

WORLD CLIMATE PROGRAMME

DATA and MONITORING



WORLD CLIMATE PROGRAMME – WATER

**DETECTING TREND AND OTHER CHANGES
IN HYDROLOGICAL DATA**

Zbigniew W. Kundzewicz and Alice Robson (Editors)

WCDMP – 45

WMO/TD-No. 1013

(Geneva, May 2000)

UNITED NATIONS EDUCATIONAL
SCIENTIFIC AND CULTURAL
ORGANIZATION

WORLD METEOROLOGICAL
ORGANIZATION

The WCP implemented by WMO in conjunction with other international organizations consists of four major components:

The World Climate Data and Monitoring Programme (WCDMP)

The World Climate Applications and Services Programme (WCASP)

The World Climate Impact Assessment and Response Strategies Programme (WCIRP)

The World Climate Research Programme (WCRP)

World Meteorological Organization
Case postale N 2300
1211 Geneva
Switzerland

World Climate Data and Monitoring Programme

Telephone: (+41-22) 730 81 11

Telefax: (+41-22) 730 80 42

Telex: 41 41 99

Email: wcdmp@gateway.wmo.ch

World Wide Web: <http://www.wmo.ch/web/wcp/wcdmp/wcdmp.html>

NOTE

The designations employed and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the World Meteorological Organization concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Editorial note: This report has for the greater part been produced without editorial revision by the WMO Secretariat. It is not an official publication and its distribution in this form does not imply endorsement by the Organization of the ideas expressed.

TABLE OF CONTENTS

List of participants in the Wallingford Workshop, 2-4 December 1998	iii
Other contributors to the volume	vi
Photograph of participants in the Wallingford Workshop	vii

PART I - CORE MATERIAL

CHAPTER 1	SETTING THE SCENE	1
	Zbigniew W. Kundzewicz and Alice Robson	
CHAPTER 2	GUIDELINE TO ANALYSIS	11
	Alice Robson	
CHAPTER 3	HYDROLOGICAL DATA FOR CHANGE DETECTION	15
	Paul Pilon, Zbigniew W. Kundzewicz and David Parker	
CHAPTER 4	EXPLORATORY / VISUAL ANALYSIS	19
	Howard Grubb and Alice Robson	
CHAPTER 5	STATISTICAL METHODS FOR TESTING FOR CHANGE	49
	Alice Robson, Andras Bardossy, David Jones and Zbigniew W. Kundzewicz	

PART II— SPECIAL TOPICS

CHAPTER 6	DETECTING CHANGES IN EXTREMES	89
	Alice Robson and Francis Chiew	
CHAPTER 7	TESTS FOR CHANGES IN FLOW REGIMES	93
	Hege Hisdal	
CHAPTER 8	SPATIAL / REGIONAL TRENDS	101
	Harry F. Lins	
CHAPTER 9	TESTING FOR CHANGE IN VARIABILITY AND PERSISTENCE IN TIME SERIES	107
	Geoffrey G.S. Pegram	
CHAPTER 10	SEGMENTATION	113
	Pierre Hubert	
CHAPTER 11	CRITERIA FOR THE SELECTION OF STATIONS IN CLIMATE CHANGE DETECTION NETWORKS	121
	Paul Pilon	
CHAPTER 12	PHASE RANDOMISATION FOR CHANGE DETECTION IN HYDROLOGICAL DATA	133
	Maciej Radziejewski, Andras Bardossy and Zbigniew W. Kundzewicz	
CHAPTER 13	SIMULTANEOUS ESTIMATION OF TRENDS IN MEAN AND VARIANCE	141
	Witold G. Strupczewski	

APPENDIX I	GUIDELINES: SOFTWARE	147
	Felix Portman	
APPENDIX 2	HYDROSPECT — SOFTWARE FOR DETECTING CHANGES IN HYDROLOGICAL DATA	151
	Maciej Radziejewski and Zbigniew W. Kundzewicz	
GLOSSARY	153

LIST OF PARTICIPANTS IN THE WALLINGFORD WORKSHOP, 2-4 December 1998

Mr A. Bardossy
Institut für Wasserbau
Lehrstuhl für Wasserbau und Wasserwirtschaft
Universität Stuttgart
D-70550 Stuttgart
Germany
Tel: (49 711) 685 47 52/47 53
Fax: (49 711) 685 46 81
Email: bardossy@iws.uni-stuttgart.de

Mr M. Bonell
Division of Water Sciences
UNESCO and IHP Secretariat
1 Rue Miollis
75700 Paris Cedex 15
France
Tel: (33 1) 45 68 39 96
Fax: (33 1) 45 68 58 11
Email: m.bonell@unesco.org

Mr F. Chiew
Cooperative Research Centre for Catchment Hydrology
Department of Civil and Environmental Engineering
University of Melbourne
Parkville
Victoria 3052
Australia
Tel: (61 3) 934 466 44
Fax: (61 3) 934 462 15
Email: fchs@civag.unimelb.edu.au

Mr H. Grubb
Department of Statistics
Reading University
Whiteknights Road
PO Box 240
Reading RG6 6FN
United Kingdom
Tel: (44 1 18) 931 65 73
Fax: (44 1 18) 975 31 69
Email: h.j.grubb@reading.ac.uk

Ms H. Hisdal
Norwegian Water Resources and Energy Administration
PO Box 5091, Majorstua
0301 Oslo
Norway
Tel: (47 22) 95 91 33
Fax: (47 22) 95 92 16
Email: hhi@nve.no

Mr P. Hubert
Ecole des Mines de Paris
35 Rue St Honore
77305 Fontainebleau – Cedex
France
Tel: (33 1) 64 69 47 40
Fax: (33 1) 64 69 47 03
Email: hubert@clg.ensmp.fr

Mr D. Jones
Institute of Hydrology
Maclean Building
Crowmarsh Gifford
Wallingford
Oxon, OX10 8BB
United Kingdom
Tel: (44 1 491) 83 88 00
Fax: (44 1 491) 69 24 24
Email: daj@mail.nwl.ac.uk

Mr Z.W. Kundzewicz
Research Centre of Agricultural and Forest Environment
Polish Academy of Sciences
Bukowska 19
60-809 Poznań
Poland
Tel: (48 618) 47 56 01
Fax: (48 618) 47 36 68
Email: zkundze@man.poznan.pl

Mr J. Kuylenskierna
ERM Dynamo
Sturegatan 46
100 41 Stockholm
Sweden
Tel: (46 8) 667 88 68
Fax: (46 8) 662 22 30
Email: johan.kuylenskierna@dynamo.se

Mr H.F. Lins
US Geological Survey
436 National Centre
Reston, Va 22092
USA
Tel: (1 703) 648 57 12
Fax: (1 703) 648 50 70
Email: hlins@usgs.gov

Mr D. Parker
Meteorological Office
Hadley Centre for Climate Prediction and Research
London Road
Bracknell
Berks RG12 2SY
United Kingdom
Tel: (441 344) 85 66 49
Fax: (44 1 344) 85 48 98
Email: deparker@meto.gov.uk

Mr G. Pegram
Civil Engineering
University of Natal
Durban 4041
South Africa
Tel: (27 31) 260 30 57
Fax: (27 31) 260 14 11
Email: Pegram@eng.und.ac.za

Mr P. Pilon
Environment Canada
Atmospheric Environment Service
75 Farquhar Street
Guelph
Ontario N1H 3N4
Canada
Tel: (1 519) 823 42 02
Fax: (1 519) 826 20 83
Email: paul.pilon@ec.gc.ca

Mr F. Portmann
Global Runoff Data Centre (GRDC)
Bundesanstalt für Gewässerkunde/Federal Institute of Hydrology
Kaiserin-Augusta-Anlagen 15-17
D-56068 Koblenz
Germany
Tel: (49 261) 1306 52 18
Fax: (49 261) 1306 52 80
Email: portmann@bafg.de

Ms A. Robson
Institute of Hydrology
Maclean Building
Crowmarsh Gifford
Wallingford
Oxon OX10 8BB
United Kingdom
Tel: (44 1 491) 83 88 00
Fax: (44 1 491) 69 24 24
Email: Alice.J.Robson@unixa.nerc-wallingford.ac.uk
arj@unixa.nerc-wallingford.ac.uk

OTHER CONTRIBUTORS TO THE VOLUME:

Mr M. Radziejewski
Faculty of Mathematics and Computer Science
Adam Mickiewicz University
Poznań
Poland
E-mail: maciejr@amu.edu.pl

Mr W.G. Strupczewski
Institute of Geophysics
Polish Academy of Sciences
Ks. Janusza 64
01-452 Warsaw
Poland
Tel: (48 22) 69 15 853
Fax: (48 22) 69 15 915
Email: wgs@igf.edu.pl



Participants in the International Workshop on Detecting Changes in Hydrological Data held in the Institute of Hydrology, Wallingford, UK from 2 to 4 December 1998.

Front row, from left to right: Mr G. Pegram, Mr J. Kuylenstierna, Mr H. Lins, Ms H. Hisdal, Ms A. Robson, Mr D. Parker, Mr F. Chiew.

Middle row, from left to right: Mr P. Hubert, Mr D. Jones, Mr M. Bonell, Mr P. Pilon.

Back row, from left to right: Mr A. Bardossy, Mr H. Grubb, Mr Z.W. Kundzewicz, Mr F. Portmann..

CHAPTER 1

SETTING THE SCENE

Zbigniew W. Kundzewicz & Alice Robson

Structure of this report

This document is the result of an International Workshop on Detecting Changes in Hydrological Data that was held in Wallingford from 2 to 4 December 1998. The primary aim of this document is to serve as a handbook for practitioners and numerate scientists who need to undertake analyses of trend and other changes in hydrological data. The following chapters contain recommendations and general guidelines for the detection of change and should serve a broad audience. The document will also be useful to statisticians, but it is not aimed solely at them. Where appropriate, references are provided for those willing to undertake further, more detailed studies.

This report begins with core material and then moves on to outline more advanced and specialist topics. The core material is presented in Chapters 1-5. The remainder of this introductory chapter sets the scene for studies of change in hydrological series. It is followed by an overview chapter that is particularly aimed at those who are relative newcomers to the topic of trend detection. Its function is to assist the reader in solving her or his particular problem. It introduces some of the most fundamental concepts and suggests how to approach the study of change.

Data is the backbone of any attempt to detect trend in long time series and Chapter 3 discusses the types of data that are available and the issues relevant to selecting suitable series. It is followed by a chapter on methods of exploratory data analysis i.e. visual approaches to studying data. These are techniques that let the data speak for themselves and that encourage exploration and improved understanding of the data. Exploratory data analysis is a critical component of any statistical analysis of time series; here methods that are particularly appropriate for hydrological series are described.

Chapter 5 provides the main theoretical component of this document. It discusses different methods of testing for change e.g., parametric and distribution-free approaches. It provides guidance as to recommended testing procedures, to the underlying assumptions, and to interpretation of test results.

The remaining chapters of this guide cover a number of specialist and more advanced topics. These will be highly relevant to some studies, and not appropriate in other circumstances. In most cases, each chapter provides an outline of the topic and summarises possible approaches to analysis, giving references to more detailed studies. Topics covered include hydrological extremes, seasonality, regional and multivariate approaches, changes in persistence and variability, segmentation methods, design of a dedicated observation network for change detection, methods for looking at simultaneous changes in mean and variance, and phase randomisation methods for series with autocorrelation.

A brief discussion of available software packages is also offered. The specialist nature of the topics that are covered here means that there is no one user-friendly package available that allows use of all the methods that are discussed below.

Why undertake an analysis of change?

Detection of trends in long time series of hydrological data is of paramount scientific and practical significance. Water resources systems have been designed and operated based on the assumption of stationary hydrology. If this assumption is incorrect then existing procedures for designing levees, dams, reservoirs, etc. will have to be revised. Without revision there is a danger that systems are over or under designed and either do not serve their purpose adequately or are overly costly.

Studies of change are also of importance because of our need to understand the impact that man is having on the “natural” world. Urbanisation, deforestation, emissions of greenhouse gases,

changes in agricultural practice and dam construction are just a few examples of anthropogenic activities that may be altering important aspects of the hydrological cycle.

The principal water-related problems have been always related to having too much water (floods) or too little water (low flows or droughts). This means that studying changes in characteristics of hydrological extremes is of major importance.

Climate variability and change

A key question at present is to understand climate change impacts: is there evidence of climate change in hydrological series? Can we distinguish climate change from natural variation in the climate? Can climate change be distinguished from other progressive environmental changes such as urbanisation? In order to address the issue of climate change it is essential to understand the meaning of the term “climate change”, as opposed to “climate variability”. Since various bodies dealing with climate change use slightly different meaning of this notion, some exemplary definitions are listed in the Appendix to this Chapter. Loosely speaking “climate change” can be thought of as describing a long-term underlying shift in the climate, whilst “climate variability” encompasses the natural variation in our climate that would be seen even in the absence of any underlying long-term change.

In the case of river flows, the problem of detecting a climate change signature is very complex. Increasing concentrations of greenhouse gases in the atmosphere cause temperature rise, which, in turn, enhances potential and actual evapotranspiration. What goes up must come down, therefore some precipitation growth is expected. Runoff is the difference between rainfall and evaporation, both of which are increasing, so the net effect on runoff is not intuitively clear.

The strong natural variability of river flow makes it difficult to detect any underlying climate (greenhouse) signature. Flow fluctuations have been occurring naturally since the dawn of history, some of these changes being rather stronger than anything seen recently. A further problem is how to distinguish change linked to climate change from changes arising from land use change signature.

The scope of hydrological change studies

There are many different ways in which changes in hydrological series can take place. A change may occur abruptly (step change) or gradually (trend) or may take more complex forms. Changes can be seen in mean values, in variability (variance, extremes, persistence) or in the within-year distribution (e.g. changing seasonality and river flow regimes). Abrupt changes can be expected as a result of a sudden alteration within the catchment e.g. reservoir construction, installing water diversions, etc. They can also inadvertently arise from changes to gauging structures, or to rating curves (stage-to-flow relationships), or to observation methods. Gradual hydrological changes typically accompany gradual causative changes such as urbanization, deforestation, climate variability and change. Although climate change is often thought of in terms of progressive trend, it is also possible for it to result in a step-like change because of complex dependencies on non-linear dynamic processes that feature cumulative effects and thresholds.

There are a huge variety of hydrological data that it is possible to analyse for trend and step change. These may be collected at a range of temporal intervals: continuous, hourly, daily, monthly, annually, or sampled irregularly. Data records contain either instantaneous values (e.g., of flow, stage) or totals for a time interval (e.g., precipitation). Data may also pertain to different spatial scales, from point or experimental plot to large areas (including the Globe).

Studies of hydrological change are typically complicated by factors such as missing values, seasonal and other short-term fluctuations (climate variability) and by lack of homogeneity (e.g. due to changes in instruments and observation techniques). In some cases, there are further problems because of censored data and data series that are not sufficiently long.

Approaches to testing for change

There are many approaches that can be used to detect trends and other forms of non-stationarity in hydrological data. In deciding which approach to take it is necessary to be aware of which test procedures are valid (i.e. the data meets the required test assumptions) and which procedures are most useful (likely to correctly find change when it is present).

Parametric testing procedures are widely used in classical statistics. In parametric testing, it is necessary to assume an underlying distribution for the data (often the normal distribution), and to make assumptions that data observations are independent of one another. For many hydrological series, these assumptions are not appropriate. Firstly hydrological series rarely have a normal distribution. Secondly, there is often temporal dependence in hydrological series, particularly if the time series interval is short (e.g. today's flow tells us quite a bit about what tomorrow's flow is likely to be). If parametric techniques are to be used, it may be necessary to (a) transform data so that its distribution is nearly normal and (b) restrict analyses to annual series, for which independence assumptions are acceptable, rather than using the more detailed monthly, daily or hourly flow series.

In non-parametric and distribution-free methods, fewer assumptions about the data need to be made. With such methods it is not necessary to assume a distribution. However, many of these methods still rely on assumptions of independence. More advanced approaches must therefore be used for daily or hourly series. A very useful class of non-parametric tests are permutation tests. They are based on changing the order (shuffling) of data points, calculating statistics, and comparing these with the observed test statistics.

Even within the basic categories above it is necessary to choose tests that are appropriate for the situation. Some tests are very good at detecting a very specific type of change, other tests may be good at picking up any one of a broad range of possible changes. Since one does not know the pattern of variability beforehand, using a number of tests is sensible.

The current picture

A number of long time series of observations of hydrological variables from across the world have been studied to determine whether there is evidence of climate change in these series. Most of these series are for river flows and lake levels, and several analyses have been performed. At present, there is no conclusive evidence of a climate change signature. Findings reported in specific works cannot be generalized and to a large extent appear to be localised chance occurrences. A similar picture holds for analyses of flood and drought series. It must be remembered that although no strong evidence of climate change has been found in hydrological series, this is not proof against climate change. It seems that longer data series will be required before it becomes possible to identify any climate change effects and it remains important to continue collecting data and to undertake further studies.

The future

This guide is focused on presenting recommended methods based on current day knowledge. However, there are clearly areas in which further developments are needed. These can be broadly broken down into the need for further data, and a need for developments in statistical methodology.

A key requirement in terms of data is to obtain and extend long high-quality worldwide data series. National Hydrological and Hydrometeorological Services in many countries are facing severe financial stringencies and there is a need to ensure that valuable stations, often in remote non-populated areas are not closed down. These stations may contribute in an important way to studies on detection of changes in flow characteristics. Chapter 11 discusses related aspects in more detail and also considers the problems of obtaining data from pristine catchments, where rivers have not been regulated nor major land use changes have occurred.

A viable extension of studies of observed long time series of instrumental hydrological data is to obtain and analyse proxy information, such as tree rings, coral increments, lake deposits, etc. Studying isotopic composition of ice in deep boreholes allowed the scientists to reconstruct the

conditions of remote past with an amazingly accurate resolution. As the processes leading to the build-up of proxy archives are natural, homogeneity of these data has not been influenced by changes in instrumental data collection (in contrast to many hydrological records).

There remains ample scope for improvements in analytical methods. Some possibilities include

- Improved guidance as to which tests are best able to detect change under a realistic set of assumptions.
- Development of multivariate methodologies that look at flows alongside other hydrological variables
- Development and application of improved regional methodologies

WCP-Water context

The present document and the International Workshop on Detecting Changes in Hydrological Data, held in Wallingford from 2 to 4 December 1998 are contributions to a project entitled “Analyzing Long Time Series of Hydrological Data and Indices with Respect to Climate Variability and Change” of the World Climate Programme - Water (WCP-Water Project A.2). The WCP-Water is an international endeavour aimed at studies of links between climate and water. It is jointly implemented by a number of international agencies (notably the World Meteorological Organization, WMO and the United Nations Educational, Scientific and Cultural Organization, UNESCO), and national institutions. WCP-Water Project A.2 capitalizes on research financed at the national level and the international context makes it possible to encourage national efforts, to compile information on national research in the area and to blend the results produced there.

In the former stages of the project there was a worldwide data collection initiative during which the Secretary-General of the WMO contacted National Hydrological and Hydrometeorological Services of WMO Member countries, requesting them to forward long time series of river flow data from non-regulated rivers. A software package TIMESER was developed in WMO Secretariat and used worldwide as a tool for analysis of properties of a long time series of hydrological records. The above data base and the software served for analysis undertaken within the Project A.2 of some 200 long time series of hydrological data sets. It was found that, for a substantial number of data series tested, there was some evidence that the mean and/or variance of flows could be changing. A number of project meetings have been held and a number of reports produced. In the project, it was agreed that the methodology of tests needs further work and that a comparison of available tests would be useful. The current report attempts to make recommendations on best practice given current day knowledge.

It is hoped that the Project A.2 will continue to stimulate further international activities worldwide in the strong and important growth area of trend detection in hydrological data.

References

Cavadias, G. S., 1992. A survey of current approaches to modelling of hydrological time-series with respect to climate variability and change. WCASP-23, WMOIFD-No. 534, Geneva, Switzerland.

IPCC (Intergovernmental Panel on Climate Change), 1996. Climate Change 1995: The Science of Climate Change. Contribution of WGI to the Second Assessment Report of the

Intergovernmental Panel on Climate Change (ed. by Houghton, J. T., Meira Filho, L. G., Cailander, B. A., Harris, N., Kattenberg, A. & Maskell, K.), Cambridge University Press, Cambridge, U.K., 572 pp.

WMO, 1988. Analyzing long time series of hydrological data with respect to climate variability. Project description, WCAP-3, WMOTFD-No. 224, Geneva, Switzerland.

PART I

CORE MATERIAL

Appendix

DEFINITIONS OF CLIMATE VARIABILITY AND CHANGE

Definitions of climate change and climate variability given by World Climate Research Programme, WCRP; Intergovernmental Panel on Climate Change, IPCC and Framework Convention on Climate Change, FCCC) are given below.

(1) Climate change

WRCP usage (WMO, 1988, p. 3):

“Climate change defines the difference between long-term mean values of a climate parameter or statistic, where the mean is taken over a specified interval of time, usually a number of decades”.

IPCC usage (IPCC, 1996, P. 48):

“Climate change as referred to in the observational record of climate occurs because of internal changes within the climate system or the interactions between its components, or because of changes in external forcing either for natural reasons or because of human activities. It is generally not possible to make clear attribution between these causes. Projections of future climate change reported by IPCC generally consider only the influence on climate of anthropogenic increases in greenhouse gases and other human-related factors.”

FCCC usage (IPCC, 1996, p. 48):

“A change of climate which is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and which is in addition to natural climate variability observed over comparable time periods.”

(2) Climate variability

WMO (1988, p. 3):

“the extremes and differences of monthly, seasonal and annual values from the climatically expected value (temporal means). The differences are usually termed anomalies.”

WMO (1988, p. 4):

“climate variability can be regarded as the variability inherent in the stationary stochastic process approximating the climate on a scale of a few decades, while climate change can be regarded as the differences between the stationary processes representing climate in successive periods of a few decades”.

CHAPTER 2

ANALYSIS GUIDELINES

Alice Robson

2.1 Aim

The aim of this chapter is to give a picture of how the various stages of an analysis fit together and provide an overview of the core methods that are recommended in this report. It also serves as a gentle introduction to some of the most important concepts. This chapter attempts to remind scientists that to analyse a dataset requires much more than just a blanket application of statistical tests.

2.2 The process of analysis

The main stages of an analysis are

- Obtaining and preparing a suitable dataset
- Exploratory analysis of the data
- Application of statistical tests
- Interpretation of the results

The process of analysis is one of iteration, development and refinement. Figure 2.1 summarises how an analysis typically proceeds and shows its cyclic nature. At any stage, the analyst needs to be open minded and to recognize when the results have further implications that are themselves worthy of exploration. Thus, for example, a series of flow data on a river may be found to show trend. This then raises the question of whether the trend is due to urbanisation or to climate change. To address these questions it may be helpful to obtain and examine rainfall data and/or data from other nearby non-urbanized catchments.

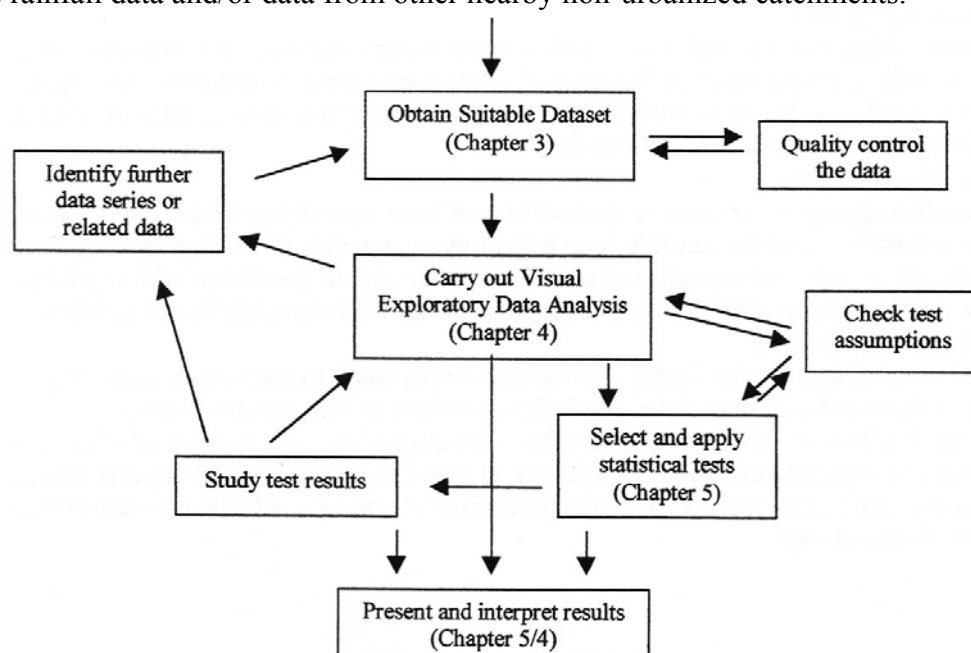


Fig. 2.1. Flow chart illustrating the main stages in a statistical analysis of change.

2.3 Obtaining a suitable dataset

In many studies, a specific dataset is the focus of study and the question is whether the data shows any evidence of trend or other change. In other cases, one has a particular question in mind and is seeking the right data to best answer the question. Even when there is a specific dataset of interest, it is still important to consider other available sources of data. For example, when investigating change in flow series it is often helpful to obtain rainfall data too. Chapter 3 gives further details on the many types of data that are of relevance to detecting change in hydrological data and of some possible sources of such data.

Obtaining a suitable dataset sounds straightforward but, in practice, it can require care and skill. There are many important aspects that may need to be considered when obtaining and preparing data. These include

- *Quality of data*
Data should be quality controlled before commencing an analysis of change. The analyst should, however, never assume that the data is set in stone and should always be on the lookout for further data problems. A very frequent problem in long hydrological series is that methods of measurement have often changed over time: it is advisable to investigate possible changes in data collection methods
- *Length of record*
Data series should be as long as possible. Short data series can be strongly affected by climate variability which can give misleading results (Section 5.7.2). For investigation of climate change, a minimum of 50 years of record is suggested - even this may be not be sufficient.
- *Missing values and gaps*
Missing values and gaps in a data series make analysis harder and raise questions of data quality (see above). It is important to consider whether gaps are truly random, or whether they are perhaps associated with major flooding making the remaining data unrepresentative. Many of the methods recommended in this report can still be applied to incomplete data series provided that the gaps are not too extensive and that they occur randomly.
- *Frequency of data*
Hourly, daily, monthly and annual data series are commonplace. In a few cases (e.g. flood series), the data may be irregular. Very frequent data contains more information but can also be harder to analyse both computationally and because more restrictive assumptions must be made (see also Section 2.5).
- *Use of summary measures*
It is often appropriate to analyse time series that have been derived from the raw data. For example, it may be sensible to calculate mean monthly flows from a daily flow series, or to derive an annual maximum flow series. Again the choice will depend on the question that is being addressed and the depth to which the analysis can be taken.
- *Use of transformation*
Hydrological data is often highly skewed and non-normal. In such cases, data analysis can sometimes be assisted if the data is first transformed (see also Section 4).

Further detail on the above issues, together with discussion on a number of other data related aspects is presented in Chapter 3. Chapter 11 also discusses important aspects related to selecting data series, and applies this to the problem of choosing a suitable network of data for study of climate change.

2.4 Exploratory data analysis

Exploratory data analysis (EDA) is a very powerful graphical technique that is a key component of any data analysis. Full details of exploratory data analysis, including a range of examples and illustrations, are given in Chapter 4. Exploratory data analysis is itself an iterative process and it should be used at more than one stage of an analysis. Its first use is to examine the raw data. This may identify further interesting aspects of the data, such as seasonality, which in turn invite further investigation. Exploratory data analysis also has an important role in helping to check out test assumptions. For example, having fitted a trend to the data, exploratory data analysis can be used to examine the residuals to check for independence (Section 4). In some cases, this may mean that the model needs to be altered then revisited using EDA. Finally EDA can provide a very valuable means of presenting both the data and the results in a way that maximises understanding and impact.

2.5 Statistical analysis

Chapter 5 contains a detailed description of recommended approaches to analysing change in hydrological series. Here, a few of the most fundamental concepts are described by way of introduction. The aim is to give a general background understanding of the main ideas, but not to provide the detail, which is left to Chapter 5.

A key recommendation made in this report is that resampling methods be used for testing hydrological data. *Resampling methods* are methods that use the data to help determine significance levels. They are very useful in the context of hydrological data because they do not require distributional assumptions to be made. There are two main types of resampling: permutation testing and bootstrapping methods.

A *permutation test* works by shuffling the data very many times. Consider a time series of data with a possible trend. One measure of the trend is the regression gradient: an example of a possible *test statistic*. Suppose first that there is no underlying trend in the data. If that is true, then it should not matter very much if data is reordered - the regression gradient should not change very much. Each time the data is shuffled as part of the permutation test, the selected test statistic (in this case the regression gradient) is recalculated. At the end of all the shuffling, we have generated a distribution of possible values of the test statistic under permutation, the *permutation distribution*. The permutation distribution usually depends on the data and must be recalculated for each data set. If there is no trend, then we would expect that the observed test statistic (regression gradient) for the original data is not very different to any of the generated test statistic values, i.e. it is somewhere in the middle of the permutation distribution. So to test for trend, the observed test statistic (regression gradient) is compared with the permutation distribution. If the gradient is larger (or smaller) than almost all the values in the permutation distribution, we conclude that a trend is present. Conversely, if the original gradient is somewhere in the middle of the permutation distribution, we conclude that there is no evidence of trend.

There are two main steps in applying a permutation test.

- *Choosing a suitable test statistic.*

Permutation testing is a very flexible approach and many test statistics can be used within the permutation framework. Different statistics tend to pick out different features of the data, so it can be helpful to use more than one, e.g. a linear regression gradient (or equivalent statistic) to look at trend, and a statistic such as Buishands' Q to look for step change.

- *Calculating the permutation distribution for the test statistic*

The permutation distribution is derived using the data (as described above). It must be recalculated for each data set and each test statistic. Once known it can be used to determine significance levels.

Bootstrapping approaches are similar to permutation techniques. The main difference is that instead of reordering the data, the new data series are generated by sampling with replacement. For example, for a series of 50 values, a bootstrap sample would take 50 values at random from the original series: the resulting series might perhaps include 3 lots of the original first value, but no instances of the last value.

Although use of resampling approaches avoids the need to make distributional assumptions (e.g. that the data is normally distributed), it still requires some assumptions to hold. An important remaining assumption is that the data are independent. Frequently measured hydrological data (daily or hourly data in particular) are typically not independent: they show dependency from one value to the next (if flow is high today it is likely to be high tomorrow). This type of dependency is referred to as serial dependency, temporal dependency or autocorrelation. If it is present, but conveniently ignored, then it can result in inaccurate significance levels. A very simple technique to avoid problems of dependence is to average or aggregate (e.g. monthly data may often be treated as being independent whereas daily data cannot). However, data with serial dependency can instead be tested using resampling methods. In this case, the data is permuted or bootstrapped in blocks (e.g. all values within a year are kept together). With this approach the dependency structure within each block is built into the test and independence assumptions are thus no longer violated. Note that there are also alternative ways of tackling this problem (Section 5.5.2).

2.6 Interpretation

The final interpretation of the results brings together the information gained out of all stages of the analysis. Thus it combines information about how the data was obtained, historical information about the catchment or region, graphical information gained from exploratory data analysis and the statistical test results. It is often appropriate to present results and the interpretation using both graphs and tables.

Great care is often needed when interpreting results. For example, with short data series there may be a statistically significant trend, but in combination with other information the most sensible conclusion could be that the trend is a chance occurrence that would probably not have been seen if a longer record had been available.

Chapter 5 provides more details on interpretation of test results.

2.7 Taking things further

The above summary outlines only the most fundamental ideas used in an analysis of hydrological change. These topics are covered in much greater detail in Chapters 3-5.

In many situations it may be appropriate to apply more specialised methods. The choice of which method to use will depend on the data and the particular problem. For example, it might be useful to consider multivariate techniques, or to use approaches tailored to regional data or to investigate seasonal aspects. Chapters 6-13 provide introductory information on many important special topics that may be valuably used alongside the basic techniques described above.

CHAPTER 3

HYDROLOGICAL DATA FOR CHANGE DETECTION

Paul Pilon, Zbigniew W. Kundzewicz & David Parker

Hydrology is the study of water, dealing with the entirety of the hydrological cycle that is including liquid, solid and gaseous water.

Hydrological data serve many purposes, including the documentation, detection and quantification of climate variability and change, and improving our understanding of the climate system and the links between the climate system and other systems. Hydrological data are extensively used in various endeavors of process modeling, assessment of impacts and development of response strategies.

The Global Climate Observing System (GCOS) is an international framework for development of a comprehensive long-term global observing system aimed at improving our understanding of the climatic system and its interactions with other systems. It provided (GCOS, 1997, p. 14) an extensive but not necessarily exhaustive list of the key hydrological variables whose observation is required for climate purposes. The list embraces such variables as atmospheric water content near the surface (relative humidity), biogeochemical transport from land to oceans, evapotranspiration, discharge (runoff), groundwater storage fluxes, precipitation, sediment load at large river mouths, soil moisture and surface water storage. Moreover, a number of variables related to the cryosphere were listed. Some of these include snow cover, its depth and water equivalent, lake and river freeze-up and break-up dates, and characteristics of ice caps and glaciers. Each variable of concern was discussed in GCOS (1997) in the context of identifying users, rationale of observations, frequency and spatial resolution, accuracy, measurement methods, present status and required actions.

Since it is neither possible nor desirable to measure everything, as this would constitute continuous measurement of many hundreds of variables, everywhere, GCOS (1997) suggested a five-tier hierarchical sampling scheme embracing a wide range of cases. On one extreme there are a few locations with frequent observations of a large number of variables, while on the other extreme there are many locations where a few variables are measured infrequently. A plan for terrestrial climate-related observations provides a rationale for the structure and implementation of the initial global observing system (OCOS, 1997).

Different hydrological indices might be of interest due to their ability to reflect changes in climatology and for improving our understanding of the links between climate and hydrology. Examples of such variables read:

- Long historical records of ice cores, sediment cores, streamflow, groundwater levels, etc.
- Easy to monitor variables such as percentage of land area covered by snow, timing of river freeze-up, break-up date of ice, number of ice covered days, number of ice free days, etc.
- Extreme events (frequency and severity of floods and droughts, which need not necessarily be a consequence of extreme climatic conditions) and their characteristics such as the number of incidences of independent flood events within a hydrological year or season and the cumulative deficit below a prescribed threshold such as the n annual discharge for a hydrological year or season.
- Seasonal mean flow (depending on season, more or less influenced by both temperature and precipitation)
- Monthly mean flow (depending on season, more or less influenced by both temperature and precipitation)
- Spring flood volume and the duration in days of the spring flood event.

Timing of seasonal events is also seen as important index of change. Such aspects include:

- start of snowmelt season (mainly influenced by spring time temperature)
- snowmelt flood-peak (mainly influenced by spring time temperature)
- maximum flow (influenced by temperature and precipitation)
- timing of freeze-up of rivers/lakes (mainly influenced by autumn temperature, but also wind)
- timing of ice break-up for rivers/lakes (mainly influenced by spring temperature, but also affected by e.g. snow-cover and wind) having an annual resolution minimum flow.

The selection of sites and data for monitoring and analyzing climatic change and variability depends on the specific objectives of the study (cf. Chapter 11). For example, if the objective is to search for a greenhouse component in a hydrological process subject to substantial natural variability, it is essential to eliminate other influences. Therefore, data unaffected by local human influences would be selected. Otherwise it would be necessary to reconstruct a natural flow series - the difficult process of flow naturalization. By contrast, other studies will use data in basins where there have been known modifications to the hydrological processes and seek to assess their effects.

The selection of data for studying climatic variability and change requires an appreciation of the processes affecting the hydrological cycle and of the causal relationships between variables and processes. The temporal and spatial intervals between observations are important aspects of the record and must be chosen to match the processes. For example, if it is desired to monitor the maximum streamflow, monthly averages would be too coarse. It is also easier to analyze observations made at regular time intervals (e. g. daily, weekly) while some data such as water quality are typically collected at irregular intervals.

The user needs to gain an appreciation of the characteristics of the data that are to be used in analyses. This embraces the information about how the measurements were made and how the data were generated. These could constitute what is referred to as metadata. Information is also required about conditions in the basin and how they may have been modified, both spatially and temporally.

The data may have been obtained from a sequence of different instruments of different accuracy, entailing the possibility of biases in distributions and mean values. Other biases may have arisen from modifications to the catchment hydrology (e.g. change of land use, urbanization, deforestation, river regulation). There may also be gaps in the record. Gaps in the data may have been filled, and the assumptions used in filling the data will invariably affect the results (e.g. infill techniques by use of historical values under assumption of stationarity, or interpolation based on values of neighbouring points in time and space). The data may also have been modified and quality controlled (e.g. via outlier detection). Knowledge of the process for controlling the quality of the data is essential in the context of the “garbage in garbage out” (GIGO) syndrome. Knowledge of the accuracy of the data and its inherent suitability for analytical purposes is an integral aspect of the process.

Some variables are observed, while others are derived (data products) by the use of some models of varying degrees of complexity. For example, the data may have been generated using a hydrological model to assimilate available observations. Often, hydrological variables are measured point-wise, yet in order to quantify hydrological processes these variables need to be extended to cover areas (catchment or rectangular grid) and a difficult issue of transforming point values into spatial aggregates comes about. Some other variables of essential importance, such as area! evapotranspiration, are not directly measured but are derived from observations of other variables.

Some hydrological variables are measured in a time point (e.g., instantaneous value of river stage), others are measured in a time aggregated form (e.g., daily precipitation). If the data are provided in derived form the derivation, e.g., spatial averaging, may have filtered the data and may have removed or modified some features of interest.

The issue of representativeness of data for large areas also needs to be considered. At present, there is a certain bias due to the existence of many rain gauges in the cities and in mountain valleys worldwide. This poses a difficult problem of extending point values to give spatially representative coverage, particularly in under-represented geographical regimes. There is also the apparent problem of the local representativeness of a site's data. An important example is the climate station in the urban environment to represent ambient conditions.

In order to assess climate variability and change and their impacts within the hydrosphere, it is necessary to analyze long time series of observational data. In many cases, long time series of instrumental data are not available. This may make it necessary to extend the time coverage by blending data from different sources of different quality, incorporating pre-instrumental data (e.g. qualitative records in chronicles and archives, proxy data and palaeodata). Fortunately, data are increasingly available in electronic form. It is well recognized, however, that additional efforts are required to digitize remaining manuscript data, although some activity is ongoing in this area.

In recent years, data of hydrological value have been obtained using remote sensing from satellites. As these records rarely exceed 20 years in coverage, they are of limited value for detection of long-term changes. In addition, satellite data sets are often very voluminous (gigabytes) and therefore are more difficult to process. But, a major advantage of such data is that they are usually geographically expansive and can therefore be used, with the aid of appropriate techniques, to interpolate between in-situ measurements. Remotely sensed data can also be used to help understand process hydrology. Note that satellite data may possess their own biases owing to changes of instruments, calibrations, orbits, and atmospheric transmission.

Global archives of both in-situ and satellite hydrological data and metadata are being developed. Such archives, which are also referred to as data centres, include the already well established Global Runoff Data Centre in Koblenz, Germany, and the Global Precipitation Climatology Centre in Offenbach, Germany. In addition, many data are held in national archives or in research institutions; their inclusion in the global archives is a desirable objective.

The Global Runoff Data Centre (GRDC), set up in the Federal Institute of Hydrology in Koblenz, Germany in 1988, is an implementation of a World Climate Programme - Water (WCP-Water Project A.5 - Collection of Global Runoff Data). It provides a general service for the collection and storage of internationally available sets of daily and monthly river flow data at the global scale and the generation of data products. The data bank comprises data from over 3500 stations, from nearly 150 countries.

Discharge data are collected under the following criteria (WCASP, 1997):

- large rivers with average discharge greater than 100 m
- basins with catchment areas greater than 1 000 000 km
- basins with more than 1 000 000 inhabitants and basins of high socio-economic importance
- basins with internal drainage
- long-time series of runoff(WCP-Water Project A.2)
- undisturbed areas up to 5000 km
- runoff into the oceans (WCP-Water Project A.8, GEWEX, GEMS/Water, GCOS, GTOS).

Aside from these criteria, discharge data are collected on a project basin, e.g. for regional hydrological analysis such as the ACSYS project.

Requests for data, and data products, can be made in writing to the GRDC. The charges requested for data cover the costs of handling, diskettes, packing and postage. The charges could be waived if the requesting body was a contributor of data to GRDC.

The Global Precipitation Climatology Centre (GPCC) set up in the German Weather Service in Offenbach, Germany in 1988 is a component of the Global Precipitation Climatology Project (GPCP) integrated in the Global Energy and Water Cycle Experiment (GEWEX). It covers the functions of collection, quality control, correction and gridding of precipitation data at the global scale, measured by raingauges. The GPCC data cover over 40 000 hydrometeorological stations from more than 130 countries. The Centre has contributed to the WCP-Water Project B.6 — Precipitation of Monthly Global Gridded Precipitation Data Sets.

The spatial coverage of data assembled in data centres is still far from being satisfactory. The reasons being that in some countries, data collection programmes are either weak or non-existent. Even in other countries where data are being collected, there is a reluctance to provide large sets of hydrological data to others. Under such circumstances, the WMO's call for free and unrestricted exchange of hydrological data and products deserves considerable attention. Systematic and comprehensive global observations are much needed and should be made available to all nations. Hydrological networks worldwide are not adequately funded and, at the global scale, they are shrinking rather than expanding. In many cases, this leads to breaking of the continuity of hydrological records, even when there exist valuable long-time-series of hydrological data for a given location.

Chapter 11 also provides more detailed information on the appropriateness of data and criteria for the possible selection of sites for analysis.

References

GCOS, 1997. GCOS/GTOS Plan for Terrestrial Climate-Related Observations, Version 2.0, Global Climate Observing System. GCOS-32, WMO TD - No. 796, UNEP / DEIA / TB.. 97-7, Geneva, Switzerland (130 pp.).

WCASP (World Climate Applications and Services Programme), 1997. Final Report of the Seventh Planning Meeting on World Climate Programme — Water (Koblenz, Germany, 13-16 May 1997). WCASP — 45, WMO TD — No. 854, Geneva, Switzerland.

CHAPTER 4

EXPLORATORY / VISUAL ANALYSIS

Howard Grubb & Alice Robson

4.1 Introduction

4.1.1 What is exploratory data analysis?

Exploratory data analysis (EDA) involves using graphs to explore, understand and present data and is an essential component of any statistical analysis.

Exploratory data analysis is an iterative process. At each stage, graphs are plotted and then refined so that the important features of the data can be seen clearly. Often patterns or features emerge that need further exploration. These might include seasonal variation, correlation or a problem with some data values. Because it is an exploration of the data, no two analyses will be the same.

4.1.2 When is exploratory data analysis needed?

Exploratory data analysis is needed whenever data is being examined, or a statistical analysis is undertaken. It is an underused technique. A study of non-stationarity that does not include a thorough EDA of both the data and the results is not complete.

Exploratory data analysis can and should be used at more than one stage of an analysis. It is particularly important to use EDA before statistical tests are applied: without a proper understanding of the data, test results can be meaningless. EDA is also invaluable when it comes to understanding, interpreting and presenting the results of a statistical analysis, e.g. to examine residuals, trend gradients and significance levels.

4.1.3 Why use exploratory data analysis?

Exploratory data analysis allows a much greater appreciation of the features in data than tables of summary statistics and statistical significance levels. This is because the human brain and visual system is very powerful at identifying and interpreting patterns. It is often able to see important features, structures or anomalies in a data series that would be very difficult to detect in any other way. Just looking at the data can change initial preconceptions, can alter the questions that it is sensible to ask, and can uncover important aspects that would never otherwise have been found.

A well-conducted EDA is such a powerful tool that it can sometimes eliminate the need for a formal statistical analysis. Alongside EDA, statistical tests become a way of confirming whether an observed pattern is significant, rather than a means of searching through data.

Some examples of aspects of the data that EDA is likely to uncover include:

- temporal patterns (e.g. trend or step-change)
- seasonal variation
- regional and spatial patterns
- data problems (outliers, gaps in the record etc.)
- correlations (between variables or sites).

If identified, many of the above can be further explored using EDA. EDA can also be used to examine issues such as

- independence and autocorrelation
- statistical distribution of data values

- details of the seasonal structure.

EDA is often useful in identifying data quality problems. However, it is not a substitute for proper quality control of data.

4.1.4 Who should do exploratory data analysis?

Exploratory data analysis can be used by anyone who is trying to understand data. It is not difficult and it does not require great expertise. It does require care and thought and a willingness to probe further. A good EDA requires sensible use of properly produced graphs (see Appendix A. 1 on style issues). The information gap between a good graph and a poor graph should not be underestimated.

There are currently many software packages that implement various tools for exploratory data analysis (see Appendix A.2 for relevant points about software). These range from spreadsheet packages that are widely accessible and allow a good start to be made, particularly with dynamic plots linked to the data, through to more flexible and powerful packages for which a level of programming ability is needed (see Appendix A.2 on software). A number of the graphs shown in the following sections were produced using spreadsheet graphics.

4.1.5 How to go about an exploratory data analysis

Exploratory data analysis involves:

- plotting graphs
- studying the graphs
- re-plotting graphs to improve the display of important features
- identifying further graphs that are needed
- iterating through the above.

In the following pages a toolbox of graphical approaches is presented. The list is not intended to be exhaustive, but specifically selects graphs that have been found to be of particular use for examining changes in hydrological time series. Not all graphs will be useful for all data sets and some situations may demand novel combinations or modifications of the basic graphs as well as more specialised graphs for specific applications. Deciding which graphs to look at is a matter of judgement and experimentation.

4.1.6 Chapter Overview

The remainder of this chapter looks in more detail at a variety of graphs. Section 4.2 presents some of the most fundamental graphical approaches - ones that are likely to be used most frequently in EDA for hydrological time series. Subsequent sections look at these and other methods in more detail and present graphs that address particular aspects of the data (e.g. seasonality). These sections also give some guidance on issues to be aware of when producing a graph. Two appendices summarise some issues on graphing style and on currently available software.

4.2 Fundamental components of an EDA

This section introduces some of the most useful techniques for looking at hydrological data. These techniques allow the user to rapidly view and assess the available data. Here, as in the rest of this chapter, it will often be the case that only some of the approaches are applicable for a specific data set, and graphs must be selected according to need.

The techniques can be applied to both the raw data and, at a later stage in the analysis, to derived quantities such as summary statistics, residuals and test results. Use of these graphs should allow identification of the most important features of the data (Section 4.3). In

particular, they enable visual assessment of any trend or step-change — and can indicate how great trend is relative to overall variation.

4.2.1 The time series plot

Most hydrological data is in the form of time series — observations of a variable recorded sequentially through time. The most fundamental plot for examining these is a time series plot of the data values against time, since the ordering in time is a key characteristic of the data, particularly when interest is in changes. Although the concept is simple, it can require some skill to produce a time series plot that best displays the features of the data (Fig. 4.1). For example, it may be necessary to plot the data on more than one scale or to transform the data. If the data series is very long, the display may be improved by spreading the data series over several plots, or by plotting summary statistics. More details are given in Section 4.3.

Interpretation of the time series plot is often aided by adding a smoothing curve (and sometimes a regression line) to follow the general trend in the data. Care must be taken to ensure that the level of smoothing is suitable for the data (see Section 4.3).

