18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

Decision Support System for Temporal Trend Assessment of Water Quality Data

Sunil Tennakoon[1], David Robinson[2] and Shuangyang Shen[1]

[1] *eWater Cooperative Research Centre, Department of Environment and Resource Management No 80, Meiers Road, EPA Building, Indooroopilly, QLD 4075.*
[2] *Victorian Environment Protection Authority (EPA), Centre for Environmental Sciences, Ernest Jones Drive, Macleod VIC 3083*
Email: sunil.tennakoon@epa.qld.gov.au

Abstract: The current interest in water quality improvement plans across the country and the wide range of stakeholder involvement have imposed a new paradigm for developing scientifically robust and technically sound software tools for water quality monitoring and assessment. Trend detection is an essential part of water quality assessment, which can help with identifying causal factors for the impairment of water quality, confirming the effectiveness of management actions and establishing a need for management intervention. Discovery of temporal water quality trends is a complex process due to the characteristics of water quality data (non-normal distribution of data, seasonality, flow interactions, missing values, values below the limit of detection, and serial correlation). The preparation of time series data sets and the checking for suitability of available data for trend detection are essential prerequisites for trend assessment.

We have designed and developed a user friendly decision support system (DSS) for water quality trend assessment to assist and improve water quality data analysis by various stakeholders including catchment groups and regional bodies. This DSS is a component of a water quality data analysis package developed by eWater CRC. The trend analysis DSS contains steps of data validation, consideration of detection limits, time series data management, statistical testing, modelling and decision-making to help explain the trends. This tool can import time series data from a variety of formats and databases. The tool comes with advanced visualisation capabilities and includes a number of data smoothing techniques. Using this tool, users are able to perform explanatory data analysis before stepping into complex modelling processes. The DSS has been designed using contemporary science and state of the art DSS technology, which includes interactive tools, knowledge bases, statistical tools, decision-making processes, and a comprehensive collection of relevant information and help materials.

The prototype has been evaluated and validated using historical water quality data collected from Queensland Department of Environment and Resource Management and the Victorian EPA. This DSS will be released nationally through the eWater CRC ecological management and restoration product development program.

Keywords: *Water quality, Trend analysis, Time series, Decision support systems*

1. INTRODUCTION

In recent times, a significant amount of resources have been allocated for water-quality monitoring, assessment and improvement. This is mainly because of on-going deterioration in many aquatic ecosystems, and issues related to environmental values and protected assets. Various agencies and natural resource management groups are involved in water-quality management activities and can benefit from tools that support the design of monitoring plans and assessment of water-quality data. eWater Cooperative Research Centre (eWater CRC) is well placed to assist this process through its research and product development programs in collaboration with a range of research partners and water managers. The CRC has designed and developed a software package named "Water Quality Analyser" (Tennakoon 2008) for water-quality data processing, analysis and assessment; the water-quality trend analysis tool is a major component of this package.

The detection and estimation of temporal or spatial trends is important for many environmental studies or monitoring programs. The identification of trends in water quality can also be used to either confirm the effectiveness of management actions or to establish a need for management intervention. Many water-quality monitoring networks have been set up with the primary objective of detecting temporal trends in water quality (ANZECC 2000). However detection of trends in water quality is not a trivial process as water quality can vary spatially and temporally for many quite natural reasons. Natural variation must be taken into account in the design of any water-quality monitoring program if the data are to be useful for trend detection or any other purposes.

Trend assessment of water quality involves a number of steps such as data checking, data processing, visualisation, exploratory data analysis (EDA), mathematical modelling and explaining. This process can take a significant amount of time and effort. A lack of knowledge or experience, or misuse of data, can generate inaccurate results. Therefore we have designed and developed a Trend Tool which is a decision support system (DSS) for data checking, data preparation, visualisation and exploratory data analysis to support the entire water-quality trend assessment process. It has been designed based on an existing trend tool developed in the CRC for Catchment Hydrology (Chiew and Siriwardena, 2005) which contains robust statistical techniques for testing trends, change and randomness of time series data. The new Trend Tool, however, emphasises the pre-processing of data and exploratory analysis of water-quality trends rather than complex statistical and mathematical processes for quantifying the significance and magnitude of trends. The new tool comes with powerful visualisation capabilities and a few selected standard statistical tests for water-quality trend detection, including some customised report generation ability (Figure 1).

