*ln(X1))**

The macro called **Clim12d.ins** calculates the kinetic energy (KE) for each period of the 50 storms into X71-X120, calculates the total KE for each storm in X61 and then summarises the data in X61. The calculations are restricted to 5 minute periods with an intensity, i >1. In general, the low intensity values contribute little to the KE, and some lower limit must be set. The formula cannot be used for the many very low values in this set of data because they give negative values to the energy!

Fig. 12.3h Macro to calculate KE for 50 storms

Submit ⇒ Run Macro select **Clim12d.ins** from **Climatic library**



Column	X61
No. of observations	44
Minimum	2 joules/msq
Maximum	2880 joules/msq
Range	2878 joules/msq
Mean	529 joules/msq
Std. deviation	704 joules/msq
Median	179 joules/msq

The results in Fig. 12.3h show that there were 44 storms with a peak intensity >1. They had a median KE of about 180 and a maximum of nearly 2900 joules/m^2.

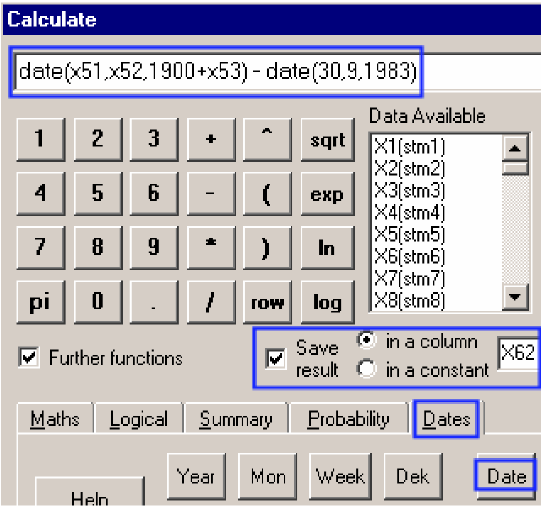
12.3.4 Units of analysis

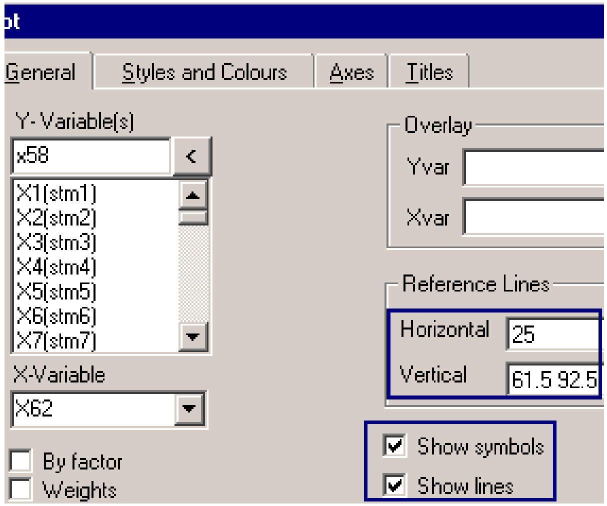
The 'unit' of analysis for the storm data is now considered briefly. There are two levels in the data, one is the 5-minute level of columns X1-X50 and the other is the storm level, with data in X51-X59 and X61. Much of the work so far has been on the calculation of storm level data from the 5-minute data. Four questions are considered to illustrate the concept of different units, all related to the proportion of rain that is at an intensity of >25mm per hour. Some can be answered from the data already generated, while others require more analysis.

1) What proportion of storms has at least 5 minutes of rain at an intensity >25mm/hour?
2) What proportion of days in December has (at least 5 minutes of) rain at an intensity >25mm/hour?
3) What proportion of rain falls at an intensity >25mm per hour?
4) What proportion of rainy periods has rain at an intensity >25mm per hour?

Fig. 12.3i Plot of maximum intensity v date

Manage ⇒ Calc **Graphics ⇒ Plot**





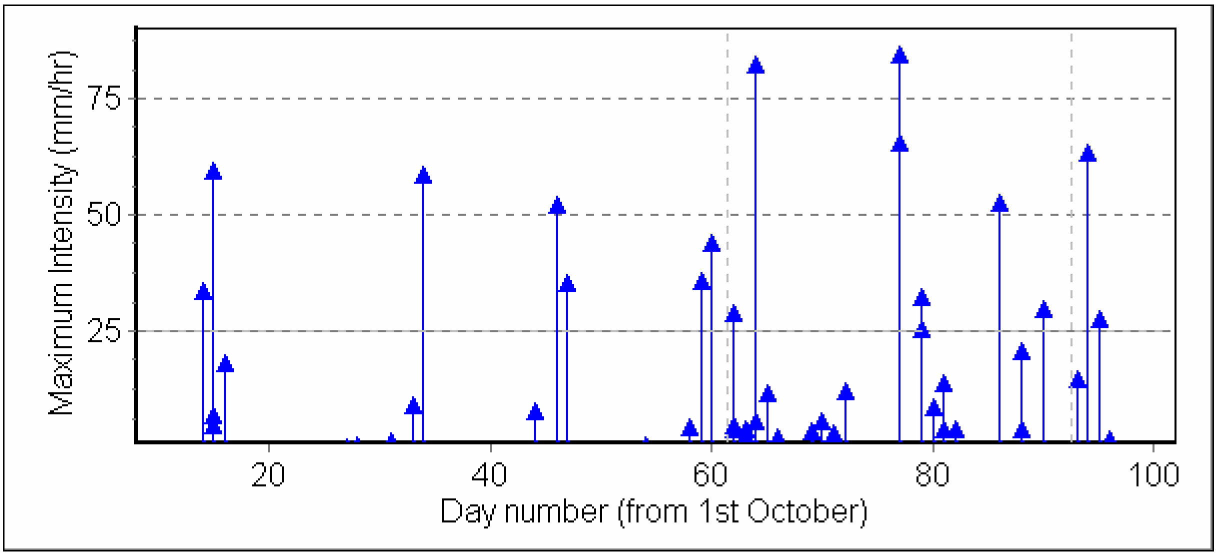
The graph in Fig. 12.3g is one way of answering the first two questions. This plots the maximum intensity against the date. The day (x51), month (x52) and year (x53) columns have first to be combined to give a day number. One way is

: x62 = date(x51, x52, 1900+x53) – date(30, 9, 1983)

This gives 1st October 1983 as day 1. The dialogue is shown in Fig. 12.3i.

For the first question, where the unit is the storm, Fig. 12.j shows that 17 out of the 50 storms, i.e. about 30% **of storms**, included rainfall at an intensity of >25 mm/hour.

Fig. 12.3j Resulting graph

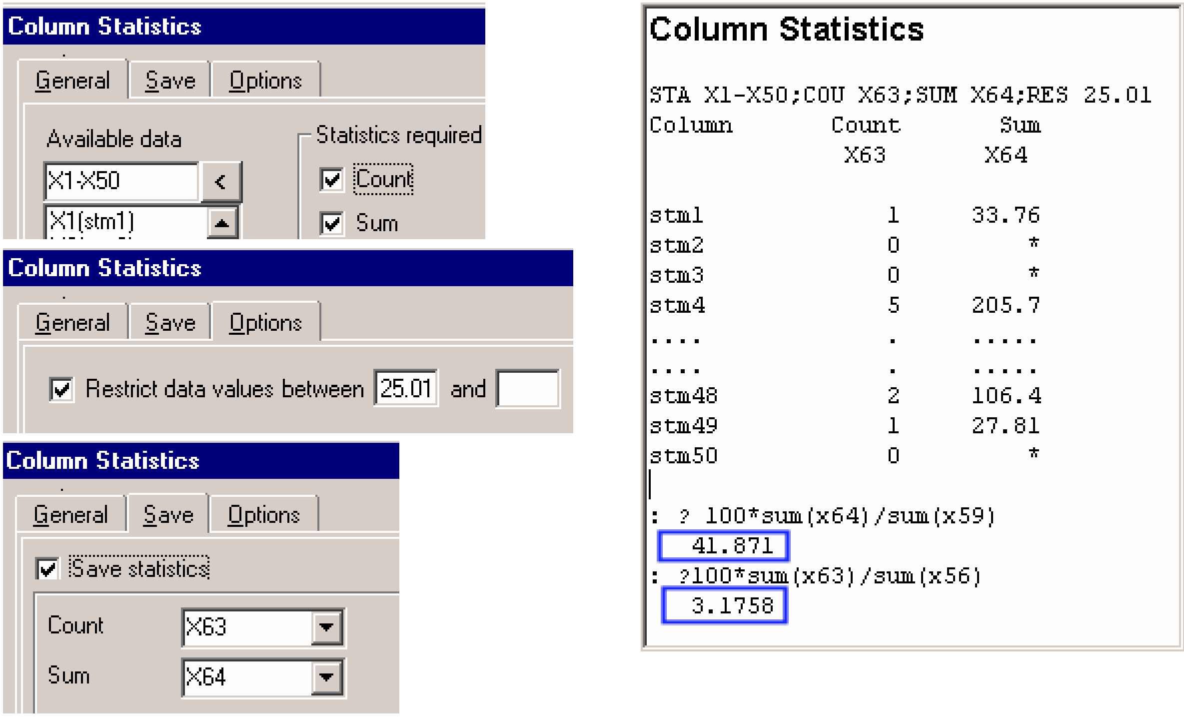


For the second question, the unit is a day. Fig. 12.3j shows that there were 8 storms in the 31 days of December with rainfall intensity > 25mm/hr. However, 2 days each had 2 of these storms, hence there were 6 December days with storms > 25mm/hr, i.e. about 20% **of days**, just over one day per week.

The other two questions need the calculation of further summaries. Use **Statistics ⇒ Summary ⇒ Column Statistics**, Fig. 12.3k, restricting attention to the values in X1-X50 that are >25mm/hr.

Fig. 12.3k Summary statistics for third and fourth units

`Statistics` ⇒ `Summary` ⇒ `Column Statistics`



Then a mm of rain is the unit for the third question, so consider the totals in X64 compared to the overall rainfall totals in X59 for the third question.

: ? 100 * sum(x64)/sum(x59)

The result, shown in Fig. 12.3k, is that 42% of the rainfall falls at an intensity of >25mm/hr.

For the final question, compare the number of periods at high intensity, in X63, with the total count of wet periods.

: ?100 * sum(x63)/sum(x56)

The result, also in Fig. 12.3k, is that only 3% of wet periods were at high intensity.

The above analyses provided 4 different answers concerning rain intensity greater than 25mm per hour. The answers range from 3% of the periods to 15% or 30% of the storm, to 42% of the rain. We should not look for the right answer because each corresponds to a different question. They act as a reminder that analyses must relate to the precise objectives of the study.

12.3.5 The analysis of circular data

The final stage, looks briefly at the time of day at which the rainfalls begin. The data are in X54, recorded as minutes after midnight. They are first transformed to hours, then the analysis begins with a simple graphical display, using a stem-and-leaf plot, Fig. 12.3i.

Fig. 12.3l An initial look at the time of the start of the rains

Manage ⇒ Calc (or type command)

X65 = X54/60

	X51*	X52*	X53*	X54*	X65
	day	month	year	start	
1	14	10	83	793	13.22
2	15	10	83	77	1.28
3	15	10	83	939	15.65
4	15	10	83	1125	18.75
5	16	10	83	867	14.45
6	27	10	83	66	1.10
7	27	10	83	1142	19.03
8	28	10	83	115	1.92
9	31	10	83	953	15.88
10	2	11	83	800	13.33
11	3	11	83	700	11.67
12	13	11	83	947	15.78
13	15	11	83	1029	17.15
14	16	11	83	141	2.35
15	23	11	83	1097	18.28
16	27	11	83	1182	19.70

Graphics ⇒ Stem and Leaf

Stem and Leaf

Data to be displayed: x65

☑ Do not trim extreme values

```
STEM & LEAF DISPLAY FOR X65
 Min=0  Max=23  No. obs.=50
23 is represented by 2:3 3

   5     0:0-1 01111
   7     0:2-3 23
   8     0:4-5 5
  11     0:6-7 667
  13     0:8-9 88
  18     1:0-1 01111
(12)     1:2-3 223333333333
  20     1:4-5 4445555
  13     1:6-7 6677
   9     1:8-9 88899
   4     2:0-1 011
   1     2:2-3 3
```

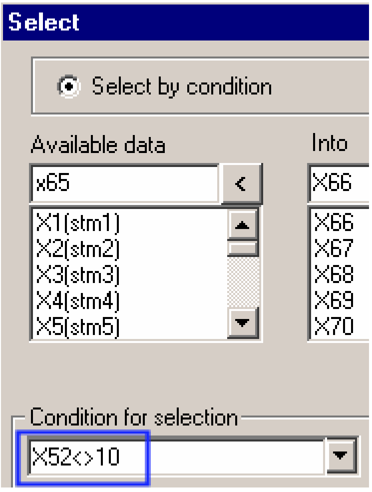
The results in Fig. 12.3l show that most rainfalls start in the early afternoon, however, there is a slight hint of bimodality, with 5 observations starting just after midnight. Although the sample size of just 50 storms is small, there is a general point, that bimodality usually has an explanation. When dealing with structured data, the complication can often be explained by some element of the structure.

Here an examination of 'Month', in X52, and the data in X65, Fig. 12.3l, shows that three of these five observations are from the beginning of the season, i.e. in October. When the October data are removed the bimodality disappears, Fig. 12.3m.

The next step is to give some summary statistics for X65 (or X66) and care is needed, because the data are **circular**. To take a simple example, suppose there were just 2 rainfalls, which start at 11pm and 1am. The mean is midnight. If they are coded in the same way as X65 and X66, they would have the values 23 and 1, giving a mean of 12, i.e. midday! Fig. 12.3n shows this problem in Instat, where the **Statistics ⇒ Summary ⇒ Describe** dialogue gives a mean of 12 o'clock. There is also a correspondingly large standard deviation, for these two observations, though they are close together on the circle of 24 hours.

Fig. 12.3m Remove data in October

Manage ⇒ Reshape ⇒ Select



Graphics ⇒ Stem and Leaf for X66

```
 STEM & LEAF DISPLAY FOR X66
 Min=0  Max=23  No. obs.=41
23 is represented by 2:3 3

   2     0:0-1 01
   4     0:2-3 23
   5     0:4-5 5
   8     0:6-7 667
  10     0:8-9 88
  15     1:0-1 01111
 (11)    1:2-3 22333333333
  15     1:4-5 4455
  11     1:6-7 6677
   7     1:8-9 889
   4     2:0-1 011
   1     2:2-3 3
```

Fig. 12.3n Incorrect analysis of a simple example of circular data

Enter data into X67

	X65	X66	X67
1	13.22	13.33	1
2	1.28	11.67	23
3	15.65	15.78	

Stats ⇒ Summary ⇒ Describe for X67

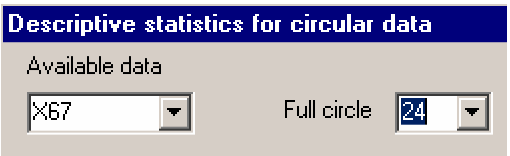
Descriptive Statistics

```
Column                   X67

 No. of observations      2
 Minimum                  1 hours
 Maximum                 23 hours
 Range                   22 hours
 Mean                    12 hours
 Std. deviation      15.556 hours
```

The solution, in Instat, is the **Statistics ⇒ Summary ⇒ Circular Statistics** dialogue, specifying that the **Full circle** is **24** (Fig. 12.3o).

Fig. 12.3o Analysis of circular data

Statistics ⇒ Summary ⇒ Circular Statistics



Descriptive statistics for circular data

```
Circular statistics for X67
Full circle =             24
Number of observations     2
 Circular mean            24
 Circular variance    0.0341
 Standard deviation   1.0058
```

This correctly gives the mean as 24hr and a low standard deviation of 1hr. The interpretation of the circular mean and standard deviation is similar to that in ordinary statistics. The circular variance is different, and is on a 0 to 1 scale, with 0 being no variation and 1 being the maximum possible. Here the variance is 0.034, a low value, reflecting the closeness of the data points 23 and 1 on the circle. The analysis of circular, or directional data, is described in various books, for example Mardia (1972), Batschelet (1981).

The summary statistics for the times of the start of the 50 storms are given in Fig. 12.3p. They show a mean start time of about 2:40pm. The circular variance is 0.66 and the circular standard deviation is 5.6 hours. Fig. 12.3p also shows that omitting the October data changes the summary statistics only slightly.

Fig. 12.3p Circular statistics for the starting times of the storms

Statistics ⇒ Summary ⇒ Circular Statistics for X65 and X66

Descriptive statistics for circular data

```
Circular statistics for  X65

 Full circle =            24
 Number of observations   50
 Circular mean            14.606
 Circular variance        0.6571
 Standard deviation       5.5887
```

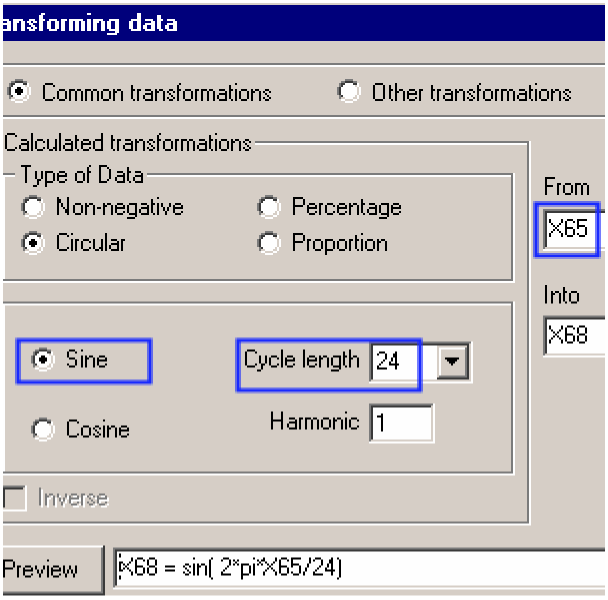
Descriptive statistics for circular data

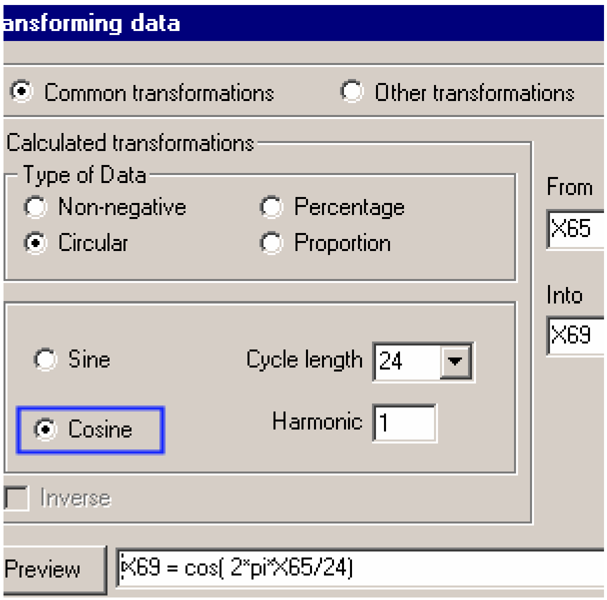
```
Circular statistics for X66

 Full circle =           24
 Number of observations  41
 Circular mean           13.888
 Circular variance       0.6206
 Standard deviation      5.3176
```

We now examine whether there is a relationship between the time when the storm starts and its other properties. Take, as an example, the duration of the storm, which was calculated as the count*5 minutes and saved in X60. Again, the data must be analysed slightly differently, because the starting times are circular, however no special commands are needed. Instead, transform the starting times in **X54** to a 2π range and take the sine and cosine terms. Fig. 12.3q shows how this is done with the **Manage ⇒ Manipulate ⇒ Transform** dialogue. The sine-transformed data are put into X68 and the cosines to X69. This dialogue generates transformations that could alternatively be done with the **Manage ⇒ Calculations** dialogue.

Fig. 12.3q Transforming circular variables for correlation and regression

Manage ⇒ Manipulate ⇒ Transform





The circular correlation is then defined as the multiple correlation between the linear variable, here the duration of the storm, and the circular variables, i.e. the sine and cosine terms.

Before doing the regression check briefly on the shape of the dependent variable. Fig. 12.3r uses boxplots to examine the shape of the distribution of the storm duration. The distribution is reasonably symmetrical after a log transformation.

Now use the **Statistics ⇒ Regression ⇒ Multiple** dialogue as shown in Fig. 12.3s.

Fig. 12.3r Transforming the data and then drawing boxplots

Manage ⇒ Calculations
(or type command – as shown below)

```
: X68 = sin( 2*pi*X65/24)
: X69 = cos( 2*pi*X65/24)
: x70 = sqr(x60)
: x71 =  ln(x60)
  WARN: Overwrite data in X71
:
```

Accept the warning to overwrite X71
Type:
: REMove X72-X120
to free the remaining columns

Graphics ⇒ Boxplot of **X70** and **X71**

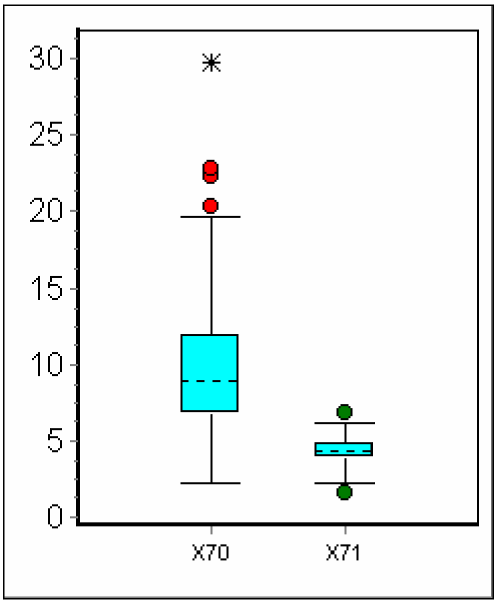


Fig. 12.3s Fitting a regression model

Statistics ⇒ Regression ⇒ Multiple

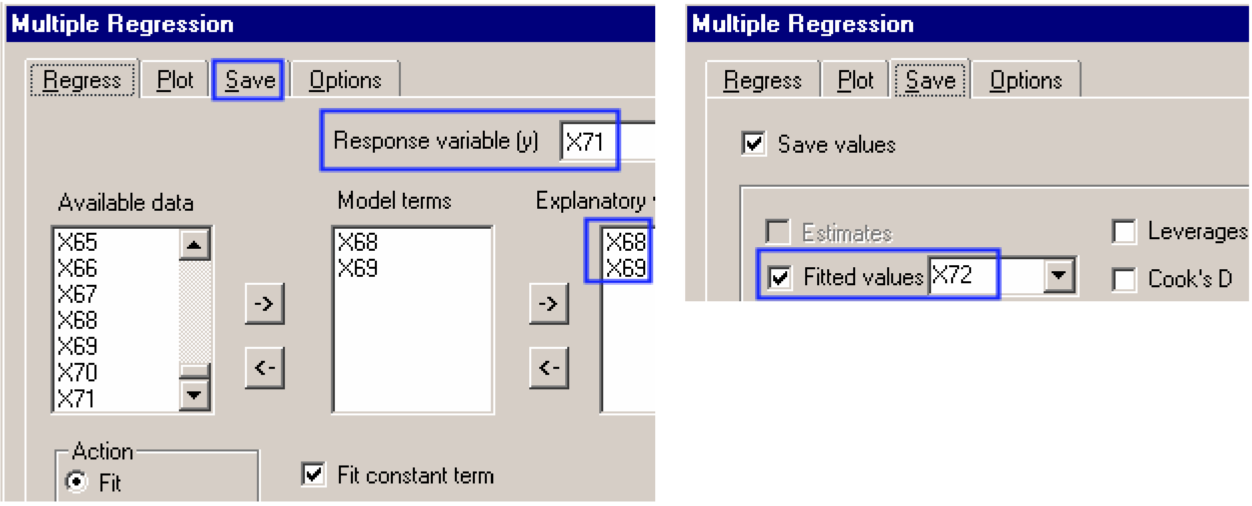


Fig. 12.3t Fit of the regression model

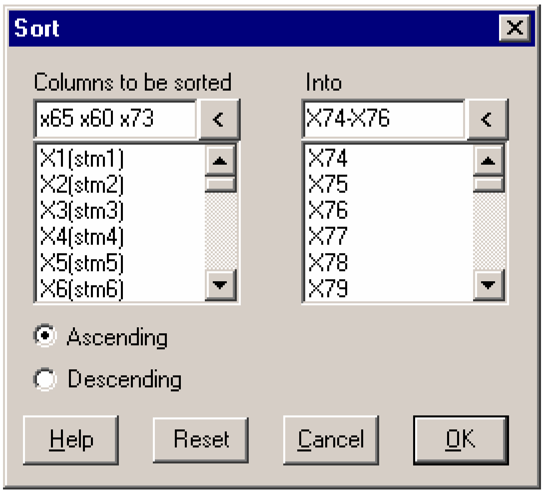
```
Multiple Regression

ANOVA for regression of X71 on X68 X69
-------------------------------------------------------------------
Source       df           SS            MS      F value     Prob>F
-------------------------------------------------------------------
Regression    2        15.5394        7.7697      10.38      0.0002
Residual     47        35.1706       0.74831
-------------------------------------------------------------------
Total        49          50.71
-------------------------------------------------------------------
R-squared = 0.3064  (adjusted = 0.2769)
:
```

The results in Fig. 12.3t show that the duration of the storms is clearly related to their time of start. In the dialogue shown in Fig. 12.3s, the fitted values are saved, so the results can be plotted.

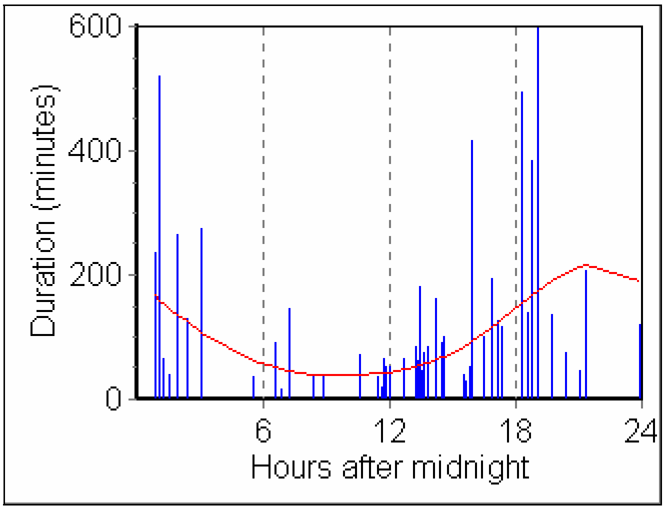
Fig. 12.3u Plots of transformed storm data

`Manage ⇒ Calc x73 = exp(x72)`

`Manage ⇒ Manipulate ⇒ Sort`



`Graphics ⇒ Plot of x75, x76 versus x74 (lines only, vertical for x75)`

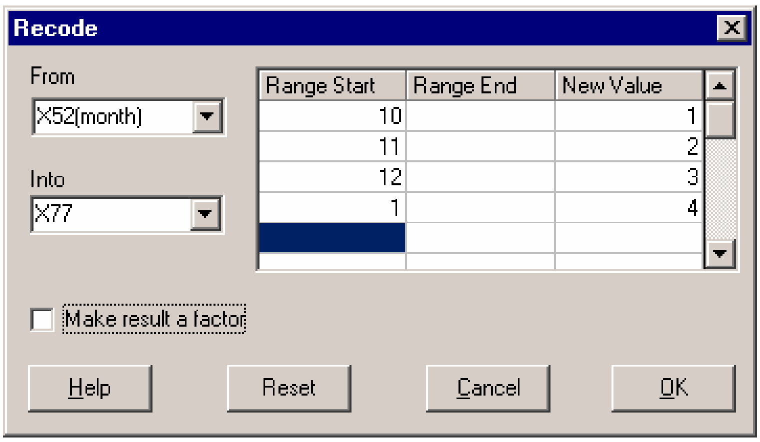


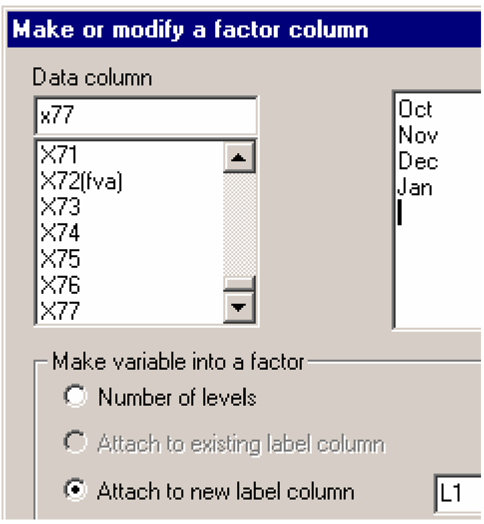
The fitted values are "back-transformed" and the data sorted into ascending order of the starting time, for the plot, as shown in Fig. 12.3u. The plot indicates that the mean duration is longer in those storms that start at night.

This is an interesting result and one might speculate on whether there are meteorological reasons for this. However, the lesson from the bimodality discussion (see Fig. 12.3l) is still relevant. It showed that the October rainfalls tended to be the ones that started in early morning and noting this aspect of the structure helped in the study of the starting times. Here this aspect was not included.

To investigate, the data in x52 are recoded, made into a factor, Fig. 12.3v, and included in the graph, Fig. 12.3w.

Fig. 12.3v Preparing the data on the time of year

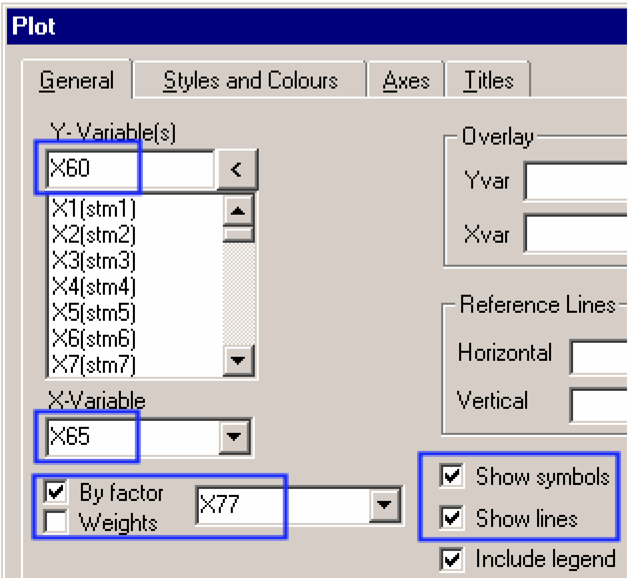


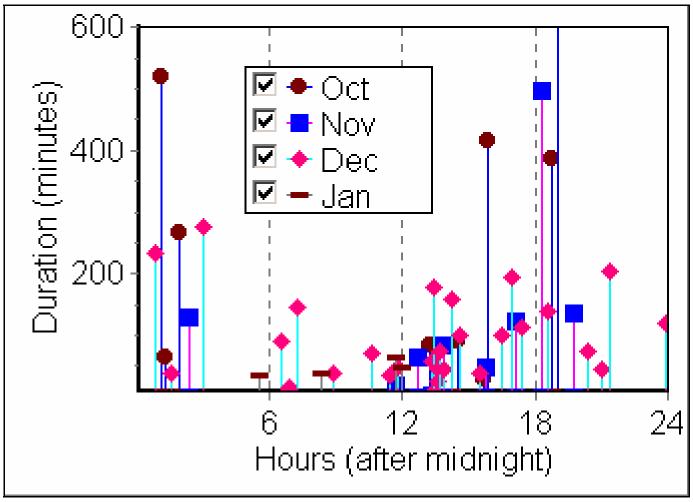


The graph in Fig. 12.3w shows that the majority of the long, nighttime storms are again in October. Hence is the relationship really one of nighttime starts or of early season rainfalls, or both? More data are needed to unravel the situation.

Fig. 12.3w Plotting the data by month

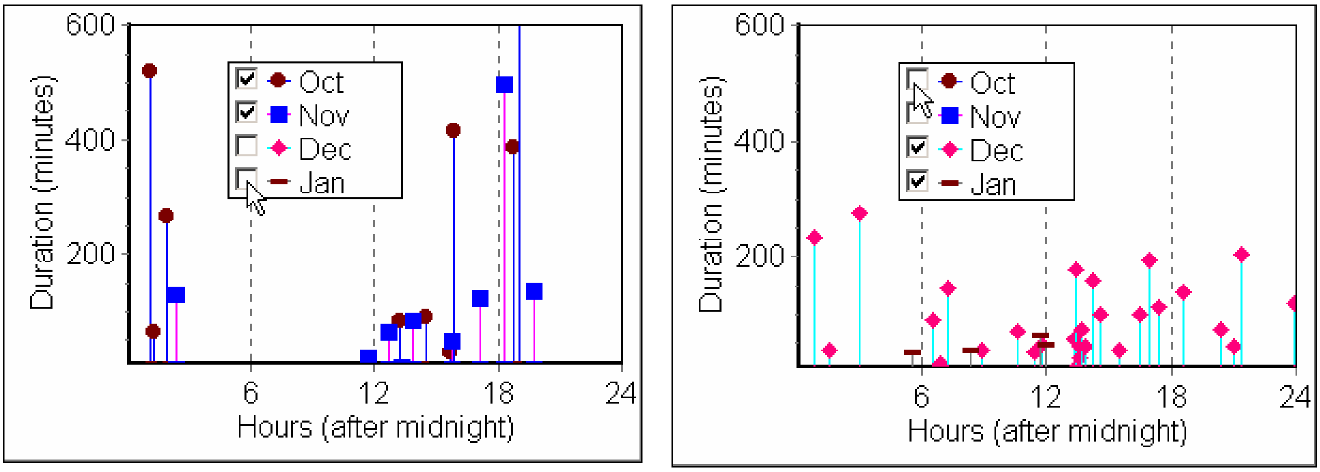
(Specify vertical lines for x60)





However, splitting the plot and showing October and November rainfalls separately from December and January provides some indication. The relation between duration and time of start seems to be an early-season feature.

