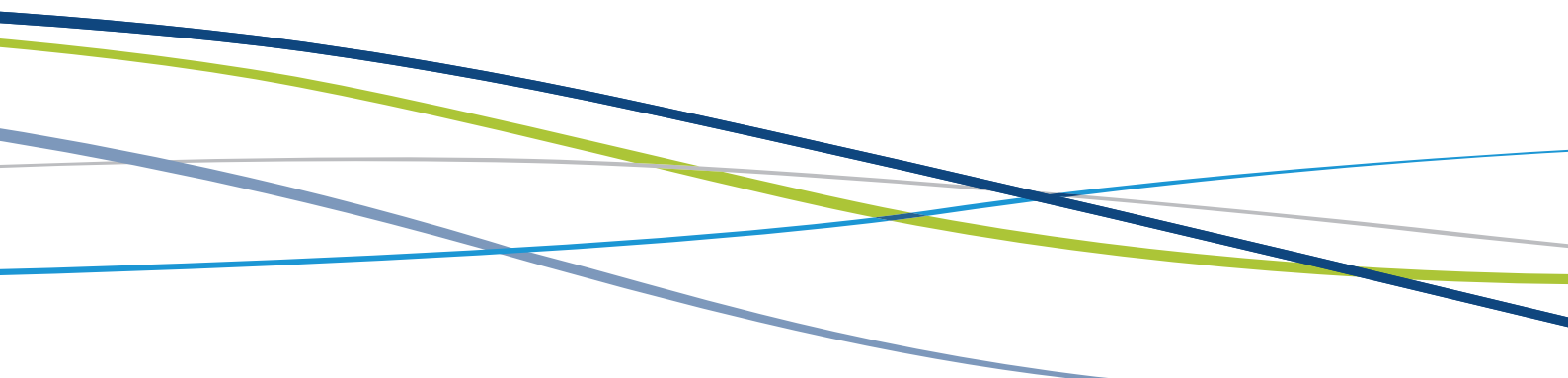


# Water Quality Analyser



## User Guide

Based on Water Quality Analyser 2.0.0  
September 2011

## Document History

Date	Author	Revision	Description
September 2011	<ul style="list-style-type: none"><li>• Sunil Tennakoon</li><li>• Jason Shen</li><li>• Yanning How</li><li>• Michael Emery</li></ul>	2.0.1	Updates to reflect latest WQA release
October 2010	<ul style="list-style-type: none"><li>• Sunil Tennakoon</li><li>• Jason Shen</li><li>• Michael Emery</li></ul>	2.0.0	Updates to reflect latest WQA release
August 2008	<ul style="list-style-type: none"><li>• Sunil Tennakoon</li></ul>	1.0.0	Creation of User Guide

## Copyright Notice

© eWater CRC, 2011

## Legal Information

All rights reserved. No parts of this work may be reproduced in any form by any means – graphic, electronic, or mechanical, including photocopying, recording, taping or information storage and retrieval systems – without the written permission of the publisher.

Products that are referred to in this document may be either trademarks and/or registered trademarks of the respective owners. The publisher and the author make no claim to these trademarks.

While every precaution has been taken in the preparation of this document, the publisher and the author assume no responsibility for errors or omissions, or for damages resulting from the use of programs and source code that may accompany it. In no event shall the publisher and the author be liable for any loss of profit or any other commercial damage caused or alleged to have been caused directly or indirectly by this document.

## Citing this Document

You can cite this document as:

eWater Cooperative Research Centre 2011, Water Quality Analyser v2.0.0 User Guide, eWater Cooperative Research Centre, Canberra

ISBN: 978-1-921543-47-0

## Acknowledgements

eWater CRC acknowledges and thanks all partners to the CRC and individuals who have contributed to the research and development of this software product

The prototype of WQA software modules were developed under the National Action Plan for Salinity and Water Quality Program in Queensland. The DERM Caring for Country Community Coastcare Project funded by the Australian Government also provided some financial support.

**The Water Quality Analyser Team gratefully acknowledge contributions from the following organisations (in alphabetical order)**

ACT DECCEW

Australian and New Zealand Environment and Conservation Council and the Agriculture and Resource Management Council of Australia and New Zealand

CSIRO

Department Environment Resources and Management (DERM – QLD)

Office Environment and Heritage (NSW)

NSW Office Water

Griffith University

La Trobe University

Murray Darling Basin Authority

Melbourne University

Melbourne Water

Monash University

National Health and Medical Research Council, Australia

QLD DPIF

South Australia Water

South Australia EPA

SARDI

Sydney Catchment Authority

SKM

Southern Rural Water –SRW

University of Canberra

Vic Department Sustainability and Environment

VIC EPA

WBMBMT

**The Water Quality Analyser Team gratefully acknowledge contributions from the following people (in alphabetical order)**

Sylvain Arene

Dr Francis Chiew

Michael Emery

Brendan Farthing

Prof David Fox

Tory Grice

Yanning How

Sally Hunter

Dr Nick Marsh

Ann Milligan

Ruth O'Connor

Dr Ralph Ogden

Dr Ian Ramsay

David Robinson

Dr Kirsten Shelly

Dr Quanxi Shao

Jason Shen

Dr Andy Steven

Dr Sunil Tennakoon

# Contents

<b>1</b>	<b>Software User Documentation</b>	<b>1</b>
<b>2</b>	<b>Introduction</b>	<b>2</b>
2.1	Overview	2
2.1.1	Data Management and Visualisation Module	3
2.1.2	Loads Tool	4
2.1.3	Trends Tool	5
2.1.4	Guidelines Tool	6
2.1.5	Dashboard	6
2.1.6	eGuides	7
2.2	Features	9
2.3	Audience	9
2.4	References and training	10
2.5	Related documents	11
2.6	Product components	11
2.7	Version history	12
2.8	Limitations and caution notes for users	12
<b>3</b>	<b>Getting Started</b>	<b>13</b>
3.1	Installation and Uninstallation	13
3.2	Navigation	13
3.3	Toolbars	14
3.3.1	Menu bar	14
3.3.2	Standard toolbar	14
3.4	Panels	15
3.4.1	Projects panel	15
3.4.2	Properties panel	15
3.4.3	Chart Layers Panel	16
3.4.4	Panel Layout	16
3.4.4.1	Unpinning a panel	16
3.4.4.2	Floating a panel	17
3.4.4.3	Docking a floating panel to a dock area	18
3.4.4.4	Hiding or showing a panel	19
3.4.4.5	Resetting the window layout	19
3.5	Data Operations	20
<b>4</b>	<b>Procedures</b>	<b>21</b>
4.1	Project Management	21
4.1.1	Creating a new project	21

4.1.2	Creating a new set	22
4.1.3	Navigating projects, sets and time series	23
4.1.4	Editing Project Properties	24
4.1.5	Editing set properties	25
4.1.6	Editing Time series properties	26
4.1.7	Deleting a project	27
4.1.8	Deleting a set	27
4.1.9	Deleting a time series	27
4.1.10	Moving a Time series into another set	28
4.1.11	Creating or editing indicators and units	28
4.2	Time Series Management	28
4.2.1	Importing a time series	28
4.2.2	Importing a Time series with data from several locations	33
4.2.3	Opening and closing Time series	34
4.2.4	Creating an empty Time series	34
4.2.5	Editing Time series properties	34
4.2.6	Deleting a Time series	34
4.2.7	Viewing a Time series log	35
4.2.8	Saving a time series	35
4.2.9	Saving a Time series under a different name	35
4.2.10	Viewing historical time series	35
4.2.11	Editing and saving a historical time series	36
4.2.12	Merging two or more time series	36
4.2.13	Copying and pasting sample results to and from external applications	38
4.2.14	Exporting a time series	38
4.2.15	Exporting a chart image	38
4.3	Time Series Visualisation	39
4.3.1	Layers	39
4.3.1.1	Opening a time series	39
4.3.1.2	Overlaying two or more time series	39
4.3.1.3	Activating a Time series layer	39
4.3.1.4	Temporarily hiding a time series	40
4.3.1.5	Removing a time series	40
4.3.1.6	Customising the appearance of a time series	40
4.3.2	Chart	41
4.3.2.1	Zooming with the mouse wheel	41
4.3.2.2	Zooming with click and drag	41
4.3.2.3	Zooming to an existing selection	42
4.3.2.4	Resetting the zoom level	42
4.3.2.5	Panning through a zoomed time series	42
4.3.2.6	Setting the default range for an axis	42
4.3.3	Interaction	43
4.3.3.1	Displaying a Time series in XY plot mode	43
4.3.3.2	Displaying scale breaks	43
4.3.3.3	Displaying the chart legend	43
4.3.3.4	Selecting sample results on the chart	44
4.3.3.5	Selecting sample results using the grid	44
4.3.3.6	Inverting a selection	45
4.3.3.7	Cancelling a selection	45

