WQA: an integrated DSS and statistical package for water quality data management, processing and analysis

S. Tennakoon[a,b] , I. Ramsay[a,b] , S. (Jason) Shen[a,b] and N. Christiansen[a]

[a] *Queensland Department of Environment and Resource Management, Dutton Park, Brisbane*
[b] *eWater CRC, The Australian National University, Canberra*
Email: sunil.tennakoon@derm.qld.gov.au

Abstract: Increasing concern regarding deterioration of freshwater and marine systems such as the Murray Darling Basin and Great Barrier Reef has prompted the need for better water quality and land management across the country. Whilst there has been an increased awareness of the need for better monitoring and assessments of water quality, there are limited tools available to assist water managers to quickly and easily analyse, assess and report on monitoring data. A major survey conducted in Australia revealed that most organisations spend a great deal of time and effort dealing with water quality information, and often require assistance to undertake valid data analysis and assessment. Therefore, Queensland Department of Environment and Resource Management (DERM) in collaboration with the eWater Cooperative Research Centre (CRC) has designed and developed a software package to streamline the data management, processing, analysis and reporting of water quality data and related information. The target end user for this tool includes a broad range of people from State and Local Governments, industry, consulting, natural resources management groups, community groups, land holders and research institutions.

The software package is known as "Water Quality Analyser" (WQA), and has been designed using contemporary science and state of the art decision support software (DSS) technology. The software includes interactive tools, databases, knowledge bases, statistical approaches, decision-making processes and a comprehensive collection of relevant support information. The software was recently upgraded from the version first released in 2009. The latest software package contains the following five modules:

- **Data Management and Visualisation Tool:** This is a new module of the software that performs data import and export for various databases and file formats; visualisation of time series data; processing, cleaning, infilling, re-sampling and data transformation, and exploratory analysis;
- **Guideline Tool:** This feature allows calculation of locally relevant water quality guideline values; test datasets against guideline values to provide a statistically sound indication of the condition of a site; and holds updated national and state specific water quality guidelines in a local database;
- **Loads Tool:** The enhanced load calculations include a variety of methods for estimating pollutant loads in streams using time-series flow and limited concentration data for both long-term and single events. It now includes a function to calculate event mean concentration (EMC);
- **Trend Tool:** The new trend tool detects water-quality trends, and includes statistical tests for change and randomness in water-quality data and other time-series data; and
- **eGuide:** This feature provides access to a number of commonly referred to water-quality related documents (e.g. ANZECC publications) in a searchable electronic format.

The software has been upgraded based on feedback from users including state agencies, councils and land care groups. It has undergone extensive testing and verification covering WQA's applicability, useability, algorithms and encapsulated knowledge. Currently DERM scientists are using an enhanced version of WQA for calculating stream pollutant loads and estimating EMC values for catchment modelling purposes. The enhanced tool can also be used to compare modelling and monitoring data. The software will be commercially released by eWater Innovation (eWI), a subsidiary of eWater CRC, and can be accessed via eWater CRC toolkit (toolkit.net.au) website.

Keywords: *Water quality, Data analysis, Guidelines, Pollutant loads, Trend analysis*

1.   INTRODUCTION

Poor land management and other anthropogenic activities have contributed to significant degradation of soil and aquatic ecosystems in many places in Australia. This is evidenced by increasing desertification of arable lands in the Murray Darling Basin and weakening ecosystems in the Great Barrier Reef. Improved water quality management is crucial in order to avoid further degradation, and ensure protection of these fragile ecosystems. In recent times, considerable resources have been allocated across Australia for water quality observation and management activities. Many government organisations, industries, catchment groups, volunteers, land holders and communities are involved in water quality monitoring and management projects. As a result, a massive amount of water quality related information is being collected and stored in various database systems across the country. A recent survey conducted by the eWater CRC revealed that handling, processing, analysing and reporting of this data has become a major issue for many stakeholders. This is mainly due to some inherent problems in water quality data such as spatial and temporal variability, extreme values, inadequate samples, data gaps, lack of replications, flow dependencies, serial correlations, non-normal distributions, detection limits and the cost involved with monitoring. Further, variations in quality control and quality assurance measures applied across various monitoring regimes is an added complication. Consistent and transparent data analysis procedures are vital for valid assessments, comparability of results and making defensible decisions. Therefore, the Queensland Department of Environment and Natural Resource Management (DERM) in collaboration with the eWater CRC has designed and developed statistical decision support systems to streamline the water quality data analysis and reporting process. This software package is known as Water Quality Analyser (WQA), and provides novel approaches to handle water quality data and information.