Fig. 12.3x Separate plots for the early and later season rainfalls



12.3.6 Recommendations

The main objective of this case study was to suggest some ways in which the work on this type of data could proceed. Precise objectives of a future study were not given, and hence any suggestions will clearly be of a general nature.

Some suggestions have been made above, within the analyses. In particular, a data exploration phase is important, together with facilities to resolve problems concerning the data. Much of the time will be spent on data management, rather than statistics, and a good knowledge of the software is needed to conduct the data management aspects efficiently. The analysis must be done with care, for example concerning the way to deal with zeros, Fig. 12.3c and the potential problems concerning the analysis of circular data, Figs 12.3k and 12.3l. Finally, and most important, data should always be analysed taking account of the structure.

There is much that can be done with this type of data, which complements the analysis of the daily records, covered in the rest of this guide. More data are needed, because 50 storms is a small number, even though there are detailed data on each storm. One problem is that a major use of these data is to study erosion, and there are often only a few heavy storms each year. The second reason for more data is that the analyses above showed clearly the need to consider aspects such as the time of year, or starting time of the storm. Thus overall statistics, from all 50 storms, such as were derived above (e.g. 42% of the rainfall is at an intensity of >25mm/hr), must be treated with caution. They are initial indicators, but their values may depend on other factors, such as the time of year.

There are four ways in which more data could be provided.

- The first is to have more detailed data on the individual storms, for example one minute or 10-second data. This is only needed if the study is particularly concerned with measurements of short period intensity.
- The second is to have information on more storms and this is particularly important. The information on a few thousand storms over a reasonable set of years should enable the data to be analysed even allowing for different aspects of the structure. If the preparation of the data is done manually, then one possibility would be to record all storms in terms of the summary statistics (columns X51-X54 above), but only give the detailed profile for storms with a rainfall total above a certain threshold.
- The third is to have more ancillary data on the meteorological events leading up to the storm. Here this is just the date and time of the storm, in X51 to X54, but wind direction, pressure and type of rainfall etc., may also be available. The reason for these data is to avoid the use of variables like "time of year" as explanatory variables. They are usually only "proxy" variables and the real explanation is concerned with the meteorological situation, which may be different, depending on the season.
- Finally there may be data on multiple sites within a zone. Here the hope will be that the pattern of storms is similar at neighbouring sites, hence summary statistics, such as the proportion of the rainfall that is erosive, can be applied to all sites within the region. Data on a few sites are therefore needed to check this aspect, if results are to be used for a region.

In the past, these types of data requirement would be daunting. However, the study above shows that it would be quite feasible to analyse more data. It may be more convenient to use a more powerful statistics package than Instat, that could cope with the thousands of detailed profiles in a single worksheet.

The structure of the data would usually be as shown above, namely with data at a detailed level (as in X1-X50), plus other variables (as in X51-X54) at the storm level. Then, after some analyses of the individual profiles, the main management of the data would use the profiles to derive important summary statistics, at the storm level. Examples are the derived data in X56 to X59. These summary data are then analysed, depending on the objectives of the study.

The analyses above represent a few of the ways the data can be summarised and analysed, but there are many other possibilities. For example, summary statistics on rainfall intensity are often based on the maximum intensity over a 30 minute, rather than a five minute period. Within Instat, the **Start of the Rains** dialogue makes it easy to find these maximum intensities. Studies on the chance itrains at a particular time of day (in a particular season) are also sometimes useful.

Ideally, as here, the analyst will have continued access to the detailed data (X1-X50) and not be limited to just the summary data. This makes it easy to respond to new questions. For example, if there was a request to consider the proportion of rainfall that was erosive, but with erosive defined as I>20mm/hr, rather than the 25mm/hr considered above, it takes only a few minutes to revise the commands from Fig. 12.3k to give the answer, which here is 49%.

Chapter 13 – Modelling Daily Rainfall data

13.1 Introduction

This chapter introduces an alternative approach to the analysis of rainfall data. This introductory section is used to motivate the approach. Then there is a simple example of the modelling process and later sections cover the methods of fitting and using this approach in more detail.

A comprehensive analysis of rainfall data requires access to the full daily data. With the small example from Samaru that has been used repeatedly, there are 11 years of daily data in x1-x11. The analyses so far are depicted in Fig. 13.1a.

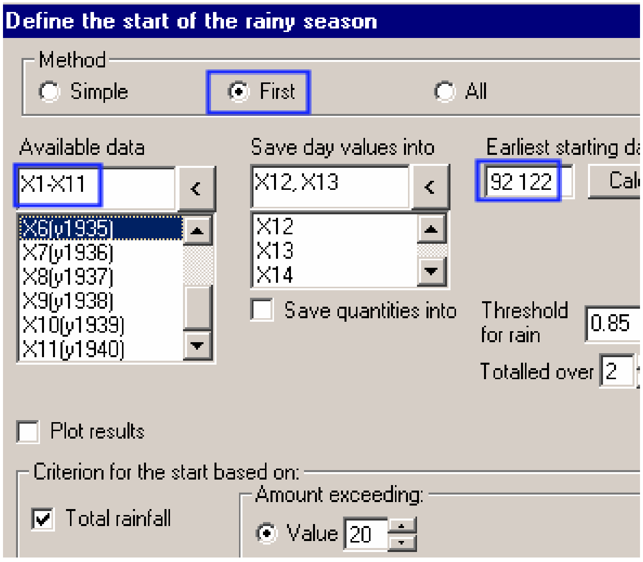
Fig. 13.1a Extracting information from the daily rainfall data

```
      x1 x2         x11                 Events:
     ------------------
1    |                 |
2    |                 |   ------ rainfall totals          (Chapter 5)
     |                 |   ------ the start of the rains      (Chapter 6)
     |                 |   ------ dry spells         (Chapter 6)
     |                 |   ------ the end & length of season  (Chapter 6)
366  |                 |   ------ a crop/water index          (Chapter 10)
      -----------------
```

Each of the chapters cited in Fig. 13.1a and others, have looked at **events** of importance for particular applications. As an example consider events concerned with the start of the rains. Fig. 6.3b is repeated as Fig. 13.1b. This shows how the date of the start of the season was calculated, for two definitions. The two columns of data that were produced, each have 11 observations. This is obvious, because there were with 11 years of data. Each of the derived columns contains one number for each year of data. In this case it is the date of the start.

Fig. 13.1b Simple start of the rains

File ⇒ Open From Library ⇒ Samsmall.wor
Climatic ⇒ Events ⇒ Start of the Rains
Manage ⇒ Column Properties ⇒ Format ⇒ x12 and x13 as Day of year



	X12	X13
1	17 Apr	12 May
2	27 Apr	10 May
3	24 Apr	12 May
4	04 Jun	04 Jun
5	25 Apr	03 May
6	13 May	13 May
7	22 Apr	02 May
8	05 May	05 May
9	27 Apr	16 May
10	13 May	13 May
11	27 Apr	06 May
12		

Figs 6.3j and 6.3k give further examples of alternative definitions for the start of the rains, while Figs 6.4a, 6.4c, 6.5a, 6.5d and 6.5g show the same feature, i.e. columns of length 11, for dry spells and for the end of the rains.

We call this the **direct method** of analysis. It is "direct" because, for any event of interest, its value each year is calculated. The next step is to analyse the subsequent columns of data. Thus, with any **events**, for example for Fig. 13.1b, the analysis starts with the daily data matrix containing about 366 * 11 = 4000 observations. From this matrix one number is extracted per year, for each event of interest, to give the columns of length 11. The subsequent analyses use the data that have been extracted.

There is nothing wrong with this approach, but it does have limitations that warrant consideration of an alternative method, described in this chapter. The rainfall data are extremely variable from year to year. This variability means that the extracted columns of data are also very variable. For example, consider the subsequent analysis of the data in X12 of Fig. 13.1b. The mean date of the start of the rains by this definition was day 123 (2nd May) and the standard deviation was 14 days, i.e. 2 weeks.

Now, moving to statistical inference, suppose the objective is to estimate the **true** mean date of the start, using the 11 years as a sample. The estimate is the sample mean, i.e. 2nd May. An estimate of its precision is also needed, and this is given by the standard error of the mean. The standard error is estimated to be **$s/\sqrt{n} = 14/\sqrt{11}$** = about 4 days. Hence 95% confidence limits for the true mean are roughly 2nd May ± 8 days.

The major problem with this simple method of analyses is the low precision of the estimates.

What can be done about this limitation? One solution is to insist on long records. WMO recommend 30 years at least and we have repeatedly stressed that the short record used in this guide was simply to illustrate the method of analysis. In the above example, if the same mean and standard deviation had been from a record of 56, rather than 11 years, then the standard error would have been $s/\sqrt{n} = 14/\sqrt{56}$ = about 2 days and the confidence limits would have been roughly 2nd May ± 4 days. With the luxury of 100 years of data the standard error would be less than 1.5 days. Hence, with long records, simple methods of analysis are adequate.

We do not always have such long records and therefore now consider what can be done to improve the precision of estimates. The case study in , Section 12.1, using data from Kurunegala, Sri Lanka, compared 5 methods of analysis. This is reviewed briefly here, to indicate the type of approach that is necessary.

In this case study there were 34 years of daily data and the task was to estimate the 30% of the weekly rainfall total, for each week of the year. There were therefore 52 events of interest and the first step was to calculate the weekly totals. This resulted in 52 columns of data, each of length 34, that were used in the subsequent analysis.

The first method was the simplest and was the same as described in Chapter 7 for decade and monthly totals. For any week, the observed 30% point of the data was calculated and used as the estimate. The resulting estimates were shown in Fig. 12.1a. Taking week 40 as an example, the estimate was 5.45mm and the methods described in Chapter 11, Section 2 are used in Fig. 13.1c to evaluate the 95% confidence limits.

From Fig. 13.1c, the approximate 95% points are 0.6mm to 20.6mm. These are very wide limits, demonstrating the imprecision of this simple method and Methods 2 to 5 in this case study showed how the situation could be improved. The main key to the improvements was to use data from neighbouring weeks to **smooth** the estimates.

The **Climatic ⇒ Process** dialogue, in Fig. 13.1d, was used for all these methods. Here we focus on the difference between Methods 4 and 5. In both cases, gamma distributions were fitted to the weekly totals, as shown in Chapter 11, Section 3. The two methods differed only on **when** the smoothing was applied. In Method 4 the estimates of the 30% points were calculated and then smoothed. Method 5 consisted of smoothing one stage earlier in the process, namely the estimates of the gamma distributions are smoothed. Then the 30% points were calculated and the earlier smoothing results in these estimates being smooth automatically.

If smoothing is such a good idea, could it be applied earlier in the process? In this chapter we smooth at the start, i.e. even before calculating the 7 day totals. Instead of calculating and then smoothing, we will smooth and then calculate.

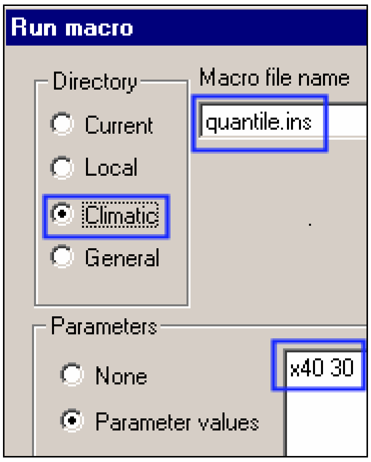
The benefits from this approach could be huge. What is proposed is to use the daily data to estimate the parameters of a model for the pattern of rainfall on a daily basis. Thus, even with just

11 years of data, there are 4000 observations to estimate the parameters of the model. This is a large amount of data, even allowing for the fact that perhaps 75% of the observations are zero.

Fig. 13.1c Confidence limits for the 30% point for week 40 from Kurunegala data

File ⇒ Open from Library ⇒ Kurun7.wor

Submit ⇒ Run Macro ⇒ quantile.ins

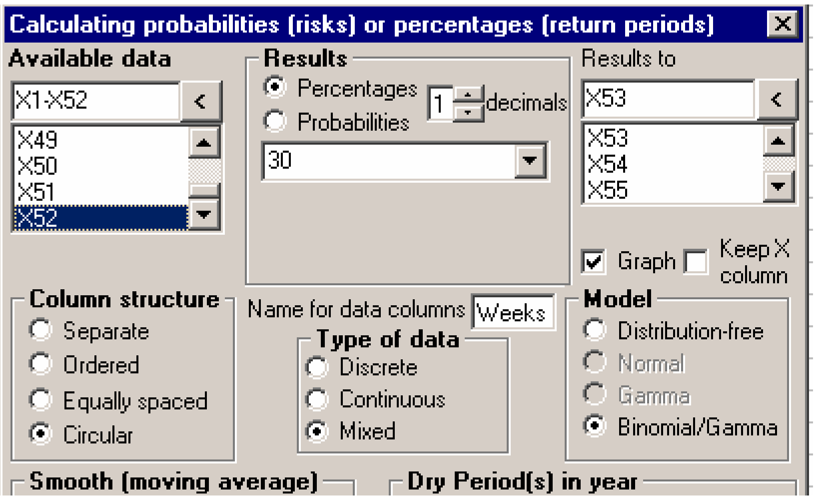


Row	Lo	Hi	%lo	%up	%	Dlo	Dhi
1	5	15	3.3	97.3	94	0.6	20.06
2	4	15	1.2	97.3	96.1	0.25	20.06
3	5	16	3.3	98.9	95.6	0.6	20.57
4	4	16	1.2	98.9	97.7	0.25	20.57

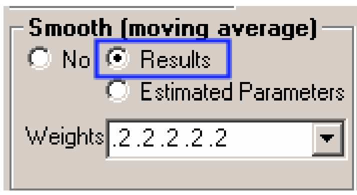
Estimate of 30% point
5.455

Fig. 13.1d Alternatives to estimate 30% points for case study in Chapter 12

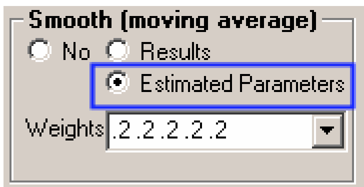
Climatic ⇒ Process



Method 4



Method 5



Another way to compare this proposal with the direct approach is as follows. The direct approach, as depicted in Fig. 13.1a, involves dipping repeatedly into the daily data matrix to extract a value each year, for each event of interest. This new proposal is to extract all the important information, which will be synthesised into one model. Then relevant results are extracted from the model. We will be using a "smooth" model and hence the results will also automatically be smooth.

We have said **the benefits could be huge**, but what exactly does this mean? In statistical terms, benefits are measured by comparing standard errors, and these can be translated directly into the years of data that are required for the same precision. The study by Sooriyarachchi (1989), on which the case study in Chapter 12 was based, quantified the benefits. Even conservatively, Method 5 was **more than twice as efficient** as Method 1, i.e. it needed less than half the years for the same precision. And the new Method 6, proposed here, is at least twice as efficient again. Hence, users should get the same precision from records of 20 years, using this new modelling approach, which need 100 years with the simple methods described earlier in this guide.

A further benefit of smoothing at the daily level is that the results for **any calculated events** will automatically be smooth. With other methods it is not clear, for example, how to get smoothed estimates for events such as the mean date of the start of the rains, from the data given in Fig. 13.1b.

The description above is intended to generate the interest for readers to continue with this chapter, but it should also have raised some questions. Two obvious questions are as follows:

- Are there no problems with the new method?
- If the benefits are huge, why aren't the methods used routinely? Why, for example, are they left to Chapter 13 of this guide?

These questions are considered in turn. The direct methods that have been used so far have the advantage that they are simple. They also make few assumptions. Their problem is the lack of precision, but the advantages of simplicity and lack of assumptions are also important.

Hence, the case study of Chapter 12, introduced complications in moving from Method 1 to Method 5. For example, for Method 5 some understanding is necessary of both the techniques of smoothing and of fitting gamma distributions. The results are also only useful if the assumptions are correct. Methods 4 and 5 estimates the 30% point assuming a gamma model for the weekly totals. If a gamma model is not appropriate, then the estimates are not so useful.

There are examples of inappropriate models being used. For example, in Kowal and Knabe (1974) a normal model was fitted to 10 day totals and 20% and 80% points of the decadal data were estimated from the models. The normal model was clearly not appropriate at the beginning and end of the rainy season and the resulting estimates are consequently of little value.

Here we believe that the gamma model was appropriate, and its goodness of fit was also tested. Hence, in Chapter 12, the gains in precision outweigh the complexity and extra assumptions and we therefore suggested that Method 5 be adopted. Similarly, there are many situations where the extra precision from the new methods proposed in this chapter will outweigh the disadvantages and hence where the new methods could usefully be adopted.

And so to the second question which is roughly "What is the catch? If the methods are so useful, why aren't they used more?"

There is no catch, or anything particularly controversial about the methods. They have been proposed by various authors and were written up in the agriculture literature in 1982, in the meteorological literature in 1983 and in the statistical literature in 1984. Some references are included at the end of this guide. They have been adopted, with one particularly usable method given by Woolhiser *et al*. (1989). Until recently however the methods have rarely been used perhaps primarily because they have not been made sufficiently "user friendly". This chapter and the corresponding facilities in Instat are part of our efforts to put this right. The Marksim software (ref) also uses the same ideas.

So, this is the last climatology chapter in this guide, users do need to have a good understanding of the simpler methods, covered in the earlier chapters, and to be motivated to consider an alternative approach.

13.2 Fitting a simple model

This section uses a simple example to illustrate the modelling approach. The stages in the analysis are the same as in the more complete examples considered later in this chapter.

The 11-year record from Samaru is used again and the task is to estimate the risk of a long dry spell after sowing. The result to be produced is the proportion of years with no long dry spell in the 30 days after sowing, in terms of the sowing date. The results in Fig. 6.5d showed that if a crop could withstand a 10 day dry spell (but no longer) then waiting until about day 118 (27th April) before sowing, would give an estimated risk of just 20%, i.e. one year in five, that the crop would fail. If the crop could not withstand a dry spell of more than 7 days then the earliest sowing should be about 10 days later, i.e. not before 7th May.

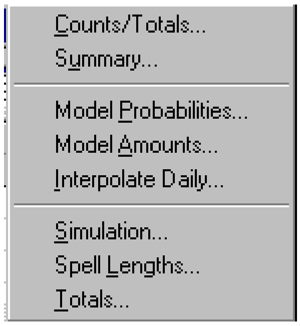
Here the modelling approach is used to produce equivalent estimates.

13.2.2 An example

There are three stages in the analysis. The first is to summarise the data, ready for the model fitting. This uses the **Climatic => Markov Modelling** menu (Fig. 13.2a).

Fig. 13.2a Markov modelling menu

Climatic ⇒ Markov Modelling



Stage 1: Preparation

Stage 2: Fit the model

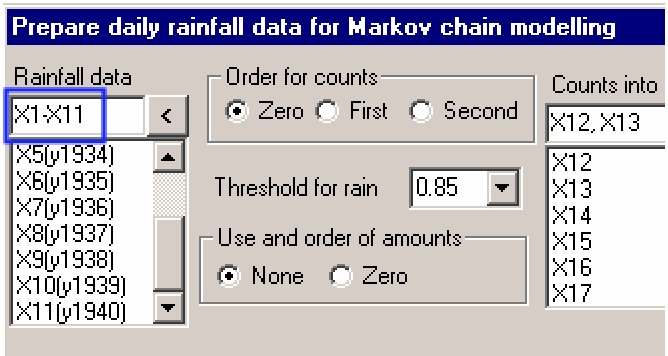
Stage 3: Use the fitted model

Open the worksheet **samsmall.wor**. The first menu option is **Counts/Totals**. Complete the dialogue as shown in Fig. 13.2b.

Fig. 13.2b Counts/Totals dialogue

File ⇒ Open (From Library) ⇒ samsmall.wor

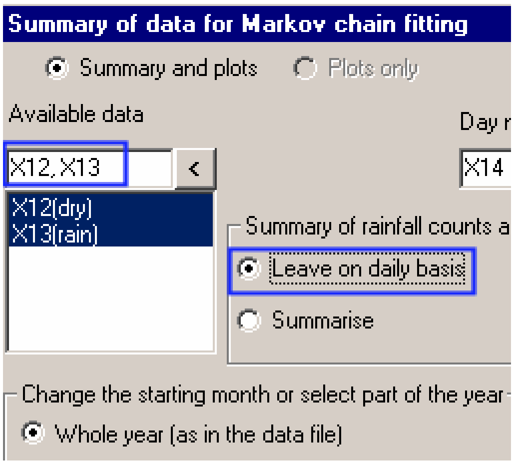
Climatic ⇒ Markov Modelling ⇒ Counts/Totals

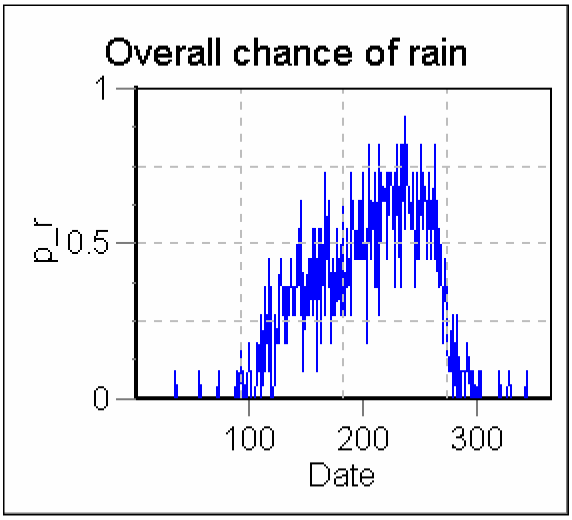


	X12	X13
	dry	rain
121	11	0
122	10	1
123	8	3
124	9	2
125	9	2
126	9	2
127	7	4
128	6	5
129	8	3
130	8	3
131	7	4

Fig. 13.2c Prepare dialogue

Climatic ⇒ Markov Modelling ⇒ Prepare





In this example, the **Threshold for rain** is left at 0.85, which defines all days with less than 0.85mm as dry.

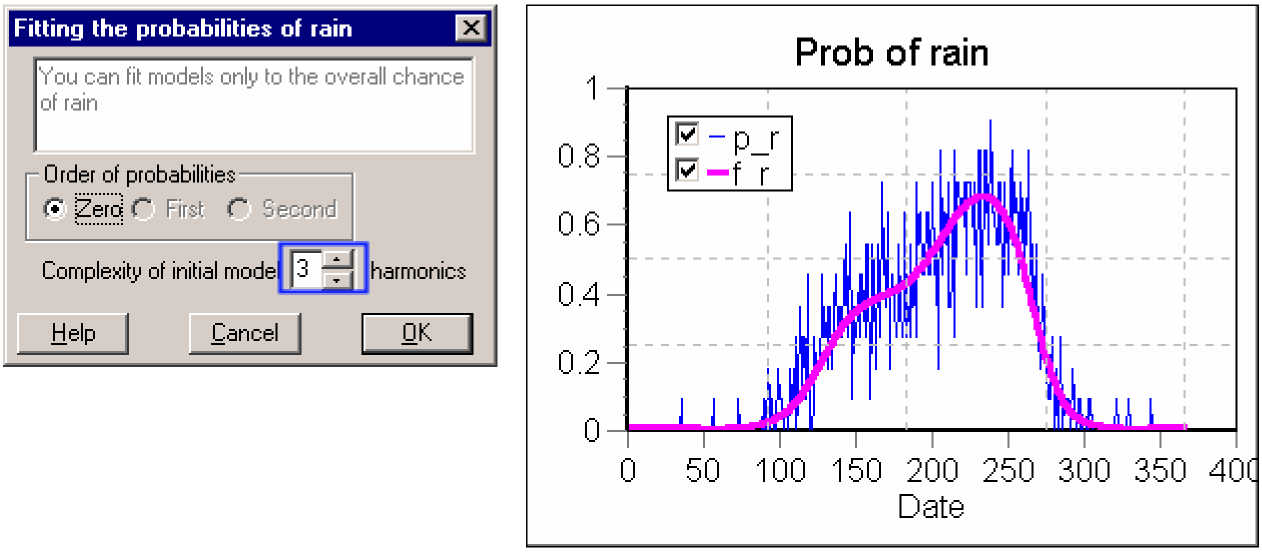
Part of the resulting data in X12 and X13 are shown in Fig. 13.2b. X12 is now named as '**dry**' and X13 as '**rain**'. They contain the number of occasions, in the 11 years, that each day is dry or rainy. For example, in Fig. 13.2b we see that day 122 (1st May) was rainy in only 1 of the 11 years, while 2nd May was rainy in 3 of the years.

The **Prepare** dialogue is the second in this stage. It sets up the columns needed to fit the models and also produces a graph of the chance of rain (Fig. 13.2c). Set the option in this dialogue to leave the data on a daily basis.

Stage 2 uses the **Model Probabilities** dialogue. You have to specify the complexity of the initial model and a previous fitting we found that it needed 3(Fig. 13.2d).

Fig. 13.2d Model Probabilities dialogue

Markov Modelling ⇒ Model Probabilities



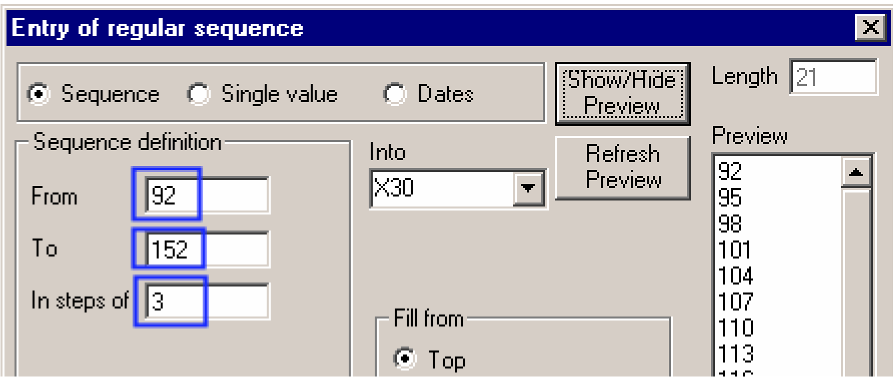
The plot in Fig. 13.2d is displayed. You are asked whether you wish to (A)dd or (D)rop terms from the model or stop running the model (Fig. 13.2f). As the 'f_r' line fits reasonably well. type **N**.

This concludes the second stage. To recap:

- The first stage was to prepare the data for the model fitting.
- The second stage was to fit the model. This produced the fitted probabilities, which are given in a column called 'f_r'.

Fig. 13.2e Enter data for spell lengths

Manage ⇒ Data ⇒ Enter Regular Sequence

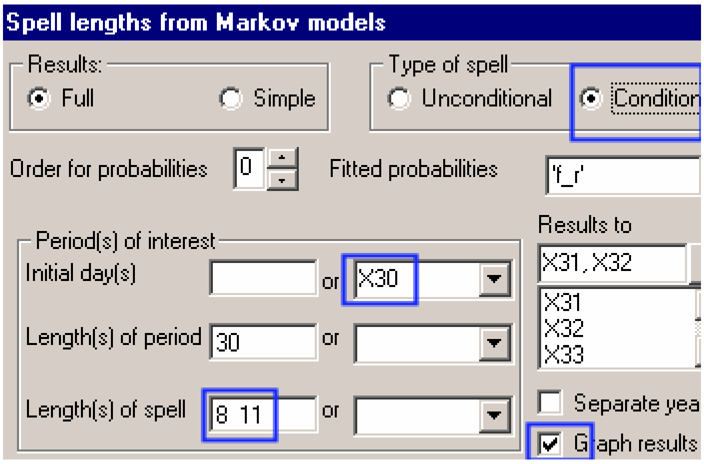


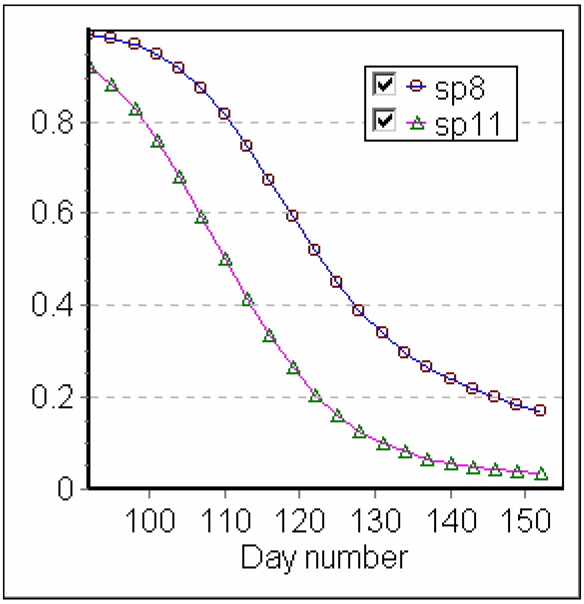
Stage 3 is to use the fitted model, here to estimate the risk of long spell lengths. Spell lengths are to be estimated for April and May. First enter the starting dates of the spells, from day 92 (April 1st) to day 152 (May 31st) in steps of 3 (Fig. 13.2e).

Complete the Spell lengths dialogue as shown in Fig. 13.2f. The spell lengths are **Conditional** on rain the day before, the **Fitted probabilities** calculated in Fig. 13.2c are in the column **'f_r'** and we wish to calculate over periods of 30 days for spell lengths of **8** and **11** days.

Fig. 13.2f The risk of a long dry spell in the 30 days following planting

Climatic ⇒ Markov Modelling ⇒ Spell Lengths





This third stage has illustrated one way that the models can be used.

13.2.3 More on the fitting process

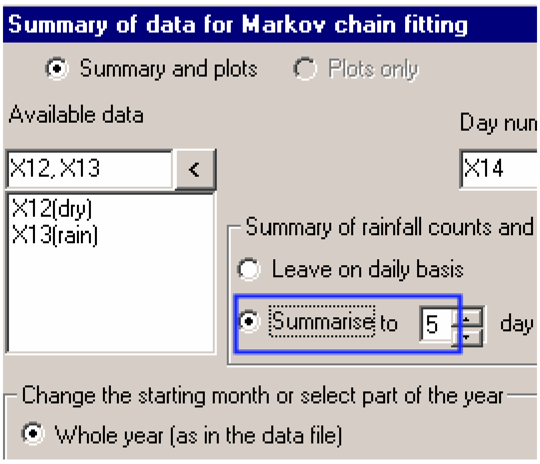
The analysis is now repeated, but fitting the model in a slightly different way.

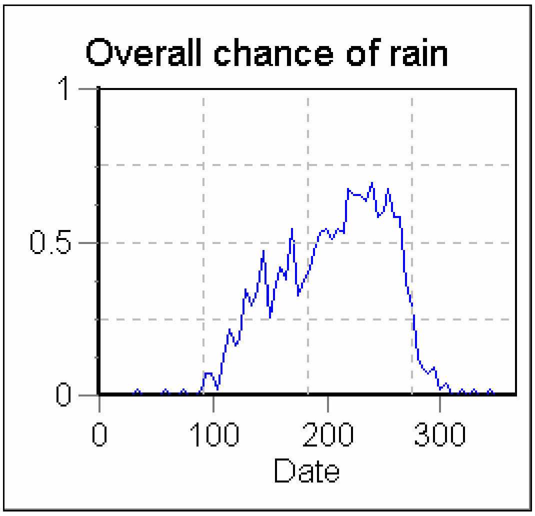
Fig. 13.2g Summary over 5 days

Manage ⇒ Remove(Clear) ⇒ x12-x60

Climatic ⇒ Markov Modelling ⇒ Counts/Totals as in ***Fig. 13.2b***

Climatic ⇒ Markov ⇒ Prepare



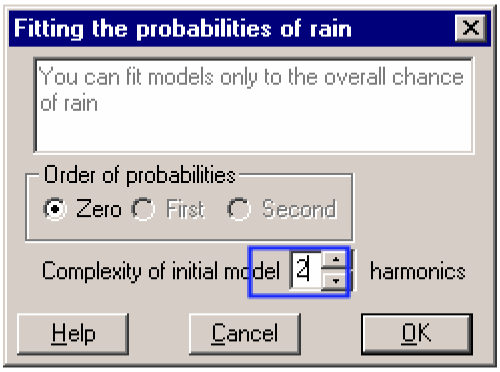


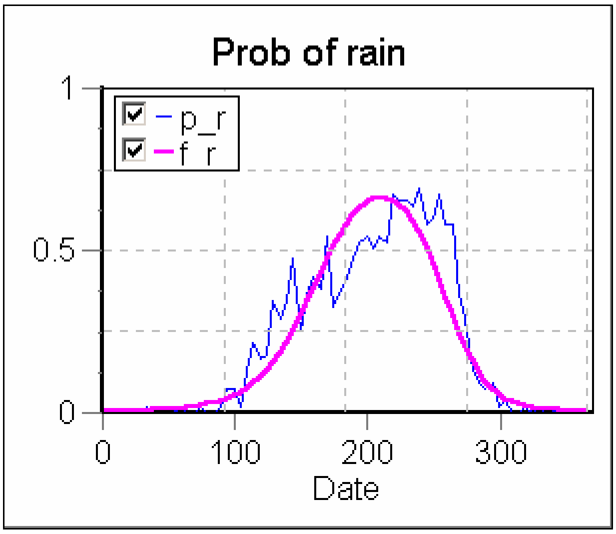
Begin by **Manage => Remove (Clear)** to clear the working columns, **X12-X60**, from the previous fit. Next use **Climate = Markov Modelling => Counts/Totals** again and accept the dialogue as shown earlier in Fig. 13.2b. Then complete the **Prepare** dialogue asking for a summary over 5 days (Fig. 13.2g).