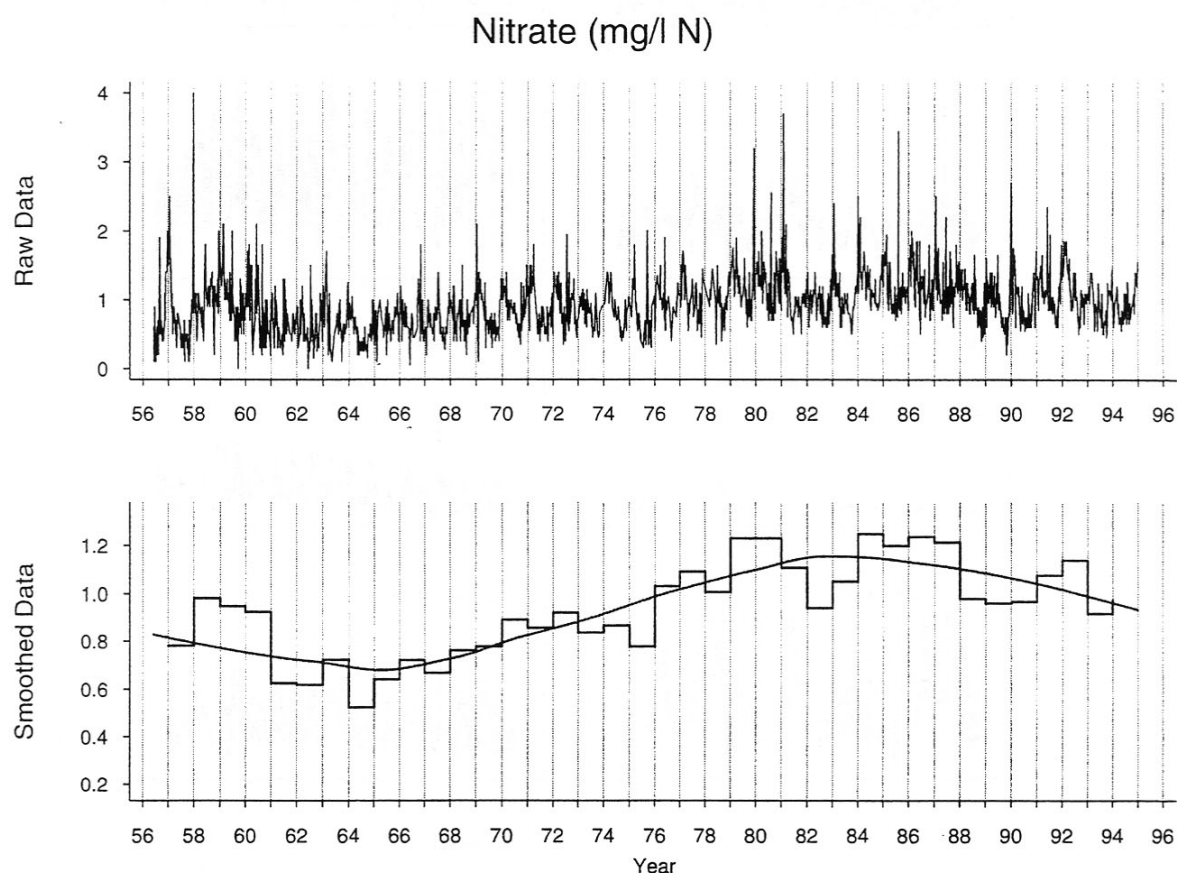


Fig. 4.1. Nitrate concentrations (mg/i) for the river Tweed, Southern Scotland. The upper graph shows a 40 year time series of concentrations recorded every month, from which it can be seen that there is some seasonal structure, plus a number of outliers. The lower graph shows a locally weighted smoothing curve and the annual averages. The lines are plotted separately from the data in order that the long-term variation is emphasized. The graphs suggest that average nitrate concentrations peaked in the early 1980's.

4.2.2 Multiple time series plots

When data from several sites or variables are available, it can be informative to examine the series together, e.g. presenting data for several sites within a region on a single page (Fig. 4.2). Patterns become more apparent when many time series are shown beside one another, and it can be easier to detect data problems and to identify whether behaviour is similar between sites. Note that it is clearer to plot multiple series on separate graphs than on a single one, which would quickly become cluttered for more than a few series.

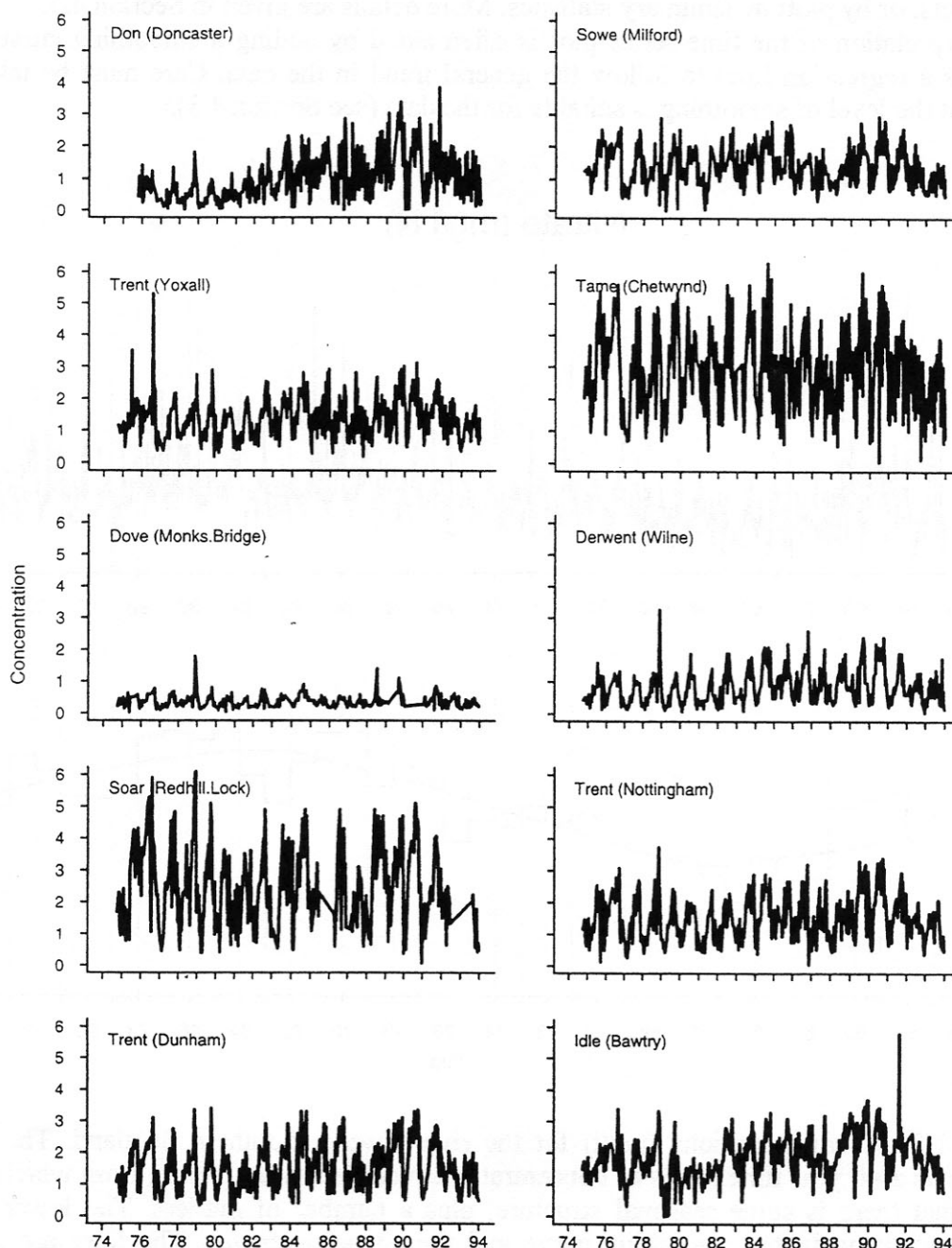


Fig. 4.2. Orthophosphate concentrations (mg/i) for 10 rivers within the Humber catchment area, England. The data sampling interval varies from weekly to monthly intervals and changes over time and from site to site. Plotting on a common scale clearly illustrates the very different concentrations present in these rivers. A seasonal component to the variation can be seen and this might merit further investigation. It would appear that concentrations

have risen most on the Don, a highly industrialised river system. Most of the time series show decreased orthophosphate concentrations since 1990, possibly related to reductions in phosphates used in domestic washing powders (a major source of orthophosphate). Some longer-than-average periods without sampling can be seen on the Soar. For regular data, it would be preferable to show a break in the time series line where there is a gap in the record.

For multiple time series plots it is desirable to use a common scale for all time axes — this presents the data in the clearest way and makes it possible to rapidly assess the similarities and differences between sites. If possible it is also best to use a common scale on the y-axes (Fig. 4.2). It may however be necessary either to allow the y-axes scales to vary, or to use a method of standardisation, in order that the variations in the series are visible for all sites. As with single time series plots, a common enhancement is to add smoothing curves and regression lines to multiple time series plots (Fig. 4.3).

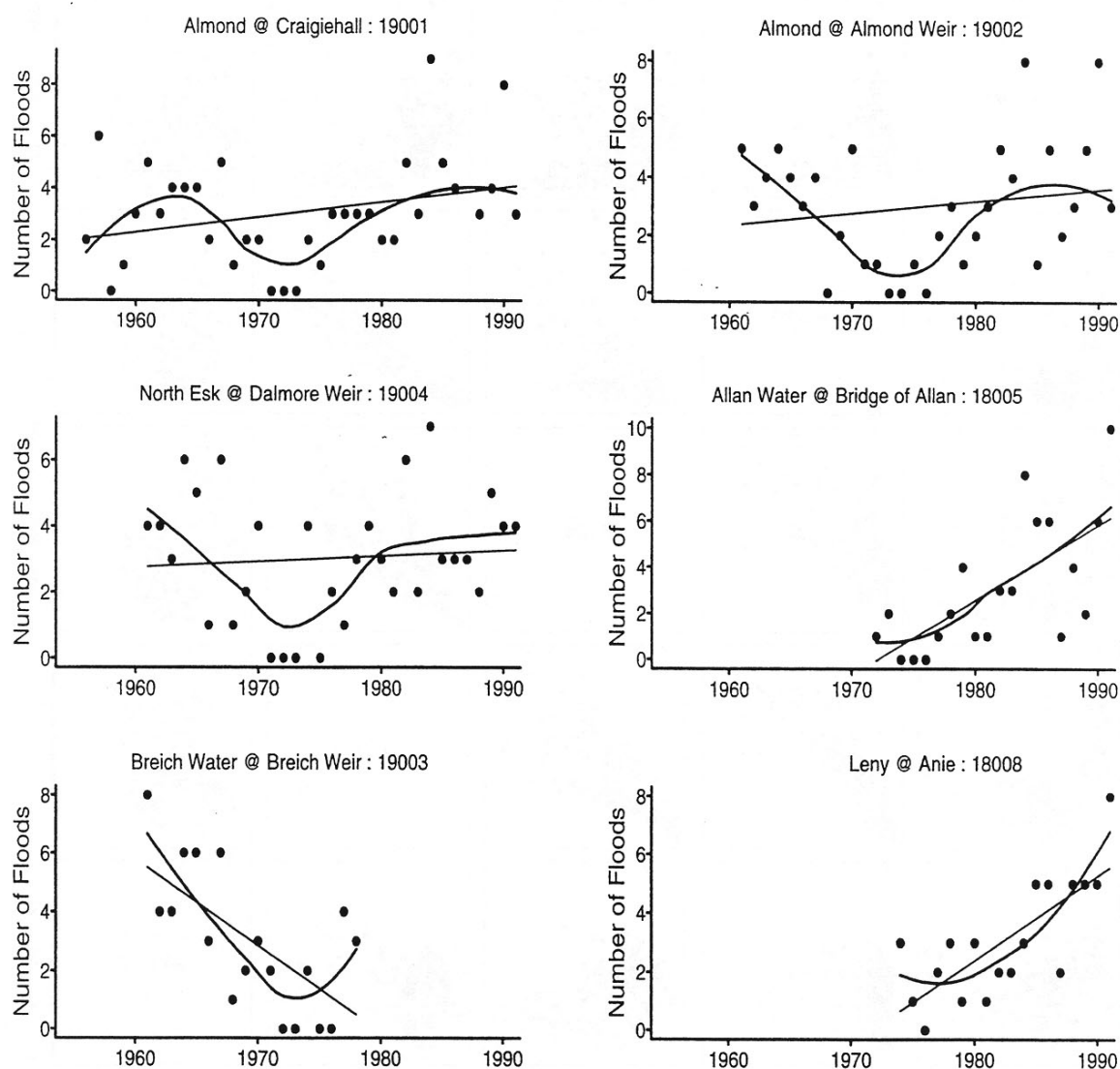


Fig. 4.3. Number of floods per year for 6 rivers in Southern Scotland. A locally-weighted smoothing curve and a regression line have been fitted to the data. A common time scale is used in all graphs. The graphs suggest a similar pattern of behaviour at all sites, i.e. more flooding in the early 1960's and the late 1980's, less flooding in the mid 1970's. For the three longest records this results in no overall trend, but for the shorter records, a significant trend

can be seen (upwards for the earlier Breich Water record, downwards for the later Allan Water and Leny). It seems reasonable to conclude that trends at the short record sites would probably not have been seen if the full period of data had been available, i.e. these are trends that are unlikely to continue into the future.

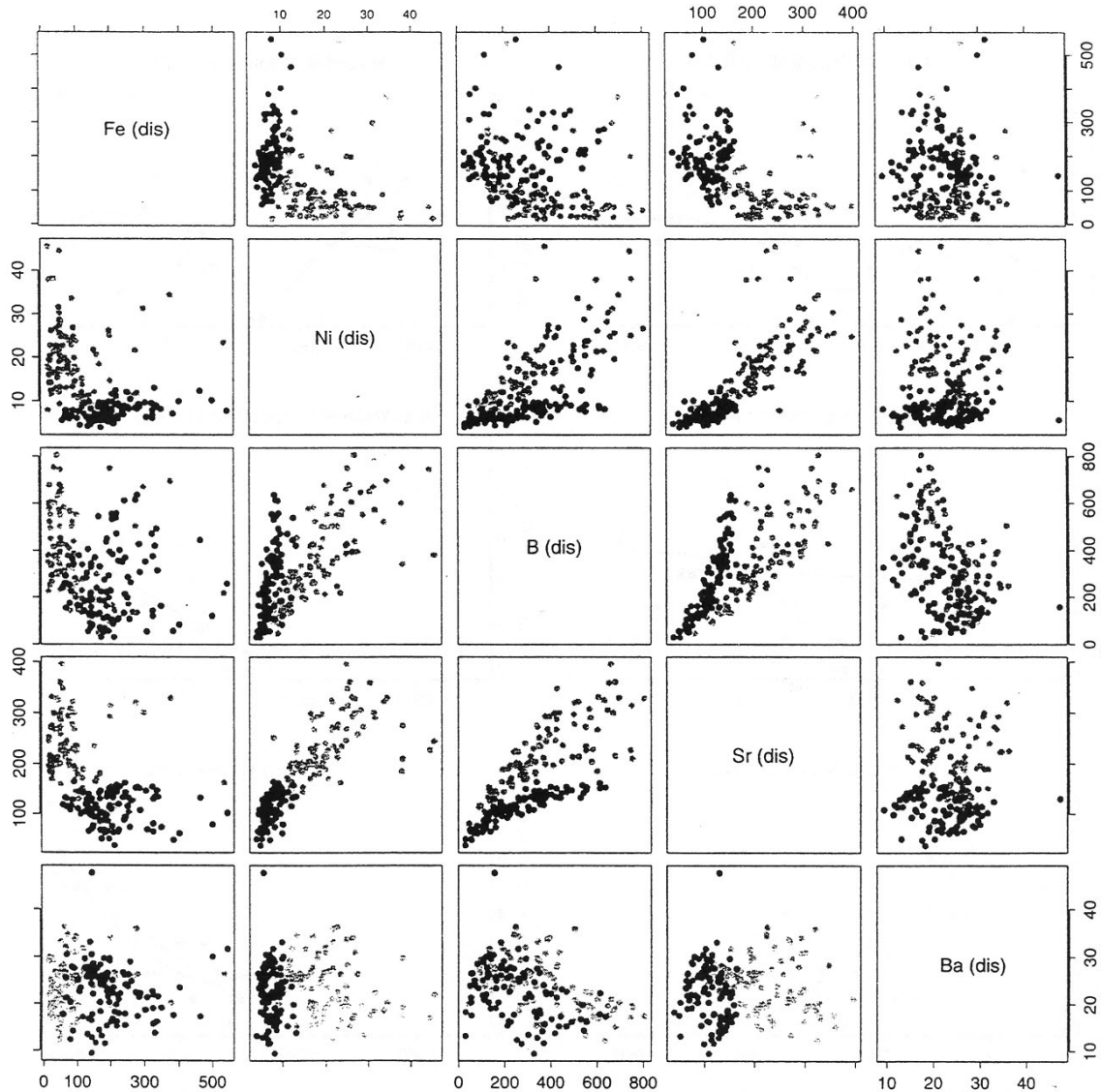


Fig. 4.4. Scatter plot matrix showing the relationships between the time series of dissolved concentrations of five metals measured at two different sites in Northern England (shown in grey and black), with apparently different relationships. More sites could be shown if colour was used. The graphs show (a) positive correlations between B, Ni and Sr, (b) negative correlation between Fe and the three metals listed in (a), (c) little or no relationship between Ba and the other metals, (d) that the ratio of B to Sr is different at the two sites (the gradients seen on the B:Sr graphs are different at the two sites).

Note that a multiple time series plot is an example of a “small multiples” plot (Tufte, 1983), i.e. a large plot containing repeated graphical units. Small multiples can be useful for presenting other types of information too e.g. maps, bar charts etc (see under simple spatial plots below).

4.2.3 Scatterplots

When time series for two variables or sites are recorded at coincident, or close to coincident time points, a scatterplot of pairs of values at these time points can be used to assess common variation between the sites or variables. This simply involves plotting a point for each variable (or site) for each common time point on separate axes. The relationship between the variables (or sites) is then displayed — a tendency to increase together can be easily seen and a smoothing line can help draw the eye towards this.

When the number of series increases, scatterplot matrices (Fig 4.4) can be used to show all pairs of variables (or sites). As the name suggests these consist of a block of scatterplots presented in matrix form, each variable (or site) being plotted against all other variables (or sites).

Scatterplot matrices are most often used when different variables have been recorded (at one or many sites), but can also be used to look at a single variable measured at different sites. Scatterplot matrices are useful for identifying relationships between variables or between sites (including correlation) and for spotting outliers. Colour and interactive visual tools (such as brushing; Cleveland, 1993) can also be profitably used with scatterplot matrices.

As before, a smoothing curve can aid interpretation by highlighting the general relationship between the variables. Residuals from this curve can also indicate which points deviate from a relationship.

4.2.4 Simple spatial plots

When information is available on the spatial location of data from multiple sites, a spatial plot can give a better indication of variation in these dimensions. In their simplest form, a summary measure, such as the mean value or a trend gradient, is plotted at the geographical location of each site (Fig. 4.5). Care may be needed in choice of plotting position! map projection (see Appendix A.1). The magnitude of the value plotted may be denoted by symbol size or by use of a colour scale. Spatial plots are good for identifying spatial patterns and anomalies. Note that this basic type of plot does not apply any smoothing to the plotted values, which might affect the patterns. It will be most clear for relatively sparse spatial data. For dense spatial sampling, some kind of spatial smoothing may help to reduce the clutter, although suitable choice of symbol and colour scale can also provide a clear picture.

It is difficult to display more than a few variables on a single map, so that for multivariate data, each variable can instead be plotted on a separate map — in “small multiple” form if appropriate (Fig. 4.6).

If the continuous spatial variation of a variable measured at different sites is of interest, then simple displays of summaries as symbols on maps (as above) may be inadequate. Instead a “surface” or contour plot representing levels of the variable will be needed (e.g. Fig. 4.7). These require some form of regional smoothing for display (unless the sites are regularly-spaced), which assumes that sites that are close together are more likely to be similar than sites that are far apart. Various regional smoothing methods are available. They include kriging (Isaacs & Srivastava, 1988), which makes explicit assumptions about the covariance between measurements at different sites, and loess in 2-D (Cleveland & Devlin, 1988), which uses a simple weighting of the distances between sites. For variables with a continuous scale of measurement some form of regional smoothing can be useful to show general, spatial trends as the data locations become denser.

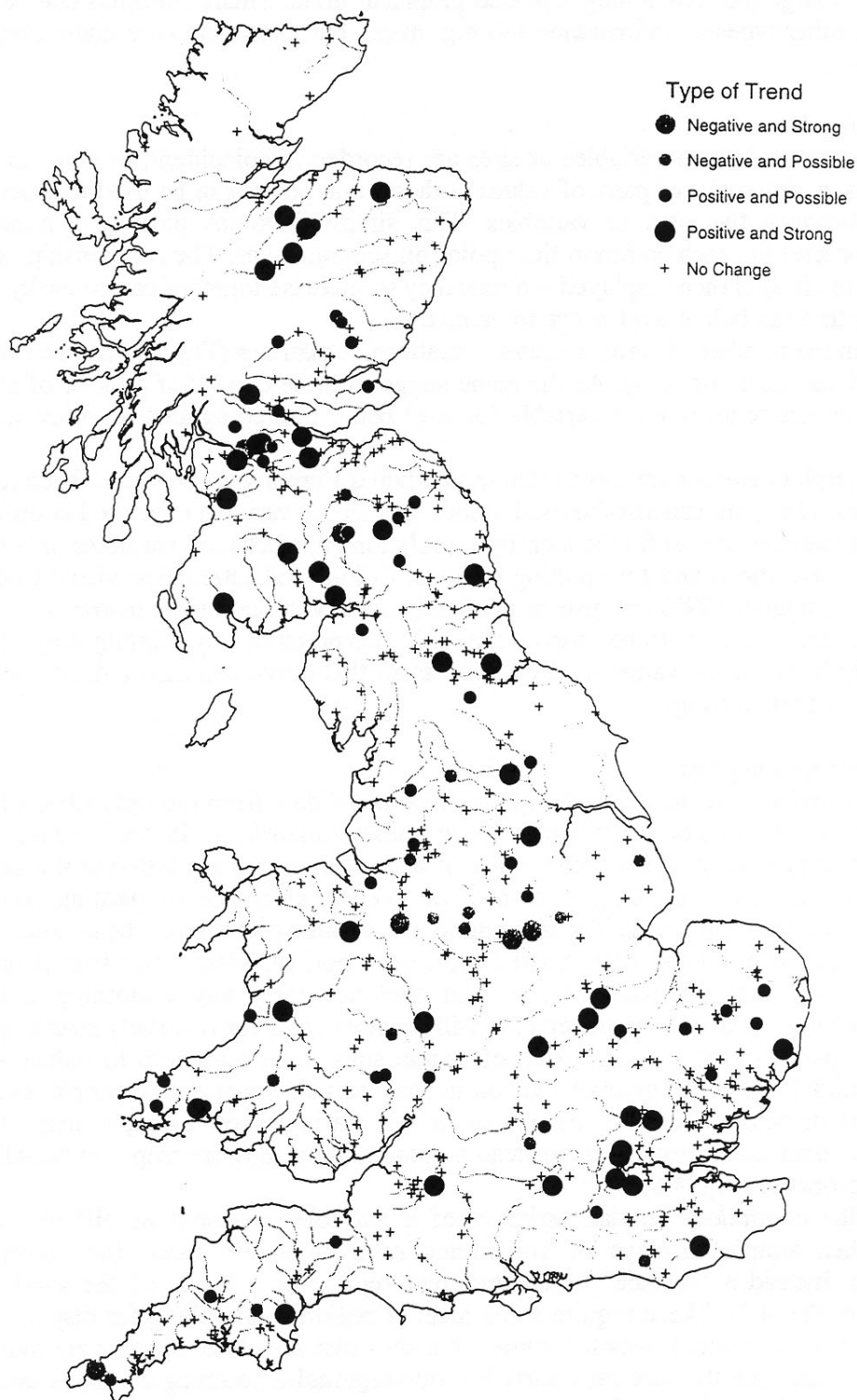


Fig. 4.5. Map showing the results from an analysis of trends in UK flood data. The length of the data series, range from 15 to over 100 years depending on the site. Most series end between 1980 and 1990. Around 600 sites were tested. The symbols have been categorised into five levels and indicate the strength and direction of the trends. There are more incidences of increased flooding than decreased flooding, particularly in Scotland and the South East of England.

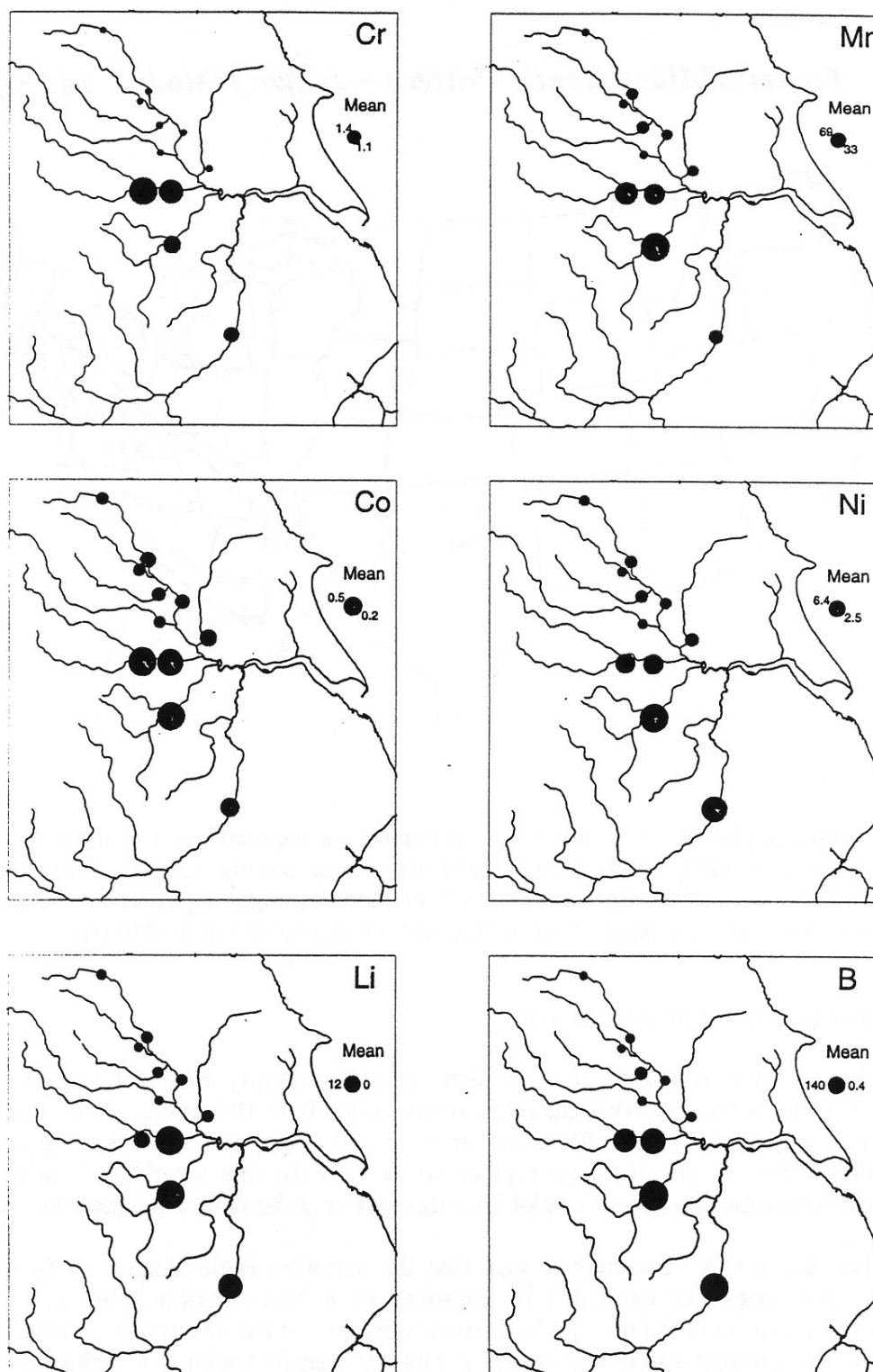


Fig. 4.6. Maps of mean concentrations ($\mu\text{g/l}$) for 6 chemical determinands measured on rivers in the Humber Estuary, eastern England for the period 1983 to 1985. The map shows an area of about 160 km with rivers and coastline marked. At each site, the symbol area represents the average total concentration and is broken into dissolved (black) and particulate (grey) components. The northern rivers show relatively low concentrations (they are rural and with low population density). The southern rivers are affected by domestic and industrial effluents. Chromium has particularly strong industrial sources relative to the background levels seen in the North.

Eastern/Mid-Atlantic Pattern – January (Mode 1, 16.7%)

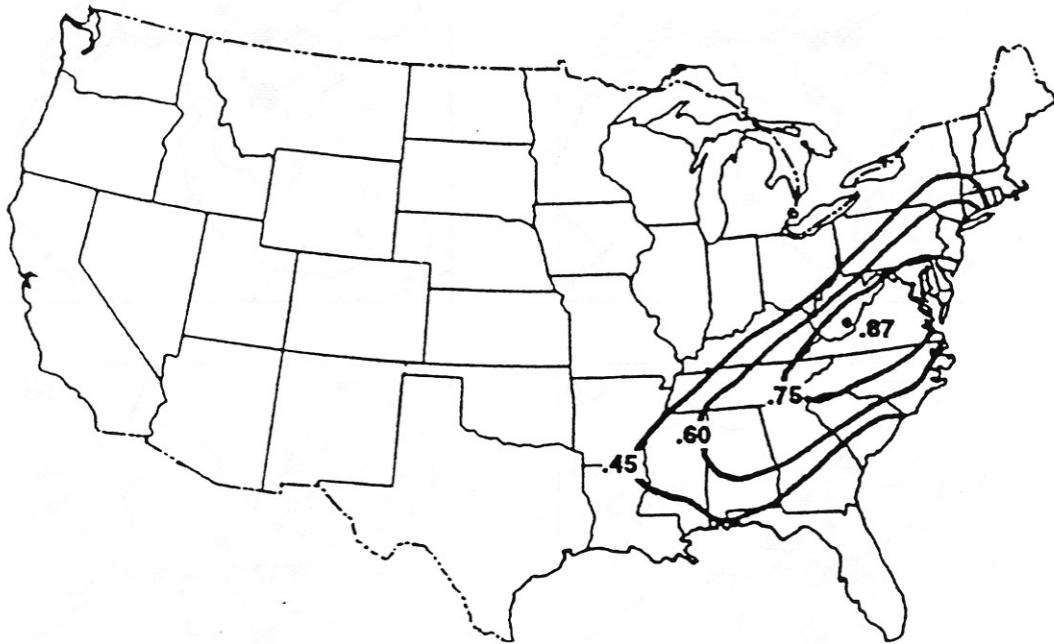


Fig. 4.7. Contour plot of a smoothed interpretation of a regional trend analysis (Lins, 1997) applied to mean monthly streamflow at 559 stations across the United States for January between 1941 and 1988. High values indicate how strongly significant the increase in streamflow was across the region (correlation with an estimated regional trend).

4.3 Further points for time series plots

When plotting time series graphs, the data should generally be displayed as either (i) individual points connected with lines if there are up to about 100 values, (ii) connected lines, if there are many values (Fig. 4.8), or (iii) unconnected points if the data is irregular. If there are missing values in an otherwise regular series then the line should be broken at these points. It is sometimes necessary to plot lines through irregular data when there are very many values.

When the scatter of values is such that the variation is dominated by (a few) large values, a time series plot can often be improved by a data transformation. The goal is to visualise the variation in the data in the clearest way, so that the interesting variation occupies as much of the plotting space as possible. A suitable transformation will reduce the extreme variation. For hydrological series the most common transformation is to take logarithms of the data (Fig. 4.9). NB, use of logarithms is only possible if all data values are positive.

Fig. 4.9 also illustrates the need to adjust the plot aspect ratios (the ratio of height to width) depending on the data variation, and the length of the series (Appendix A.1). Sensible choice of aspect ratio can mean that important features of the data are more easily seen (Fig. 4.9). For very long series, it may be sensible to split the data to produce plots of more manageable size. Cleveland (1993) and (1994) has other examples of the use of aspect ratios to reveal the features of data variation.

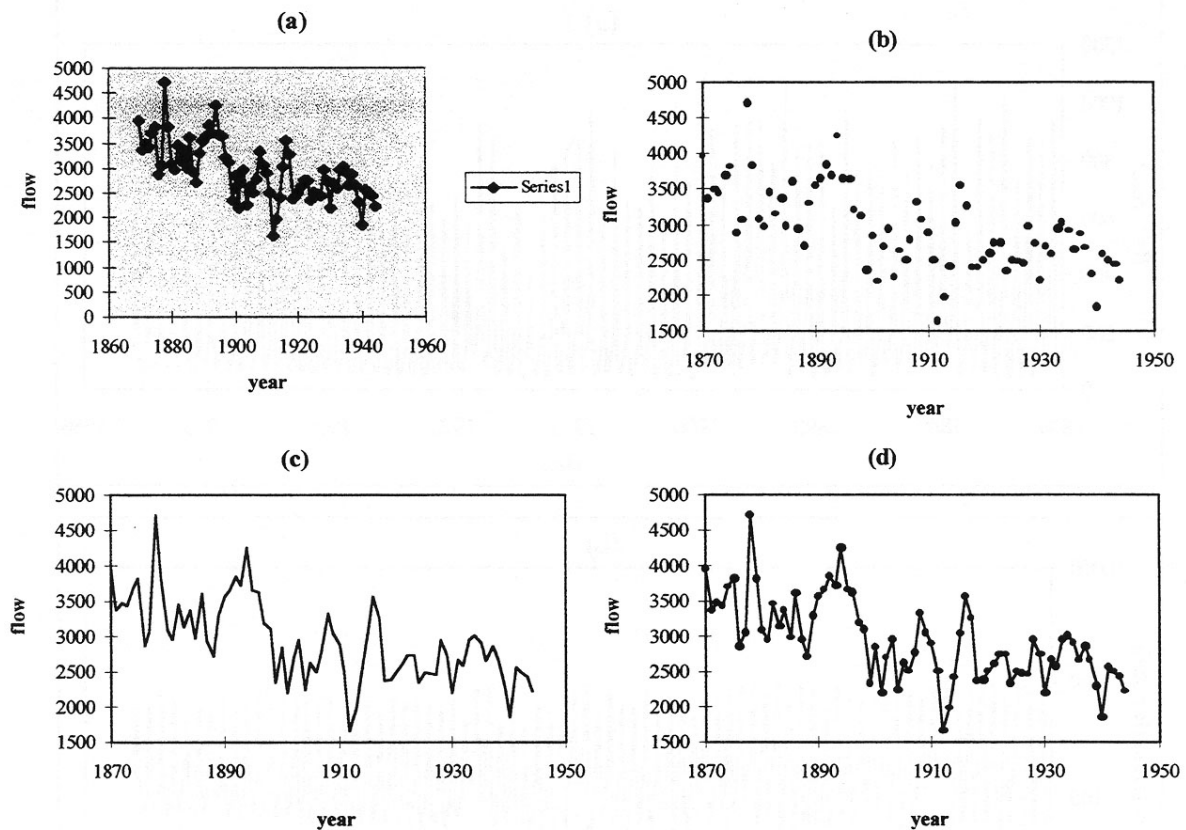


Fig. 4.8. (a) - (d) show basic time series plots of the annual flow (cms) in the river Nile at Aswan, Egypt from 1870 to 1944 (75 years). (a) is the default plot produced by the spreadsheet package and makes poor use of the available space. Plots (b) - (d) are alternatives, with plot (c) or (d) being the best choice in this particular instance. Plot (b) shows points but loses the continuous nature of time. The use of lines in (c) is recommended for a large number of regular data points (with the line broken when there is a break in the data), while a combination (d) is useful for up to about 100 points. Note also that the range of some variables may be far from zero, so that the y-axis may need to be broken e.g. plots (b) - (d) do not include zero, so as to give maximum range on the plot to the variation of interest.

It is usually helpful to add a smoothing line to a time series data plot. This highlights the general local trend (changes in level) in the data. A smoothing line is ideally obtained using some form of robust smoothing (such as a running median or loess, locally-weighted regression — Cleveland & Devlin, 1988 and Cleveland et. al., 1988). Most plotting software is capable of producing a smoothing line to go through regular data e.g. running medians are straightforward to calculate in a spreadsheet. For irregular data more sophisticated smoothing methods will be needed — some implementations of loess are suitable.

To see more subtle types of change, particularly in variation, it can often be powerful to remove the dominant source of variation and plot the residuals (e.g. Fig. 4.10 (e)). This is particularly true if the reason for the underlying variation is well known (e.g. seasonality). This is a technique that can be valuably used in more complex situations, e.g. see the seasonal decomposition examples below.

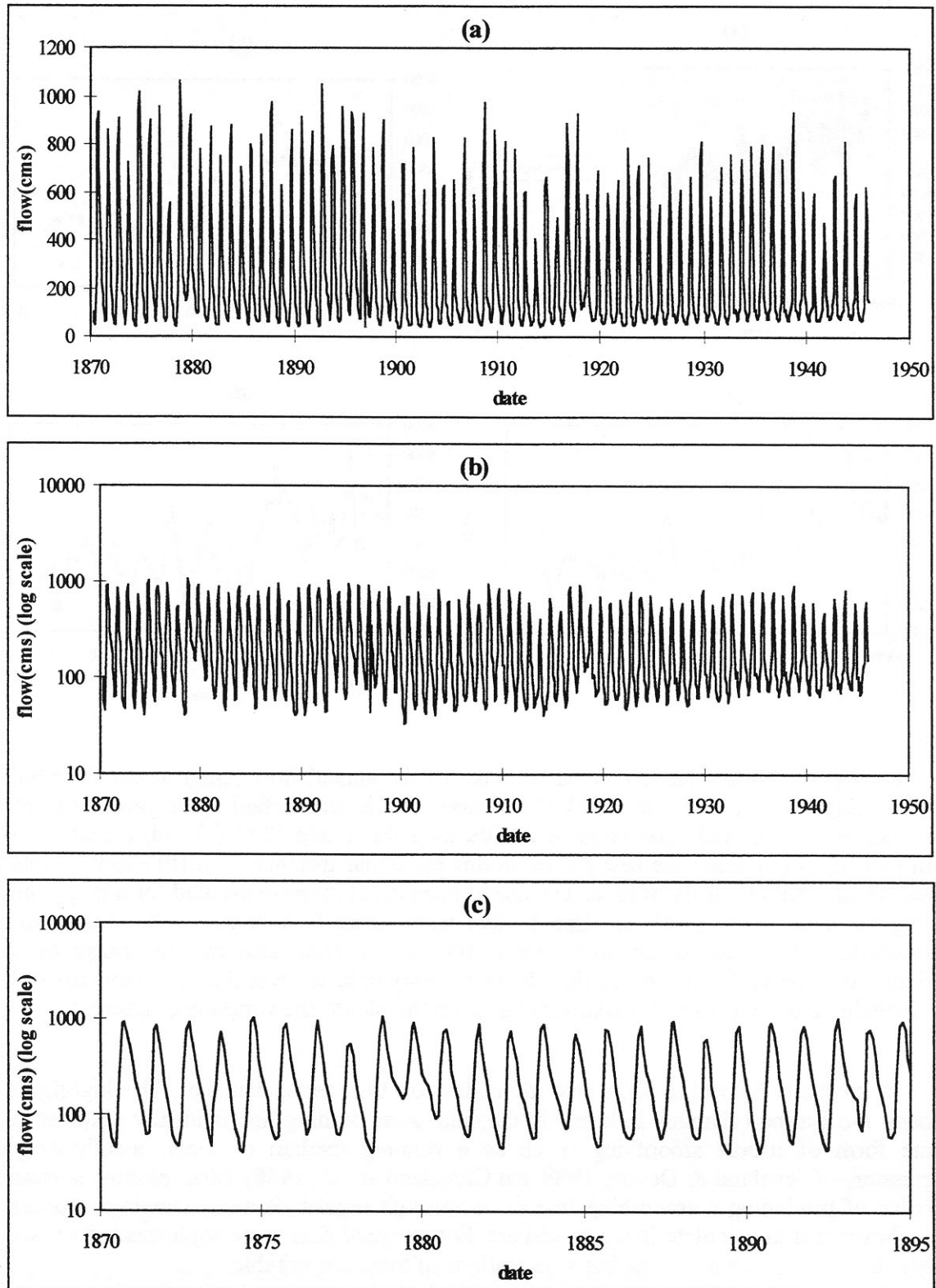


Fig. 4.9. (a)-(b) show the monthly mean flow (cumecs) of the river Nile at Aswan, Egypt (March 1870 to December 1945 – 910 values). (a) shows the flow on the original scale, where the variation is dominated by the large values. (b) shows the flow on a logarithmic (base 10) scale, this evens-out the high and low variation and gives a better idea of the long time-scale fluctuations in the series. (c) shows the first 25 years (again on log scale), it now becomes possible to see the asymmetry of the seasonal pattern (this was not visible in plots (a) and (b) above).

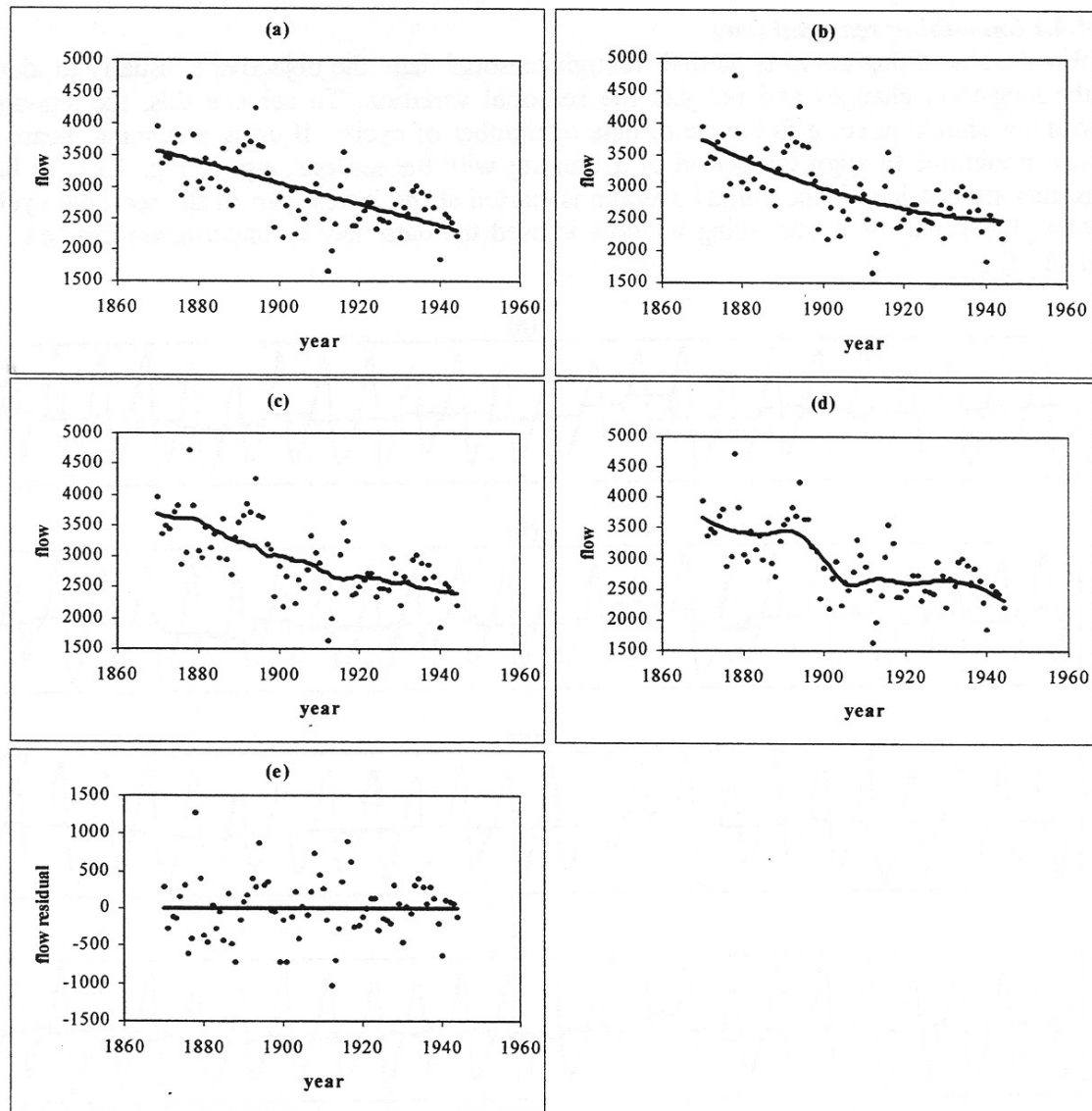


Fig 4.10. Four types of trend line fitted to the annual Nile data from Fig. 4.8 (a) shows a straight linear regression line, (b) a fitted quadratic, (c) a running median and (d) a loess smooth. The line in (d) most closely adapts to the variation in the data and suggests a step-change rather than a continuous trend. (e) shows the residuals from the trend line in (d) giving an indication of changing variability in different portions of the series.

4.4 Seasonal variation

Hydrological data often contains noticeable cycles – e.g. a seasonal cycle will usually be apparent in monthly data, while daily cycles might be observed in hourly measurements. There may be other approximately cyclic variations at longer time-scales, such as modes of climate variability. These will generally not be as regular (i.e. with fixed periods) as seasonal variation and should be considered separately.

To study seasonal variation requires regular or nearly regular observations, e.g. one a month or one a season, more frequent observations being preferable. It is usually helpful to extract the seasonal patterns and then to examine the variation about these general patterns in more detail (see below). If the seasonal period is fixed and known (it is generally annual) any display needs to take account of this.

4.4.1 Smoothing seasonal data

When a smoothing curve is plotted through seasonal data, the objective is usually to identify the long-term changes and not just the seasonal variation. To achieve this, the smoothing window should cover a full cycle of data or number of cycles. If using a running mean, it is also preferable to align the period of averaging with the seasons, e.g. in Fig. 4.11 (b) better results are obtained if the annual average is started at the lowest part of the seasonal cycle in May. If too narrow a smoothing window is used the data may be under-smoothed (e.g. Fig. 4.11 (d)).

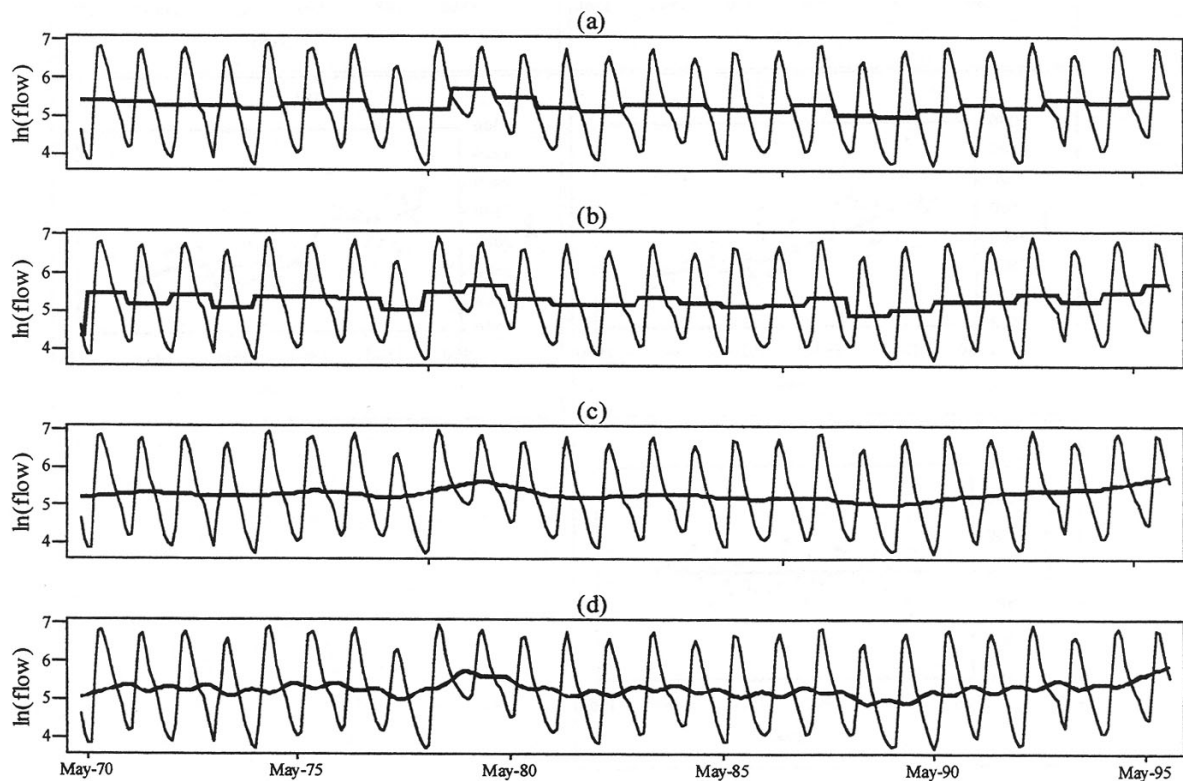


Fig. 4.11. Smoothing lines fitted to monthly Nile flow data of Fig. 4.9. (a) is an annual average (on the log scale), (b) is an annual average starting at the lowest part of the seasonal cycle in May - it gives clearer results, (c) is a loess smooth with a reasonable degree of smoothing, (d) is a loess smooth with too little smoothing, showing seasonal as well as longer-scale variation.