4.3.3.8	Adding to an existing selection . . . . .	45
4.3.3.9	Subtracting from an existing selection . . . . .	45
4.3.3.10	Adding or subtracting a single sample result from an exist- ing selection . . . . .	45
4.3.3.11	Selecting baseflow . . . . .	45
4.3.3.12	Viewing properties of a selection . . . . .	45
4.3.3.13	Viewing properties of a sample result . . . . .	45
4.3.3.14	Switching to fullscreen mode . . . . .	46
4.4	Time Series Processing . . . . .	46
4.4.1	Editing the properties of a sample result . . . . .	46
4.4.2	Deleting a sample result . . . . .	47
4.4.3	Deleting a group of sample results . . . . .	47
4.4.4	Trimming a time series . . . . .	47
4.4.5	Infilling gaps in a time series . . . . .	47
4.4.6	Resampling a time series . . . . .	48
4.4.7	Aggregating a time series . . . . .	49
4.4.8	Aligning a Time series to another Time series or a fixed time step	50
4.4.9	Shifting the dates of a time series . . . . .	51
4.4.10	Transforming a time series . . . . .	52
4.4.11	Processing outliers . . . . .	53
4.4.12	Processing detection limits . . . . .	53
4.4.13	Processing conflicting sample results . . . . .	54
4.5	Statistical Analysis . . . . .	55
4.5.1	Viewing basic statistics . . . . .	55
4.5.2	Viewing a histogram or frequency plot . . . . .	56
4.6	Load Estimation . . . . .	57
4.6.1	Introduction . . . . .	57
4.6.2	Overview . . . . .	57
4.6.3	Selecting a flow and concentration Time series for load analysis	60
4.6.4	Selecting an analysis period . . . . .	61
4.6.5	Manually selecting load methods . . . . .	62
4.6.6	Using the decision helper to select load methods . . . . .	63
4.6.7	Viewing load estimates . . . . .	66
4.6.8	Viewing calculated numbers . . . . .	66
4.6.9	Viewing additional information . . . . .	66
4.6.10	Saving load estimates . . . . .	67
4.6.11	Input data preparation for events . . . . .	67
4.6.12	Method selection for events . . . . .	68
4.6.13	Creating a new unit for load estimate output . . . . .	69
4.7	Trend Tool . . . . .	69
4.7.1	Introduction . . . . .	69
4.7.2	Overview . . . . .	70
4.7.3	Features . . . . .	71
4.7.4	Audience . . . . .	71
4.7.5	Limitations and cautionary notes for users . . . . .	72
4.7.6	Data requirements . . . . .	72
4.7.7	Exploratory analysis . . . . .	72
4.7.8	Selecting an analysis period . . . . .	74
4.7.9	Trend Tests selection . . . . .	75
4.7.10	Running the tests . . . . .	76

4.7.11	Results screen	76
4.7.12	Description of statistical tests in TREND	79
4.7.13	References	87
4.7.14	Acknowledgement	87
4.8	Dashboard	88
4.8.1	Introduction	88
4.8.2	Features	88
4.8.3	Using dashboard	89
4.8.4	Adding a chart to Dashboard	89
4.8.5	Removing the Chart from Dashboard	90
4.8.6	Renaming a chart	90
4.8.7	Copying the chart and pasting onto a document	90
4.8.8	Copying a chart image and pasting onto Dashboard	91
4.8.9	Saving the chart image	91
4.8.10	Saving the Dashboard	91
4.8.11	Open previously saved binary file/Dashboard	91
4.8.12	Opening a new Dashboard	92
4.8.13	Clearing the Dashboard	92
4.8.14	Re-arranging the tiles on the Dashboard	92
4.8.15	Configuring the view of the Dashboard	92
4.8.16	Closing the Dashboard	93
4.9	Guideline Tool	93
4.9.1	Introduction	93
4.9.2	Overview	93
4.9.3	Water quality guideline database	94
4.9.4	Applying water quality guidelines	94
4.9.5	Features	96
4.9.6	Audience	96
4.9.7	Related documents	96
4.9.8	Data requirements	97
4.9.9	Product components	97
4.9.10	Limitations and caution notes for users	97
4.9.11	Using Water Quality Guideline Tool	98
4.9.12	Running the software	98
4.9.13	Search guideline library (database)	98
4.9.14	Testing (apply) water quality guidelines or targets	100
4.9.15	Using a user defined value for testing	101
4.9.16	Deriving water quality guidelines	104
4.9.17	Calculating guidelines using biological effects data	104
4.9.18	Calculating guidelines using reference data	106
4.9.19	Add user defined values to the database	109
4.9.20	Set water quality targets using proportional improvement	111
4.9.21	References	111
4.10	eGuides	112
4.10.1	procedures-eguides-introductionIntroduction	112
4.10.2	Overview	112
4.10.3	Features	113
4.10.4	Audience	113
4.10.5	Limitations and caution notes for users	113
4.10.6	The Contents Tab	118

4.10.7	The Breadcrumb Trail . . . . .	119
4.10.8	The Navigation Toolbar . . . . .	120
4.10.9	The HTML Help Toolbar . . . . .	120
4.10.10	Print a Help topic . . . . .	121
4.10.11	Sort searched items . . . . .	123
4.10.12	Find definitions for technical terms . . . . .	124
4.10.13	Copy and paste selected information into a document . . . . .	124
4.10.14	Exit eGuides . . . . .	125
<b>5</b>	<b>Troubleshooting</b> . . . . .	<b>127</b>
5.1	Loads tool shows zero total load . . . . .	127
5.2	Application does not start . . . . .	127
<b>6</b>	<b>Appendices</b> . . . . .	<b>128</b>
6.1	Appendix 1 - Biological approach of guideline setting . . . . .	128
6.2	Appendix 2 - Procedures for deriving guidelines using reference data (local or regional) . . . . .	130
6.3	Appendix 3 - Pollutant loads estimation in rivers or streams . . . . .	131
6.4	Appendix 4 - Decision tree with preferred load estimation methods . . . . .	149
6.5	Appendix 5 - Events load estimation . . . . .	150
6.6	Appendix 6 - Protocols for the optimal measurement and estimation of nutrient loads . . . . .	156
6.7	Appendix 7 - Assessment of Water Quality Data For Temporal Trends . . . . .	164
6.7.1	Introduction . . . . .	164
6.7.2	Water quality variability . . . . .	165
6.7.2.1	Spatial Variation . . . . .	165
6.7.2.2	Temporal Variation . . . . .	165
6.7.3	MAIN STEPS IN ASSESSMENT OF WATER QUALITY DATA FOR TRENDS . . . . .	166
6.7.3.1	Data Validation . . . . .	167
6.7.3.2	Visual Assessment and Exploratory Data Analysis . . . . .	170
6.7.3.3	Statistical Tests For Testing The Significance Of A Trend . . . . .	172
6.7.3.4	EXPLAINING THE TRENDS . . . . .	173
6.7.4	References . . . . .	174

---



## Chapter 1

# Software User Documentation

[Introduction](#)

[Getting Started](#)

[Procedures](#)

[Troubleshooting](#)

[Appendices](#)

## Chapter 2

# Introduction

[Overview](#)

[Features](#)

[Audience](#)

[References and training](#)

[Related documents](#)

[Product components](#)

[Version history](#)

[Limitations and caution notes for users](#)

### 2.1 Overview

Water Quality Analyser (WQA) is an integrated collection of data analysis and assessment tools. The Data Management and Visualisation Module functions as the central hub of data flow and visualisation, while other modules perform specific scientific analysis. The overall architectural diagram is shown below, followed by short description of each module.

- Data management and visualisation
- Loads Tool
- Trend Tool
- Guidelines tool
- Dashboard
- eGuides

The overall architectural diagram is shown below following by short description of each module.

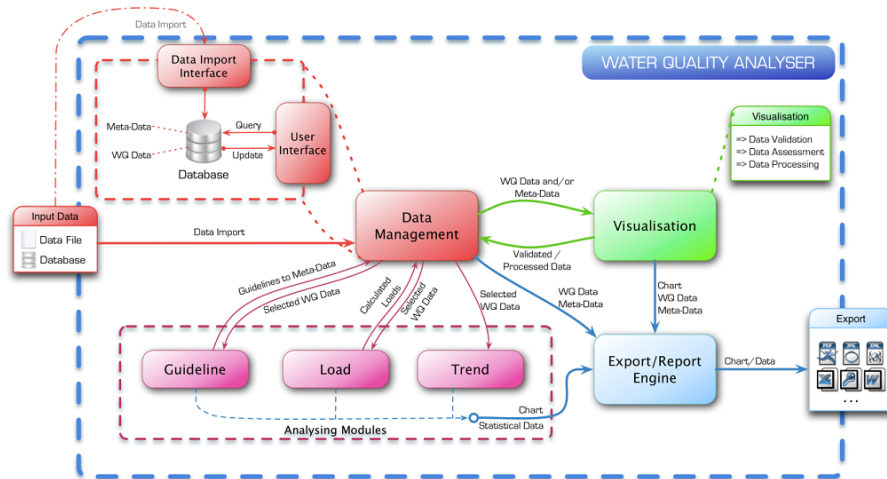


Figure 2.1: WQA Architectural Diagram

### 2.1.1 Data Management and Visualisation Module

The Data Management and Visualisation Module is a tool designed to import, store, export and visualise water quality data. It functions as a data store and simplifies the process of preparing data for scientific analysis. It consists of a local file-based database, a visual chart component and several processing tools. Each processing tool provides feedback about the state of a time series and a set of recommended actions to perform on the time series.



