Institute of Hydrology

Centre for Ecology & Hydrology

Natural Environment Research Council

Centre for Ecology and Hydrology

Component Institutes
Institute of Freshwater Ecology
Institute of Hydrology
Institute of Terrestrial Ecology
Institute of Virology & Environmental Microbiology

Natural Environment Research Council

A review of methods for detecting changes in hydrological data

Alice Robson
December, 1997

Prepared by:

Institute of Hydrology
Crowmarsh Gifford,
Wallingford
OXON
OX10 8BB

Under the Special Service Agreement with WMO, Geneva, 6 October, 1997. Ref 16.322/a/CNS

1.0 Introduction

This report provides a review of methods for examining changes such as trends and breaks in hydrological data series. It addresses the following requirements as agreed under the WMO special service agreement under project WCP-Water A.2

(a) prepare a basic report reviewing the methods and software packages currently available and under development for the investigation of trends, cycles and breaks in time series of hydrological data.

(b) prepare recommendations as to those methods which the WMO might recommend for use by members and by the Global Runoff Data Centre.

The report sets the scene for the WMO/UNESCO Workshop on trends and changes in hydrological data to be held in late 1998. It discusses statistical methods and software and makes preliminary recommendations regarding the planned workshop.

2.0 Methods for detecting changes in hydrological data

2.1 Hydrological change

Change in hydrological time series can take place in a variety of ways. In order to assess possible changes in a data series, the following may need to be considered

- gradual/progressive trend in the mean (usually assumed to be monotone)
- step change in the mean (a jump in the series)
- changes in variability
- seasonal variation
- climatically driven fluctuations occurring over the medium term (10 to 30 years)
- changes in persistence

Usually it is important to distinguish whether changes are directly linked to man's influence or whether they have a climatic origin. Step changes are usually due to man (eg. from building a reservoir). Trend can arise from either natural or unnatural causes (eg. progressive urbanisation). Seasonal, cyclic and other fluctuations are most likely to have a climatic origin. Under a situation of climate change, there may be a trend in the mean, the extremes or changes in variability or persistence.

Numerous statistical methods are available for testing for *trend* and for *step change* (see below). In these cases, the main issue for the practitioner is one of selecting the most appropriate methods for the particular circumstance. Statistical methods also exist for testing for *change in variability*, although such tests are rather less abundant.

Seasonality in data is frequently of concern in that it may need to be accounted for when carrying out tests for trend and step change. Some statistical methods allow for seasonality. For others, it may be sensible to "remove" the seasonal component of the data before applying tests. Seasonality may also be an issue if trends occur during particular seasons, or if the timing of a seasonal event is important (eg snowmelt).

Handling aspects such as *fluctuation* (other than seasonal variation) is rather more difficult, particularly where series are of only moderate length. In many parts of the world the climate shows fluctuation across periods of 10 to 20 years. It would be convenient to characterise these fluctuations as cycles, and some authors have attempted to do so (eg Currie and O'Brien, 1990). However, such fluctuations may well arise out of chaotic behaviour of the system (see in particular, Burroughs, 1992). On balance there is little conclusive evidence of climatic cycles over the medium term. If cycles are sought they are often found, but they are unlikely to persist. In view of this, formal testing for cycles is probably best treated with extreme caution within change detection methods.

Climatically driven fluctuations in hydrological time series pose a serious problem for interpretation of statistical tests for trend and step change. If an observed trend is in reality just part of a fluctuation then it is unlikely to persist, and ought not to be viewed as evidence of climatic change. Addressing this problem is not easy but we discuss promising approaches in section 2.4.

2.2 Data types and properties

Testing for trend and other forms of change is dependent on both the properties and resolution of available hydrological data. Examples of relevant hydrological series include, amongst others, flows, water level, rainfall and temperature. Such variables can be represented in a number of ways including

(1) daily series
(2) monthly series
(3) extreme value series eg annual maximum, Peaks-Over-Threshold data, drought indexes
(4) series describing river regime eg seasonal distribution, timing of snowmelt

With each form of series and each type of data there are different aspects which need to be taken into consideration. Examples include:-

- *distributional assumptions*. Many statistical tests require distributional assumptions to be made. Assumptions of Normality are likely to be reasonable when temperature data are considered, but are inappropriate when extremes (types 3-5 above) are in question.