4.4.2 Seasonal patterns

If the seasonal patterns are of interest they can be extracted by grouping all observations within each season together — e.g. for monthly data, collect together all Januarys, all Februarys etc. This can be easily achieved within most software packages by grouping on a variable that indicates the month. The plots below have been produced in the statistical package MINITAB (see Appendix A.2).

Seasonal patterns can be viewed in relation to the within-season variability by combining a line-plot of the underlying average seasonal pattern with boxplots that show the variation within the seasons. Fig. 4.12 shows that the flow always peaks in September, while the within-month variation is higher in the low flow parts of the year (February-July) as might be expected. Better representations of the seasonal variation can be obtained if the seasonal *residuals* are used, for this the average annual levels are first removed from the data,

and then the seasonal pattern and seasonal variation in the residuals are examined (Fig. 4.12 (c), (d)). The annual level can be obtained from either a running mean or a smoothing curve. Plotting seasonal residuals will give a better representation of the seasonal pattern, since longer-term variation between the years has been removed.

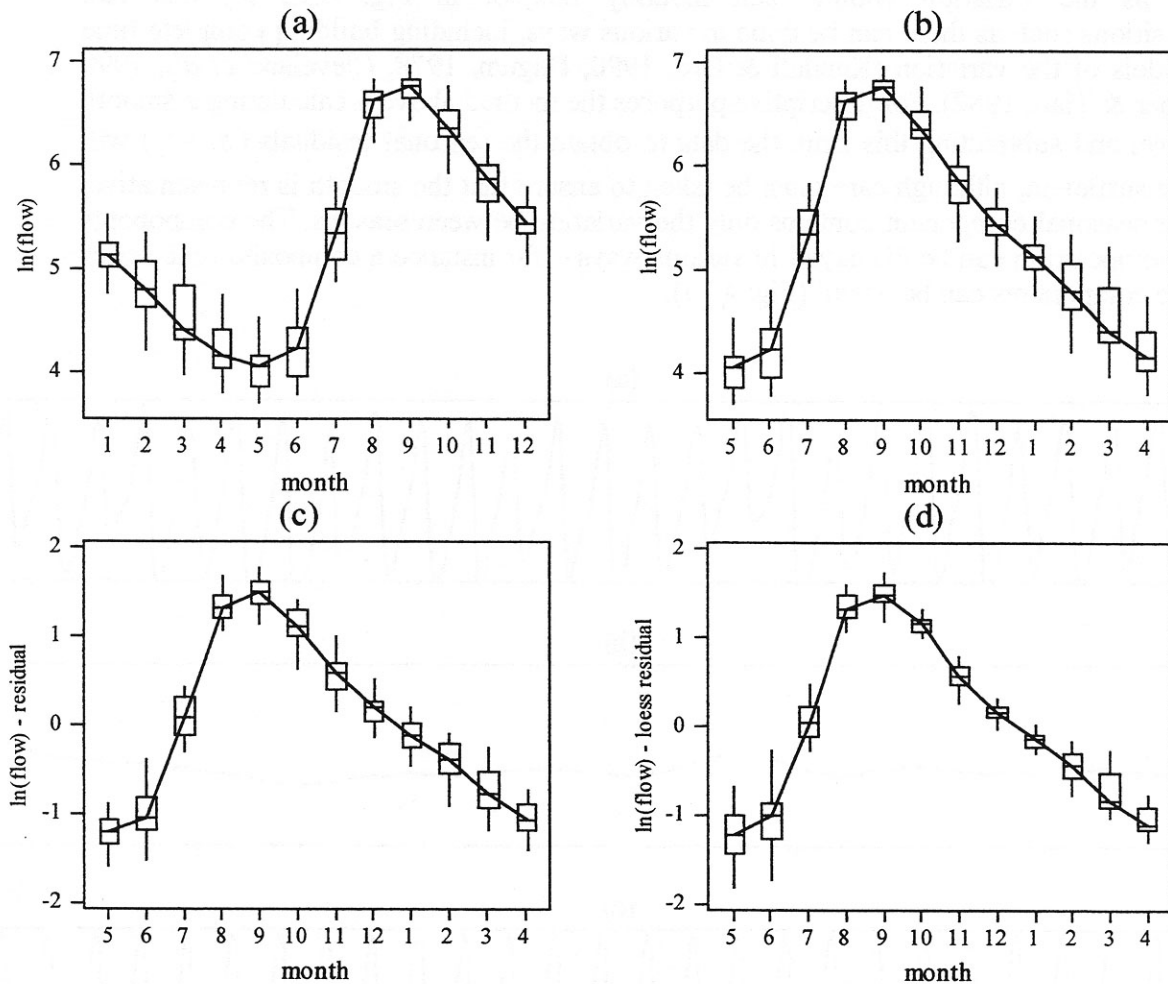


Fig. 4.12. Seasonal boxplots of the monthly Nile flow data of Fig. 4.9, showing the variation for each month of the year - each boxplot shows median (joined with line), lower and upper quartiles in the main box, indicating the main variation in the numbers, while the 'whiskers' show the full range of the data variation for that month, over all of the years of the data. Plot (a) shows that the seasonal pattern peaks in the late summer. By starting the seasonal pattern in May (b), a clearer representation of the seasonal shape, that emphasizes the asymmetry, can be seen. The lower graphs look at seasonal variation after removing the average annual variation shown in Fig. 4.11 (b) above. Plot (c) uses the annual average from a running mean, starting in May, and plot (d) uses the annual level as estimated from the smoothing curve in Fig. 4.11 (c) above. Plots (c) and (d) give better representations of the seasonal pattern.

4.4.3 Seasonal decomposition

The seasons extracted from Fig. 4.12 (c) and (d) above have had the effect of year-to-year variations in the series removed. This is equivalent to decomposing the data y_t into:

$$y_t = I_t + S_t + e_t \quad (4.1)$$

where l_t is the estimate of the annual mean or loess smooth at time t , s_t is an estimate of the seasonal effect, and e_t is a residual error term. The seasonal effect is represented as the deviation from an annual level and takes a fixed value for each season (month in this example). For example, the seasonal effect could be represented using the medians of the boxplots in Fig. 4.12 (c) or (d). The residual errors or remainders, e_t can be seen, in this example, as the variation within each monthly boxplot in Fig. 4.12 (c) and (d). Decompositions such as these can be done in various ways, including building complete time series models of the variation (Kendall & Ord, 1990, Dagum, 1978, Cleveland *et al.*, 1990 and Hillmer & Tiao, 1982). For descriptive purposes the method above — calculating a smooth annual level and subtracting this from the data to obtain the seasonal residuals ($s_t + e_t$) will usually be sufficient, although care must be taken to ensure that the smooth is representative, so that the seasonal component contains only the variation between seasons. The components of the decomposition can be displayed in various ways — for instance a composite time series plot of the components can be useful (Fig. 4.13).

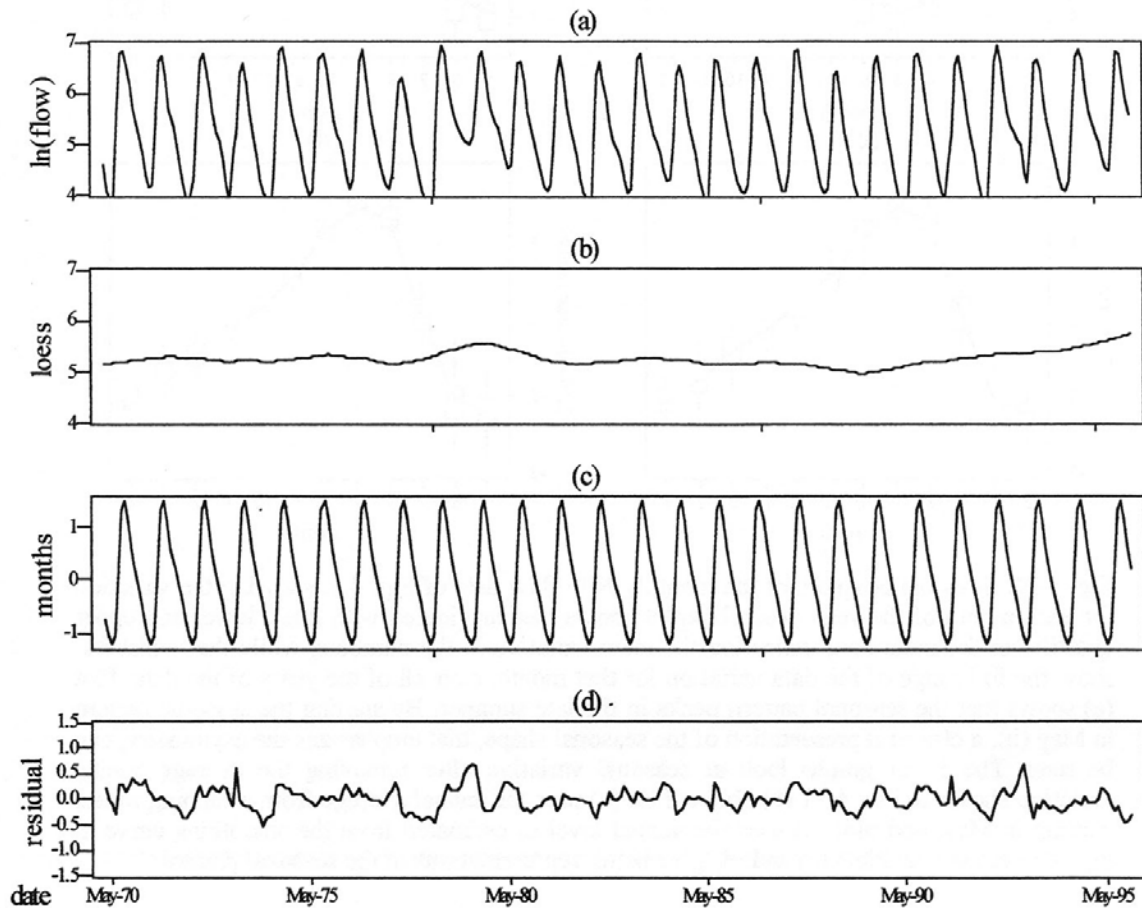


Fig. 4.13. Trend, seasonal and remainder decomposition of monthly Nile flow data (a), showing long-term changes in level (b), regular seasonal variation (c) and residual variation (d). This illustrates how the data is broken down into the various components, so that (a) = (b) + (c) + (d). Note the comparable ranges of the y-axes of each subplot.

4.4.4 Seasonal sub-series

Individual components of the decomposition can be explored in various ways, e.g. seasonal residuals can be summarised by boxplots as above, or individually to see changes in each seasonal sub-series. Fig. 4.14 shows the seasonal residual ($s_t + e_t$) for each July (recall that s_t is constant for any particular month) from the monthly Nile flow data.

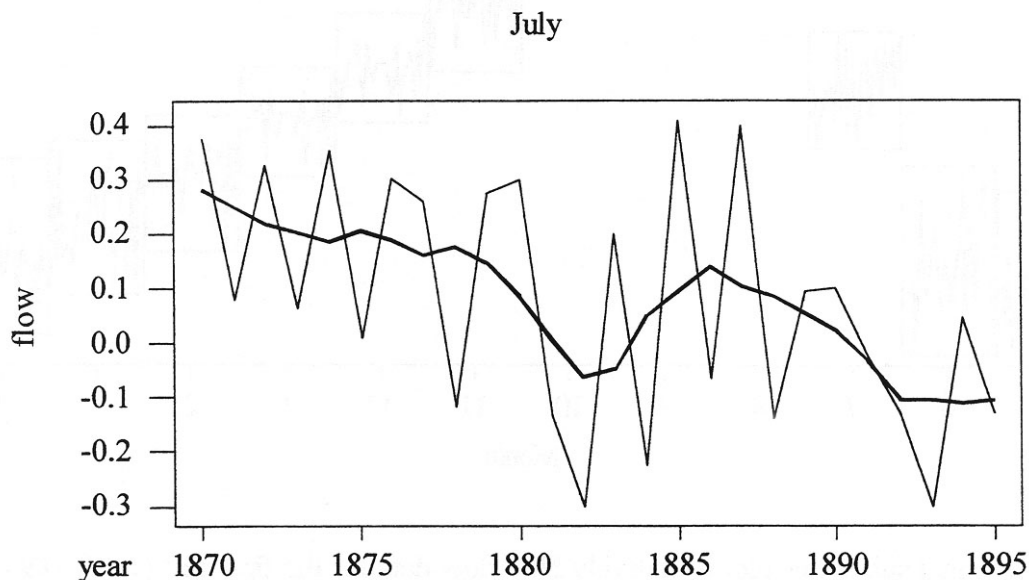


Fig. 4.14. Sub-series of July seasonal component of monthly Nile flow data, with smooth trend, showing an apparent decrease in July flows through the period, relative to the rest of the year.

Seasonal sub-series can be displayed for all of the seasons together using a composite plot built out of components such as Fig. 4.14. This gives a plot that shows the overall pattern of seasonal variation as well as sub-series of changes within each season (Fig. 4.15). Any changes in annual level are removed and the position and size of the 12 monthly boxes indicate the relative seasonal median and variation for each month. The seasonal sub-series is shown within each monthly box as in Fig. 4.14.

Fig. 4.14 and Fig. 4.15 show changes in seasonal pattern occurring over a 25-year period. It is also of interest to examine the full 75-year record. Fig. 4.16 shows the changes in June seasonal residual for these 75 years. It appears that June flows have been increasing relative to the other months, even taking account of annual fluctuations — i.e. June has been becoming less extreme (low). Fig. 4.17 shows a composite sub-series plot for the full 75-year series, which suggests a similar effect in May and April.

Note the outlier that occurs in month 11 (November 1896; Fig. 4.17). This is probably a data coding error; the recorded value is 43.3, but a more sensible value for this month would be 433 (the median value in November is 265).

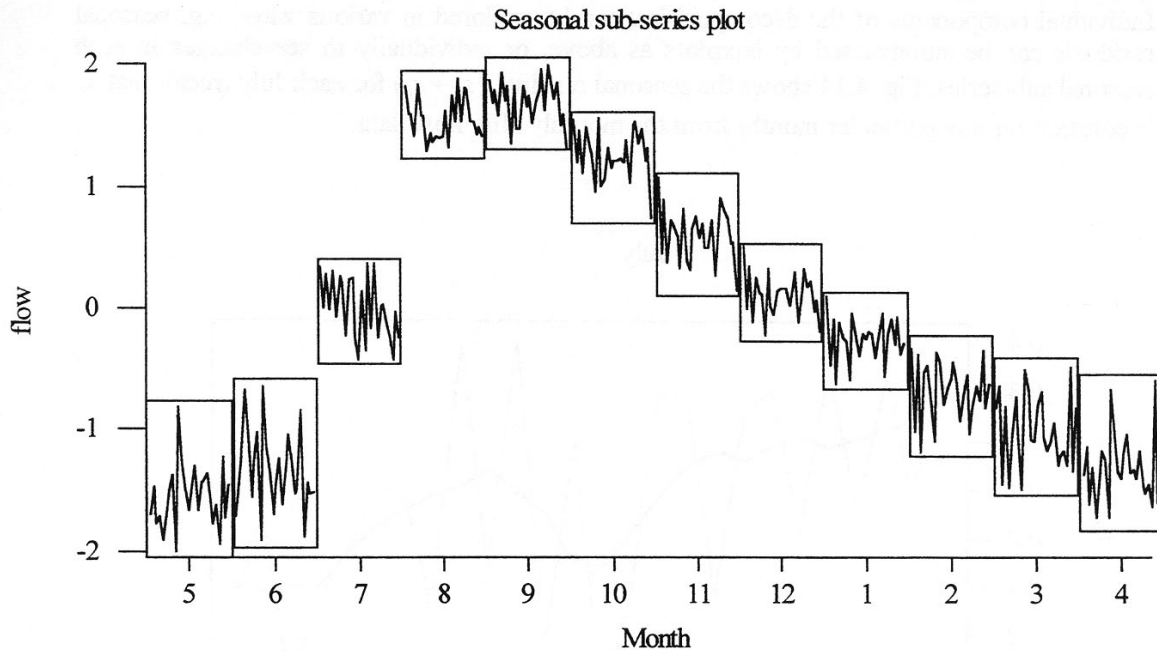


Fig. 4.15. Seasonal sub-series plot of monthly Nile flow data for the first part (1870-1895) of the record. The position and size of the 12 monthly boxes indicate the relative seasonal median and variation for each month. Within each box the seasonal sub-series for each month are shown as in Fig. 4.14. This is effectively ordering the values in the boxplots of Fig. 4.12 as time series. The plots show the general seasonal shape (mid-point of boxes) and trend and variation within each season (month).

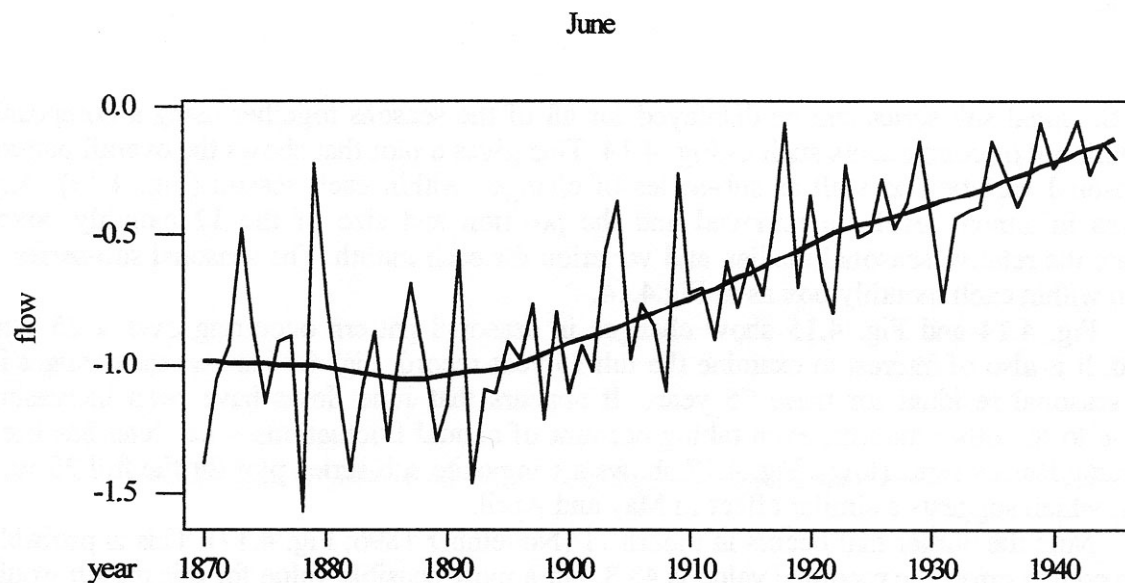


Fig. 4.16. June seasonal sub-series and smooth trend for complete monthly Nile flow data from 1870-1945. The June flows can be seen to be increasing towards the annual median (0 residual) through the years.

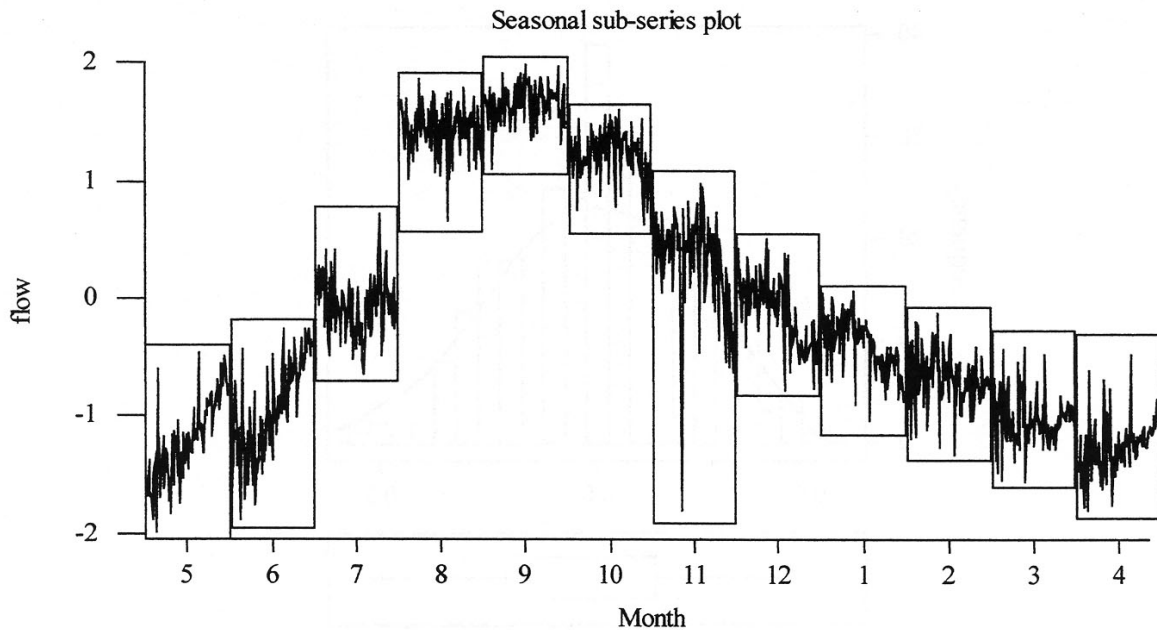


Fig. 4.17. Seasonal sub-series plot for complete monthly Nile flow data from 1870-1945. May, June and April show patterns of increasing flow towards the annual median. Other months seem stable relative to the median. A possible outlier can be identified in November.

4.5 Residual analysis (checking test assumptions)

Most tests for detecting change make some assumptions about the data, or more particularly about the *residuals* from a suitable model for the data. For example, a test based on linear regression has a model that the mean changes linearly over time, the main assumptions are that the residuals are independent and are normally distributed. Other types of test will make differing assumptions. Exploratory data analysis is a valuable way of checking that test assumptions hold.

4.5.1 Distributional assumptions

Many common statistical tests make assumptions about the distribution of the residuals. For example linear regression assumes a normal distribution. If a normal distribution is assumed, but the residuals actually follow a distribution that is skewed or heavy-tailed (e.g. there are some very large values) then this can affect the significance level of the test, and render the result incorrect.

Note that many of the tests recommended in this report are distribution-free tests, i.e. tests that do not require assumptions to be made about the distribution of the residuals. In these cases the test results are not sensitive to the distribution of the residuals

Boxplots, histograms and quantile plots can be used to examine the distribution of residuals. Histograms show the general shape of the distribution (Fig. 4.18). Quantile or q plots simply plot the data values against their rank, or against the equivalent quantiles from a reference distribution e.g. the normal (Fig. 4.19). If the quantile plot gives a straight line then the data can be assumed to come from the required distribution. If the quantile plot deviates significantly from a straight line then this indicates departure from the assumed distribution and indicates which part of the data deviates from this distribution. Quantiles display the data without smoothing, which can be useful. For example, plots of the untransformed Nile data (Fig. 4.20 - 4.22) show asymmetry, confirming the need for the log transformation.

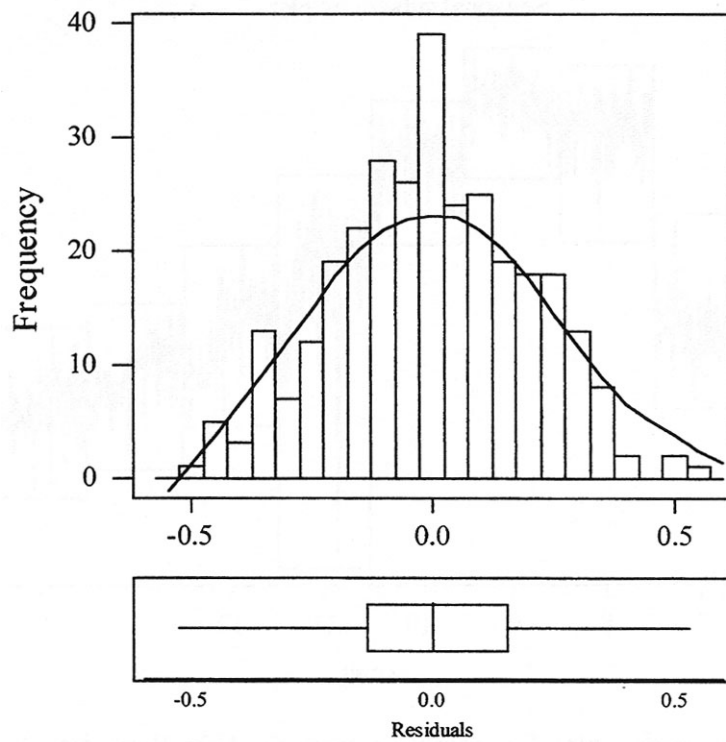


Fig. 4.18. Histogram (with smooth estimate) and boxplot (below) of residuals from seasonal model for monthly Nile flow data showing reasonable symmetry.

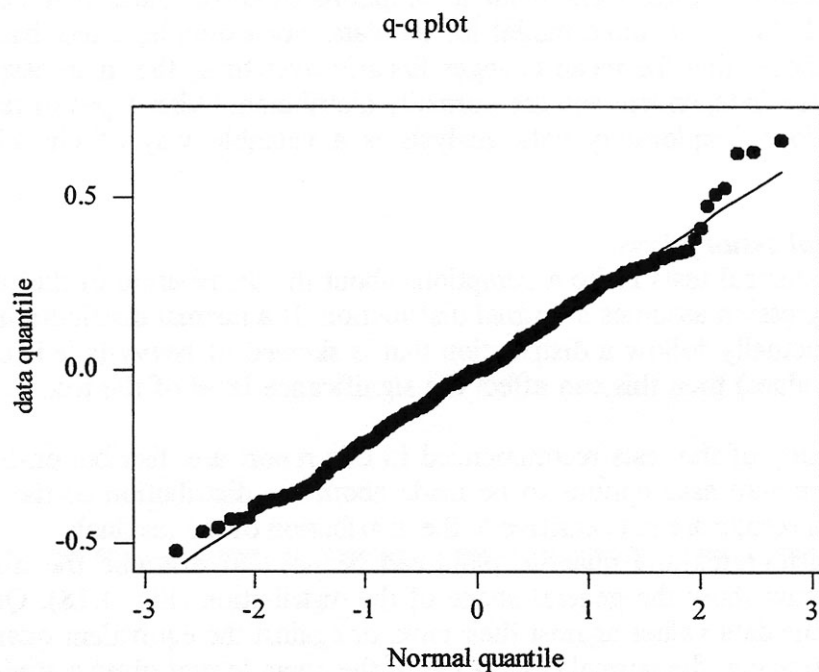


Fig. 4.19. Quantile-quantile (normal) plot of residuals from seasonal model for monthly Nile flow data. This distribution is very slightly skewed — some of the high and low values are higher than the equivalent quantiles of a normal distribution (line).

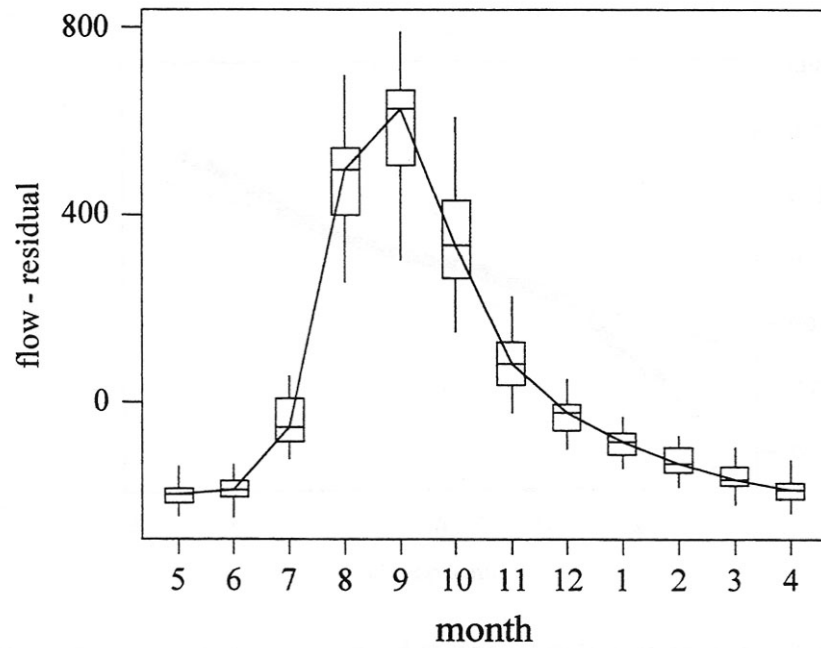


Fig. 4.20. Seasonal pattern of flow (original scale) of monthly Nile flow data. This is the equivalent of Fig. 4.12 for the original data - the asymmetry and heteroscedasticity indicates non-normality and confirms that a transformation (in this case log) is required.

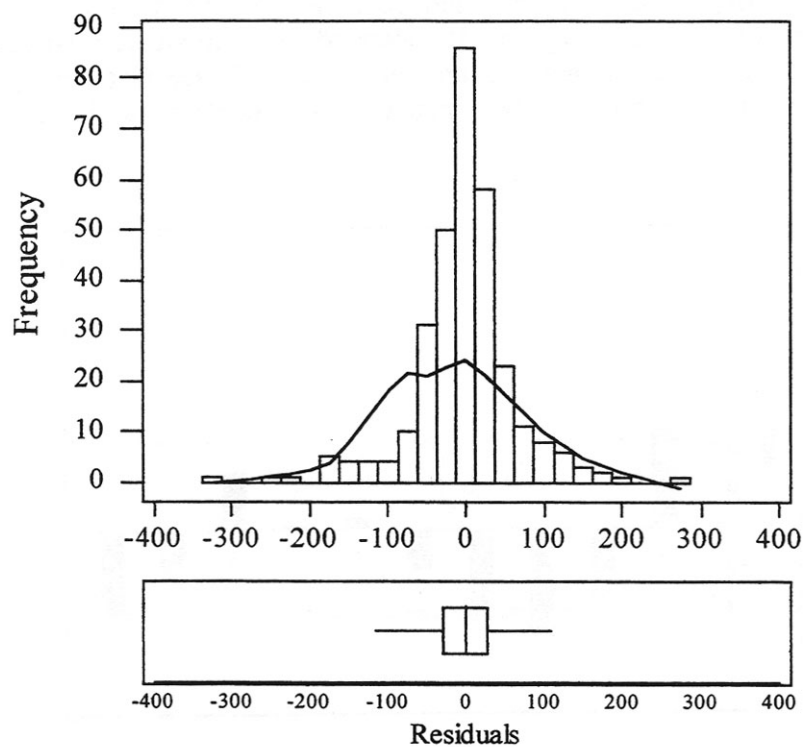


Fig. 4.21. Histogram (with smooth estimate) and boxplot of residuals from seasonal model for monthly Nile flow data (original scale). This appears to show a longer-tailed distribution than a normal.

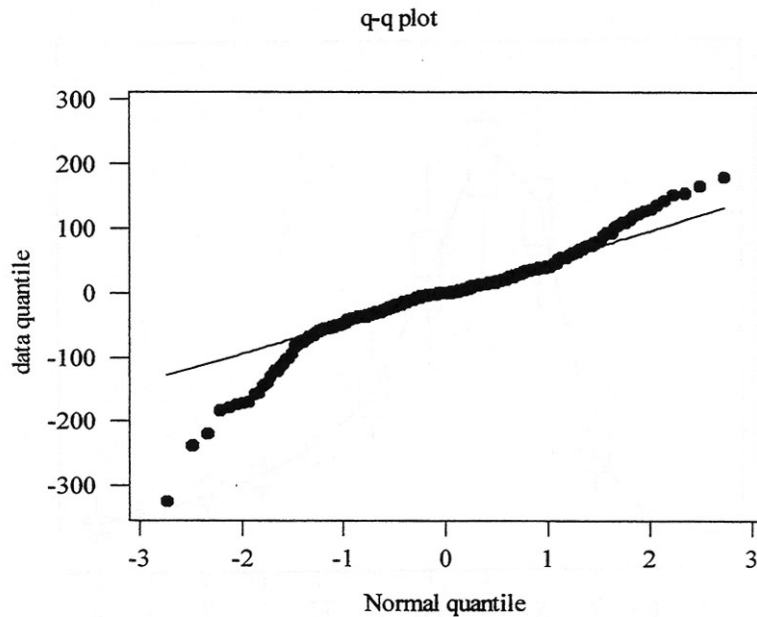


Fig. 4.22. Quantile-quantile (normal) plot for residuals from seasonal model for monthly Nile flow data (original scale). These residuals do not follow a normal distribution — both tails, particularly the lower, are longer than the equivalent normal quantiles.

4.5.2 Independence

Most tests for detecting change assume independence of the sample values. The tests presented in this report are no exception. If independence cannot be assumed then block-bootstrap and block-permutation methods are recommended (see Chapter 5). These avoid independence assumptions by building the dependencies in the data into the test.

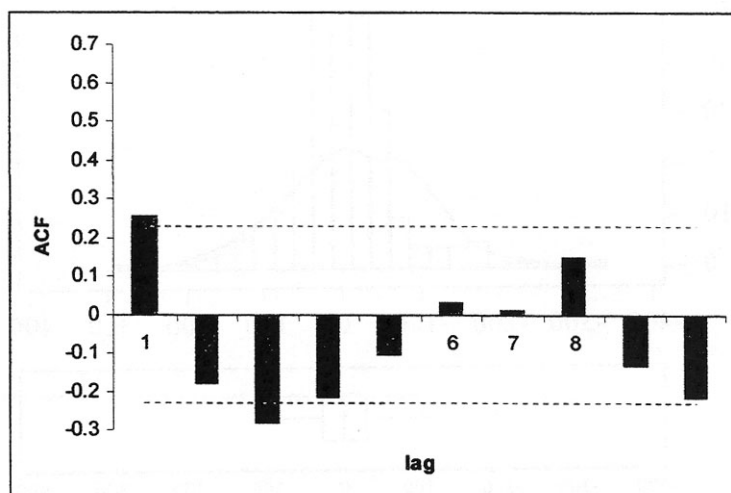


Fig. 4.23. Autocorrelation of annual Nile data showing apparent positive correlation between observations up to 8 years apart. Values that lie above (or below) the dotted lines indicate that there is a significant correlation at the lag (in years) indicated. If the data values were independent then most correlation values should lie between the dotted lines.

Independence means that knowing the current value of a variable provides no information about what the next value will be. This clearly does not hold for time series data, which are usually correlated due to being observed frequently or seasonally. For example, knowing today's flow provides a lot of information about what tomorrow's flow is likely to be. One way to quantify the extent of the correlation (dependence) is to calculate the autocorrelation function (ACF — see e.g. Kendall & Ord, 1990). Autocorrelation is a measure of the correlation of a variable with itself, but with the time shifted. For example, a lag 1 autocorrelation for a daily series is the correlation between the series and the same series but moved 1 day. The lag 2 autocorrelation is the correlation with a time difference of 2. The autocorrelation plot shows the correlations at a series of lags (Fig. 4.23 — Fig. 4.26). If autocorrelation is present at one or more lags then the data is not independent. A sample variogram (Isaaks & Srivastava, 1989) can be used in a similar way to the autocorrelation plot when data is irregularly spaced.

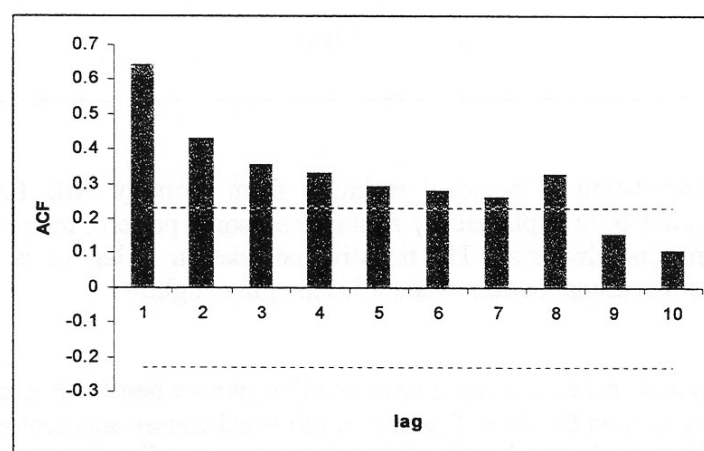


Fig. 4.24. Autocorrelation of annual Nile residuals after removing the trend. The correlation has reduced markedly, leaving only small dependency between values up to 3 years apart.

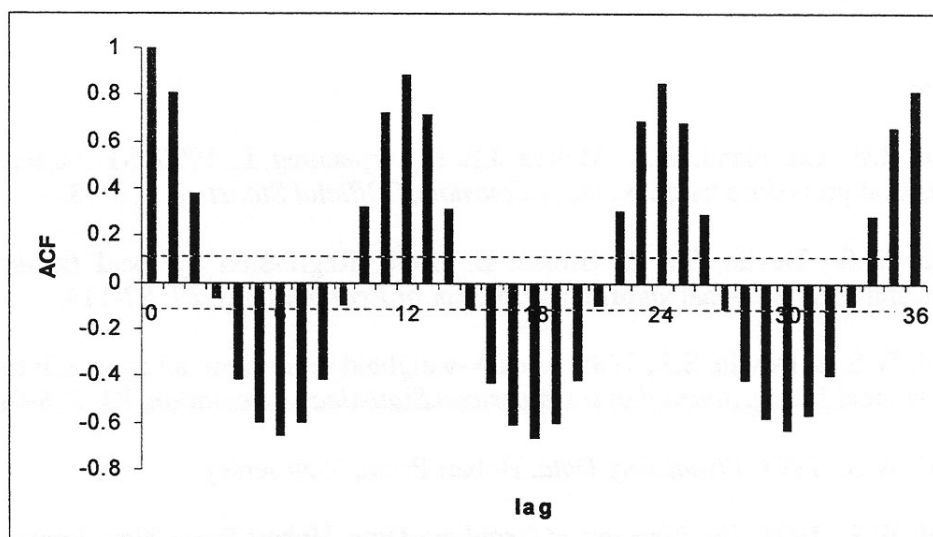


Fig. 4.25. Autocorrelation of monthly Nile flow data showing very strong seasonal (12 monthly) correlation indicative of strongly seasonal data. It can be seen that each month is similar to the same month the previous year (lag 12 months).

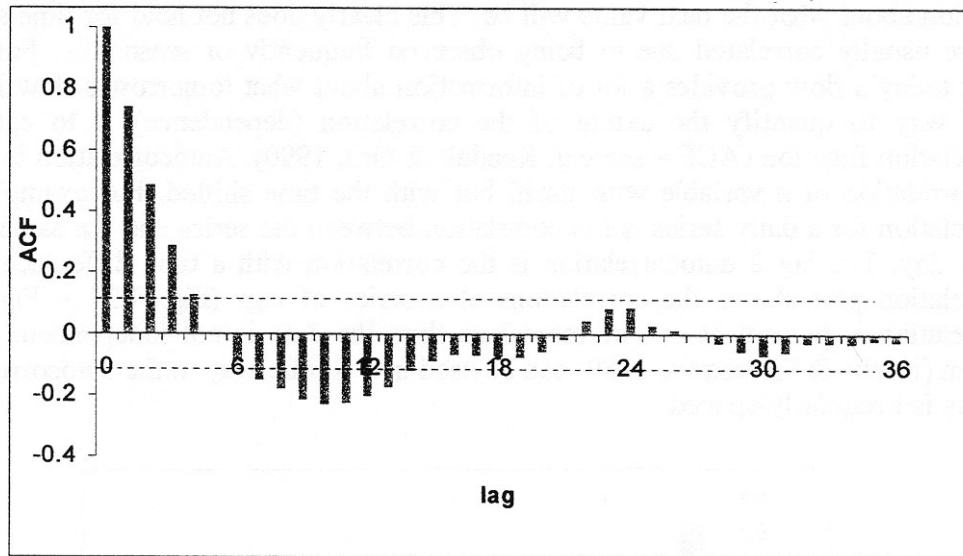


Fig. 4.26. Autocorrelation of seasonal residuals from monthly Nile flow data. Once the seasonal variation has been explained by a simple seasonal pattern, the remaining correlation dies out after three months or so. The negative correlation at lag 12 is most likely due to assuming a constant seasonal pattern when it is changing slightly.

Care is required in interpreting autocorrelation plots when there is an underlying trend or long-term fluctuation in the data. This is because trend causes apparent autocorrelation, e.g. compare Fig. 4.23 (raw data) with Fig. 4.24 where the underlying fluctuation (estimated by a smoothing curve) has been removed and the residuals show much less correlation.

Autocorrelation can also be used for data on shorter time-scales, e.g. for monthly data the seasonal influence is apparent (Fig. 4.25). However, if the ACF of the residuals from the seasonal decomposition of Fig. 4.13 is examined, it shows greatly reduced correlations, although still significant from month to month (Fig. 4.26).

References

- Cleveland, R.B., Cleveland, W.S., McRea, J.E. & Terpenning, I., 1990. STL: a seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, **6**, 3-73.
- Cleveland, W.S., Devlin, S.J. & Grosse, E., 1988. Regression by local fitting: methods, properties and computational algorithms. *Journal of Econometrics*, **37**, 87-114.
- Cleveland, W.S. & Devlin, S.J., 1988. Locally-weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association*, **83**, 596-610.
- Cleveland, W.S., 1993. *Visualizing Data*. Hobart Press, New Jersey.
- Cleveland, W.S., 1994. *The Elements of Graphing Data*. Hobart Press, New Jersey.
- Dagum, E.B., 1978. Modelling, forecasting and seasonally adjusting economic time series with the X-11 ARIMA method. *The Statistician*, **27**, 203-216.

Hilmer, S.C. & Tiao, G.C., 1982. An ARIMA-model-based approach to seasonal adjustment. *Journal of the American Statistical Association*, **77**, 63-70.

Isaaks, E.H. & Srivastava, R.H., 1989. *Applied Geostatistics*. Oxford University Press, Oxford.

Kendall, M. & Ord, J.K., 1990. *Time Series*. Edward Arnold, London, 3rd edition.

Lins, H.F., 1997. Regional streamflow regimes and hydroclimatology of the United States. *Wat. Resour. Res.*, **33** (7), 1655-1667.

Tufte, E.R., 1983. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT.

Additional references

Harvey, A.C., 1990. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.

Hastie, T.J. & Tibshirani, R.F., 1990. *Generalized Additive Models*. Chapman and Hall, London.

Mosteller, F. & Tukey, J.W., 1977. *Data Analysis and Regression*. Addison-Wesley, Reading, MA

Tufte, E.R., 1990. *Envisioning Information*. Graphics Press, Cheshire, CT.

PLOTTING PRACTICE AND STYLE

Some general points to consider for all plots (where relevant):

1. Axis ranges — these need not always include zero and for multiple graphs should be common if the range of variation makes this possible. Also, if there are a few extreme points that dominate the variation; consider transforming or removing these points to see the variation in the rest of the data.
2. Fill the plot area — the aim is to display the variation in the data, so the axes, aspect ratio and titles should be controlled to give most of the space on the plot to the interesting variation.
3. Aspect ratio — the shape of the plot can be varied and correct choice of the aspect ratio (height to length) can make a big difference to the features that are shown. Try differently shaped plots to see the effect — generally make the plots longer when the data series is long. It may be necessary to split a time-series plot into a number of consecutive segments if the series is very long.
4. Labelling, titles, legends — these all help to identify aspects of a graph. Generally, axes should be clearly labelled, with sufficient tick marks to identify points. Axis titles and an overall title are also useful, but are not necessary on each plot in a small multiple with common ranges. Legends are only necessary when different lines or points need to be distinguished. Care needs to be taken that the axes are labelled at sensible intervals, particularly if dates are involved.
5. Gridlines — these can help to read off values, or compare them easily with a target, but too many of them make the display cluttered and they should normally be shown as dotted or greyed.
6. Symbols, size, number — symbols for points should be small enough to avoid cluttering the plot, but large enough to distinguish between different types, if present. For data that falls into categories, or is otherwise discrete, perhaps due to measurement rounding, it can be helpful to “jitter” points (adding a small amount of noise) to move them slightly away from the common value.
7. Points vs lines — lines should only be used when there is a connection between the points, such as consecutive in time or when there are a large number of points. Gaps should be left for missing values.
8. Use of colour — this is often over-used. Essentially it is only necessary if used to display another “dimension” on the graph, for which it can be powerful. Always consider how many different elements of variation are being displayed to determine whether colour is necessary. It can often be clearer to produce multiple black-and-white plots than to try and put all of the information on a single one with categories in different colours.
9. Map co-ordinates - 1 minute of latitude and longitude are quite different at high latitudes, so ideally positions should be converted to kilometres from a reference point, or for large regions a correct map projection should be used.
10. Smoothing lines - These can be running means or locally weighted smoothing lines (Cleveland & Devlin, 1988). Care must be taken that the degree of smoothing is appropriate for the data, e.g. if seasonality is present the smoothing window must cover a number of years.
11. Small-multiples - there are some style issues that are important for “small multiple” graphs. For example, maximising each graph compared to the axes, keeping the spacing between each graph small, using a uniform x-axis scale, sometimes just labelling the axes on the outside edges.

Appendix A.2

SOFTWARE NOTES

Possible software packages range from spreadsheet packages (such as Excel) through desktop statistical packages (such as MINITAB) to powerful statistical programming languages (such as S-PLUS and its freeware version, R).

Spreadsheet packages can allow many of the basic plots to be produced. However, care is often needed. Beware of the following:

- Time is a continuous variable and so should not be plotted as categories (which is the default in some, for example, spreadsheet packages), particularly if observations are made at irregular intervals, as can often be the case for hydrological series.
- There are sometimes limitations on the number of points that can be displayed by a package.

More complex plots will require software with some programming capability — though often only fairly basic. However, the power to decompose the variation into interesting components and to highlight subtle types of change, which cannot be seen in a simple time series plot of the data, makes this well worthwhile.

Complex spatial displays are not available in all software packages and few will have satisfactory methods of smoothing, or displaying sub-plots at spatial locations, making some programming expertise essential.

The plots shown in this section were produced using a combination of Excel, MINITAB and S-PLUS. The plots of seasonal components were produced in the statistical package MINITAB, using a simple macro based upon ‘recording’ the code for a standard plot, and modifying this code. The coding is reasonably straightforward and gives complete control over layout, titles etc. Excel was used for the simpler time series plots. S-PLUS was used for the small multiples and some of the spatial maps.

CHAPTER 5

STATISTICAL METHODS FOR TESTING FOR CHANGE

Alice Robson, Andras Bardossy, David Jones & Zbigniew W. Kundzewicz

5.1 Introduction

This chapter deals with statistical methods for formal testing of change. These methods are intended to be used as one part of the process of statistical analysis, i.e. alongside data quality control, visual exploratory data analysis etc. An overview of how the various stages of a statistical analysis fit together is given in Chapter 2.

The chapter begins by summarising some of the basic statistical concepts needed for testing change (Section 5.2) and then introduces a selection of statistical methods that are available for testing for change (Section 5.3). It looks in detail at distribution-free methods, particularly resampling methods such as permutation and bootstrapping (Section 5.4). Distribution-free methods can be applied even if data is strongly non-normal - as is typically the case for hydrological series. Section 5.5 examines the assumptions that are required for statistical testing and considers the issue of independence. This is a critical assumption that is commonly required even within distribution-free testing approaches. Methods for assessing and coping with dependencies in hydrological data are discussed. Section 5.6 summarises choice of tests/test statistics. Section 5.7 considers interpretation of test results and how to determine the cause of change. The final section provides a summary.

5.2 Setting the scene

5.2.1 *Some basics of statistical testing for change*

This section provides an overview of some of the main statistical concepts and terminology required for the statistical testing of change. For further details on statistical testing the reader should refer to a standard introductory statistical textbook (e.g. Chatfield, 1970) or to texts such as Hirsch et al. (1992) and Helsel & Hirsch (1992).