Figure 2.2: WQA Data Management and Visualisation Module

### 2.1.2 Loads Tool

The determination of constituent loads or loading rates from rivers and streams is not a trivial task. Reducing estimation errors requires the application of special load estimation techniques to compensate for the relative lack of concentration data and large diversity of sampling methods. The 30-plus alternative load estimation methods make assumptions about the behaviour of pollutant concentrations in-stream during times when water quality isn't sampled. This tool presents nine of the most common methods for long-term load calculation and four methods for estimating loads from storm events. It can also calculate event mean concentration (EMC) values for use in catchment modelling exercises.

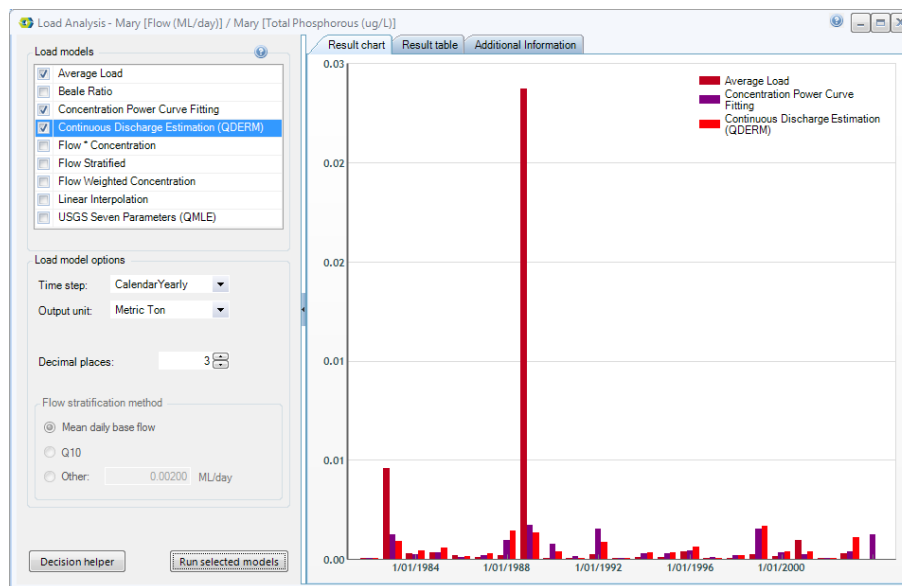


Figure 2.3: WQA Loads Tool

### 2.1.3 Trends Tool

The Trend Analysis tool allows statistical testing of trend, change and uncertainty in water quality and other time series data. It is a major enhancement of the original trend tool developed by the CRC for Catchment Hydrology, based on feedback from CRC partners. The current version presents the following 13 statistical tests:

- Spearman's Rho (non-parametric test for trend)
- Linear Regression (parametric test for trend)
- Distribution-Free CUSUM (non-parametric test for step jump in mean)
- Cumulative Deviation (parametric test for step jump in mean)
- Worsley Likelihood Ratio (parametric test for step jump in mean)
- Rank-Sum (non-parametric test for difference in median from two data periods)
- Student's t (parametric test for difference in mean from two data periods)
- Median Crossing (non-parametric test for randomness)
- Turning Points (non-parametric test for randomness)
- Rank Difference (non-parametric test for randomness)
- Autocorrelation (parametric test for randomness)
- Seasonal Mann-Kendall

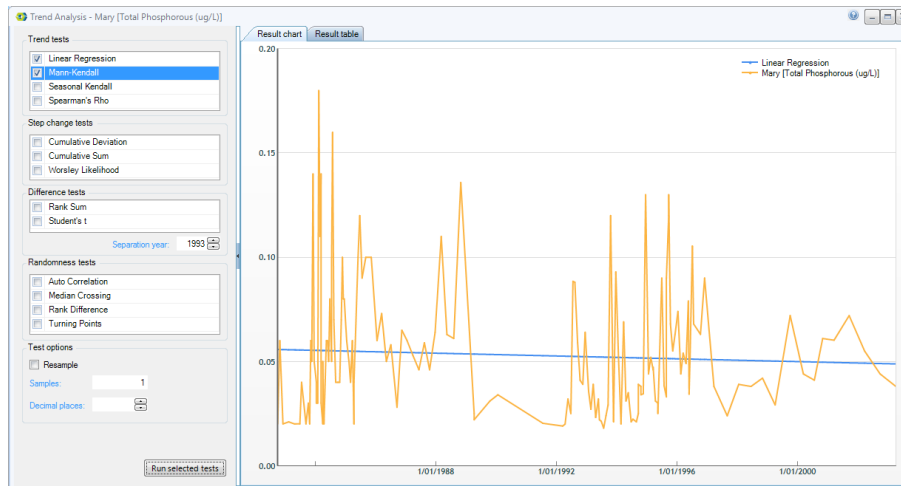


Figure 2.4: WQA Trends Tool

### 2.1.4 Guidelines Tool

The guidelines tool is a statistical tool developed to: calculate locally relevant guideline values; store guideline values in a searchable database for later recall; and test new datasets against guideline values to provide a statistically sound indication of the health of a site. The procedures used in this tool are based on the methods in ANZECC & ARMCANZ 2000. Both referential and biological approaches are available for setting up guidelines, depending on type of available data. The tool can also assist in the development of water quality targets by providing a method to calculate proportional improvements in a measured variable. The Guideline Tool has been built with a user-friendly interface for easy operation and the system comes with comprehensive help materials for operations as well as technical information for use in deriving local water quality guideline values.

Guidelines				
Guideline	Type	Value	Indicator	Result

Search for Guideline Remove From List Create New Guideline Create New Target... Test Guideline Save Results

Figure 2.5: WQA Guidelines Tool

### 2.1.5 Dashboard

The Dashboard provides an interactive display area that allows multiple charts to be presented on a single panel for analysis of statistical data. Users can also easily compare and contrast different charts at a glance.

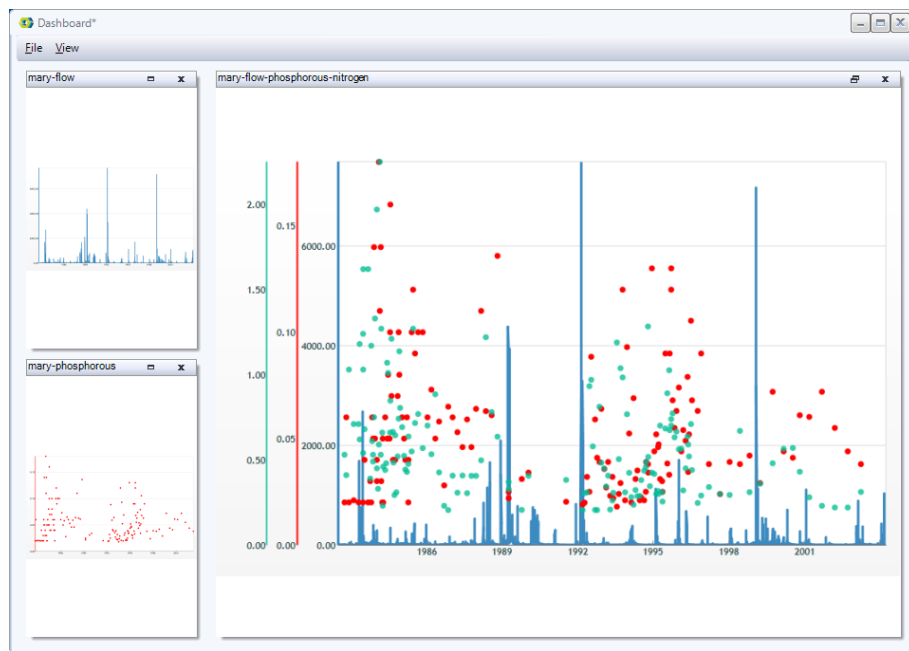


Figure 2.6: WQA Dashboard

### 2.1.6 eGuides

eGuides is an electronic document which consists of a number of commonly referred to water quality guideline documents. The current version of eGuides contains the following documents.

- ANZECC/ARMCANZ 2000 Monitoring & Reporting Guidelines
- ANZECC/ARMCANZ 2000 Water Quality
- NHMRC 2005 Recreational Water Quality
- Queensland Water Quality Guidelines