Water Quality Analyser is an integrated package of data processing, analysis and assessment tools (Figure 1), and contains a time series data management module, pollutant load estimation tool, water quality trend assessment utility, water quality guideline tool and electronic versions of commonly used water quality related documents. WQA has been designed with unique interfaces for efficient and easy access to data and information. Water quality data can be imported or exported to various databases and file formats. The data management and visualisation module acts as the central hub for data flow, and assists in visualising and processing data. The processed data can be used in other modules to perform specific scientific analysis, for generating reports or use as model inputs. WQA has undergone a number of testing and evaluation processes to meet the functionality and interface requirements outlined by user feedback. The features improved since the previous version include: visual and interactive data processing, exploratory data analysis, trend assessment, direct data exchange with external spread sheets, ability to work on multiple time series and quick start help for beginners. This paper describes the key features of the latest version of WQA, and how the software has been designed, tested and adopted. The software has been released through eWater Innovation (eWI), a subsidiary of eWater CRC, and can be accessed via eWater CRC toolkit (toolkit.net.au) website.
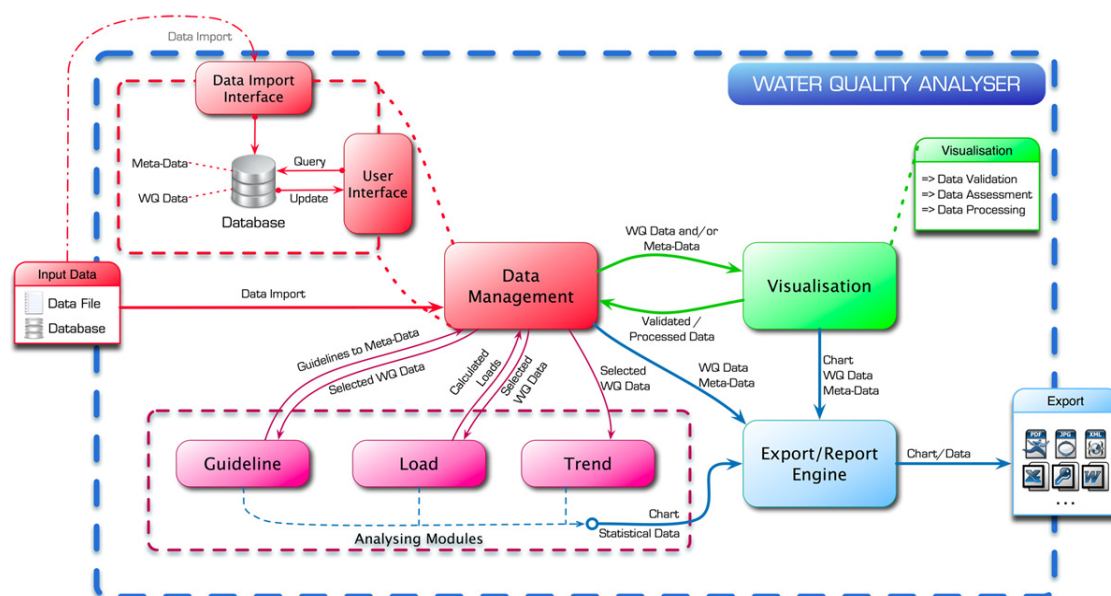

Figure 1. Architecture of the latest version of Water Quality Analyser

2. USER REQUIREMENTS

At the beginning of the product planning stage, a comprehensive survey was conducted across Australia to identify key end-user requirements. Details of the survey, a summary of gathered information and the particulars of proposed products are given in Tennakoon et al. (2007a). The water quality and ecosystems health domain is fairly complex and large, and as a result the product requirements expressed by the partners were beyond the scopes of the original proposed projects. Therefore, the product development team prioritized the user requirements to the following areas:

- Common time series data processing and analysis techniques;
- Consistent, transparent data analysis and comparisons;
- Estimation of pollutant loads and load based assessment;
- Setting of load based targets;
- Water quality guideline based assessment & derivation of locally specific guidelines;
- Spatial and temporal water quality trend assessment;
- Legislative requirements of water quality reporting;
- Defensible decision making; and
- Standard reporting templates and guides for interpretation of outputs.