Types of change

Change in a series can occur in numerous ways: e.g. steadily (a trend), abruptly (a step-change) or in a more complex form. It may affect the mean, median, variance, autocorrelation or almost any other aspect of the data.

The most widely used tests for change look for one of the following

- Trend in the mean or median of a series
- Step-change in the mean or median of a series.

There are also some tests that look for a general change in distribution (Section 5.6). Trend and step change are special cases of a change in distribution. Tests for a change in distribution are generally not particularly powerful: i.e. if trend is present it would be best detected by a test for trend. However, such tests may be useful as a general check for evidence of change.

Testing for more complex types of change and for measures other than the mean/median generally requires use of advanced techniques such as Maximum Likelihood (see also Section 5.3). Typically these techniques can be difficult to apply and are beyond the scope of this report.

Hypotheses

The starting point for a statistical test is to define the null and alternative hypotheses; these are statements that describe what the test is investigating. The null and alternative hypotheses are usually framed in terms of the types of change described above. For example, to test for trend in the mean of a series the null hypothesis (H_0) would be that there is no change in the mean of a series, and the alternative hypothesis (H_1) would be that the mean is either increasing or decreasing over time. To test for step-change in the mean of a series, the null hypothesis would again be that there is no change in the mean of the series, but the alternative hypothesis would be that the mean of the series has suddenly changed.

The starting point for statistical testing is to assume that the null hypothesis is true, and then to check whether the observed data are consistent with this hypothesis. The null hypothesis is rejected if the data are not consistent.

Test statistic

The test statistic is a means of comparing the null and alternative hypotheses. It is just a numerical value that is calculated from the data series that is being tested. A good test statistic is designed so that it highlights the difference between the two hypotheses. A simple example of a test statistic is the linear regression gradient: this can be used to test for a trend in the mean. If there is no trend (the null hypothesis) then the regression gradient should have a value near to zero. If there is a large trend in the mean (the alternative hypothesis) then the value of the regression gradient would be very different from zero. More formally, to carry out a statistical test it is necessary to compare the observed test statistic with the expected distribution of the test statistic under the null hypothesis. The significance level of a test statistic expresses this concept more formally.

Significance level

The significance level is a means of measuring whether a test statistic is very different from values that would typically occur under the null hypothesis. Specifically, the significance level is the probability of a value as extreme as, or more extreme than the observed value, assuming “no change” (the null hypothesis). In other words, significance is the probability that a test detects trend when none is present.

A possible interpretation of the significance level might be:

- Significance level $>10\%$ - very little evidence against the null hypothesis (H_0)
- 5 % to 10 % - possible evidence against H_0
- 1 % to 5 % - strong evidence against H_0
- below 1 % - very strong evidence against H_0 .

Note that when reporting results the actual significance levels should normally be quoted (e.g. a significance level of 6.5 %).

For many traditional statistical methods, significance levels can be looked up in reference tables or calculated from simple formulae, providing the required test assumptions apply. In general, the significance level can be found if the distribution of the test statistic under the null hypothesis (i.e., assuming the null hypothesis is true) is known or can be estimated. One case where this distribution is usually easy to determine is where the data are independent and normally distributed. Resampling methods provide an alternative, robust method of estimating the test statistic distribution in a general case.

Power and errors

There are two possible types of error that can occur in a test result. The first is that the null hypothesis is incorrectly rejected (type I error) - the significance level expresses the probability of this error. The second is that the null hypothesis is accepted when the

alternative hypothesis is true — type II error. A test which has low type II error probability is said to be powerful. In general more powerful tests are to be preferred. The power of the test is the probability of correctly detecting a trend when one is present.

5.2.2 Cautionary words

Poorly understood data gives poor results

Statistical tests can be easily misapplied unless the data is thoroughly understood. A prerequisite before undertaking any formal statistical testing, is that the data should be quality controlled (Chapters 2, 3) and that an exploratory analysis (Chapter 4) should have been carried out. Statistical testing is a clear case where the “Garbage In Garbage Out” principle applies.

Inappropriate test assumptions are dangerous

If the assumptions made in a statistical test are not fulfilled by the data then test results can be meaningless. For example, many statistical tests are founded on an assumption that the data being tested are normally distributed. If the data follow a strongly non-normal distribution then the test results cannot be trusted. Another common assumption that can lead to highly misleading test results if ignored is that data values are independent. Many hydrological data sets either show autocorrelation (correlation from one time value to the next: also referred to as serial correlation or temporal correlation) or spatial correlation (correlation between sites) and therefore data values are not independent. It is very important to understand what restrictions apply to a particular statistical test, and in what situations it is valid to apply the test.

A statistical test provides evidence not proof

Statistical tests give results that are expressions of probability and not certainty. There is always the chance that the null hypothesis was true when a test result suggests it should be rejected. Similarly, if the null hypothesis is accepted, then this result says only that the available evidence does not contradict the null hypothesis, it is not proof that the null hypothesis is true.

Each statistical test frames only a very specific question

There is no universal test that can prove that a series is truly free of any change. For example, a test result that shows there is no conclusive evidence of a trend in the mean does not establish that the variance of the same series is unchanged, or that frequency and magnitude of the extremes are unchanged.

Tests can be sign for the ‘wrong’ reason

Even if a test result is significant it does not prove that the hypothesised change has taken place. For example, if there has been a marked step change in a data series then it is likely that a test for trend will give significant results — even though there is no trend. Often a test can only be correctly interpreted if it is examined alongside plots of the data, and with some understanding of possible causes of change.

Sign is not the same as importance

A test result may be highly significant (i.e. provide strong evidence against the null hypothesis) but the size of the observed change may be so small that it is of no importance. Conversely, an important level of change might not be significant because noise in the data means it cannot be statistically distinguished from the null hypothesis. In such cases it is important to recognise that acceptance of the null hypothesis does not mean very much, that further information is needed, and that the question may need to be reformulated.

5.2.3 The components of testing for change

The main stages in statistical testing are

- Decide what type of series/variable to test depending on the issues of interest (e.g. monthly averages, annual maxima, deseasonalized data etc)
- Decide what types of change are of interest (trend/step-change)
- Check out data assumptions (Section 5.5)
- Select one or more tests/test statistics that are appropriate for each type of change (Using more than one is good practise: Section 5.6).
- Evaluate significance levels, using resampling methods if needed (Section 5.3, 5.4)
- Investigate and interpret results (Section 5.7).

5.3 Approaches to testing for change

5.3.1 Introduction

There are very many statistical methods that can be used to look for various types of change in a data series. This section attempts to provide a brief overview.

Two pieces of terminology are frequently used to distinguish types of test. The most useful of these is whether the test is distribution-free or distribution-dependent i.e. whether it is necessary to assume a particular distribution for the data when carrying out the test. Tests are also commonly referred to as either parametric or non-parametric. A test is said to be parametric if the change evaluated by the test can be specified in terms of one or more parameters. Linear regression is an example of a parametric test. Tests that are based on ranks (see below) are considered non-parametric because although they detect change, they do not quantify the size of change. Most non-parametric tests are also distribution-free tests.

Historically many statistical tests either assumed normally distributed data and were distribution-dependent parametric tests, or used data ranks and were distribution-free non-parametric tests. Recent developments in statistics and in the availability of computing power have increased the range of possibilities. In particular it is now possible to construct parametric tests that are distribution-free. As will be seen below, almost any parametric distribution-dependent test can be adapted and used in a distribution-free way.

For the purposes of this report, it is helpful to view the choice of a statistical test as being composed of two parts:

- Selecting the test statistic
- Selecting a method for determining the significance level of the test statistic.

By viewing the process in these two parts it becomes possible to separate out the issue of how to select a test statistic from that of how to evaluate the significance level. The resampling methods presented in Section 5.4 provide a very flexible methodology that allow significance levels to be estimated for any choice of test statistic. This means that traditional statistical tests can be adapted for application to hydrological series by extracting the test statistic but using resampling methods to determine significance.

5.3.2 Distribution-free approaches to statistical testing

The majority of hydrological series are non-normally distributed and it therefore makes sense to use distribution-free testing methods. The following approaches are distribution-free:

- *Rank-based tests*
Rank based tests are tests that use the ranks of the data values (not the actual data values). A data point has rank r if it is the r^{th} largest value in a data set.

There are a number of widely used and useful rank-based tests (see Section 5.6). Most rank-based tests assume that data are independent and identically distributed. Rank based tests have the advantage that they are robust and usually simple to use. They are usually less powerful than a parametric approach.

- **Tests using a normal scores transformation**
There are many tests for change that rely on assumptions of normality. Such tests are generally not suitable for direct use with hydrological data. However, they can be used if the data are first transformed. The normal scores transformation results in a data set that has a normal distribution. It is similar to using the ranks of a data series, but instead of replacing the data value by its rank, r , the data value is replaced by the typical value that the largest value from a sample of normal data would have (the r^{th} normal score). Thus the normal scores value for the r^{th} largest value in a series of length N is given by

$$r^{th} \text{ normal score} = \Phi^{-1} (r - 0.5)/N \quad (5.1)$$

where Φ^{-1} is the inverse of the cumulative distribution function of the normal distribution.

The advantages of using normal scores are that the original data need not follow a normal distribution, and the test is relatively robust to extreme values. Normal scores tests are likely to give slightly improved power for detection of change relative to equivalent rank-based tests.

- *Tests using resampling approaches*
Resampling methods are methods that use the data to determine the significance of a test statistic. They include methods such as permutation and bootstrapping. Resampling methods are very useful techniques for testing hydrological series and are described in detail in Section 5.4. Resampling methods can be applied to almost any test statistic and are an alternative way of obtaining significance levels. The advantages of resampling methods are that they are flexible and robust and that when used with parametric test statistics they allow the degree of change to be measured. Resampling tests are relatively powerful, e.g. for large samples, permutation tests can be shown to be as powerful as the most powerful parametric tests (Bickel & Van Zwet, 1978). Furthermore, resampling methods can be adapted to test data which are not independent (see Section 5.5).

5.3.3 Other approaches to testing for change

There are many other ways of testing for change. These are beyond the scope of this chapter, but are mentioned for completeness. One group of approaches requires the specification of a distributional form and of a model for change. These are therefore useful techniques for modelling complex types of change. They include

- **Maximum likelihood estimation**
Maximum likelihood estimation is a very powerful testing approach. However, formulating and solving the maximum likelihood equation(s) is often non-trivial and can require considerable expertise.
- **Bayesian methods**
Bayesian methods have a number of attractive features e.g. they provide a measure of the uncertainty of an estimate of change. However, they can be complex to apply and require distributional assumptions to be made (Lee, 1997).

- **Time series methods**
Time series methods are potentially useful because they build in an autocorrelation structure to the data (i.e. a link between the current observation and successive observations; cf. Chatfield, 1996). However, most time series methods often require elimination of any trend component before an autocorrelation structure is modelled and they are often reliant on knowing the distribution of the data.

Some other approaches that involve less complete modelling include the following:

- *Data generation methods*
Data generation methods work by producing a large number of artificial data series and using these to evaluate significance levels. Resampling is one type of data generation technique. An alternative advanced data generation method is the “phase randomisation”, technique which is described in Chapter 12. This is a method of generating data series that preserves the autocorrelation structure of the data.
- *Smoothing methods*
Smoothing methods include methods such as locally weighted regression (lowess). These techniques are often used informally as part of a visual data analysis (see also Section 4.3).
- *Methods for looking at changes in variance and correlation (persistence)*
Most test procedures assume that the variance of the data remains constant. Some methods for testing for changes in variance and persistence are described in Chapter 9 and 13.
- *Segmentation*
Segmentation is a Bayesian technique that looks for multiple step changes in a series. It is described in Chapter 10.

5.4 Statistical testing by resampling methods

Resampling methods such as permutation testing or the bootstrap are robust methods for estimating the significance level of a test statistic. A useful practical text on resampling methods and permutation tests is provided by Good (1993). Efron & Tibshirani (1998) and Davidson & Hinkley (1997) describe bootstrapping methods. Resampling methods are very useful for testing hydrological data because they require relatively few assumptions to be made about the data, yet they are also quite powerful tests.

5.4.1 Understanding resampling

The basic idea behind resampling methods is very straightforward. Consider testing a series for trend: a possible test is the regression gradient. If there is no trend in the data (the null hypothesis) then the order of the data values should make little difference. Thus shuffling (permuting) the data series should not change the gradient very much. Under a permutation approach the data are shuffled very many times. After each shuffle (permutation) the test statistic is recalculated. After very many permutations, the original test statistic is compared to the generated test statistic values. If the original test statistic is rather different from most of the generated values then this suggests that the ordering of the data affects the gradient and thus that there was trend. If the original test statistic lies somewhere in the middle of the generated values then it seems reasonable that the null hypothesis was correct (the order of

the values does not matter, so there is no evidence of trend). In other words, if an observer (or in this case, the statistical test) can distinguish between the original data and the resampled (permuted) data, then the observed data are judged not to satisfy the null hypothesis.

5.4.2 Permutation and the bootstrap

The bootstrap and permutation methods are two slightly different approaches to resampling the data. In permutation methods the data are reordered, each of the data points in the original data series appearing once in each resampled (generated) data series. In bootstrap methods, the original data series is sampled with replacement to give a new series with the same number of values as the original data. With this method, the generated series may contain more than one of some values in the original series and none of other values. In both cases, the generated series has the same distribution as the empirical (i.e., observed) distribution of the data.

The bootstrap is generally but not always, less powerful than a permutation test (Good, 1993). However, bootstrap methods are often to be preferred where a test is looking for change in variance. Further, permutation tests cannot be applied with test statistics that do not change when the data are permuted, e.g. tests for which the test statistic is the mean or median. The tests given here can be used with either method. In general, bootstrap methods are more flexible than permutation methods and can be used in a wider range of circumstances.

5.4.3 Determining significance level

To determine the significance level, the data are resampled, by either permutation or bootstrapping, a large number of times, S . For each of these generated series, the test statistic, T , is calculated to give S artificial values of T . These are then ordered as

$$T_1 \leq T_2 \leq \dots \leq T_S \quad (5.2)$$

If the original test statistic is T_0 and

$$T_k \leq T_0 \leq T_{k+1} \quad (5.3)$$

then the probability of the test statistic being less than or equal to T_0 under the null hypothesis is approximately

$$p = \frac{k}{S} \quad (5.4)$$

p may also be estimated as

$$p = (k+0.5)/(S+1) \quad (5.5)$$

or even

$$p = (k+1)/(S+2) \quad (5.6)$$

Assuming that large values of T indicate departure from the null hypothesis, the significance level for this test is then

$$100 * 2 \min(p, 1-p)\% \quad (5.7)$$

(assuming a two-sided test, i.e., a test in which the direction of change is assumed unknown).

5.4.4 Number of resamples

The number of samples that need to be generated depends on the level of significance required and on the degree of change seen in the data. Typically around 100 to 2000 samples might be generated. The larger the number of resamples, the more accurate the estimate of significance. More resamples will be required to accurately determine significance levels of 1% than significance levels of 10%. A simple approach to check whether the sample size is sufficient is to rerun a test a few times and check that the required percentiles of the generated test statistic values are not varying too much.

For permutation testing, all permutations could, theoretically, be evaluated. However, typically there are too many to be evaluated (for a series of length n there are $n!$ permutations) and a random selection of possible permutations is used instead.

Note that if confidence intervals are required then sample sizes of 199, 999, 1999 etc. give exact confidence intervals (Faulkner & Jones, 1999 Appendix 2) e.g. for 199 samples, the 95% confidence interval is given by the 5 largest and smallest values; for 1999 samples, the 95% confidence interval is given by the 50 largest and smallest values.

5.4.5 Resampling when data are not independent

The above resampling methods are applicable only in the case where it can be assumed that the data are independent. For series with serial dependency, or series with seasonal structure, different techniques should be used.

If data show serial correlation, or additional structure such as seasonality, then it is necessary that the generated series should replicate this structure. A straightforward means of achieving this is to permute, or bootstrap the data in blocks. For example, for a 40 year series of monthly values, it would be sensible to treat the data as consisting of 40 blocks of one year. Each year's worth of data is left intact and is moved around together as a block — thus maintaining the seasonal and temporal dependencies within each year. The 40 blocks are then reordered many times. The resampled series would then preserve the original seasonality and serial correlation seen in the data. It is important that the size of the blocks should be sensibly selected. If there is seasonality then the block should contain an integral number of seasonal cycles. If there is autocorrelation then the block should be chosen so that data points one block apart are approximately independent. For the block-bootstrap there are also other more sophisticated methods of resampling such as the wild block bootstrap (Shao & Tu, 1995). Section 5.5 discusses issues related to dependence more generally and points out other alternative approaches.

Note that block-bootstrap and block-permutation methods can be used with rank-based tests, which, although distribution-free, would otherwise depend on assumptions of independence.

Note also that blocking methods can be useful when there is spatial dependency in a set of multi-site data that is to be tested as a group. In this case, the usual choice of blocks would be to group data across all sites that occurred in the same time interval. Experience suggests that allowance for multi-site dependencies can be very important for estimated significance levels (e.g. Robson *et al.*, 1998).

5.4.6 Summary of method for resampling

The basic method for carrying out a permutation or bootstrap test is as follows

- Select one or more suitable test statistics (see Section 5.6)
- Calculate the test statistic for the observed data, Resample the data series many times (e.g. 1000) to generate new data series and recalculate the test statistic for each of these series using blocking methods if appropriate
- Estimate the significance levels using Section 5.4.3.

5.5 Understanding Test Assumptions

5.5.1 Importance

It is very important to check that test assumptions are approximately true when testing data for change. The violation of a test assumption can result in surprisingly inaccurate significance levels. For example, data that is assumed to be independent when it is not, could result in a significance level of 5% when in reality it should only be 25%.

In general the assumptions that are made in a statistical test are linked to the method used to estimate significance levels, and not to the test statistic. Most test statistics can be tested using resampling approaches in order to minimise assumptions.

5.5.2 Types of assumption

Three types of assumption are commonly made when carrying out statistical tests:

The form of the distribution (e.g. normally distributed)

Not all tests make assumptions about the underlying distribution of the data. Tests that avoid assuming a distribution are called distribution-free (Section 5.3).

Constancy of the distribution (i.e. all data points have an identical distribution)

Most basic statistical tests assume, under the null hypothesis that the distribution of the data does not change. This assumption is violated if there are seasonal variations or any other cycles in the data, or if there is an alteration over time in the variance or any other feature of the data that is not part of the test.

If there are seasonal cycles in the data, then the options are either to (1) deseasonalize the data, i.e. estimate the seasonal structure and remove this from the data series, or (2) to use a testing approach that allows for seasonality. Possible approaches that allow for seasonality include

- use of block bootstrap and block permutation methods (see Section 5.4)
- use of maximum likelihood / time series methods (see Section 5.3)
- seasonal Kendall test (see Appendix C).

Section 4.4 looks at methods of investigating and handling seasonality.

If the variance or some other feature of the data is changing over time then the problem of testing for change becomes a much more complex one. Some specialist methods for tackling this type of problems are described in Chapters 9 and 13.

Independence

The assumption of independence is frequently violated by hydrological series and can have a very big effect on estimated significance levels.

Data values can be said to be independent if they are completely unrelated to one another. For many hydrological series, this is not the case: e.g. knowing the flow in the river today, tells one quite a bit about what tomorrow's flow is likely to be — so these data values are dependent. However, knowing today's flow does not usually say very much about what the flow will be in a year's time — thus these values are independent. When successive values show dependency this is known as autocorrelation, serial correlation or temporal dependency.

When there are data from more than one site, assumptions of independence may also be violated because of spatial correlation (e.g. the flow is high in one catchment, it is also likely to be high in an adjacent catchment). This is also sometimes referred to as spatial dependency.

The more frequent the data points the more likely it is that there will be important serial correlation in the data. As a rough guideline, daily hydrological series are usually

strongly correlated, annual series are often approximately independent, and monthly values are intermediate.

Most common statistical tests do not allow for serial correlation in the data. If serial correlation is present then possible options are

- Use block permutation or block bootstrap methods (Section 5.4)
- Decrease the frequency of the data series (e.g., by calculating monthly or annual averages)
- Use time series/phase randomisation/multivariate methods that build in serial correlation
- For data with seasonality, consider the modified seasonal Kendall test (Appendix C).

5.5.3 Methods for checking assumptions

The principal method for assumption-checking is to use visual techniques such as are described in Section 4.5. They include

- Histograms and normal probability plots — to examine distribution
- Time series plots - to spot time dependent patterns or possibly changes in variance
- Autocorrelation plots.

Assumption checking may need to be carried out both prior to and after application of tests. For example, if a trend is detected, then the trend should be estimated and removed from the data, and the residuals checked for autocorrelation and for constancy of distribution.

Use of visual methods for assumption checking will usually be sufficient, however formal tests are also available for checking some assumptions, e.g. tests for normality of data and tests for data independence (e.g. Bartlett's test, Appendix F).

5.6 Choosing tests and test statistics

5.6.1 Introductory notes

This section provides summaries of common tests for change. These are intended to be used as a source of test statistics for use within resampling methods. Further details on most of these tests are given in Appendices A-F.

5.6.2 Choosing which test statistics to use

Sections 5.6.5-5.6.7 list a number of common tests that can be used to test for change. For most studies, it is recommended that more than one of these tests should be used. Criteria to be aware of when selecting test statistics are

- Type of change that is of interest
- Power of test — more powerful tests are to be preferred.
- Different types of test — some tests are very similar to one another and it is best to choose a selection of tests that are not too similar
- Whether test is for a known or unknown change-point time (see Section 5.6.4).

If a resampling technique is to be used, it is possible to either construct a test statistic to test for a particular type of change, or to extract a suitable test statistic from almost any other test for change.

5.6.3 Choosing how to evaluate significance levels

The tests given in Section 5.6.5 to 5.6.7 fall into two main groups

- Rank-based tests - inherently distribution-free
- Tests that assume a normal distribution.

Tests assuming a normal distribution are provided with the intention that sign levels should be evaluated using resampling methods. For this, the test statistic is calculated as usual, but the significance level is obtained by resampling, rather than by referring to tabulated values. An alternative would be to use the normal scores transformation (Section 5.3).

Rank-based tests can be applied directly to the data, providing the data meet assumptions of independence and constancy of distribution.

If data violate independence assumptions, then it is recommended that block resampling methods are used to obtain significance levels (for both of the above types of test). Note that all tests assume that, under the null hypothesis, the distribution of data values does not change with either time or space (Section 5.5.2).

5.6.4 Testing when the time of change is unknown

Tests for step-change and tests for a change in distribution work by dividing the data into two parts and comparing the parts. In the tests listed below, some assume a known time of change others assume an unknown time of change.

In most cases, the time of change is unknown and tests assuming an unknown time of change are preferable. It is also possible to adapt a test that assumes a known time of change. A recommended method is to replace the original test statistic by two new test statistics. These are (1) the maximum value of the original test statistic when it is calculated for all possible change-point times, (2) the time at which this maximum occurs. These modified test statistics can then be evaluated using resampling methods.

5.6.5 Tests for step change

1. *Median change point test / Pettitt's test for change.* This is a rank-based test for a change in the median of a series with the exact time of change unknown (Siegel & Castellan, 1988; Pettitt, 1979). The test is considered to be robust to changes in distributional form and powerful relative to tests such as the Wilcoxon-Mann-Whitney test (see below).

2. *Wilcoxon-Mann-Whitney test / Mann-Whitney test / Mann test / Rank-sum test.* This test is a rank based test that looks for differences between two independent sample groups (Siegel & Castellan, 1988; WMO, 1988; Helsel & Hirsch, 1992). It is based on the Mann-Kendall test statistic - see below, but is calculated for subsets of the series in order to detect the point of change in the mean (Chiew & McMahon, 1993). In its basic form it assumes that the time of change is known. When the time of change is unknown, use of the median change-point test is recommended.

3. *Distribution-free CUSUM test.* This is a rank-based test in which successive observations are compared with the median of the series (Chiew & McMahon, 1993; McGilchrist & Woodyer, 1975). The test statistic is the maximum cumulative sum (CUSUM) of the signs of the difference from the median (i.e. the CUSUM of a series of plus or minus ones) starting from the beginning of the series.

4. *The Kruskal-Wallis test.* The Kruskal-Wallis test (Sneyers, 1975) is a rank-based test for equality of sub-period means. It can also be used to test for equality of sub-period variability.

5. *Cumulative deviations and other CUSUM tests.* The cumulative deviation test (Buishand, 1982) is based on the rescaled cumulative sum of the deviations from the mean. The test is relatively powerful in comparison with other tests (e.g. Worsley likelihood ratio test; Buishand, 1982) for a change-point that occurs towards the centre of the time series. The

basic test assumes normally distributed data. Other CUSUM based tests (using Bayesian and likelihood methods) are described in Buishand (1984).

6. *Student's t-test*. This is a standard parametric test for testing whether two samples have different means. In its basic form it assumes normally distributed data and a known change-point time.

7. *The Worsley likelihood ratio test*. The Worsley likelihood ratio test (Worsley, 1979) is similar to Student's t-test but is suitable for use when the change-point time is unknown. It assumes normality.

5.6.6 Tests for trend

1. *Spearman's rho*. This is a rank-based test for correlation between two variables that can be used to test for a correlation between time and the data series (Siegel & Castellan, 1988). Spearman's correlation is a rank-based version of the usual parametric measure of correlation (the Pearson product moment; Sprent, 1989).

2. *Kendall's tau / Mann-Kendall test*. This is another rank-based test which is similar to Spearman's rho (same power and still based on ranks) but using a different measure of correlation which has no parametric analogue.

3. *Seasonal Kendall test*. The seasonal Kendall test is a version of the Mann-Kendall test that allows for seasonality in the data (Hirsch *et al.*, 1982). There is also a modified seasonal Kendall test that additionally allows for some autocorrelation in the data (Hirsch & Slack 1984).

4. *Linear regression*. The test statistic for linear regression is the regression gradient. This is one of the most common tests for trend and in its basic form assumes that data is normally distributed.

5. *Other robust regression tests*. There are a number of robust methods for estimating trend in series. These could potentially be used as alternative measures of the change. For example, in least absolute deviation regression, the gradient is that which minimises the sums of the deviations of the points from the fitted line (Bloomfield & Steiger, 1983). Other robust means of estimating the rate of change include M-estimates of regression and trimmed regression (Rousseeuw & Leroy, 1987).

5.6.7 Tests for a change in distribution

1. *Kolmogorov-Smirnov test*. This test can be used to decide whether two samples have the same distribution. it is a distribution free approach. In its basic form it assumes that the time of change is known. The test statistic is based on the maximum difference between the distributions of the data before and after the change-point.

2. *Cramer-von-Mises test*. This is a second distribution-free statistic which is similar to the Kolmogorov-Smirnov test, except that it uses a different way of measuring the difference between distributions.

5.7 Interpreting results

5.7.1 Understanding test results

No statistical test is perfect, even if all test assumptions are met. A 5% significance level means that we will make an error 5% of the time: i.e. if the null hypothesis was true then 1 in 20 test results will have a significant (and incorrect) result. It is important to remember this when interpreting results.

Often, there will be many test results to be examined. For example, a typical approach to testing for change is to apply a selection of tests e.g. a few tests for trend and a few tests for step-change. Interpretation of multiple test results can be complex. In some cases the results from these tests will generally be in agreement, but in other cases there will be differences between the tests. Some suggestions for interpretation of such results are

- Use visualisation methods. Where there are multiple test results per site, it is easier to interpret tables of results if the tables show symbols rather than figures. For example, use of circles with circle size indicating significance level (3 to 5 classification bands are recommended) and circle colour indicating direction of change can be effective (Robson & Reed, 1999). If there are multiple sites, then plotting significance levels or size of change on a geographical map is helpful (see also Chapter 4)
- Take care in interpreting significance levels. The presence of a single significant test result may only be weak evidence of change — even if this test is highly significant. If many of the tests are significant then this provides stronger evidence of change. However, if tests are very similar then multiple significant values are not an extra proof of change.
- Examine the test results alongside graphs of the data, and with as much historical knowledge about the data as possible. For example, if both step-change and trend results are significant, and historical investigations reveal that a dam was built during the period, and this is consistent with the time series plot, then a reasonable conclusion is that the dam caused a step change.
- Look out for patterns in the results that may indicate further structure e.g. regional patterns in trends. These may suggest that further investigation is needed.

When many tests are applied to a single series it is usually only possible to draw qualitative conclusions (an overall significance level is not appropriate) - this is usually sufficient for most purposes.

5.7.2 Interpreting change

If test results suggest that there is a significant change in a data series, then it is important to try to understand the cause. Although the investigator may be interested in detecting climate change, there may be many other possible explanations.

Causes of change

Common causes of change include:

- Changes directly caused by man (urbanisation, reservoirs, drainage systems, water abstraction, land-use change etc).
- Natural catchment changes (e.g. changes in channel morphology)
- Climate variability
- Climate change
- Problems linked to data.
-

Examples of problems linked to the data that can cause apparent change in a data series are:

- Typographical errors

- Instruments malfunctioning (zero-drift, bias)
- Change in measurement techniques/instrumentation/ instrument location
- Change in accuracy of data / changes of data units
- Changes in data conversions (e.g. altered rating equations).

Data should always have been quality controlled before starting an analysis, but even with good quality control, some problems may be missed and it is helpful to be open minded at any stage of an analysis.

Gathering additional information

The best way to improve understanding of change is to gather as much information as possible. Examples include:

- Historical information about changes in the catchment, land-use change etc.
- Historical information about data collection methods etc.
- Data from nearby sites — if data from other nearby sites show similar patterns then the cause is probably widespread (e.g. linked to climate, or to extensive land-use change).
- Related variables — information on temperature and rainfall can help determine whether changes in flow can be explained by climatic factors.
- Data that extends record lengths — a primary problem with many hydrological records is that they are too short. If related data can be obtained that extends to a longer period then this may be of assistance.

Climate variability and climate change

It is very important when interpreting test results to understand the difference between climate change and climate variability (see also Chapter 1: Appendix). Climate variability is the natural variation in the climate from one period to the next. Climate change refers to a long-term alteration in the climate.

Climate variability appears to have a very marked effect on many hydrological series. This has two important effects.

- *Climate variability can cause apparent trend*
Climate variability can easily give rise to apparent trend when records are short - these are trends which would be expected to disappear once more data had been collected. Some examples of this can be seen in Figure 4.3. Because of climate variability, records of 30 years or less are almost certainly too short for detection of climate change. it is suggested that *at least 50 years of record is necessary* for climate change detection.
- *Climate variability obscures other changes*
Because climate variability is typically large, it can effectively obscure any underlying changes either due to climate change or to anthropogenic causes, such as urbanisation.

Advice on choosing suitable series for study of climate change is given in Chapter 11.

5.8 Summary

This chapter gives recommendations for statistical analysis with the aim of providing a relatively straightforward approach to testing for change that is likely to yield reliable and useful results in a wide variety of situations. These recommendations are not the only approach to testing, but they should provide a good starting point in most cases.

It is recommended that hydrological data should be tested using distribution-free methods, i.e. using methods that do not rely on assumptions being made about the distribution of data. This is because hydrological data are often non-normal. Use of distribution-dependent methods is possible, but will generally require more advanced statistical treatment.

The simplest distribution-free methods are rank based methods (Section 5.3). A related approach is to apply a normal scores transformation to the data. This transformation results in a series that is normally distributed and to which tests suitable for normal data can then be applied (Section 5.3). Both these approaches are rapid and straightforward, but they are only suitable if the data series are independent (Section 5.5).

Resampling techniques, such as permutation and bootstrapping allow a very flexible approach to distribution-free testing. They are powerful tests and can be used to test a very wide variety of statistical hypotheses (Section 5.4). Resampling methods have the advantage that they provide a robust means of estimating significance levels for almost any test statistic without the need for distributional assumptions. Resampling techniques can also often be used even when the data are not independent, by making use of approaches such as block-permutation or the block-bootstrap. A variety of test statistics for use with resampling approaches are described in Section 5.6 and the Appendices.

Care is always needed when interpreting test results. There are many causes of change other than climate change, and it is often the case that climate variability can obscure other possible changes in the data. Understanding the catchment and the data is a prerequisite for sensible interpretation of results.

References

- Bickel P.M. & Van Zwet, W.R. 1978. Asymptotic expansion for the power of distribution-free tests in the two-sample problem. *Annal. Statist.* 6, 987-1004 (corr. 1170-1171).
- Bloomfield, P. & Steiger, W.L., 1983. *Least Absolute Deviations: Theory, Applications and Algorithms*. Birkhauser, Boston, Mass.
- Buishand, T.A., 1982. Some methods for testing the homogeneity of rainfall records. *J. Hydrolo.*, 58, 11-27.
- Buishand, T.A., 1984. Tests for detecting a shift in the mean of hydrological time series. *J. Hydrol.*, 73,51-69.
- Cavadias, G., 1992. A study of current approaches to modelling of hydrological time-series with respect to climatic variability and change. World Climate Programme-Water, Project A2. WCASP -23 WMOITD - No 534.
- Chatfield, C., 1970. *Statistics for Technology*. Penguin.
- Chatfield, C, 1996. *The Analysis of Time Series: An Introduction*. 5th Ed. Chapman and Hall, 283pp.
- Chiew, F.H.S. & McMahon, TA, 1993. Detection of trend or change in annual flow of Australian rivers. *Int. J. of Climatology.*, 13,643-653.

- Davidson A.C. & Hinkley D.V., 1997. *Bootstrap Methods and Their Application*. Cambridge University Press, 582pp.
- Efron B. & Tibshirani, R.J., 1998. *An Introduction to the Bootstrap*. Chapman and Hall. 436pp.
- Faulkner, D.S. & Jones D.A., 1999. The FORGEX method of rainfall growth estimation III: Examples and confidence intervals. *Hydrology and Earth System Sci.* 3(2), 205-212.
- Gilbert, R.O., 1987. *Statistical Methods for Environmental Monitoring*. Van Nostrand Reinhold, N.York 320pp.
- Good, P., 1993. *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. Springer—Verlag. 228 pp.
- Helsel, D.R. & Hirsch, R.M., 1992. *Statistical Methods in Water Resources*. Studies in environmental science 49. Elsevier. 529pp.
- Hirsch, R.M., Alexander, R.B. & Smith, R.A., 1991. Selection of methods for the detection and estimation of trends in water quality. *Water Resour. Res.*, 27,5,803-813.
- Hirsch, R.M., Helsel, D.R., Cohn, T.A. & Gilroy, E.J., 1992, Statistical analysis of hydrologic data. Chapter 17, In: *Handbook of Hydrology* ed. by D.R. Maidment, McGraw-Hill, New York
- Hirsch, R.M. & Slack, J.R., 1984. A non parametric test for seasonal data with seasonal dependence. *Water Resour. Res.*, 20,727-732.
- Hirsch, R.M., Slack, J.R. & Smith R.A., 1982. Techniques of trend analysis for monthly water quality data. *Water Resour. Res.*, 18, 107-121.
- Kendall, M.G., 1975. *Rank Correlation Methods*. London: Griffin.
- Kim, P.J. & Jennrich, R.I., 1973. Tables of the exact sampling distribution of the two sample Kolmogorov-Smirnov criteria. In: *Selected Tables in Mathematical Statistics*, Volume 1. ed. By H.L. Harper & D.B. Owen, Providence, Rhode Island: American Mathematical Society.
- Lee, P.M., 1997. *Bayesian Statistics. An Introduction* (2 Ed). Arnold, 344pp.
- Lehmann, E.L., 1975. *Nonparametrics: Statistical Methods Based on Ranks*. McGraw-Hill, 457
- Maritz, J.S., 1981. *Distribution Free Statistical Methods*. Chapman and Hall, 1-264.
- McGilchrist, C.A. & Woodyer. K.D., 1975. Note on a distribution-free CUSUM technique. *Technometrics*, 17(3), 321-325.
- Mendenhall, W., 1983. *Introduction to Probability and Statistics*, Sixth Edition, PWS Publishers, Duxbury Press Boston, Mass.

- Pettitt, A.N., 1979. A non-parametric approach to the change point problem. *Applied Statistics*, 28, 126-135.
- Robson, A.J., Jones, T.A. & Reed, D.W., 1998: A study of national trend and variation in UK floods. *Int. J. Climatology*, 18, 165-182.
- Robson, A.J. & Reed D.W., 1999. *Statistical Procedures for Flood Frequency Estimation. Flood Estimation Handbook*, Volume 3. Institute for Hydrology, Wallingford, UK 388pp.
- Rousseeuw, P.J. & Leroy, A.M., 1987. *Robust Regression And Outlier Detection*. Wiley, New York
- Siegel, S. & Castellan, N.J., 1988. *Non Parametric Statistics for The Behavioural Sciences*. McGraw -Hill, 399pp.
- Shao, J. & Tu, D., 1995. *The Jackknife and Bootstrap*. Springer, 516pp.
- Sneyers, R., 1975. Sur l'analyse statistique des series d'observations, Technical Note No. 143, World Meteorological organization, Geneva, 192pp.
- Sprent, P., 1989. *Applied Non-Parametric Statistical Methods*. Chapman and Hall, 259pp.
- Srikanthan, R., McMahon, T.A. & Irish, J.L., 1983. Time series analyses annual flows of Australian streams, *J. of Hydrol.* **66**, 213-226.
- Srikanthan, R. & Stewart B.J., 1991. Analysis of maximum daily rainfall with regard to climatic change. International Hydrology and Water Resources Symposium, Perth 2-4 October 1991.
- Statistical Sciences*, 1995. Splus Verion 3.3 supplement, Seattle: StatSci, A Division of MathSoft, Inc. Chapter 3.
- World Meteorological Organisation, 1988. Analysing Long Time Series of Hydrological Data with respect to climate variability and change, WCAP-3, WMO/TD-No 224, 12pp.
- World Meteorological Organisation, 1966. Climate Change. Technical Note No 79, WMO No 195, TP100.
- Worsley, K.J., 1979. On the likelihood ratio test for a shift in location of normal populations. *J. Am. Stat. Assoc.*, **74**, 365-376.

Appendix A

NOTES FOR APPENDICES B-E

Appendices B-F contain details of various statistical test procedures. For each test, the text describes the *type of test* i.e., whether rank-based or making normal assumptions. It then gives details on the required *test-statistic*. Finally, details of how to determine significance levels under the stated test assumptions are given. Note that unless otherwise mentioned, *all tests make the assumption that the data are independent and that the distribution of the data is constant*. If the data meet all test assumptions then the method given under *determining significance levels* can be used to obtain significance levels. Otherwise a resampling method will usually be more appropriate.

Where tests are applied using resampling methods then only the test statistic is used since significance levels are obtained by resampling.

Notation List

- CS — cumulative sum
- d_f — number of degrees of freedom
- H_0 — null hypothesis
- N — number of observations n number of runs
- R — rank
- r_1 — lag-one autocorrelation
- S, S_x, S_y — sample standard deviation
- S_{xy} — sample covariance
- sgn — sign function (for definition see eq. 5B.2)
- T — number of observations
- t_j — number of ties of extent j
- $t_\alpha, t_{\alpha/2}$ — critical value of test statistic
- Var — variance
- x, y, z — random variable
- $\bar{x}, \bar{y}, \bar{z}$ — sample mean
- Z_m — sample median
- α — significance level A
- μ — mean value
- ρ — correlation coefficient
- ρ_s — sample correlation coefficient
- σ — standard deviation

Other tests

Further listings of tests for changes in hydrological data can be found in Cavadias (1992), WMO (1966), WMO (1988), Srikanthan *et al.* (1983) and Srikanthan & Stewart (1991).

Appendix B

TESTS FOR STEP-CHANGE

See also the explanatory notes in Appendix A.

Median change point test / Pettitt test for a change:

Type of test: rank-based and distribution-free for an unknown time of change.

Test statistic: For a series of T observations, the test statistic is defined as:

$$T_s = \max_{1 \leq t \leq T} |U_{t,T}| \quad (5B.1)$$

where U is defined as:

$$U_{t,T} = \sum_{i=1}^t \sum_{j=t+1}^T \text{sgn}(z(i) - z(j)) \quad (5B.2)$$

and sgn denotes the sgn function (1 for positive, 0 for zero and -1 for negative arguments).

Determining significance levels: For a significance level, α , the null hypothesis is rejected if

$$\alpha > \exp\left(\frac{-6K_T^2}{T^3 + T_2}\right) \quad (5B.3)$$

This is based on formulae for the probability of K_T based on Bernoulli experiments.

Comments: The change point time can be estimated as the time t when the maximum K_T occurs. Modified versions of this test are also available. For example the test statistic

$$K_T^* = \max \left| \frac{U_{t,T}}{\sqrt{Tt - t^2}} \right| \quad (5B.4)$$

is an alternative indicator for the change point. For this test statistic significance levels should be estimated using resampling methods.

Rank-sum test:

The rank-sum test is also known as the Wilcoxon-Mann-Whitney or Mann-Whitney test.

Type of test: This test is rank-based and distribution-free. It evaluates whether two independent data groups are different: the null hypothesis H is that the medians of the two groups are equal (under the assumption of identical distribution of the two populations). In the context of testing for step change, the time of change is assumed known and the series is divided into two groups (before and after the change point time) and these groups are compared. The method can be adapted for an unknown time of change as described in Section 5.6.4.

Test statistic: To compute the rank-sum test statistic (Hirsch et al., 1992):

- (i) Assign ranks to all the data, from 1 (smallest) to N(largest). In the case of ties (equal data values) use the average of ranks.
- (ii) Split the data into two groups of size m and n. Compute a test statistic S as the sum of ranks of the n observations in the smaller group.
- (iii) Compute the theoretical mean and standard deviation of S under H for the entire sample:

$$\mu = n (N+1) / 2 \quad (5B.5)$$

$$\sigma = [n m (N+1) / 12]^{1/2} \quad (5B.6)$$

The standardised form of the test statistic Z_{rs} is computed as:

$$Z_{rs} = \begin{cases} (S - 0.5 - \mu) / \sigma & \text{if } S > \mu \\ 0 & \text{if } S = \mu \\ (S + 0.5 - \mu) / \sigma & \text{if } S < \mu \end{cases} \quad (5B.7)$$

Determining sign levels: Z_{rs} is approximately normally distributed and so significance levels can be obtained from normal probability tables. Thus, for a significance level of α , reject H_0 if $|Z_{rs}| > Z_{1-\alpha/2}$ where $Z_{1-\alpha/2}$ is the 1- $\alpha/2$ point of the standard normal probability distribution.

Comments: A correction for data with ties is given in Hirsch et al. (1992) and Siegel & Castellan (1988).

Distribution-free CUSUM test:

Type of test: This is a rank-based test in which elements of the time series are compared with the median. it is distribution-free and for an unknown time of change (McGilchrist & Woodyer, 1975).

Test statistic: The test statistic is the cumulative sum of the signs of the difference from the median (i.e. plus ones for values greater than the median and minus ones for values less than the median):

$$TS = \frac{2}{N} \text{Max} |CS_k| \quad k = 1, \dots, N \quad (5B.8)$$

where

$$CS_k = \sum_{i=1}^k \text{sgn}(z(i) - Z_m) \quad (5B.9)$$

and Z_m is the sample median.

Determining significance levels: The test statistic is equivalent to the Kolmogorov-Smirnov test for the equality of distribution of the following two random variables:

- (i) times at which observations greater than the median occur
- (ii) times at which observations less than the median occur

and the standard algorithms for the Kolmogorov-Smirnov test can be used for determination of the percentage points of the test statistic (cf. Chiew & McMahon, 1993).

Kruskal-Wallis test:

Type of test: rank-based distribution-free test for equality of sub-periods. In the context of testing for step-change, it assumes a known change-point time.

Test statistic: Let N be the number of data in the time series and let the series be sub-divided into m sub-periods of length n_j ($j = 1, 2, \dots, m$). Let R_{ij} be arank of the i^{th} observation in the j^{th} sub-period and

$$R_j = \sum_{i=1}^{n_j} R_{ij} \quad (5B.10)$$

The test statistic is:

$$TS = 12 \sum (R_j^2 / n_j) / [N(N+1)] - 3(N+1) \quad (5B.11)$$

Determining significance levels: Under the null hypothesis of equal sub-period means, this statistic follows the Chi-square distribution with $(m-1)$ degrees of freedom.

Comments: In the case of t_j ties in the j^{th} sub-period the value of TS should be divided by:

$$1 - \left[\sum_{j=1}^m (t_j^3 - t_j) \right] / (N^3 - N) \quad (5B.12)$$

The Kruskal-Wallis test can also be applied, using the same formulae, to detect differences in the variances of the sub-periods, if ranks of the quantities $|X_i - \bar{X}|$ are used where \bar{X} is the mean for the complete period. See Sneyers (1975).

Cumulative deviations test:

Type of test: This test assumes that data are normally distributed.

Test statistic: The test statistic is defined as:

$$Q = \max_k |CS_k^*| \quad (5B.13)$$

where

$$CS_k^* = \frac{CS_k}{\sqrt{\frac{1}{T} \sum_{t=1}^T (z(t) - \bar{z})^2}} \quad (5B.14)$$

and CS_k is the cumulative sum

$$CS_k = \sum_{t=1}^k (z(t) - \bar{z}) \quad \text{for } k = 0, \dots, T \quad (5B.15)$$

Determining significance levels: Critical values for this test were found by Buishand (1982).

Student's t-test:

Type of test: This test assumes normally distributed data. The null hypothesis (H_0) is that the means of two independent groups of data are equal. The alternative hypotheses are either that the means are non-equal (two-tailed test) or that one mean is higher than another (one-tailed test). In the context of step-change, the t-test corresponds to assuming a known time of change and testing whether the mean shifted at this point.

Test statistic: If the variances of the two groups are assumed equal then the test statistic is given by:

$$t = \frac{(\bar{x} - \bar{y})}{S \sqrt{\frac{1}{n} + \frac{1}{m}}} \quad (5B.16)$$

where \bar{x} and \bar{y} are the sample means for two groups of data, where n and m denote the number of observations in each group, and S is the sample standard deviation (assumed equal for the two groups).

Determining significance level: The test statistic follows a t-distribution. The rejection region is

$$t > t_\alpha \quad \text{for one-tailed test} \quad (5B.17)$$

and

$$|t| > |t_{\alpha/2}| \quad \text{for two-tailed test} \quad (5B.18)$$

The critical values of t , t_α and $|t_{\alpha/2}|$ can be found in statistical textbooks (e.g., Mendenhall, 1983), for $(n+m-2)$ degrees of freedom.

Comments: There is an alternative version of the t-test in which the two samples are not assumed to have equal variances. This test can be found in standard statistical texts.

Worsley likelihood ratio test:

Type of test: This test assumes that the data are normally distributed and that the change-point time is unknown. It tests for a step-change in a series.

Test statistic: The test statistic is

$$W = \frac{V \sqrt{T-2}}{\sqrt{1-V^2}} \quad (5B.19)$$

where

$$V = \max |S_k^*| \quad (5B.20)$$

and

$$S_k^* = \frac{\sqrt{k(T-k)}CS_k}{S_z} \quad (5B.21)$$

and CS_k is the cumulative sum, as defined for the cumulative deviations test above.

Determining significance level: Critical values for different significance levels for this test have been derived by Worsley (1979).

Appendix C

TESTS FOR TREND

See also the explanatory notes in Appendix A.

Spearman's rho:

Type of test: Spearman's ρ (rho) is a rank-based test which determines whether the correlation between two variables is significant. The null hypothesis is that there is no association between the rank pairs (Mendenhall, 1983). It can be tested in a two-tailed (unknown direction of change) or a one-tailed-mode (known direction of change).

Test statistic: The test statistic is the correlation coefficient which is obtained in the same way as the usual sample correlation coefficient but using ranks:

$$\rho_s = S_{xy} / \sqrt{S_x S_y} \quad (5C.1)$$

where

$$S_x = \sum_{i=1}^n (x_i - \bar{x})^2 \quad (5C.2)$$

$$S_y = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (5C.3)$$

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (5C.4)$$

and x_i , \bar{x} , y_i , \bar{y} refer to ranks.

If there are no ties in the data then there are simpler expressions that can be used to obtain rho (Mendenhall, 1983; Siegel & Castellan, 1988).

Determining significance levels: For small values of N , the significance level of the test statistic can be looked up in tables (e.g., Mendenhall, 1983; Siegel & Castellan, 1988). For samples with more than 20 values, the quantity $\rho\sqrt{N-1}$ is approximately normally distributed (with mean of zero and variance of one).