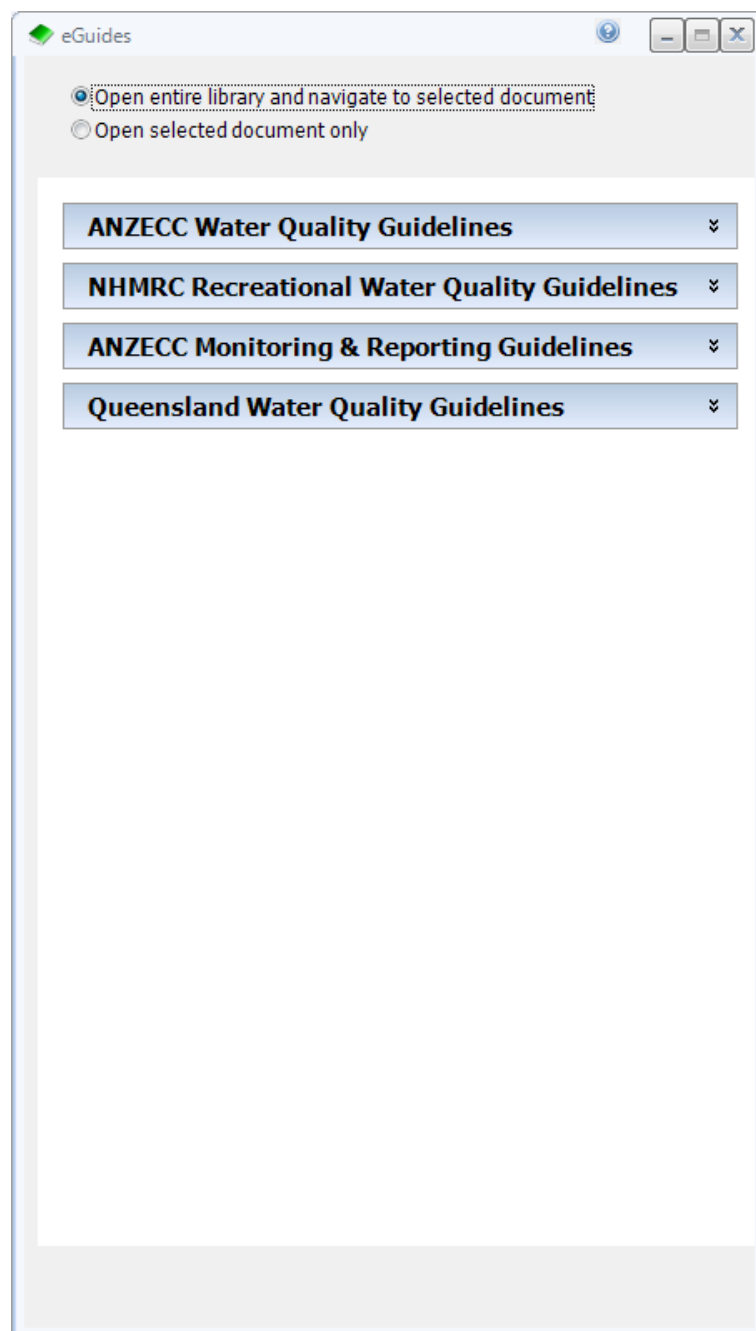


Figure 2.7: WQA eGuides



## 2.2 Features

The Water Quality Analyser has a wide range of end-users including community groups and volunteers. Therefore, user-friendliness and easy handling of both inputs and outputs are important for the success of these utilities. The Water Quality Analyser is designed with a user-friendly interface and no special computer skills are required to run the tools. Enhancements in this version include:

- Visual and interactive data processing component
- Built in quick start help system for beginners
- Ability to export output as a spreadsheet or image
- Ability to export output as a spreadsheet or image
- Collection of processing tools with visual feedback
- Ability to work with multiple time series using Chart layers
- Presents multiple charts on a single panel

The electronic documents (eGuides) in standard Microsoft Windows Help format provide a valuable knowledge base for water quality data analysis and assessment.

## 2.3 Audience

A wide range of stakeholders were recognised as end-users, and they are summarised in the following table.

Use	Primary users	Secondary users
Pollution / regulation	<ul style="list-style-type: none"> <li>• State regulator</li> <li>• Local government</li> <li>• NRM bodies</li> <li>• Industry clients</li> </ul>	<ul style="list-style-type: none"> <li>• Industry</li> <li>• Technical staff</li> <li>• Consultants</li> <li>• Commercials</li> </ul>
Monitoring design, assessment and reporting	<ul style="list-style-type: none"> <li>• State agencies</li> <li>• NRM bodies</li> <li>• NRM groups</li> </ul>	<ul style="list-style-type: none"> <li>• Councils</li> <li>• Catchment groups</li> <li>• Community groups</li> <li>• Water watch</li> <li>• Land holders</li> </ul>
SOE Reporting / NPI (National Pollutant Inventory)	<ul style="list-style-type: none"> <li>• Federal Government</li> </ul>	<ul style="list-style-type: none"> <li>• State government</li> <li>• Local government</li> </ul>
Research and Education	<ul style="list-style-type: none"> <li>• Research organisations</li> <li>• Educational institutes</li> </ul>	<ul style="list-style-type: none"> <li>• Scientists</li> <li>• Technical staff</li> <li>• Students</li> </ul>

## 2.4 References and training

See <http://www.ewatercrc.com.au/training> for upcoming training courses or to register an interest in a course. For Technical information and software bugs please contact following person/s:

Dr Sunil Tennakoon

Principal Scientist

Water Quality & Aquatic Ecosystem Health

Environment and Resource Sciences

Queensland Department of Environment and Resource Management

Block B East Level 1, 41 Boggo Road

---

Dutton Park, QLD 4102, Australia

Phone: +61 7 3170 5558

Fax: +61 7 3170 5799

Email: [sunil.tennakoon@derm.qld.gov.au](mailto:sunil.tennakoon@derm.qld.gov.au)

## 2.5 Related documents

The list of user manuals produced for Water Quality Analyser utilities is as follows:

- User guide for time series data manager
- User guide for guideline tool
- User guide for loads tool
- User guide for trend analysis tool
- User guide for dashboard
- Built-in help for eGuides use.

A comprehensive context-sensitive help system has been built into the Water Quality Analyser to provide technical and user guides to end-users.

## 2.6 Product components

Water Quality Analyser contains the following modules:

- Data management and visualisation module
- Loads tool
- Trends tool
- Guidelines tool
- Dashboard
- eGuides

The installation file can be downloaded from the Toolkit website,

<http://www.toolkit.net.au>.

---

## 2.7 Version history

Date	Author	Revision	Description of change
September 2011	<ul style="list-style-type: none"><li>• Sunil Tennakoon</li><li>• Jason Shen</li><li>• Yanning How</li><li>• Michael Emery</li></ul>	2.0.1	Updates to reflect latest WQA software release
October 2010	<ul style="list-style-type: none"><li>• Sunil Tennakoon</li><li>• Jason Shen</li><li>• Michael Emery</li></ul>	2.0.0	Updates to reflect latest WQA software release
August 2008	<ul style="list-style-type: none"><li>• Sunil Tennakoon</li></ul>	1.0.0	Creation of user guide

## 2.8 Limitations and caution notes for users

- This package is being delivered "as is", and eWater Ltd makes no warranty as to its use, reliability or performance.
  - Users assume all risks associated with the quality, performance, installation and use of all utilities in Water Quality Analyser including, but not limited to, equipment, loss of data or software programs, or interruption to operations.
  - Users are solely responsible for determining whether Water Quality Analyser is appropriate for their own use or purpose and they assume all risk for any loss or damage resulting directly or indirectly from their use of or inability to use the software.
  - Users should be aware of input data requirements.
-

## Chapter 3

# Getting Started

[Installation and Uninstallation](#)

[Navigation](#)

[Toolbars](#)

[Panels](#)

[Data Operations](#)

### 3.1 Installation and Uninstallation

To install the software, **double-click** the supplied exe file and follow the wizard. If **Microsoft .NET Framework 4** is not installed on the host system, download and install it. Once the Water Quality Analyser software installation is complete, a start menu item appears in the **All Programs -> DERM -> WQA** submenu. **Double-click** the WQA shortcut to launch the software.

Minimum software requirements:

- Windows XP Service Pack 3
- Windows Installer 4.5
- .NET Framework 4.0
- SQL Server Compact 3.5 SP2
- 40MB free disk space

### 3.2 Navigation

The Water Quality Analyser interface consists of five main components:

- *Toolbars*: Contains a collection of tools for data operations and management
- *Projects panel*: Contains a list of projects, set and time series imported into the WQA database
- *Properties panel*: Contains information about the currently selected item
- *Chart layers panel*: Contains a list of all opened time series
- *Chart area*: Provides an interactive, visual representation of the opened time series

### 3.3 Toolbars

#### 3.3.1 Menu bar

The **menu bar** consists of seven main menu items: File, Edit, View, Process, Analyse, Tools and Window. The **File** menu provides data operations commands for time series data such as Import, Export and Save. The **Edit** menu provides options relating to the currently selected or active item as well as editing operations such as undo, redo and delete. The **View** menu contains a list of options dealing with the visibility of panels on the interface, the appearance of the chart and additional information windows such as the log menu. The **Process** menu contains a list of available processing tools relating to one or more time series. The **Analyse** menu contains a list of analysis tools such as a basic statistics menu, the loads tool, trends tool, guidelines tool and eGuides tool. The **Tools** menu contains administrative tools for database import and export, addition of indicators, units and other data types, application theme options and software updates. The **Window** menu provides tools to change the window layout.



Figure 3.1: Menu Bar

#### 3.3.2 Standard toolbar

The standard toolbar consists of a set of icons below the menu bar. The standard toolbar icons provide quick shortcuts to tools which would ordinarily be accessed through the menu bar. Each set of icons is split into groups representing the relevant category.



Figure 3.2: Standard Toolbar

## 3.4 Panels

### 3.4.1 Projects panel

The **Projects panel** displays a list of all projects, sets, time series and historical time series stored in the local database. It provides options to open, edit or delete time series. **Double-clicking** a time series in the projects panel opens it for processing. The items in the projects panel are displayed in a hierarchical view in the form of **Project** -> **Set** -> **Time series** -> **Historical time series**. To expand a hierarchy level, **left-click** the expansion arrow to the left of an item name. The expansion arrow points down to indicate that the hierarchy level is opened. To collapse a hierarchy level, **left-click** the expansion arrow again.