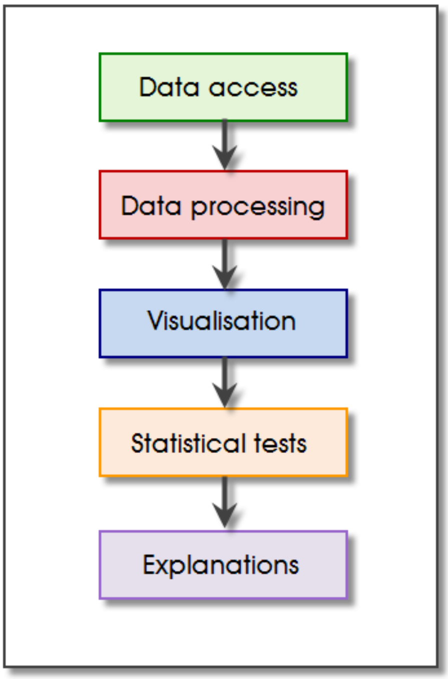

Figure 1 - Basic logical structure of the trend tool

The software contains a comprehensive help system including guidance for selecting methods appropriate to the quantity and quality of available data. A user-friendly user interface has been developed for easy operation, with the facility of visualising predicted and estimated outputs. The software tool was tested and evaluated for its usability and possible errors during the prototype stage. The Trend Tool will be available through the eWater CRC Toolkit website (www.ewatercrc.com.au/toolkit).

2. WATER-QUALITY VARIABILITY: IMPLICATIONS FOR TREND ASSESSMENT

The normal constituents of water (water quality) normally exhibit great variability. This variability can be caused entirely by natural factors, or can be a result of human factors such as point and diffuse sources of pollution, or a combination of both. The concentrations of diffuse water-quality constituents are often correlated with stream flow, land use, management practices and geology. The causes of the relationships may vary from constituent to constituent (Hirsch et al., 1982). The detection of water-quality trends in indicator concentrations is often complicated by the presence of flow-related variability in water-quality records. Flow-related variability may be large relative to the magnitude of change in water quality resulting from human activity in the catchment. Assessment of water-quality trends needs to make allowance for possible flow interactions by focussing on tests for trend in flow-independent water-quality concentrations, or flow-adjusted concentrations, or at least to consider trends in flow and their possible effects on trends in concentration.

Spatial Variability: Spatial variability can potentially be confounded with temporal variability and the contributing factors to spatial variability need to be understood if they are to be controlled. There should be no systematic changes in the spatial distribution of samples over time if confounding of spatial variability and temporal variability is to be avoided. A water body's geographical position within a catchment or on a broader regional scale can correlate strongly to climate and rainfall distribution, which in turn can affect water quality. Adequate coverage of spatial variation, including affected sites and sites in natural condition, is essential not only for descriptive purposes but also to include such information as, for example, geographic patterns in trends, which can give important clues to causes.

Temporal Variation: Cyclic variation in water quality can operate on a number of scales such as daily, seasonal and over a number of years, decades or even longer. The daily cycle in light and temperature can have a strong effect on many water-quality parameters (including biotic factors which can affect water quality). Therefore one needs to ensure that trends in the time of day of sampling are not a significant factor that could be producing consistent differences between sites or at the same site over time. To control the influence of the time of day of sampling a water-quality monitoring program needs to either standardise sampling times or, better still, to randomise sampling times through all of or part of the day.

To adequately define seasonal cycles or within-year variations requires samples to be taken frequently over a number of years, and it is important to ensure that all seasons of the year are sampled with equal intensity. Monthly and quarterly sampling frequencies are often used. Whatever the sampling interval used it is important to randomise or spread the sampling times throughout that interval. The longer the interval the more important it is to pay attention to randomisation of the sampling time, and the longer it takes to accumulate sufficient data to describe seasonal and inter-year variation and detect trends.

Sampling Frequency: The frequency of sampling is the single most critical aspect of water-quality network design affecting the power of a program to detect trends. Lettenmaier (1978) recommends that the best sampling frequency for a trend detection program would be at least monthly (for statistical power) and not more than fortnightly due to considerations of serial correlation. Investigations done for the review of the Victorian water-quality monitoring network (VWQMN) indicated that at least 8 years of monthly data is needed to detect a change of one standard deviation in the mean (EWQMC 1996). With five years of data only trends of high magnitude would be detectable. Data sets with longer periods are required for clear examination of trends, to observe long-term cycles and to distinguish one from the other.

3. MAIN STEPS IN ASSESSMENT OF WATER-QUALITY DATA FOR TRENDS

The required quantity and quality of data for trend analysis is often not available, and significant effort is required to process/validate the available data before it can be used for trend assessment. A systematic approach is required to assess and process water-quality data for temporal trends to avoid potentially confounding factors. The following steps are recommended as an efficient process that minimises false trends being identified and repeated back-tracking to eliminate spurious data points.