Kendall's tau:

Type of test: This is a distribution-free rank based test which is based on an alternative measure of correlation known as Kendall's correlation coefficient or Kendall's τ (tau). Like Spearman's rho, it is robust to the effect of extreme values (i.e. to highly skewed hydrological data) and to deviations from a linear relationship.

Test statistic:

The test statistic is obtained as follows (Hirsch et al., 1982):

- (i) Arrange the n data pairs $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, in order of increasing x -value.
- (ii) Divide all $n(n-1)/2$ ordered pairs of y_i values into P cases, where $y_i > y_j (i > j)$ and M cases, where $y_i < y_j (i > j)$.

- (iii) Define a test statistic $S = P - M$
- (iv) If $n > 10$, compute the standardized test statistic Z :

$$Z = \begin{cases} (S - 1) / \text{Var}(S)^{1/2} & S > 0 \\ 0 & S = 0 \\ (S + 1) / \text{Var}(S)^{1/2} & S < 0 \end{cases} \quad (5C.5)$$

where

$$\text{Var}(S) = n(n + 1)(2n + 5) / 18 \quad (5C.6)$$

Determining significance levels: The standardised test statistic Z is approximately normally distributed. Thus, the null hypothesis is rejected at significance level α if

$$|Z| > Z_{1-\alpha/2} \quad (5C.7)$$

where $Z_{1-\alpha/2}$ is the value of the standard normal distribution with the probability of exceedance of $\alpha/2$.

Comments: The Kendall's correlation test lends itself well to applications for censored data. For tied data the expression for variance takes the form

$$\text{Var}[S] = \left[n(n - 1)(2n + 5) - \sum_{i=1}^n t_i i(i - 1)(2i + 5) \right] / 18 \quad (5C.8)$$

where t_i is the number of ties of extent i .

For example, if the time series of interest is: $\{ 5, 3, 6, 7, 7, 3, 2, 3, 9, 6 \}$. Then the t_i values are:

- $t_1 = 3$ as there are three untied elements in the series (2, 5, 9);
- $t_2 = 2$ as there are two sixes and two sevens, i.e. two ties of extent two in the series;
- $t_3 = 1$ as there are three threes, i.e. one tie of extent three in the series.

The Kendall's correlation coefficient, a measure of the strength of the correlation, can be calculated as

$$\tau = 2S / [n(n - 1)] \quad (5C.9)$$

It attains values from the interval $(-1, +1)$, where the sign indicates the slope and the absolute value indicates the strength of the relationship.

Seasonal-Kendall test and modified Seasonal-Kendall test:

Type of test: The seasonal Kendall test is a rank-based distribution-free test for trend in data with seasonality. It was introduced by Hirsch *et al.* (1982) for identifying and quantifying changes in water quality data. It is based on Kendall's tau and combines the Kendall's tau measures for each of the months. It can be adapted to apply to non-monthly seasonal data (eg. quarters of the year).

Test statistic: Let the complete sample X be subdivided into sub-samples X_1 through X_{12} (one for each month), where each subsample X_i contains the n_i annual values from month i

$$\begin{aligned} X &= (X_1, X_2, \dots, X_{12}) \\ X_i &= (X_{i1}, X_{i2}, \dots, X_{ini}) \end{aligned} \quad (5C.10)$$

The null hypothesis H_0 for the seasonal Kendall test is that X is a sample of independent random variables and that X_i is a subsample of independent, and identically distributed random variables.

The test statistic S_i is defined:

$$S_i = \sum_{k=1}^{n_i-1} \sum_{j=k+1}^{n_i} \text{sgn}(x_{ij} - x_{ik}) \quad (5C.11)$$

Now,

$$E[S_i] = 0 \quad (5C.12)$$

and

$$\text{Var}[S_i] = \left[n_i(n_i - 1)(2n_i + 5) - \sum_{t_i} t_i(t_i - 1)(2t_i + 5) \right] / 18 \quad (5C.13)$$

where t_i the extent of a given tie in month i . See Hirsch *et al.* (1982) for further details.

Determining significance levels: The above test statistic is approximately normally distributed.

Comments: The modified seasonal Kendall test uses a more complex measure of the variance of S_i which builds in an allowance for autocorrelation. See Hirsch & Slack (1984) for further details. The test is suitable for seasonal data with a moderate level of autocorrelation.

The gradient of trend can be estimated using the seasonal Kendall slope estimator (Hirsch *et al.*, 1982).

Linear Regression:

Type of test: This is a parametric test that assumes normally distributed data. It is used here to test for linear trend by the linear relationship between time and the variable of interest.

Test statistic: The regression gradient is estimated by:

$$\hat{b} = \frac{\sum_{t=1}^T (t - \bar{t})(z(t) - \bar{z})}{\sum_{t=1}^T (t - \bar{t})^2} \quad (5C.14)$$

while

$$\hat{a} = \bar{z} - \hat{b}\bar{t} \quad (5C.15)$$

The required test statistic is

$$S = \frac{\hat{b}}{\hat{\sigma}} \quad (5C.16)$$

where

$$\hat{\sigma} = \sqrt{\frac{12 \sum_{t=1}^T (z(t) - \hat{a} - \hat{b}t)^2}{N(N-2)(N^2-1)}} \quad (5C.17)$$

Determining significance level: This test statistic follows a Student distribution with $T-2$ degrees of freedom under the null hypothesis. The application of this test assumes that the errors (deviations from the trend) are independent and follow the same normal distribution with 0 mean.

Appendix D

TESTS FOR CHANGES IN DISTRIBUTION

See also the explanatory notes in Appendix A.

Kolmogorov-Smirnov test:

Type of test: This is a distribution-free test for a general change in distribution. The test merely states that the two distributions are different — it does not specify whether the change is an increase or decrease in the mean or due to a change in the variance or extremes. In the context of testing for change, the data is divided into two, assuming a known change-point time, and the test is used to compare the two parts.

Test statistic: Suppose that the investigated time series changed at time m for a series of N observations. The test statistic is:

$$D = \sqrt{m(N-m)} \max_x |F_1(x) - F_2(x)| \quad (5D.1)$$

where $F_1(x)$ and $F_2(x)$ are the empirical distributions for the two parts of the data, i.e.

$$F_1(x) = (\text{No values in sample } 1 \leq x) / m$$

and $F_2(x) = (\text{No values in sample } 2 \leq x) / (N-m)$

Determining significance levels: The critical values for D can either be taken from textbooks, or assessed directly using re-sampling.

Appendix E

TESTS FOR RANDOMNESS

See also explanatory notes in Appendix A.

Runs test:

Type of test: This is a rank-based test for randomness.

Test statistic: Runs are defined as a set of consecutive observations above or below the median (WMO, 1988). If the data are randomly distributed then the expectation and the variance of the number of runs are, respectively:

$$E[n_r] = 1 + n / 2 \quad (5E.1)$$

$$Var[n_r] = n(n - 2) / [4(n - 1)] \quad (5E.2)$$

where n is the number of data in the series and n_r is the number of runs. The following test statistic can be used:

$$\{n_r - E[n_r]\} / \{Var[n_r]\}^{1/2} \quad (5E.3)$$

Determining significance levels: The above test statistic is approximately normally distributed, $N(0,1)$.

Kendall's turning point test:

Type of test: This is a rank-based test for randomness of a data series.

Test statistic: The idea of this test is to calculate the number of turning points in a series. This is done by defining a new series I :

$$I(t) = \begin{cases} 1 & \text{if } Z(t) > z(t-1) \text{ and } z(t) > z(t+1) \\ -1 & \text{if } z(t) < z(t-1) \text{ and } z(t) < z(t+1) \\ 0 & \text{else} \end{cases} \quad (5E.4)$$

The test statistic is derived from:

$$Q = \sum_{t=1}^T I(t) \quad (5E.5)$$

For large T values and an independent stationary time series (without change) Q follows a normal distribution with the expectation:

$$E[Q] = \frac{2T - 4}{3} \quad (5E.6)$$

and the variance:

$$\text{Var}[Q] = \frac{16T - 29}{90} \quad (5E.7)$$

The series can thus be tested by calculating the test statistic

$$C = \frac{Q - \frac{2T - 4}{3}}{\sqrt{\frac{16T - 29}{90}}} \quad (5E.8)$$

Determining significance levels: C has an approximately normal $N(0,1)$ distribution and significance levels can be found in the usual way.

Appendix F

TESTS FOR INDEPENDENCE

See also the explanatory notes in Appendix A.

Bartlett's test for autocorrelation:

Type of test: This is a parametric test which assumes the data to follow a normal distribution.

Test statistic: The independence of the series is checked using the lag-1 autocorrelation of the series.

$$r_1 = \frac{\sum_{t=1}^T (z(t) - \bar{z})(z(t+1) - \bar{z})}{\sum_{t=1}^T (z(t) - \bar{z})^2} \quad (5F.1)$$

The null hypothesis is that the autocorrelation is zero — meaning that the subsequent data in the time series are independent. The corresponding test statistic is:

$$S = \frac{r_1 \sqrt{d_f}}{\sqrt{1 - r_1^2}} \quad (5F.2)$$

Where d_f is the degrees of freedom — estimated as:

$$d_f = (T - 3) \frac{1 - r_1^2}{1 + r_1^2} \quad (5F.3)$$

Determining significance levels: The test statistic S follows a Student distribution with $T-3$ degrees of freedom. Thus the statistical testing is carried out by calculating S and comparing its value with the critical value of the Student distribution. This means that the hypothesis of independence is rejected if: $|S| > s_{\left(1-\frac{\alpha}{2}\right)}$ with $s_{\left(1-\frac{\alpha}{2}\right)}$ being the $1 - \frac{\alpha}{2}$ quantile of the Student distribution with d_f degrees of freedom.

Von Neumann ratio test:

Type of test: This is a non-parametric test for independence. It should only be applied for time series of at least 30 points. The null hypothesis is that the series consists of independent elements.

Test statistic: The basis for the test statistic is R :

$$R = \frac{T \sum_{t=1}^T (z(t+1) - z(t))^2}{(T-1) \sum_{t=1}^T (z(t) - \bar{z})^2} \quad (5F.4)$$

with \bar{z} being the mean of the series z .

This R follows a normal distribution with expectation:

$$E(R) = \frac{2T}{T-1} \quad (5F.6)$$

and a variance:

$$Var(R) = \frac{4(T-2)}{(T-1)^2} \quad (5F.7)$$

In order to test the independence the standardised test statistic C is calculated:

$$C = \frac{R - \frac{2T}{T-1}}{\sqrt{\frac{4(T-2)}{(T-1)^2}}} \quad (5F.8)$$

Determining significance levels: The test statistic C has a normal distribution. Thus if

$|C| > c_{\left(1-\frac{\alpha}{2}\right)}$ with $c_{\left(1-\frac{\alpha}{2}\right)}$ being the $1 - \frac{\alpha}{2}$ quantile of the standard normal distribution $N(0,1)$

then the hypothesis that the series consists of independent observations is rejected.

Example of testing for independence

Annual discharges of the Neckar river are considered. The following table shows the annual mean discharges at the Rottweil gauge in m³/s.

Year	Discharge	Year	Discharge
61	4.17	76	2.77
62	4.28	77	5.83
63	3.88	78	6.59
64	3.20	79	5.52
65	7.98	80	5.18
66	6.13	81	5.80
67	4.49	82	6.63
68	7.13	83	6.11
69	5.51	84	4.59
70	7.63	85	3.56
71	2.48	86	7.29
72	3.51	87	6.62
73	4.47	88	7.31
74	4.37	89	3.36
75	4.55	90	4.51

Assuming that the annual discharges follow a normal distribution, the Bartlett test for the autocorrelation was applied. The estimated lag-1 autocorrelation is 0.081. From this the number of degrees of freedom d_f is:

$$d_f = (30 - 3) \frac{1 - 0.081 \cdot 0.081}{1 + 0.081 \cdot 0.081} = 26.65$$

Thus the test statistic S is:

$$S = \frac{0.081 \sqrt{26.65}}{\sqrt{1 - 0.081 \cdot 0.081}} = 0.420$$

A significance level of 5 % is chosen — this means that the critical value of the Student distribution (available as tables in standard statistical texts) is 2.06. As S does not exceed the critical value the hypothesis that the data are independent cannot be rejected.

In the case one cannot be sure about the distribution of the individual values the non-parametric von Neumann test can be applied. The test statistic R is:

$$R=1.880$$

From this the standardised variable C is calculated as:

$$C = \frac{1.880 - \frac{2 \cdot 30}{29}}{\sqrt{\frac{4 \cdot 28}{29 \cdot 29}}} = -0.518$$

The critical value of the standard normal distribution for the significance level of 5% is 1.96. As the absolute value of the calculated C does not exceed this value the hypothesis of independence cannot be rejected by this test.

PART II

SPECIAL TOPICS

CHAPTER 6

DETECTING CHANGES IN EXTREMES

Alice Robson & Francis Chiew

6.1 Introduction

6.1.1 *What is an extreme event?*

An extreme event is simply an event that is uncommon or rare. Testing for trend or other non stationarity in extremes is particularly difficult because there is typically only limited data available. The hydrological extremes are droughts and floods and these are discussed below. Similar principles apply to other types of extremes, e.g. temperature, wind speed.

6.1.2 *Importance of changes in extremes*

Knowledge about changes in extreme events is needed because of the devastating consequences associated with very extreme events. Increases in flooding, or increased thought severity, generally has a greater impact than a shift in average conditions. Climate change could potentially impact extreme events, but currently there is little evidence over whether such effects are occurring. Unless very long hydrological data sets become available, it will remain difficult to determine if climate change is affecting extremes.

6.1.3 *Is there evidence that extremes are changing due to climate*

Recent studies examining trends in extremes of flows have not generally found conclusive evidence of change due to climate (e.g. Lins, 1999; Robson *et al.*, 1997). It must be remembered that this means that we cannot prove that change is taking place. It is possible that changes are occurring but that we do not yet have sufficient data for it to be detectable. The picture is somewhat different for studies of rainfall. Here there are a number of studies that suggest that rainfall has become more variable, and that rainfall intensity and the frequency of high intensity rainfall has increased in some areas (see Section 5.3 Nicholls *et al.*, 1996 for a summary).

6.2 *Choosing a suitable data series*

A critical stage in testing for trend or step change is to obtain a suitable data series. Because extremes are infrequent, it is generally necessary to construct a data series that specifically picks out extremes. For example, in the case of floods, it would be inappropriate to use a daily or hourly flow series because the data set would be dominated by values that are normal flows. Instead, it is usual to use a dataset that contains only peak flows. The following two sections describe the most common types of data series used for studying extreme events. Ideally we should look at more than one type of data when testing for non-stationarity. Each type of data provides a different perspective on how extremes may be changing

6.2.1 *Annual maxima and minima series*

An annual maxima series is obtained by taking the largest value in each year or season of interest. For example, the annual maximum flood series is an annual series of the largest flow value for each year. Annual maxima and minima series are generally straight forward to obtain and are a useful and practical way of summarising extremes in data series. However, they may exclude information on some extremes (e.g. if more than one extreme event occurs within a year), or may contain values that do not relate to extreme events (e.g. if no extreme

event occurs within a year). In some instances, annual maxima values may be even zero, e.g. if a stream is dry for a year.

The annual maxima or minima series should be derived from an appropriate underlying data series. If extremes extend over long periods (e.g. droughts and low flows), a daily flow series is likely to be adequate. If an extreme event happens rapidly (e.g. a flood peak) then the annual maxima series will ideally be derived from instantaneous or hourly maximum flows; the extreme event could be missed with daily data.

6.2.2 Peaks-over-threshold (POT) or partial duration series

A peak-over-threshold (POT) series, also called a partial duration series (PDS), contains all events larger than a threshold value. A POT series has advantages over annual maxima series in that all major events are included (not just the largest in a year) and all data points in the POT series are extreme events (Stedinger *et al.*, 1993). A POT series provides a much fuller description of the nature of extreme events and is likely to result in a greater ability to detect change. Furthermore, POT series can be used to examine whether there is a change in either the magnitude or frequency of extreme events

In practice, the POT or partial duration series can be more difficult to obtain than annual maxima (see below) and care is often needed in analysing such series because they are irregular (there may be many events in one year and none in another).

To obtain a POT or PDS series, it is necessary to

(1) Identify a suitable threshold

The threshold should be such that the average number of events per year is within reasonable bounds. Often the threshold is selected to give a fixed average number of events per year. It is also possible, but less common, to select a threshold that corresponds to criteria (e.g. bank level).

(2) Determine whether events are independent

It is important that POT events are independent of one another (there should not be multiple peaks corresponding to the same event). Judging independence can be complex and requires suitable criteria to be available. If there are multiple peaks corresponding to the same event then usually only the largest peak is retained in the POT series. Judging independence is complex and depends on local circumstances and the type of data (e.g. see below).

6.3 Data series for floods

For floods, use of both annual maxima and POT data sets is suggested. For POT data it is recommended that more than one threshold be used when analysing the data. For example, testing a POT series containing an average of 1 event/year provides information about changes in large floods, and testing data containing three events/year also gives knowledge about changes in moderate events. Testing a variety of series gives a better picture of what is occurring.

For POT data it is necessary to define independence rules, to ensure that flood peaks are from separate events. An approximately independent series can sometimes be obtained using simple hydrological criteria (e.g. see Bayliss & Jones, 1993) but can also be derived using more sophisticated statistical approaches (e.g. Davison & Smith, 1990; Dixon & Tawn, 1995).

6.4 Data series for droughts

Testing for trends in drought is non-trivial because it is often the duration of the drought that is critical (in contrast to floods where flood magnitude is usually more important). Furthermore, severe droughts may span a number of years, which means that much longer data sets are required for trend/step-change detection to be useful.

There are various definitions of drought. Hydrological drought typically refers to periods of below normal streamflow or depleted reservoir storage (Beran & Rodier, 1985). The annual series of seven-day low-flow (lowest discharge over seven consecutive days) is commonly used as a low-flow series.

Data series for “droughts” may also be obtained from a flow series by considering periods when the flow is below a certain threshold. The truncation level can be standardised by defining it using a certain percentile of the flow duration curve - this method of standardisation gives the same average number of days having drought at each site (Zelenhasic & Salvai, 1987). Trends can then be tested on the annual and/or partial series of drought duration (number of days flow is below threshold), drought severity (flow volume below the threshold) and the number of times in a year that flow is below the threshold. However, like the POT and PDS series for floods, it is difficult to judge whether consecutive droughts are independent.

This can be overcome by using a deficit series, defined as the local minimum storage in a semi-infinite reservoir between spills (Pegram *et al.*, 1980). The flow series is used in a storage behaviour analysis (McMahon & Mein, 1978) which simulates storage fluctuations in a reservoir subject to a given sequence of inflow (the flow series) and a given draft (a constant value, say 70% of the mean inflow, can be used here to represent water use, evaporation and other reservoir losses). In the analysis, the next day's storage is calculated as the present storage plus inflow less draft. The deficit series is obtained by having a finite reservoir capacity and not allowing the reservoir to empty (the capacity can be set as zero datum - the storage is therefore always negative). The deficits (minimum storage between spills) in the series are statistically independent because they are the realisations of a renewal process (Pegram *et al.*, 1980).

Meteorological drought and agricultural drought are two other commonly used drought definitions. Meteorological drought is defined as periods during which the actual moisture supply cumulatively falls short of the climatically appropriate moisture supply, and agricultural drought is defined as periods when the soil moisture is inadequate to meet evapotranspirative demands. The Palmer Drought Severity Index (PDSI) is commonly used to reflect a prolonged and abnormal moisture deficiency (Palmer, 1965). The PDSI is computed as a function of the difference, accumulated through time, of the actual rainfall and the CAFEC rainfall (climatically appropriate for existing conditions of evaporation and other components of the water balance). It ranges from -4 (extreme drought), 0 (normal condition) to +4 (extreme wet period). A continuous PDSI series can be obtained by analysing either the weekly or monthly water balance.

6.5 Data series for seasonal data

If a data series is strongly seasonal or if changes relating to a particular season are important, then it may be appropriate to consider a seasonal extreme series. For this, the annual maxima and POTIPDS series are derived using just the data for the required season.

6.6 Guidelines for testing for non-stationarity

The various series described here can be tested for trends and/or step-change using the permutation and bootstrap methods detailed in Chapter 5.

6.7 Discussion

Because extreme events are rare it is very important to use long records, more so than with other hydrological data. This is particularly true for droughts because they may extend to periods of a year or more. Detection of effects due to climate change is likely to require much longer data sets than detection of effects with a clear anthropogenic cause.

Whenever non-stationarity is detected it is very important to try and establish the likely cause. It usually is helpful to examine other related hydrological series. For example, if trend is seen on a catchment that has experienced land-use change, examining a rainfall series can provide information about whether climatic conditions have been steady across the period of record.

A further way of gaining insights into the nature and causes of non-stationarity is to look at extremes from a regional perspective (Chapter 8, also Robson *et al.*, 1997). Use of the graphical approaches described in Chapter 4 should be considered.

References

- Bayliss, A.C. & Jones, R.C., 1993. Peaks-over-threshold flood database: summary statistics and seasonality. Report 121, Institute of Hydrology, Wallingford, U.K. 61 pp.
- Beran, M.A. & Rodier, 1985. Hydrological Aspects of Drought. UNESCO, Paris.
- Davison, A.C. & Smith, R.L., 1990. Models for exceedances over high thresholds. *J.R. Statist. Soc. B.*, **52**, 393-442.
- Dixon, M.J. & Tawn, L.A., 1995. Extreme sea-levels at the UK A-Class sites: Optimal site-by-site analyses and spatial analyses for the east coast. Proudman Oceanographic Laboratory, 1-298.
- Lins, H.F. & Slack, J.R., 1999. Streamflow trends in the United States. *Geophys. Res. Letters*, **26**, 2, 227-230.
- McMahon, T.A. & Mein, R.G., 1978. *Reservoir Capacity and Yield*. Elsevier, New York.
- Nicholls, N., Gruza, G.V., Jouzel, J., Karl, T.R., Ogallo, L.A. & Parker, D.E., 1996. Observed climate variability and change. In: *Climate Change 1995. The Science of Climate Change*. Eds. Houghton, J.T., Meira Filho, L.G., Callander, B.A., Harris, N., Kattenberg, A. and Maskell. Cambridge University Press.
- Palmer, W.C., 1965. Meteorological Drought. U.S. Department of Commerce, Weather Bureau Research Paper No. 45.
- Pegram, G.G.S., Salas, J.D., Boes, D.C. & Yevjevich, V., 1980. Stochastic properties of water storage. Colorado State University, Hydrology Paper 100, 48 pp.
- Robson, A.J., Reed D.W. & Jones, T.K., 1997. Trends in UK floods, In: Proceedings of 32nd MAFF conference of River and Coastal Engineers, Keel University. 12 pp.
- Stedinger, J.R., Vogel, R.M. & Foufoula-Georgiou, E., 1993. Chapter 18: Frequency analysis of extreme events. In: *Handbook of Hydrology*, ed. by Maidment, D., McGrawHill, 65pp.
- Tallaksen, L.M., Madsen, H. & Clausen B., 1997. On the definition and modelling of streamflow drought duration and deficit volume. *Hydrol. Sci. J.*, 42(1), 15-34.
- Zelenhasic, E. & Salvai, A., 1987. A method of streamflow drought analysis. *Wat. Resour. Res.*, **23**, 156-168.

CHAPTER 7

TESTS FOR CHANGES IN FLOW REGIMES

Hege Hisdal

7.1 Introduction

A number of factors such as increased demands for water in general and requirements for increased levels of service (e.g. reliability of supply) have put focus on the quantity and quality of water resources. Flow regimes describe the average seasonal pattern of river flow. This pattern is important to water resource management and biological cycles in a river. Changes in regime types could have serious negative and positive consequences for a number of water management issues including drinking water supply, hydropower production, irrigation, reservoir design and management, river pollution and ecological aspects. It should be stressed that a redistribution of water throughout the year might occur without influencing the annual discharge.

An example of a positive consequence of changed distribution of runoff during the year in the Nordic countries is described in S (1998). An expected climate change leading to increased winter runoff and reduced spring floods will allow more reservoir capacity to be used for attenuation, thus reducing spillage and increasing total production. It is important to highlight however, that even though climate changes have the potential to affect water resources availability; studies have shown that considerable differences might arise even within a narrow range of feasible future climates (Arnell, 1992 a).

Hydrological regimes reflect climatic and physiographic conditions of a catchment. A changed regime would therefore indicate natural or man-induced changes in the climate or the environment of a catchment. Hydrological time series describing regimes could be based on seasonal mean flow or monthly mean flow. The timing of a seasonal event is an important characteristic of a regime.

The regularity of seasonal patterns is another important aspect in this context because operational water management relies upon a certain stability of a river flow regime. The average pattern can be stable, demonstrating the same river flow from year to year, or unstable with alternating seasonal patterns.

This chapter includes a short description of flow regime classifications. Several variables characterise river flow regimes and could be tested for changes. Examples of variables and tests are listed.

7.2 Flow regime classification

A river flow regime describes the average seasonal behaviour of flow. Several classifications exist some covering the whole world others more locally adapted. Two examples of traditional classifications are due to Lvovich (1938) and Pardé (1955) both qualitative based on genetic source (e.g. snowmelt, rainfall) and flow distribution within the year. A classification according to Pardé is carried out for a number of catchments in Australia (Ward, 1984). The major drawbacks of these traditional classifications are their time-consuming nature and their dependence on subjective experience and judgement.

When large amounts of data both in time and space are to be classified, an automatic, computer-adapted method is necessary. An example of a global regime classification is given by Haines *et al.* (1988). A Scandinavian Working Group on Flow Regimes developed a

robust definition of flow regimes for Scandinavia (Gottschalk *et al.*, 1979). The classification distinguishes between three main sources of flow: water from snowmelt, rain and mixed and defines the timing of maximum and minimum mean monthly flow. The principles of this classification, given in Table 7.1, have been used as a basis for the first computerised flow regime classification of Scandinavia (Krasovskaia & Gottschalk, 1992).

Table 7.1. Principles of Scandinavian flow regime classification (from Krasovskaia *et al.*, 1994).

High water		
H1 Dominant snowmelt high water	H2 Transition to secondary high water	H3 Dominant rain high water
Low water		
L1 Dominant low water in winter	L2 Transition zone, two low water periods in different seasons	L3 Dominant summer low water

Later the same scheme was generalised and adapted to river flow series in western Europe (Krasovskaia *et al.*, 1994). A total of 13 regime types were defined, of which 4 were transitive. Examples of the most common European flow regimes according to this classification are found in Fig. 7.1. Flow regime classifications are traditionally presented by placing the monthly hydrograph on a geographical map, showing the location of the stations. With a large number of stations this is problematic. A solution is to use some form of spatial interpolation and map the regimes on a gridded basis. Examples of gridded maps based on the 13 regimes types discussed above can be found in Arnell *et al.* (1993).

A river flow regime classification could be used as a basis for defining hydrological regions (see Chapter 8 on regionalisation).

7.3 Variables describing river flow regimes

Important in studies of changes in river flow regimes is the data requirement (see Chapter 3 on data). The type of hydrological variable to be analysed will be limited by the available temporal resolution of the time series. River flow regimes often cover large regions and changes in regimes might form regional patterns. Therefore a broad spatial coverage of data series is important. In the context of assessing climate variability and change and possible impacts on river flow regimes, long time series of observational data are needed. These aspects have to be focused when choosing variables to be analysed.

Several variables are used to describe river flow regimes and changes in regimes could be found analysing different hydrological parameters as:

- ✎ seasonal mean flow
- ✎ monthly mean flow
- ✎ timing of seasonal events as:
 - start of snowmelt
 - snowmelt flood-peak
 - maximum flow
 - freezing of rivers
 - ice break up
 - minimum flow
- ✎ stability of regimes

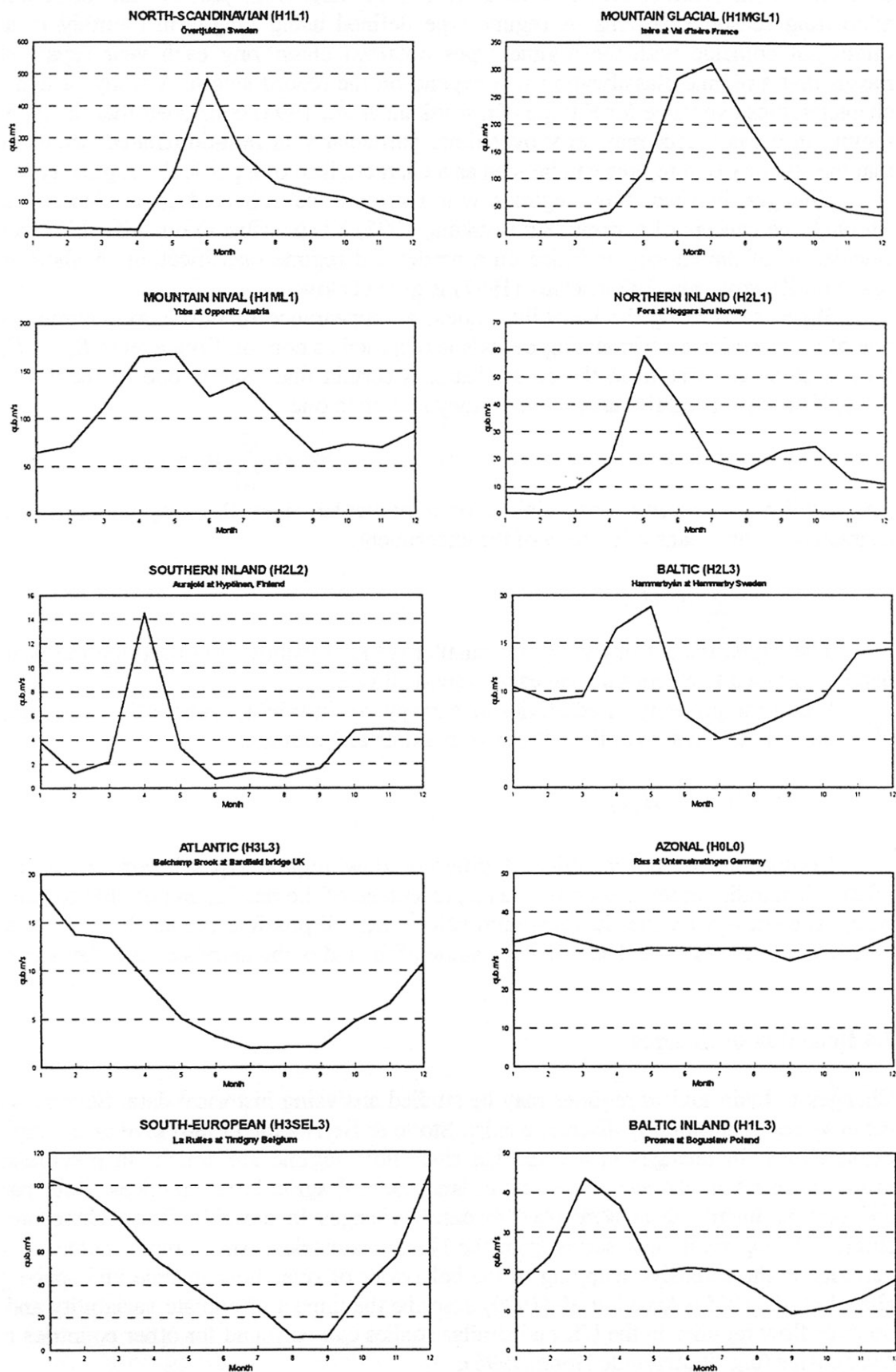


Fig.7.1. Examples of the most common European flow regimes (Krasovskaia *et al.*, 1994).

The stability of a certain river flow regime, i.e. the regularity of seasonal patterns is an important aspect in this context, and requires a further explanation. The average pattern can be stable, demonstrating the same variation in river flow from year to year, or unstable with alternating seasonal patterns. A regime type defined using long-term monthly mean flows might not coincide with the regime types obtained classifying each year separately. This means that a regime classification will depend on the record length. A study of hydrological characteristics over time for Europe (Krasovskaia *et al.*, 1993) concluded that the type of flow regime in a catchment may vary over time, particularly in rain-dominated catchments, and that the stability of a regime can be seen as a characteristic of a particular regime type.

Changes in river flow regimes will therefore be reflected in a changed stability. Through the concept of entropy (Krasovskaia, 1995) it is possible to quantify this change. The calculation of the entropy is based on a predefined regime classification. A short summary based on Krasovskaia & Gottschalk (1997) is given below.

When calculating the instability index, an appearance of flow maxima/minima within one of n respective discriminating periods is regarded as coming from events E_1, \dots, E_n which form a complete system in the sense that it is certain that exactly one of them will occur. Thus, if their probabilities are p_1, \dots, p_n , they add up to one:

$$p_i = P(E_i) \quad i = 1, \dots, n \quad \text{where} \quad \sum_{i=1}^n p_i = 1 \quad (7.1)$$

The stability of maxima and minima of a particular river flow regime, respectively, is characterised by means of entropy of the experiment:

$$H = -\sum_{i=1}^n p_i \ln(p_i) \quad (7.2)$$

The higher the entropy value, the smaller is the probability of observing the flow regime pattern assigned to a series during each individual year.

Using the property of additivity of entropy, an instability index of a flow regime type HR is calculated as a sum of the entropy of maxima and minima

$$H = H_{MAX} + H_{MIN} \quad (7.3)$$

Comparison of the stability of different flow regime types becomes easier when a relative instability index is used (e.g. as a percentage of the maximum possible for this regime type). The entropy reaches its maximum value when all possible events, E_1, \dots, E_n , are equally probable, $p_1, \dots, p_n = 1/n$. The lower the value of the index the more stable is the regime type.

7.4 Detection of changes

Changes in hydrological regimes may be studied analysing historical data. Several studies of mean seasonal or monthly discharge exist. Stolte & Herrington (1984) give an example where explanations for changes in a Canadian river flow regime are sought in precipitation and evapotranspiration changes and shifts in land use and agricultural practices. Both parametric and non-parametric tests were used to detect changes in monthly flow. There are several studies of long river flow series from the Nordic countries. An example including seasonal patterns, is the extensive mapping of the behaviour of river flow in time and space given in Hisdal *et al.* (1995). Arnell *et al.* (1990) describe the impact of climate variability and change

on river flow regimes in the UK and similar studies can be found for other countries or single river basins (e.g. Wateren-de Hoog, 1995).

Another approach is the modelling of river flow under the assumption of changes or fluctuations in the climate. Nemec & Schaake (1982) utilised a deterministic conceptual rainfall-runoff model to estimate the sensitivity of water resources systems to climate variations. Other examples of impact assessments using rainfall-runoff models are given in Arnell (1992 b), Krasovskaia & S (1997) and S (1998). The latter describes expected changes in runoff regimes in different altitude zones (mountain, lowland etc.) due to alternative future global change scenarios and the resulting impacts on the Nordic system for hydroelectric power production.

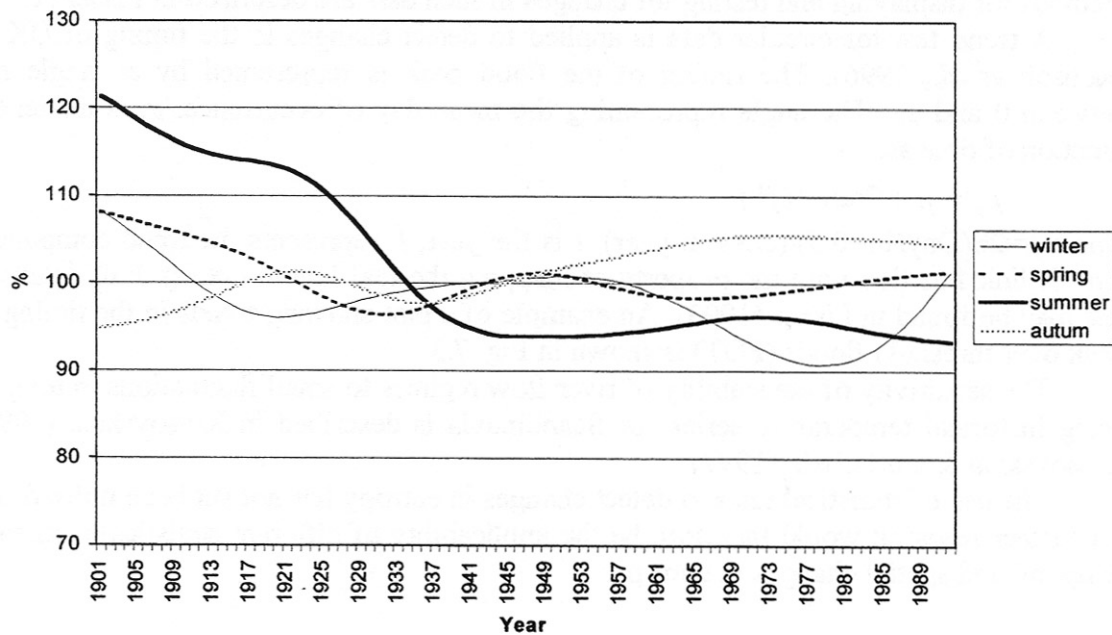


Fig. 7.2. Smoothed seasonal mean regional discharge values from north-western Norway. The values are given as a percentage deviation from the long-term mean value.

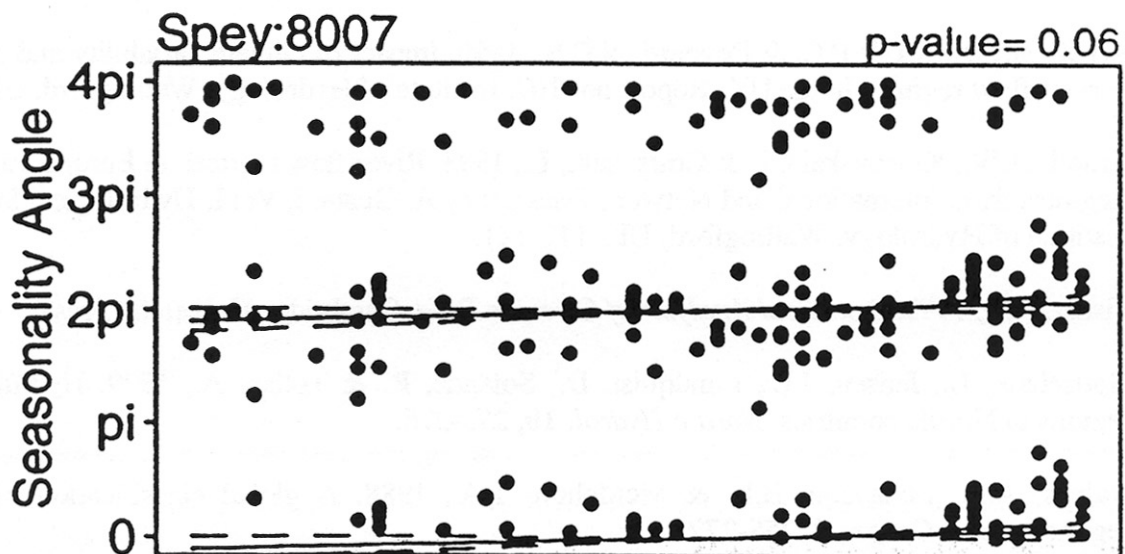


Fig. 7.3. Trend in seasonality for POT (station Spey, UK). The water year is shown on the x-axis whilst the angle representing the season is shown on the y-axis, confidence intervals for the fitted trends shown in dotted lines. Source: Robson & Reed (1996).

In general, the exploratory/visual analyses described in Chapter 4 and the tests for changes recommended in Chapter 5, can be used to detect changes in river flow regimes. Fig. 7.2 shows an example of smoothed mean seasonal flow in north-western Norway (Hisdal *et al.*, 1995).

An important supplement is to regard the timing of seasonal events as circular data (especially recommended if the event occurs around the turn of the year). Different statistical methods for displaying and testing for changes in such data are described in Fisher (1993).

A trend test for circular data is applied to detect changes in the timing of UK floods (Robson *et al.*, 1996). The timing of the flood peak is represented by an angle ranging between 0 and 2π . The angle representing the mean day of occurrence, μ , is modelled as a function of time as:

$$\mu t = \mu + 2 \tan^{-1}(\beta t) \quad (7.4)$$

where $\mu = 2\pi(\text{DayNo} + 0.5)/(\text{days in year})$, t is the year, β represents the trend component and \tan^{-1} is link function with the property of mapping the real line to $(-\pi, \pi)$. Full details of the test may be found in Fisher (1993). An example of a plot showing trends in the timing of UK peak over threshold floods (POT) is shown in Fig. 7.3.

The sensitivity of the stability of river flow regimes to small fluctuations in temperature using historical temperature series for Scandinavia is described in Krasovskaia (1996) and Krasovskaia & Gottschalk (1997).

The use of statistical tests to detect changes in entropy has not yet been utilised. A topic for further research would therefore be the applicability of different statistical tests to detect temporal and spatial changes in entropy.

References

- Arnell, N., 1992 a. Impacts of climatic change on river flow regimes in the UK. *J. of Inst. of Water and Environ. Manag.*, **6**(4), 432-442.
- Arnell, N., 1992 b. Factors controlling the effects of climate change on river flow regimes in a humid temperate environment. *J. of Hydrol.*, **132**, 321-342.
- Arnell, N., Brown, R.P.C. & Reynard, N.C.S., 1990. Impact of climate variability and change in river flow regimes in the UK. Report no. 107, Institute of Hydrology, Wallingford, UK.
- Arnell, N.W., Krasovskaia I. & Gottschalk, L., 1993. River flow regimes in Europe. In: Flow Regimes from International and Network Data ed. by A. Gustard, Vol I, Hydrological Studies. Institute of Hydrology, Wallingford, UK, 112-121.
- Fisher, N.I., 1993. *Statistical Analysis of Circular Data*. Cambridge University Press.
- Gottschalk, L., Jensen, J.L., Lundquist, D., Solantie, R. & Tollan, A., 1979. Hydrological regions in Nordic countries. *Nordic Hydrol.* **10**, 273-276.
- Haines, A.T., Finlayson, B.L. & McMahon, T.A., 1988. A global classification of river regimes, *Appl. Geogr.*, **8**, 255-272.
- Hisdal, H., Erup, J., Guómundsson, K., Hiltunen, T., Jutman T., Ovesen, N.B. & Roald, L.A., 1995. Historical runoff variations in the Nordic countries, NHP Report No 37.
- Krasovskaia, I., 1995. Quantification of the stability of river low regimes. *Hydrol. Sci. J.*, **40**(5), 587-598.

Krasovskaia, I., 1996. Sensitivity of the stability of river flow regimes to small fluctuations in temperature. *Hydrol. Sci. J.*, **41**(2), 251-264.

Krasovskaia, I., Arnell, N. & Gottschalk, L., 1994. Flow regimes in northern and western Europe: development and application of procedures for classifying flow regimes. In: FRIEND: Flow Regimes from International and Network Data, ed. by P. Seuna, A. Gustard, N.W. Arnell & G.A. Cole, Proc. 2nd. FRIEND Conf. UNESCO, Braunschweig, Germany, October 1993, 185-193, IAHS Publ. no. 221.

Krasovskaia, I. & Gottschalk, L., 1992. Stability of river flow regimes. *Nordic Hydrol.*, **23**, 137-154.

Krasovskaia, I. & Gottschalk, L., 1997. Stability of river flow regimes. In FRIEND Flow Regimes from International Experimental and Network Data, Third Report: 1994-1997, ed. By G. Oberlin, Cemagref, Lyon, France, 79-86.

Krasovskaia, I., Gottschalk, L., Arnell, N. & Wateren-de Hoog, B., 1993. Variations in regime class over time. In: Flow Regimes from International and Network Data, ed. by A. Gustard, Vol. I, Hydrological Studies. Institute of Hydrology, Wallingford, Oxfordshire, UK, 131-138.

Krasovskaia, I. & S.N.R., 1997. Sensitivity of the stability of Scandinavian river flow regimes to a predicted temperature rise. *Hydrol. Sci. J.*, **42**(5), 693-711.

Lvovich, M.I., 1938. Opyt klassifikatsii rek SSSR (Experience of classification of rivers of the USSR). Trudy (316, Leningrad (in Russian).

Nemec, J. & Schaake, J., 1982. Sensitivity of water resources systems to climate variation. *Hydrol. Sci. J.*, **27**, 327-343.

Pardé, M., 1955. *Fleuves et Rivières*. Collin, Paris.

Robson, A. & Reed, D.W., 1996. Non-stationarity in UK flood records. Flood Estimation Handbook Note 25, Institute of Hydrology, Wallingford, UK.

Stolte, W.J. & Herrington, R., 1984. Changes in the hydrologic regime of the Battle river basin. Alberta, Canada, *J. of Hydrol.*, **71**, 285-301.

Saelthun N.R. (ed.), 1998. Climate change impacts on runoff and hydropower in the Nordic countries. Nordic Council of Ministers, TemaNord 1998, 552.

Ward, R.C., 1984. Some aspects of river flow in northern New South Wales, Australia. *J. of Hydrol.*, **71**, 31-51.

Wateren-de Hoog, B., 1995. The effect of climate variability on discharge as dependent on catchment characteristics in the Upper Loire basin, France. *Hydrol. Sci. J.*, **40**(5), 633-646.

CHAPTER 8

SPATIAL/REGIONAL TRENDS

Harry F. Lins

The assessment of regional trends in hydrological conditions can be approached from two distinct perspectives, one inherently univariate and one multivariate. The univariate approach involves testing for trends at individual sites and then grouping or regionalizing sites having similar test results. The multivariate approach differs in that regions are first identified from the hydrological time series collected at multiple sites, and a new derived time series for each region is then tested for trends.

The former is more applicable in instances where the analyst wants to preserve as much of the temporal information at a single site as possible, while also identifying adjacent sites exhibiting similar behavior. A common area where such an application is seen is in engineering design analysis. The latter is more useful in applications where the goal is to emphasize the temporal behavior of coherent regional patterns of variability; as might be the case in a climatic or hydroclimatic analysis. For any given set of hydrological records, however, the basic conclusions drawn from either approach should be the same.

Moreover, consideration of trends in a regional framework necessarily involves a geographical perspective. The plotting of trend results on a map is, therefore, essential to communicating effectively the pattern or dimensions of a regional trend. How the results are mapped must also be considered. If the density of the observing network is low, then it may be most meaningful to plot the trend results as point symbols on the map. If the network is dense, or the quantity being mapped has a gradient, then the results may be contoured. At times it may also be desirable to couple a map with a time series plot. The point is, the analyst must decide on an appropriate graphical design for meaningfully and objectively depicting the spatial characteristics of the trend results. Examples of alternative approaches to evaluating and illustrating trends in a spatial/regional framework, along with assessments of their relative strength and weakness, follow.

Example 1. Map of trend test results for a collection of stations

One of the most common approaches to assessing spatial or regional trends is to apply a test for trend as described in Chapter 5, to the hydrological time series collected at a number of individual sites. The test results (such as trend direction and significance level) are then mapped and interpreted in terms of their spatial distribution. An example of this approach was used by Smith et al. (1996) in a study of trends in 14 physical and chemical variables measured monthly at 77 river sites in New Zealand between 1989 and 1993. Mapped results for two of the variables, flow (discharge) and water temperature, appear in Fig. 8.1. The information depicted on the maps includes, for each sampling site, an identifier, a trend direction, and the trend significance level. The site identifier is presented as a three-digit alphanumeric code. The trend direction is depicted by a triangle, pointing upward for an increasing trend and downward for a decreasing trend. A solid circle is used to denote a site having no trend. The significance of the trend is illustrated by the shading of the triangle, which is solid when the p-value is less than 0.05, and open when the p-value is greater than 0.05 but less than 0.10.

The map for flow (Fig. 8.1A) indicates that nearly all sites in the northern half of New Zealand's North Island experienced a statistically significant decrease in discharge during the period of record. On the South Island no regional cluster of trends was evident, although an increasing trend in flow was indicated at several stations. The authors explained the regional

decline in northern North Island flows as reflecting a contemporaneous rainfall decline over the same area that had been independently documented by others.