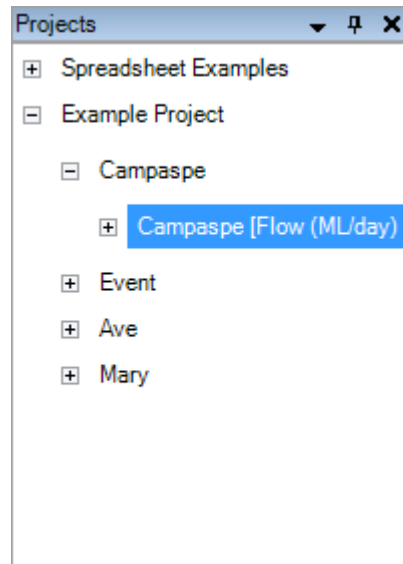


Figure 3.3: Projects Panel

### 3.4.2 Properties panel

The **Properties panel** displays a list of contextual information about the most recently selected item. When a project, set, time series or sample result is selected, the properties panel will display a concise list of information pertaining to that item. The information in the properties panel is for viewing only. In most cases, the properties can be edited through the **Edit** -> **Properties** button.

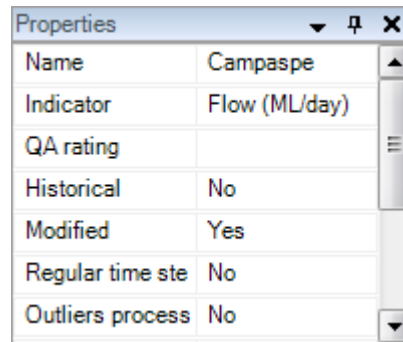


Figure 3.4: Properties Panel

### 3.4.3 Chart Layers Panel

The **Chart Layers panel** displays a list of all opened time series. Each time series is associated with a colour-coded layer. Each layer item contains a vertical bar indicating the colour of the layer, a checkbox to indicate whether the layer is visible, the name of the layer and a colour swatch button to view or edit the appearance of the layer. Operations such as selection or deletion of points in the chart only apply to the active layer. Only one layer can be active at a time. The active layer is indicated by the use of bold text, and a thicker axis line in the chart.

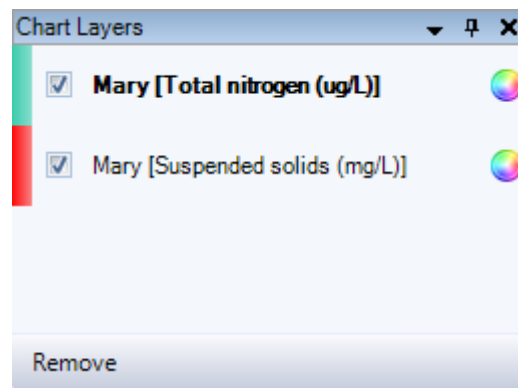


Figure 3.5: Chart Layers Panel

### 3.4.4 Panel Layout

#### 3.4.4.1 Unpinning a panel

By default, the **Projects panel**, **Properties panel** and **Chart layers panel** are pinned to the main interface. To unpin a panel, click the **pin** button in the top right corner of the panel. A tab appears with the panel name on the docked edge. Move the mouse away from the



panel to allow it to slide off the main interface. **Left-click** or **hover** over the tab to show the panel again.

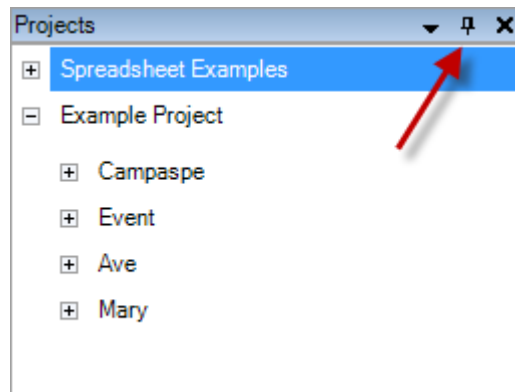


Figure 3.6: Panel Pin Button

#### 3.4.4.2 Floating a panel

To float a panel at any position in the main interface, **left-click** and **drag** the **title bar** of a panel into the centre of the main interface. The panel undocks and floats over the main interface. The floating panel can be moved using the same procedure.

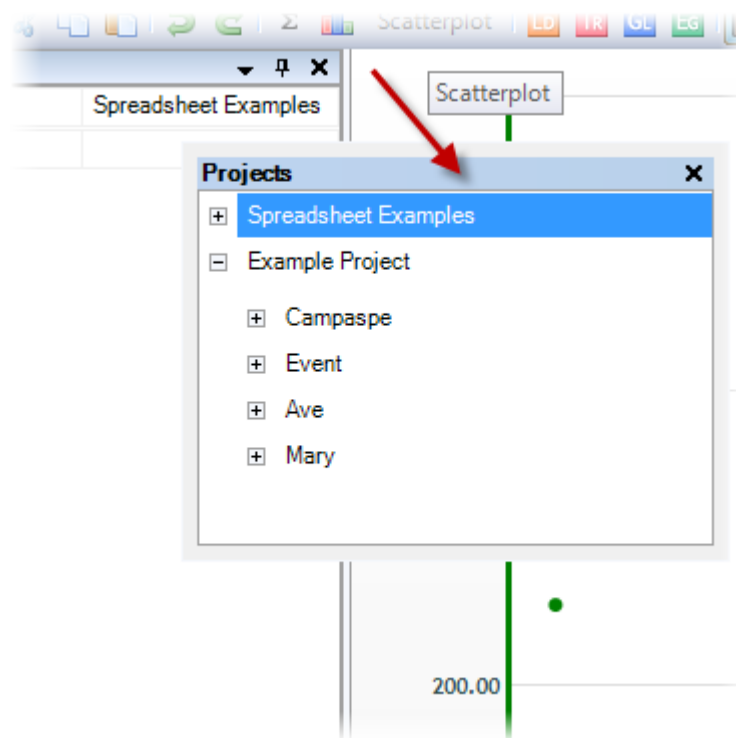


Figure 3.7: Panel Title Bar

#### 3.4.4.3 Docking a floating panel to a dock area

To dock a floating panel to a dock area, **left-click** and drag the **title bar** of a panel to an area over the main interface. The **docking indicators** appear. Release the mouse button over a docking indicator to dock to that location.

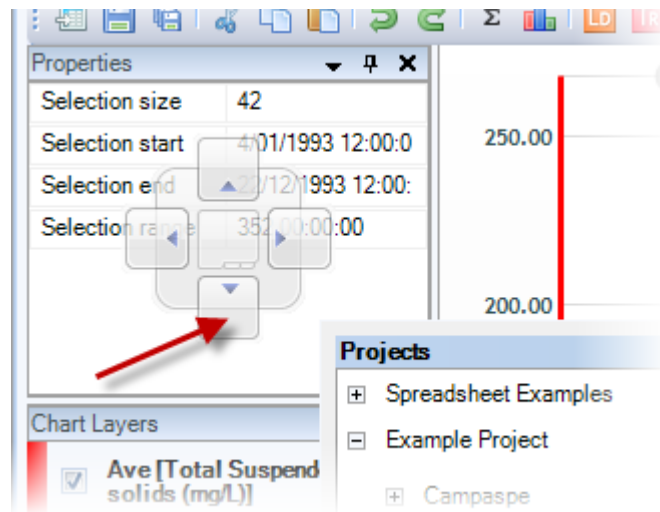


Figure 3.8: Docking a Panel

#### 3.4.4.4 Hiding or showing a panel

To hide a panel, click the **x** button in the top right corner of the panel. To show a panel again, click the panel name in the View menu.

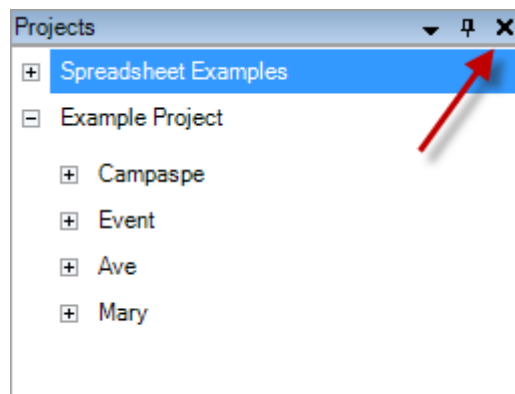


Figure 3.9: Panel Close Button

#### 3.4.4.5 Resetting the window layout

Resetting the window layout causes the toolbars and panels to be reset to the default positions. To reset the window layout, click the **Window -> Reset Window Layout** button.

## 3.5 Data Operations

All imported time series are stored in a local database. Two or more users can share databases by using the database export and import tools. To export a database, click the **Tools** -> **Export Database...** button. The database can be stored on the local machine or to external media for sharing. To import a database, click the **Tools** -> **Import Database...** button and select a database. Importing a database replaces the current database. To temporarily view a database without replacing the local database, click the **Tools** -> **Change Database...** button.