The graph in Fig. 13.2g is similar in shape to that in Fig. 13.2c, but its shape is clearer. Next use **Model Probabilities** as shown in Fig. 13.2h, fitting the initial default of 2 harmonics.

Fig. 13.2h Fitting a model with 2 harmonics

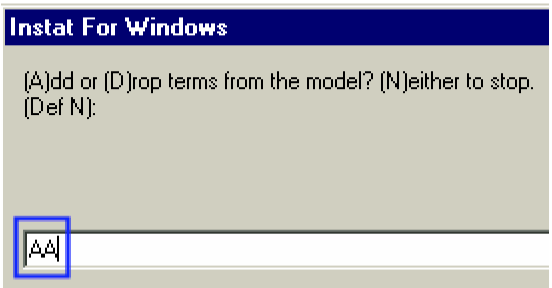
Climatic ⇒ Markov Modelling ⇒ Model Probabilities

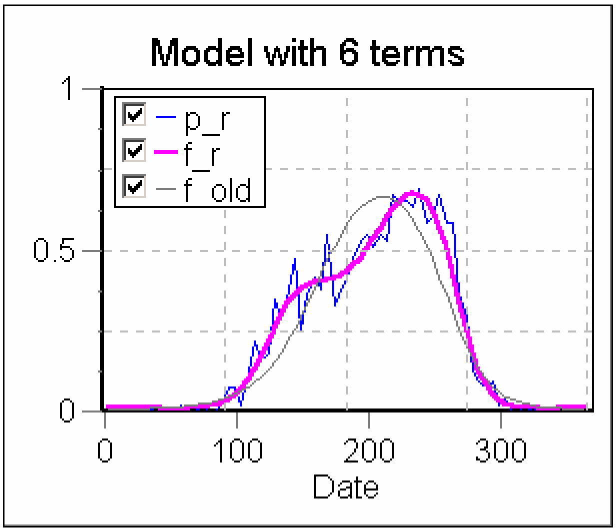




In Fig. 13.2h this model does not fit well. You are invited to (A)dd or (D)rop terms from the model. Typing "AA", Fig. 13.2i, adds two terms, which takes the model to three harmonics. (It add a sine and a cosine term to the regression equation.)

Fig. 13.2i Adding terms for a model with 3 harmonics, (6 terms)





The fitted curve, called "f_r" in Fig. 13.2i fits much better. Add a further harmonic, Fig. 13.2j, to check whether this fits even better.

The 4th harmonic makes almost no difference, so type DD, Fig. 13.2j to return to the provious model. Then type N, or just press OK to accept the model with 6 terms.

The fitted model shows that the chance of rain increases from less than 10% on day 100 (April) to a 70% chance of rain in August / September.

After this fitting on a 5-day basis, the **Interpolate Daily** dialogue, (Fig. 13.2k) must be used, before repeating **Stage 3** (Fig. 13.2e and 13.2f) to estimate the risk of long spell lengths,

Fig. 13.2j Checking on a model with 4 harmonics (8 terms)

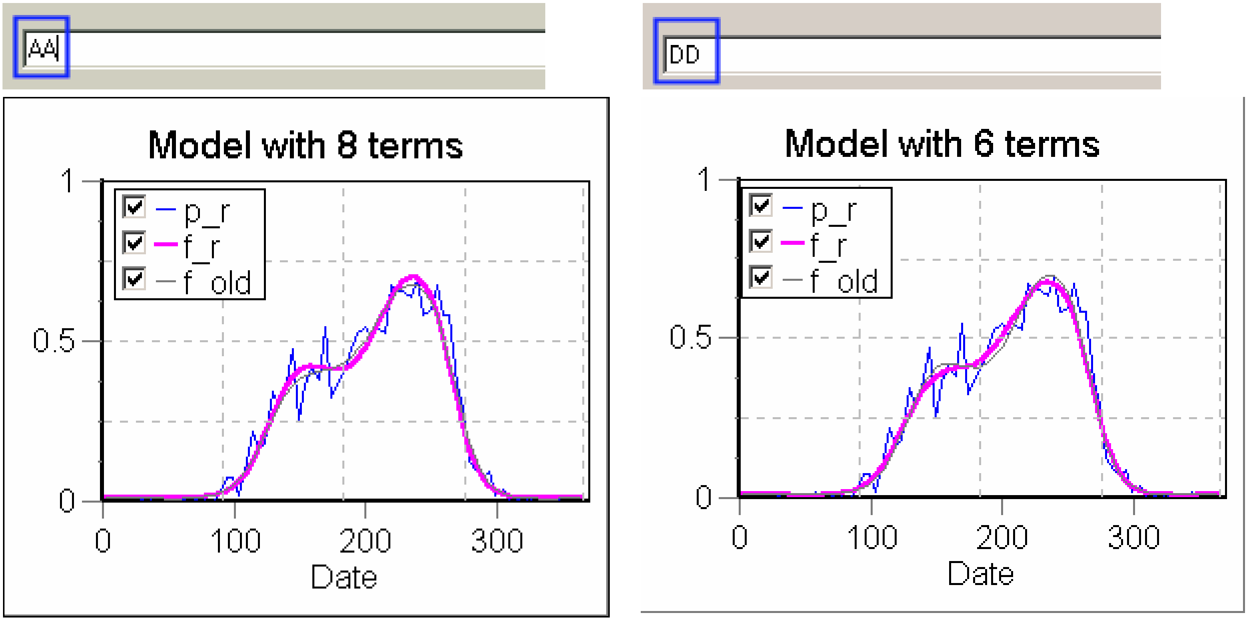
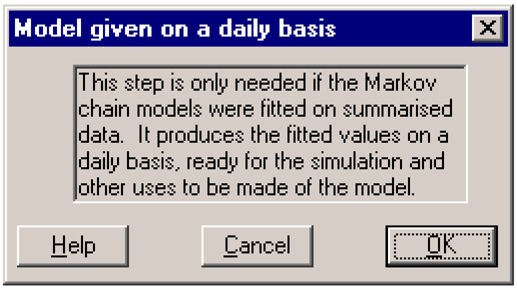


Fig. 13.2k Interpolate Daily dialogue



13.3 The preparatory stage

The previous section considered the fitting and use of a simple model. The process was split into three stages and these next three sections are each devoted to one stage. Here the preparation stage is described.

The first stage starts with the **Counts / Totals** dialogue. In the last section, in Fig. 13.2b , it was used to::

- put the number of dry days on each day of the year into x12
- and the number of rain days into x13.

Rain was defined as a day with more than 0.85mm.

Considering just the two categories of **dry** and **rain** is called a **zero-order Markov chain**. First-order or second-order models are often needed and are explained with examples in this Section.

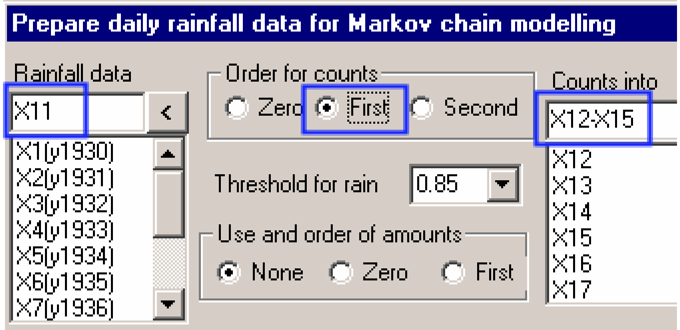
13.3.2 First and second order chains

For a first-order model the **Counts/Totals** dialogue is completed as shown in Fig. 13.3a. As this is just for explanation, only the data in X11 are summarised.

Some results are in Fig. 13.3a. x12 to x15 contain the 4 categories of day named 'dd, 'dr, 'rd and 'rr. Take day 156 (4th June) as an example. In this year it was dry and the previous day was also dry. It therefore gives a 1 in the category 'dd. The next day (5th June) was rainy with 15.24mm, with the previous day dry. It therefore gives a 1 in the category 'rd. The following day, 158, was dry, with the previous day rainy. It is therefore in the category 'dr. And so on.

Fig. 13.3a First order Markov chain for 1 year from samsmall.wor

Manage ⇒ Remove(Clear) ⇒ x12-x60
Climatic ⇒ Markov Modelling ⇒ Counts/Totals

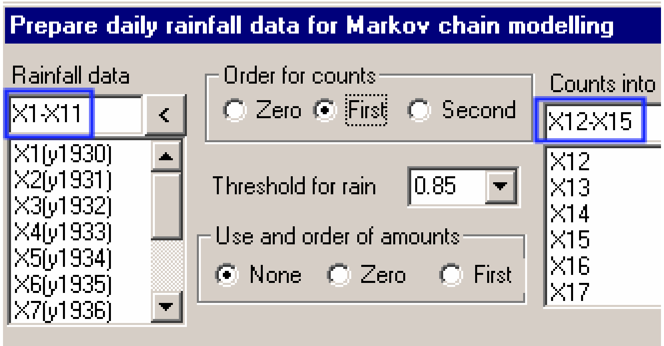


	X11*	X12	X13	X14	X15
	y1940	dd	dr	rd	rr
155	0	0	1	0	0
156	0	1	0	0	0
157	15.24	0	0	1	0
158	0	0	1	0	0
159	0.76	1	0	0	0
160	0	1	0	0	0
161	0	1	0	0	0
162	0	1	0	0	0
163	20.57	0	0	1	0
164	0	0	1	0	0
165	19.05	0	0	1	0
166	12.45	0	0	0	1

The dialogue and results for the 11 years are shown in Fig. 13.3b.

Fig. 13.3b Result of first order chain for 11 years

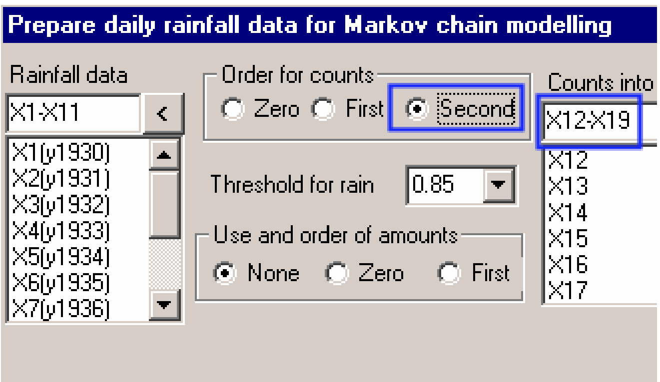
Manage ⇒ Remove(Clear) ⇒ x12-x15
Climatic ⇒ Markov Modelling ⇒ Counts/Totals



	X12	X13	X14	X15
	dd	dr	rd	rr
122	10	0	1	0
123	8	0	2	1
124	6	3	2	0
125	8	1	1	1
126	8	1	1	1
127	7	0	2	2
128	4	2	3	2
129	4	4	2	1
130	5	3	3	0
131	6	1	2	2
132	6	3	1	1
133	6	1	3	1

Fig. 13.3c Fitting second-order Markov chain for 11 years

Manage ⇒ Remove(Clear) ⇒ x12-X15
Climatic ⇒ Markov Modelling ⇒ Counts/Totals



	X12	X13	X14	X15	X16	X17	X18	X19
	ddd	ddr	drd	drr	rdd	rdr	rrd	rrr
122	10	0	0	0	1	0	0	0
123	8	0	0	0	2	0	1	0
124	6	0	2	1	2	0	0	0
125	6	2	1	0	0	1	1	0
126	7	1	1	0	1	0	0	1
127	6	1	0	0	2	0	1	1
128	4	0	0	2	3	0	2	0
129	3	1	3	1	1	1	0	1
130	2	3	2	1	2	1	0	0
131	3	3	1	0	2	0	2	0
132	5	1	2	1	1	0	0	1

Taking day 123 (2nd May) as an example, 8 of the years were in the category **'dd**, i.e. they were dry and the previous day was also dry. Of the 3 rainy days, 1 was in the category **'rr** and the other 2 were **'rd**. These data can be compared with those from Fig. 13.2b where there were just two categories, called 'dry (with 8) and 'rain (with 3). In Fig. 13.3b they have been subdivided into 4 categories, depending on the state of the previous day.

- So, the values earlier in 'dry are the sum of those in 'dd and 'dr.
- Those in 'rain are the sum of 'rd and 'rr.

Fig. 13.3c considers a Markov chain of second-order, i.e. looking back 2 days. Then there are 8 categories, ranging from **'ddd** (dry and the two previous days were also dry) to **'rrr**.

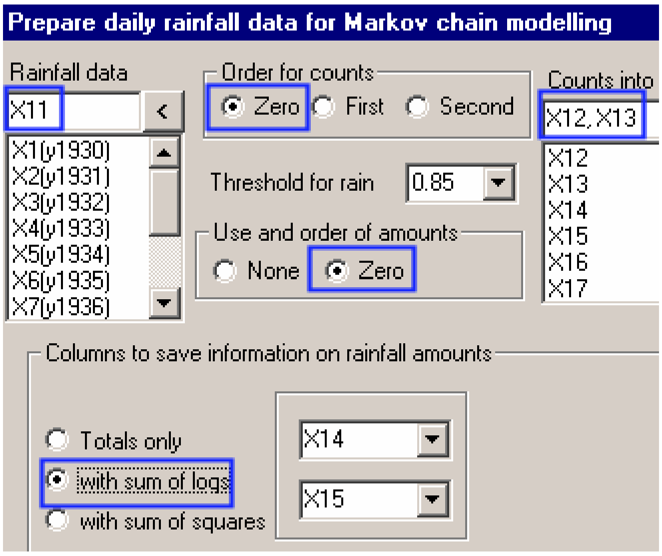
These summaries provide the information required to fit models to the **probabilities** of rain.

13.3.3 Rainfall amounts

The rainfall **amounts** on rainy days are also modelled. The Markov dialogue therefore also provides the initial summary of the amounts. Again, for explanation see Fig. 13.3d which first shows the information for a single year.

Fig. 13.3d Summary of the amounts of rain for 1 year

Manage ⇒ Remove(Clear) ⇒ x12-x19
Climatic ⇒ Markov Modelling ⇒ Counts/Totals



	X11*	X12	X13	X14	X15
	y1940	dry	rain	tr	lr
155	0	1	0	0	0
156	0	1	0	0	0
157	15.24	0	1	14.39	2.666533
158	0	1	0	0	0
159	0.76	1	0	0	0
160	0	1	0	0	0
161	0	1	0	0	0
162	0	1	0	0	0
163	20.57	0	1	19.72	2.981633
164	0	1	0	0	0
165	19.05	0	1	18.2	2.901422
166	12.45	0	1	11.6	2.451005
167	0	1	0	0	0

In Fig. 13.3d, take day 157 (June 5th) as an example. This had 15.24mm of rain. This value, has the threshold of 0.85mm subtracted, giving a value of (15.24-0.85)=14.39mm in the column named **'tr**. The shape parameter of the gamma distribution also has to be estimated. This needs the sum of the logs of the rainfall amounts, if the method of maximum likelihood is being used for the estimation. The sum of the logs, here ln(14.39)=2.667, is in the column called **'lr**.

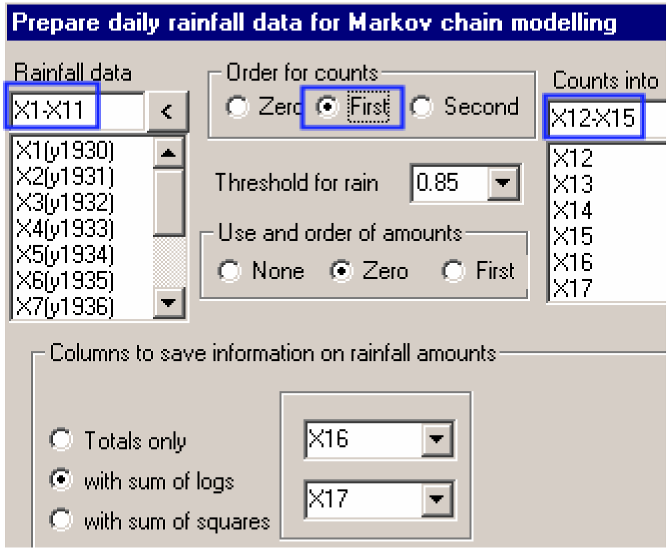
Fig. 13.3e shows the summary of the 11 years of data, requesting first-order counts plus zero order amounts and log amounts. Thus, with a threshold of 0.85, on day 123 there were 3 rainy days. The total rainfall on these three days was 29.7mm and is in the column called **'tr**. The corresponding sum of the logs (to base e) are in **'lr**.

The most complex model, with the current version of Instat is second-order occurrences of rain and first-order amounts.

Fig. 13.3e Summary of the first-order counts and the amounts of rain for 11 years

Manage ⇒ Remove(Clear) ⇒ x12-x15

Climatic ⇒ Markov Modelling ⇒ Counts/Totals



	X12	X13	X14	X15	X16	X17
	dd	dr	rd	rr	tr	lr
122	10	0	1	0	3.98	1.38
123	8	0	2	1	29.70	6.20
124	6	3	2	0	31.32	4.74
125	8	1	1	1	9.22	2.65
126	8	1	1	1	16.84	3.47
127	7	0	2	2	22.25	2.89
128	4	2	3	2	42.74	9.58
129	4	4	2	1	18.78	5.22
130	5	3	3	0	39.87	7.62
131	6	1	2	2	92.11	11.55
132	6	3	1	1	2.87	0.59
133	6	1	3	1	66.69	9.75
134	7	2	0	2	26.24	4.70

13.3.4 The Prepare dialogue

The second menu in the preparatory stage, calls the **Prepare** dialogue (Fig. 13.3f).

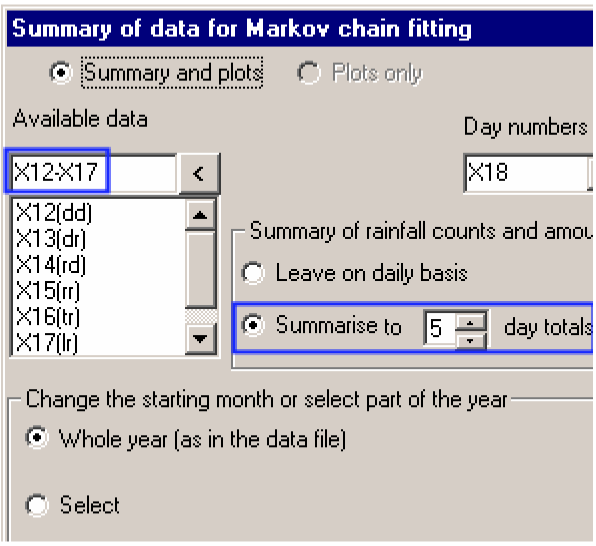
There are four possible reasons for using the **Prepare** dialogue:

- To set up a column called 'Date that gives the day number in the year
- To group the output from the Markov dialogue over a set number of days
- To change the starting date of the year
- To select a subset of the year for processing

Behind the scenes, this dialogue also sets up all the columns needed to fit the models.

Fig. 13.3f The Prepare dialogue sets up the columns for the model fitting

Climatic ⇒ Markov Modelling ⇒ Prepare



	X12	X13	X14	X15	X16	X17	X18
	dd	dr	rd	rr	tr	lr	Date
21	53	1	1	0	15.15	2.72	104
22	41	7	6	1	52.21	10.20	109
23	37	6	9	3	175.47	26.93	114
24	36	10	7	2	142.71	19.78	119
25	40	5	7	3	91.06	18.45	124
26	26	10	12	7	215.75	36.86	129
27	30	9	9	7	184.24	28.69	134
28	18	18	17	2	216.26	36.50	139
29	15	14	18	8	338.82	52.58	144
30	27	14	11	3	207.56	30.03	149
31	20	15	14	6	301.52	40.16	154
32	20	12	15	8	283.74	46.85	159
33	16	18	15	6	261.53	45.44	164

In Fig. 13.3f, the data have been grouped over 5 days and also shows some of the results. The columns of counts and the amounts have simply been totalled over successive 5 day periods and the **'Date** column has been generated that gives the average day number of each group of data.

Here, with a threshold of 0.85, the 25th row has a total count of 40 in the 'dd category, followed by 5, 7 and 3 in 'dr, 'rd and 'rr.

- Hence there were 10 (7+3) rainy days in this 5-day period, giving a proportion of 10/55=18% of rain days.
- The ‘Date column shows that the day number is around day 124 (3rd May). Thus, this is the 5- day period from 1 to 5 May.
- In the same row of data, the rainfall total is 91.06mm. This gives a mean rainfall per rainy day of 91.06/10 =9.1mm.

This grouping may seem odd, in a chapter on the modelling of the daily data. However the Markov dialogue has been used on the daily data and the grouped is afterwards. Some studies, e.g. Virmani et al. (1978) have grouped the data, over 7 day periods, **before** evaluating the probabilities. That is different, and is then considering wet or dry weeks.

There are two reasons for grouping before fitting the models. The first is that the observed data are then easier to compare visually with the potential models. On a daily basis, the observed proportions show a rather messy cloud of data, e.g. Fig. 13.2b. The second reason is that the approximate method of fitting the models, used in Instat, is more valid.

The next facility is to change the starting day in the year. Where there is winter rainfall, it may be convenient to consider the year from 1st September. This is not as crucial a step here as with the direct analyses, because the models fitted to the full year are **circular**. That is, they have the property of being **smooth**, over the end of the year. Hence, events that go over the end of the year can be considered from the fitted model.

Thus this change is needed only if it is more convenient to examine the potential models over the full season, or if the analysis is only for a part of the year and that part goes over the end of the calendar year.

The final facility is to select a subset of the year for the analysis. The analysis in Section 13.2 showed that this may not be necessary, even if the site is dry for part of the year. However, it is useful if only a part of the year is of particular interest. The model over the full year can be quite complicated and may not be an equally good fit for all parts. Concentrating on a few months may yield a simple model that is appropriate for a specific application.

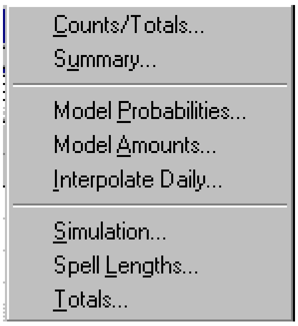
13.4 Fitting the Markov chain models

This section describes the facilities in Instat for fitting the models to the probabilities of rain and to the rainfall amounts. Technical details are minimised.

This stage uses the three dialogues shown in Fig. 13.4a under Stage 2. A model is first fitted to the chance of rain and then to the amounts. When the fitting is not on a daily basis, the third dialogue interpolates, so the fitted model gives the results for each day.

Fig. 13.4a Markov modelling menu

Climatic ⇒ Markov Modelling



Stage 1: Preparation

Stage 2: Fit the model

Stage 3: Use the fitted model

The objective is a model that encapsulates all the relevant information from the rainfall data. If successful, it should be such that simulated data, using the model, is indistinguishable in structure from real data.

The model has two components. The first component is the set of equations that fit the chances of rain through the year. In the simple model that was fitted in Section 13.2, there was just a single equation for the overall chance of rain. Here both the chance of rain given the previous day was dry and given it was rainy are considered. Thus there will be two curves to fit. At some sites it is necessary to consider the two previous days. Then there are four curves.

The fitting of a single curve corresponds to a **zero-order Markov chain**. A zero-order chain is one that has no memory. The fact that yesterday was dry does not affect the chance of rain today. A first-order chain has only one day of memory. If the chain is **first-order**, then the fact that yesterday was dry may affect (i.e. change the probability) that today is rainy. However, with a first-order chain, the extra information that the day-before-yesterday was also dry does not further change the probability of rain today.

With a second-order chain the memory extends two days, but no more. And so on. In the current version of Instat the Markov command is limited to second-order chains.

The chance(s) of rain change, depending on the day of the year. This is allowed for, by using Instat's regression facilities to fit curves to the probabilities. In the literature this is called a non-stationary chain. Many studies have avoided this by assuming constant probabilities for each period (usually a month). This is not desirable for most of the applications that are considered here.

While the first part of the model is for the chance(s) of rain, the second part is a model of the rainfall amounts on rain days. The simplest assumption is that the daily rainfall amounts follow an exponential distribution. The exponential distribution has just a single parameter, which is the mean. Hence, in this case, the second component of the model is the equation of the curve fitted to the mean rain on rainy days through the year.

There are two possible extensions. The first is that the distribution of rainfall amounts may be different on the first rainy day, compared to subsequent rain days. This is similar to the the chances of rain. If the memory extends one day, then two curves are fitted to the mean rain per rain day, depending on whether the previous day was dry, or not.

In principle, we could go further back, as with the probabilities. In the current version of Instat, the maximum complexity that is modelled is just a single day of memory, i.e. two curves, for the amounts.

The second extension models the rainfall amounts with a gamma, rather than an exponential distribution. (See **Chapter 11, Section 3** for a description of the gamma distribution.) Then the shape parameter has to be estimated as well as the mean. This is permitted within Instat, but with the limitation that only a single value for the shape parameter can be estimated for each curve, i.e. it is assumed constant through the year.

Therefore, the full model for a given site consists of:

- one or more curves for the chance(s) of rain,
- plus curve(s) for the mean rain per rain day,
- plus estimates of the shape parameter of the gamma distribution(s).

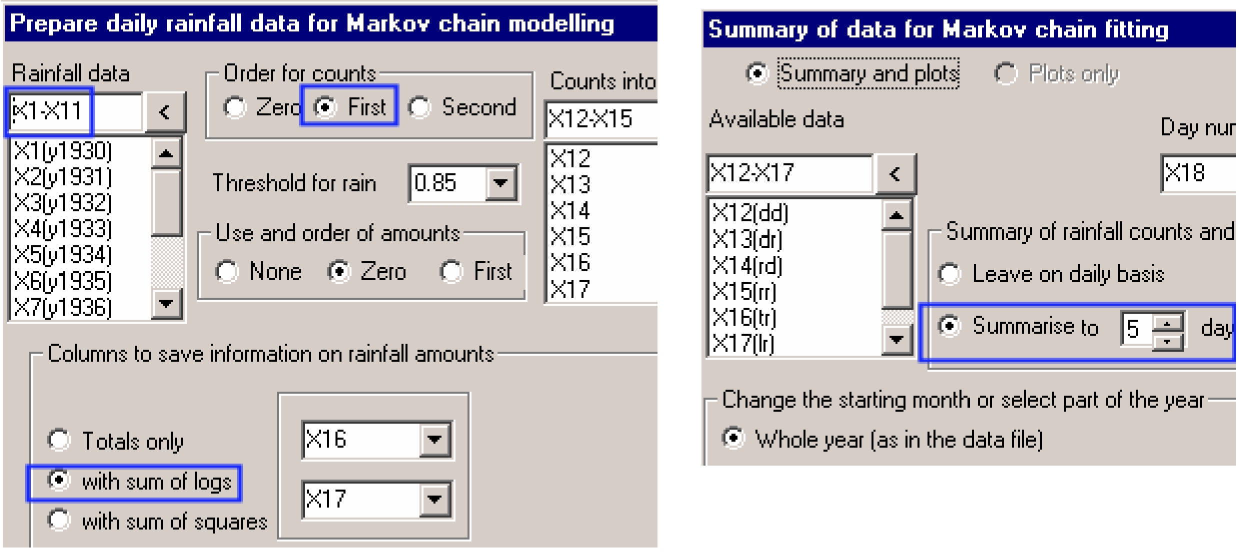
13.4.2 Example 1 – Data from Samaru

The first stage, described in Section 13.3 is repeated. Assume, as shown in Fig. 13.4a, that the Markov command is to give first-order counts, plus zero-order totals and log totals.

Fig. 13.4b The Markov and Prepare dialogues

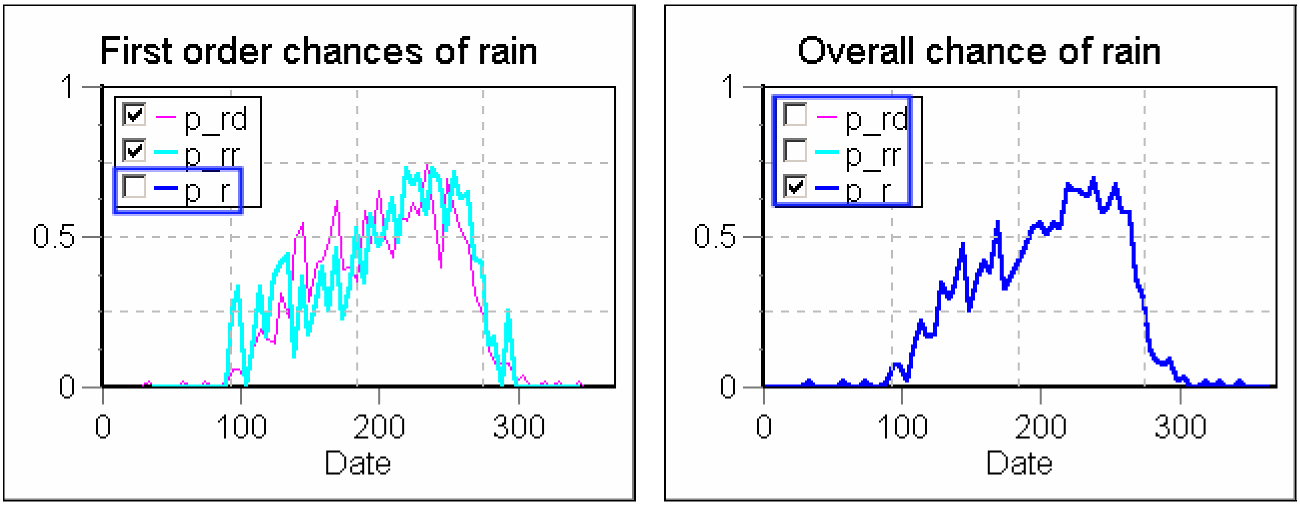
Manage ⇒ Remove(Clear) ⇒ x12-x60

Climatic ⇒ Markov ⇒ Counts/Totals **Climatic ⇒ Markov ⇒ Prepare**



Then use the **Prepare** dialogue to summarise the data on a five-day basis, to graph the chance(s) of rain, and to set up the variables needed for the fitting (Fig. 13.4b). The graphs are shown in Fig. 13.4c.

Fig. 13.4c Graph (split into two) to show the chances of rain

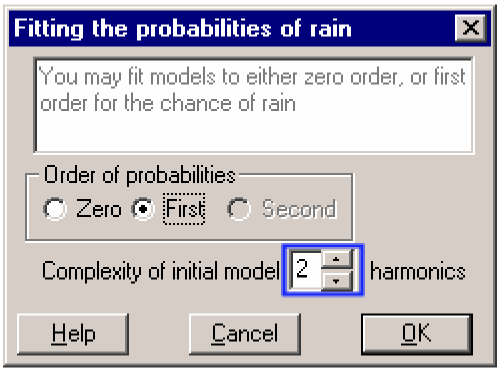


The second stage starts with the **Model Probabilities** dialogue. In this example fit first-order probabilities and accept the starting default of 2 harmonics (Fig. 13.4d).

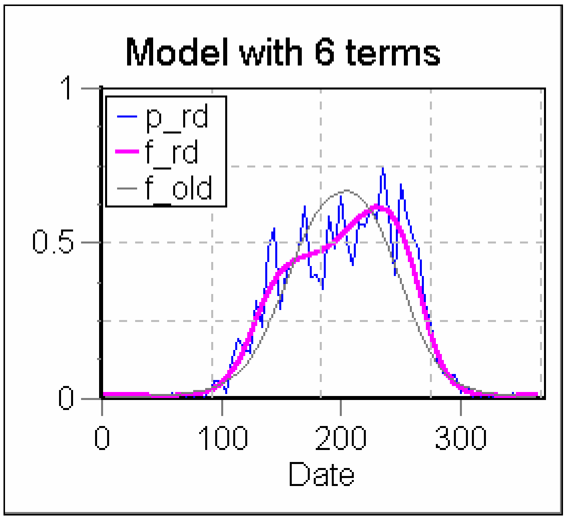
Add two terms to the model for the chance of rain given dry as shown also in Fig. 13.4c. Then model the chance of rain given rain. Here accept the initial model with 2 harmonics.