- *seasonality and autocorrelation (persistence)*. Daily and monthly series generally show seasonal effects and significant autocorrelation (ie persistence from one time period to the next). This needs to be allowed for when carrying out statistical tests. Annual data typically show less of such effects although persistence occurs in series such as ground water levels and even some river flows.

- *missing data*. Some statistical methods can be applied even if there are gaps in the data. In other cases, infilling of series may be required and this in itself can be a complex and time-consuming procedure that may not always be practical.

- *spatial effects*. Spatial aspects may need to be addressed in some cases. This is perhaps most important in understanding droughts since both temporal and spatial aspects contribute.

- *series irregularity*. Some statistical tests can be applied to both regular and irregularly spaced time series, others are only suited to regular data series. Peaks-Over-Threshold (POT) data provides an example of an irregular series; although irregularity increases complexity, in this case the POT data allows changes in frequency as well as magnitude to be assessed.

The above examples underline the need to apply different statistical tests to different data series. However, some statistical tests are applicable in a wider range of circumstances; preferential use of such methods would mean that a similar portfolio of tests could be used on a range of data types.

2.3 Statistical methods

There are numerous statistical tests available for examining trends and breaks in data. Cadavias (1992) lists a substantial number of tests that been used in hydrological studies and provides references. Robson and Reed (1992) also review tests applicable to detecting trend and step change in floods.

To evaluate all possible tests is beyond the scope of the current report. Instead, we consider statistical tests in various broad categories and discuss advantages and disadvantages.

Parametric tests

Parametric tests are tests which are based on a distributional assumption, eg that data are normally distributed. A simple case is that of ordinary linear (least squares) regression. Such tests are widely available, easily used and, if assumptions are met, are powerful. The tests generally require assumptions of independence and of identically distributed observations. Assumptions of independence would not hold for most daily series and many monthly series because of the autocorrelation in the series.

Non-parametric tests

Non-parametric tests require rather less stringent assumptions than parametric tests. Common examples include Kendall's Tau and distribution-free CUSUM tests (Chiew and McMahon, 1993). Many non-parametric tests use the ranks (ordering) of the values rather than the actual observations and thus avoid the need to assume a distribution. Such methods are often much safer to use and are termed *robust* tests. Nevertheless, they are not devoid of assumptions; in particular, the assumption of independence generally remains and must not be ignored. Even with robust procedures, daily and monthly series are unlikely to meet all the assumptions.

Whilst it is not possible to discuss all available tests, we make special mention of the seasonal Kendall test (Hirsch et al, 1982, Hirsch and Slack, 1984). This test allows for seasonality in the data and for a certain level of autocorrelation. It is therefore well suited to application to monthly environmental data and has rightly seen widespread use in the hydrological world.

Permutation tests

Permutation tests are really a special case of non-parametric tests, but they are of particular value in many hydrological applications and therefore merit separate consideration.

Permutation tests involve using the data to determine the significance level of a selected test statistic (Maritz, 1981). This requires permuting the data many times.

To understand how a permutation test works, consider the example of a linear regression of annual temperature on year. Assume, to begin with, that temperature has not changed over time. If that is true then it should not matter very much if the original dates were re-ordered - ie, the gradient of the regression line should not change very much. To carry out a permutation test, the data are shuffled very many times and each time the relevant test statistic (in this case the regression gradient) is calculated. This generates the distribution of possible values of the test statistic under permutation, called the *permutation distribution*. The permutation distribution usually depends on the data and must be recalculated for each data set. To test for a 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.

Like the other robust tests, permutation tests require few distributional assumptions. However, they offer advantages in that a carefully applied permutation test can allow for dependencies in the data such as seasonality and autocorrelation; this type of data structure can be incorporated by permuting the data in blocks of years. Permutation tests can also be very valuable when examining trend at a regional, rather than at a site-specific level (Robson et al., 1998).