---

## Chapter 4

# Procedures

[Project Management](#)

[Time Series Management](#)

[Time Series Visualisation](#)

[Time Series Processing](#)

[Statistical Analysis](#)

[Load Estimation](#)

[Trend Tool](#)

[Dashboard](#)

[Guideline Tool](#)

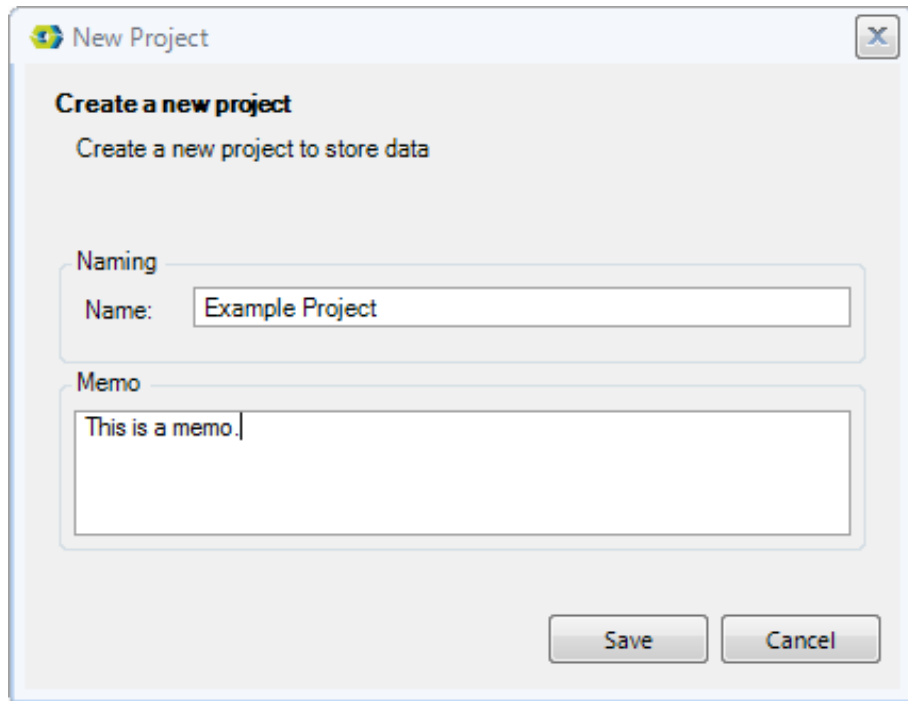
[eGuides](#)

### 4.1 Project Management

The Water Quality Analyser has a four level hierarchy consisting of projects, sets, Time series and historical time series. Under a **Project** you can store a number of sets and under a set a number of Time series can be stored.

#### 4.1.1 Creating a new project

To create a new project, click the **File** -> New Project... button. The **New Project** form appears. The name and a brief memo (500 characters or less) can be specified. Click the **Save** button to save the empty project. The **Project** appears in the **Projects panel**.



**New Project**

**Create a new project**  
Create a new project to store data

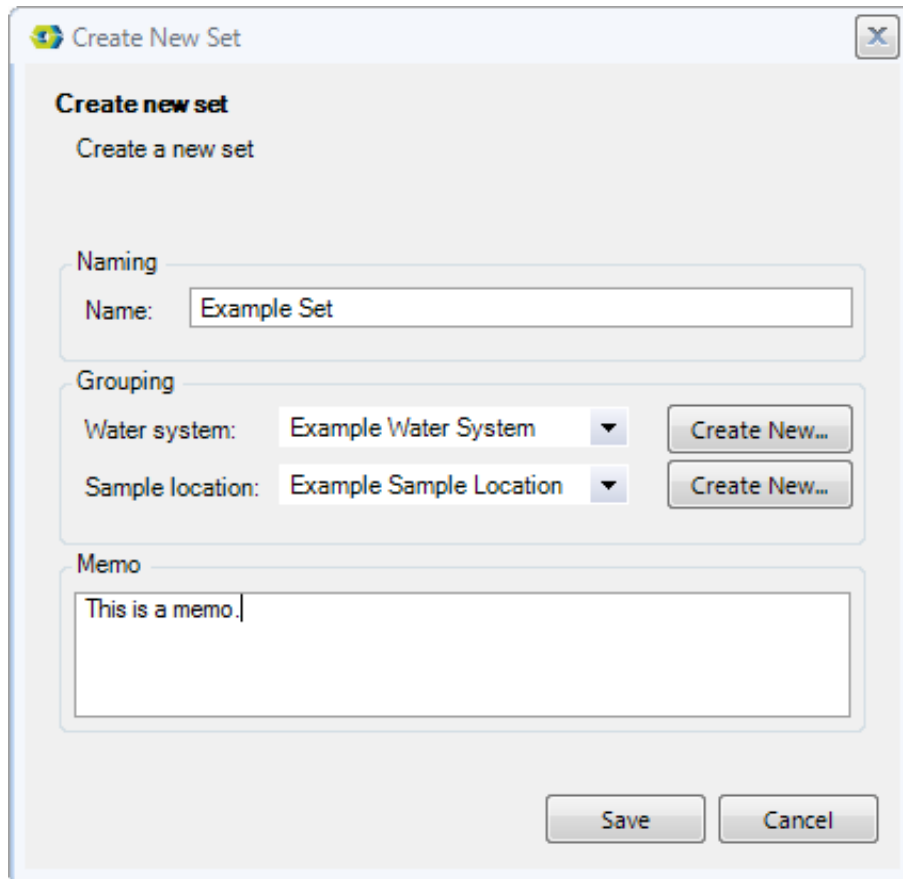
**Naming**  
Name:

**Memo**

Figure 4.1: New Project Form

#### 4.1.2 Creating a new set

To create a new set, first ensure a *Project* is selected in the *Projects panel*. Click the *File -> New Set...* button. The *Create New Time Series* form appears. The name, water system, sample location and a brief memo (500 characters or less) can be specified. Click the *Save* button to save the empty set. The set appears in the *Projects panel* under the selected project.



**Create new set**  
Create a new set

**Naming**  
Name:

**Grouping**  
Water system:    
Sample location:

**Memo**

Figure 4.2: New Set Form

### 4.1.3 Navigating projects, sets and time series

The **Projects panel** has a hierarchy of four levels. The top level is the **Project** level. A **Project** is used to hold a group of related sets. Each set is associated with a sample location. A set can contain one or more **Time series** from the same sample location. A **Time series** can contain zero or more historical time series. The four hierarchy levels (**Project** -> **Set** -> **Time series** -> **Historical time series**) can be accessed by **left-clicking** the expansion icons in the projects panel.

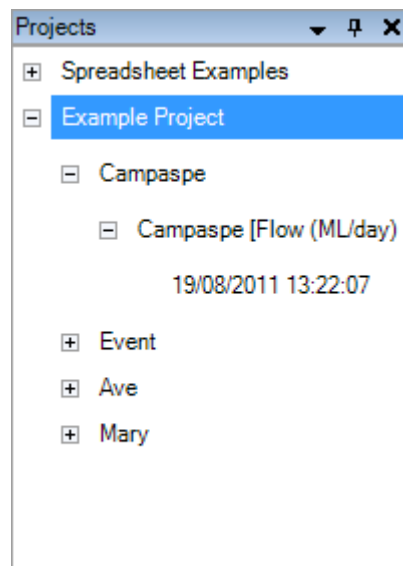
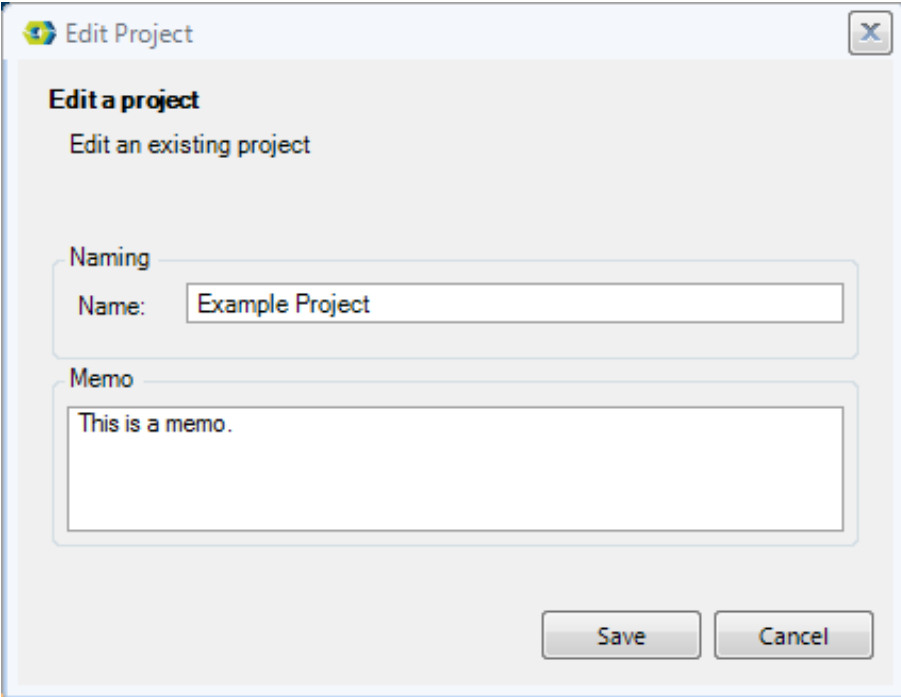


Figure 4.3: Projects Panel Expanded View

#### 4.1.4 Editing Project Properties

To edit the properties of a project, **left-click** the **Project** in the **Projects panel**, then click the **Edit -> Properties** button. An **Edit Project** form appears with fields to edit the name and to include a brief memo (less than 500 characters) associated with the project. Click the **Save** button to save any changes.