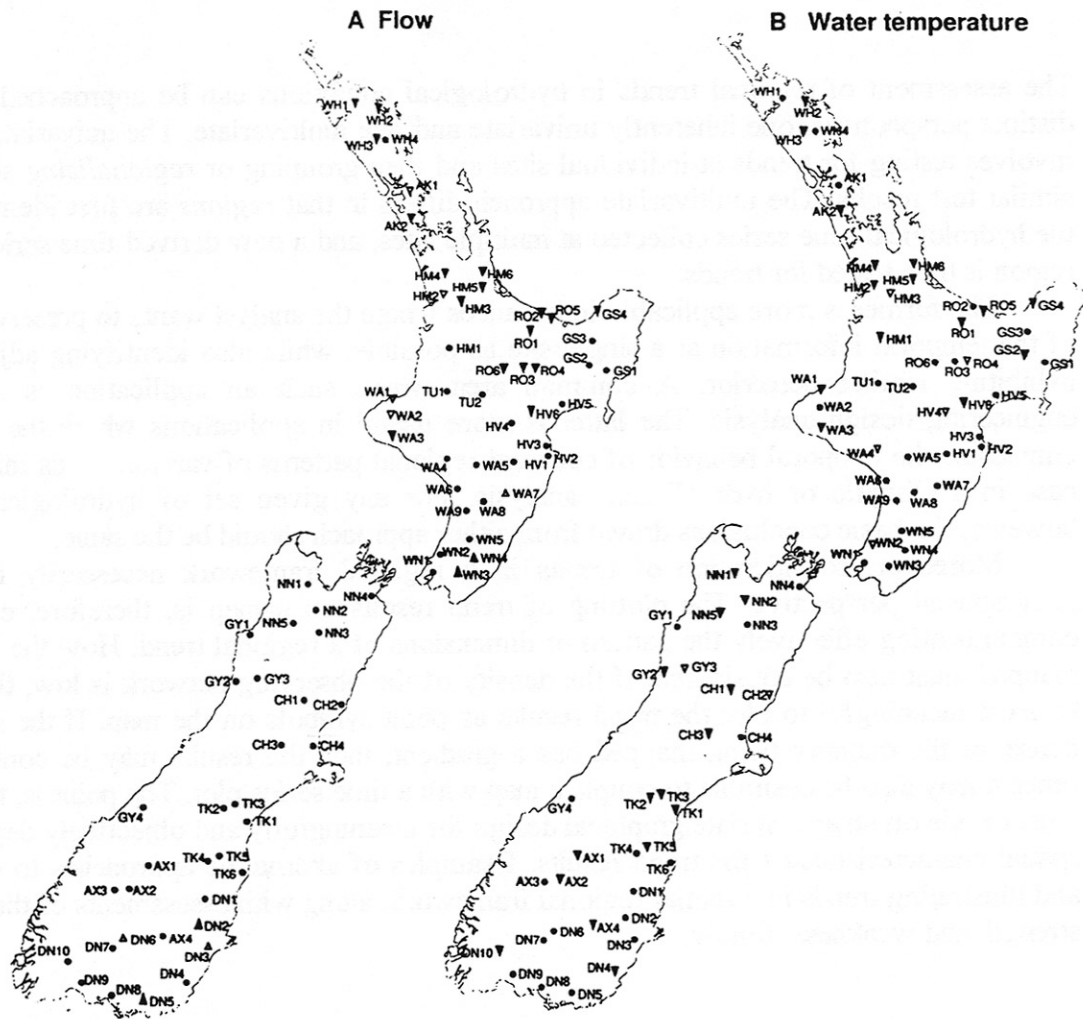


Fig. 8.1. P -values for streamflow (A) and water temperature (B). Upward trends are denoted by ▲ ($P < 0.05$) and △ ($0.05 < P < 0.10$); absence of trend is denoted by • ($F > 0.10$); downward trends are denoted by ▼ ($P < 0.05$) and ▽ ($0.05 < P < 0.10$). Source: Smith *et al.* (1996).

A very different pattern of trends was found for water temperature (Fig. 8.1B). Rather than a regionally-specific distribution, as with flow, water temperature was found to be decreasing across the entire country. Downward trends were calculated at approximately half of the 77 sites, and the declines were relatively uniform spatially. This suggests a national, as opposed to regional, trend that the authors report as being consistent with two very large-scale events: 1) changes in the Southern Oscillation Index; and 2) the eruption of Mt. Pinatubo. The two events coupled to reduce solar insolation and lower air temperatures across all of New Zealand near the end of the hydrologic trend analysis period, leading to the observed water temperature decline.

Advantage of this approach: It is possible to visualize the presence or absence of trends (including significance level information) at individual locations within a spatial domain (i.e., river basin, state country, continent etc.). Provides a good basis for comparing trend test results of multiple variables and for identifying potential causative factors. Applicable over a range of network densities.

Disadvantage: The observed and mapped trends are monotonic representations of change over the entire period of record. Detailed information on intra-period transitions, such as interannual or interdecadal variations, is usually not depicted. Information on trend magnitude, if available, is difficult to depict.

Example 2. Map with plots of area-averaged time series and trends

In some applications, it is desirable to show the actual time series (or a variant thereof) and a trend line for each at several sites or areas. A good example, comparing recent changes in evaporation for several regions in the former Soviet Union and the United States, appears in Nicholls *et al.* (1996; Fig. 8.2).

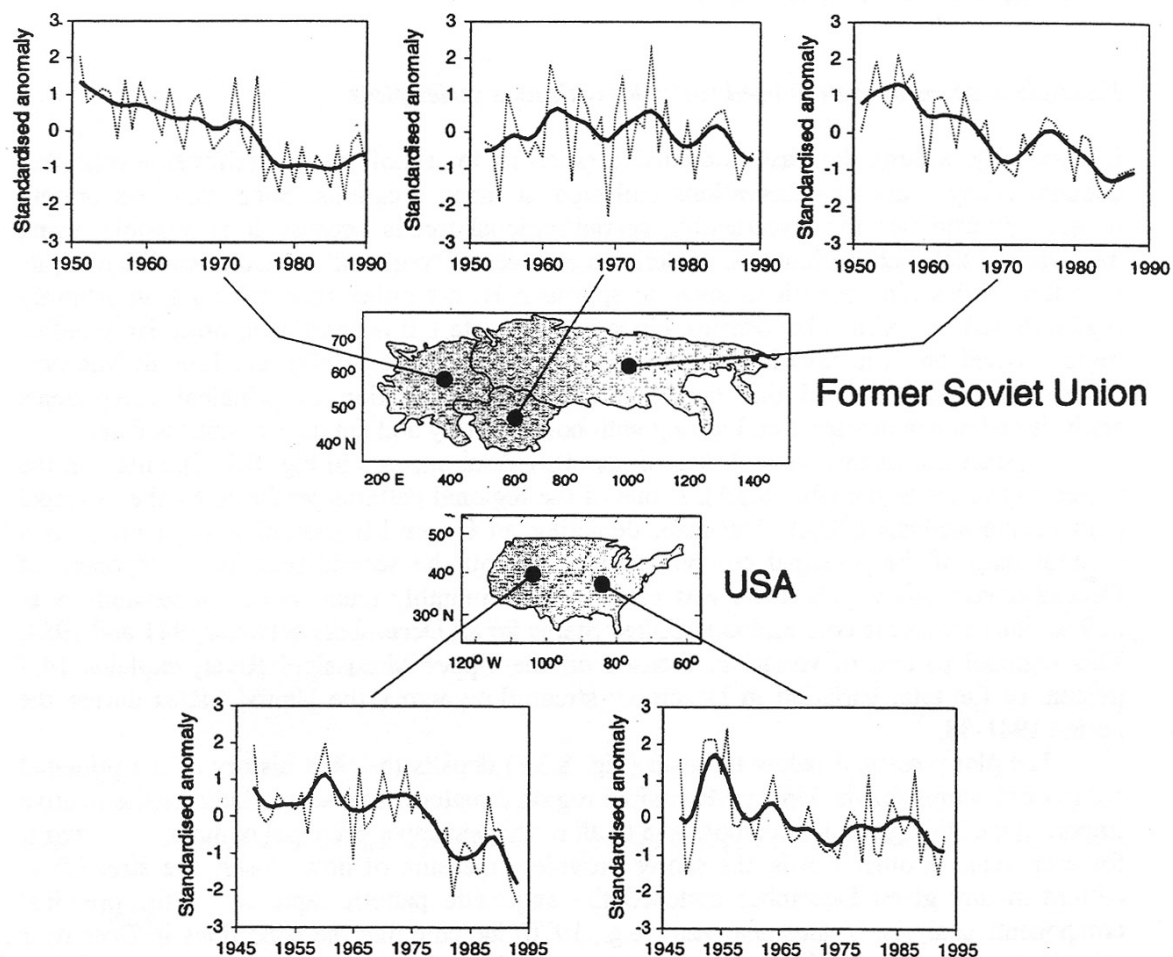


Fig. 8.2. Area-averaged standardized anomalies of evaporation derived from pan evaporimeters for sectors in the former Soviet Union in the warm season of the year, and for the USA. The dashed lines represent interannual variations and the smoothed curves suppress variations on time-scales of a decade or less. Source: Nicholls *et al.* (1996).

In the figure, it is easy to see that warm season evaporation declined over the European and Siberian sections of the former Soviet Union between the early 1950s and late 1980s, while no systematic trend occurred in central Asia and Kazakhstan. Significantly, the rapid decline in evaporation in 1976 in the European sector occurred contemporaneously with a documented decrease in the diurnal temperature range across the country. Decreases in evaporation are also apparent in the United States, where a similar decrease in the diurnal temperature range was also documented. In this particular example, explicit trend tests were not performed. The authors chose to depict temporal change using smoothed curves without applying standard tests for trend. This is really a matter of choice for the analyst depending on the objective. One could very easily apply this same approach and incorporate standard parametric or non-parametric trend tests.

Advantage of this approach: Provides a very simple and flexible means of depicting changes in a variable through time; allows quick inter-site or inter-regional comparison; facilitates visual comparison with the time series of other variables to evaluate possible contemporaneous patterns or relationships.

Disadvantage: Difficult to use as a page-sized presentation of results when there are many site or regional time series to display.

Example 3. Maps of regionalized variables with time series plots

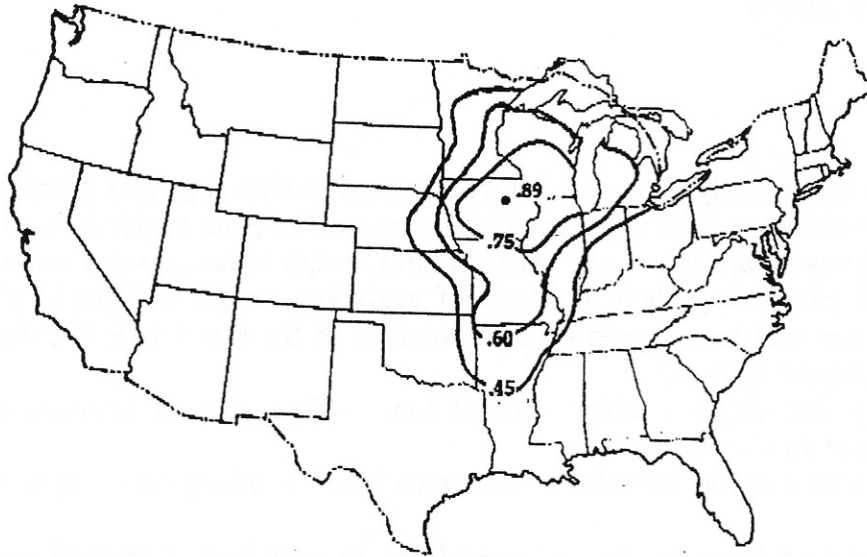
Occasionally, a large database is available (such as for runoff or groundwater levels) that contains many years of observations collected at many locations. Such data sets present unique opportunities for investigating spatial/regional trends because it is possible, using multivariate statistical techniques, to derive a new set of “regional” variables each having its own time series. The benefit to such an approach is that rather than selecting an arbitrary region (based on political boundaries, climatic zones, etc.), it is possible to objectively define regions based on their covariability through time. Lins (1985, 1997) and Lins & Michaels (1994) have demonstrated this approach using a combination of principal components analysis and nonparametric trend testing with both monthly and annual streamflow data.

An example of this work, based on monthly data, appears in Fig. 8.3. The map on the upper side of the figure (Fig. 8.3A), is one of the regional patterns produced by the principal components analysis (PCA). The map, depicting an Upper Mississippi River pattern, is a contour map of the principal component loadings on the second principal component of December streamflow. The PCA was performed on monthly mean values of streamflow at 559 stations across the conterminous United States for all Decembers between 1941 and 1988. This regional pattern of variation, focused on the Upper Mississippi River, explains 14.5 percent of the total variation in December streamflow across the United States during the period 1941-88.

The plot presented below the map (Fig. 8.3B) depicts the time history of the principal component scores for the Upper Mississippi regional pattern. The scores indicate the relative importance of this pattern, in comparison to all other December principal component patterns, for each year. In other words, the scores provide a measure of how closely the streamflow pattern in any given December matched the aggregate pattern captured by the principal components analysis. Values near zero (e.g., 1970) indicate that the conditions in December of that year didn't closely match the mapped principal component pattern. In contrast, relatively high values (e.g., 1983) indicate that the observed conditions matched the PC pattern very closely. It is apparent from the plot that the Upper Mississippi region experienced increasing streamflows during the month of December between 1941 and 1988.

By applying the Mann-Kendall test to the principal component scores, Lins (1994) documented that a statistically significant increase did, indeed, exist for this pattern.

A *Upper Mississippi Pattern -- December (mode 2, 14.5%)*



B

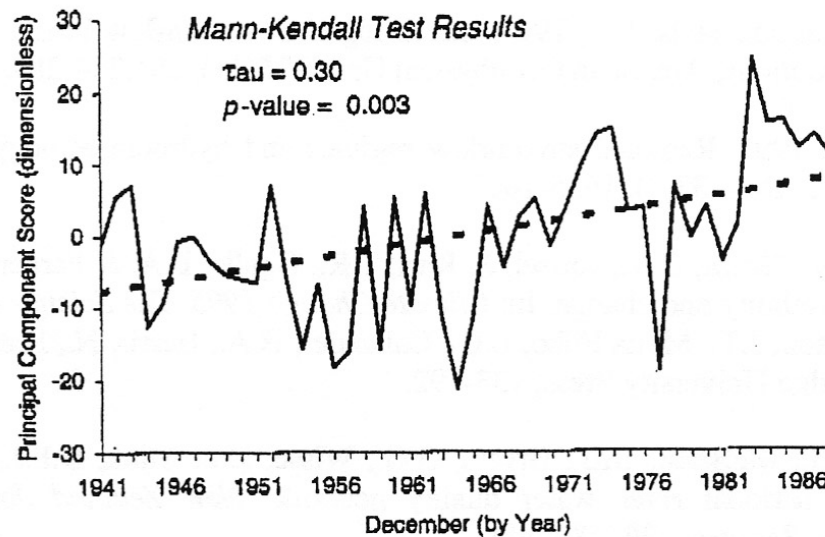


Fig. 8.3. The Upper Mississippi pattern of streamflow variability in December (A). Contoured values are principal component loadings; contour interval is 0.15. Temporal variation in the principal component scores for the Upper Mississippi pattern (B), with results of application of Mann-Kendall test to the scores, and fitted trend line (dashed). Source: Lins & Michaels (1994); Lins (1997).

Advantage of this approach: Provides a comprehensive characterization of spatial/regional trends based on the inherent regional variability in the data; facilitates good graphical presentation (both in map and plot form).

Disadvantage: Requires relatively large spatial time series having no missing values. Is a more complex analysis and requires the analyst to exercise additional care in interpreting and explaining the results.

Discussion

In preparing for an analysis of regional trends, analysts should give as much thought to their problem's spatial aspects as to its temporal elements. The point to this form of trend analysis is intrinsically geographic and one must carefully weigh techniques that ensure and enhance geographical consistency, both in terms of analytical design and graphical output. Such consistency can best be achieved by incorporating, to the extent possible, three elements in each regional trend analysis:

- 1) a map that displays either detailed trend information or provides a geographical reference for the trend results;
- 2) time-series plot that includes a fitted trend line, smoothing curve, or trend test statistic; and
- 3) a robust quantitative estimate of trend direction, magnitude, and significance.

References

- Lins, H.F., 1985. Streamflow variability in the United States: 193 1-78. *J. of Climate and Applied Meteorology*, **24**(5), 463-471.
- Lins, H.F. & Michaels, P.J., 1994. Increasing U.S. streamflow linked to greenhouse forcing. *Eos, Transactions, American Geophysical Union* **75**(25), 28 1,284-285.
- Lins, H.F., 1997. Regional streamflow regimes and hydroclimatology of the United States. *Wat. Resour. Res.*, **33**(7), 1655-1667.
- Nicholls, N., Gruza, G.V., Jouzel, J., Karl, T.R., Ogallo, L.A. & Parker, D.E., 1996. Observed climate variability and change. In: *Climate Change 1995. The Science of Climate Change*, ed. by: Houghton, J.T., Meira Filho, L.G., Callander, B.A., Harris, N., Kattenberg, A. & Maskell, K. Cambridge University Press, 133-192.
- Smith, D.G., McBride, G.B., Bryers, G.G., Wisse, J. & Mink, D.F.J., 1996. Trends in New Zealand's national river water quality network. *New Zealand Journal of Marine and Freshwater Research*, **30**, 485-500.

CHAPTER 9

TESTING FOR CHANGE IN VARIABILITY AND PERSISTENCE IN TIME SERIES

Geoffrey G.S. Pegram

9.1 Introduction

Non-stationarity can take many forms. For sequences of independent random variables these forms include

- shift in mean from one level to another
- trend, linear or otherwise
- heteroscedasticity (change in variability).

The above may be tested using conventional tests.

For time series that exhibit dependence, non-stationarity becomes harder to test. In these cases, possible forms of non-stationarity are

- shift in mean (*)
- trend (*)
- heteroscedasticity (#)
- change in correlation structure (a).

The first two of these (*) are first order problems. They may be treated using the same techniques as for independent sequences, but with appropriate adjustments for correlation. This is elaborated on below. The second pair of problems (#) fall into the group of second order statistical testing. Possible approaches to testing for changes in heteroscedascity, and changes in the serial dependence structure or *persistence* of time series are considered in the following sections. It should be noted that there is currently very little in the literature on these two issues.

9.2 Techniques employed

This section details suggested statistical methods, tests and graphical approaches that may be of use when testing for change in variability and persistence in time series. Use of the methods is illustrated via examples in the final section.

9.2.1 Symmetrizing transform (Box-Cox)

It is often useful to symmetrize the data before testing (e.g. prior to application of Moolman's test (Moolman, 1985)). This is important for parametric tests that are often based on the assumption of normality, but is less important for non-parametric tests. The transform is:

$$x_T = \begin{cases} (x_T - 1) / T \\ \ln(x) \text{ if } T = 0 \end{cases} \quad (9.1)$$

Choose T so that $(\text{Mean}(x_T) - \text{Median}(x_T))/d_F \cong 0$, where d_F is the fourth spread of x_T , i.e. the distance between the lower and upper quartiles.

9.2.2 Moolman 's modification of the t-test

It is not obvious how to test for change in mean or trend in a time series exhibiting serial correlation. Moolman's (1985) contribution was to adapt the usual t-test for shift in means or trend development to normal AR(1) processes with parameter ϕ . He derived the relation that

$$\alpha(\phi) = 1 - \Phi[z_\alpha(1 - \phi)] \quad (9.2)$$

where z_α is the upper critical 100 α % value under independence, $\Phi(\cdot)$ is the cumulative standard normal and $\alpha(\phi)$ is the significance level of the dependent series. He found the t-test more powerful than the U-test (Wilcoxon, 1945; Mann & Whitney, 1947) in detecting shift and trend and that normality was approached by $n = 70$ (see also the example in the final section of this chapter). This technique provides an alternative to the non-parametric tests presented elsewhere in this report.

9.2.3 Testing for variability and persistence

The sample variance and first serial covariance of a zero-mean sequence are given by:

$$s^2 = \sum x_i^2 / n \quad \text{and} \quad (9.3)$$

$$\phi s^2 = \sum x_i x_{i+1} / n \quad (9.4)$$

where ϕ is the first serial correlation coefficient. These are summary (global) statistics and are averages of individual elements computed from the sample.

A possible approach to testing for changes in variability and persistence is to consider the terms $v_i = x_i^2$ and $c_i = x_i x_{i+1}$. The $\{v\}$ and $\{c\}$ series are plotted, and non-parametric tests are applied to the segments of the series that are considered to be different.

An additional statistic that can be used for testing persistence, useful especially when the sample is suspected of being drawn from a non-stationary process, is based on the variogram at lag-one which is given by:

$$g = \sum (x_i - x_{i+1})^2 / n \quad (9.5)$$

where the sequence $\{x\}$ does not need to be zero-mean. As with the $\{v\}$ and $\{c\}$ series, the $\{g\}$ series, made up from the individual elements of g , can be plotted and then non-parametric tests used to detect change. Indications are that, in tests on persistence, the elemental variogram g is more powerful than the covariance c for detecting change in correlation in a time series (see example in final section of this chapter).

A multivariate extension particularly useful for testing for change in a basin or region, is to compute the sequence $V_t = \text{cov}(x_t, x_{t+1})$, which is the covariance between the vectors of successive observations of streamflows (monthly or annual) for example. This can be generalised to the variance and the variogram if desired.

9.2.4 Grouping by threes to visualize and test for variability

A fairly simple method of visualizing the variation within a time series is to successively group, and average by threes, the elemental statistics such as $\{v\}$, $\{c\}$ or $\{g\}$. This is easily achieved using a spreadsheet. The averages of each of the successive sequences is equal to the global variance, covariance or variogram of the sample. The successive grouping of the elements assists in visually detecting shifts and trends in these statistics.

The approach is illustrated below for the case of the covariance, c . Let

$$C_i = x_i x_{i+1}, \quad i = 1, 2, \dots, n-1 \quad (9.6)$$

and define

$$R3_j = (c_{3j-1} + c_{3j} + c_{3j+1})/3, \quad j = 1, 2, \dots, (n-1)/3 \quad (9.7)$$

$$R9_j = (R3_{j-1} + R3_j + R3_{j+1})/3, \quad j = 1, 2, \dots, (n-1)/3 \quad (9.8)$$

etc.,

with the final sequence containing at least 3 groups. For example, if $n = 244$, the last grouping will be $R81$ with three elements in it. Of course, the cascade can be computed with any number of values forming the sequences of separate partial sums, but the symmetry associated with the odd numbers centres the successive sums at the mid-points of the intervals they summarize which is nice for plotting.

Once such a plot has been made, then it is relatively straightforward to decide whether ϕ is stationary or not. Where the R values appear to exhibit a trend, this can be treated by parting the series in two and comparing the two estimates of the serial correlation from the two sub-samples (found by averaging the covariances c_i), using a t-test with Moolinan's modification for correlation *between the c_i values*. A more convenient way may be to use a non-parametric test on parts of the sequence $\{c\}$.

9.2.5 Windowing

An alternative to grouping by threes is a moving window wherein the statistic of interest is calculated. This is particularly appropriate in the calculation of the Hurst coefficient, for example. A development and exposition of this technique is given by Radziejewski & Kundzewicz (1997).

9.2.6 Exponential filter

To examine change in the occurrence of flood-producing rainfall, one needs to extract the annual maxima of the accumulation of rain, not the 1 hour or 1 day maxima. Exponentially filtering, with a mean of 5 days for example, will yield maximum accumulations which are more appropriate for flood analysis. These can then be examined for change.

9.2.7 The bootstrap

A method which has promise when used in combination with the sequences of elemental variances, covariances or variograms is the bootstrap. it is used here to compare the slopes of the linear trend lines fitted to the $\{v\}$, $\{c\}$ or $\{g\}$ values. Because of linearity, these trends are almost identical to the trend lines fitted through the derived R values (equations 9.7 and 9.8). Again, the method is illustrated for the case of the covariance.

- 1 The first step is to take the series in question and estimate the linear trend-line of c by least-squares.
- 2 The next step is to analyse the time series, assuming that it is stationary, and fit the appropriate model $ARMA(p, q) : p, q = 0, 1 \text{ or } 2$, employing the usual Box-Jenkins approach. Take the residuals $\{\alpha_{ij}\}$ which (under the null hypothesis) form an independent sequence derived from this analysis. From these, by sampling with replacement, generate a large odd number (say 101) of bootstrap sequences $\{x_{ij}\}^*$ of the same length as the original sample, using the parameters of the ARMA model estimated from the original sample. Hereon one can choose how to test for trend (see below for an example based on pairwise covariances using the grouped by threes approach).
- 3 In exactly the same way as in the original sample, calculate the bootstrapped $\{c_{ij}\}^*$ pairwise covariances and fit linear trend lines through them with slopes b^* . Do this 101 times.

- 4 Accumulate all the 101 b^* values and compute their mean, median, standard deviation and fourth spread.

The resulting statistics give the distribution of b under the null hypothesis of the sequence being stationary and provide a guide as to whether b measured in the original sample is significantly different from zero.

9.3 Examples

In this section, the methods described above are illustrated using a set of examples.

9.3.1 An artificial sequence

Fig. 9.1 is a plot of an artificial, standardized sequence of length 244, which was generated with a first serial correlation coefficient, r , that increases linearly from 0.0 to 0.4 over the length of the sequence. This sequence will be used to examine how to detect trend. It can be seen from the trace in the figure that the differences between successive values reduce as the correlation increases, as expected.

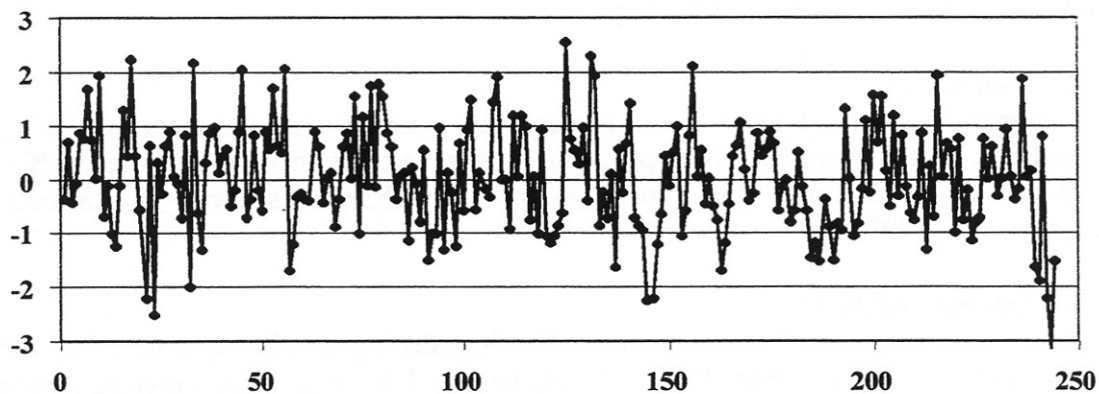


Fig. 9.1. Artificial standardized normal time series generated with r increasing linearly from 0.0 to 0.4.

Fig. 9.2 shows the pairwise covariances $c_i = x_i x_{i+1}$ for the sequence shown in Fig. 9.1. The trend line is fitted using linear least squares. Fig. 9.2 also shows the c_i values successively grouped by threes (3, 9, 27, 81) with a linear trend fitted through the R_{27} values (averaged values of the successive 27-long sub-sequences). This gives a close approximation to the underlying trend.

The bootstrap resampling of the trend-line fitted through the pairwise covariances of 101 sequences computed using the algorithm outlined above yields the following statistics of the 101 slopes:

- Mean 0.00015
- Standard deviation 0.00128
- Upper 95% confidence limit 0.00277

The observed value for the original series is 0.0017. Comparing this with the above statistics shows that the trend of the covariances is not significant at this level. This is confirmed by examining the list of slopes of the bootstrap samples, for which 10 of the 101 slopes exceeded the sample value.

Fig. 9.3 shows the pairwise lag-one elements (all positive) of the variogram grouped by threes. It can be seen that the slope of the trend line is negative and confirms that the higher the persistence, the smaller the differences between successive values. The bootstrap samples of the slopes of the elemental variogram values gives a lower 95% confidence limit of -

0.00449 which is just above the sample value of -0.0048. Only 2 of the 101 bootstrapped slopes were below this value, lending support to its significance.

From this exploratory calculation on an artificial sample, it seems that the sequence of variogram elements is more sensitive to the detection of trend than is that of the covariance.

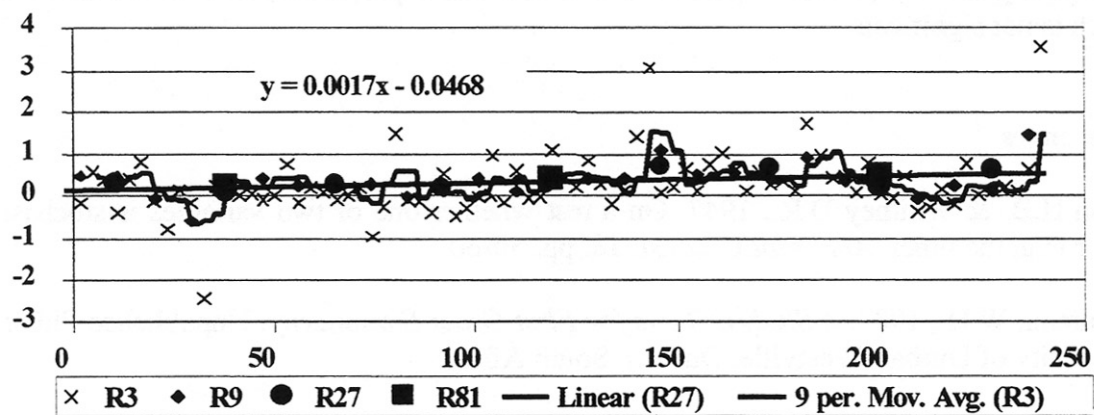


Fig. 9.2. Pairwise covariances grouped by threes for the sequence in Fig. 9.1.

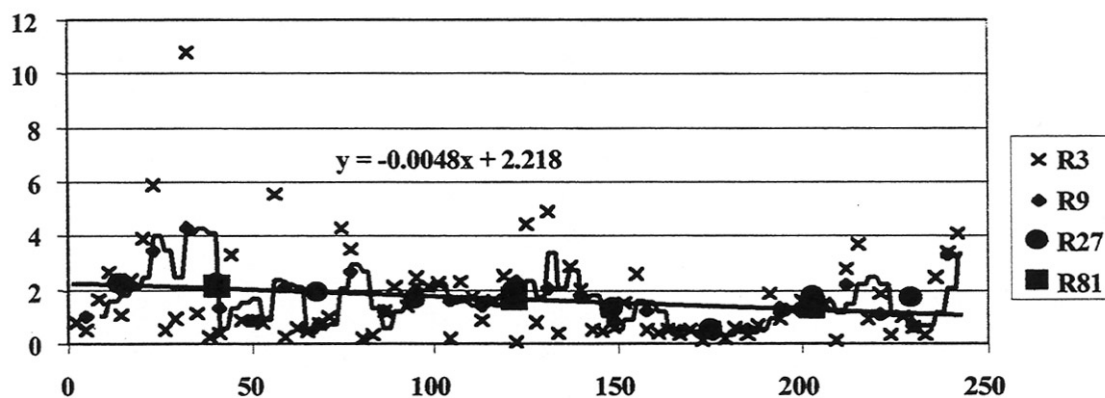


Fig. 9.3. Pairwise lag-one variogram grouped by threes for sequence in Fig. 9.1.

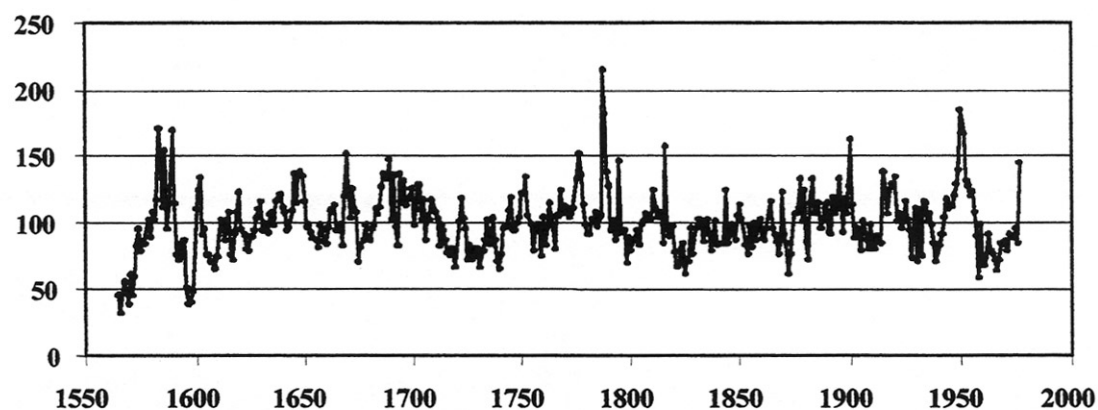


Fig. 9.4. Tree-ring data from South Africa.

9.3.2 Tree ring data from South Africa

Figure 9.4 presents a sequence of raw tree ring data that Moolman used in his thesis (1985). Performing a Box-Cox symmetrizing transform with exponent 0.459 yields a value of 0.619 for the serial correlation. To test whether the mean of the first 20 years is the same as the remainder under persistence, Student's t-test is calculated assuming unequal variances. This yields a one-tailed probability of exceedence of 0.0011 which has a z-value of 2.29. Multiplying this by (1-0.619) gives 0.872 which has a probability of exceedence of 0.192, which is not significant.

References

- Mann H.B. & Whitney D.R., 1947. On a test whether one or two variables is stochastically larger than the other. *Ann. Math. Statist.* **18**, pp. 50-60.
- Moolman, W.H., 1985. *A Study of Tests for First Order Stationarity*. Unpublished PhD thesis, University of Durban-Westville, Durban, South Africa.
- Radziejewski, M. & Kundzewicz, Z.W., 1997. Fractal analysis of flow in the river Warta. *J. of Hydrol.*, vol. **200**, pp 280-294.
- Wilcoxon, F., 1945. Individual comparisons by ranking methods. *Biometrics*, **1**, pp. 80-83.

CHAPTER 10

SEGMENTATION

Piere Hubert

10.1 Introduction

A possible manifestation of nonstationarity in time series is the existence of some modification of their statistical parameters, and especially a sudden change of the mean. Series with such a change may exhibit a strong temporal persistence, with high values of the Hurst coefficient, but poorly fit any autoregressive model.

Some classical tests, (Pettitt, 1979; Buishand, 1982) help in detecting a possible change point of the mean so that the original nonstationary series can be split into two stationary sub series. The Bayesian procedure defined by Lee & Heghinian (1977) supposes the *a priori* existence of a change of the mean somewhere in the series and yields at each time step an *a posteriori* probability of mean change.

Yet these classical approaches seek one change point in the original series. To go further and to explore multiple singularities, a segmentation procedure of time series has been developed (Hubert, 1997). It yields an optimal partition, from a least squares point of view of the original series into as many subseries as possible. The Scheffe test of contrasts ensures that all differences between two contiguous means remain simultaneously significant. The main problem has been to master the combinatory explosion while exploring the tree of all possible segmentations of a series.

10.2 A procedure of series segmentation

10.2.1 Definitions

Given a time series composed of n numerical values:

$$x_i \quad i = 1, 2, \dots, n$$

A series x_i , $i_1 \leq i \leq i_2$ where $i_1 \geq 1$ and $i_2 \leq n$ is called a segment of the initial series. Each division of the initial series into m segments constitutes an m -order segmentation of this series. Thus, given a particular m -order segmentation of the series, and given i_k , $k = 1, 2, \dots, m$, the rank in the initial series of the extreme end of the k -th segment (by convention, we will pose $i_0 = 0$),

$$i_0 = 0 < \dots < i_k < \dots < i_{m-1} < i_m = n$$

one can note $n_k = i_k - i_{k-1}$ the length of the k -th segment, and X_k its mean (local mean):

$$X_k = \left(\sum_{i=i_{k-1}+1}^{i=i_k} x_i \right) / n_k \quad (10.1)$$

Let

$$d_k = \sum_{i=i_{k-1}+1}^{i=i_k} (x_i - X_k)^2 \quad (10.2)$$

and define the quantity:

$$D_m = D(i_1, i_2, \dots, i_m) = \sum_{k=1}^{k=m} \sum_{i=i_{k-1}+1}^{i=i_k} (x_i - X_k)^2 = \sum_{k=1}^{k=m} d_k \quad (10.3)$$

as the quadratic deviation between the series and the considered segmentation. This deviation depends only, for a given series, on the adopted segmentation.

For $m = 1$ and $m = n$, there is only one possible segmentation:

$$D_1 = D(i_1) = n\sigma^2 \quad i_0 = 0 \quad i_1 = n \quad (10.4)$$

where σ is the standard deviation of the initial series and:

$$D_n = D(i_1, i_2, \dots, i_n) = 0 \quad i_0 = 0 \quad i_1 = 1, \dots, i_n = n \quad (10.5)$$

For any order m between 1 and n , there exist several possible segmentations.

10.2.2 Enumeration of the segmentations of a series

Given $N(n, m)$ the number of m -order segmentations of a series of length n . The number of segmentations is equal to the number of:

$$N(n, m) = C_{n-1}^{m-1} = \frac{(n-1)!}{(m-1)!(n-m)!} \quad (10.6)$$

The number of possible segmentations is very high. So, it is definitely impossible to look for an optimal segmentation by simple enumeration of all possible cases and it is necessary to define an optimization algorithm.

10.2.3 Optimization algorithm

The m -order segmentations of series of length n can be organized like the branches of a tree. The length of the first segment can take on a value between 1 and $n-(m-1)$ because the initial series, diminished from the first segment should be divided into $m-1$ segments. The choice of the first segment length constitutes the first level of branching. n_1 is this first choice. We are now faced with the problem of the $m-1$ order segmentation of a $n-n_1$ length series. n_2 will be the second segment length, the value of which will be bounded by 1 and $n-n_1-(m-2)$. One can thus continue until the choice of the length of the $(m-1)$ -th segment, $(m-1)$ -th and last level of branching because the length of the m -th segment is then entirely defined as:

$$n_m = n - n_1 - n_2 \dots - n_{m-1} \quad (10.7)$$

The branching can be generated systematically by increasing first the length of the segment corresponding to the deepest possible level. Taking into account the number of segmentations of a time series, even of modest size, one will bypass exhaustive exploration of the branching and the corresponding combinatory explosion by means of a *branch and bound* type algorithm.

This algorithm permits one to obtain the optimal segmentations of successive orders of several tens of terms in less than one minute on a PC permitting thus a conversational use. However, it is unable at the present state to tackle series of much more than hundred terms like the dendrochronological ones. The combinatory explosion dominates then, but

introducing some heuristics about a first guess of the optimal segmentation value would improve the algorithm, shortening dramatically the segmentation tree exploration.

10.2.4 Validity test of the segmentations

The described algorithm makes it possible, for a given order, to determine the optimal segmentation in the least squares sense. This procedure should be completed by the introduction of a constraint applying to the segmentations produced. This will be acceptable only if the means of two contiguous segments are significantly different.

$$X_k \neq X_{k+1} \quad k = 1, 2, \dots, m-1 \quad (10.8)$$

A complete segmentation being produced, this can be tested using the concept of contrast introduced by Scheffe (1959).

The Scheffe test has been integrated in the optimization algorithm in order to verify the validity of the segmentations concerning the entire chronological series under study. During exploration of the m -order segmentations branching of the complete series, a segmentation whose square root deviation from the series is inferior to the weakest square root deviation already obtained is not retained as a new optimal provisional segmentation except if the null hypothesis of the Scheffe test is rejected at a chosen level of confidence for all the above defined contrasts. The optimal solution supplied by our algorithm will thus be necessarily valid according to the Scheffe test.

The Scheffe test will supply a means of limiting the order of segmentation. The simple criterion of deviation does not permit it because the rest of the deviations between the complete series and the best segmentation is not increasing with m . Considering a series of length n , the optimal m -order segmentation can thus never be preferable to the optimal $(m+1)$ -order segmentation. One would then be led to pursue the process of segmentation up to the n order, for which the deviation is zero, but which does not present any interest at the level of the interpretation because it brings one back to the series under study. However, if during the $(m+1)$ -segmentation process, no segmentation produced shows itself to be valid according to the Scheffe test, one retains the optimal m -order segmentation as the best segmentation of the proposed series.

This use of the Scheffe test nevertheless suffers from a drawback as pointed out by Bernier (1993). It would be applied to segments randomly chosen independently of the data. It is applied to dates of change determined a posteriori and there is a risk that a number of such changes can be fallaciously significant, so it appeared useful to make some simulations in order to have a better idea of the segmentation procedure credibility.

10.2.5 The credibility of the procedure

The procedure of segmentation can be regarded as a test of stationarity, the affirmation that *the series under study is stationary* constituting the null hypothesis of this test.

If the procedure does not produce any acceptable segmentation of the order greater than or equal to 2, the null hypothesis will be accepted. It will be rejected in the opposite case, but there obviously exists a risk of the first kind error (which consists of rejecting the null hypothesis when it is valid), as a consequence of the same risk in the application of the Scheffe test to the different contrasts of a segmentation. However, no attempt to attribute a level of significance to this test of stationarity will be made because such a level of significance, although being a function of that of the Scheffe test, does not depend on it in a simple way. For a given significance level of the Scheffe test, one can only propose an empirical observation of the first kind errors when applying the procedure of segmentation to synthetic stationary series. A hundred series of 50 normally distributed random values

according to $N(0,1)$ have been tested in order to establish a reference base for the hydrometeorological series of a comparable length which will be further studied.

When the Scheffe test is done at the 0.05 significance level, 47 of the 100 synthetic stationary series are segmented which signifies that the procedure rejects the hypothesis of stationarity for these series. If the Scheffe test is applied at the 0.01 significance level, this number is no more than 11, which means that the *noise* associated with the procedure is reduced to an acceptable level. Thus, this significance level will be used further unless otherwise stated.

In our first study of hydrometeorological series the main danger was to admit the non stationarity of series which were in fact stationary. The purpose of the above simulation approach is to determine, in this case, the significance level of the procedure for a given significance level of the underlying Scheffe test. But this approach does not exhaust the subject; first of all because it concerns only a simulation and is not a demonstration; next because it does not say anything about the significance level of the procedure applied to non-stationary series (composed of several stationary sequences) the analysis of which also carries, obviously, a risk of error of the first kind. Such simulations would have to be further pursued.

10.3 Applications

The segmentation procedure was first applied (Hubert *et al.*, 1989) to West-African hydrometeorological series (33 rainfall series plus Senegal and Niger discharge). A previous study (Hubert & Carbonnel, 1987), using the Bayesian procedure of Lee & Heghinian (1977) suggested that the end of the sixties was a major change point for West African hydrometeorological annual time series, the mean then decreasing abruptly by about 30 %, but apart from this major change, some series exhibited a relatively minor change at the end of the forties.

These observations led us to devise the previously described segmentation procedure. Its applications to our series yielded very consistent results all over the region. Five successive and contrasted climatic phases have been determined in this region covering the time span from 1905 to 1985: before 1923 (dry), from 1923 to 1935 (wet), from 1936 to 1950 (dry), from 1951 to 1970 (wet), and after 1970 (dry). This study showed regional consistence of the results, local random variations of rainfall series being rubbed out

The study of African rivers discharge (Fig. 10.1) has been updated and enlarged (Hubert & Carbonnel, 1993; Hubert *et al.*, 1998). The segmentation procedure appears quite robust and the addition of new data does not modify the change points already determined.

The segmentation procedure has been used in other studies regarding African climate such as Cbaouche (1988), Moron (1992), Laraque *et al.* (1997), Paturel *et al.* (1997), Servat *et al.* (1997) and applied to Tunisian rainfall series by Kebaili-Bargaoui (1990). It was also used concurrently with other methods by Slivitzky & Mathier (1993) in a study devoted to the twentieth century climatic changes in the Laurentian great lakes. Some applications have also been devoted to East European rainfall time series in Romania and Bulgaria (Carbonnel & Hubert, 1994; Carbonnel *et al.*, 1994) and to climate evolution in Bolivia (Ronchail, 1996).

The Segmentation Procedure software is freely available on the web at the address: <http://www.cig.enscm.fr/~hubert> with instructions for use in English, French and Portuguese and a sample dataset. The Segmentation Procedure has also been included in various software packages such as TSA1 of the Norwegian Hydrological Service and KHRONOSTAT developed by IRD-ORSTOM (See the Software chapter in this volume).

10.4 Conclusion

The purpose of this paper was mainly to present a first view of the segmentation procedure. This procedure has been assessed by Bernier (1993) who pointed out that the point changes determined by segmentation were optimal according to the general Bayesian procedure he proposed. Cavadias (1992, 1993) quoted this segmentation procedure in his survey of current approaches to modelling of hydrological time series with respect to climate variability and change.

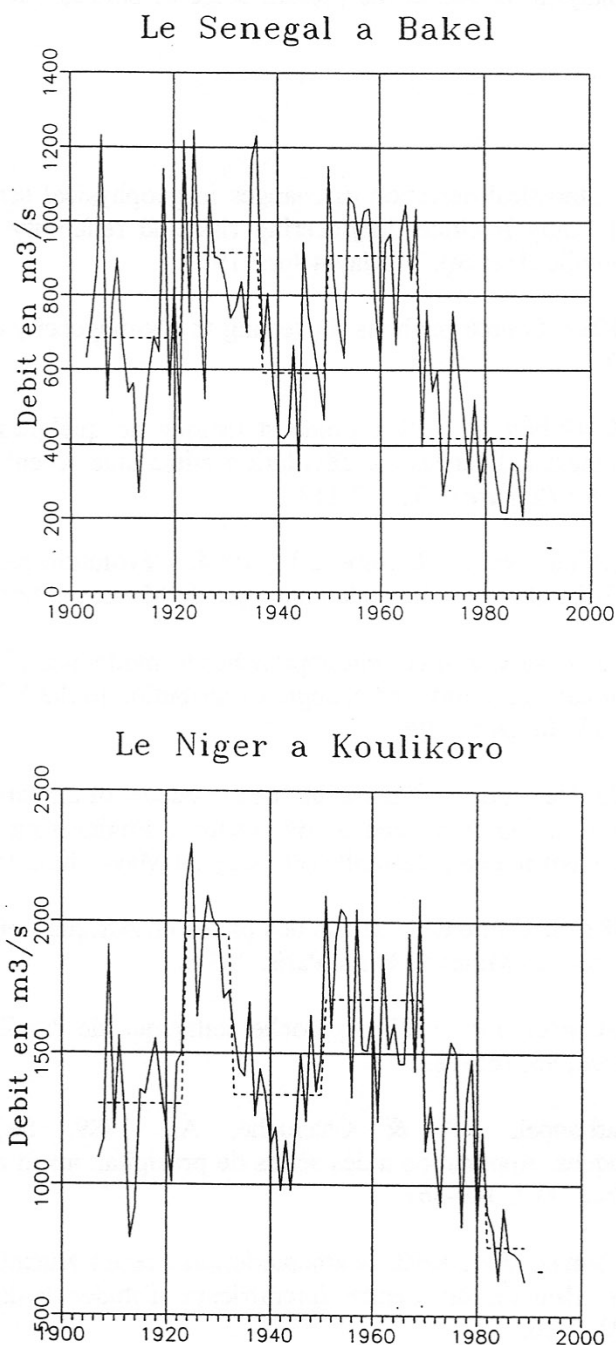


Fig. 10.1. Optimal segmentations of the mean annual discharge for rivers Senegal at Bakel (1903-1993) and Niger at Koulikoro (1907-1992). The procedure automatically stops at the five order for Senegal data and at the six order for Niger data.

From experience gathered in the course of different applications to various time series, the segmentation procedure appears to give reliable results, comparable to those of classical and Bayesian methods. The basic algorithm would be improved in order to process long time series or series exhibiting missing values and more simulations would clarify its significance level estimation. Nevertheless, the segmentation procedure appears to be a useful robust tool for preliminary analysis needed at the present stage of studies into climate variability and change.

References

Bernier, J., 1993. Statistical detection of changes in geophysical series. Contribution to the NATO Advanced Study Institute, Engineering risk and reliability in a changing physical environment, Deauville (France), 24 May-4 June 1993.

Buishand, T.A., 1982. Source methods for testing the homogeneity of rainfall records. *J. of Hydrol.*, **58**, 11-27.

Carbonnel, J.P. & Hubert, P., 1994. Etude statistique de quelques séries pluviométriques bulgares et roumaines; incidences sur l'évolution climatique récente, *Romanian Journal of Hydrology and Water Resources*, **1**, 107-113.