- Data Validation
- Visual Assessment/ Exploratory Data Analysis
- Statistical Testing
- Modelling and explanation of the trends.

3. 1 Data validation

As the first step, preliminary evaluation of the data should be carried out using graphical representation followed by some basic statistical (quantities) tests. By reviewing the data both graphically and numerically (Figure 2), it is possible to learn the "structure" of the data and thereby identify appropriate approaches and limitations for trend analysis. Then a number of important data processing steps need to be undertaken before performing the actual statistical test for trend detection. Outlier-checking, dealing with detection limits (Hirsch and Slack, 1984), consideration of sampling time and frequency, dealing with missing values and removing extra values are some of the important steps in this data-validation process. Data processing is done in the data management module available in Water Quality Analyser software package.

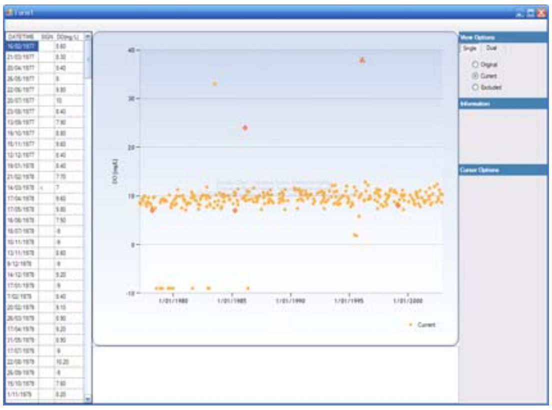

Figure 2 – Numeric and graphical data visualisation

3.2 Visual assessment and exploratory data analysis

Visual assessment and exploratory data analysis (EDA) is an essential component of any trend assessment. The human brain and visual system is very powerful at identifying and interpreting patterns, and a well conducted EDA and visual assessment may eliminate the need for a formal statistical analysis. Sole reliance on statistical test results can be meaningless without a proper understanding of the data. EDA involves using graphs to explore, understand and recognise patterns. EDA is an iterative process where graphs are plotted and refined so that important features of the data can be seen clearly.

3.3 Statistical analysis

Once a possible trend has been identified from visual analysis of the data, a statistical test can be used to resolve uncertainty about a possible trend. Trend assessment should never rely on an uncritical application of statistical tests to large amounts of data without data checking and some visual assessment. Parametric or non-parametric statistical tests should be selected based on data distribution. Erroneous results and conclusions can result if the background assumptions of the particular statistical test are not met (normal or non-normal distribution). Serial correlation in the data also violates the assumption of independence of the data (no short-term correlation between samples). When serial correlation occurs the “p” values calculated using statistical tests will be too low and trends may be falsely identified (Helsel et al., 2006).

3.4 Explaining the trends

Identification of a trend is a first step, but probably of more importance is explaining a trend. Being able to explain a trend means that one can properly assess its significance and suggest appropriate management actions to correct it, if needed. A subjective assessment would need to be based upon a comprehensive knowledge of the water-body and its catchment and the history of potential disturbances and influences to the water-body. In such instances a trend may be considered evidence for a particular cause and effect relationship. Such an assessment may provide sufficient insight for the development of management responses. However, often a more quantitative explanation of a trend is required. Such an explanation of a trend requires a model which links variation in the indicator in question with a number of other variables. This is a complex task requiring not only information collected with the water-quality information (e.g. streamflows, rainfall, other indicators and environmental information) but the ability to link in with other data sets, e.g. through GIS.

4. SYSTEM DESIGNS AND DEVELOPMENT

The Trend Tool software has been designed for water-quality trend analysis using time series of water-quality data. The features available in the tool include data access and processing, visualisations, exploratory data analysis, statistical tests and facilities for generating reports.

4.1 Data processing

A comprehensive time series data-processing facility has been built into this system. Users are able to perform these data-processing activities interactively with the visualisation facility. A few important data-processing facilities built into this tool are described below.

Data Import/Export: The tool can access data from compatible data formats which include the commonly used *.csv format. Selected indicators from a data series can be presented visually with the capability to annotate or exclude certain data. After processing, a number of data-export features in various formats are also available. Metadata of time series data sets can also be stored and shared using a MS SQL compact file-based database. This can store information about water systems, including geo-data (absolute and relative), indicator group, indicator, water type-specifics, guidelines, etc. This information is useful in the assessment of water-quality trends.