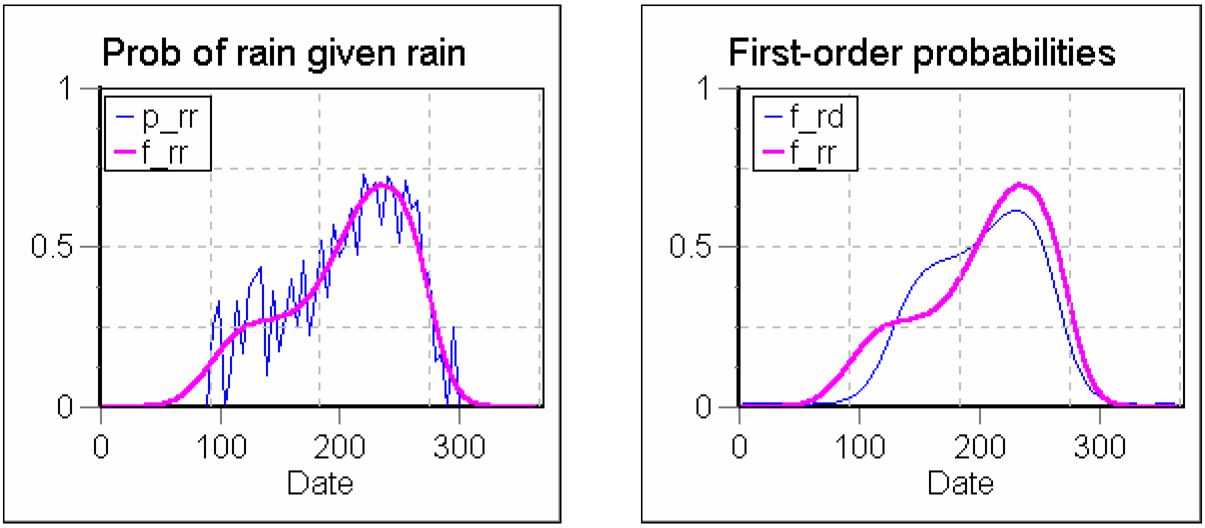
Fig. 13.4d Fitting first-order probabilities

Climatic ⇒ Markov Modelling ⇒ Model Probabilities



AA to add 2 terms to the first model.
Then **N** to accept model for **rain given dry.**
N to accept model for **rain given rain.**





This gives a model for the chances of rain, shown in the last plot in Fig. 13.4d.

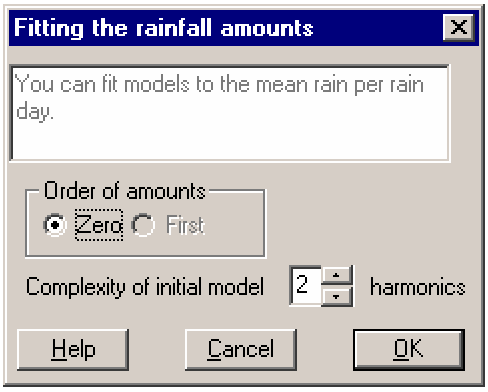
The next step uses the **Model Amounts** dialogue to display the mean rainfall(s) per rain day, set up the variables needed for fitting of the amounts and estimate the shape parameter of the gamma distribution (Fig. 13.4e).

Fig. 13.4e shows some of the results from running the **Model Amounts** dialogue. The output depends on the options for the amounts that were used with the **Counts / Total** dialogue (Fig. 13.4b).

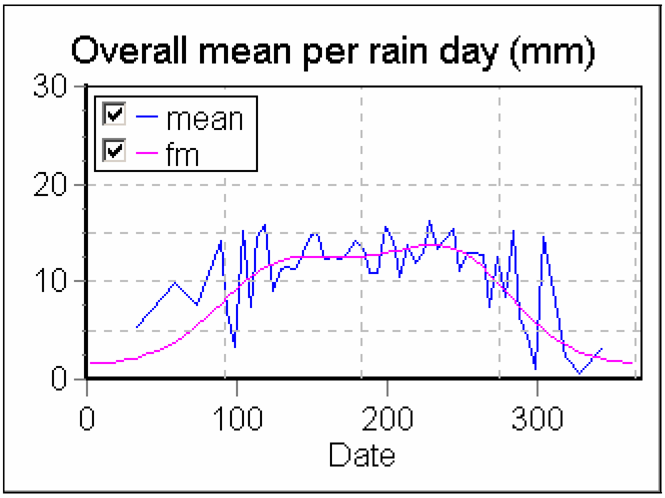
The shape parameter is estimated for the gamma distribution of amounts. The value here is 0.93, i.e. just less than 1 and this is typical for daily data. The simpler exponential distribution corresponds to a gamma distribution with k=1.

Fig. 13.4e Modelling amounts

`Climatic ⇒ Markov Modelling ⇒ Model Amounts`



`Overall maximum likelihood estimate of k = 0.931`



Use **`Manage => Worksheet Information`** to see the columns that have been generated by the fitting process, Fig. 13.4f.

Fig. 13.4f Column information

`Manage ⇒ Worksheet Information ⇒ Columns(X)`

Column	Na...	Length
X12	dd	73
X13	dr	73
X14	rd	73
X15	rr	73
X16	tr	73
X17	lr	73
X18	Date	73
X19	_d	73
X20	p_rd	73
X21	_r	73
X22	p_rr	73
X23	dry	73
X24	rain	73

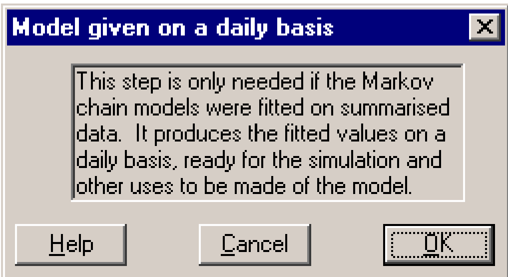
Column	Name	Length
X25	tot	73
X26	p_r	73
X27	circ	73
X28	ds1	73
X29	dc1	73
X30	ds2	73
X31	dc2	73
X32	ds3	73
X33	dc3	73
X34	ds4	73
X35	dc4	73
X36	ds5	73
X37	dc5	73

Column	Name	Length
X40	l_r	73
X41	f_rd	73
X42	f_old	73
X43	estrd	7
X44	f_rr	73
X45	estrr	5
X46	mean	73
X47	dev	73
X48	temp	73
X49	estk	1
X50	lm	73
X51	fm	73
X52	esta	5

The final step in this stage is to interpolate the results to a daily basis. This uses the **`Interpolate Daily`** dialogue as shown in Fig. 13.4g. A lot of calculation occurs "behind the scenes".

Fig. 13.4g Interpolate the data to a daily basis

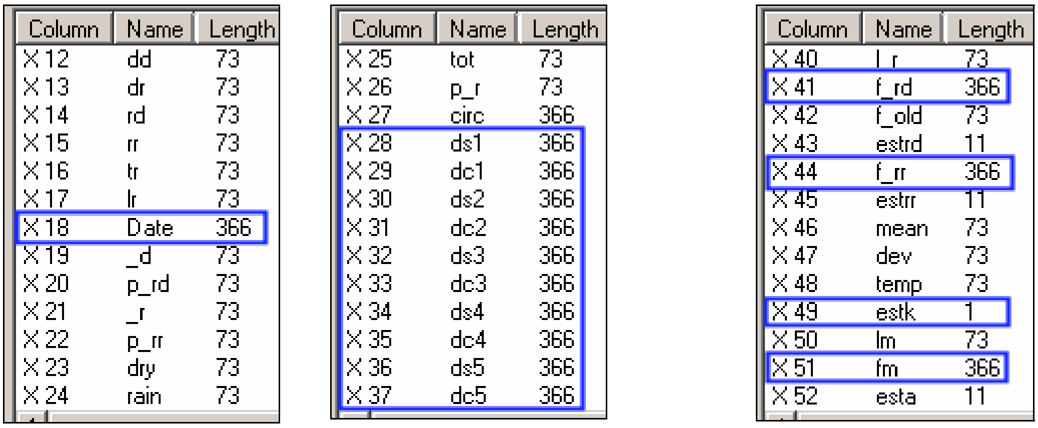
`Climatic ⇒ Markov Modelling ⇒ Interpolate Daily`



The main visible difference is that the fitted columns are changed from length 73 to length 366, compare Fig. 13.4h with Fig. 13.4f.

Fig. 13.4h Column information – for comparison with Fig. 13.4f

`Manage ⇒ Worksheet Information ⇒ Columns(X)`



The Markov Modelling has generated about 40 columns in the worksheet. The names should not be changed arbitrarily during the fitting process, otherwise the system will not work.

To recap, the system to fit the models is as follows:

- Use the **`Counts/Totals`** dialogue to run the Markov command.
- Use the **`Prepare`** dialogue to generate the columns for the fitting and possibly to give a summary (over 5 days).
- Use the **`Model Probabilities`** dialogue to model the chances of rain
- Use the **`Model Amounts`** dialogue to model the mean rain per rain day.
- Use the **`Interpolate Daily`** dialogue (if necessary) to convert columns to length 366.

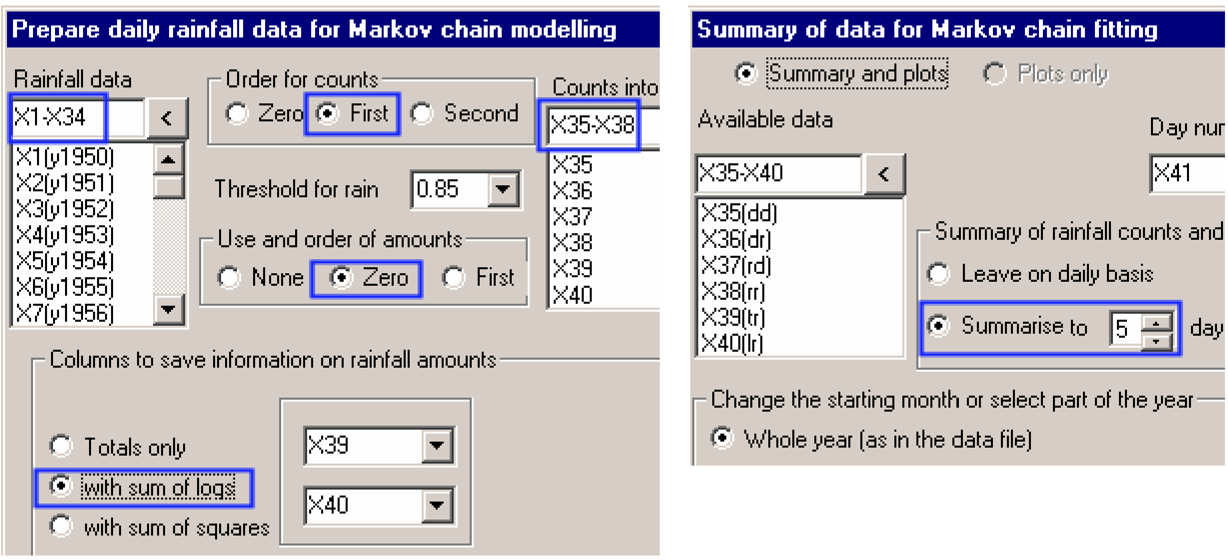
Save the **samsmall.wor** worksheet with the fitted model. It will be used in **Section 13.5**.

13.4.3 Example 2 – Data from Kurunegala

A second example uses the data from Sri Lanka that was analysed in the case study in Chapter 12 and mentioned in Section 13.1. The preparation stage is shown in Fig. 13.4i.

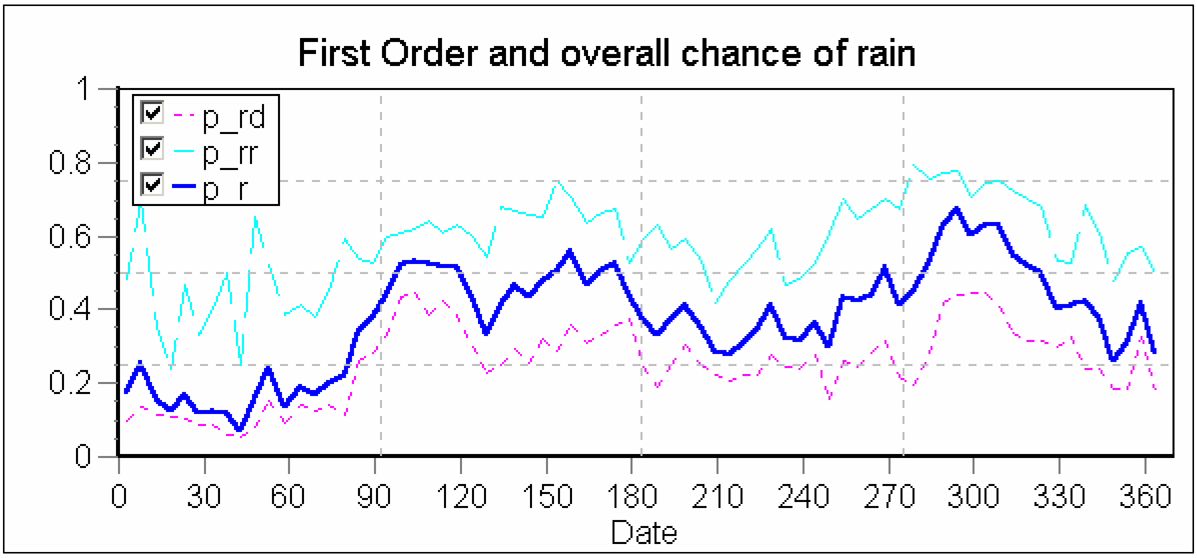
Fig. 13.4i Fitting a model to data from Kurunegala, Sri Lanka

`File ⇒ Open From Library ⇒ Kurunega.wor`
`Climatic ⇒ Markov Modelling ⇒ Counts/Totals`



The plot in Fig. 14.4j shows the model will be very different from that for Samaru. The pattern is clearly bimodal and the chance of rain given rain is much larger than that of rain given dry.

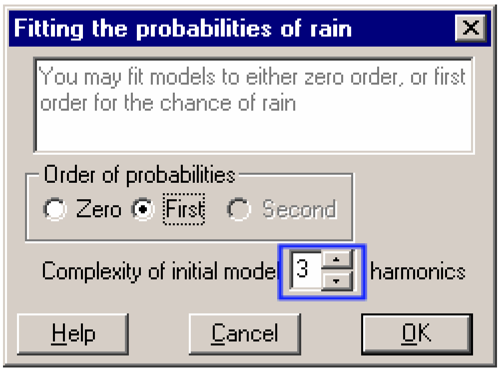
Fig. 13.4j Plot of the chances of rain

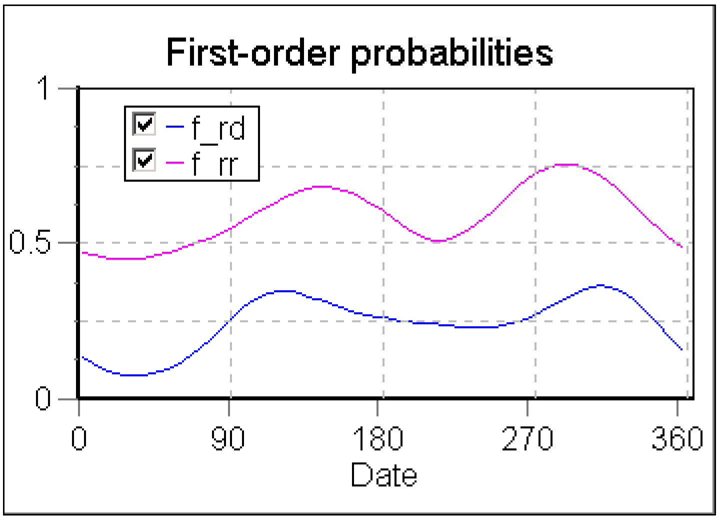


The model for the chance of rain is now fitted, starting with 3 harmonics, and accepting the initial model. Fig. 13.4k shows the fitted curves.

Fig. 13.4k Fitting a model to the probabilities of rain

Climatic ⇒ Markov Modelling ⇒ Model Probabilities

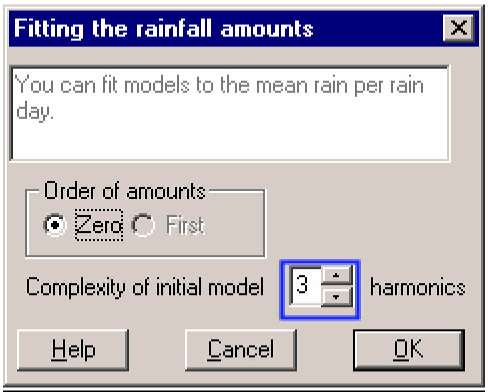


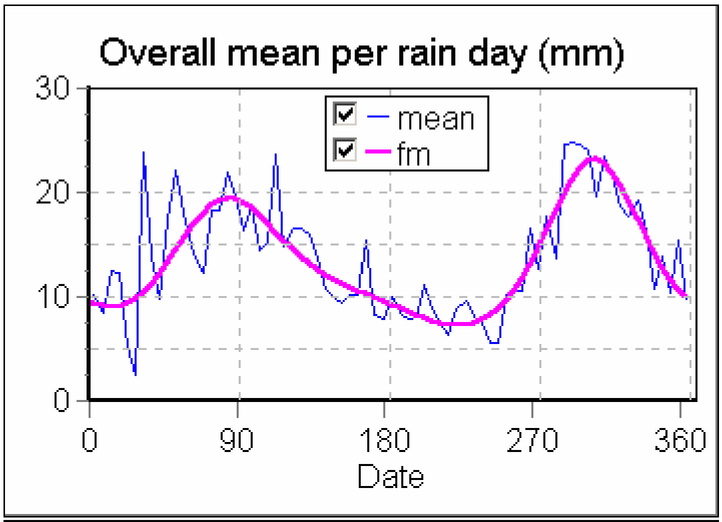


The model for the mean rain per rain day is fitted, again starting with 3 harmonics and accepting the resulting model. The fitted curve is shown in Fig. 13.4l, and the estimated shape parameter was estimated as 0.69.

Fig. 13.4l Model Amounts

Climatic ⇒ Markov Modelling ⇒ Model Amounts

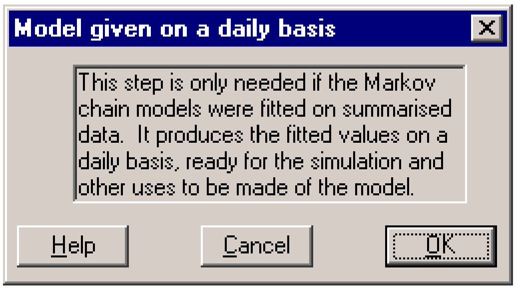




Finally use the Interpolate dialogue, Fig. 13.4m to put the fitted model onto a daily basis.

Fig. 13.4m Interpolate the data to a daily basis

Climatic ⇒ Markov Modelling ⇒ Interpolate Daily



Use **File => Save**, to save the **kurunega.wor** worksheet with the fitted model. It will be used in **Section 13.5**.

The ease of use of this system is intended to encourage users to consider adopting the modelling approach. With these examples of fitted models, we now explain why it should give more precise results than the direct approach, considered in previous chapters.

In the two examples considered in this section, the full model consists of about 15 coefficients, given in the equations for 'estrd, 'estrr and 'esta, plus a value for k. This is typical, in that an adequate model usually has between 12 and 30 coefficients. Now consider the situation with 40 years of daily data. This gives over 14,000 daily values from which to derive the estimates. Thus the curves are fitted precisely and the resulting estimates will therefore also be precise.

What though of the situation, such as in this section, where we only use a short record from Samaru? As shown here, it is still not difficult to fit a specified model. What is more difficult, with short records, is to assess the complexity of the model that is required. What we recommend therefore is as follows. In any given region, use the long records to evaluate what type of model is needed, e.g. first or second order. Then, with the short records, from neighbouring stations, assume generally that the same complexity can be used.

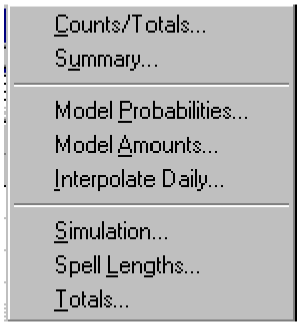
What are the risks from this modelling approach? They are basically that the model is not appropriate. Then results have great precision, but they are wrong! There are various reasons why the model could be inappropriate. It could be that a Markov chain type of model should not be fitted; or the class of model is suitable, but the model is of inappropriate order or complexity. Thus, as with any modelling approach, the model should be chosen with care. This is easier if the fitting process is interactive, as is the case within the system developed here.

13.5 Using the models

The third stage is to use the models. There are 3 dialogues in the **Climatic ⇒ Markov Modelling** menu as shown in Fig. 13.5a.

Fig. 13.5a Markov modelling menu

Climatic ⇒ Markov Modelling



Stage 1: Preparation

Stage 2: Fit the model

Stage 3: Use the fitted model

- **Simulation** - used to simulate rainfall data from the model. If just the probabilities of rain are given, then the rainfall occurrence will be simulated. If the means through the year are also given, then the rainfall amounts will be simulated.
- **Spell Lengths** - the modelling equivalent to the **Climatic => Events => Spells** dialogue. It uses the model of rainfall occurrence, i.e the probabilities of rain, and gives the probabilities of dry spells of specified lengths. It can give either conditional or unconditional probabilities.
- **Totals** - used to give percentage points and probabilities for the total rainfall in specified 5, 7 or 10 day periods.

The last two dialogues both use **recurrence relations**. They derive **exact** results from the model. Often, however the events to be studied, or the model that is fitted, are too complex for the recurrence relations. Then the simulation dialogue is used.

13.5.2 Example 1 – Samaru

For the first example the columns that are needed from the model are transferred, into a new worksheet, Fig. 13.5b. This step is partly to simplify the explanations. It is often useful, but is not an essential step.

The 3 dialogues need the fitted probabilities and (possibly) the mean rain per rain day for each day of the year. If gamma models are used for the rainfall amounts, then they also need a column with the estimate(s) of the shape of the gamma distributions.

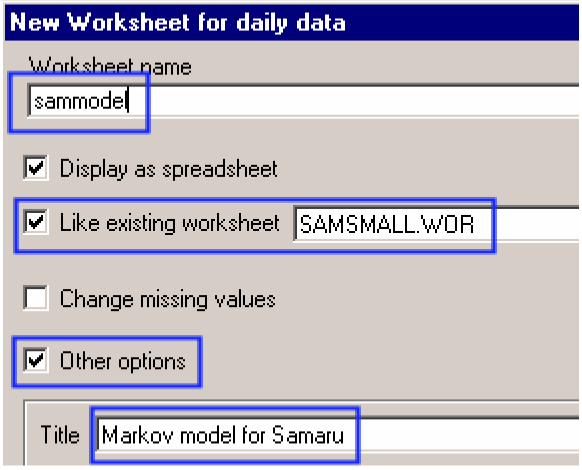
Fig. 13.5b Data columns with the fitted model

File ⇒ Open Worksheet ⇒ Samsmall

Find columns containing the fitted model (brackets show ours)

'Date(X18) 'f_rd (X41) 'f_rr (X44) 'fm (X51) 'estk (X49)

Climatic ⇒ Manage ⇒ New **Manage ⇒ Data ⇒ Duplicate**



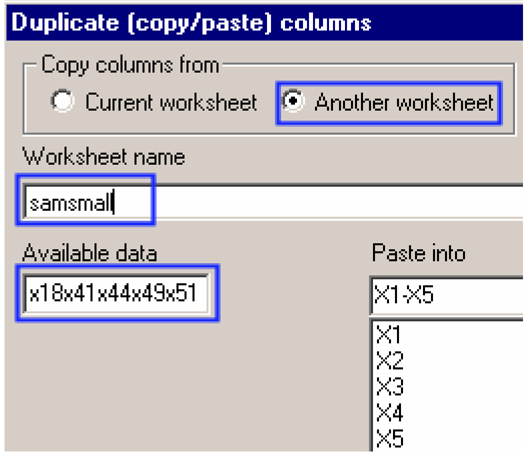
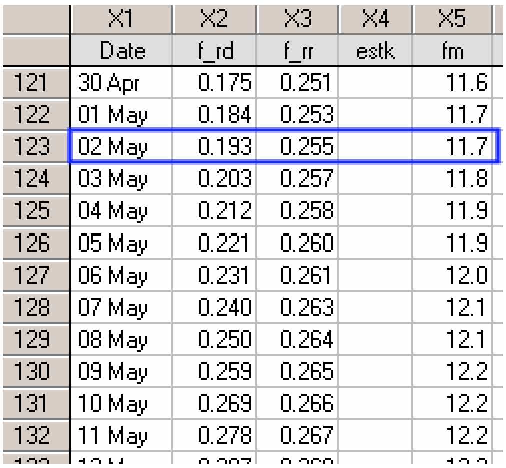


Fig. 13.5c shows the fitted model for a few of the days. For example, on 2nd May, the estimated chance of rain given the previous day was dry is 0.193. Similarly 'f_rr is 0.255 on that day and the mean rain, when 2nd May is rainy, is 11.7mm. You can check the information in the new worksheet is sufficient for the model, by plotting the probabilities and means against the date column, see for example Fig. 13.5c. This gives the same plot as shown earlier in Fig. 13.4d .

Fig. 13.5c Model in new worksheet



	X1	X2	X3	X4	X5
	Date	f_rd	f_rr	estk	fm
121	30 Apr	0.175	0.251		11.6
122	01 May	0.184	0.253		11.7
123	02 May	0.193	0.255		11.7
124	03 May	0.203	0.257		11.8
125	04 May	0.212	0.258		11.9
126	05 May	0.221	0.260		11.9
127	06 May	0.231	0.261		12.0
128	07 May	0.240	0.263		12.1
129	08 May	0.250	0.264		12.1
130	09 May	0.259	0.265		12.2
131	10 May	0.269	0.266		12.2
132	11 May	0.278	0.267		12.2

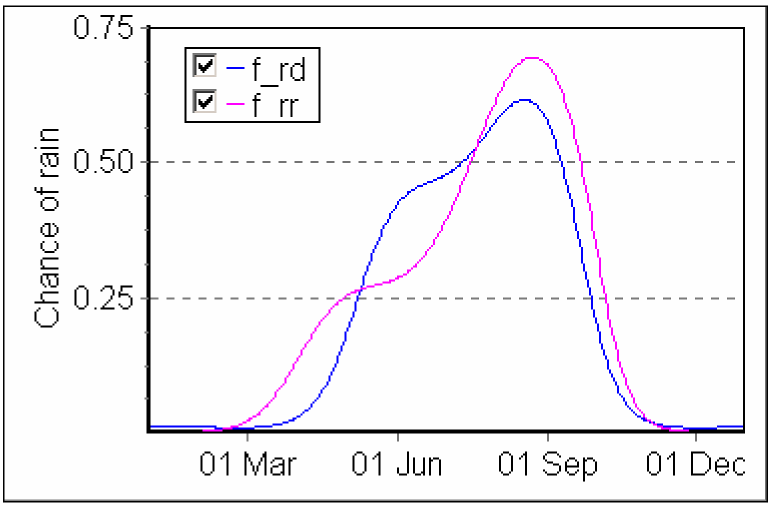


Fig. 13.5d gives the column names for the different types of model that can be fitted. These names are optional. The 3 dialogues, described in this Section, assume those names by default. If there are no names, or the names are different, then users must specify which column corresponds to each element of the model.

Fig. 13.5d Names for different types of models

Probability of rain	**Column names**
Zero order	f_r
First order	f_rd f_rr
Second order	f_rdd f_rdr f_rrd f_rrr
Amounts	
Zero order	fm
First order	fmd fmr
Shape parameter	estk

In the list of dialogues in Fig. 13.5a, **Simulation** is often needed and is therefore given before the other two dialogues. However simulation is a last resort and the **Spell Length** and **Totals** dialogues are to be preferred when they can be used. They give numerical results that correspond to simulating an infinite number of years.

The first examples use the **Climatic ⇒ Markov Modelling ⇒ Spell Lengths** dialogue. Start with the model for Samaru, given in Section 13.4, (or using Fig. 13.5b above), and calculate the risk of a dry spell of more than 7 or 10 days in the 30 days following a rain day. These are the same events that were considered for the longer record from Samaru in Chapter 7.

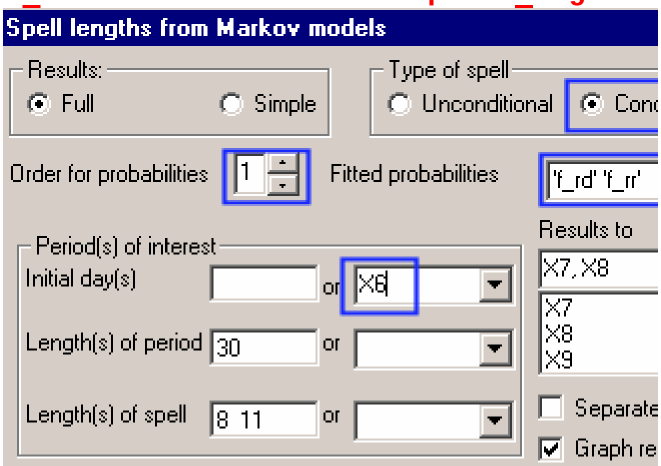
The dialogue and output are in Fig. 13.5e. If the model was not copied into a new worksheet, then different columns to X6, X7 and X8 shown in Fig.13.5e will be used.

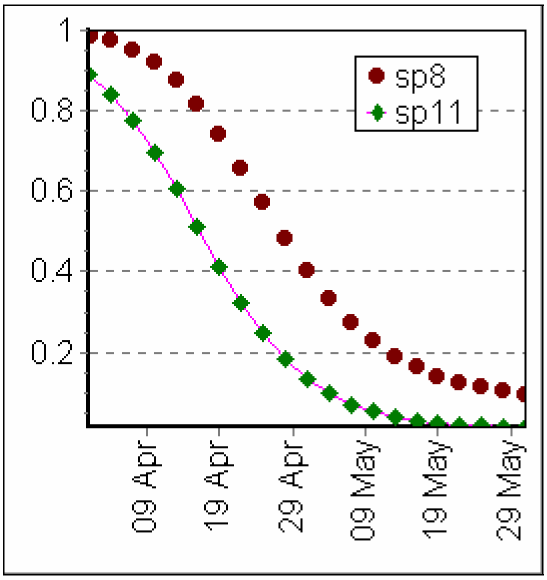
Fig. 13.5e Estimating dry spells in the 30 days following a rain day

Manage ⇒ Data ⇒ Regular Sequence

`Enter data from` **92 to 153** `in steps of` **3** `into` **X6**

Climatic => Markov => Spell Lengths





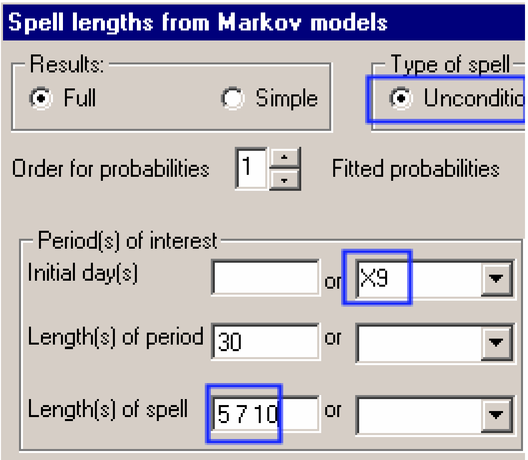
The results show that the chance of a dry spell of more than 7 days, within the 30 days following planting, has dropped to about 0.4 by the beginning of May and is below 0.2 by 19th May. This type of result can help to determine planting strategy.

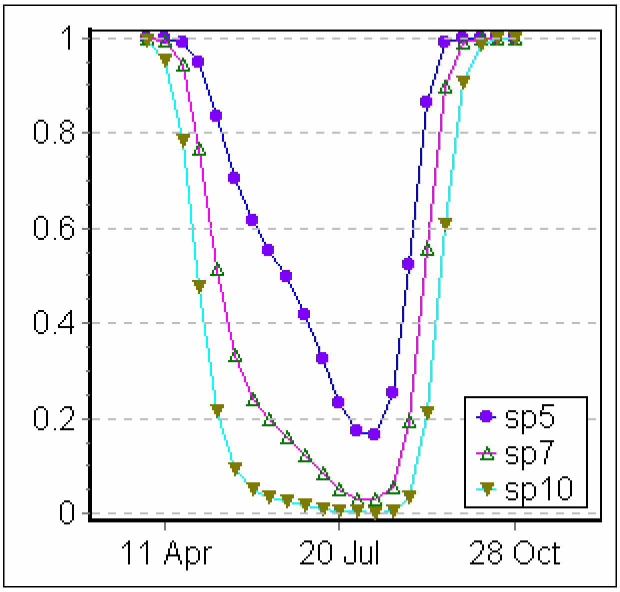
Fig. 13.5e plots the results as points as well as with a line. The points emphasise that they have been evaluated for each of the 21 starting dates in x6. The probabilities change smoothly, because the fitted model is smooth.