Based on these priority areas, a number of original products were identified and developed including data management utilities, a loads tool, a guideline tool and a trend assessment tool. However, with the users' feedback, all of these tools were subsequently combined into a standalone software package to facilitate easy installation and maintenance. The combined beta version of WQA (Version 1) was released in 2009 for testing and evaluation. A series of workshops and training sessions were conducted during the beta testing period. The following requirements for WQA enhancement were identified from this process:

- Data import and export mechanisms for corporate databases, and other types of data formats;
- Centralised data management and processing;
- Enhanced data visualisation;
- Estimation and visualisation of uncertainty;
- Exploratory data analysis, data smoothing and enhanced trend analysis;
- Ability to compare monitored and modelled data;
- Ability to handle data from multiple sites with multiple indicators;
- An updated water quality guideline database;
- Additional methods for load calculation (e.g. USGS method);
- Greater report generation capacity; and
- Numerous changes in the user interface and navigation.

WQA was subsequently improved based on this feedback, and the enhanced, where WQA (Version 2) was produced. The load calculation module was also enhanced as part of a Federal Government caring for county sponsored project with DERM.

3. SYSTEM DESIGN

The overall aim of WQA development was to produce software that could support and streamline water quality data processing, analysis, assessment and reporting activities. This package has been designed using contemporary science and latest information technologies. The eWater CRC product development protocols were followed in developing the software, and this also required the package to be extensively tested and verified for accuracy and performance. The required knowledge, information and data for product development and upgrades were captured from a number of sources including engagement of experts from CSIRO Mathematics Informatics and Statistics, the University of Melbourne, the Environmental Protection Authority of Victoria, the National Park Service in Colorado and Queensland DERM.

3.1.   Development Environment: The software was written in the C# language using the Microsoft .NET Framework. Microsoft .NET Framework Version 4.0 Redistributable Package (x86) is required to install and run this software. The .NET Framework 4.0 allows creating applications within TIME, The Invisible Modelling Environment (Rahman et al., 2005). All the source codes used in this software have been committed to the CSIRO TIME Source Code Repository. Visualisation capability in WQA has been improved by using third party software called "Infragistics NetAdvantage" for .NET (www.infragistics.com) which is a tool set for Microsoft platform developers.

3.2.   **Usability and Interface:** WQA has been designed to cater for a wide range of end-users including community groups and volunteers. Therefore, user-friendliness and ease of handling inputs and outputs are important for uptake of the software. WQA tools were designed with built-in intelligence and user-friendly interfaces, and therefore no special training or programming skills are required to setup or run the software unlike many other statistical packages. WQA interfaces are compatible with other eWater products, and have the same look and feel. Users are able to customise the interfaces and the appearance of graphics to suit their requirements and comfort. The modules of the enhanced WQA are discussed in more detail below.

4.   MODULES OF THE ENHANCED WQA

4.1.   Data Management

The data management and visualisation module in WQA has been significantly enhanced. New features include improved importing/exporting, visualising and processing of time series data (Ramsay et al., 2009). This module consists of a local file-based database which is versatile, and allows sharing data with other applications. The data import facility provides options to import data from various formats such as Excel files, SQL, Oracle or FoxPro (Hydstra) data sources. The upgraded system can handle time series data from multiple sites and indicators at different scales and units simultaneously. The enhanced visual charting component provides facilities for interactive data processing and exploratory data analysis. This has the capability to plot multiple "Y" axes of time series data, and can be used for comparing monitored and modelled data. Data processing capabilities include filling gaps, aggregation, re-sampling, aligning and data transformations (Figure 2). Statistical procedures are also available for detecting outliers and dealing with detection limits. Detection limits are identified by the presence of less-than or greater-than symbols in the imported data (e.g. "< 0.05" is a lower detection limit), and a number of replacement methods are available. During data processing, modified data sets can be stored in a series of historical files, and at any stage the user can go back and retrieve historical data, and verify all the modifications made. This makes the data processing work transparent, and keeps the original data set intact. The tool allows direct data copy and paste facility into external applications such as Excel using the clipboard. The working databases can be exported or imported which allows sharing of data sets or the results with other users.