**Edit Project**

**Edit a project**  
Edit an existing project

**Naming**  
Name:

**Memo**

Figure 4.4: Edit Project Form

#### 4.1.5 Editing set properties

To edit the properties of a set, *left-click* the set in the **Projects panel**, then click the **Edit -> Properties** button. An **Edit Set** form appears with fields to edit the name, water system, sample location and to add a brief memo (less than 500 characters) associated with the set. Click the **Save** button to save any changes.

**Edit Set**

**Edit existing set**  
Edit an existing set

**Naming**  
Name:

**Grouping**  
Water system:    
Sample location:

**Memo**

Figure 4.5: Edit Set Form

#### 4.1.6 Editing Time series properties

To edit the properties of a time series, *left-click* the **Time series** in the **Projects panel**, then click the **Edit -> Properties** button. An **Edit Time Series** form appears with fields to edit the name, indicator, **Project** and set associated with the time series. Click the **Save** button to save any changes.

**Edit Time Series**

**Edit existing time series**  
Edit an existing time series

**Naming**  
Name:

**Grouping**

Indicator:	<input type="text" value="Total Suspended solids (mg/L)"/>	<input type="button" value="Create New..."/>
Project:	<input type="text" value="Example Project"/>	<input type="button" value="Create New..."/>
Set:	<input type="text" value="Ave"/>	<input type="button" value="Create New..."/>

Figure 4.6: Edit Time series Form

#### 4.1.7 Deleting a project

To delete a project, **left-click** the **Project** in the **Projects panel**, then click the **Edit -> Delete** button. A confirmation dialog appears to confirm the deletion. Any sets will be deleted along with the project.

#### 4.1.8 Deleting a set

To delete a set, **left-click** the set in the **Projects panel**, then click the **Edit -> Delete** button. A confirmation dialog appears to confirm the deletion. Any **Time series** will be deleted along with the set.

#### 4.1.9 Deleting a time series

To delete a time series, **left-click** the **Time series** in the **Projects panel**, then click the **Edit -> Delete** button. A confirmation dialog appears to confirm the deletion. Any historical **Time series** will be deleted along with the time series.

---

#### 4.1.10 Moving a Time series into another set

To move a **Time series** into another set, **left-click** the **Time series** in the **Projects panel**, then click the **Edit -> Properties** button. The **Edit Time Series** form appears. The **Grouping** box displays options to modify the **Project** and set associated with the time series. Select a new set in the **Set** combo box, then click the **Save** button. The **Time series** is moved to the new set.

#### 4.1.11 Creating or editing indicators and units

To create or edit an indicator, unit, water system, sample location, primary water type or secondary water type, click the **Tools -> Manage Metadata...** button. The **Edit Indicators and Units** form appears. Clicking a tab shows the data associated with the name of the tab. To create a new item, click the **Create New...** button. To edit an existing item, either click a cell to edit the value or click the **Edit...** button to view an editing form for the active row.

Name	Short Name	Is Flow	Unit
% Alien	% Alien	<input type="checkbox"/>	%
1,1-Dichloroethene	1,1-Dichloroethene	<input type="checkbox"/>	ug/L
1,1-dichloroethene	1,1-dichloroethene	<input type="checkbox"/>	mg/L
1,1,2-trichloroethane	1,1,2-trichloroethane	<input type="checkbox"/>	ug/L
1,2-dichlorobenzene	1,2-dichlorobenzene	<input type="checkbox"/>	ug/L
1,2-dichlorobenzene	1,2-dichlorobenzene	<input type="checkbox"/>	mg/L
1,2-Dichloroethane	1,2-Dichloroethane	<input type="checkbox"/>	ug/L
1,2-dichloroethane	1,2-dichloroethane	<input type="checkbox"/>	mg/L
1,2-dichloroethene	1,2-dichloroethene	<input type="checkbox"/>	mg/L
1,2,3-trichlorobenzene	1,2,3-trichlorobenzene	<input type="checkbox"/>	ug/L
1,2,4-trichlorobenzene	1,2,4-trichlorobenzene	<input type="checkbox"/>	ug/L
1,3-dichlorobenzene	1,3-dichlorobenzene	<input type="checkbox"/>	ug/L
1,3-dichlorobenzene	1,3-dichlorobenzene	<input type="checkbox"/>	mg/L
1,4-dichlorobenzene	1,4-dichlorobenzene	<input type="checkbox"/>	ug/L

Figure 4.7: Edit Indicators and Units Form

## 4.2 Time Series Management

### 4.2.1 Importing a time series

The standard method of accessing data in the Water Quality Analyser is the **Time Series Import Wizard**. To import one or more time series, click the **File -> Import...** button. The **Time series import wizard** appears.

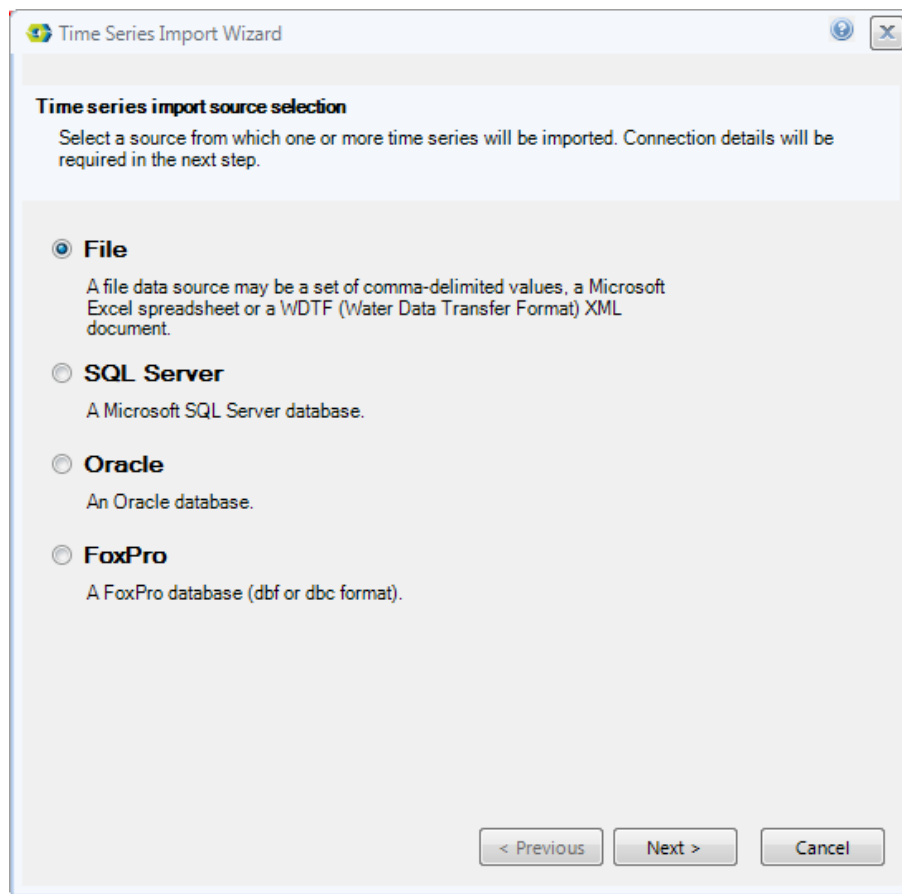
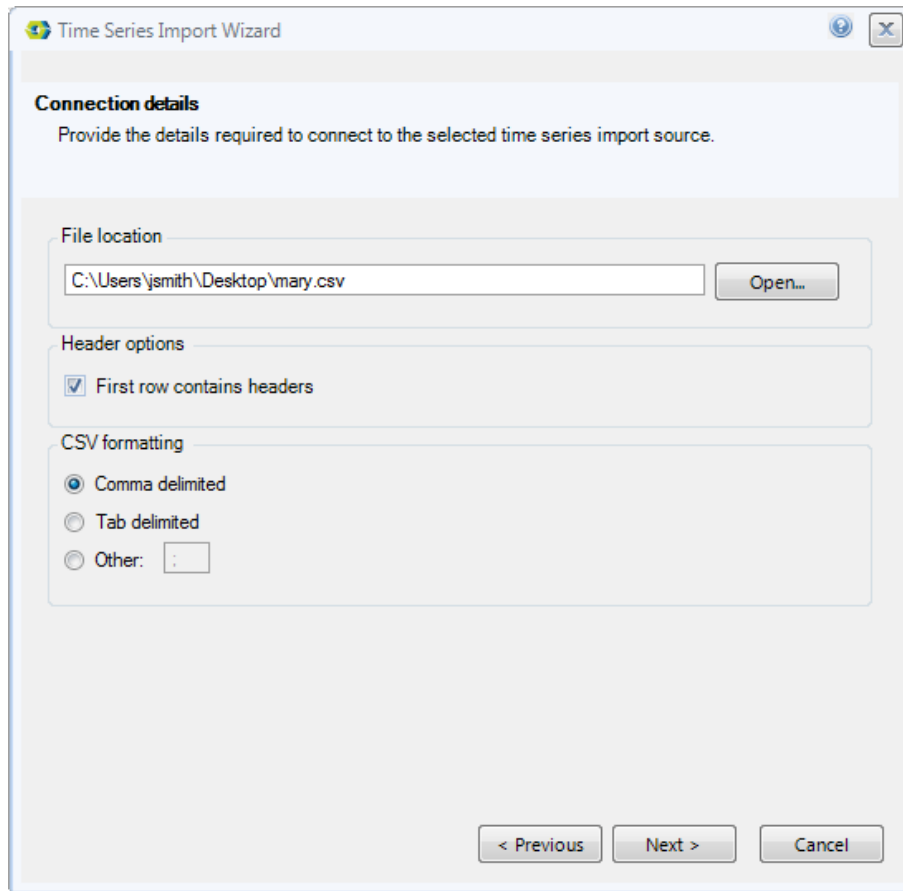


Figure 4.8: Import Source Selection

The **Import source selection** step provides options to choose a File, SQL Server, Oracle or FoxPro data source. In most cases the **File** option is appropriate. Click the **Next** button to continue to view the **Connection details** step.