The overall risk of dry spells through the season can also be calculated, as illustrated in Fig. 13.5f. If a crop is sensitive to drought at a particular growth stage, then this type of plot can help to time planting to minimise such risks. The vertical reference lines on the graph correspond to the 1st May, 1st July and 1st September. Using the second reference line, the graph shows, for example, that the chance of a 10 day dry spell in July is very low, but the chance of a 5 day dry spell is about 0.5. Sometimes dry spells are needed for drying purposes and this type of plot also allows the timing of such events to be planned.

Fig. 13.5f Unconditional risk of dry spells in the following 30 days through the season

`Manage ⇒ Data ⇒ Regular Sequence`
`Enter data from 92 to 305 in steps of 10 into X9`





13.5.3 Example 2 - Kurunegala

The **Totals** dialogue uses similar recurrence relations to those for **Spell Lengths**. It gives the distribution of the number of rain days in each period. Percentage points of the total rainfall in each period can also be calculated.

The model for the occurrence of rain can be zero, first or second-order. The model for the amounts is more limited. Only zero-order amounts can be used. Also a single 'average' mean is calculated for each period. This will normally be adequate if results, i.e. probabilities or percentage points, are required for 5, 7 or 10 day periods. The command can be used with periods as long as a month, but the results should be treated with caution, particularly at the start and end of the season, when the mean rain per rain day often changes markedly through the month.

The example is a model from Kurunegala, Sri Lanka, for which a first-order Markov chain model was fitted in Fig. 13.4k and 13.4l. These are the same data used in the case study in Section 12.2, where 5 methods of analysis were compared. The results here show Method 6.

The Totals dialogue needs a column that give the number of days in each week of the year, and one with the starting day number of each week. There are various ways this can be done. One way is shown in Fig. 13.5g. It produces the required columns in x78 and x79.

Fig. 13.5g Enter data to use in Totals dialogue

File ⇒ Open Worksheet ⇒ Kurunega.wor
(assume the model fitted as in Section 13.4 – using columns up to x74)

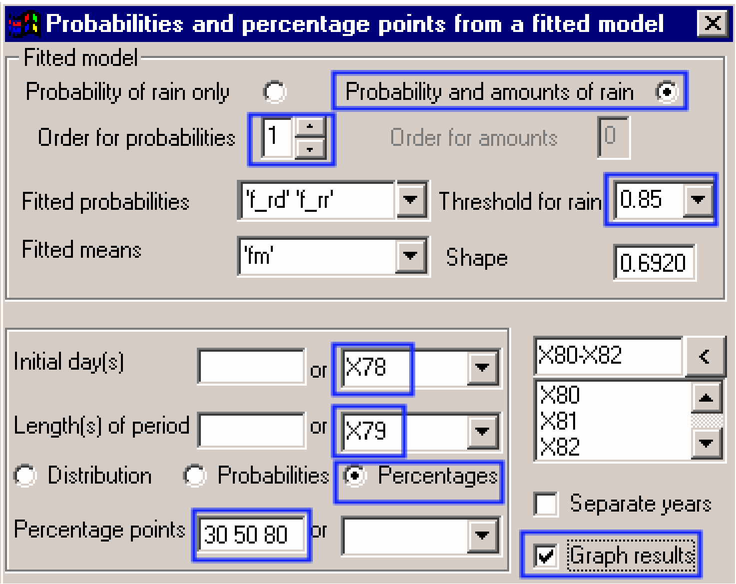
```
remove x75-x79

note x78 and x79 are of length 52
note x78 is starting day number of each week
note x79 is the number of days in the week

note x75-x77 are working columns of length 366

enter x75; data (1]366)
make x76;week
x77 = diff(x76)
note x77 = 1 for each row that starts a week
note x77 = 0 for the other rows

select x75;into x78;if (x77>0)
stats x76;by x76;counts into x79
```

	X75	X76 - F	X77	X78	X79
1	1	1	1	1	7
2	2	1	0	8	7
3	3	1	0	15	7
4	4	1	0	22	7
5	5	1	0	29	7
6	6	1	0	36	7
7	7	1	0	43	7
8	8	2	1	50	7
9	9	2	0	57	8
10	10	2	0	65	7
11	11	2	0	72	7
12	12	2	0	79	7
13	13	2	0	86	7
14	14	2	0	93	7
15	15	3	1	100	7

The columns calculated in Fig. 13.5g are now used in the **Climatic ⇒ Markov Modelling ⇒ Totals** dialogue as shown in Fig. 13.5h. The 30% point is calculated and also the 50% and 80% points of the weekly totals.

Fig. 13.5h Estimating 30% 50% and 80% points of weekly rainfall totals

Climatic ⇒ Markov Modelling ⇒ Totals




```
% points from a fitted model

Row    30%    50%    80%

  1    0.0    5.7    26.0
  2    0.0    2.9    20.8
  3    0.0    1.4    18.1
 ..    ...    ....   ....
 12    9.1   28.0    74.6
 13   14.5   35.0    82.6
 14   19.2   40.3    88.5
 ..   ....   ....    ....
 51    5.5   16.7    44.3
 52    3.4   13.0    37.6
```

The estimates are given in the output window, and in the worksheet. For example, from Fig. 13.5h, in week 12 the 30% point is estimated to be 9mm, and the 50% and 80% are 28 and 75mm.

The results are also plotted in Fig. 13.5i. They look reasonable, but it would also be useful to compare them with the simpler methods of estimating the 30% and other percentage points, used in Chapter 12.

Fig. 13.5i The estimated percentage points

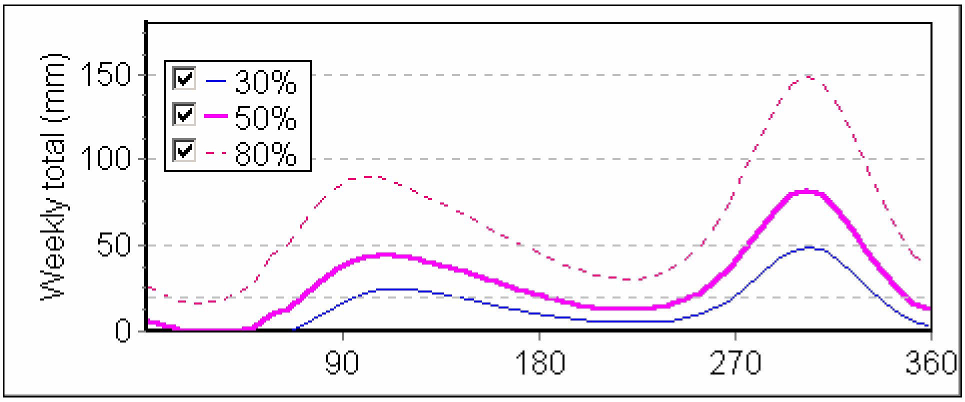
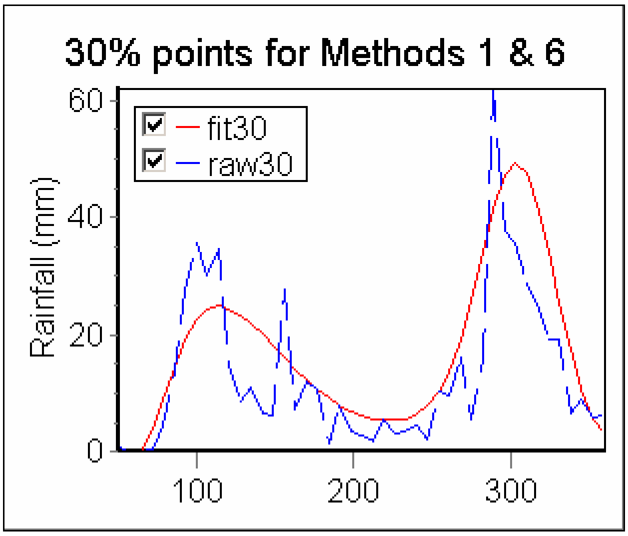


Fig. 13.5j shows the instructions for the simplest method again, for the 30%, 50% and 80% points. Refer back to Fig. 12.1a if the instructions in Fig. 13.5i are not clear.

Fig. 13.5j Calculating 30%, 50% and 80% points

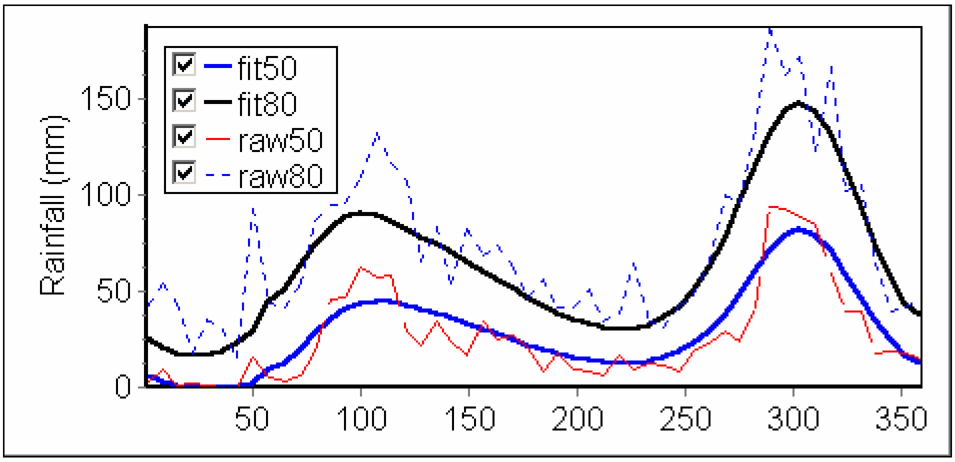
```
Note Comparison of methods 1 and 6
Note Initially for the 30% points
Note Assumes x78 for week number
Note and x80 for estimated 30% point

restore
warn off
open @kurun7
sta x1-x52;per 30 x53;per 50 x54;per 80 x55
save
open @kurunega
input @kurun7 x53-x55;into x83-x85
name x83 'raw30
name x80 'fit30
line x80 1 1 1
line x83 1 2 4
plot x80 x83 x78;xaxis 1 360;
tit "30% points for Methods 1 & 6"
```



The model seems excellent. Note though that the main peak in the observed data is sharp and is in week 42, where the observed 30% point was 62.1mm. The estimates from the model are less pronounced. Should the simple summary of the data, or the results from the model, be used? One way to assess this is to compare the results for other percentage points, because if the sharp peak in the simple summary is real, then it should be apparent for all percentage points.

Fig. 13.5k shows the same results for the 50% and 80% points. With this as further information, you have to decide which method is preferable, when the objective is to estimate the 30% point.

Fig. 13.5k 50% and 80% points using Methods 1 and 6



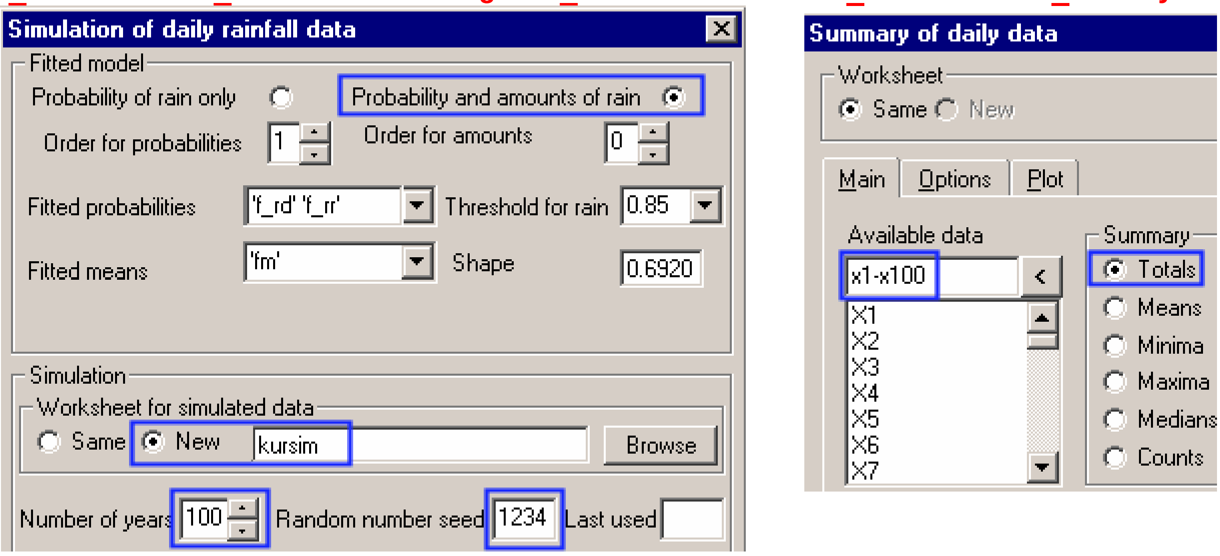
13.5.4 Simulation for Kurunegala

The **Climatic ⇒ Markov Chains ⇒ Simulation** dialogue simulates daily rainfall data, using the fitted model. Two tasks are considered. The first is to estimate the distribution of the total annual rainfall using the model for Kurunegala in Sri Lanka. The second is to estimate the distribution of the start of the rains, using the model for Samaru. In each case the results are compared with those from the direct analysis.

The **Simulation** dialogue is shown in Fig. 13.5l. The model for Kurunegala was used and 100 years of data were simulated. They have been put into a new worksheet called **Kursim.wor**. An initial random number seed of 1234 was also used.. This is not normally done, but it permits readers to use the same seed, and hence generate the same data used here.

Fig. 13.5l Simulating 100 years of data and calculating annual totals

Climatic ⇒ Markov Modelling ⇒ Simulation   Climatic ⇒ Summary



The total annual rainfall is also calculated for the 100 years, Fig. 13.5l and put into x101. To compare with the observed data, the worksheet Kurunega.wor is opened again. The 34 annual totals are calculated similarly, and the column is then copied over into x102 in the same worksheet.

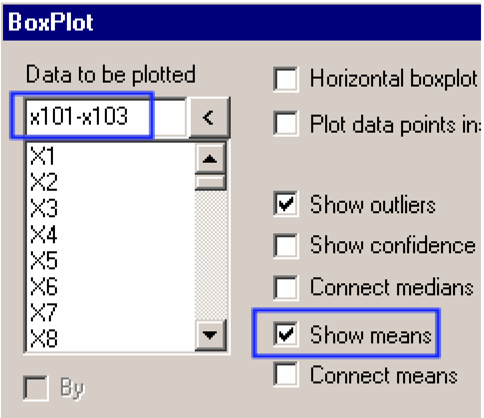
The boxplot for the rainfall totals from x102 showed, there was one extraordinary year when the total was greater than 4000mm. This was 1977, whereas no other year had more than 3000mm. Therefore the data omitting that year were put into X103. The dialogue is Manage => Reshape => Select and the command generated was:

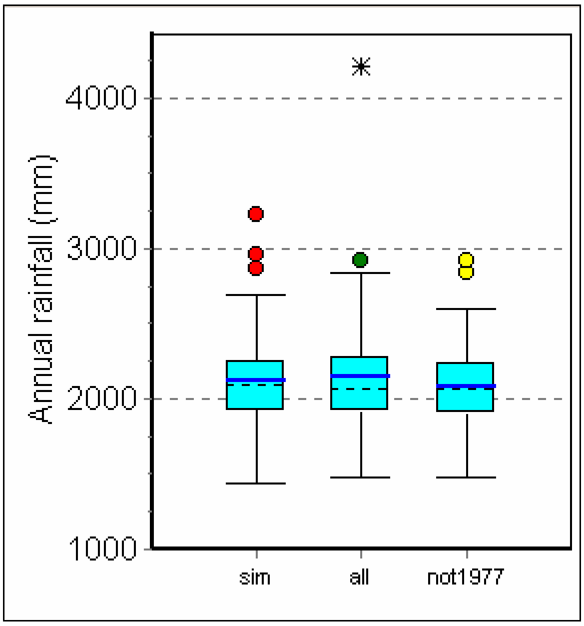
: select x102 ; into x103; if x102>4000

Boxplots of the simulated and actual totals are given in Fig. 13.5m. The odd year (1977) stands out clearly, and has made the mean considerably higher than the median, for x102.

Fig. 13.5m Boxplots of the simulated and actual annual totals

Graphics ⇒ Boxplot

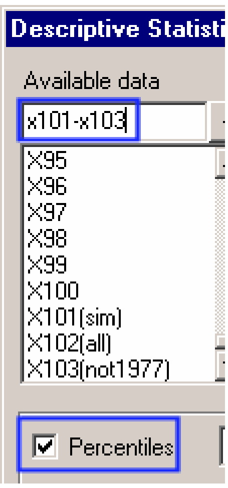




The numerical comparison in Fig. 13.5n confirms that the simulated series has a similar distribution to the actual data. The extremes, i.e. minimum and maximum, from the simulated data are wider than the actual data (after leaving out 1977) and this is to be expected. The extremes from 100 years would normally be more extreme than those from 34 years. The mean, median, standard deviation and percentiles are all similar.

Fig. 13.5n Comparison between data

Statistics ⇒ Summary ⇒ Describe - include 10,20,50,80 and 90% points



Simulated data

Descriptive Statistics

Column	sim
No. of observations	100
Minimum	1445 mm
Maximum	3232 mm
Range	1787 mm
Mean	2116 mm
Std. deviation	331 mm
10th percentile	1636 mm
20th percentile	1870 mm
50th percentile	2097 mm
80th percentile	2375 mm
90th percentile	2576 mm

Real data (omitting 1977)

Descriptive Statistics

Column	not1977
No. of observations	33
Minimum	1484 mm
Maximum	2924 mm
Range	1440 mm
Mean	2080 mm
Std. deviation	349 mm
10th percentile	1550 mm
20th percentile	1825 mm
50th percentile	2072 mm
80th percentile	2360 mm
90th percentile	2568 mm

Close inspection of the results shows, however, that there are small differences in the distributions that are probably real. The actual data are slightly more variable than the simulated data. Thus the actual years are a little more different than has been predicted from the model. Perhaps this is just the particular 100 years that was simulated? A different 100 years could be simulated by following the procedure in Fig. 13.5m again, but omitting or changing the "Random number seed" that was used in the Simulate dialogue. The results will be different, but the same pattern is likely.

It is tempting to query whether this difference is throughout the year, or just in particular seasons. This could be investigated by getting statistics for monthly, or other periods, rather than the annual totals. However, results on a weekly basis are already available from the **Climatic => Markov Modelling => Totals** dialogue, see Figs 13.5h and 13.5j. The **Totals** dialogue is like simulating an infinite number of years and the results show the same tendency. (Examining Fig. 13.5j carefully, shows that the observed values for the 30% point tend to be slightly below the fitted values. In Fig. 13.5k, for the 80% point, they tend to be slightly above). Hence, if we were grading the model for Kurunega, it might get a mark of **'good' but not 'very good'**. We return to this point after looking at the model for Samaru.

13.5.5 Simulation for Samaru

For Samaru a different model is used to the one fitted in Section 13.4. The same type model will be fitted, but on a daily basis. The steps are as shown in Fig. 13.5o and are those used in Figs 13.4b – 13.4e, ***except*** that the **Prepare dialogue** leaves the summary data on a daily basis.

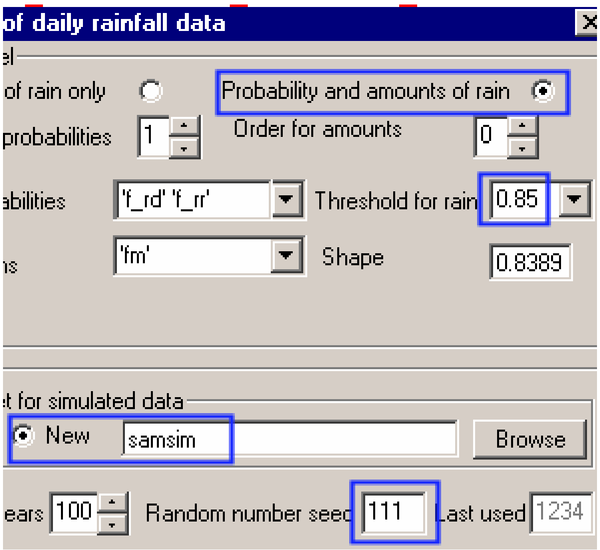
Fig. 13.5o Dialogues for Markov model of Samaru

Dialogue	Comment
File ⇒ Open	Samsmall.wor
Manage ⇒ Remove	X12-x60
Climatic ⇒ Markov ⇒ Counts/Totals	1^{st} order counts and zero order amounts
Climatic ⇒ Markov ⇒ Prepare	Leaving data on a daily basis
Climatic ⇒ Markov ⇒ Model Probabilities	3 harmonics for rain given dry 2 harmonics for rain given rain
Climatic ⇒ Markov ⇒ Model Amounts	2 harmonics

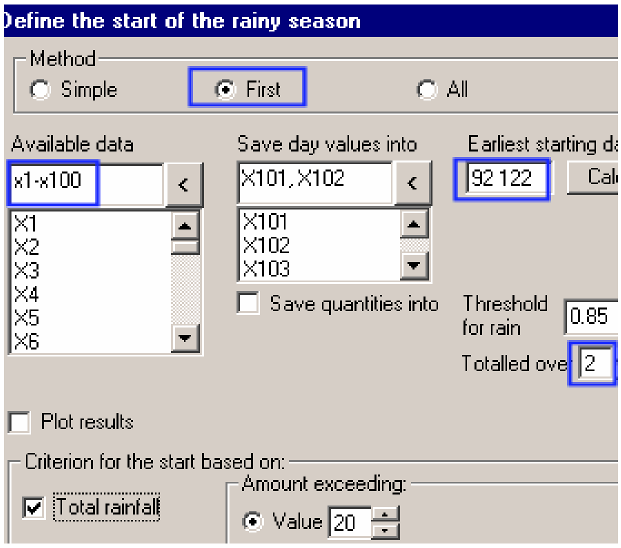
After following the instructions in Fig. 13.5o the dialogue in Fig. 13.5q is used simulate 100 years of data. Then the starting dates are calculated for the same two definitions used in Chapter 6, i.e. the date of the start is defined as the first day from 1^{st} April, and from 1^{st} May with more than 20mm in a 2-day period.

Fig. 13.5q Simulation of Samaru data and generation of starting dates

Climatic ⇒ Markov ⇒ Simulation



Climatic ⇒ Events ⇒ Start

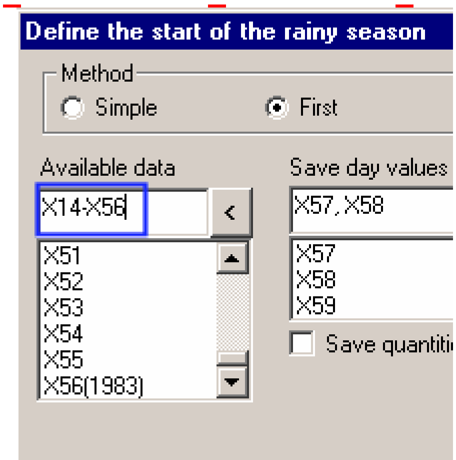


In this case a more severe test of the model is conducted. The model above was fitted using the 11 years of data from 1930 to 1940. The data from Samaru are available to 1983, so the results from the model are compared with the direct summary of the data, using just the remaining 43 years from 1941.

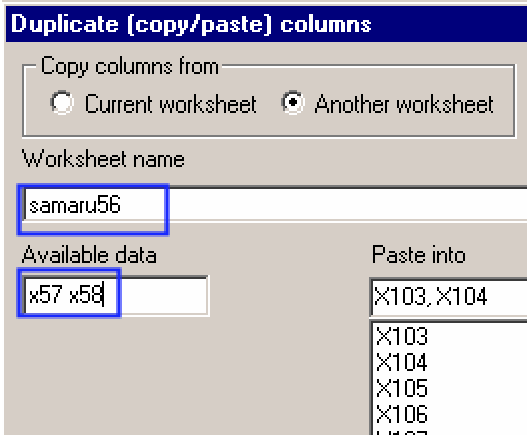
Fig. 13.5r Using a large set of data from Samaru

File ⇒ Open Worksheet ⇒ Samaru56.wor

File ⇒ Open ⇒ Samaru56.wor
Climatic ⇒ Events ⇒ Start



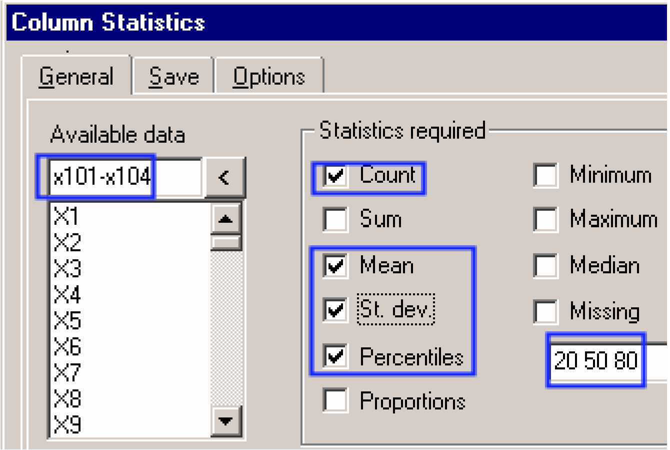
File ⇒ Open ⇒ Samsim.wor
Manage ⇒ Data ⇒ Duplicate



The dates of the start are found for x14-x56 (1941-83), Fig. 13.5r using the same definition as for the simulated data, Fig. 13.5q, and the results copied over to the worksheet with the simulated data. The simulated and actual starts are then summarised, Fig. 13.5s.

Fig. 13.5s Start dates from the simulation to 1930-1940 and observed 1941-83

	X101	X102	X103	X104
	SimApril	SimMay	April	May
1	16 Apr	03 May	17 May	17 May
2	01 Apr	25 May	29 Apr	01 May
3	10 May	10 May	19 May	19 May
4	14 May	14 May	08 Jun	08 Jun
5	18 May	18 May	06 Jun	06 Jun
6	23 May	23 May	08 May	08 May
7	23 Apr	05 May	13 May	13 May
8	31 May	31 May	18 Apr	10 Jun
9	28 May	28 May	01 May	01 May
10	27 Apr	23 Jun	15 May	15 May
11	10 May	10 May	04 Apr	21 May
12	06 May	06 May	10 May	10 May
13	24 May	24 May	04 May	04 May
14	15 May	15 May	23 Apr	07 May



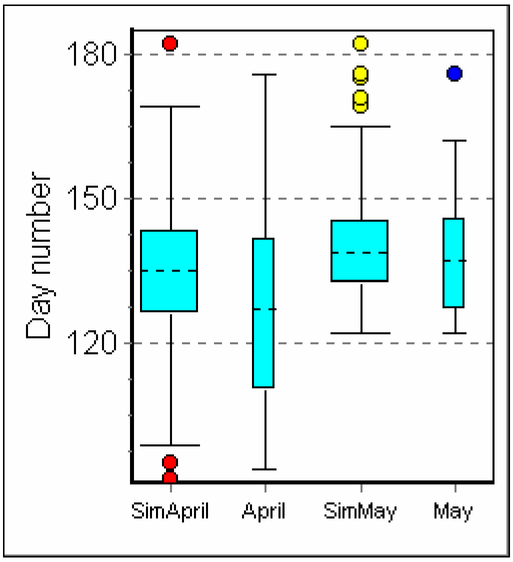
The results are close for the planting from May, Fig. 13.5t. The distribution from the model fitted to the data from 11 years in the 1930's is very similar to the actual distribution for the years 1941 to 1983. This is an excellent demonstration of the potential of the modelling approach.

However, comparing the two columns of simulated results, (x101 and x102 in Fig. 13.5s) shows that the model predicts that the mean sowing date from a possible April start is on average just 6 days earlier than the mean from the May starts. The actual data, for 1941-83, show a mean difference of about 12 days. The model does not therefore appear to be predicting enough heavy rainfalls in April.

Fig. 13.5t Summary values and boxplots for the start dates

Column Statistics

Column	Count	Mean	SDE	20%	50%	80%
SimApril	100	134.9	16.94	120.2	135	145.8
April	43	127.2	20.14	109	127	144.6
SimMay	100	141.2	13.18	131	139	150.8
May	43	138.9	13.52	126.8	137	152.4



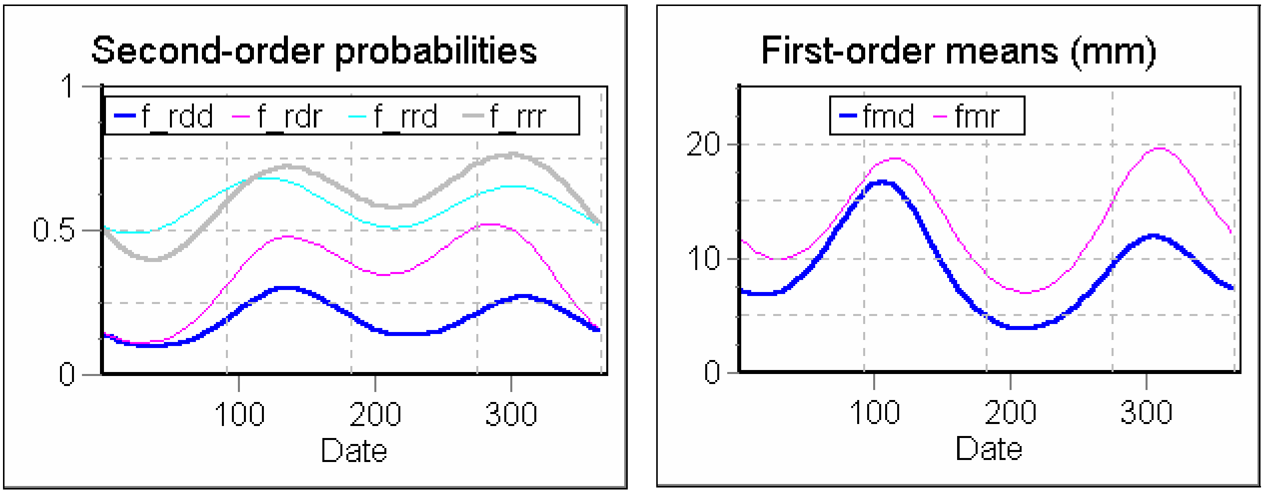
This model for Samaru therefore again gets a mark of 'good' but not 'very good'!

13.5.6 Review of the models

We have shown various ways in which they can be applied and it is important to assess whether their defects are a problem with the modelling approach, or just in the models that have been chosen.

The model for Kurunega is considered first. For simplicity the model used was a first-order Markov chain for the probabilities of rain and zero-order for the amounts, see Fig. 13.4k. Many sites in Sri Lanka and India require a second- order chain and Fig. 13.5u shows the curves resulting from fitting a second- order chain to the probabilities of rain. Particularly when the previous day was dry, there is a clear difference between the curves. This second-order model would result in more variable data, because spell lengths tend to last longer than with a first-order chain. The second graph in Fig. 13.5u shows that the model for the amounts was also over simplistic. The distribution of rainfall amounts per day is about 5mm higher when the previous day was rainy, than on the first rainy day of a spell.

Fig. 13.5u Fitting a second order model



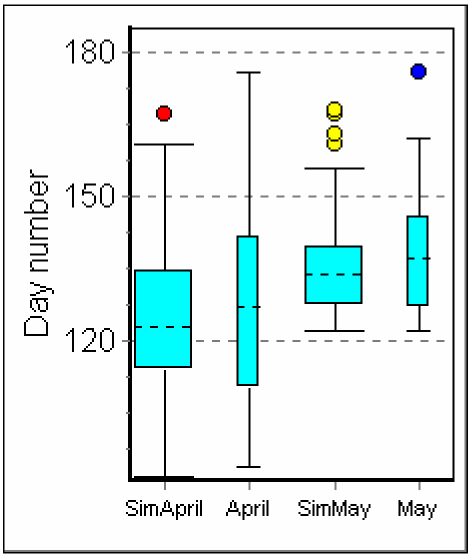
Using this model with second-order probabilities and first-order amounts, gives a better fit to the annual totals and we award this model a mark of **very good**.

For Samaru, the model used for the simulation in Fig. 13.5o was fitted in a daily basis. This was again a first-order Markov chain for the probabilities and zero-order for the amounts. Here the problem appears to be different. The complexity of model is adequate, it is the fitting within Instat that is the problem.

Fitting on a daily basis is recommended, with the more powerful packages, such as Genstat. However, Instat uses a normal approximation for the fitting and, with such a short record, there are problems with this approach, particularly at the beginning and end of the season, when there are few rainy days. If Instat is the only package available, then fitting on a 5 day basis is preferable. That is what was done in Section 13.4. Using that model for the simulation (with the same random number seed) gives the equivalent results to Fig. 13.5t in Fig. 13.5v.