Time series based tests

Time series techniques include methods such as ARMA and spectral methods and tests based on these approaches (see Cadavias, 1992 for various examples). Time series techniques allow structure such as autocorrelation in the data to be incorporated. Usually, regular and complete data are a requirement and Normal variation is assumed. Time series techniques may be extended to incorporate an input variable, such as rain, as well as the usual output variable.

Modelling based methods

Modelling based methods can be very varied. Time series models in which input variables are used are one form of modelling. Alternatively hydrological modelling techniques may be used, for example, to link rainfall and flows. In general, modelling methods tend to be complex and are difficult to apply consistently across numerous and varied sites. However, for specific sites valuable information may be obtained from modelling. A modelling approach to assessing trends should be regarded as a task for which considerable expertise is required.

An intermediate position might be to make use basic use of covariate information but avoid detailed modelling. At its simplest this might involve fitting a linear relationship between two variables and looking for trends in the residuals. Alternatively multivariate approaches might be considered.

2.4 Other components of detecting change

Application of statistical tests to a data series forms only one component of a full statistical analysis. Where particular statistical tests are to be recommended, it is also appropriate that other stages of the analysis should be laid out. Aspects that are likely to merit consideration are

(1) selection of suitable data
(2) exploratory data analysis.
(3) identification of suitable tests
(4) checking of test assumptions
(5) application of statistical tests
(6) interpretation of test results
(7) taking a regional perspective

(1) Selection of suitable data

To detect changes, data series generally need to be long and of good quality. For trend detection, records should be at least 30 years long and preferably considerably longer. If the objective is to test for climate change, it is critical that records are free of anthropogenic effects. Removal of the effects of anthropogenic change is very difficult. It may be necessary to restrict analysis to pristine sites, ie. rivers free of regulation, urbanisation and major land use change.

(2) Exploratory data analysis

Exploratory data analysis allows the analyst to obtain a good understanding of the data through graphical methods. Amongst other things, it could include use of:-

Time series plots: Extremely useful in understanding the data and allowing outliers and gaps in the data to be readily recognised. Such plots are often more informative than use of summary statistics (which can be included as part of the time series plot).

Smoothing techniques. Smoothing techniques are now widely used for hydrological applications (Hirsch, 1991; Bradley and Potter, 1992; Robson and Neal, 1996). The addition of a smoothed curve to the raw data visually enhances the data and focuses the eye on important underlying features. Smoothing techniques can also be used to decompose a series into trend, long term fluctuation, seasonal cycles and random components (Cleveland, 1979; Statistical Sciences, 1993). Use of a smoothed curve can highlight medium-term fluctuations in the data. Possible trends can appear insignificant or irrelevant once compared to the general underlying level of fluctuation (Robson and Neal, 1996; Robson and Reed, 1997). This effect is most marked for shorter records and decreases as record lengths increase.

Regional plots. Examining single-site data in isolation from other sites in the locality often overlooks important information. For example, a possible outlying data point may be seen to have been a widespread phenomenon once compared with other data for sites in a region. Comparison with other sites may also lead to identification of a man-made change or a data reporting problem (eg where a step change effect is seen in a series but does not appear at an upstream site). Application of smoothing techniques to sites in a region may show common underlying fluctuations - suggesting that these are climatic effects that extend over a wider area. Regional comparisons are not difficult to carry out. A page of small time series plots using identical time scales is often sufficient.

(3) Identification of suitable tests

Selection of which tests to use depends on many factors. These include the study objectives, the availability of data, properties of the data, requirements for robustness, ease of use, and the power of a test.

For each main type of data (see section 2.2), it should be possible to provide a set of basic recommendations . Documentation on the underlying assumptions, the recommended scope of a test and its likely ability to detect change will allow users to adapt the recommendations to particular circumstances.