Data Validation: Loaded time series data can be visually assessed and validated by removing any outliers and dealing with detection limits. Basic statistics are automatically calculated and constantly updated while the user is working with the data. Statistical graphs such as moving averages, smoothing curves and various other visual aids (e.g. frequency plot, box plot, etc.) can be created to help in the exploratory assessment. The cleaned-up data can then be exported, saved or directly used for further analysis. Data can also be aggregated to a user-specified time interval for trend analysis.

Visualisation: The time series plot is the most useful visual tool for analysing trend or change. The variable of interest can be plotted against time as a scatter or line plot, and a trend line can be fitted to the data (Figure 3). Techniques for fitting trend lines include moving average, linear regression, quadratic regression, and LOWESS smoothing (Cleveland, 1979). Visual inspection of plots of the raw or transformed data together with smoothed curves superimposed can clearly indicate the type, direction and magnitude of a trend as well as revealing long term cycles and other patterns in the data such as linear, monotonic, curvilinear, step trend and trend reversals.

Figure 3 – Data smoothing using the LOWESS method

4.2 Statistical analysis

The Trend Tool contains a number of statistical techniques for testing trend, change and randomness of time series data tests. A formal statistical test can be useful to resolve uncertainty about a trend or non-trend identified through visual analysis. However, a trend assessment should never rely on statistical tests alone. The choice of test will depend on the data distribution, occurrence of extreme values or outliers, non-detect values, missing values, the possible presence of serial correlation, and the preference of the investigator.

Linear regression: Simple linear regression is used to assess the statistical significance of the trend line fitted to the data; that is, it can be used to test if the slope of the trend line is significantly different to zero.

Kendall and Seasonal Kendall: Kendall's Tau (Kendall, 1975) is used as a basic non-parametric test for trend testing. Its only background assumption is that the random variable is independent and identically distributed. It is a robust test that can be used with data that is non-normally distributed, has missing values, and has values below the detection limit, serial correlation, and non-linear but monotonic trends. In the Seasonal Kendall test only data pairs from the same season are compared. This has the advantage that the background assumption can now be significantly relaxed because the random variable need only be identically distributed in like seasons (Hirsch et al., 1982). Season can be defined by the investigator, but is usually individual months when monthly time series data are used. The effects of serial correlation are minimised in the Seasonal Kendall test. The Seasonal Kendall test was modified (Hirsch and Slack, 1984) to account for serial correlation. Hirsch and Slack (1984) suggest using the modified test if there are more than ten years of data, as it commonly takes this much data to detect serial correlation, if present. If no serial correlation is present then the modified test is less powerful than the un-modified Seasonal Kendall test.

Although generally a very robust test, if the trends in the data set are not monotonic (i.e. they change direction) or there are opposing trends in different seasons then the power of the test will be greatly weakened because such opposing trends will cancel out in the testing procedure.

Testing for a difference between two periods: Sometimes a comparison of step changes is required, rather than linear or monotonic trend assessment (Hirsch et al., 1991). This occurs when there is a natural break in the data between two data collection periods that is greater than about one-third of the total data collection period, or when a specific event has occurred at a specific time that could affect water quality, such as commissioning of a sewage treatment plant. Tests for the difference between two means (Student's t test) or a non-parametric version that tests for a difference between two medians (Rank Sum test or Mann-Whitney test) could be used. To use the Student's t test the data from each period must be normally distributed and have the same variance.

4.3 User Interface

The Trend Tool user interface has been designed to be user friendly and it is highly graphically driven. The uploaded time series data are plotted in the central chart window. A data grid is also presented side-by-side, on the left hand side. The available context-sensitive toolbox panel provides view options of the chart, data statistics, select, zoom, pan and other required manipulations for customising the chart. The users are able to select data points on the chart to be included or excluded, or to tie them to a detection limit. Generation of data-smoothing curves, data-aggregation, and statistical trend analysis can be performed simply by using options available in the screen (Figure 4). A standard help system has been built-in to the system to assist users, explaining basic data-processing details and concepts of trend detection.