Fig. 13.5v Comparing simulated and actual results after fitting on a 5-day basis

Column	Count	Mean	SDE	20%	50%	80%
SimApril	100	124.7	14.59	112.2	123	137
April	43	127.2	20.14	109	127	144.6
SimMay	100	135.2	10.17	126	134	141
May	43	138.9	13.52	126.8	137	152.4



The statistics for the April planting are now about 10 days earlier than those for the second definition, which matches those for the real data, from 1941-1983.

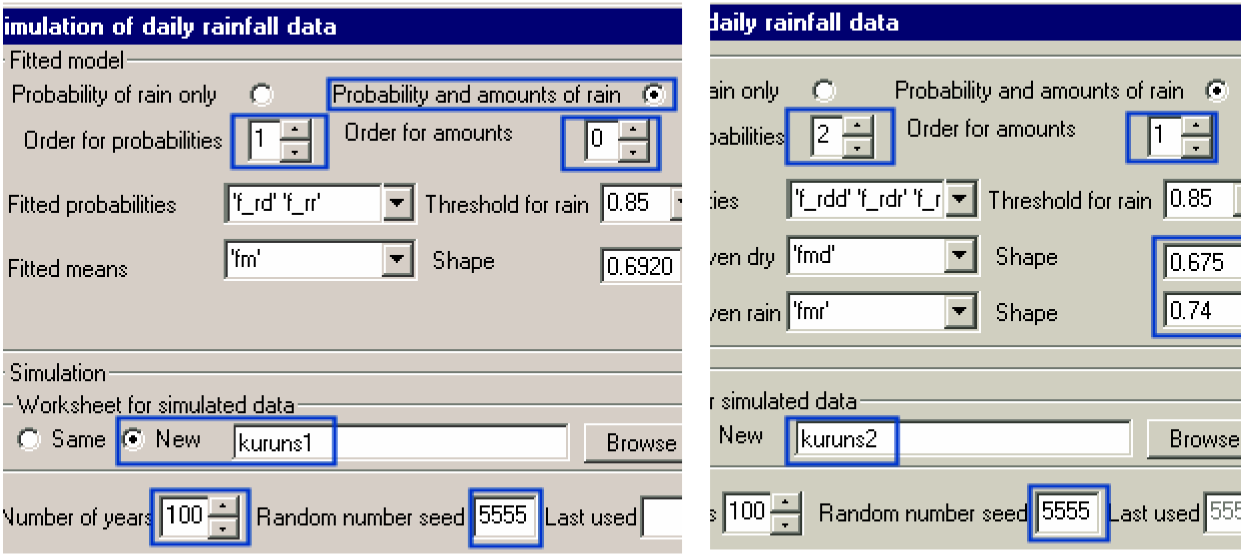
A second comparison of the real data and the model predictions from the model fitted on a five-day basis was for the April totals. It is again emphasised that the model was fitted to the data for just the 11 years, from 1930 to 1940, and the test data are from 1941 to 1983. The model gives a mean of 31.8mm and standard deviation of 25.7mm, compared with 36.6mm and 30.5mm for the 1941-1983 data. These results are close. If necessary the two columns could be compared for both mean and spread (standard deviation) using the **Statistics ⇒ Simple Models ⇒ Normal Two Samples** dialogue. The differences are not statistically significant.

This new model for Samaru, seems to earn a mark of **very good**!

13.5.7 Efficient comparisons

It is often important to compare two or more simulation models. Here they are for the same site, Kurunegala, and examples of first and second order models are compared. Usually they will be for different sites or for different periods at the same site, perhaps to investigate climate change, or to compare El Nino and ordinary years.

Fig. 13.5w Simulating 100 years from 1st and 2nd order models for Kurunegala



Each model is fitted and 100 years of data are simulated, see Fig. 13.5w. The key point in Fig. 13.5w is that the data from the 2 models are highly correlated, **because the same seed for the random number generator was used in each case**. This enables the models to be compared with high precision.

Fig. 13.5x shows some of the daily data from the simulation. The first three columns are from the first model and the others are from the second. In any year the data show that when the first model has rain, then often there is rain on the same day in the second model.

Fig. 13.5x Daily data for three years

	First order			Second order		
1	0.0	0.0	0.0	0.0	0.0	0.0
2	1.7	0.0	0.0	2.8	0.0	0.0
3	6.1	0.0	0.0	8.2	0.0	0.0
4	10.8	0.0	0.0	13.9	0.0	0.0
5	3.9	0.0	0.0	5.6	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.9	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	0.0	0.0	0.0
11	0.0	0.0	0.0	0.0	0.0	0.0
12	0.0	0.0	2.2	0.0	0.0	1.3
13	0.0	0.0	0.0	0.0	0.0	0.0
14	0.0	0.0	0.0	0.0	0.0	0.0
15	0.0	0.0	0.0	0.0	0.0	0.0
16	0.0	0.0	0.0	0.0	0.0	0.0
17	0.0	0.0	0.0	0.0	0.0	0.0
18	0.0	0.0	0.0	0.0	0.0	0.0
19	6.4	0.0	3.8	4.2	0.0	2.3
20	0.0	0.0	7.7	0.0	0.0	8.1
21	14.7	0.0	0.0	16.9	0.0	0.0
22	0.0	0.0	0.0	0.0	0.0	4.0

Fig. 13.5y January rainfall totals

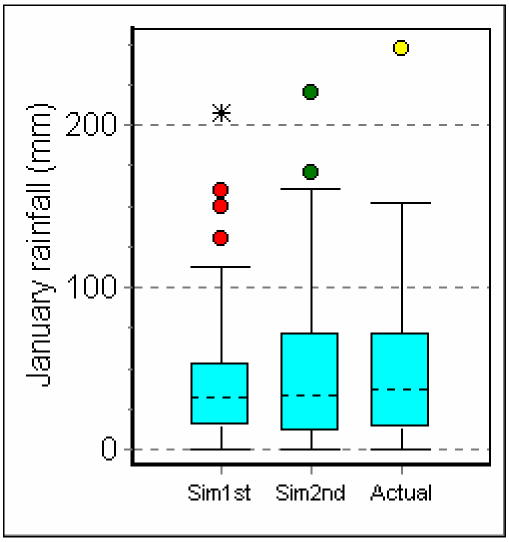
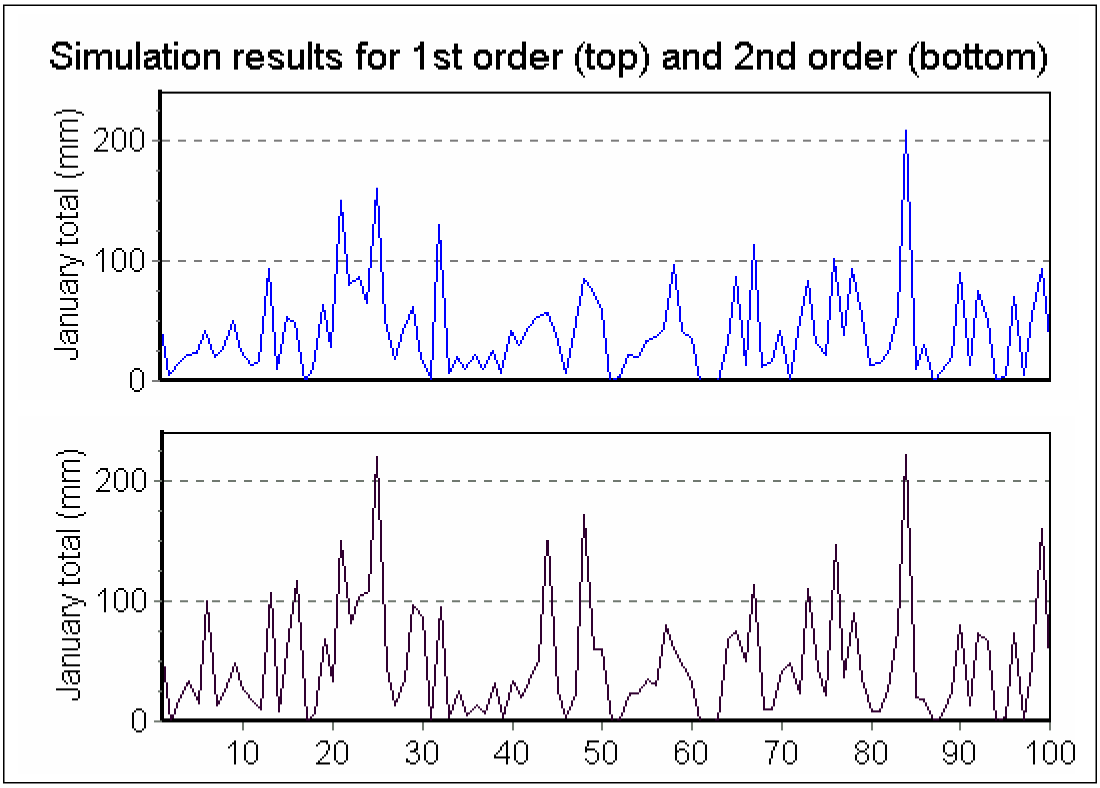


Fig. 13.5y considers the results for the January rainfall totals for the real data, compared to the two simulation models. The boxplots show that the median is about the same in each case, but the year-to-year spread is more closely followed with the second-order chain.

Fig. 13.5z January rainfall totals from 100 years of simulated data



The correlation in the way the daily data were generated (using the same seed), carries over to derived results, such as the January rainfall totals. This is shown in Fig. 13.5z which gives the totals for each of the 100 years. The values are clearly different, but a high January in one model tends also to be high in the other. The correlation between the two series is about 0.9.

This correlation provides a considerable variance reduction when the models are compared. Without it, about 700 years of data from each model, would have had to be simulated to have the same precision. The general ideas of variance reduction in simulation modelling are described in various texts, for example Morgan (1983).

13.6 Conclusions

Our aim in this guide has been to encourage the full use of climatic data. In this chapter, the modelling approach has stood up well, providing, of course, that an appropriate model is fitted. Hence we believe that the methods described here, of fitting a model to the daily data and then deriving results from this model, should become part of this full use of the data.

Sections 13.4 and 13.5 has stressed that the fitting of these models is limited, in Instat, because the regression facilities, though powerful, are only for the ordinary regression models. Ordinary means models where the residuals are from a normal distribution. The ideal modelling system would be one that is, at least as easy to use as that developed in Instat, and which uses a powerful statistics package. By powerful we mean one that includes facilities for fitting generalised linear models.

Some of the extra features that would be available in such a system.

- **Proper fitting of the models**
  Binomial regression models would be used for the chances of rain. There is then no difficulty in coping with the zeros in the data. Gamma regression models would be used for the amounts.
- **Comparison of different models**
  It would be easy to test the complexity of model required. For example, for Samaru Fig. 13.4d, are the two curves **really** needed, or would a single curve be sufficient?
- **Models can be simplified where possible**
  For example in Fig. 13.4k the two curves for the rainfall amounts have been fitted separately, but they look fairly parallel. Do these separate curves, which have a total of 11 parameters, fit significantly better than the model with parallel curves, which would have just 7 parameters?
- **Comparison of models for different time periods - climatic change**
  Where long records are available the curves can be fitted to different subsets of the data. They can then be compared to check for evidence of climatic change. If there is an indication of climatic change, the increased precision of the modelling approach makes it feasible to just use the recent years for planning purposes.
- **Comparison of models for different sites**
  This is an important extension. We have concentrated here on the models fitted to data from a single site, but where data are available for many sites in the same country or region, the models for the different sites may be simply related. Perhaps the models depend primarily on the geographical co-ordinates of the sites. Then a model could be found and results derived, just by knowing its position.

There is much that can be done.

Chapter 14 – Stretching the Software

14.1 Introduction

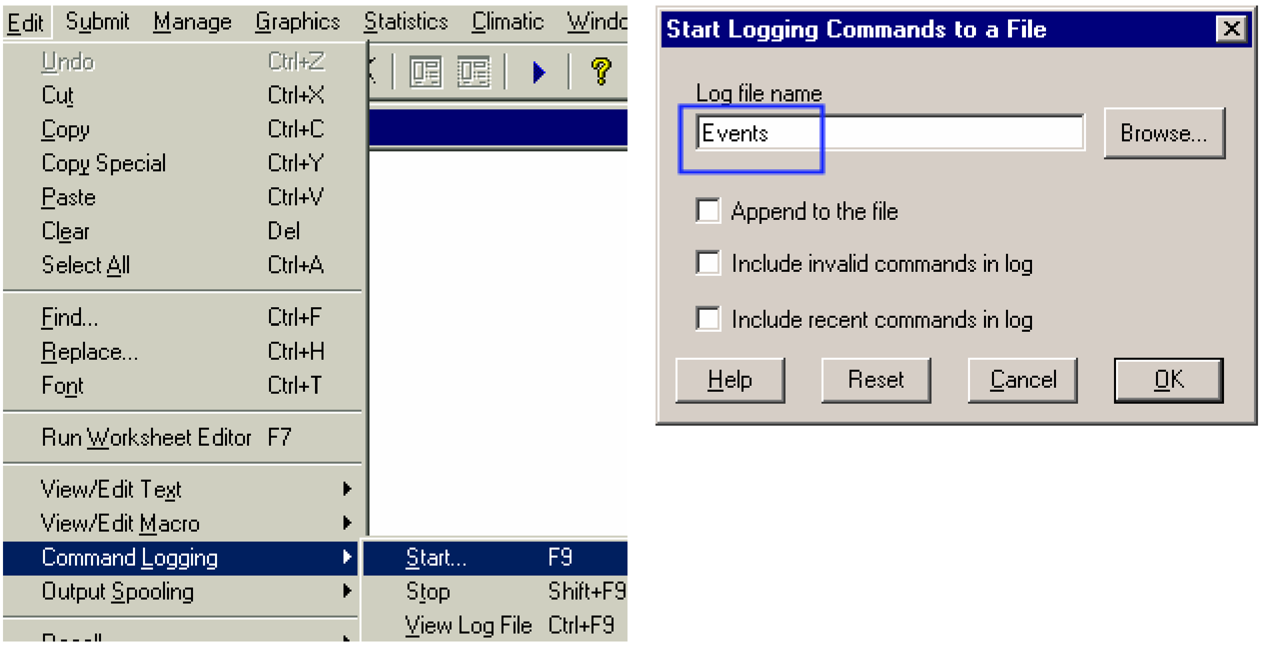
The Instat dialogues have been sufficient for many analyses. However there were instances where they were not enough. For example there can be complications in importing data as described in Chapter 3. In each of Chapters 5 to 7 we have emphasised that analysis must correspond to the clients' needs and not be limited by the dialogues, and Chapter 8 showed that the analysis of temperature records was assisted through various Instat macros.

Before Windows, learning a statistics package implied learning a new "command language". These commands are needed now, to construct "macros" or little "programs". The situation changed once the facilities of Windows are added. The copy/paste and similar dialogues in many statistics packages provide a sort of universal language. Hence it is easy to move from one statistics package to another, or from a spreadsheet to a statistics package.

However, there is a price to be paid for this initial ease of use. It is important that users take advantage of the dialogues and do not become imprisoned by them. A simple example is used to show how to avoid this imprisonment. It starts with the facility of keeping a **log file** to review how the dialogues correspond to commands.

Use **Edit** ⇒ **Command Logging** ⇒ **Start** as shown in Fig. 14.1a.

Fig. 14.1a Start a log file of commands



Now go through the steps given in Fig. 14.1b. This repeats the type of analysis from Chapter 7, Section 7.2, where we open a worksheet with monthly data and produce summaries and a graph.

Fig. 14.1b Dialogues used in Chapter 7

Menu	Details
File ⇒ **Open Worksheet**	Open the worksheet **sammonth.wor**
Manage ⇒ **Data** ⇒ **Clear(Remove)**	Remove **x14-x50**
Manage ⇒ **Data** ⇒ **Regular Sequence**	Use values **1 to 12**
Manage ⇒ **Column Properties** ⇒ **Name**	Name **x14** as **Month**

Stats ⇒ Summary ⇒ Column Stats	Use **x1-x12** Produce 20, 50, and 80% points in X15-X17
Graph ⇒ Plot	Produce **lines** and **symbols**

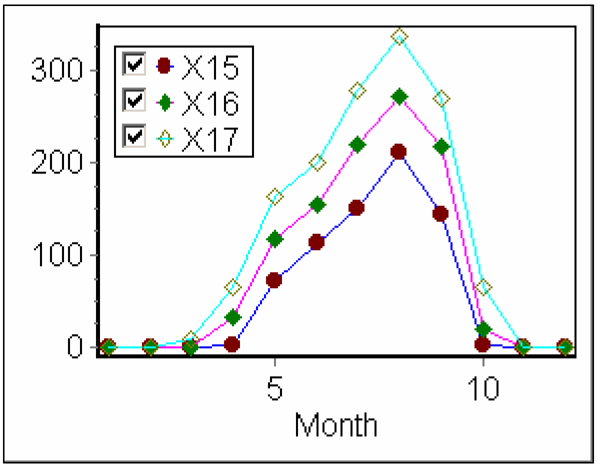
This should produce a graph, roughly as shown in Fig. 14.1c. It does not have to be exactly the same.

Now use **Edit ⇒ Command Logging ⇒ View Log File** to display the set of commands, roughly as in Fig. 14.1c. Then use either **<Shift> F9**, or **Edit ⇒ Command Logging ⇒ Stop**, to close the log file.

Fig. 14.1c Results and the Log file

Edit ⇒ Command Logging ⇒ View Log File

Edit ⇒ Command Logging ⇒ Stop




```
LOG @Events.log
OPEn @SAMMONTH.WOR;TEMP
DED ON
REMove x14-x50;lab
NAMe x14 'Month'
STA X1-X12;PER 20 X15;PER 50 X16;PER 80 X17
LINe X15 1 1 4
LINe X16 1 1 5
LINe X17 1 1 6
SYMbol X15 11 2 9
SYMbol X16 12 2 10
SYMbol X17 13 2 11
PLOt X15-X17 'Month'
```

Fig. 14.1c shows that a log file has acted as a sort of **translator** from the dialogues to the corresponding Instat commands. This is useful in its own right, because it provides a record of the analyses we have done. This example also shows:

- When using a new statistics package now, you can start with the Menus and use commands later.

You can start by **reading** commands (rather than having to write them). That is often easier, see Fig. 14.1c.

- Then you can start by **editing commands** produced from the dialogues, rather than writing them from scratch. This is what is to be done now.

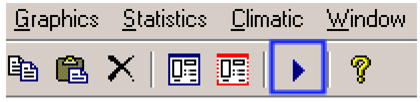
The log file is a useful starting point for the task of **macro-writing**. A **macro**, is essentially just a set of commands and that is just what is in the log file. It is sometimes called a “program”. If you have never written a program before you may be impressed how easily you have become a programmer!

To edit the commands they are first copied into a macro file, with **Edit ⇒ Edit Macro ⇒ New**.

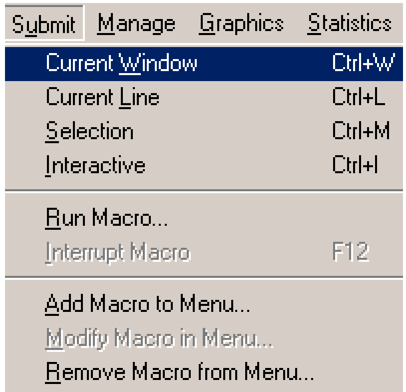
Fig. 14.1d Create a macro file

Edit ⇒ Edit Macro ⇒ New

```
NOTe Simple macro to plot monthly % points
OPEn @SAMMONTH.WOR;TEMP

REMove x14-x50;lab
ENTer X14; DATa (1]12)
NAMe x14 'Month'
STA X1-X12;PER 20 X15;PER 50 X16;PER 80 X17
LINe X15 1 1 4
LINe X16 1 1 5
LINe X17 1 1 6
SYMbol X15 11 2 9
SYMbol X16 12 2 10
SYMbol X17 13 2 11
PLOt X15-X17 'Month'
```



Submit ⇒ Current Window



Return to the Log file and use **Edit ⇒ Copy** to copy everything from the log file into the clipboard. Then go to the empty macro file and use **Edit ⇒ Paste** or **<Ctrl> V** to paste the contents into this file (Fig. 14.1d).

If there were any mistakes in the dialogues, and hence in the log file, these can be edited out. In Fig. 14.1d the first line from Fig. 14.1c has been removed, and replaced by a NOTe – this just makes the macro more readable. It is ignored by Instat. The third line from Fig. 14.1c has also been deleted.

Now check the macro file is the current window and then run the macro, using either of the ways shown in Fig. 14.1d. This should give the graph in Fig. 14.1c, just as was done with the dialogues earlier.

Now, still with the macro as the current window, use **File ⇒ Save** or **File ⇒ Save As**, shown in Fig. 14.1e, to save the macro. Call it **sammonth.ins**. The "ins" extension is to identify the file as a macro.

Section 14.2 shows different ways that this file and other files of commands (macros) can be run. Instat's Climatic Library that we have used in earlier chapters is re-introduced. Section 14.3 adds the idea of constructing a **local library** for the useful macros from your own site.

Section 14.4 describes the writing of simple macros, mainly using examples of a seasonal analysis, introduced in Chapters 5 and 6.

Sections 14.5 and 14.6 shows how to add some generality to the macros. For example, the macro above is limited to the particular station. For example, it would be more flexible if the macro would ask for the name of the station to process.

Section 14.7 returns to the construction of a local library, and in particular to the way of adding a HELP files to a macro. This is useful, even just for you to remember how to use the macros on a later date. It is essential, if you hope that others will be able to use your macros. Section 14.8 describes an example of a macro to calculate the day length at any place and time of year.

Finally, we consider how ambitious to become in writing macros. This discussion is at a more general level, than Instat, because almost all spreadsheet and statistics software includes similar or more powerful facilities. For example many people extend Excel, because of the powerful Visual Basic language that is part of the package. While the main theme of this chapter is to prevent users being imprisoned by the Windows style-dialogues, it is also important to guard against the risk of "re-inventing the wheel". It can take a long time to write macros that are of general use. It doesn't take many days before the development cost may exceed the cost of obtaining alternative software that already contains the facilities being programmed.

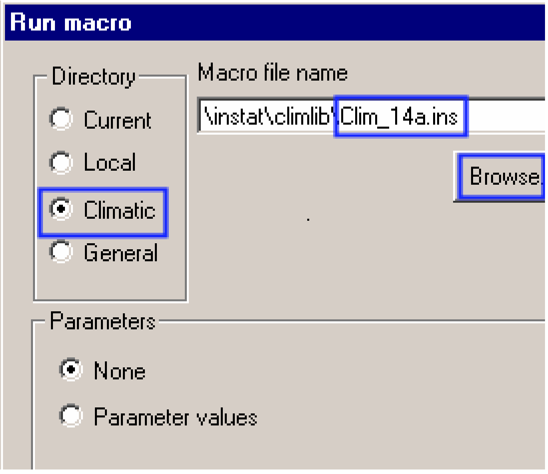
14.2 Using the Instat Climatic library

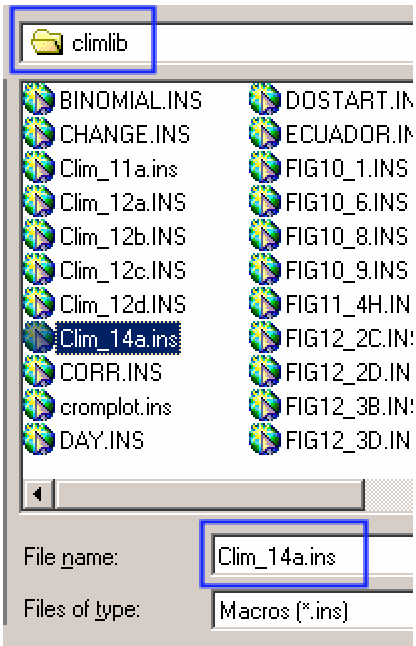
Using macros that other people have written is much easier than writing them. This section describes ways of using the macros provided with Instat. We consider mainly macros from Instat's climatic library. If you are following these operations in Instat, you may want to clear the output window before starting.

The first method is to use the **Submit => Run Macro** dialogue as shown in Fig. 14.2a. As an example we use our own version of the simple macro described in the last section. It is in the climatic library and called **Clim_14a.ins**, so in Fig. 14.2a, use **Browse** to find the file. Alternatively, if you use your own version, created above, then stay in the current directory. The macro runs when you press **OK**.

Fig. 14.2a Run a macro file

Submit ⇒ Run Macro ⇒ Browse

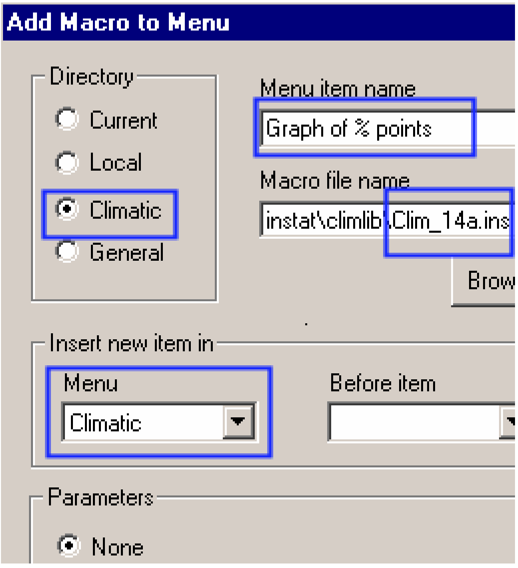




You may wish to close any worksheets, before trying the second approach. This time use **Submit => Add Macro to Menu** dialogue and complete it as shown in Fig. 14.2b. The macro file name is **Clim_14a.ins** as in Fig. 14.2a. Now also give a descriptive name, for example "**Graph of % points**". Then use the option to add the item to the bottom of the Climatic menu. Pressing OK does not seem to do anything, but if you examine the **Climatic** menu, Fig. 14.2b, you will see that an extra item has been added.

Fig. 14.2b Add a macro to a menu

`Submit => Add Macro to Menu`



Now this macro can be run, just by clicking on the menu item. It will also be there when you start Instat again. This is therefore a good option if you plan to use the same macro repeatedly. It is also useful if a beginner to computing is to use the macro later. They no longer even need to know that it was a macro.

If you have made a mistake in the way you have put this item on the Climatic menu then use **Submit => Remove Macro from Menu**.

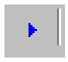
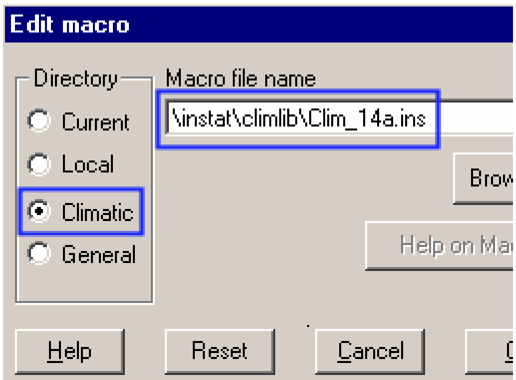
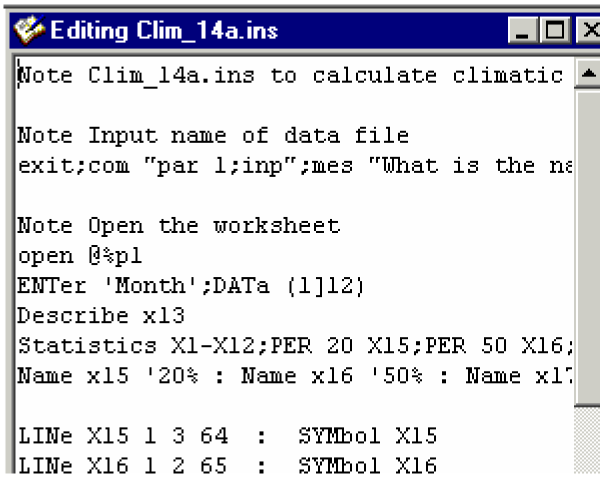
For the third method use **Edit => Edit Macro => Open** as shown in Fig. 14.2c. When you press **OK** a new window opens, containing the contents of the macro. This can now be run with **Submit => Current Window**. Short cuts are to type **Ctrl+W** or to click on the  symbol in the toolbar.

Fig. 14.2c Edit and run a macro

`Edit => View/Edit Macro => Open`





This is the method to use while developing a macro. It also allows submission of just parts of the programme if appropriate.

Finally, you could type the name of the macro into the Commands and Output window. A macro needs to have an @ sign in front of it, so typing

: **@Clim_14a**

gives the same result as the other 3 methods. An alternative is to type

: **exec @Clim_14a**

In this case the commands are listed in the output window as the macro is run. This is not normally needed, but if the macro gives an error, then this method would allow you to see which Instat command was giving the problem.

This fourth method shows that Instat treats macros similarly to other commands. Macros are sets of commands that are executed together for convenience, and this indicates that they could also include other macros within them.

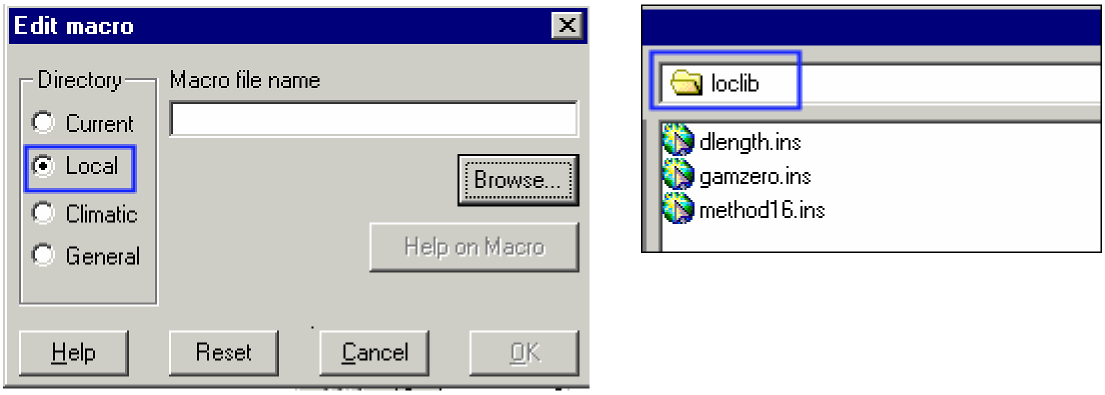
14.3 Using the Instat local library

Additional macros may also be provided locally. They are normally stored in a **local library**. When new versions of Instat are released, the macros in the global libraries may change, but those in the local library are the responsibility of staff at your site.

The general concept is that it is not necessary for all users of a (statistics) package to be equally proficient. If a few staff (perhaps those in a computing or statistics centre) can become expert, then they can help by making the package easier for others to use. This help can include the provision of courses and the writing of macros. It may also be that some macros that are provided in the global libraries, could be improved and the improved versions can then be stored in the local library.

An example of a local library, containing macros for use on a recent course is in Fig. 14.3a.

Fig. 14.3a Macros in a local library



The next sections of this chapter show how a macro can be written. Section 14.4 shows how basic macros can be constructed. Further ideas are introduced in Sections 14.5 and 14.6 that enable quite general programs to be constructed, using Instat's commands.

14.4 Writing a simple macro

Section 14.2 has shown that it is easy to execute a macro that is already written. In this section we show that once you have some experience of Instat, it is also straightforward to write a simple macro.

The first step is to be clear on the commands that would be typed if you were working interactively. You could use Instat's dialogues and then copy the log file as described in Section 14.1, but here the Instat commands are typed. The calculation of the start of the rains for Samaru is taken as an example. The commands are as follows:

```
open @samsmall
remove x12-x15
RUN X1-X11;RANge 0.85;SUM 2 20;
        FIRst 92 122;DAYs X12 X13
x14=x13-x12
select x14;into x15;if x14 > 0
```

The second step is to decide on a name for the macro. We will call it **start1.ins**. Next, the commands are typed into a text file. This can be done independently of Instat, but here the file is prepared within Instat, using **Edit => Edit Macro => New**. Type the commands as shown in Fig. 14.4a and save the file.

Fig. 14.4a Writing a macro

Edit => View/Edit Macro => New

Editing Commands

```
open @samsmall
remove x12-x15
RUN X1-X11;RANge 0.85;SUM 2 20;FIRst 92 122;DAYs X12 X13
x14=x13-x12
select x14;into x15;if x14 > 0
```

```
Year      1        2

y1930   108      133
y1931   118      131
y1932   115      133
y1933   156      156
y1934   116      124
y1935   134      134
y1936   113      123
y1937   126      126
y1938   118      137
y1939   134      134
y1940   118      127

Number of cases = 7
```

By default the macro will be given the extension .INS

N.B. This file will be created in the working directory, e.g..\Working, it does will not overwrite any of the macros provided by Instat, which are in the directories ..\Instat\inslib and \Instat\climlib. Instat macros all have the extension .INS.