4.2.   Water Quality Guidelines

Water quality data can be used to evaluate the physical, chemical, and biological characteristics of a water body (e.g. stream, river, lake, estuary, or ocean) and potential impacts (human health, ecosystem conditions, environmental values). Assessments are generally made against water quality guidelines, water quality objectives or water quality targets (ANZECC/ARMCANZ, 2000). The guideline module in WQA has been designed for calculating and applying water quality guidelines using the procedures specified in ANZECC/ARMCANZ (2000) and the Queensland Water Quality Guidelines (QWQG, 2009). The latest version of WQA comes with an updated database containing default guideline/trigger values from ANZECC/ARMCANZ (2000), the new Queensland guidelines, and some

Figure 2. Data processing and visualisation in the enhanced version of WQA

state specific guidelines and water quality objectives values. Users can search the database for guidelines or water quality objectives, and these can be used for testing against time-series data. The tool can calculate locally relevant water quality guidelines using reference data, as recommended by ANZECC/ARMCANZ (Tennakoon et al., 2008). WQA also contains an option to estimate biological effects guidelines or toxicants trigger values. These values are calculated using aquatic toxicity data obtained from laboratory studies (Shao, 2000). Calculation of protecting concentrations (trigger values) is computed by fitting a distribution called Burr type III as recommended by ANZECC/ARMCANZ. The reference data approach uses the non-
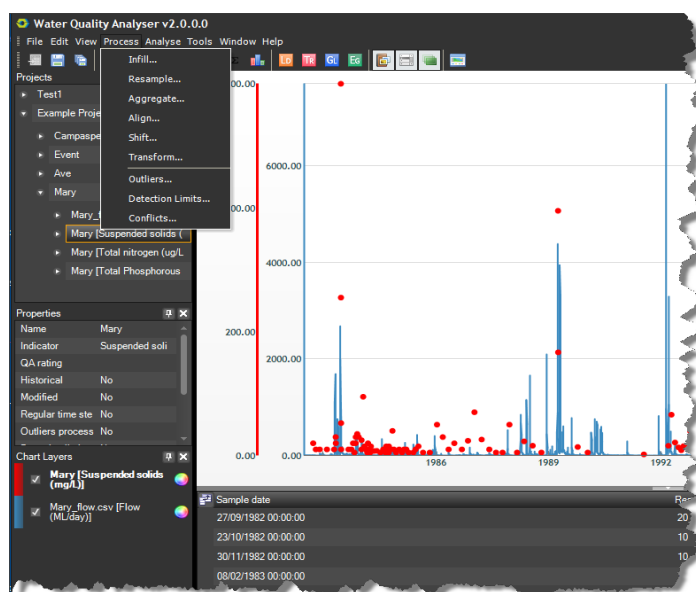
parametric statistical procedures for deriving guideline values and its confidence intervals (Payton et al., 2003). The tool can also be used to set water quality targets using a proportional improvement approach. The current version of the water quality guideline tool contains following components:

- Water quality guideline database and search facilities;
- Water quality guideline assessment;
- Guideline derivation using biological effects data;
- Guideline derivation using reference site data;
- Managing and updating the database;
- Water quality target setting and comparison.

4.3. Pollutant Load Estimation

Determining constituent loads in a river or stream is generally complicated due to the relatively sparse nature of concentration data and different sampling methods applied. Special load estimation techniques should be applied to reduce estimation errors (Fox, 2004). The enhanced loads module in WQA presents a number of commonly accepted methods for long-term and storm event load calculation (Figure 3) including the USGS seven parameter method (Cohn et al., 1992). A built-in function is available to calculate event mean concentration (EMC) values which are useful in catchment modelling exercises. WQA uses time series of both flow and concentration data to calculate loads. Users can either use an entire time series or selected time intervals for estimating loads. Selecting an appropriate method for a particular data set is not an easy task as there is no objective way of choosing a method. Often users adopt past experience or expert advice for selecting a calculation method. WQA presents a decision support framework to guide the user on the most appropriate method of load estimation for a given situation. Estimation of uncertainty of calculations is also a major concern, and this version of WQA presents uncertainty estimates for a couple of methods (average flow * average concentration and flow weighted concentration) used (Fox, 2005). Historical results indicate that there are major differences in the load estimates produced when different load estimation methods are used. Therefore the consistency in data collection and the method of calculation is vital for valid comparisons. More details of the load module can be found elsewhere (Tennakoon et al., 2007b). Load based targets and limits have been established in many catchments or river systems as a part of managing the conditions of receiving waters and halt decline. Detection of temporal water quality trends is important to help assess the likelihood of achieving aspirational targets and monitoring the effectiveness of management practises.
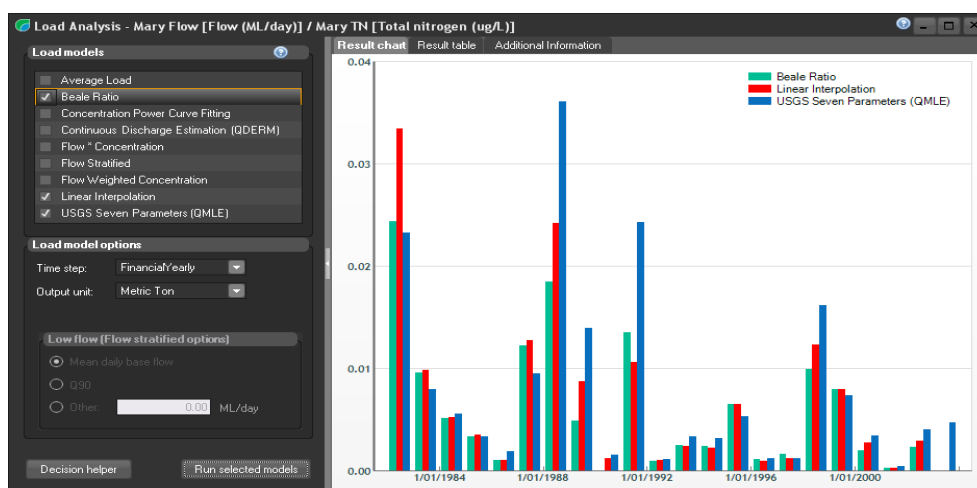