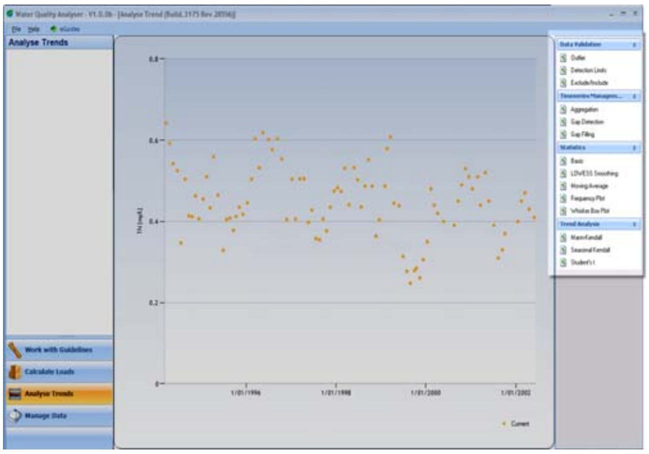

Figure 4 – Trend tool user interface

4.4 Code Development

The Trend Tool is built on the latest Microsoft .NET Framework 3.5 and coded in C#. The Microsoft .NET Framework 3.5 is redistributed with the software installer to ensure that users who do not already have the .NET framework installed can easily install and run the Trend Tool. Most of the component parts of the Trend Tool, including data graphical validation and trend analysis control have been committed to TIME (The Invisible Modelling Environment) (Rahman et al., 2005), a repository maintained by CSIRO, making the codes reusable in other eWater CRC software projects.

5. SOFTWARE EVALUATION AND TESTING

The Trend Tool was verified to ensure that all of the algorithms were properly represented in computer codes, so that it estimates as intended. A number of workshops were conducted during the design and development process, and feedback from various eWater CRC partners was used to upgrade the functionality and user interface. The prototype was tested using actual field data accessed from Victorian and Queensland Department of environment and resource management.

The targeted end-users for this trend tool are people involved in water-quality management and improvement activities. These users include research institutes, environmental groups, State and local governments, regional natural resource management organisations and industry groups. Having focused on end-user requirements and capabilities, the software was developed for easy operation by selecting available options on the screen. In particular, the structure of the required input is easy to follow and the output is easy to comprehend with some visualisation capabilities. A thorough user manual and help system that describes the required input data structures and the handling of outputs in detail is also provided.

6. CONCLUSIONS

The Trend Tool will be a valuable planning and operating tool for catchment groups, project leaders and government agencies, for detecting and assessing water-quality trends over different time scales. The Trend Tool offers a number of techniques for data checking, data processing and visualisation. The tool also comes with powerful statistical capabilities for detecting trends, identifying changes and testing for randomness of time series of water-quality data or other hydrological data sets. This tool will help users with data preparation, exploratory data analysis and identification and visualisation of trends. However, the complex water-quality modelling process required for explanation of the underlying causes of trends is considered beyond the focus of this version. The Trend Tool will be released at a national level under eWater CRC’s ecological management and restoration product development program.

ACKNOWLEDGMENTS

The authors would like to acknowledge Dr Ian Ramsay at Queensland EPA and Dr Nick Marsh at CSIRO Land and Water for their valuable suggestions and guidance during the development and testing of water-quality trend tool.

REFERENCES

ANZECC (2000), Australia and New Zealand Guidelines for Fresh and Marine Water Quality. Paper No 4. Australian and New Zealand Environment and Conservation Council and the Agriculture and Resource Management Council of Australia and New Zealand.

Chiew, F., and Siriwardena, L. (2005), Trend, trend and change detection software, User guide. CRC for Catchments Hydrology, University of Melbourne, Victoria.

Cleveland, W.S. (1979), Robust locally weighted regression and smoothing scatter plots. *Journal of American Statistical Association* Vol 74, (368) 829 – 836.

EWQMC, (1996), Environmental Water Quality Monitoring Committee “Testing The Waters” Review of Victorian Environmental Water Quality Monitoring: Draft Technical Discussion Paper, September 1996.

Helsell, D.R., Mueller, D.K., and Slack, J.R. (2006), Computer program for the Kendall family of trend tests: US Geological Survey Scientific Investigations Report 2005-5275, 4p.

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 Resources Research*, Vol 27, 803-813.

Hirsch, R.M., and Slack, J.R. (1984), A nonparametric trend test for seasonal data with serial dependence: *Water Resources Research* No. 20, p. 727-732.

Hirsch, R.M., Slack, J.R., and Smith, R.A. (1982), Techniques of trend analysis for monthly water quality data: *Water Resources Research* No. 18, p. 107-121.

Kendall, M. (1975), Multivariate Analysis. Charles Griffin and Company, London.

Lettenmaier, D.P. (1978), Design Considerations for Ambient Stream Quality Monitoring. *Water Resources Bulletin*, Vol. 14(no. 4), 804-902.

Rahman, J.M., J.M. Perraud, S.P. Seaton, H. Hotham, N. Murray, B. Leighton, A. Freebairn, G. Davis and R. Bridgart, (2005),. Evolution of TIME. MODSIM 2005, International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2005, 603-703. ISBN: 0-9758400-2-9.

Tennakoon, S.B (2008), Water Quality Analyser, Software User Guide. eWater CRC, Innovation Centre, University of Canberra, Canberra.