You are now ready to test macro, by selecting **Submit ⇒ Current Window**. If there are no mistakes, this should Open the Samsmall worksheet and calculate that there were 7 of the 11 years where the planting date was different for the two definitions, Fig. 14.4a.

This storing of commands is effectively programming. If you have never programmed before, then **start1.ins** is your first program! If there were mistakes then you need to return to the editor to make corrections. Otherwise the initial programming task is finished.

The **start1.ins** program has introduced a second way Instat can be used. If another set of data has to be analysed, then this macro could be edited for the new set of data, rather than using the dialogues or typing the commands again, as is the case when Instat is used interactively.

Often, when an initial program has been constructed, some improvements can be made. Here some changes are shown in Fig. 14.4b.

The reason to turn **Warnings** off is that Instat will normally ask you if you are about to overwrite data. This is usually inconvenient when using a macro. With **Warnings OFF**, Instat does not ask. The **REStore** command restores the state of the WARning and any other flags, that are changed in the macro, to their original settings, once the macro has finished.

With Warnings off, the line in Fig. 14.4b**,** to remove x12-x15 is not needed and so it is removed. Commands are also added to display the final result more clearly.

To make these changes, either continue editing the file, if it is still open, or use **Edit ⇒ Edit Macro ⇒ Open** and select **start1.ins**. Type the additional commands so the Edit window looks similar to Fig. 14.4b. Run it, as before, by typing **Submit ⇒ Current Window**. When run it should display the result as:

7 of 11 years, or 64 percent, had later planting.

Fig. 14.4b Editing a macro

(If the macro file was closed) **Edit ⇒ Edit Macro ⇒ Open ⇒ start1.ins**

Editing start2.ins

```
restore
warn off
open @samsmall

RUN X1-X11;RANge 0.85;SUM 2 20;FIRst 92 122;DAYs X12 X13
x14=x13-x12
select x14;into x15;if x14 > 0
K1=count(x14)
K2=count(x15)
K3=100*K2/K1
display K2 "of " K1 "years, or " K3 "percent, had later planting";
parallel; hea; field 3 3 3 10 3 28; fix 0
```

This macro can then be saved, possibly with a different name, in Fig. 14.4b the name **start2.ins** is used.

14.5 Adding flexibility

14.5.1 System integers

Nine system integers are available, %1, %2, ..., %9. They are usually set with the calculate command and may then replace any integer anywhere within a command. For example typing

: %1 = 4

: DISplay X%1

in the **Commands and Output** window displays the data in X4.

These system integers are normally used in macros so they can be written more generally, without assuming data are in specific columns.

As an example, the macro from the previous section is further extended. The resulting macro, called **Start3.ins** is shown in Fig. 14.5a.

Fig. 14.5a Using system integers in a macro

```
restore
warn off
%1=11
%2=%1+1
%3=%1+2
%4=%1+3
%5=%1+4
open @samsmall
RUN X1-X%1;RANge 0.85;SUM 2 20;FIRst 92 122;DAYs X%2 X%3
x%4=x%3-x%2
select x%4;into x%5;if x%4 > 0
K1=count(x%4)
K2=count(x%5)
K3=100*K2/K1
display K2 "of " K1 "years, or " K3 "percent, had later pl
parallel; hea; field 3 3 3 10 3 28; fix 0
```

Fig. 14.5b Changes with another worksheet

Editing start3.ins

```
restore
warn off
%1=56
%2=%1+1
%3=%1+2
%4=%1+3
%5=%1+4
open @samaru56
RUN X1-X%1;RANge 0.85;SU
x%4=x%3-x%2
select x%4;into x%5;if x
K1=count(x%4)
K2=count(x%5)
K3=100*K2/K1
```

There are 11 years of data so %1 = 11 is on line 3 of the macro in Fig. 14.5a. Then %2 to %5 are calculated. The line

RUN x1-x**11**;… ; DAYs X**12** X13 becomes
RUN x1-x**%1**;…;DAYs X**%2** X**%3**

and so on.

The advantage of this becomes clear if later the analysis is to be repeated for another station. As an example, suppose this analysis is then to be used on the longer file from Samaru, sam56.wor. The changes are shown in Fig. 14.5b.

14.5.2 Parameters

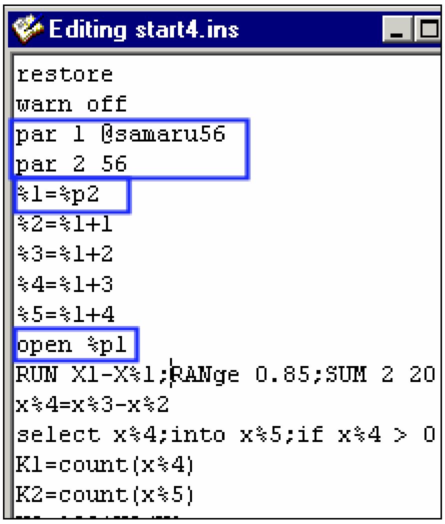
One limitation of the system integers, %1 %2 ... %9 is that they have to be whole numbers. It is also possible to set **parameters** P1, P2, ... P9 that can contain strings. They are normally set with the PARameter command, e.g

: **PAR 1 K2**

would set parameter P1 to be "K2". Then K2 is substituted anytime that %P1 appears in a command. Hence

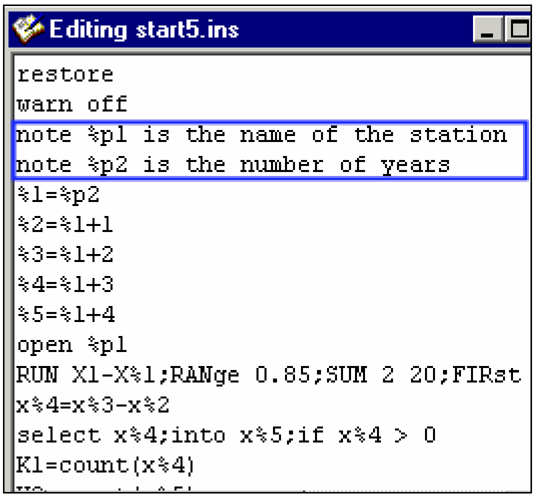
: **DIsplay %P1** is the same as : **DIsplay K2**

Fig. 14.5c Using parameters in a macro



Editing start4.ins

```
restore
warn off
par 1 @samaru56
par 2 56
%1=%p2
%2=%1+1
%3=%1+2
%4=%1+3
%5=%1+4
open %p1
RUN X1-X%1;RANge 0.85;SUM 2 20
x%4=x%3-x%2
select x%4;into x%5;if x%4 > 0
K1=count(x%4)
K2=count(x%5)
```



Editing start5.ins

```
restore
warn off
note %p1 is the name of the station
note %p2 is the number of years
%1=%p2
%2=%1+1
%3=%1+2
%4=%1+3
%5=%1+4
open %p1
RUN X1-X%1;RANge 0.85;SUM 2 20;FIRst
x%4=x%3-x%2
select x%4;into x%5;if x%4 > 0
K1=count(x%4)
```

Fig. 14.5c shows the changes made in the macro called start4.ins. Now the station can be changed merely by editing the two lines that set the parameters.

Some commands set parameters automatically. This includes the Execute command. If it is given with arguments it automatically sets P*i* to be the *i*th argument. This is used in **Start5.ins** to make **Start4.ins** more general.

It can be run by typing

: **@Start5 @samsmall 11**

to analyse **11** years of data from the worksheet **Samsmall**

Although the generalisation has made the macro less readable, the macro is now much more powerful. For example to analyse 2 datasets, each with a different number of years of data, one has only to type

: **@Start5 @Samsmall 11**
: **@Start5 @Kano 30**

If necessary, these lines could themselves be put in a file called Nigeria.ins and the stations analysed by typing

: **@Nigeria**

14.6 More on macro writing

The Climatic library macro, RAINSUM, is used to illustrate further concepts of macro writing. It summarises daily rainfall data and then plots the 20, 50 and 80 percentage points of the totals.

It can be called in any of the ways described in Section 14.2. As a command, if given as

: **@RAINSUM**

it prompts the user for the relevant inputs. Experienced users, could call it by

: **@RAINSUM @SAMSMALL "DEC" 11 @SAMDEC**

In this case it is run without prompts to the user.

Fig. 14.6a Partial listing of Rainsum.ins in the Instat Climatic library

Editing rainsum.ins

```
note RAINSUM.INS edited 8/7/2005 RDS
restore : warn off : echo off
loop 1; if narg=0
  display "Input of information for SUMMARY macro"
  exit;com "par 1;inp";mes "@What is the name of the worksheet with the daily data?"
  disp" The data are assumed to be in columns X1 ... "
  exit;com "par 2 MON;inp";mes "MONth, DECades, WEEks or PENtades (Default MON)?"
  exit;com "par 3;inp";mes "How many columns of daily data are there?"
  exit;com "par 4;inp";mes "What will you call the worksheet with the summary data"
loop 1
NOTe Open the worksheet and check it has enough columns
open @%p1
%1=num(3)
%2=%p3*2
exit ;mess "Not enough columns in the worksheet";if %1<%2
...  ...  ... ...  ...  ... ...  ...  ...
NOTe change the following lines if different % points are wanted
stats x1-x%4;per 20 x%5;per 50 x%6;per 80 x%7
name x%5 '20% : name x%6 '50% : name x%7 '80%
...  ...  ... ...  ...  ... ...  ...  ...
plot x%5-x%8;title "Percent points for %p1 %p2 data";href x%9
```

Aspects described are those that have not been covered in the previous sections.

The EXIt command is used to suspend the running of a macro. This may be to allow the user to supply information as shown in Fig. 14.6a. For example, the command

exit; com "par 3; input ";mes "How many columns of data are there?"

exits temporarily from the macro to allow the information to be copied to parameter **p3**. If it had been given as

exit; com "par 3 30; input ";mes "How many columns of data (default 30)?"

then this provides a default. Thus, with the command given as "par 3 30; input", if no input is given, 30 is transferred to the parameter **p3**. In Fig. 14.6a the second parameter has a default setting of “**MONths**”

EXIt can also be used to exit permanently from a macro if an error is detected. For example,

exit ;mess "Not enough columns in the worksheet"; if %1<%2

The LOOp command may be used with any of the three subcommands ;IF ;REPeat and ;WHIle. It allows sections of a macro file to be omitted or repeated more than once.

In Fig. 14.6a there is a section

loop 1; if narg=0

display "Input of information for SUMMARY macro"

....

....

loop 1

The variable, NARG is set automatically when a macro is called. If the macro is called by

: **@SAMRAIN**

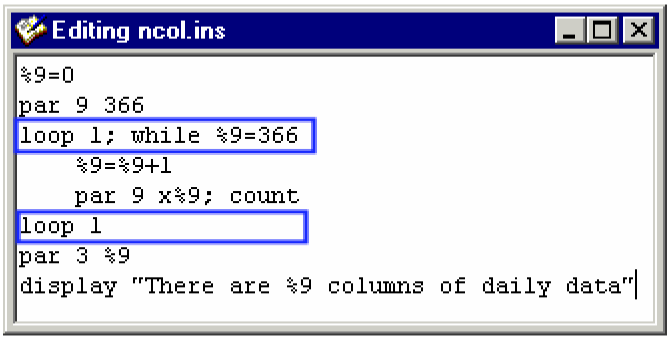
then NARG would be set to 0, while

: **@RAINSUM @SAMSMALL "Dec" 11 @SAMDEC**
would set NARG=4.

Testing the value of NARG is the way the macro know whether to prompt the user for the information it requires (i.e. has the macro been called with arguments?).

The **LOOping** facility is so important that a further example is given. use. The command **LOOp; WHIle**, is illustrated in the small macro given in Fig. 14.6b, which stores the number of columns with daily data, i.e. of length 366 in a worksheet. It assumes these are the first columns in the worksheet.

Fig. 14.6b Illustrating LOOp; WHIle


Editing ncol.ins

```
%9=0
par 9 366
loop 1; while %9=366
    %9=%9+1
    par 9 x%9; count
loop 1
par 3 %9
display "There are %9 columns of daily data"
```

This could be used within the **rainsum.ins** macro, if the stations contained differing numbers of years to be analysed, i.e. X1 - .... and the subsequent columns did not have 366 observations. Then, in **rainsum.ins**, the third parameter could be omitted and **rainsum.ins** could contain the commands

open @%p1
@ncol

This example also shows that macros can call further macros, as needed.

It is always a good idea to make worksheets with more columns than are needed for the data, because this allows the remaining columns to be used for temporary storage. The functions NUM and LEN can be used in a CALc command to find the number and maximum length of data structures in a given worksheet. For example, the line in Fig. 14.6a

%1=num(3)

is used to check that the worksheet has enough columns for the analysis.
The arguments in NUM and LEN can be 0, 1, 2, or 3, corresponding to S(tring), L(abel), K(Constant) or X(Columns).

Finally, macros will often require the user to have access to details of the current state of the worksheet, etc. This is usually achieved using the **PARameter** command, for example
: **PAR 3; WOR**
would put the name of the current worksheet into the parameter **p3**.

Many macros can be constructed as adaptations of files that are already in the Instat library. To transfer a macro from the climatic library to the current working directory for editing, use Windows Explorer and copy the file from ….\Instat\Climlib to …\My documents\Working, or your current working directory.

Then edit the local version by **Edit ⇒ Edit Macro ⇒ Open** and selecting the macro from the current directory.

When a macro is called, Instat looks first in the current working directory, then in the local library, then in the climatic library, and finally in the general library. Hence there is no need to change the name of the macro.

Alternatively, you may wish to give your adapted macro a new name, because its function is slightly different to the library macro and you therefore want access to both of them.

14.7 Adding a help file

Once a macro has been written, it is useful to add a help file that describes how it can be used. This is sensible, even if the macro is only for your personal use, and is even more important if you are providing the macro for others, perhaps in a local library.

Constructing a help file is easy. It should normally have the same name as the macro and the extension must be AID. To construct it, use **Edit ⇒ Edit Text ⇒ New**. When you have typed the information into the Editing window, use **File ⇒ Save** and name the file **<macroname>.AID**, for example **sumrain2.aid**.

With the extension AID, this file may be displayed when the associated macro is selected and the **Submit ⇒ Run Macro** or the **Edit ⇒ View / Edit Macro ⇒ Open Macro** dialogues are used. For example, if you selected the macro **sumrain2.ins**, then clicking Help would display the text in **sumrain2.aid**.

By default, macros and help files, that you write, are stored in the current directory that you are using. You may later wish to transfer them to a local library. This concept of a local library is useful for two reasons. First it provides a way to organise your files. Thus, it separates the macro files from the data files. Scientists often keep data files in different directories, depending on the application, and the macros in the local library will be accessible from any other directory.

The second reason for having a local library is where a group of staff within an institute, wish to share the macros that they have written. Then it becomes useful that one person becomes responsible for managing the local macro library and other staff send them their macros.

One use of local help files is for sites where the main language is not English. Because the help files are prepared using a simple editor, they can easily be written in any language. For example, at some Instat sites in West Africa the important macros and associated help files from the main library have been translated into French and put into the local library.

The local library is in a sub-folder, called **….\Loclib**, of the **…..\Instat** folder. Hence, if you have been using your own folder initially, putting macro and help files into the local library merely involves copying the files into this folder.

14.8 An example – day lengths

Fig. 14.8a has been written to calculate the day length at a specified latitude for any time of the year.

Fig. 14.8a First part of the macro length.ins, to calculate day lengths

Edit ⇒ View/Edit Macro ⇒ Open ⇒ dlength.ins in the Climatic Library

```
Note Macro dlength.ins - Roger Stern 11/07/2005

Note Calculate daylength from column with day of year at any given Latitude
Note from formula in Sellers W.D.,1965. Physical Climatology.
Note University of Chicago Press,Chicago Il
Note Adapted from Genstat procedure written by R.J. Reader & K. Phelps,
Note Biometrics Department, Horticulture Research International, Wellesbourne,
Note Warwick. CV35 9EF, UK.

restore : warn off : echo off
par 2 0:par 3 0
loop 1; if narg=0
  display "Input of information for dlength macro"
  exit;com "par 1;inp";mes "Which column has the day numbers, e.g. x1?"
  exit;com "par 2 ;inp 0";mes "Latitude, in degrees e.g. 52, -30 (default 0)?"
  exit;com "par 3;inp 0";mes "And the minutes (default 0)?"
loop 1
```

The macro is written so if it is called without arguments, it will prompt the user for the information it needs. Otherwise it just proceeds. An example is in Fig. 14.8b, where it was called by typing

: @dlength x1 5020

The day numbers were input into x1 in a new worksheet, see Fig. 14.8b. A copy was put into x2, so x1 could be formatted as a day number in the year. Then running the macro puts the results into x3 (the first empty column) and also displays them in the Output window.

Fig. 14.8b Using the dlength macro

	X1	X2	X3
	day	dayno	dayhrs
1	01 Jan	1	8.12
2	31 Jan	31	9.25
3	29 Feb	60	10.93
4	01 Apr	92	12.95
5	01 May	122	14.70
6	01 Jun	153	15.88
7	01 Jul	183	15.90

Commands and Output

```
@dlength x1 50 20

Latitude 50 degrees 20 minutes


Row          day        dayhrs

  1            1          8.12
  2           31          9.25
  3           60         10.93
  4           92         12.95
  5          122         14.70
  6          153         15.88
  7          183         15.90
```

14.9 How far should software be stretched?

It is easy to write simple macros in Instat. Their use in various chapters has shown that this is sometimes desirable if you wish to analyse data fully in a way that is not tedious.

Learning to write simple macros is also an excellent way of gaining experience in the use of the software.

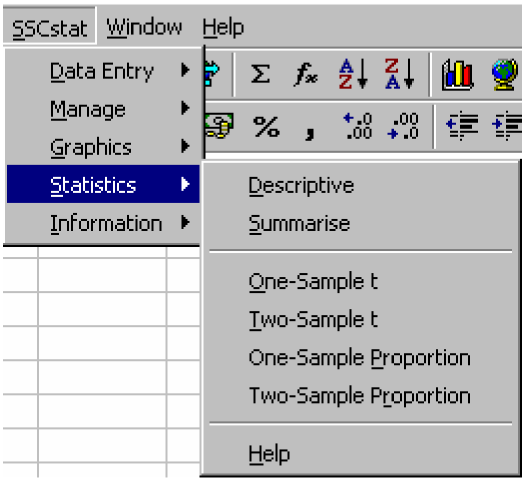
However, making the macros into general tools that can be shared, is a much more time-consuming business. We caution that users should not become over-ambitious. In general macros have two main purposes. The first is to assist on data entry or manipulation and the second is to extend the statistical capabilities. Most of the macros in this guide are of the first type and we suggest that, with Instat, this will normally be the case.

There is nothing wrong in writing macros to extend Instat's statistical features, for example as shown in Section 14.8, but if it takes much time, then the approach should be questioned. One key question is whether the proposed macro is already a standard facility in another package. A further question is whether Instat is the best environment for this macro. To help with this second issue we consider the programming facilities within some other packages.

Almost all the common statistics packages allow macros to be written. They differ in the power of their language and also in the extent to which the resulting macros can be added to the Windows system and distributed to others. Excel epitomises software that caters more comprehensively than Instat for additions to be provided.

With software such as Excel it is possible to add extra facilities with full dialogues, including help files and add these to the Excel menus. We have used these facilities to provide an add-in, called SSC-Stat, as shown in Fig. 14.9a. Many others have done the same and the Excel menu in Fig. 14.9a also shows a second extra menu that corresponds to a simple statistics package called StatPlus (Berk and Carey, 2000).

Fig. 14.9a Excel menu for SSC-Stat



However, if you decide to use Excel for program writing, the first question posed above still stands. Make sure you are not becoming forced into writing yet another statistics package. Plenty have already been written!

All the statistics packages, such as Minitab, Systat, Genstat, S-Plus and R, include good language facilities. Once you know the package, then S-Plus and R are perhaps the quickest for developing new applications and many libraries of routines have been written for them. Genstat is another package with very powerful language capabilities.

Chapter 15 – Conclusions

One aim in the climatic components of Instat and in writing this climatic guide has been to help users make full use of their climatic data. This is in support of agricultural research and in the many other fields that need an analysis of climatic data as part of the work.

In agricultural research, agroclimatology can provide part of the initial assessment of the suitability of different sites for cropping strategies that have been evaluated at research institutes, or in on-farm studies. Variables, such as soil type are also important, but climatic constraints are a major component of such an assessment. Within the (semi-arid) tropics, rainfall is the key climatic element. Hence a comprehensive analysis of rainfall data is a crucial component of the planning phase in agricultural research.

The main change in the recent versions of this climatic guide has been the addition of Chapters 10 to 14. The emphasis in these new chapters has been more on statistics than computing. Here we have provided guidance on how to conduct a good analysis and mentioned common pitfalls that have limited the usefulness of some studies in the past. We hope that this guidance will help users in their future work and look forward to feedback on the ideas presented. With this guide freely available over the web, the future material can also include readers' reactions, etc.

There is always more that can be done. A major change in the version of Instat introduced in 1997 was the improved macro facilities and this should also enable users to add their own macros where necessary. Chapter 14 was written to encourage users to extend the climatic library and we would welcome further macros that extend Instat's facilities further, or that make the existing facilities easier to use.

Future developments of Instat depend primarily on the users requirements and it was their demands that led to the recent versions of Instat within Windows and to the rewriting of this guide for the Windows version.

There should also be more development work, over the next few years, on providing climatic facilities within some of the powerful statistics packages, like Genstat, Systat, SAS, Matlab or R. Here we hope that Instat provides some indication, to the developers, on the climatic facilities and documentation that they could provide.

Appendix 1 – Day Numbers

Day numbers for the days of the year, starting from 1st January as day 1

Month / Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	1	32	61	92	122	153	183	214	245	275	306	336
2	2	33	62	93	123	154	184	215	246	276	307	337
3	3	34	63	94	124	155	185	216	247	277	308	338
4	4	35	64	95	125	156	186	217	248	278	309	339
5	5	36	65	96	126	157	187	218	249	279	310	340
6	6	37	66	97	127	158	188	219	250	280	311	341
7	7	38	67	98	128	159	189	220	251	281	312	342
8	8	39	68	99	129	160	190	221	252	282	313	343
9	9	40	69	100	130	161	191	222	253	283	314	344
10	10	41	70	101	131	162	192	223	254	284	315	345
11	11	42	71	102	132	163	193	224	255	285	316	346
12	12	43	72	103	133	164	194	225	256	286	317	347
13	13	44	73	104	134	165	195	226	257	287	318	348
14	14	45	74	105	135	166	196	227	258	288	319	349
15	15	46	75	106	136	167	197	228	259	289	320	350
16	16	47	76	107	137	168	198	229	260	290	321	351
17	17	48	77	108	138	169	199	230	261	291	322	352
18	18	49	78	109	139	170	200	231	262	292	323	353
19	19	50	79	110	140	171	201	232	263	293	324	354
20	20	51	80	111	141	172	202	233	264	294	325	355
21	21	52	81	112	142	173	203	234	265	295	326	356
22	22	53	82	113	143	174	204	235	266	296	327	357
23	23	54	83	114	144	175	205	236	267	297	328	358
24	24	55	84	115	145	176	206	237	268	298	329	359
25	25	56	85	116	146	177	207	238	269	299	330	360
26	26	57	86	117	147	178	208	239	270	300	331	361
27	27	58	87	118	148	179	209	240	271	301	332	362
28	28	59	88	119	149	180	210	241	272	302	333	363
29	29	60	89	120	150	181	211	242	273	303	334	364
30	30		90	121	151	182	212	243	274	304	335	365
31	31		91		152		213	244		305		366

Day numbers for the days of the year, starting from 1st July as day 1

Month	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
Day												
1	1	32	63	93	124	154	185	216	245	276	306	337
2	2	33	64	94	125	155	186	217	246	277	307	338
3	3	34	65	95	126	156	187	218	247	278	308	339
4	4	35	66	96	127	157	188	219	248	279	309	340
5	5	36	67	97	128	158	189	220	249	280	310	341
6	6	37	68	98	129	159	190	221	250	281	311	342
7	7	38	69	99	130	160	191	222	251	282	312	343
8	8	39	70	100	131	161	192	223	252	283	313	344
9	9	40	71	101	132	162	193	224	253	284	314	345
10	10	41	72	102	133	163	194	225	254	285	315	346
11	11	42	73	103	134	164	195	226	255	286	316	347
12	12	43	74	104	135	165	196	227	256	287	317	348
13	13	44	75	105	136	166	197	228	257	288	318	349
14	14	45	76	106	137	167	198	229	258	289	319	350
15	15	46	77	107	138	168	199	230	259	290	320	351
16	16	47	78	108	139	169	200	231	260	291	321	352
17	17	48	79	109	140	170	201	232	261	292	322	353
18	18	49	80	110	141	171	202	233	262	293	323	354
19	19	50	81	111	142	172	203	234	263	294	324	355
20	20	51	82	112	143	173	204	235	264	295	325	356
21	21	52	83	113	144	174	205	236	265	296	326	357
22	22	53	84	114	145	175	206	237	266	297	327	358
23	23	54	85	115	146	176	207	238	267	298	328	359
24	24	55	86	116	147	177	208	239	268	299	329	360
25	25	56	87	117	148	178	209	240	269	300	330	361
26	26	57	88	118	149	179	210	241	270	301	331	362
27	27	58	89	119	150	180	211	242	271	302	332	363
28	28	59	90	120	151	181	212	243	272	303	333	364
29	29	60	91	121	152	182	213	244	273	304	334	365
30	30	61	92	121	153	183	214		274	305	335	366
31	31	62		123		184	215		275		336	

Appendix 2 – Importing Daily Climatic Data

This appendix contains examples of different datasets that have been supplied on the SIAC (Statistics in Applied Climatology) courses and on the agro-meteorological courses run by WMO. Our thanks go to the many participants who provided the data.

The datasets have come from many different data entry systems and have sometimes had to be edited, so that they are in a format that can be read by Instat. Instructions on how to do this and the Instat macros required to import the daily data are given.

A brief description of the different data formats is given below. Choose the one that looks most appropriate for your data and then follow the more detailed instructions.

Example	Data Format	Method	Country
A2.1	Recording sheets	HONDURAS.INS	Honduras
A2.2	Clicom output	Clicom dialogue	Belize
A2.3	File with several elements	MEXICO.INS	Mexico
A2.4	Free format : days * months	VEN.INS also MARKSIM.INS	Venezuela
A2.5	Free format : months * days	THAI.INS also ZAMBIA.INS	Thailand Zambia
A2.6	Free format : monthly	MONGOLIA.INS	Mongolia
A2.7	Fixed format : days * months	MADAGASC.INS	Madagascar
A2.8	Fixed format : days * elements	ECUADOR.INS	Ecuador
A2.9	Comma delimited (CSV) format	OZ.INS	Australia

A2.1 Honduras

These data are on recording sheets. Two years of data were typed into a file, called **Honduras.dat**. The layout was designed for Instat. To input, commands were added to the file and the name changed to **Honduras.ins**, see Fig. A2.2a.

Fig. A2.2a Data from Honduras

```
 NOTe Create a file with 400 rows - this allows for mistakes
 NOTe the data if more than 366 numbers entered!!

 CREate @HONDURUS; MISS 9999 8888 9988 ; COL 50 400
Title: 2 years of data from Hondurus
ENTer 'y81
 31(0)
 0 0 1.4 0.6 7(0) 1 5(0) 0.3 27.3 3.2 8(0) 9988
 4(0) 4.5 8(0) 0.5 0 0.6 18.4 1.1 13(0)
 ... ... ... ... ...
 3(0) 8.5 3.3 12(0) 31.8 .1 12(0)
 EOD

ent 'y82
5(0) .3 .1 9(0) .5 .9 0 1.9 11(0)
0 .3 3(0) 1 1.2 .4 9(0) .4 10(0) 9988
... ... ... ... ...
4(0) 2 .9 9(0) .1 8(0) .4 .3 0 2.6 3(0)

DAY X1 X2; CODe 0
```

A2.2 Belize – Clicom output

The data from Belize have been exported from Clicom as described in Chapter 3, Section 7.

The datafile, BELIZE.DAT contains two years of daily precipitation data, (element 005), and one year of both maximum temperature (002) and minimum temperature (003).

Fig. A2.1a Data from Belize as exported from Clicom - BELIZE.DAT

```
     Site  Element Year Month         Daily data
      ↓        ↓     ↓    ↓         ↓
001,PUNTAGOR ,002, ,1984-01,23.9,,23.9,,.... January  Max Temp 1984
001,PUNTAGOR ,003, ,1984-01,15.6,,15,,15....          Min Temp
001,PUNTAGOR ,005, ,1984-01,0,,0,,0,,0,,....          Precip.
001,PUNTAGOR ,002, ,1984-02,28.3,,28.9,,.... February Max Temp
001,PUNTAGOR ,003, ,1984-02,17.2,,18.9,,....          Min Temp
001,PUNTAGOR ,005, ,1984-02,0,,17.5,,0,,....          Precip.
.........................................
001,PUNTAGOR ,002, ,1984-12,31.1,,30.8,,.... December Max Temp
001,PUNTAGOR ,003, ,1984-12,18.9,,18.3,,....          Min Temp
001,PUNTAGOR ,005, ,1984-12,0,,46.5,,0,,....          Precip.
001,PUNTAGOR ,005, ,1985-01,23.1,,0,,0,,.... January  Precip.  1985
001,PUNTAGOR ,005, ,1985-02,0,,0,,2.3,,1.... February Precip.
............................................
001,PUNTAGOR ,005, ,1985-12,3.3,,0.8,,6.3... December Precip.
```

To import the data into Instat, use the Climatic ⇒ Manage ⇒ Import Daily dialogue after using Climatic ⇒ Manage ⇒ New, to create a new worksheet.

Alternatively the commands in Fig. A2.2b are in a macro called **Belize.ins**.

Fig. A2.2b Importing data in Clicom export format

```
CREate @BELIZE; MISsing 9999 8888 9988; COLumns 100 366;
      TITle "Belize Data for 1984, 1985"
NOTe put the file Belize.DAT into the working directory
NOTe To import 2 years of precipitation data
CLIcom @BELIZE.DAT; ELEment 005
NAMe X1 'Rain84 : NAMe X2 'Rain85

NOTe To import 1 year of maximum temperature data
CLIcom @BELIZE.DAT; ELEment 002; INTo X3
NAMe X3 'Tmax84

NOTe To import 1 year of minimum temperature data:
CLIcom @BELIZE.DAT; ELEment 003; INTo X4:
NAMe X4 'Tmin84

NOTe To display BELIZE data:
DAY X1-X4; CODe 0
```

Sometimes Clicom data in the form given in Fig. A2.1 have been put into Excel first. This makes it difficult to import into Instat, as the spacing is affected. In such instances the simplest solution is usually to get another set of data from Clicom and avoid Excel.

A2.3 Mexico – entry from yearly datafiles

The files, MEXICO83.DAT and MEXICO84.DAT contain not only precipitation data, but also other climatic elements. This typifies data where there are multiple files, with one for each year. A macro, called mexico.ins, is written to import them all.

Fig. A2.3a Importing data from Mexico

```
CREate @MEXICO;col 50 366; miss 9999 8888 9988;
       title "Two years of data from separate files"

NOTe In MEXICO83.DAT the 5th col. has the precipitation data
READ 'y83; FILe @MEXICO83.DAT; FIEld 5
NOTe In 1983, a non-leap year, there is no value for Feb. 29th
NOTe   So this is INSerted
INSert 60 '83; DATa 9988

NOTe In MEXICO84.DAT the 6th col. has the precipitation data
READ 'y84; FILe @MEXICO84.DAT; FIEld 6

DAY 'y83 'y84; CODe 0
```

Fig. A2.3b The Mexico data files

```
Filename: MEXICO83.DAT
Year   Month    Day               Precipitation
 |       |       |                  5th field
 ↓       ↓       ↓                      ↓
83       1       1        1.30        0.00         5.0        13.5
83       1       2        5.40        0.00         5.5        19.5
83       1       3        3.70        0.00         2.5        19.5
...........................................................
83       2      27        4.50        0.00        -4.5        20.5
83       2      28        5.90        0.00        -2.0        23.0    No Feb 29th
83       3       1        6.20        0.00         1.5        25.5
83       3       2       11.20        0.00         1.0        23.0
...........................................................
83      12      30        2.80        0.00         0.5        17.5
83      12      31        2.80        0.00        -4.0        19.0
-------------------------------------------------------------------
Filename: MEXICO84.DAT
Year    Month     Day                       Precipitation
 |       |         |                        6th field
 ↓       ↓         ↓                            ↓
84       1       1        3.99        4.16         0.0        -2.5
84       1       2        2.85        3.04         0.0        -2.0
...........................................................
84       2      28        5.39        5.53         0.0        -1.0
84       2      29        5.19        5.29         0.0        -0.5   Leap Year
84       3       1        5.28        4.96         0.0         8.0
84       3       2        6.10        6.12         0.0         3.0
...........................................................
84      12      30        3.03        3.25         0.0         4.0
84      12      31        3.44        3.52         0.0         5.0
```

A2.4 Venezuela

Description of original data file

There are extra rows of text at the start and end of each year. This extra text is ;OMItted by the REAd command in VEN.INS. Several years of data are in one file.