Figure 3. Estimating pollutant loads in the enhance version of WQA

4.4. Trend Assessment

The original trend tool developed by the CRC for Catchment Hydrology (Chiew and Siriwardena, 2005) was upgraded using feedback received from CRC partners, and then packaged into the original version of WQA. The user feedback revealed that 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 was required to validate and process water-quality data for temporal

trend detection to avoid potential confounding factors (Tennakoon et al., 2009). Therefore, the tool was upgraded to facilitate the process of data cleaning, processing, visual assessment and exploratory data analysis. The user can then use validated data for statistical testing or in other complex modelling processes. The new tool contains procedures for statistical testing of trends, delineating step changes and testing randomness in water quality data or other time series data.

4.5.   Other Utilities

eGuides: eGuides is an electronic document which consists of a number of commonly referred to water quality documents such as ANZECC/ARMCANZ (2000) and the Queensland Water Quality Guidelines (2009). These documents have been compiled into a standard "HTML" version of Windows Help Systems, and provide easy and quick access to information. Users can select the document that they would like to manually browse, or select the 'search' tab to search all the guides using key words. The searched items can be viewed, copied to another document or printed out for later reference.

Dashboard: The dashboard utility has been designed to store charts, and assist in comparing results and preparing reports. The Dashboard provides an interactive display area that allows multiple charts to be presented on a single panel; so users can easily and quickly compare and contrast different charts.

5.   TESTING & REVIEWS OF PERFORMANCE

Testing and evaluation was carried out at different stages of product development to assess accuracy and performance of the software suite. This was done internally by validation and verification procedures. All the algorithms were tested by the software development team using manual calculations through spreadsheets. A number of workshops were conducted with potential users for refining the tool. End-user feedback was used to upgrade functionality and the user interface. The prototype testing and evaluation were conducted by DERM field officers using real world data collected from regional areas in Queensland. The enhanced load module was tested and evaluated with the DERM load monitoring team, and feedback and comments were collected using an online bug tracking system (bugzilla). A well planned usability test was performed to identify any problems with the interaction between users and the tool. The observed usability issues were rectified. An independent external numerical testing process was also carried out to confirm the accuracy of estimates and statistics.

5.1.   Documentation: The documents produced during product development and testing included: project planning and survey documents, programmers manual, test reports, online bug reporting system and user manuals. A number of peer reviewed conference papers were produced to help validate the science, knowledge and the procedures used in the WQA software package. As many organizations and individuals were involved in the development process, the intellectual properties used in this tool were identified and registered in the eWater CRC intellectual property registry. An agreement was signed between DERM and eWater CRC for commercialising the software.

5.2.   Current Users: The current users of this tool include a broad range of DERM staff, eWater CRC catchment modelling teams, Great Barrier Reef load monitoring team, South East Queensland event monitoring team, local councils in South East Queensland, Griffith University, and South Australian Research and Development Institute. As a part of Reef Water Quality Improvement Plan, the DERM Spatial and Scientific Information Management for Reef (SSIMR) Project has designed a tool called DARTS (Data Recording Tool for Science) to collect paddock to reef data. WQA will be linked to DARTS for data processing, visualisation and analysis to assist with this process. The other main group of potential users are catchments authorities, regional bodies, community groups and government organisations.