(4) Checking test assumptions

For each test, it is necessary to know the main assumptions being made and to verify that these are not badly violated. Issues that may require consideration are autocorrelation, normality, independence, medium range fluctuation and seasonal structure. Simple plots and tests can be used to carry out these checks. Guidance will be needed on how to go about these checks and how to interpret them (eg how non-normal does data have to be before certain tests become inappropriate).

(5) Application of statistical tests

Details of the basic methodology will be required for each of the tests.

(6) Interpretation of test results

Given modern computing resources it is often possible to apply a range of possible tests to different aspects of a data series eg mean level, extremes, variability. The user is likely to need to interpret a large number of results. Effective presentation can aid interpretation considerably and there is scope for providing much greater guidance on this aspect. Robson and Reed, 1996 provide examples of methods for presenting results from multiple statistical tests.

Other advice on interpretation will also be needed. For example, common sense should always be used; statistical significance does not always imply that an observed change is important; if you apply enough tests to enough sites then some significant results will occur by chance.

(7) Regional perspectives

Interpretation of trends can often be much improved by looking at the regional picture. If many sites in an area show similar changes this strengthens the evidence. If a trend is solitary it may derive from chance occurrence. In some cases, regional analysis may be appropriate and recommendations for useful approaches could be made.

3.0 Software

3.1 Software types

Numerous software programs abound and it is not possible to describe all the statistical packages. Lists of statistical software can be found on some web-sites (eg see hhtp:/milkyway.stats.gla.ac.uk/cti/links_stats/software.html). Here we outline the broad types of software that might be considered.

General purpose statistical packages

These vary from fairly limited PC packages aimed at the semi-numerate scientist with menu-driven analysis, through to sophisticated statistical programming languages. Packages that are designed to be simple to use will generally not provide specialist trend tests. Packages that provide a programming environment allow the user to implement any method they choose, but require a much higher level of expertise to be used (eg Splus, SAS). Some programmable packages would allow an expert to set up tests and then to present these in a user friendly menu-driven format which would be accessible to the non-expert.

Specific-task statistical packages

There are a number of statistical packages which address a special area of statistics eg time series packages. Some of these are of value in particular circumstances but no single one is likely to meet the full requirements for trend detection.

Specialist statistical hydrology packages

These are likely to be specially designed software packages and might range from commercial ventures which are menu-driven through to provision of coded routines eg the FORTRAN program developed under WMO in 1988 for testing trends. It is unlikely that commercially produced software packages will contain all the statistical methods which might be considered for recommendation.

3.2 Comments

It should be remembered that methods that WMO might wish to recommend are likely to be quite specialist. A choice may need to be made between

(1) developing a package which offers the desired features in a user-friendly environment

(2) leaving the software to the user - which will necessitate the user having a certain level of programming expertise

(3) limiting recommendations to available user-friendly software; this may prove very restrictive

Cost and available hardware will also need to be considered.

4.0 Suggestions as to the scope of the workshop

The proposed workshop stems from the seventh WMO planning meeting in Koblenz (WMO, 1997).

> *In discussions on the search for signals of variability and change in hydrological time series where WCP-Water Project A2 is relevant, it was noted that there are a number of methods and software packages available for investigating trends, cycles and breaks*

in records. Consequently it was suggested that a small workshop be organised to bring together those experts in this particular field to provide guidance on which of these methods is appropriate to apply in certain circumstances. It was suggested that such a workshop might be held in the UK and that the Institute of Hydrology should be approached in anticipation of WMO and UNESCO support.

The thoughts given below are intended primarily as the basis for discussion in order to better formulate the scope of the proposed trends workshop.

(1) The basic goal of the workshop is to come up with recommendations. The recommendations should be aimed at being accessible to a reasonably numerate scientist. The workshop will allow exchanges of experience and expertise but this should not be the primary objective.

(2) Recommendations should take the form of general guidelines and should go further than just providing a list of tests. An outline of possible contents for these guidelines is given in section 2.4.

(3) The workshop should include methods for detecting trend in:-

mean values
extremes (floods and droughts)
variance and persistence
seasonal aspects (eg timing of snowmelt)

A secondary issue will be methods for detecting breaks. Tests for cycles were mentioned in the original remit (see above) but it is not recommended that these form part of the workshop. The workshop may need to discuss how to handle medium to long term fluctuation in the data.