Daily rainfall data in format:

Rows = days of month
Columns = months

i.e. 31 rows by 14 columns (Day, Jan, Feb, Mar, ...Dec, Day)

There are blank fields in months that have < 31 days. These blank fields are padded with -1's to make each column the same length for the REAd command to work.

1. Edit your dataset until it has a similar format to VEN.DAT
2. Edit as appropriate for your data, the <u>Underlined</u> text in VEN.INS
3. Enter INSTAT and run the macro VEN.INS

The macro asks about the name and format of the data file, Fig. A3.4a.

When the macro has finished running, there will be a worksheet and the results from the DAY command. The macro automatically chooses a name for these files, based on the name of the data file. For example, if the data file is VEN.DAT, the worksheet will be called VEN.WOR. The results are for the user to check that the data have been entered correctly.

This format is similar to the layout of data from the MarkSim (2002) software. There is therefore a second macro, called marksim.ins, which is an alternative. This uses fixed format input, so the "missing" days at the end of some months do not need to be padded. A sample datafile is called palmira-.gen.

Fig. A2.4a Running the macro to import data from Venezuela

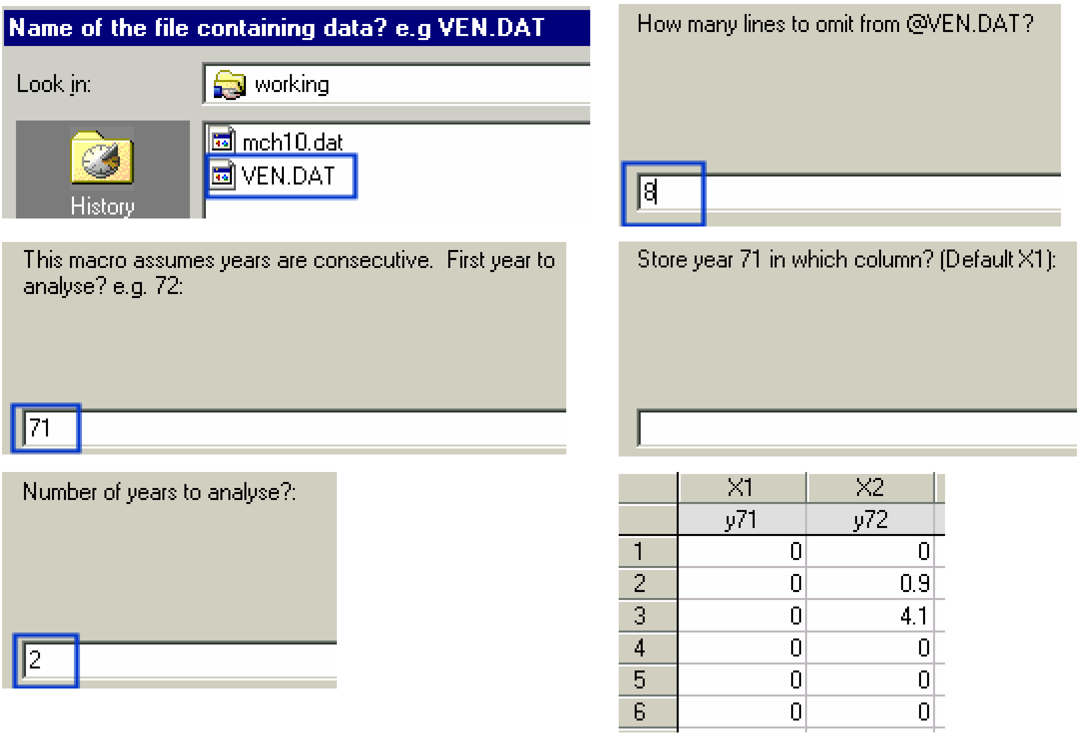


Fig. A2.4b Daily data from Venezuela - original file - VENEZ.DAT

```
DIRECCION DE HIDROLOGIA Y METEOROLOGIA
                                        PRECIPITACION 08000-0800 (mm).
ESTACION: SANTA CRUZ EDAFOLOGI  SERIAL: 0417   Zona: 02  Edo: ARAGUA   AÑO: 1971
Latitud: 10□10'00"  Longitud: 67□29'15"  Altitud: 444 m  Ins: 06/64  Elim: 00/00
***************************************************************************
DIA   ENE   FEB   MAR   ABR   MAY   JUN   JUL   AGO   SEP   OCT   NOV   DIC
***************************************************************************
  1    .0    .0    .0    .0    .0  19.6    .0  10.2    .0    .6   9.1    .7
  2    .0    .0    .3    .0    .0    .7    .6    .0   1.2    .0    .0   7.6
  3    .0    .0   6.0    .0   3.0   9.0    .0    .1    .0    .0    .0    .6
  4    .0    .0  24.7    .0   1.8    .0   1.0   5.6   3.0    .0    .0    .0
  5    .0    .0    .0    .0    .0   5.5    .0   3.1  19.0   2.9   4.4   3.9
  6    .0    .0    .0    .0    .0    .0    .0    .0  10.9    .0   3.2    .0
  7    .0    .0    .0    .0    .0    .0   5.5    .0   5.4    .0  14.0    .0
  8    .0    .0    .0  11.2    .0   8.3  38.0   8.6  11.9    .0    .0    .0
  9    .0    .0    .0    .8    .0  11.1  10.7   2.2    .0    .0    .0    .0
 10    .0   4.3    .0    .0  39.6   4.9    .3   5.5    .0   2.3    .0   8.0

 11    .1    .0    .0    .0    .0   3.9  16.0  42.3    .0    .0    .7    .0
 12    .0    .0    .0    .0   2.4    .0  25.0   3.7   2.9   1.6    .2    .0
..  .. ..   ..   .. ..   ..  .. ..   ..  .. ..   ..  .. ..   ..  .. ..   ..
 30    .0          .0    .0    .0    .0    .0   4.2  11.0   6.0  33.5    .0
 31    .0          .0          .8        10.6    .0          .2          .4

TOT 10.6    4.3  32.7  17.8 149.3 115.6 260.5 228.9 105.4 113.2  75.8  22.8

MAX  9.6    4.3  24.7  11.2  39.6  23.2  38.8  42.3  19.0  25.6  33.5   8.0
DIAS  20     10     4     8    10    15    26    11     5    22    30    10

TOTAL ANUAL:  1136.9     MAXIMA:   42.3    EL    11/ 8
..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..  ..
DIRECCION DE HIDROLOGIA Y METEOROLOGIA
                                        PRECIPITACION 08000-0800 (mm).
ESTACION: SANTA CRUZ EDAFOLOGI  SERIAL: 0417   Zona: 02  Edo: ARAGUA   AÑO: 1972
Latitud: 10□10'00"  Longitud: 67□29'15"  Altitud: 444 m  Ins: 06/64  Elim: 00/00
***************************************************************************
DIA   ENE   FEB   MAR   ABR   MAY   JUN   JUL   AGO   SEP   OCT   NOV   DIC
***************************************************************************
  1    .0   1.5    .0    .0   9.0  25.0   1.1  13.9  32.8    .0    .0    .0
  2    .9    .0    .0    .0    .0    .0    .0    .0    .0    .4    .1    .0
```

Fig. A2.4c Daily data from Venezuela - edited file - VEN.DAT

```
M. A. R. N. R.  -  D. G. S. I. C. A. S. V.                                      line 1
DIRECCION DE HIDROLOGIA Y METEOROLOGIA
                                        PRECIPITACION 08000-0800 (mm).
ESTACION: SANTA CRUZ EDAFOLOGI  SERIAL: 0417   Zona: 02  Edo: ARAGUA   AÑO:
1971
Latitud: 10□10'00"  Longitud: 67□29'15"  Altitud: 444 m  Ins: 06/64  Elim:
00/00
***************************************************************************
DIA   ENE   FEB   MAR   ABR   MAY   JUN   JUL   AGO   SEP   OCT   NOV   DIC
***************************************************************************
  1    .0    .0    .0    .0    .0  19.6    .0  10.2    .0    .6   9.1    .7     line 9
  2    .0    .0    .3    .0    .0    .7    .6    .0   1.2    .0    .0   7.6
  3    .0    .0   6.0    .0   3.0   9.0    .0    .1    .0    .0    .0    .6
  4    .0    .0  24.7    .0   1.8    .0   1.0   5.6   3.0    .0    .0    .0
  5    .0    .0    .0    .0    .0   5.5    .0   3.1  19.0   2.9   4.4   3.9
  6    .0    .0    .0    .0    .0    .0    .0    .0  10.9    .0   3.2    .0
  7    .0    .0    .0    .0    .0    .0   5.5    .0   5.4    .0  14.0    .0
  8    .0    .0    .0  11.2    .0   8.3  38.0   8.6  11.9    .0    .0    .0
  9    .0    .0    .0    .8    .0  11.1  10.7   2.2    .0    .0    .0    .0     blank line
 10    .0   4.3    .0    .0  39.6   4.9    .3   5.5    .0   2.3    .0   8.0     erased
 11    .1    .0    .0    .0    .0   3.9  16.0  42.3    .0    .0    .7    .0
..  .. ..   ..   .. ..   ..  .. ..   ..  .. ..   ..  .. ..   ..  .. ..   ..
 30    .0    -1    .0    .0    .0    .0    .0   4.2  11.0   6.0  33.5    .0     -1 inserted
 31    .0    -1    .0    -1    .8    -1  10.6    .0    -1    .2    -1    .4
                                                                                blank line to
TOT 10.6    4.3  32.7  17.8 149.3 115.6 260.5 228.9 105.4 113.2  75.8  22.8     end data input

MAX  9.6    4.3  24.7  11.2  39.6  23.2  38.8  42.3  19.0  25.6  33.5   8.0
DIAS  20     10     4     8    10    15    26    11     5    22    30    10
.................................................................
```

A2.5 Thailand – using macros

Description of original data file

Daily rainfall data in format:

Rows = months (*The number of rows is (number of years * 12)*)
Columns = days

N.B. There are extra columns at the start of each row, so formatted input is needed. Thus, there are 34 columns (station, year, month, day 1 ... day 31)

The blank fields in months with < 31 days, was padded with -1's to make each row is complete.

1. Edit the dataset if necessary, so it has a similar format to THAI.DAT, *Figs A2.5a* and *2.5b*
2. Edit the macro as appropriate for your data
3. Run the macro THAI.INS

You will be asked questions about the name and format of the data file, as in *Fig. A2-4a*.

When the macro has run, there will be a worksheet with the data plus results from the DAY command. The macro automatically chooses a name for these files, based on the name of your data file. For example, if the data file is THAI.DAT, the worksheet will be called THAI.WOR. The results are so the data can easily be checked.

Fig. A2.5a Daily data from Thailand - original file - THAILAND.DAT

```
Station
 |       Year                                                        Blank fields where
 |       |  Month                                                months have < 31 days
 ↓       ↓    ↓                                                                  ↓
814556011951 1   0   0   0  ..  ..   0   0   2   0   0   0   0   0   0
814556011951 2   0   0   0  ..  ..   0   0   0   0   0   0   0
814556011951 3   0   0   0  ..  ..   0   0   0   0   0   0   0   0   0   0
.......................................................................
814556011952 1   0   0   0  ..  ..   0   0   0   0   0   0   0   0   0   0
814556011952 2   0   0   0  ..  ..   0   0   0   0   0   0   0   0
814556011952 3   0   0  33  ..  ..   0  43   1 106  24  95  27   0   1 221
.......................................................................
81455601195211   0   0   0  ..  ..   0   0   0   0   0   0   0   0   0
81455601195212   0   0  26  ..  ..   0   0   0   0   0   0   0   0   0   0
```

Fig. A2.5b Daily data from Thailand - edited file - THAI.DAT

```
Station
 |       Year                                                    -1's inserted where
 |        |  Month                                             months have < 31 days
 ↓        ↓   ↓                                                                ↓
814556011951 1   0   0   0  ..  ..   0   0   0   2   0   0   0   0   0   0
814556011951 2   0   0   0  ..  ..   0   0   0   0   0   0   0  -1  -1  -1
814556011951 3   0   0   0  ..  ..   0   0   0   0   0   0   0   0   0   0
.......................................................................
814556011952 1   0   0   0  ..  ..   0   0   0   0   0   0   0   0   0   0
814556011952 2   0   0   0  ..  ..   0   0   0   0   0   0   0   0  -1  -1
814556011952 3   0   0  33  ..  ..   0  43   1 106  24  95  27   0   1 221
.......................................................................
81455601195211   0   0   0  ..  ..   0   0   0   0   0   0   0   0   0  -1
81455601195212   0   0  26  ..  ..   0   0   0   0   0   0   0   0   0   0
```

Data from Zambia had a similar format, Fig. A2.5c. Hence the macro for Thailand was later adapted and one written called ZAMBIA.INS. These data already had –99 where the months were short, so the data files did not need to be edited. A similar file, with comma-delimited values, was sent from Kenya: the importing macro is called KENYA.INS.

Fig. A2.5c Daily data from Zambia - LIVING.DAT

```
'LIVING01' 1950  1   14.7     .0    1.8  .. .. .. ..  9.7     .0     .0     .0     .0
'LIVING01' 1950  2     .0   16.5    2.0  .. .. .. ..  8.1    3.8  -99.0  -99.0  -99.0
'LIVING01' 1950  3     .0    2.0     .3  .. .. .. ..   .0     .8    2.5     .5     .0
'LIVING01' 1950  4    2.8   13.0    6.3  .. .. .. ..   .0     .0     .0     .0  -99.0
'LIVING01' 1950  5     .0     .0     .0  .. .. .. ..   .0     .0     .0     .0     .0
```

A2.6 Mongolia

This macro shows how the data can be checked before being processed, to make sure that the original data are correct.

The data file MONGOLIA.DAT has (2 years * 12 months * 35) numbers. Each month of 35 values consists of: Site, Year, Month, 31 rainfall data, Monthly total. The macro MONGOLIA.INS checks

a) Site numbers are all the same
b) Year numbers are consecutive
c) Months go from 1-12
d) Total rainfall for each month is equal to the sum of the monthly data

Your dataset should look something like MONGOLIA.DAT
You will be asked questions about the name and format of the data file, as in Fig. A2.4a.

When the macro has finished running, there are results from the DAY command to check that the importing has worked correctly. The macro automatically chooses a name for the file, based on the name of the data file. For example, if the data file is MONGOLIA.DAT, the worksheet will be called MONGOLIA.WOR.

Fig. A2.6a Daily data from Mongolia

```
Site Year Month  January data  January total = 44mm
  ↓   ↓    ↓           ↓
 240  84   1   0   0   0   0   0   0   0   9   4   3   0   0
   0   0   0   0   1   0   0   1   0   0   0   0  17   0   1
   8   0   0   0  44 240  85   1   0   0   0   0   0   5   0
   0   0   0   0   0   0   0   0   0   0   4   0   0   0   0
   0   0   0   0   0   0   0   0   0   9 240  84   2   0   0
   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0   0   0   0   0   3   0   2   0   0   5
 240  85   2   4   0   0   0   0   0   0   0   0   0  16  41
   0   0   0   0   0   0   0   0   0   0   0   0   0   0   5
   0   0   0   0  66 240  84   3   0   3   0   0   0   0   2
   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
  31   0   0   0   0   0   0   0   0  36 240  85   3   0   0
   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
 240  84   4   0   0   0   0   2   0   0   0   0   0   0   0
   0   0   0   5   0   0  56   4   0   0   0   0   0   0   4
   0   0   0   0  71 240  85   4   0   0   0   0   0   0   2
   0   0   0   0   0   0   0   0   0  51   0   0   0   0   0
   0   0   0   0   0   0   0   0   0  53 240  84   5  15   0
   0   0   0   0   0   0   0   0   0   0   0   0  36   0   0
   0   0   0   0   0  82   0   0   0   0   3   0 200   0 336
 240  85   5   0   0   0   0   0   1   7   0   1   1   1   1
   0   0  75   7   0   0   0   0   0   4   0   0   0   1   0
   1   0 130   1 231
 240  84   6   0   0   0  37 254  36  17  16   0   0   0   0
   0   2   0   0  47  11   0   0   0   0   0   0 110 303  92
 103   0   5   1 1034 240  85   6 122   0   0   0   0   0   1
  51  49   0  27 159  28   0   0   0   0   0   0  52   0  12
   2   0  63  59   1   0   1   1   0 628
```

A2.7 Madagascar

The rainfall data file MADAGASC.DAT is in fixed format:

Columns 1-6 Site code
Columns 7-10 Year
Columns 11,12 Month
Column 13 Row number (1, 2 or 3)
Columns 14-57 11 data values, each of width 4. In rows numbered 1 and 2 the 11th value has to be discarded, so each month has 31 values. The original data were multiplied by 10, for easier data entry. Thus they have to be divided by 10 to get the actual value.

1. Your dataset should look something like MADAGASC.DAT
2. Edit MADAGASC.INS as appropriate for your data

When the macro is run, there will be questions about the name and format of the data file, as in Fig. A2.4a.

When the macro has finished running, there will be a worksheet and results, containing output from the DAY command, so the importing can be checked. The macro automatically chooses a name for these files, based on the name of your data file. For example, if the data file is MADAGASC.DAT, the worksheet will be called MADAGASC.WOR.

Fig. A2.7a Daily data from Madagascar

```
Site  Year       (Columns 14-57 contain 11 data values, each of
 |      | Month   width 4. e.g. 0004, 0006. In rows numbered 1 & 2
 |      |   | Row  the 11th value is discarded, so each month has
 ↓      ↓   ↓ ↓    31 days)
534155195201100040006001000030121014100000000002480294000  Jan.52
534155195201200000000000450135070903970056017602000238000  row 2
534155195201300000134050000140000012601800104011502150024  row 3
534155195202101170102070902040113010501440000000000000000  Feb.
534155195202200000000000000116024700000000000000000250000
534155195202300200000000000000000000000000000000000000000
534155195203100000000000000000141013000000000000001900000  March
534155195203202810140001400000240032000510000000000000000
534155195203300000000000000000089000001800000000000000000
.........................................................
.........................................................
.........................................................
534155195312102140023018000000000201300000002510000000000  Dec.53
534155195312200260000000320000000020004000000000005000000000
534155195312300230000000100050023000000000000000120000014
```

A2.8 Ecuador

The data from Ecuador are in one file, ECUADOR.DAT. They are in columns with readings every eight hours for precipitation, and daily Tmax and Tmin. The data have to be summed to get the daily precipitation. A check is also made to ensure that the data are for a full year and that Tmax is > Tmin. There are two stages to the data input.

The data for the first year are read into a worksheet, checks are made that the data are complete, missing values are recoded, the 8 hourly precipitation are summed and the three elements (Precipitation, Tmax and Tmin) are transferred to worksheets **ECUAPREC.WOR** and **ECUATEMP.WOR**, before the next year is processed.

1. Your dataset should look something like ECUADOR.DAT
2. Edit **ECUADOR.INS** as appropriate for your data.

When the macro is run, you will be asked about the name of the data file, as in Fig. A2.4A.

When the macro has finished running, there will be two worksheets and results from the DAY command. The macro automatically chooses a name for the file, based on the name of your data file. For example, if the data file is **ECUADOR.DAT**, the rainfall worksheet will be called **ECUAPREC.WOR** and the temperature worksheet will be **ECUATEMP.WOR**.

Fig. A2.8a Daily Data from Ecuador - ECUADOR.DAT

```
SiteYear
 |   | Month Precipitation TMax TMin
 |   |  | Day  (8 hourly)   |    |
 ↓   ↓  ↓ ↓   ↓      ↓   ↓    ↓    ↓
M00390 1 1   .0    .0    .7 19.4  4.3     January
M00390 1 2   .0    .0    .0 19.4  4.6
M00390 1 3   .0    .0    .0 21.2  4.0
M00390 1 4   .0    .0    .4 20.4  3.0
M00390 1 5   .0    .0    .0 19.5  3.0
M00390 1 6   .0    .0    .0 99.9  3.1     Temperature missing values = 99.9
M00390 1 7   .0    .0   5.1 20.5  1.7
M00390 1 8   .0    .0  19.1 18.7  7.5
M00390 1 9   .0    .0  14.8 17.2  6.8
...................................
M003901218   .0    .0    .0 21.0  6.5      December
M003901219   .0    .0    .0 20.2  6.4
M003901220999.9999.9999.9 20.4  5.4      Precip. missing data = 999.9
M003901221   .0    .0   2.2 21.4  3.8
M003901222   .0    .0   4.7 18.4  6.5
M003901223   .0    .0    .0 18.9  7.9
M003901224   .0    .0888.8 18.3  8.0
M003901225   .0    .0  11.9 20.6  5.9
M003901226   .0    .0   2.0 19.6  4.0
M003901227   .0    .0   1.8 17.8  6.0
M003901228   .0    .0   3.4 19.2  5.0
M003901229   .0    .0  21.8 19.2  4.1
M003901230   .0    .0  13.7 15.6  6.5
M003901231   .0    .0    .6 15.4  7.4
m00391 1 1   .0    .0    .1 17.8  3.6     Year 2
m00391 1 2   .0    .0    .0 18.0  4.0
...................................
M003911229   .0    .0    .6 17.0  7.4
M003911230   .0    .0   1.1 19.4  5.0
M003911231   .0    .0    .0 19.3  5.0
M00392 1 1   .0    .0    .0 20.4  3.9     Year 3
M00392 1 2   .0    .0   6.3 19.8  6.5
...................................
M00392 228   .0    .0    .3 19.3  7.1
M00392 229   .0    .0    .0 19.0 10.6     Leap year
M00392 3 1   .0    .0    .0 22.2  7.8
...................................
M003921229   .0    .0    .0 20.4  3.1
M003921230   .0    .0   4.1 19.8  7.0
M003921231   .0    .0  19.8 17.1  6.6
```

A2.9 Australia

The data are in the file OZ.DAT in comma delimited format.

Field 1 Site code
Field 2 Year
Field 3 Month
Field 4 Precipitation
Fields 5,7,9 Blank
Field 6 TMax
Field 8 Tmin

1. Your dataset should look something like OZ.DAT
2. Edit OZ.INS as appropriate for your data.

When the macro is run, there are questions about the name and format of your data file. Within the macro, a check is made that TMax is greater than Tmin.

When the macro has finished running, there will be a worksheet and results from the DAY command, so a check can be made that the data have been imported correctly. The macro automatically chooses a name for these files, based on the name of your data file. For example, if the data file is OZ.DAT, the worksheet will be called OZ.WOR.

Fig. A2.9a Daily data from Australia

```
Site    Year
                Month        Precipitation
                     Day                        TMax         TMin
 87100,1963,  01, 01,       0.0,    ,29.7,    ,   9.9, Jan. 1963
 87100,1963,  01, 02,       0.0,    ,29.7,    ,   9.9,
 87100,1963,  01, 03,       0.0,    ,30.8,    ,   6.6,
 87100,1963,  01, 04,       0.0,    ,31.9,    ,  14.3,
 87100,1963,  01, 05,       0.0,    ,29.7,    ,  13.2,
.........................................................
 87100,1963,  01, 28,       0.0,    ,23.1,    ,   6.6,
 87100,1963,  01, 29,       0.0,    ,22.0,    ,  12.1,
 87100,1963,  01, 30,       0.0,    ,19.8,    ,  12.1,
 87100,1963,  01, 31,       0.0,    ,23.1,    ,   7.7,
 87100,1963,  02, 01,       0.0,    ,26.4,    ,   8.8, Feb. 1963
 87100,1963,  02, 02,       0.0,    ,29.7,    ,   9.9,
 87100,1963,  02, 03,       0.0,    ,30.8,    ,   8.8,
.........................................................
 87100,1963,  02, 27,       0.0,    ,36.3,    ,  15.4,
 87100,1963,  02, 28,       0.0,    ,37.4,    ,  15.4, Non-leap
 87100,1963,  03, 01,       0.0,    ,37.4,    ,  16.5, Year
 87100,1963,  03, 02,       0.0,    ,35.2,    ,  17.6,
.........................................................
 87100,1963,  10, 30,       0.0,    ,28.6,    ,  28.6,
 87100,1963,  10, 31,       0.0,    ,33.0,    ,  17.6,
 87100,1963,  11, 01,      9999,    ,9999,    ,  9999, Nov. 1963
 87100,1963,  11, 02,      9999,    ,9999,    ,  9999,  Missing
 87100,1963,  11, 03,      9999,    ,9999,    ,  9999,  =9999
 87100,1963,  11, 04,      9999,    ,9999,    ,  9999,
.........................................................
 87100,1963,  12, 31,       0.0,    ,31.9,    ,   9.9,
 87100,1964,  01, 01,       0.0,    ,31.9,    ,   9.9, Jan. 1964
 87100,1964,  01, 02,       0.0,    ,31.9,    ,  12.1,
 87100,1964,  01, 03,       0.0,    ,31.9,    ,  11.0,
.........................................................
 87100,1964,  02, 27,       0.0,    ,33.0,    ,  13.2,
 87100,1964,  02, 28,       0.0,    ,33.0,    ,  13.2,
 87100,1964,  02, 29,       0.0,    ,33.0,    ,  14.3, Leap Year
 87100,1964,  03, 01,       0.0,    ,35.2,    ,  14.3,
.........................................................
 87100,1964,  12, 29,       0.0,    ,30.8,    ,  11.0,
 87100,1964,  12, 30,       0.0,    ,30.8,    ,  11.0,
 87100,1964,  12, 31,       0.0,    ,29.7,    ,   9.9,
```

References

R.G. Allen, R.G.,L.S. Pereira, L.S., Raes, D., and Smith, M. (1998). Crop Evapotranspiration – Guidelines for Computing Crop Water Requirements. FAO Irrigation and Drainage Paper 56.

Baradas, M.W. and Sutrisno, M. (1981). Forecasting monthly area harvested for wetland paddy in Java. Working paper 20, GOI/UNDP/WMO project INS/78/042. BMG, Jakarta.

Batschelet, E. (1981). Circular Statistics in Biology. Academic Press, London, 371pp.

Direction Nationale de la Météorologie, République du Mali, 1998. Extension de l'assistance météorologique opérationnelle au monde rural. Direction Nationale de Météorologie, Bamako.

Doorendos, J. and Pruitt, W.O. (1977). Crop Water Requirements. FAO Irrigation and Drainage Paper 24, revised edition, 144pp.

FAO, (1978). Report on the Agro-ecological zones project. Food and Agriculture Organization of the United Nations, Rome.

FAO (1986). Early Agrometeorological Crop Yield Assessment. FAO Plant Production and Protection Paper 73, 150pp.

Franquin, P., and Rijks, D. (1983). Index des paramètres météorologiques et des informations agrométéorologiques qui en découlent. Bull. OEPP 13: 322-331.

Frère, M. and Popov, G. (1979). Agrometeorological Crop Monitoring and Forecasting. FAO Plant Production and Protection Paper 17, 64pp.

Friesland, H., Kersebaum K.C., and Lopmeier, F.J. (1998). Operational use of irrigation models using medium range weather forecasts. COST 711 report. Deutsche Wetterdienst, Offenbach.

Gerbier, N., and Remois, P. (1977). Influence du climat sur la qualité et la production du vin de Champagne. Monographie 106. Météorologie Nationale, Paris.

Gommes, R.A. (1983). Pocket Computers in Agrometeorology. FAO Plant Production and Protection Paper 45, 140pp.

Haan, C.T. (1977). Statistical Methods in Hydrology. Iowa State University Press, 378pp.

Huda, A.K.S., Ghildyal, B.P., Tomar, V.S. and Jain, R.C. (1975). Contribution of climatic variables in predicting rice yields. Agricultural Meteorology 15: 17-86

Kowal, J.M. and Knabe D.T. (1972). An Agroclimatological Atlas of the Northern States of Nigeria. Ahmadu Bello University Press, Zaria, Nigeria.

Manning, H.L. (1956). The statistical assessment of rainfall probability and its application in Uganda agriculture. Proc. R. Soc. Lond. B 144: 460-480.

Mardia, K.V. (1972). Statistics for Directional Data

Mellaart, E.A.R., Kruger, J., Holmes, M., and Human, N. (1999). Windbreak research for improving fruit quality. In: CAgM report 77, Contributions from members on operational applications in agrometeorology for the International Workshop "Agrometeorology in the 21st Century, Needs and Perspectives", held in Accra, February 1999. WMO, Geneva.

Morgan, B.J.T. (1983). Elements of Simulation. Chapman and Hall, 351 pp.

da Mota, F.S. and da Silva, J.B. (1980). A weather-technology model for rice in Southern Brazil. In Proceedings of a Symposium on the Agrometeorology of the Rice Crop, 235-238. IRRI, Los Baños, Phillipines.

Payne, W.A. (1997). Managing yield and soil water use of pearl millet in the Sahel. Agronomy Journal 89(3)

Penman, H.L. (1948). Natural evaporation from open water, bare soil and grass- Proc. Roy. Soc. (London) 193: 120-145.

Rainey, R.C., Browning, K.A., Cheke R.A., and Haggis, M.J. (1990). Migrant pests: Progress, Problems and Potentialities. The Royal Society, London.

Raman, C.V.R. (1974). Analysis of commencement of monsoon rains over Maharashtra State for agricultural planning. Scientific Report 216, India Meteorological Department, Pune, India.

Rijks, D. (1976). Agrometeorology. In: M.H. Arnold (Editor), Agricultural research for development. Cambridge University Press, Cambridge.

Rijks. D. (1977). The conservation and utilization of water. Phil. Trans. R. Soc. Lond. B 278: 583-593.

Rijks, D. (1986). The environment - assessing the problems, finding solutions. In: Ted J. Davis (Editor), Development of rainfed agriculture under arid and semiarid conditions. The World Bank, Washington D.C.

Rijks, D. (1987). Agrometeorology, a tool in land and water management. In: J. Kozub (Editor), Land and water resources management. Economic Development Institute of the World Bank, Washington D.C.

Rijks, D. (1994). Agroclimatic characterization. In: J.F. Griffiths (Editor), Handbook of Agricultural Meteorology. Oxford University Press, New York, Oxford.

Rijks, D. (2000). The clients for agrometeorological information. Agric. For. Meteorol. 103, 27-42

Rijks, D., and Gbeckor-Kove, N. (1990). Agrometeorological information for effective irrigation scheduling. Acta Horticulturae 278: 833-840.

Sivakumar, M.V.K., Maidoukia, A. and Stern, R.D. (1993). Agroclimatology of West Africa: Niger. ICRISAT Information Bulletin 5, Niamey, Niger, 108pp.

Smith, M. (2000). The application of climate data for planning and management of sustainable rainfed and irrigated crop production. Agric. For. Meteorol. 103, 99-108.

Stigter, C. (1994). Management and manipulation of microclimate. In: J.F. Griffiths (Editor), Handbook of Agricultural Meteorology. Oxford University Press, New York, Oxford.

Sooryarachchi, M.R. (1989). Estimation of probable rainfall for irrigation and water management in the tropics. MSc (Biometry) dissertation, Department of Applied Statistics, The University of Reading, UK.

Snyder R.L. (1985). Agricultural and Forest Meteorology, Volume 35: 353-355.

Stern, R.D. and Coe, R. (1984). A model fitting analysis of daily rainfall data. J. Royal Statistical Society (A), 147: 1-34.

Stern, R.D., Dennett, M.D. and Dale, I.C. (1982). The analysis of daily rainfall measurements to give agronomically useful results, 1: Direct Methods. J. Experimental Agric., 18: 223-236. Cambridge University Press.

Stern, R.D., Dennett, M.D. and Dale, I.C. (1982). The analysis of daily rainfall measurements to give agronomically useful results, 2: A Modelling Approach. J. Experimental Agric., 18: 237-253. Cambridge University Press.

Stern, R.D. and Dale, I.C. (1983). Statistical methods for tropical drought analysis based on rainfall data. WMO (PRTM) Report 9, 59pp.

Strydom, J. (1999). An agrometeorological information service to wine farmers. In: CAgM report 77, Contributions from members on operational applications in agrometeorology for the International Workshop "Agrometeorology in the 21st Century, Needs and Perspectives", held in Accra, February 1999. WMO, Geneva.

Traore, K., Diarra, B., and Konare, K. (1992). Elaboration et mise-en-oeuvre de projets pilotes en agrométéorologie. In: La radio rurale et la diffusion des informations agrométéorologiques. WMO, Genève

Virmani S.M., Sivakumar, M.V.K. and Reddy, S.J. (1978). Rainfall probability estimates for selected locations of semi-arid India. Research Report 1, ICRISAT, Hydrabad, India.

Woolhiser et al. (1989). Modelling rainfall: Computer program for modelling daily climatic data.

Young, C.P. (1963). A computer programme for the calculation of mean rates of evaporation using Penman's formula. Met. Mag., 92: 84-88.

Index