Carbonnel, J.P., Mihnevski, N. & Hubert, P., 1994. L'évolution récente de la pluviométrie dans le massif de Rila. Observatoire de la montagne de Moussala, Sofia (Bulgaria), **2**, 52-58.

Cavadias, G., 1992. A survey of current approaches to modelling of hydrological time series with respect to climate variability and change. Contribution to the WMO workshop on WCP Water, Geneva, 3-13 November 1992.

Cavadias, G., 1993. Detection and modelling of the impact of climatic change on river flows. Contribution to the NATO Advanced Study Institute, Engineering risk and reliability in a changing physical environment, Deauville (France), 24 May-4 June 1993.

Chaouche, A., 1988. Structure de la saison des pluies en Afrique Soudano-Sahélienne. These de doctorat de l'Ecole des Mines de Paris, Paris, 263 p.

Hubert, P. & Carbonnel, J.P., 1987. Approche statistique de l'hydroclimat de l'Afrique de l'Ouest. *J. of Hydro.*, **95**, 165-183.

Hubert, P., Carbonnel, J.P. & Chaouche, A., 1989. Segmentation des series hydrométéorologiques. Application a des series de précipitations et de debits de l'Afrique de l'Ouest. *J. of Hydrol.*, **110**, 349-367.

Hubert, P. & Carbonnel, J.P., 1993. Segmentation des series annuelles de debits des grands fleuves Africains. Bulletin du Centre Inter africain d'études hydrauliques, Ouagadougou (Burkina Faso), **92**, 3-10.

Hubert, P., 1997. Change points in hydrometeorological time series. Proc. Conf. Applications of Time Series Analysis in Astronomy and Meteorology, ed. by: Rao, Priestley & Lessi, Chapman and Hall, 399-412.

Hubert, P., Servat, E., Paturel, J.E., Kouame, B., Bendjoudi, H., Carbonnel, J.P. & Lubes-Niel, H., 1998. La procedure de segmentation, dix ans après..., *Water Resources Variability in Africa during the XXth Century*, Proceedings of the Abidjan Symposium, ed. by: Servat, E., Hughes, D., Fritsch, J.M., Hulme, M., IAHS Publication No. 252, 267-274.

Kebaili-Bargaoui, Z., 1990. Modélisation stochastique des sécheresses annuelles et pluriannuelles. These de doctorat d'état, Faculté des Sciences de Tunis, Tunis (Tunisie), 335 p.

Laraque, A., Olivry, J.C., Orange, D. & Marieu, B., 1997. Variations spatio-temporelles des regimes pluviométriques et hydrologiques en Afrique centrale du debut du siècle a nos jours. In: *FRIEND'97-Regional Hydrology Concepts and Models for Sustainable Water Resources Management*, Proc. of the Postojna (Slovenia) symposium, ed. by: Gustard, A., Blazkova, S., Brilly, M., Demuth, S., Dixon, J., Van Lannen, H., Liasat, C., Mkhandi, S. & Servat, E., IAHS Publication No. 246, 257-266.

Lee, A.F.S. & Heghinian, S.M., 1977. A shift of the mean level in a sequence of independent normal random variable. A Bayesian approach, *Technometrics*, **19**, 503-506.

Moron, V., 1992. Variabilité spatio-temporelle des précipitations en Afrique sahélienne et guinéenne de 1933 a 1990. *La Météorologie*, Paris, Série V 43-44, 24-30.

Paturel, J.E., Servat, E., Kouame, B., Lubes, H., Ouedraogo, M. & Masson, J.M., 1997. Climatic variability in Humid Africa along the gulf of Guinea. Part 2: An integrated regional approach. *J. of Hydrol.*, **191**, 16-36.

Pettitt, A.N., 1979. A non-parametric approach to the change-point problem. *Appl. Stat.*, **28**, 126- 135.

Ronchail, J., 1996. Variabilité pluridécennale des précipitations en Bolivie: essai de mise en relation avec les temperatures de surface océaniques de l'Atlantique extratropical. *Publications de l'Internationale de Climatologie*, **9**, 504-511.

Scheffe, M., 1959. *The Analysis of Variance*. Wiley, New -York, 4.77 p.

Servat, E., Paturel, J.E., Lubes, H., Kouame, B., Ouedraogo, M. & Masson, J.M., 1997. Climatic variability in Humid Africa along the gulf of Guinea. Part 1: Detailed analysis of the phenomenon in Côte d'Ivoire. *J. of Hydrol.*, **191**, 1-15.

Slivitsky, M. & Mathier, L., 1993. Climatic changes during the 20-th century on the Laurentian great lakes and their impacts on hydrological regime. Contribution to the NATO Advanced Study Institute, Engineering risk and reliability in a changing physical environment, Deauville (France), 24 May-4 June 1993.

CHAPTER 11

CRITERIA FOR THE SELECTION OF STATIONS IN CLIMATE CHANGE DETECTION NETWORKS

Paul Pilon

Many hydrological variables could be of importance to studies of climate change and variability (WMO, 1988; Peterson *et al.*, 1997; GCOS, 1998). Lawford (1992) provides the view that many variables within the natural sciences could be of importance to such studies. However, Burn (1994, P. 28) points out that many such variables may not have records available for a large number of locations nor have a sufficient length of record. Intuitively, there must be sufficient data and such data must also be of a suitable quality for them to be of benefit in studies of detection of change. He selects streamflow as the hydrological variable of choice in hydrology for detecting the impacts of climate change as “it represents a basin integrated response to hydrological inputs and therefore affords good spatial coverage.” There are, however, other variables that could be of importance for establishing the biological and chemical impacts of climate change on natural systems.

In essence, hydrometric or streamflow data are seen as very important for furthering our understanding of physical processes and their feedback within the water cycle and climate change. It has also been suggested (Pilon *et al.*, 1991; Burn, 1994) that certain responses of the hydrological cycle to climate change may be “hydromagnified”, such that elements within hydrometric data may be better able to elucidate change than traditional climatic variables. In order to ascertain which hydrometric sites may be most suitable for use in the detection, monitoring, and assessment of climate change, a set of selection criteria must be identified and applied to available monitoring stations. This paper reviews a number of efforts in the creation of specific or “specialised” observational networks, thereby illustrating selection criteria of importance to those interested in studies of climate change and variability. These case studies cover surface climate, hydrometric, and water quality observational systems. In-depth material is presented on the selection criteria used within each. One potential exception to the primary focus of the networks described below is the USGS National Stream Quality Accounting Network, wherein many of the sites are chosen to evaluate and track sufficiency of regulatory and pollution prevention measures on the environment, with some sites being potentially suitable for climate change studies. This network does, however, provide an excellent example of a national network for determining human and terrestrial impacts on the aquatic environment.

11.1 Surface climate networks

WMO (1966 and 1986) have established criteria for the WMO Reference Climatological Stations (RCS) network. The RCS network comprises sites from the existing national networks of climatological stations. Its purpose is to facilitate the detection and accurate evaluation of climate change. This network is comprised of stations with relatively long records, whose data are available, verified and “homogeneous”. The network reflects a wide distribution of sites throughout the world.

WMO (1966) defines homogeneity of climatological data as “intending to mean uniform representativeness of the data for conditions in rather large geographical areas.” The homogeneity of the data is compromised through the alteration of local conditions such as urbanisation, landscaping, reforestation, and other changes to the local environment. It is also compromised through changes in instrumentation, exposure of instrumentation, location of

instrumentation, and change of observing times. Homogeneity within the climate record is a very serious problem that jeopardises the utility of data for detection studies. Much effort is dedicated to making climatological records homogeneous through the correction of archived data. WMO (1966) suggests the use of regionally averaged climatological series as indices superior to individual station series, especially for highly variable elements such as precipitation. This regional averaging is suggested, in part, due to the complications associated with “homogenising” the data series.

WMO (1986) provides the following criteria for the preliminary identification of a candidate RCS. The criteria are:

- permanency—the existence of the site into the foreseeable future;
- location—in an environment unaffected by densely populated or industrialised areas;
- quality assurance—trained observers, reliable and calibrated equipment, regular inspections, technical servicing with back up equipment;
- longevity - records should span as long a period as possible;
- homogeneity—as few as possible significant relocation of instruments, changes of observing times, instruments, and exposure, or observing techniques;
- quality control - data should go through the strictest procedures; and
- measure a minimum set of climate elements of either mean temperature or precipitation, including preferably minimum and maximum temperature.

The RCS network aspired to have a density of 2 to 10 stations per 250,000 km². However, this density would not be considered adequate for determining the spatial coverage of precipitation.

WMO (1993) documented the efforts made in various countries regarding the establishment of a RCS network. It was reported that Canada had developed additional criteria that included at least 30 years of “continuous” data and with no data gaps exceeding 4 years. The data have to be homogeneous in maximum and minimum temperature, or correctable. Canada (Environment Canada, 1996) considered only sites with both daily temperature and precipitation records, which happened to be the vast majority of the Canadian network. During the late 1980s, it took approximately five years to complete a national review that culminated in the identification of 254 stations as members of the Canadian RCS network. An archive of corrected or adjusted observations is not available. WMO (1993) does indicate, however, that the adjustments are made to the data from stations within the US Historical Climatology Network (USHCN). Corrections to the data are applied for: time of observation bias; station changes; urban heat island effect various temperature system effects; precipitation including rain gauge under-catch; and other discontinuities. This list of corrections exemplifies the problems encountered in using climate records for detecting potential anthropogenically induced climate change.

Peterson *et al.* (1997) describe the process used within ICOS to identify candidate stations for the Global Stations Network (GSN). A multiple step process was used to create a list of about 1000 stations. This process and its selection criteria are now described. The initial step was to identify potential sources of climatological data. The sources used included the Global Historical Climatology Network (GHCN) (Vose *et al.*, 1992), which is similar to the aforementioned USHCN, and a data set from the United Kingdom (Jones, 1994). These data sets contain 7,283 and 2,525 stations, respectively, after removal of a subset of USA sites for duplicative reasons. Both sets have had some form of homogeneity corrections applied to their data series. The remaining two sources of information comprised the WMO 1961-1990 Normals stations and the lists of the RCS that had been submitted to WMO by Member states. These two sources comprised approximately 3,342 and 2,283 stations, respectively. These latter two sources may contain duplicates between them and with the previous two sources. The combined four sources contained an estimated 8,653 unique

stations. Peterson *et al.* (1997) reported that it was a difficult task to identify potential duplicates among these different sources for a variety of reasons. They devised a ranking scheme to establish the most suitable stations for inclusion in the GSN. This scheme is shown in Table 11.1.

Table 11.1. Ranking Scheme for GSN.

Category	Points	Weighting
Record length	Up to 20	- number of years since 1896 until 1996 multiplied by 0.2
Data quality or homogeneity	Up to 20	- only stations in the first two global data sources (i.e., GHCN and United Kingdom “Jones” data sets) - number of years of homogeneous data multiplied by 0.4
Data quality or homogeneity	Up to 10	- only stations contained in the WMO list of RCS - number of years of homogeneous data multiplied by 0.4
Real-time capability	Up to 20	- no weight given to stations where there have been no data received since 1990
Pristine nature	Up to 20	-20 points allocated to a rural station (population of <10,000) - 15 points for 10,000 - 50,000 - 10 points if population statistics are unknown - none, if population >50,000
Continued longevity	Up to 10	- from 1-4 points assigned if it was part of different global networks (e.g., GUAN, GAW, etc.)
Total	Up to 100	

A complex iterative approach was used to select the highest weighted stations in each geographical area, and with a certain variation in procedures to allow for mountainous conditions. The selection of the network was finalised and contained 1000 stations, having an average quality of 67.6. It is interesting to note that 18% of the final stations within the GSN were in small towns, having a population greater than 10,000, while an additional 18% were in urban centres, having a population in excess of 50,000. Given the rather low reporting rate of current network, as mentioned earlier, it is unfortunate that GCOS (1998) does not make mention as to what percentage of this reduced network may be attributed to the highly populated areas.

Environment Canada (1996) reported recent efforts in redefining its RCS network. Approximately 15 of the original 254 stations had been discontinued. In addition efforts had gone into the identification of additional stations, some of which were located near the discontinued RCS. These efforts resulted in the definition of a Canadian RCS network having approximately 300 stations. It was noted that the spatial coverage in southern Canada is quite good, but as one proceeds north, the density of stations becomes greatly reduced. Regional gaps in the north become evident. It was also noted that many of these sites are either operated by other agencies or by volunteers. Hence, contingency planning was required in order to address the problem of discontinuation of sites from the network. It was recommended that there be a two year overlap at these automated sites, however this was not carried out in most cases due to fiscal and other operational restraints.

11.2 Hydrometric networks

Wallis *et al.* (1991) describe the construction and publication for a combined climate and hydrometric data set for the continental United States. This daily data set includes 1009 US Geological Survey (USGS) streamflow stations and 1036 climatological stations, with missing values estimated. The climatological stations within this data set are a subset of the

USHCN. They removed stations from the USHCN when data were not available in electronic form, more than 20% of the data were missing, very short period of record (e.g., four years) or when monthly temperature or precipitation values were not available. They estimated missing data using ratios from the nearest neighbour, or in some cases the long-term mean was used. Their efforts reduced the 1228 temperature and 1220 precipitation sites of the USHCN to their 1036 temperature and precipitation stations.

Wallis *et al.* (1991) noted that the effects on streamflow by human activities presented somewhat of a different problem than for climate stations. They argued that most gauged streams within the United States “are affected to some extent” by human activity. The most significant effects on natural flows would be upstream containment structures, diversions, and water loss through consumptive uses. They also provided a brief history of activities within the United States to define a network of sites that would be free of water management effects and presumably would represent stable land-use conditions. The USGS developed a “benchmark” network comprising approximately 50 basins. These basins were also relatively small in size. Wallis *et al.* (1991) reported on the earlier work of Langbein & Slack (1982), who had defined a group of 200 sites to evaluate long term streamflow variations. Langbein & Slack (1982) classified stations into three classes. The first class comprised sites with no diversions or regulation. The second class included basins with diversions or storage structures accounting for less than 10% of the annual mean discharge. The third class represented the remaining sites. Wallis *et al.* (1991) reported that unfortunately many of the identified sites are no longer in service, with one site being discontinued in 1913.

Their primary concern in the identification of streamflow sites for detection studies was that the data be “as free as possible from upstream diversions and storage”. They also wanted the data to be suitable for rainfall modelling; hence, the streams had to be unaffected “at the storm response scale”, which to them implied daily data. Another aspect affecting their screening criteria was that the data within the national climatological archive are only available in electronic form since 1948. As their intent was to have comparable periods of coverage, they focused their attention on a similar period of time for the streamflow data. From the USGS daily archive streamflow data, they first screened all sites for those with 40 years of data and for which operations had started prior to 1948. This resulted in 5,000 stations being identified. They then allocated each of these into one of six classes:

- Class I - no upstream diversion or regulation,
- Class II- minimal upstream diversions or regulations,
- Class III - sites where the extent or effects were not known,
- Class IV - sites that were probably unusable as they contained substantial effects,
- Class V comprised major upstream diversions or regulation and such stations were considered practicably unusable, and
- Class VI comprised stations with substantial natural upstream storage (e.g., lakes).

In essence, only 1413 stations from Class I and 11 were retained for further consideration. Sites discontinued prior to 1978 were removed, resulting in the final list of 1009 sites. A process similar to that applied to the climate data was used to estimate missing streamflow data.

Slack & Landwehr (1992), in an update of the efforts of Langbein & Slack (1982), reviewed all USGS streamflow records for the entire USA and its protectorates, through to water year 1988. They outlined their set of criteria for the identification of candidate stations and worked closely with USGS District offices to review each potential candidate. A total of 1659 streamflow stations were identified and are contained in their Hydro-climatic Data Network (HCDN). The data and associated basin characteristics are available on CD-ROM. Slack and Landwehr present a number of maps and charts showing the locations of sites

throughout the USA. They also show figures illustrating the characteristics of these sites, such as distribution by length of record and basin size.

In the development of criteria, Slack & Landwehr (1992, p. 2) argue that “the pattern of past climate variation to be discerned in the streamflow record would be confounded by changes induced by anthropogenic activity.” They also provide an exception to this rule when the “non-climatic forcing factor” is consistent in its application over an identifiable period of record. The time interval of their data is taken such that it must be shorter than the interval of the effect of the climate-forcing factor on the streamflow characteristic. They adopt a criterion that the streamflow characteristics must be representative of the natural or stable conditions, and the flow of the site must be representative of natural conditions at least on a monthly basis. They indicate that upstream controls or diversions must not affect monthly averaged flows. They consider both active and discontinued stations and indicate that the quality of such data is consistent, as all data “are collected by nationally standardized procedures” by the District staff. All data are stored electronically in the USGS national archive.

There are six criteria that must be met for acceptance of stations into the HCDN. The first is that the data must be in *electronic form*, which does not pose a problem within their country. The second is the *breadth of coverage*. Both active and discontinued stations operating prior to 1988 are considered provided they span an entire water year, which is defined as October 1 through September 30. The third criterion is the *length of record*. There must be at least twenty years of suitable record, or less for stations in under-represented geographic or climatic areas.

The fourth criterion is the *accuracy of the records*. The accuracy of the records had to be assigned at least a value of “good”. The occurrence of a few years of a less than good rating did not disqualify a site. Accuracy reflects several factors in the process of determining the discharge for a given site. These factors include, but are not limited to: the accuracy of the stage measurement; the stability of the rating curve; the accuracy and frequency of discharge measurements to establish the rating curve; and the degree of interpretation of records.

The fifth criterion is that there must be *unimpaired basin conditions* affecting the average monthly discharge. This implies there must be: no overt adjustment of streamflow through diversions; no regulation by a control structure; no reduction in baseflow by groundwater extraction; and no change in land use that could significantly affect the monthly value of streamflow. Exceptions were allowed for period of stable record, or for control structures that did not significantly affect the utility of the data, or diversions that were either insignificant or stable with time.

The sixth and final criterion is *measured discharge values*. By this, they infer that the reported discharge values are obtained by the use of USGS national standard procedures. Within these procedures, there is the allowance for the estimation of discharge when certain recording abnormalities arise, if too many estimated values are contained within a given month, then Slack and Landwehr assign the data a quality of less than “good”. The data are, in turn, discarded due to the quality of record criterion. In addition, this criterion allows for no constructed records, which are values of flow obtained by combining known flows of other sites. There are no attempts made to fill-in missing records, other than those values that appear in the national archive as “estimated”.

Pilon *et al.* (1991) document initial efforts within Canada to establish a streamflow network suitable for climate change studies. Stations were selected that had stable land-use patterns, which would not impede the use of the data for their intended purpose. The stations had to have had no known diversions or control structures, and sites had to be listed within the national archive as having been operated for at least fifty years. This condition was not an inference as to the completeness of the record, as some stations could have less than fifty years of data, yet span a time period longer than fifty years. There was also no condition that

the station had to be currently active. The data, of course, were in electronic form. A total of 23 sites were listed as having passed these criteria. It was obvious that this did not represent suitable geographic coverage for Canada, particularly as most of the identified sites were in the southeastern portion of the country.

Environment Canada (1999) developed a series of criteria and applied them systematically to sites within their national hydrometric network. Their work has resulted in the identification of a specialised network termed the Reference Hydrometric Basin Network (RHBN). Their selection criteria and approaches follow closely these previous hydrological efforts, and could be considered to be more similar to that of Slack & Landwehr (1992), but with some modifications.

The criteria used to select stations for the Canadian RHBN include:

1. breadth of coverage (seasonal, continuous, streamflow and lake level);
2. degree of basin development;
3. no significant regulation or diversions;
4. length of suitable record;
5. longevity; and
6. accuracy of the data.

There was no need to include as a criterion the availability of the data in electronic format, as all the data in the Canadian national water archive are in electronic form. As well, there was no specific criterion established regarding the density of the network. In essence, judgement was used in including sites that failed to meet to some extent the established criteria, in under-represented geographic or ecological areas. As well, where several sites existed in close geographical proximity, judgement was used to select the best of those available, based mostly on the selection criteria of breadth of coverage, length of record, longevity, and accuracy of records.

The *breadth of coverage* criterion refers to the types of hydrometric stations to be considered in the analysis. All seasonal, continuous, and lake level stations were considered for further screening. Seasonal stations were included in the analysis, as the prairie region has a number of stations operated on a seasonal basis due to local climatological and physiographic conditions. Such sites are operated from just prior to spring break-up to late fall. It was felt these sites would prove useful for analysis of change related to this portion of the year, thereby greatly increasing the spatial coverage of the network.

Lake levels were also considered as being potentially useful for analysis of the impacts of climate on surface waters and were included in the screening process. Two designations were allowed for lakes. The first was for lakes representing closed drainage systems where no surface water channel exits the body of water. The second represents the more typical lake within an open channel system. No closed drainage systems were identified that met the established criteria and therefore are not represented in the final list of 7 lake level stations in the RHBN.

In addition, under this criterion, only observed discharge values and values estimated through the application of national standard procedures were included. This infers, similar to the HCDN, that no sites with “constructed” records were considered for inclusion in the network. Overall, this “*breadth of coverage*” criterion is different than that set in the establishment of the HCDN, in that the Canadian effort includes seasonal and lake level sites. The second criterion reflected the *degree of basin development*. Stations within the RHBN represent pristine or, as a minimum, stable land-use conditions. No systematic recording of changes in landscape is made by the national hydrological service, hence a subjective assessment was made of the percentage of basin development for each candidate site. Pristine sites were considered as those having less than 10% of the surface area modified in some fashion.

The third criterion was *no significance regulation or diversions* within the river system. The national water meta-database (HYDEX), which is available on CD-ROM, provides information on every station regarding the basin state as being either “natural” or “regulated”. This designation reflects only the physical structures within the waterways upstream of the site. It does not reflect on the use made of land within the basin nor the existence of diversions. The database does not reflect the case of the removal of control structures, an important fact given the large number of small control structures that have been decommissioned. In other words, the “natural” designation in HYDEX does not infer pristine conditions, but it does infer that there are no control structures upstream. For “regulated” systems, the question arises as to whether the degree of regulation is significant. Basins with structures controlling less than 5% of the area of a basin are included in the analysis.

The fourth criterion was *length of record*. A minimum of 20 years was set, with the provision that stations in under-represented geographic, climatic, or ecozones could be considered.

The fifth criterion was *longevity*. This criterion was to reflect the judgement of regional staff that the basin would remain in a pristine or stable state into the foreseeable future. In other words, the station must be currently active and no future activities within the basin would impair the data from its inclusion within the RHBN. This criterion was also to reflect the relative potential, although difficult in periods of fiscal restraint and budgetary decreases, of future funding. Hence, in some cases, preference between potential sites within close proximity may be given to those sites with funding secured for specific purposes such as flood forecasting.

The sixth and final criterion was the *accuracy of the records*. Regional offices of Environment Canada and the Ministère de l'Environnement et de la Faune, for the Province of Quebec, collect streamflow data across Canada to a set of well-defined national standard procedures. However, as is mostly the case in hydrometry, there is no quantitative estimate of the accuracy of a particular published streamflow value. In this initial phase of RHBN development, the accuracy of the data was qualitatively assessed by local experts based on their knowledge of hydraulic conditions at each site, such as the stability of the control and the accuracy of the rating curve. They assigned a nominal score from 1 to 5, representing excellent to poor quality data.

Data accuracy was assessed for both open-water and ice-cover conditions. This approach was taken because certain sites may possess very poor quality winter measurements, but have excellent open water data, or vice versa. Such knowledge may prove valuable in selecting the best stations for studying different seasonal conditions across the country.

Systematic application of the criteria resulted in the selection of 255 hydrometric stations, including 7 continuous lake level, 37 seasonal streamflow and 211 continuous streamflow stations. The majority of RHBN stations are found in the south of Canada. Most of the basins found in the mid-to-northern latitudes are relatively large. There are no RHBN stations north of 70 degrees latitude. Although there are gaps in some other regions of the country (e.g., northern Quebec) the RHBN covers most of Canada's major hydrologic regions. In accordance with their selection criteria, all stations have at least 20 years of record. Sixty percent of the flow stations have more than 30 years of record, while the average suitable record length is 38 years. The longest is 86 years. Basin sizes for the network of 248 flow stations range from 3.9 km² to 145,000 km². The median basin size is about 1100 km²; about 10% of the basins are greater than 20,000 km² in size; about 10% are less than 100 km².

Ice effects influence the data of over 80% of the RHBN flow stations. Most stations where data are not influenced by the presence of ice are found in the southern coastal areas.

An important aspect considered by Environment Canada was the assessment of the accuracy of each station within the selection process. Accuracy of data is considered to be of great pertinence in studies of climate change and variability. Hence, additional attention is given this aspect in the sequel.

11.3 The quantitative assessment of accuracy of hydrometric data

Environment Canada (1999) describe a novel approach, which is of potentially great utility in the assessment of the accuracy of site specific data. They indicate that in the near future, the accuracy of data collected at RRBN stations will be re-assessed using this more systematic and quantitative approach. This approach involves the determination of a single composite index, reflecting overall data quality, that takes into consideration three factors: the stability of the control, the level of accuracy of the rating curve, and the effect of ice cover. The latter aspect is of extreme importance in countries experiencing colder climates. Each of these factors is evaluated to give an individual “factor index” that is then combined to form a “composite index” for a particular station for both open water and periods subject to ice formation. It should be noted that not all sites are necessarily subject to ice formation within Canada, due to local climatological conditions or the hydraulic conditions prevailing at the site. Each factor index is treated as an independent component of accuracy and is given an index value ranging from 0 to 10, with higher values inferring better conditions.

11.3.1 *Stability of the control*

The stability of the control is considered only as a function of the stability of the bed material. A sand bed is considered relatively unstable and is given an index of 2, while a cobblestone bed is given an index of 5, and a solid rock control is given an index of 10. There is flexibility allowed in the allotment of values for certain bed types, as the stability of the channel may change with stage and other conditions.

11.3.2 *Level of accuracy of the rating curve*

The index representing the accuracy of the rating curve was chosen simply to reflect the degree of extrapolation of the stage-discharge relation necessary in the estimation of values of discharge from measures of stage. A separate index is assigned to the upper and lower portions of the curve, as the extremities of the relation may behave differently.

Extrapolations less than one times the gauged flow, on either end, are given a maximum index of 10. Extrapolations to 1.5 the nearest gauged flow are given a value of 5, while extrapolations beyond twice the nearest gauged flow are assigned a value of 2.

11.3.3 *Ice effect*

The maximum value of 10 is associated with sites where the percentage correction due to ice effect of the observed discharge in establishing the corrected discharge is less than 15%. A score of 10 reflects sites with no ice formation or with minimal effect of ice on the open-water rating curve. Corrections between 30 and 50% are assigned an index of 5, while corrections higher than 50% are assigned an index of 2.

11.3.4 *The Composite Index*

These factor indices are then combined for the differing conditions to provide a composite index reflecting the overall accuracy of the data. Tables showing the index for each factor provide a quick means of assessing what may be impacting on the quality of the records for the particular site.

Table 11.2 illustrates the derivation of the rating-curve factor index for three hypothetical stations. The minimum and maximum aspects of the curve are reflected separately, then totaled and divided by 2 to provide the factor index for accuracy of the rating curve.

Table 11.2. Derivation of the factor index for rating curve accuracy for three hypothetical stations.

Station	Minimum Rating Curve	Maximum Rating Curve	Total	Factor Index
1	2	4	6	3.0
2	10	2	12	6.0
3	2	7	9	4.5

Table 11.3 shows the factor indices and the derivation of the composite index for the three hypothetical stations. The composite index is the simple average of the factor indices. This approach provides an equal weighting to each factor. The right-most column in Table 11.3 is the resulting composite index.

Table 11.3. Derivation of the composite index for three hypothetical stations.

Station	Factor Index				Composite Index
	Stability of Control	Ice Effect	Rating Curve	Total	
1	5	4	3.0	12.0	4.0
2	10	8	6.0	24.0	8.0
3	2	10	4.5	16.5	5.5

11.4 Stream water quality networks

Alexander *et al.* (1998) describe a comprehensive collection of water quantity and quality data by the USGS including measurements for 122 physical, chemical, and biological properties of water at 680 monitoring stations. These sites are from two water quality networks. The first is the Hydrologic Basin Network (HBN) which had previously been mentioned when referencing the work of Langbein & Slack (1982), who had termed it a “benchmark” network. It comprises from 50 to 63 small, ii disturbed basins. The basins range in size from 5 to 5,200 km with a median size of 148 km. Water quality sampling was conducted from about 1967 through to 1996. Water quality sampling at HBN sites was discontinued in 1997. The second water quality network is the National Stream Water Quality Accounting Network (NASQAN).

NASQAN historically comprises up to 618 sites, ranging in size from 3 to 3 million km with a median size of 10,400 km. These sites represent mostly non-pristine basins, as the impetus for this network was to quantify long-term trends in national water quality and assess the sufficiency and effectiveness of pollution control legislation. This network was initiated in 1973, with just 51 sites. The number of sites fluctuated throughout its history and had approximately 400 stations from 1987 to 1992. The number of sites had declined to 140 sites in 1995, and in 1996 the network was reconfigured to cover 39 basins and included a broader range of parameters.

These sites within the HBN could be very valuable for assessing the potential impacts of climate change on the environment. The potential utility of the network for monitoring

change from a water quality perspective has been lost. In comparison, the NASQAN provides an excellent source of material to evaluate and assess human impacts on the quality of the water.

11.5 Discussion

From the review of the selection criteria for hydrometric networks, it is evident that previous efforts in hydrometry tended to choose somewhat similar criteria. There is, however, a major distinction that could be made between efforts in hydrometry with those of the surface climate data efforts. It is evident in the work of Vose *et al.* (1992), Jones (1994), and GCOS (1998) that climate stations reflecting direct human contamination of data, such as the heat-island effect, are acceptable. In contrast, hydrologists have paid special attention in ensuring that the data from hydrometric sites were as free as possible from direct human interference on the landscape.

To assess the potential for human interference within hydrometric data as well as for assessing the accuracy of the measurements, local expert knowledge is required. The application of selection criteria without the participation of local experts having site specific knowledge will potentially result in specialised networks containing sites of dubious quality and worth. Analysis of such networks may lead to erroneous conclusions or hamper the identification of change attributable to climate change and variability.

An important aspect that should not be overlooked is the need to obtain visibility for the hydrological network for climate change detection within the scientific and policy communities. Efforts must be made to underline the importance of the specific stations that have been selected for inclusion in such specialised networks. Without suitable monitoring of potential change and support to its scientific analyses, it will be difficult if not impossible to ascertain the characterisation of change and the sufficiency of political measures taken to deal with the implications.

References

- Alexander, R.B., Slack, J.R., Ludtke, A.S., Fitzgerald, K & Schertz, T.L., 1998. Data from selected U.S. Geological Survey national stream water quality monitoring networks. *Wat. Resour. Res.*, **34**(9), 2401-2405.
- Burn, D., 1994. Identification of a data collection network for detecting climatic change. *Canadian J. of Water Resources*, **19**(1), 27-38.
- Environment Canada, 1996. Climate network rationalization. Atmospheric Environment Service, National Weather Services Directorate, Downsview, Ontario, November, 48p. plus 116 p. of annexes.
- Environment Canada, 1999. Establishment of the Reference Hydrometric Basin Network (RHBN) for Canada — Draft. National Weather Services Directorate, Downsview, Ontario, January 21, 1999, 42 p.
- Global Climate Observing System (GCOS), 1998. Report on the adequacy of the global climate observing systems. COP4, GCOS-48, Buenos Aires, Argentina, November 2-23, 34 p.

Jones, P.D., 1994. Hemispheric surface air temperature variations: a reanalysis and an update to 1993. *J. of Climate*, **7**, 1794-1802.

Lawford, R.G., 1992. Hydrological indicators of climate change: issue or opportunity? In: Proceedings of the NHRI Workshop, No. 8, Using Hydrometric Data to Detect and Monitor Climatic Change, Environment Canada, NHRI, Saskatoon, Saskatchewan, 247 p.

Langbein, W.B. & Slack, J.R., 1982. Yearly variations in runoff and frequency of dry and wet years for the conterminous United States. 1911-79, U.S. Geological Survey Open File Report No. 82-751.

Peterson, T, Daan, H. & Jones, P., 1997. Initial selection of a GCOS surface network. GCOS 34, WMO/TD No. 799, 14 p.

Pion, P.J., Winkler, T., Harvey, K.D. & Kimmett, D.J., 1991. Hydrometric data in support of climate change studies in Canada, presented at the NATO Advanced Research Workshop on Opportunities for Hydrological Data in Support of Climate Change Studies, Lahnstein, Germany, 26-30 August.

Slack, J.R. & Landwehr, J.M., 1992. Hydro-climatic data network (HCDN): a U.S. Geological Survey streamflow data set for the United States for the study of climate variations. 1874-1988, U.S. Geological Survey Open File Report 92-129, USGS Water Supply Paper No. 2406, Reston, Virginia, 171 p.

Vose, R.S., Schmoyer, R.L., Steurer, P.M., Peterson, T.C., Heim, R., Karl, T.R. & Eischeid, J., 1992. The global historical climatology network: long-term monthly temperature, precipitation, sea level pressure, and station pressure data. ORNL/CDIAC-53, NDP-041, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Wallis, J.R., Lettenmaier, D.P. & Wood, E.F., 1991. A daily hydroclimatological data set for the continental United States. *Wat. Resour. Res.*, **27**(7), 1657-1663.

World Meteorological Organization (WMO) (Mitchell, Jr., J.M., Dzerdzeevskii, B., Flohn, H., Hofmeyr, W.L., Lamb, H.H., Rao, K.N. & Wallen, C.C.), 1966. *Climate change*, WMO - No. **195**. TP. 100, Technical Note No. 79, 79 p.

World Meteorological Organization (WMO), 1986. Guidelines on the selection of representative climatological stations (RCSs) from the existing climatological station network. WCP-1 16, WMO/TD-No. 130, 18 p.

World Meteorological Organization (WMO), 1988. Technical regulations. WMO - No. 49, Chapter D.1 - Operational Hydrology, p. D.1.1, 1-3.

World Meteorological Organization (WMO), 1993. Report of the experts meeting on representative climatological stations (RCS) and national climate data catalogues (NCC). WCDMP-No. 23, WMO - 11) No. 535, Offenbach am Main, 25-27 August 1992, 14p. plus 94 p. of annexes.

CHAPTER 12

PHASE RANDOMISATION FOR CHANGE DETECTION IN HYDROLOGICAL DATA

Maciej Radziejewski, Andras Bardossy & Zbigniew W. Kundzewicz

12.1 Introduction

The issue of change detection in hydrological data is of much practical importance. Water resources systems are typically designed and operated under the assumption of stationarity, meaning that the essential characteristics of variability of hydrological processes do not change with time. If this assumption is abandoned, existing codes of design of water resources systems, dams, levees and other water engineering works would have to be revised. Otherwise, the systems would be either underdesigned or overdesigned, i.e. either missing the target or becoming overly costly.

Many tests for trend detection have been used in studies of long time series of hydrological data. Yet, every test requires a number of assumptions to be satisfied. When underlying test assumptions are not fulfilled, acceptance and rejection regions of the test statistic cannot be rigorously determined. Therefore, such tests should be treated as methods of exploratory data analysis rather than as rigorous testing techniques.

The assumption of normality, needed in the case of parametric tests, may be an unacceptably simplifying one in the context of strongly positively skewed hydrological data. In the case of non-parametric, robust tests, one does not need to assume a distribution. Hirsch *et al.* (1991) found that non-parametric procedures offer large advantages when the data are strongly non-normal, and suffer only small disadvantages (in terms of efficiency of power) for normally distributed data.

Even though no distribution needs be assumed, non-parametric tests still make assumptions. Usually, an assumption of temporal independence must be made. When analyzing a time series of river flows, this assumption may be adequate for annual flow records. However, for shorter time intervals, such as months or days, it is not likely to hold.

12.2 Technique of phase randomization

The technique of phase randomisation is a way of generating data whilst preserving the autocorrelation structure of the raw series. The idea stems from the Fourier transform philosophy, converting an original time series into a frequency domain - amplitude and phase spectra. Consider a continuous example where the time series of observed data is given as a function $f(t)$.

By taking Fourier transform of $f(t)$ one obtains

$$F(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt = |F(\omega)| e^{i\varphi(\omega)} \quad (12.1)$$

where i is the imaginary unit, ω is the frequency, $F(\omega)$ is the Fourier transform, $|F(\omega)|$ is the amplitude spectrum and $\varphi(\omega)$ is the phase spectrum of the function $f(t)$.

The amplitude spectrum in the frequency domain, $|F(\omega)|$, is linked with the autocorrelation function of the original signal $f(t)$ in the temporal domain

The principle of the phase randomisation technique is to generate a signal while keeping the amplitude spectrum of the original signal, by changing (randomising) the phase spectrum. The generated time series has the same autocorrelation structure as the raw series. Preserving

the amplitude spectrum in the complex domain is tantamount to preserving the autocorrelation properties in the temporal domain.

The three essential steps of the phase randomisation are shown in Table 12.1.

Table 12.1. Phase randomisation.

A	Fourier transformation to the spectral domain. The Fourier transform (amplitude and phase spectra) of the resulting standardised series are computed.
B	Randomising the phases in the phase spectrum, keeping the power spectrum preserved.
C	Reverse Fourier transformation, back to the temporal domain.

The use of this method for trend detection in the present study is similar to the spirit of bootstrapping. Like bootstrapping, phase randomisation is used to obtain test significance levels and the method can be applied to any selected test statistic. Phase randomisation avoids the need to use classical formulae for rejection / acceptance ranges of the test statistics — which are based on strong simplifying (hence unacceptable) assumptions. Phase randomisation is likely to be particularly useful for testing series with strong autocorrelation in the data, e.g. daily hydrological series.

12.3 Methodology

Phase randomisation is recommended as a useful technique in its own right that avoids the restrictive assumption of independence of observed data. Moreover, as will be shown in this contribution, the phase randomisation technique lends itself well for comparing available statistical testing approaches.

In order to obtain a synthetic, trend-free series, out of a natural flow record, the following procedure was conducted. Good quality raw data was subjected to normalisation, de-seasonalisation and Fourier transformation. Keeping the amplitude spectrum preserved, the phase spectrum was subjected to randomisation. In this way a number of realisations were generated while preserving the essential properties of the raw series. After transforming back to the temporal domain, the data were contaminated with either linear trends or an abrupt jump. The changes were of various strengths measured as a ratio of the amplitude of the trend or jump, and the standard deviation of the series to which they were added. The series were subjected to a number of statistical tests.

The algorithm of the proposed methodology is shown in Table 12.2.

The assessment of significance level was made as follows. Suppose that a time series $f(t)$ is to be investigated and that a test statistic $S(f)$ is calculated. In order to assess the significance of the result, N time series with prescribed properties are generated. For each of these series the test statistic S is recalculated. Suppose they are ordered:

$$S(f_1(t)) < S(f_2(t)) < \dots < S(f_N(t)) \quad (12.2)$$

$$\text{If } S(f_m(t)) < S(f(t)) < S(f_{m+1}(t)) \quad (12.3)$$

then the probability of not exceeding $S(f)$ is

$$p = m / N \quad (12.4)$$

Ties in (12.2) and (12.3) can be handled as usual, by taking an appropriate average. Now, the result would be significant on the 95% level if the probability were above 0.975 (indicating a high result) or below 0.025 (indicating a low result). A notion of significance defined in terms of this probability will be used herein as:

$$\text{Significance} = 2p - 1 \quad (12.5)$$

The absolute value of significance corresponds directly to the significance level achieved, while the sign indicates the direction of the change.

Attention: The convention used here differs from the one in Chapter 5. Significance of 95% here corresponds to 5% in Chapter 5.

Table 12.2. Algorithm of the proposed methodology.

0	Departure point: long time series of raw data of good quality.
1	Normalization of the distribution, by replacing each term with the value which would have had the same non-exceeding probability in the Gaussian distribution. This step preserves the relative ranks of the data.
2	De-seasonalization. The annual cycle was removed by subtracting the regime and dividing the residue by seasonal standard deviations.
3	Phase randomisation (Table 12.1)
4	Artificial trends/ step changes were added to the series generated from step 3. Then the original distributional and seasonal properties of the series were restored by reversing the steps 2 and 1. The resulting series was tested for changes.
5	Comparison of test statistics for a contaminated series with those obtained for all series generated from this one by further phase randomisation, in order to replace each statistic's value by the corresponding significance. The process was repeated for all contaminated series and the mean significance was calculated for each shape, strength and statistic under study.

12.4 Comparison of tests

Hydrological change may take different forms, e. g. abrupt jump versus gradual monotonic change (trend), and may occur in the mean and / or in variability (variance, extremes, persistence). In addition, there are complicating factors such as seasonality, missing values and in some cases, censored data and problems arising from small sample sizes. The ability to detect change using a test depends on the type of change, its magnitude, the length of the series and the time when the change occurs in the series. A particular test may work well in specific situations (e. g. for a gradual trend or an abrupt jump) and not so well elsewhere. Even so, many changes observed in real data do not necessarily fall in a single category; there exist intermediate cases.

Two possible selection criteria are the power of the test and its computational efficiency. The former is a measure of how well the test detects trends — it looks at the probability of error (detection of false trend or failure to detect a real trend). The computational efficiency is of lesser importance given the massive growth in computer power, but can still be important in computationally-intensive Monte-Carlo studies and in multi-site analyses over large spatial regions.

There are many parametric and non-parametric tests for change detection. Some parametric tests can be applied in a non-parametric way by testing either the ranks or the so-called “normal scores”, i.e. the series transformed in such a way that the marginal distribution becomes normal, while the relative ranks of the values are preserved. The tests used in the present study represent a selection of some of the most commonly used statistical tests and are further described in Radziejewski *et al.* (2000). These are:

No. 1.	Mann's test (non-parametric)
No.2.	Normal scores linear regression (non-parametric)
No.3.	Spearman's rank correlation (non-parametric)
No.4.	Linear regression (parametric)
No.5.	Jump fitting to normal scores (non-parametric)
No. 6.	Jump fit to ranks (non-parametric)
No. 7.	Jump fit (parametric)

12.5 Results

12.5.1 Data used

In order to reduce problems linked to data quality, a set of river flow observations stemming from the USGS Hydro-Climatic Data Network (Slack *et al.*, 1993) was used. The subset of 202 series selected for investigations fulfilled the following criteria: sufficient length of continuous daily record (at least 60 years), lack of significant anthropogenic influence and no documented ice influence. The records of the river Greenbrier at Alderson, WV (1896-1985) and the river Mississippi at Clinton, IA (the 1879—1967 part) have served as the basis for the artificially generated series.

It is often necessary to run tests using aggregated data (typically—annual) in order to meet the usual test requirements of independence. The independence assumption is not needed here because significance levels will be determined using phase randomisation. This means that it is possible to test the full daily series, thus making maximum use of the available data.

12.5.2 Results for test comparisons using artificial series

Series contaminated with controlled artificial trends of different amplitude and form (2500 series for each change type / strength) have been created following the procedure explained in Table 12.2. The results are used to compare test performance under phase randomisation and to evaluate how detection of change varies as a function of the magnitude of the contaminating trend.

Figures 12.1—12.2 illustrate how well trend is detected for the series contaminated with changes in mean of strength from 0.1 δ to 2 δ , in 0.1 δ intervals. The changes considered are abrupt jump and linear trend, respectively. Radziejewski *et al.* (1998) report on the results for other change types, such as broken line, gradual jump and rectangular pulse.

In case of contamination with a linear trend (Fig. 12.1), tests 1—3 performed almost identically to each other and better than all the other tests. The parametric test 4 was the next best, but still, significantly worse except in cases of very strong trend. The better performance of tests 1-4 in the linear trend case is not surprising, since gradual monotone change is what these tests look for. Another close similarity is between test 5 and 6, with test 6 (jump fit to ranks) being marginally better. Again, these non-parametric tests performed better than their parametric counterpart (test 7). Tests 5 and 6 allowed the detection of changes they were not designed for, while test 7 was completely useless in this situation. The feature of detecting a broad range of changes is very important in real-life situations, where we have no information about the type of changes to be expected.

As could be expected, for contamination with an abrupt jump (Fig. 12.2) tests 5 and 6 give the best results. Again, there is a high degree of similarity between the two tests, both of them significantly outperforming test 7. Test 7 gave better results than tests 1—4 for the Mississippi-based series, but was of little use otherwise. Tests 1-4 all exhibited similar performance. The Mississippi river case, where the parametric test 4 does well, is interesting, but it could be a chance occurrence.

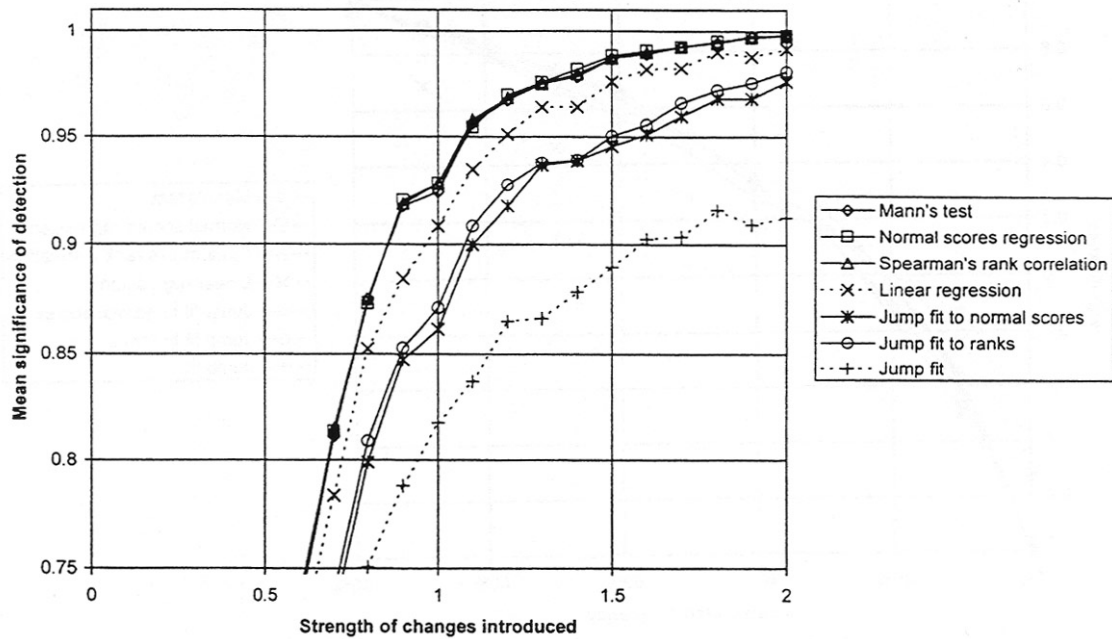


Fig. 12.1. Linear trend detectability in artificial series based on the flow of the river Mississippi in Clinton, IA, USA.

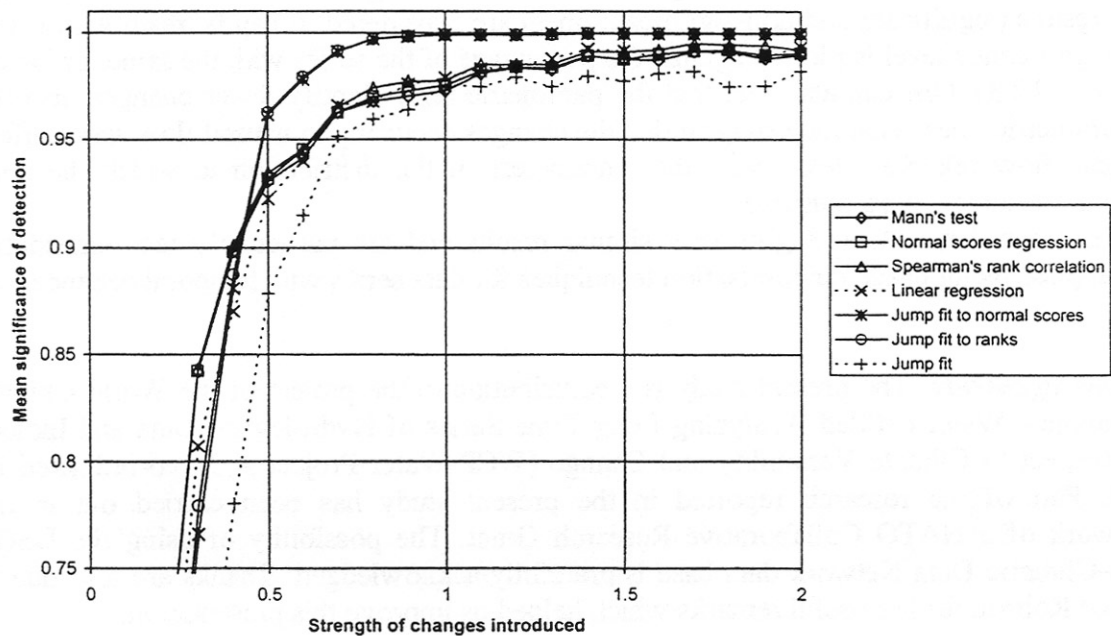


Fig. 12.2. Jump detectability in artificial series based on the flow of the river Greenbrier at Alderson, WV, USA.