The Workshop should consider changes in extremes (notably floods and droughts) as part of its remit; increased frequency of floods or droughts is often more critical to a local population than a shift in mean levels. Methods for trends in floods are fairly well known, but it should be anticipated that trend in droughts is a very complex problem and further scientific developments are likely to be required before standard procedures can be recommended.

(4) The workshop must avoid simply producing an exhaustive list of possible statistical procedures with pro's and con's. It should aim to identify a *practical selection* of tests. It may be preferable to recommend that a scientist applies a range of tests, rather than try to identify a single "best" test. The recommendations may need to be restricted to fairly simple procedures. In selecting tests it will need to balance statistical power (how good tests are), robustness, general applicability and ease of use.

The number of tests applied should probably be limited to 10-12 for any one type of series (eg monthly flows). The tests would include perhaps 4-6 tests for trend, 2-4 tests for step change and 1-2 tests for change in variability. Tests should be chosen to be reasonably robust and to be applicable in a range of circumstances. However, recommended tests need not be the same for all data types, eg. extreme values vs monthly means, or flows vs temperatures.

Guidance on appropriate methods for each type of data will be a key goal of the Workshop.

(5) The Workshop should focus on making the best use of available methods which have already seen practical application. Identification of areas where statistical development is required will be of passing concern.

(6) The methods identified should be suitable for a numerate scientist. The recommendations may however include brief comments on more complex methods which a statistical expert might wish to consider.

(7) The issue of software should be considered towards the close of the meeting after a consensus on best statistical methods has been reached. The workshop should provide guidance on available software identifying what it is capable of, the level of expertise required to use it, and the likely costs. It is unlikely that software will be available to meet all the criteria. It is probable that the Workshop will recommend that software should be developed.

References

Bradley, A.A and Potter, K.W., 1992. Flood frequency analysis of simulated flows. Water Res Res., 28, 9, 2375-2385.

Burroughs, W.J., 1992. Weather Cycles, Real or Imaginary? Cambridge University Press. 201 pp.

Cadavias, 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 WMO/TD - No 534.

Chiew, F.H.S. and McMahon, T.A., 1993. Detection of trend or change in annual flow of Australian rivers. Int. J. of Climatology., 13, 643-653.

Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. J. Am. Statis. Soc., 74, 386, 829-836.

Currie, R.G. and O' Brien, D.P., 1990. Deterministic signals in USA precipitation records. Int. J. Climatology. Vol10, 795-818.

Hirsch, R.M. and Slack, J.R., 1984. A nonparametric text for seasonal data with serial dependence. Water Res. Res., 20, 6, 727-732.

Hirsch, R.M., Alexander, R.B. and Smith, R.A., 1991. Selection of methods for the detection and estimation of trends in water quality. Water Res. Res., 27, 5, 803-813.

Hirsch, R.M., Slack, J.R. and Smith, R.A., 1982, Techniques of trend analysis for monthly water quality data. Water Res. Res. 18, 1, 107-121.

Maritz, J.S., 1981. Distribution free statistical methods. Chapman and Hall, 1-264.

Robson, A.J. and Neal, C., 1996. Water Quality trends at an upland site in Wales, UK, 1983-1993. Hydrol. Proc., 10, 183-203.

Robson A.J. and Reed, D.W., 1996. Non-stationarity in UK flood series. Flood Estimation Handbook Note 25. Report to MAFF, Project FD0409, 37 pp + appendices.

Robson, A.J., Jones, T.K., Reed, D.W and Bayliss, A.C., 1998. A study of national trend and variation in UK floods. Int J. Climatology. In press.

Statistical sciences, 1993. S-PLUS Reference Manual, Version 3.2, Seattle: StatSci, a division of MathSoft, Inc., 1993.

WMO, 1997. Seventh planning meeting on world climate programme- Water, Final Report. WCASP-45. WMO/TD-No. 854.