6.   DISCUSSION AND CONCLUSION

WQA has been redesigned to improve the efficiency of data handling and data sharing facility. The newly designed data module works as the mainstay of the tool, and provides processed data to other utilities. The statistical approaches in the trend and loads modules have significantly been enhanced. Extensive testing and trialling of the software has been undertaken. Very positive feedback was received from beta testing and the limited release of Version 1 within eWater CRC partners. WQA will assist a broad range of stakeholders in managing, processing, analysing and reporting water quality related data or any other time series data given the absence of a similar software product on the market. The time spent on performing these tasks can be significantly reduced. WQA provides consistent and transparent approach for data processing and analysis which allows valid assessments and comparability. This will assist regulators and other stakeholders with making defensible decisions and statistically sound comparisons. WQA can also be used to compare

monitored and modelled data, derive model parameters and prepare data as models inputs, although not discussed in any detail in this paper. The enhanced version of WQA (Version 2) has been recently released, and the commercial viability and the demand from national and international users have not yet been evaluated.

ACKNOWLEDGMENTS

The authors would like to acknowledge the eWater CRC for their ongoing support and collaboration with WQA development. We also acknowledge the contribution from Dr Nick Marsh and Brendan Farthing at CSIRO Land and Water, Prof David Fox at the University of Melbourne, and David Robinson at Victorian EPA. Tory Grice and Dr Kirsten Shelly at Griffith University assisted in the testing process. The prototype of WQA software modules were developed under National Action Plan for Salinity and Water Quality Program. The DERM caring for country project provided some funds for loads module enhancement.

REFERENCES

ANZECC/ARMCANZ (2000). Australia and New Zealand guidelines for fresh and marine water quality. 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.

Cohn, T.A., Caulder, D.L., Gilroy, E.J., Zynjuk, L.D., and Summers, R.M. (1992). The validity of a simple statistical model for estimating fluvial constituent loads: An empirical study involving nutrient loads entering Chesapeake Bay. *Water Resources Research,* 28(9), 2353-2363.

Fox, D.R. (2004). Statistical considerations for the modelling and analysis of flows and loads - components of load. Technical Report 02/04, Australian Centre for Environmetrics, University of Melbourne, Victoria.

Fox, D.R. (2005). Protocols for the optimum measurements and estimation of nutrient loads - error approximations, Technical Report 03/05, Australian Centre for Environmetrics, University of Melbourne. Victoria.

Payton, M. E., Greenstone, M. H., and Schenker, N. (2003). Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance? *Journal of Insect Science*, 3:34.

QWQG (2009). Queensland water quality guidelines, Version 3. Department of Environment and Resource Management, ISBN 978-0-9806986-0-2.

Rahman, J.M., Perraud, J.M., Seaton, S.P., Hotham, H., Murray, N., Leighton, B., Freebairn, A., Davis, G., and Bridgart, R. (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.

Ramsay, I., Shen, S., and Tennakoon, S. (2009). Water quality visualisation and tracking – generic decision support tool. 18th IMACS World congress MODSIM09, 13 -17 July 2009, Cairns, Australia.

Shao, Q. (2000). Estimation for hazardous concentrations based on NOEC toxicity data: An Alternative approach. *Environmetrics,* 11, 583-595.

Tennakoon, S., Ramsay, I., Marsh, N., and O'Connor, R. (2007a). The Integrated Monitoring and Assessment System (IMAS): A decision support system for water quality monitoring and assessment programs. In Oxley, L. and Kulasiri, D. (eds) MODSIM 2007, International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2007, pp 358-364.

Tennakoon, S., Marsh, N., and Arene, S. (2007b). A software tool for the estimation of pollutant loads in rivers and streams using time series data. In Oxley, L. and Kulasiri, D. (eds) MODSIM 2007, International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2007, pp 2403-2409.

Tennakoon, S., Farthing, B., and Marsh, N. (2008). A statistical tool for setting water quality guidelines and testing water quality targets. 11th International River Symposium, Brisbane Australian, 1-4 September 2008.

Tennakoon, S., Robinson, D., and Shen, S. (2009). Decision support system for temporal trend assessment of water quality data. 18th IMACS World congress MODSIM09, 13 -17 July 2009, Cairns Australia.