12.5.3 Results for natural flow series

Out of the 202 natural flow series under study, 23 showed a significant increase (at the 5% level) and only 4 a significant decrease. Tests 4 and 7 gave hardly any significant results at all with much better detection of change occurring when the non-parametric tests are used.

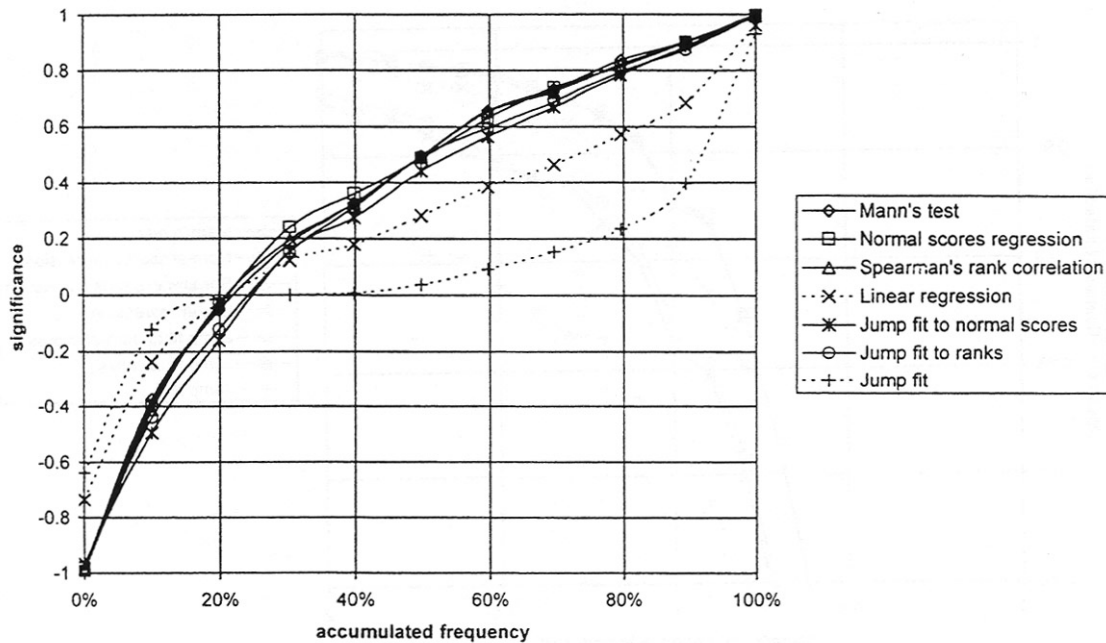


Fig. 12.3. Link between cumulative frequency and significance for natural series (a subset of 202 flow series from HCDN).

The prevailing direction of change is towards increasing flow, this is most clear when all the results (significant and non-significant ones) are considered. It can be readily observed when significance level is plotted against the percentage of the series with the same, or lower, level (Fig. 12.3). One can also see, that the parametric tests identify fewer changes than the non-parametric ones. That may suggest that the changes occurring in natural flow series affect the mean flow relatively less than other parameters of the distribution to which the non-parametric tests are more sensitive.

It seems that tests 1-3 give very similar results and are particularly recommendable when applied using phase randomisation techniques for data series with temporal dependence.

Acknowledgements The present study is a contribution to the project of the World Climate Programme - Water entitled Analyzing Long Time Series of Hydrological Data and Indices with Respect to Climate Variability and Change (WCP-Water Project A.2), co-ordinated by WMO. Part of the research reported in the present study has been carried out in the framework of a NATO Collaborative Research Grant. The possibility of using the USGS Hydro-Climatic Data Network data base is gratefully acknowledged. Thanks are also due to Dr Alice Robson for her useful remarks which helped us improve this presentation.

References

- Hirsch, R.M., Alexander, R.B. & Smith, R.A., 1991. Selection of methods for the detection and estimation of trends in water quality. *Wat. Resour. Res.* **27**(5), 803-813.
- Radziejewski, M., Bardossy, A. & Kundzewicz, Z. W., 1998. Intercomparison of tests for detection of changes in long time series of river flow. Proc. of the Second International Conference on Climate and Water, Espoo, Finland, 17-20 August 1998, Vol 3, 1120-1129.

Radziejewski, M., Bardossy, A. & Kundzewicz, Z. W., 2000. Detection of change in river flow using phase randomisation, *Hydrol. Sci. J.* (in print).

Slacic J.R, Lumb, A.M. & Maciunas Landwehr, J., 1993. Hydro-Climatic Data Network: (HCDN) Streamflow Data Set, 1874-1988, USGS (CD-ROM and Manual).

CHAPTER 13

SIMULTANEOUS ESTIMATION OF TRENDS IN MEAN AND VARIANCE

Witoki G. Strupczewski

13.1 Introduction

Investigation of trend in long time series of hydroclimatological data is mainly aimed at detection of change of the mean value. However, the change may go beyond the first statistical moment. In particular, a trend in the second moment can be of much importance to ecosystems, socio-economy and individuals. For example, increasing trend in variance of winter air temperature means that severe winters are more likely while a change in autocorrelation function affects likelihood of a sequence of severe or mild winters. The scope of this paper is constrained to a parametric method of simultaneous estimation of trend in mean and variance. It should be stressed that, even if only investigation of trend in mean is of concern, heteroscedasticity has to be taken into account. Therefore investigating a long time series of unknown character, it is best to assume the most general case, i.e. of functionally non-related trends in the both first statistical moments, and then to proceed to simpler cases ending at the stationary case, from this identifying the best-fitting model. The investigation can comprise various functional forms of trend, e.g. linear, parabolic or the periodic functions of time, while keeping in mind the need for parameter parsimony. Statistical significance of detected trends is not tested but the model showing the best fit of all competing models is taken as the optimal one.

13.2 Discussion of methods

The Maximum Likelihood (ML) method is the most theoretically rigorous approach to estimation of trends in both mean and variance in the sense that, under appropriate distributional assumption, it produces asymptotically efficient and unbiased estimates of time-dependent moments. Strupczewski & Feluch (1998) used it in flood frequency analysis under non-stationarity. They developed the Identification of Distribution and Trend (IDT) software package which identifies an optimum flood frequency model with time dependent parameters from a class of competing models that are useful in design of structures in a changing environment. The notion of a model is understood here as a type of probability distribution function (pdf) together with a class and form of trend in the two first moments. The original parameters of the pdf are expressed in terms of time-dependent moments using relationships between moments and parameters available in statistical literature and, in effect, the trend is explicitly introduced to the moments. The IDT program estimates model parameters from the time series, derives the covariance matrix and estimates the probability distribution of exceedance with confidence intervals for any given year or a period of any length optionally located along the time axis. The Akaike Information Criterion (AIC) goodness-of-fit test is used to identify an optimum model.

One notes that the use of the ML method links the estimators of time-dependent moments with the type of pdf. Therefore the values of trend estimators depend on an assumed pdf; a shortcoming of using the ML method for trend estimation. In practice, the “true” pdf is not known and the hypothetical one can differ substantially from it. Superiority of ML-estimators of moments over those of the method of moments (MOM) relies on the distributional assumption. An effort to relax the distributional assumption in trend investigation resulted in development of the Weighted Least Squares (WLS) method

(Strupczewski & Kaczmarek, 1998), which is a generalisation of the Least Squares method for time-variable variance.

13.3 Weighted least squares method

The principle of the WLS method is based on the minimisation of sums of weighted squared deviations of observed and estimated moments, where the weights are reciprocals of their expected values (Strupczewski & Kaczmarek, 1998). The WLS, being conceptually quite distinct from the ML-method, coincides with the ML method in the case of normally distributed data. In this case a simple presentation of the WLS as a problem of the ML-estimation is possible.

The log-likelihood function in a time series subject to normal distribution with time-variable parameters has a form:

$$\log L \rightarrow -\sum_{t=1}^T \ln \sigma_t - \sum_{t=1}^T \frac{(x_t - m_t)^2}{2\sigma_t^2} \quad (13.1)$$

where $m_t = m(\mathbf{g}, t)$ and $\sigma_t = \sigma(\mathbf{h}, t)$ are time-dependent mean and standard deviation, respectively, while the \mathbf{g} and \mathbf{h} are vectors of parameters.

The conditions of maximum log L are:

$$\frac{d \log L}{dg} \Rightarrow \sum_{t=1}^T \frac{(x_t - m_t)}{\sigma_t^2} \frac{dm_t}{dg} = 0 \quad (13.2)$$

$$\begin{aligned} \frac{d \log L}{dh} &\Rightarrow -\sum_{t=1}^T \frac{1}{\sigma_t} \frac{d\sigma_t}{dh} + \sum_{t=1}^T \frac{(x_t - m_t)^2}{\sigma_t^2} \frac{d\sigma_t}{dh} = \\ &= \sum_{t=1}^T \frac{1}{\sigma_t^3} [(x_t - m_t)^2 - \sigma_t^2] \frac{d\sigma_t}{dh} = 0 \end{aligned} \quad (13.3)$$

Note that both sets of equations contain both time-dependent mean (m_t) and variance (σ_t), i.e. they need to be solved jointly unless the standard deviation can be assumed to be constant. The WLS method covers four classes of trend, i.e.:

- A** - trend in the mean only, which is the common least squares problem of trend estimation
- B** - trend in the standard deviation only;
- C** - trend both in the mean and standard deviation being functionally related, e.g. by a constant value of the variation coefficient (c_v);
- D** - non-related trend in the mean and standard deviation.

The most basic case is that of time-invariable moments, is called the stationary option (S). In this case, equations (13.2)-(13.3) reduce to the MOM equations.

The test of goodness-of-fit based on Akaike Information Criterion serves to identify an optimum model in a class of competing models of trend:

$$AIC = -2 \ln ML + 2k \quad (13.4)$$

where ML denotes the maximum likelihood for the model and k is the number of fitted parameters. The best approximating model is the one, which achieves the minimum AIC value in the class of competing models.

Equations (13.2)-(13.3) derived here for the ML-estimation of time-variable parameters of normally distributed variables are equivalent to the WLS equations and as such they remain valid for other distribution functions providing that certain conditions are fulfilled. Table 13.1 shows applicability of equations (13.2)-(13.3) for various two- and three-parameter distribution functions and various classes of trend. A constant value of the coefficient of variation is accepted for the class **C**. Six pdfs were examined, namely: Normal (**N**), two-parameter Lognormal (**LN2**), three-parameter Lognormal (**LN3**), Gamma (**P2**), three-parameter Pearson (**P3**) and GEV of the first type, i.e. Fisher-Tippett of the first type (**FT**). Shading indicates applicability of the WLS method for the case, while dark colour denotes the cases of equal weights, i.e. where both the LS method or MOM are applicable. An absence of shading indicates cases where equation (13.3) does not hold. The constraints of the WLS method do not restrict, in any significant way, its climatological or hydrological application. This is because the lack of a prior information on the functional form of the probability distribution gives a certain freedom of a distribution choice. For example, instead of the two-parameter Lognormal (**LN2**) or Gamma distribution (**P2**), three-parameter distributions with time invariant coefficient of asymmetry can be assumed (e.g. **LN3** or **P3**). With this approach, the WLS method is effectively applicable in all the above cases.

Table 13.1. Applicability of the Weighted Least Squared Estimation (**WLS**) of trend for various types of probability distribution and various classes of trend. *c* is the coefficient of asymmetry.

Model	Distribution					
Trend	N	LN2	LN3	P2	P3	FT
S						
A						
B			$c_s = \text{const}$		$c_s = \text{const}$	
C			$c_s = \text{const}$		$c_s = \text{const}$	
D			$c_s = \text{const}$		$c_s = \text{const}$	

A rough assessment of the type of probability distribution can be made using observed time series. This should be followed by a more detailed analysis performed on the series reduced to stationary conditions, e.g. assuming a normal distribution for the series ($Y_1, Y_2, \dots, Y_t, \dots$), where $Y_t = (X_t - m_t)/\sigma_t$ is the standardised variable.

For example, consider the case of a class **D** with trend of linear form, i.e. $m_t = a + bt$ and $\sigma_t = c + dt$

Then equations (13.2) and (13.3) take the following form:

$$\sum_{t=1}^T \frac{1}{(c + dt)^2} [x_t - (a + bt)] = 0 \quad (13.2a)$$

$$\sum_{t=1}^T \frac{t}{(c + dt)^2} [x_t - (a + bt)] = 0 \quad (13.2b)$$

$$\sum_{t=1}^T \frac{[x_t - (a + bt)]^2}{(c + dt)^3} - \sum_{t=1}^T \frac{1}{(c + dt)} = 0 \quad (13.3a)$$

$$\sum_{t=1}^T \frac{t[x_t - (a + bt)]^2}{(c + dt)^3} - \sum_{t=1}^T \frac{t}{(c + dt)} = 0 \quad (13.3b)$$

while the *AIC* formula reads

$$AIC = - \ln ML + 8 \quad (13.4)$$

To solve equations (13.2a-b) and (13.3a-b), a gradient method was applied with moment estimates of the stationary case as the starting vector.

Strupczewski & Mitosek (1998) demonstrate the importance of relaxing the assumption of homoscedasticity in an investigation of linear trend in annual peak flow series of Polish rivers. Out of 39 series covering the common period of 1921-1990, the stationary model (**S**) was preferred in 14 cases, while the **A**, **B**, **C** and **D** models were found to be the best in 1, 6, 17 and 1 cases, respectively. The predominant identification of the class **C** model (with constant coefficient of variation) gives evidence of a strong positive correlation of trends in the mean and in the standard deviation. According to the *AIC*, allowing for time-variable variance may significantly improve the fit of a model. In fact, it also affects the estimated trend in the mean. The average difference in the gradient of trend in the mean between class **A** (time-constant variance) and classes **C** and **D** is 26% for the 18 series in the **C** and **D** classes above. Analysing equation (13.4), one can see that its first term accounts for the criterion of a good statistical fit and its value is growing with the series length. The second term represents the doctrine of parameter parsimony in *AIC*. Therefore, the longer the time series the higher a chance of selection of multiparameter forms of trend and the class **D** instead of **C**.

13.4 Conclusions

The presented method enables estimation of the trend in the two first moments of time series and identification of optimum trend model from a set of competing models. For selection of alternative forms of trends in both moments both visual properties of a time series and its length should be taken into account. The results of the case study corroborate the need to account for time-dependent variance when investigating trend in time series of hydroclimatological data.

Acknowledgement The work reported in this study was supported in part by the Polish Committee of Scientific Research (KBN) grant No. PO4D 056 17, "Revision of applicability of the parametric methods for estimation of statistical characteristics of floods". This support is gratefully acknowledged.

References

- Strupczewski, W.G. & Feluch, W., 1998. Investigation of trend in annual peak flow series. I. Maximum likelihood estimation. *Proceedings of the Second International Conf on Climate and Water*, Espoo, 241-250.
- Strupczewski, W.G. & Kaczmarek, Z., 1998. Investigation of trend in annual peak flow series. II. Weighted Least Squares estimation. *Proceedings of the Second International Conf on Climate and Water*, Espoo, 251-263.

Strupczewski, W.G. & Mitosek, H.T., 1998. Investigation of trend in annual peak flow series. ifi. Flood analysis of Polish rivers. *Proceedings of the Second International Conf on Climate and Water*, Espoo, 264-272.

GUIDELINES: SOFTWARE

Felix Portmann

Working on time series analysis leads irrevocably to the use of software. Only in special cases manual methods replace or assist software.

Users should take into account that the **software use** for statistical analysis is threefold:

- **Data preparation:**

In the beginning, the data exist often in various forms and must be transformed into the format that the specific software can read. Perhaps missing values have to be excluded, marked or replaced by estimated ones. Instantaneous values have to be aggregated to daily, monthly, annual series and annual maximum series or partial duration series have to be generated. Either the statistical software helps with the data preparation or the user has to do it on her/his own.

- **Statistical analysis:**

At this core analysis, the data is analysed with a statistical software and results in primary form are produced. These may be in the form of tables, graphs, semantic information like rejection of hypothesis, etc. Interpretation of the results follows.

- **Presentation of results:**

Finally, the primary and conclusive results have to be compiled in a desired form, e.g. a report, presentation slides, a scientific publication, etc., and not only as plain hardcopy text or graphs on a sheet of paper. The compilation often needs additional treatment, e.g. annotation of graphs, inclusion of additional material, etc., normally done nowadays with software, too, e.g. Office software and Drawing software.

When choosing software, users should be aware of the scope of their analysis, e.g.:

- data checking for calculation of frequency statistics
- number of series to analyse
- form of presentation

For in-house data checking, especially for a small number of series, a manually organised sequence of analyses with simple output might be sufficient. When using large number of series, automatic procedures should be present. Furthermore, when the results have to be included into a final document or presented at a conference, for their inclusion with the specific editing software an accepted format (format of graphical metafile, e.g. WMF, JPG, TIF, PCX GIF, PICT, or of text, e.g. ASCII or RTF) and necessary type and quality (black & white vs. colour, colour depth, adequate and correct representation of style of text and points, lines, legends, annotations in graphics, pixel resolution, etc) have to be chosen, as retyping is not always possible or feasible. Any format of results supporting the final treatment is more appropriate than a format ensuing additional steps or even impeding the production of the final product in the case of incompatibility. Also, the reproduction form (paper vs. transparent media) and costs (black & white vs. colour, PostScript or not) should be considered. In this respect, hardware for the production of the output may be important to check, e.g. PostScript capability of printer or colour printer. Perhaps adjustments of colours in graphs have to be made to get a readable and ready-to-print or ready-to-photocopy hardcopy.

The specific circumstances can be very different, as the goals of the analysis, the organisational framework and the available hardware may be very different, but decisive, too.

As it is impossible to be comprehensive with respect to that situation, only **statistical** software will be treated in detail here, not software needed for the other components.

Users should keep in mind that any software may present errors, either open or hidden ones. Therefore, results of statistical software should not be taken for self-evident. Common sense and checking important framework conditions, e.g. sums, can avoid relying on wrong results.

Another problem with software is inappropriate or bad algorithms. E.g., often the so-called “random generators” do not generate truly random series, but biased ones. Besides that, when doing own programming, any correction for sample size should be done in the proper way.

Considering **software specifications**, the workshop experts concluded that currently no software exists that covers all recommended features; many approaches like permutation or bootstrapping have to be programmed or implemented at your own. Especially standardised time series analysis software often do not accept missing values. Estimation of those often remains a problem that has to be resolved externally from the statistical software. Therefore, the experts emphasise the need of

- **Development of a software with the desired features**

The realisation of this goal is beyond the scope of this workshop itself, as the programming effort will be considerable. WMO is informed through its participant of this central view of the workshop experts. Funding of such an enterprise could be proposed e.g. within the 5 Research & Development Framework Programme of the European Commission.

However:

- For **specific applications** solutions can be **programmed at your own** either directly, via programming libraries, or within statistical programming frameworks, e.g. SPLUS. Users should be aware that this takes some time but can result in optimised, very quick analyses within a common data framework.
- Some work can be done by **already existing special software** or even **general purpose statistical software**, e.g. SYSTAT. A selected software often presents only partial solutions. When using different software at a time, the raw data sets have to be brought into the specific data format. This requires some efforts, especially pre-processing of missing values that many programmes do not accept. Own programming may be required or very useful for the pre-processing.
- WMO has a collection of programmes within the **HOMS (Hydrological Operational Multipurpose System)**, of which many, but not all, are freely available. A list of freely accessible software for specific tasks, some of them within HOMS, is given in the following table. Please note, that the list is not exhaustive, but has been selected especially to provide information on freely or easily obtainable software which is giving high-quality output and does not normally need very sophisticated hardware. Experts designed and used these programmes, they can recommend these for specific tasks. For more details, readers may contact either the WMO internet site (<http://www.wmo.ch/web/homs/homshome.html>), the authors of the software or the author of this contribution.

Key name	Descriptive name	Author/originator institution	General and specific application	Specific test included	Input data	Output	Miss. Values	Operation remarks	Con-conditions on use	Costs	Further comments	General comment on practical applicability
Software already available												
HOMS Component I00.1.02: TMESER	Statistical analysis of long time hydrological data	WMO WCP Water Project A.2; WMO Hydrology and Water Resources Department	Selected statistics to identify long-term variability, trends and jumps, non-stationarity	Descriptive statistics; monthly ranks, AC, cumulative periodogram, variance spectrum, adjusted range, rescaled adjusted range, Hurst's coefficient, runs, trends (mean, variance), jump (mean), low pass Gaussian filter	Monthly (M), annual (A)	File: listings and alpha-numerical plot for whole series and sub-series of 5, 10, 20, 30 years	No	FORTRAN source code; sample data	No restrictions	Free of charge	Reference year may be specified; documentation	Useful especially for descriptive statistics; no automatic display of most significant test results; free use
HOMS Component I00.3.02: CHPP	Change point problem	Czech Hydrometeorological Institute (CHMI)	Detecting point changes in hydrological and meteorological series	Descriptive statistics, autocorrelation (AC), partial AC (PAC), normality, independence, periodogram, runs, ump (mean, variance)	M, A	Screen, graphics, listings also as files and to printer	No (Yes at specific tests)	Executable, MSDOS3.3+ colour screen (VGA); sample data	No restriction	Free licence	Selection of subseries, automatic display of most significant test results; documentation	Most significant test results; free use
JUMP4	Jump in mean and variance	Ruhr-Univ. Bochum, Germany (Andreas Schumann)	Jumps in time series	Descriptive statistics, jump(s); log-normal distribution	M, A	Screen, graphics, listings as file	No	Executable, MSDOS; sample data	Contact author	Contact author	Query of comparable series on screen; several series at a time	No automatic display of most significant test results; several series at a time
KHRONO-STAT	Time series analysis & segmentation	ORSTOM, France	Time series analysis	e.g. rank correlation, Pettitt, Buishand, Lee & Heghinian, Hubert	(M, A)	Screen, graphics, tables	<i>NA</i>	Windows95		Free of charge	Bilingual (F, E)	First version v.1.00
SEG	Segmentation	École des Mines de Paris (Pierre Hubert)	Optimal segmentation into stationary sub-series	Significance 0.05 or 0.01, test of residuals	Free (A)	Listings as file	No	Executable, MSDOS; sample data	Info, of author	Free (www.ci.g.ensmp.fr/~hubert/segment.htm)	Documentation in French, English, Portuguese	
SPLUS	Statistical Package	Statistical Sciences (StatSci), Inc.	Statistical package; trend, step change, graphical exploration	Range of tests (trend, step change), regressions, graphical exploration, summarising data	Various	Screen, PostScript (PS)	Various	Unix, Windows, MSDOS	Licence	GBP 1000 — 2000	Programming necessary	Very memory consuming; programming necessary

Key name	Descriptive name	Author/originator institution	General and specific application	Specific test included	Input data	Output	Miss. Values	Operation remarks	Con- ditions on use	Costs	Further comments	General comment on practical applicability
Software already available (continued)												
SSA	Singular-Spectral Analysis		Spectral analysis	Correlogram, Multi-Taper Method (MTM), Maximum-Entropy Method (MEM), Monte Carlo realisations	Any univariate time series	Screen, PS-graphics, tables, logs, vectors, series, power spectra	No	Unix commands; standard FORTRAN 77	Free	Free (www.atmos.ucl.ac.uk/tcs/ssa)	Available as source code, 5 main components may be used as stand-alone modules	Tcl/Tk-X Windows-GUI, visualisation also possible with IDL or public-domain ACE/gr and gnuplot
TREND1	Linear trends	Ruhr-Univ. Bochum (Andreas Schumann)	Linear trends in time series	Descriptive statistics, trend(s); log-normal distribution	M, A	Screen, graphics, listings as file	No	Executable, MSDOS); sample data	Contact author	Contact author	Query of similar trends on screen; several series at a time	No automatic display of most significant test results; several series at a time
TSA1	Time series analysis	Norwegian Water Resources and Energy Administr. (Lars Roald)	Time series analysis	e.g. runs, trend, AC, jumps	Daily (D), M, A, seasonal, annual maximum series (AMS)	Tables, graphics (GPGS)	Linear interp.	Unix, FORTRAN	Contact author	Contact author		
Software available within short												
KOLIBRI	Extreme value series	Costa Rica Electricity Institute (ICE) and HYDRO-CONSULT AB, financed by SIDA (Sweden)	Exploratory analysis and Extreme value analysis	Control of randomness, trends, AC, outliers	D, extreme values (Q, P)	Tables and graphs	Registered	PC software	NA	NA	Developed for data of humid tropics, manual available in English and Spanish	Will be available soon

Appendix 2

HYDROSPECT - SOFTWARE FOR DETECTING CHANGES IN HYDROLOGICAL DATA

Maciej Radziejewski & Zbigniew W. Kundzewicz

Note: The software below has been developed under the WCP-Water framework. It is free software containing implementation of a number of parametric and nonparametric tests commonly used in hydrological studies. It does not currently support the resampling methods recommended in this report. It is hoped that the software will be extended at a future date to allow resampling methods to be used.

Hydrospect is a software package for detecting changes in long time series of hydrological data. It makes use of a set of eight different tests for change detection (linear regression, Mann-Kendall distribution-free CUSUM, cumulative deviations, Worsley's likelihood ratio, Kruskal-Wallis, Spearman's rank coefficient, normal scores regression) and provides a possibility to create customized derived series, all in a user-friendly Windows environment. The source code of Hydrospect, Version 1.0, has been written in Visual C. The package delivered free to users consists of the executable programme file hydrospect.exe, three sample data files and the User's Manual (.doc file in MS Word 97).

Hydrospect is a contribution to the Project A (Analyzing Long Time Series of Hydrological Data and Indices with Respect to Climate Variability and Change) of the World Climate Programme - Water prepared for the World Meteorological Organization. This software was developed by Maciej Radziejewski under supervision of Zbigniew W. Kundzewicz.

The programme can be run on a personal computer compatible with the IBM PC standard. The required processor is 486 or Pentium and the operating system — MS Windows 95 or 98.

Creating or editing the data set is not supported within Hydrospect. The Hydrospect package follows a simple philosophy of working with the data as they stand, without modifications. Making changes to the data (for example correcting errors), can be done with another tool, e.g. a spreadsheet.

Hydrospect can read data from text files. The import feature has been designed for flexibility, under consideration of a number of possible formats. Provision to deal with missing values and lines containing comments and other information has been made.

If one wishes to study a time series with another program (e.g. to create a graph), one can save it in a text file. This is of particular use as Hydrospect allows one to perform a number of operations on a time series, i.e. derive new series from existing ones, and those derived series might be of some use elsewhere, too.

Hydrospect returns the values of the test statistic and of the significance level actually achieved. High value of the actual significance level means that the hypothesis of a lack of change is rejected in the light of evidence, i.e. change is detected on a high significance level. It is essential to emphasize that all the tests included in Hydrospect are based on strong assumptions on the time series: temporal independence for all the tests used and normal distribution for some tests. Validity of these assumptions in particular cases guides our credibility in test results, in particular in the regions of the test statistics where the hypothesis should be accepted / rejected (not communicated to the Hydrospect user) and the confidence

levels (returned to the user). If assumptions are not fulfilled, the tests can be only interpreted as exploratory data analysis tools, rather than rigorous statistical methods.

Hydrospect does not include tests for checking assumptions. In fact, the assumption of normality is rarely valid in the realm of heavily skewed hydrological data. Such data can be easily transformed to normality by computing normal scores (see below), so the tests can be applied to the transformed, normally distributed series. The assumption of temporal independence depends on the time step. A time series of annual values may consist of independent elements, whereas the same process represented by monthly or daily data is likely to show an increasing level of temporal dependence.

Standard procedures like ranking observations in a sample or removing annual regime from a series of flows can be seen as producing a new, derived time series, from an existing one. Hydrospect implements a number of such procedures. One can produce a number of derived series from any series. For example one can compute annual means and monthly means for a series of daily data as well as select a subseries consisting only of observations recorded in June. Then one can rank the annual means etc. The relationships between derived time series are presented in the left pane in the form of a tree.

A list of time series operations supported by Hydrospect includes:

- Derived series with ranks assigned to each observation.
- Aggregation resulting in division of the series into subperiods and replacing the values in each subperiod by one value, the mean, the sum, the minimum, the maximum, the median or Tukey's trimean.
- Creating a subseries of the time series of concern by restricting to a subperiod or to a certain part of the year, e. g. from December and January.
- Some tests for changes in the mean can be applied to detect changes in variance. This involves computing the distance of each value in the series from the overall mean and applying the test to the series of distances.
- Computing normal scores is supported in that the series is transformed in such a way that the marginal distribution becomes normal (with zero mean and unit standard deviation), while the relative ranks of the values are preserved.
- Deseasonalisation: the seasonal means (regime) are subtracted from each value and the remainder is divided by the seasonal standard deviation. The means and deviations are smoothed using harmonic functions.

The *Report* menu entry allows one you to create a text file containing the results of all the tests performed during a session.

References

Radziejewski, M. & Kundzewicz, Z. W., 1999. Hydrospect, Version 1.0, User's Manual. Unpublished manuscript.

The software package Hydrospect and the User's Manual are distributed free of charge by Dr Arthur J. Askew, Director of Hydrology and Water Resources Department, World Meteorological Organization, Geneva, Switzerland; (e-mail: askew_a@gateway.wmo.ch).

GLOSSARY

Acceptance region – Set of test statistic values for which the null hypothesis is accepted.

Alternative hypothesis – Counterpart to the null hypothesis (Section 5.2).

Autocorrelation – A series shows autocorrelation if there is dependency between one value and the next. This is also called serial correlation, or temporal dependence.

Biased – An estimate is biased if the average value of the estimate differs from the true value of the quantity or parameter that it estimates.

Bootstrapping – A resampling method used for testing data (Section 5.4).

Box plot – A visual data summary which shows: the smallest value, the three quartiles (see: quartiles) and the highest value (cf. Chapter 4).

Climate change – A long term alteration in the climate (Chapter 1, Appendix).

Climate variability – The inherent variability in the climate (Chapter 1, Appendix).

Confidence interval – Range of values that a parameter is likely to fall within. If the 95% confidence interval is (A, B) then there is a 95% chance that the true value lies between A and B.

Correlated – Variables are said to be correlated if they are related to each other in some way.

Correlation coefficient – A measure of the association of two variables. The correlation coefficient is the covariance of two random variables (or data sets) divided by the product of their standard deviations.

Covariance – A measure of the association between two variables. It is the expected value of the product of departures from the mean (mixed second central moment of two stochastic processes).

Critical value – Value that separates the rejection and acceptance regions for a test statistic.

Dependent – A variable is dependent if it is correlated with another variable.

Deseasonalize – Remove the seasonal variation in a time series (e.g. annual seasonal pattern).

Distribution-free approach – A method that does not require a specific probability distribution to be assumed (can be used for all distributions).

Entropy (information-theoretic) – Measure of the degree of uncertainty or disorder. (Concept, germane to thermodynamic entropy, introduced in Shannon, C. E. (1948) A mathematical theory of communication, *Bell System Tech. J.*, 27, 379-423, 623-659.

Exploratory data analysis (EDA) – Graphical display of data in order to reveal essential properties (cf. Chapter 4). Instruments of EDA are scatter plots, box plots, etc.

Extremes (hydrological) – Values that are substantially different from the ordinary or usual. (e.g. floods and droughts are extreme hydrological events).

Fourier transform – A time series technique that is used to break a series up into a sequence of sinusoidal waves.

Heterogeneity – A sample shows heterogeneity if the values come from a number of different distributions (e.g. corresponding to different underlying physical mechanisms).

Heteroscedasticity – Indicates that the variance of the data is not constant.

Homogeneity – Property where by a whole sample comes from the same distribution.

Homoscedasticity – Indicates that the variance of the data is constant.

Independent – Data points are independent if they are not related to each other. If data are correlated, or show serial correlation or spatial correlation they are not independent.

Interpolation – Estimating values of a variable between known values.

Jump: see step change.

Kurtosis – A measure of the peakiness of a distribution. See moments.

Lag – A term used in time series methods. A lag 1 observation is the observation from the previous time step. A lag 2 observation is an observation from 2 time steps before.

Loess/LOWESS /Locally weighted smoothing – A robust smoothing method (see Chapter 4).

Maximum likelihood – Method of parameter estimation in which the likelihood function is ma

Mean – The average of a series. See moments.

Median – The middle ranking value of a series. See second quartile.

Moments: Absolute moment – Expected value of a power of a random variable. Central moment - Expected value of a power of a difference of random variable and its mean. First (absolute) moment is the mean. Second central moment (about the mean) is variance. Next two central moments are called skewness and kurtosis. Mixed moments - see covariance.

Monotonic change – A change that is consistently in one direction (either always upwards or always downwards).

Multivariate analysis – A method for simultaneous analysis of a number of dependent variables (of time and / or space), including links between these variables.

Non-parametric test – A test that does not involve estimation of parameters. Rank-based tests are non-parametric.

Non-stationarity – Indicates that the distribution of a random variable is changing with time (e.g. increasing mean).

Normal scores – The expected values of a sample from the normal distribution.

Null hypothesis – Hypothesis to be tested.

One-sided test – Statistical test where the rejection region is located in one tail of the distribution.

Outliers – Values that appear unusual because they are distant from the bulk of data.
Parametric test - A test that involves estimation of one or more parameters (linear regression is a parametric test because it involves estimation of the gradient of change).

Persistence – Property of long memory of the system, whereby high or low values (excursions to high or low states) are clustered over longer time periods. Persistence is also referred to as autocorrelation or serial correlation.

Phase randomisation – Data generation technique that preserves the autocorrelation structure of the data series.

Power – A measure of how effective a test is. A test with high power is good at accepting the alternative hypothesis when the alternative hypothesis is true.

Principal component analysis – A method of analysis in which multivariate data can be simplified. A small number of new variables (principal components) can be used to represent the original dataset.

Probability density function – Function describing the distribution of data - it expresses the relative likelihood that a random variable attains different values.

Quartiles – The quartiles mark the $\frac{1}{4}$ and $\frac{2}{4}$ and $\frac{3}{4}$ positions in a dataset when the data are placed in size order. The interpretation of quartiles is as follows: *lower (first) quartile* - a value whose probability of exceedence is 75%; *second quartile (median)* - a value whose probability of exceedence is 50%; *upper (third) quartile* - a value whose probability of exceedence is 25%.

Random variable – A numerically valued function defined over a sample space. Can be thought of as something that provides observations.

Rank – The position of a data point when values are ordered by size. An observation has rank r if it is the r^{th} largest (or smallest) observation (depending on the direction of ordering).

Rejection region – If the value of the test statistics is in the rejection region, the null hypothesis is rejected.

Resampling – A method for testing data in which artificial data series are generated from the original dataset and these series are used to obtain an estimate of significance level.

Robust test – A statistical test which is not much affected by minor departures from assumptions on which it is based.

Sample size – Number of data elements in the sample.

Seasonality – Seasonal (periodic) behaviour of variable.

Segmentation – Method of finding abrupt changes in the time series by fitting a step-wise function (parameters: amplitude and time instant of a change).

Serial correlation: see autocorrelation.

Sign level – Probability of rejecting the null hypothesis when it is true (type I error).

Skewness – A measure of the asymmetry (See moments).

Smoothing – Replacement of the raw series by a more regular function of time that has less variability.

Standardization – Transformation of data by subtracting the mean and dividing by the standard deviation.

Stationarity – Property of a random variable whose statistical moments do not change with time. (Strict stationarity - first and second moments are constant with time. Weak stationarity - first moments are constant with time.)

Statistical test of a hypothesis – Formal procedure to check whether the hypothesis can be rejected in the light of available evidence.

Step change – Abrupt change in a time series.

Temporal dependence: see autocorrelation.

Test: see statistical test of a hypothesis.

Test statistic – Numerical value calculated from information contained in the sample. The test statistic is used to determine whether or not to reject the null hypothesis.

Tie – Situation where at least two elements in the data set have the same value.

Time series – A sequence through time of observed values of a variable. Time series techniques include spectral methods and ARMA (auto-regressive moving average) approaches and related tests.

Trend – Gradual change in a variable. Often assumed to be monotonic.

Two-sided test – Statistical test where the rejection region is located in two tails of the distribution.

Type I error for a statistical test – Rejecting the null hypothesis when it is true.

Type II error for a statistical test – Accepting the null hypothesis when it is false.

Univariate analysis – Analysis in which a single dependent variable is considered (of time and br space).

Unbiased (estimate) – An estimate is unbiased if its average (expected) value is equal to the quantity or parameter it estimates.

Variance – A measure of the spread in a sample/distribution. (See moments).

Variogram (called also *semivariogram*) – A technique used with spatial data to show how the covariance between points changes as a function of distance.

REPORTS PUBLISHED IN THE WORLD CLIMATE DATA PROGRAMME (WCDP)/

WORLD CLIMATE DATA AND MONITORING PROGRAMME (WCDMP) SERIES

WCDP-1	WMO REGION III/IV TRAINING SEMINAR ON CLIMATE DATA MANAGEMENT AND USER SERVICES, Barbados, 22-26 September 1986 and Panama, 29 September - 3 October 1986 (available in English and Spanish) - (WMO-TD No. 227)
WCDP-2	REPORT OF THE INTERNATIONAL PLANNING MEETING ON CLIMATE SYSTEM MONITORING, Washington DC, USA, 14-18 December 1987- (WMO-TD No. 246)
WCDP-3	GUIDELINES ON THE QUALITY CONTROL OF DATA FROM THE WORLD RADIOMETRIC NETWORK, Leningrad 1987 (prepared by the World Radiation Data Centre, Voeikov Main Geophysical Observatory) - (WMO-TD No. 258)
WCDP-4	INPUT FORMAT GUIDELINES FOR WORLD RADIOMETRIC NETWORK DATA, Leningrad 1987 (prepared by the World Radiation Data Centre, Voeikov Main Geophysical Observatory) - (WMO-TD No. 253, p. 35)
WCDP-5	INFOCLIMA CATALOGUE OF CLIMATE SYSTEM DATA SETS, 1989 edition - (WMO-TD No. 293)
WCDP-6	CLICOM PROJECT (Climate Data Management System), April 1989 (updated issue of WCP-1 19) - (WMO-TD No. 299)
WCDP-7	STATISTICS ON REGIONAL NETWORKS OF CLIMATOLOGICAL STATIONS (based on the INFOCLIMA World Inventory). VOLUME II: WMO REGION I - AFRICA - (WMO-TD No. 305)
WCDP-8	INFOCLIMA CATALOGUE OF CLIMATE SYSTEM DATA SETS - HYDROLOGICAL DATA EXTRACT, April 1989- (WMO-TD No. 343)
WCDP-9	REPORT OF MEETING OF CLICOM EXPERTS, Paris, 11-15 September 1989 (available in English and French) - (WMO-TD No. 342)
WCDP-10	CALCULATION OF MONTHLY AND ANNUAL 30-YEAR STANDARD NORMALS, March 1989 (prepared by a meeting of experts, Washington DC, USA) - (WMO-TD No. 341)
WCDP-11	REPORT OF THE EXPERT GROUP ON GLOBAL BASELINE DATASETS, Asheville, USA, 22-26 January 1990 - (WMO-TD No. 359)
WCDP-12	REPORT OF THE MEETING ON HISTORICAL ARCHIVAL SURVEY FOR CLIMATE HISTORY, Paris, 21-22 February 1990 - (WMO-TD No. 372)
WCDP-13	REPORT OF THE MEETING OF EXPERTS ON CLIMATE CHANGE DETECTION PROJECT, Niagara-on-the-Lake, Canada, 26-30 November 1990 - (WMO-TD No. 418)

Note: Following the change of the name of the World Climate Data Programme (WCDP) to World Climate Data and Monitoring Programme (WCDMP) by the Eleventh WMO Congress (May 1991), the subsequent reports in this series will be published as WCDMP reports, the numbering being continued from No. 13 (the last UWCDP report).

WCDMP-14	REPORT OF THE CCI WORKING GROUP ON CLIMATE CHANGE DETECTION, Geneva, 21-25 October 1991
WCDMP-15	REPORT OF THE CCI EXPERTS MEETING ON CLIMATE CODE ADAPTATION, Geneva, 5-6 November 1991 - (WMO-TD No. 468)
WCDMP-16	REPORT OF THE CCI EXPERTS MEETING ON TRACKING AND TRANSMISSION OF CLIMATE SYSTEM MONITORING INFORMATION, Geneva, 7-8 November 1991 - (WMO-TD No. 465)
WCDMP-17	REPORT OF THE FIRST SESSION OF THE ADVISORY COMMITTEE ON CLIMATE APPLICATIONS AND DATA (ACCAD), Geneva, 19-20 November 1991 (also appears as WCASP-18) - (WMO-TD No. 475)

WCDMP-18	CCL WORKING GROUP ON CLIMATE DATA, Geneva, 11-15 November 1991 - (WMO-TD No. 488)
WCDMP-19	REPORT OF THE SECOND CLICOM EXPERTS MEETING, Washington DC, 18-22 May 1992- (WMO-TD No. 511)
WCDMP-20	REPORT ON THE INFORMAL PLANNING MEETING ON STATISTICAL PROCEDURES FOR CLIMATE CHANGE DETECTION, Toronto, 25 June, 1992 - (WMO-TD No. 498)
WCDMP-21	FINAL REPORT OF THE CCI WORKING GROUP ON CLIMATE DATA AND ITS RAPPORTEURS, November 1992- (WMO-TD No. 523)
WCDMP-22	REPORT OF THE SECOND SESSION OF THE ADVISORY COMMITTEE ON CLIMATE APPLICATIONS AND DATA (ACCAD), Geneva, 16-17 November 1992 (also appears as WCASP-22) - (WMO-TD No. 529)
WCDMP-23	REPORT OF THE EXPERTS MEETING ON REFERENCE CLIMATOLOGICAL STATIONS (RCS) AND NATIONAL CLIMATE DATA CATALOGUES (NCC), Offenbach am Main, Germany, 25-27 August 1992 - (WMO-TD No. 535)
WCDMP-24	REPORT OF THE TENTH SESSION OF THE ADVISORY WORKING GROUP OF THE COMMISSION FOR CLIMATOLOGY, Geneva, 20-22 September 1995 (also appears as WCASP-34) - (WMO-TD No. 711)
WCDMP-25	REPORT OF THE FIFTH SESSION OF THE ADVISORY COMMITTEE ON CLIMATE APPLICATIONS AND DATA (ACCAD), Geneva, 26 September 1995 (also appears as WCASP-35) - (WMO-TD No. 712)
WCDMP-26	REPORT ON THE STATUS OF THE ARCHIVAL CLIMATE HISTORY SURVEY (ARCHISS) PROJECT, October 1996 (prepared by Mr M. Baker) - (WMO-TD No. 776)
WCDMP-27	SUMMARY REPORT OF THE MEETING OF THE THIRD SESSION OF THE CCI WORKING GROUP ON CLIMATE CHANGE DETECTION, Geneva, 26 February - 1 March 1996- (WMO-TD No. 818)
WCDMP-28	SUMMARY NOTES AND RECOMMENDATIONS FOR CCI-XII FROM MEETINGS CONVENED TO PREPARE FOR PUBLISHING THE FIFTH AND SIXTH GLOBAL CLIMATE SYSTEM REVIEWS AND FOR A PUBLICATION ON THE CLIMATE OF THE 20TH CENTURY, July 1997 - (WMO-TD No. 830)
WCDMP-29	CLIMATE CHANGE DETECTION REPORT - REPORTS FOR CCI-XII FROM RAPPORTEURS THAT RELATE TO CLIMATE CHANGE DETECTION, July 1997 - (WMO-TD No. 831)
WCDMP-30	SUMMARY NOTES AND RECOMMENDATIONS ASSEMBLED FOR CCI-XII FROM RECENT ACTIVITIES CONCERNING CLIMATE DATA MANAGEMENT, July 1997 - (WMO-TD No. 832)
WCDMP-31	REPORTS FOR CCI-XII FROM RAPPORTEURS THAT RELATE TO CLIMATE DATA MANAGEMENT, July 1997 - (WMO-TD No. 833)
WCDMP-32	PROGRESS REPORTS TO CCI ON STATISTICAL METHODS, July 1997 (prepared by Mr Christian-Dietrich Schönwiese) (WMO-TD No 834)
WCDMP-33	MEETING OF THE CCI WORKING GROUP ON CLIMATE DATA, Geneva, 30 January -3 February 1995- (WMO-TD No. 841)
WCDMP-34	EXPERT MEETING TO REVIEW AND ASSESS THE ORACLE-BASED PROTOTYPE FOR FUTURE CLIMATE DATABASE MANAGEMENT SYSTEM (CDBMS), Toulouse, France, 12-16 May 1997- (WMO-TD No. 902)
WCDMP-35	REPORT OF THE ELEVENTH SESSION OF THE ADVISORY WORKING GROUP OF THE COMMISSION FOR CLIMATOLOGY, Mauritius, 9-14 February 1998 (also appears as WCASP-47) - (WMO-TD No. 895)
WCDMP-36	REPORT OF THE MEETING OF THE CCI TASK TEAM ON CLIMATE ASPECTS OF RESOLUTION 40, Geneva, Switzerland, 10-11 June 1998 - (WMO-TD No. 925)

- WCDMP-37 REPORT OF THE MEETING OF THE JOINT CCVCLIVAR TASK GROUP ON CLIMATE INDICES, Bracknell, UK, 2-4 September 1998 - (WMO-TD No. 930)
- WCDMP-38 REPORT OF THE MEETING OF THE WMO COMMISSION FOR CLIMATOLOGY (CCI) TASK GROUP ON A FUTURE WMO CLIMATE DATABASE MANAGEMENT SYSTEM (CDMS), Ostrava, Czech Republic, 10-13 November 1998 and FOLLOW-UP WORKSHOP TO THE WMO CCI TASK GROUP MEETING ON A FUTURE WMO CDMS, Toulouse, France, 30 March-i April 1999 - (WMO-TD No. 932)
- WCDMP-39 REPORT OF THE MEETING OF THE CCI WORKING GROUP ON CLIMATE DATA, Geneva, Switzerland, 30 November-4 December 1998 - (WMO-TD No. 970)
- WCDMP-40 REPORT OF THE MEETING ON CLIMATE STATISTICS, PRODUCT DEVELOPMENT AND DATA EXCHANGE FOCUSING ON CLICOM 3.1, Geneva, 25-29 January 1999 - (WMO-TD No. 971)
- WCDMP-41 PROCEEDINGS OF THE SECOND SEMINAR FOR HOMOGENIZATION OF SURFACE CLIMATOLOGICAL DATA, Budapest, Hungary, 9-13 November 1998 - (WMO-TD No. 962)
- WCDMP-42 REPORT OF THE MEETING OF EXPERTS ON THE CLIMATE OF THE 20TH CENTURY, Geneva, 26-30 April 1999- (WMO-TD No. 972)
- WCDMP-43 REPORT OF THE. TRAINING SEMINAR ON CLIMATE DATA MANAGEMENT FOCUSING ON CLICOM/CLIPS DEVELOPMENT AND EVALUATION, Niamey, Niger, 03 May-10 July 1999, (WMO-TD No. 973)
- WCDMP-44 REPRESENTATIVENESS, DATA GAPS AND UNCERTAINTIES IN CLIMATE OBSERVATIONS, Invited Scientific Lecture given by Chris Folland to the WMO Thirteenth Congress, Geneva, 21 May 1999 - (WMO-TD No. 977)
- WCDMP-45 WORLD CLIMATE PROGRAMME – WATER, DETECTING TREND AND OTHER CHANGES IN HYDROLOGICAL DATA, Zbigniew W. Kundzewicz and Alice Robson (Editors) – (WMO-TD No. 1013)
-