The screenshot shows a window titled "Time Series Import Wizard" with a standard Windows interface (help, close buttons). The main heading is "Connection details" with a sub-instruction: "Provide the details required to connect to the selected time series import source." The form is divided into three sections: "File location" with a text box containing "C:\Users\jsmith\Desktop\mary.csv" and an "Open..." button; "Header options" with a checked checkbox "First row contains headers"; and "CSV formatting" with three radio button options: "Comma delimited" (selected), "Tab delimited", and "Other:" followed by a small text box containing a colon ":". At the bottom are three buttons: "< Previous", "Next >", and "Cancel".

Figure 4.9: Connection Details

The connection details step provides inputs to connect to the specified data source. If the **File** option was chosen, a field will be available to specify the location of a spreadsheet or comma delimited file. If the first row of the data contains labels or headers, check the **First row contains headers** checkbox. If a comma delimited **File** is opened, the **CSV formatting** box appears. Specify the type of delimiter, such as comma or tab used by the file. Click the **Next** button to continue to the **Column configuration** step.

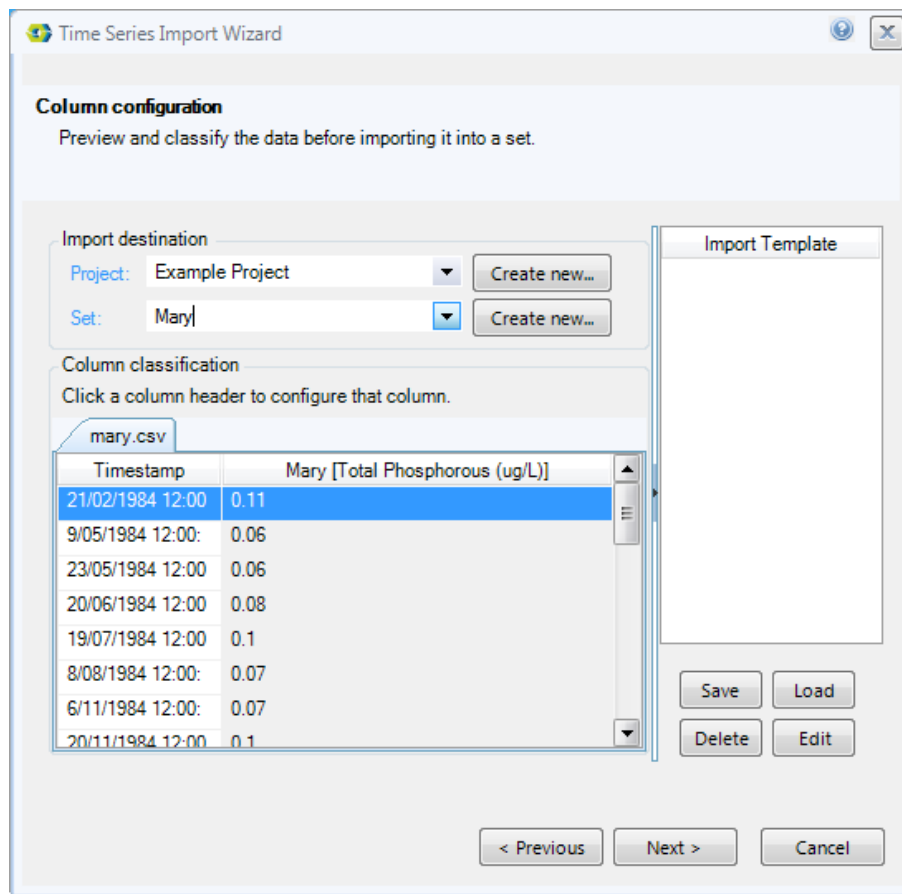


Figure 4.10: Column Configuration

The column configuration step displays a data preview in the **Column classification** box. If the import data contains multiple worksheets, they will be displayed as tabs above the data preview. To select a worksheet, **left-click** the associated tab. The data preview will show data from the selected worksheet.

Each column of the import data must be enabled and classified before data can be imported. To enable and classify a column, **left-click** the column header. The column classification popup appears.

Column Classifier

Mary [Total Phosphorous (ug/L)]

☒ Enabled

Column type: Indicator

Indicator: Flow (ML/day)

Conversion factor: 1.000000

☒ Ignore negative values

Import destination

☒ New time series

☐ Existing time series

Time Series [Total Phosphorous (ug/L)]

Conflicts

☒ Keep newer points

☐ Keep older points

☐ Keep all points

Close

Figure 4.11: Column Classifier

Check the **Enabled** checkbox to enable the column. Specify the column type (indicator, time stamp or depth value). An indicator column has additional options allowing you to specify an indicator name and a conversion factor (optional: for converting data between units) before data import, as well as an option to ignore negative values. The **Import destination** box provides options for importing data into a new **Time series** or an existing time series. The **Conflicts** box provides three options to deal with conflicting sample results when importing into an existing **Time series** (keep newer, keep older, and keep all). If the **Keep all** option is selected, conflicting sample results can still be processed using the **Conflict processor** after the **Time series** is imported.

When a time stamp column and at least one indicator column is enabled and configured, click the **Next** button to continue to the **Filter Sample Results** step.



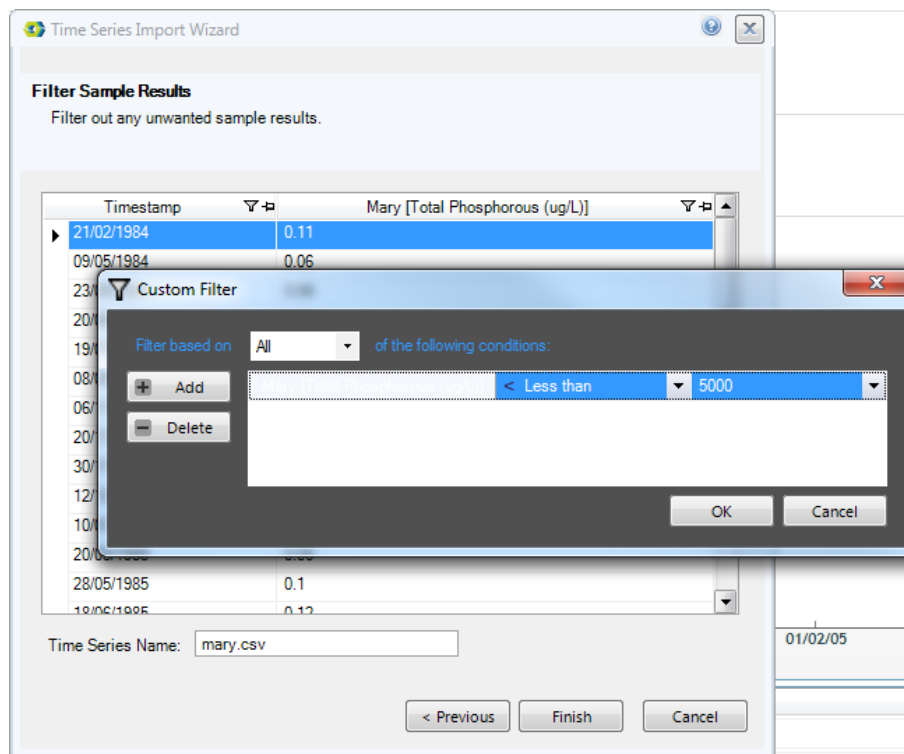


Figure 4.12: Filter Sample Results

This final step provides tools to filter the data. Click the **Filter** button at the right of a column header to view basic filtering options. Click the **(Custom)** option to view the **Custom Filter** form. One or more conditions can be added as a custom filter using the **Add** button. Each filter parameter accepts an operator (left combo box) and a value (right combo box). Click the **OK** button to apply the filter. The preview window displays only the sample results which satisfy the filter conditions. The name of the time series can be changed in **Time Series Name** textfield. Click the **Finish** button to import the data. The imported **Time series** appears in the **Projects panel** under the specified **Project** and set name.

#### 4.2.2 Importing a Time series with data from several locations

If a time series contains data from several locations, it can be imported by identifying the location columns during the **Column classification** step of the **Time Series Import Wizard**. **Left-click** the column header which contains the location name. The column classification popup appears. Check the **Enabled** checkbox and set the column type to "location". When the data is imported, a set will be created for each location and the time series will be divided into these sets.

### 4.2.3 Opening and closing Time series

To view or edit an imported time series, it must be open in the **Chart layer panel**. To open a time series, expand the **Project** and set in the **Projects panel**. **Double-click** the **Time series** name to open it. The **Time series** will appear in the chart and the chart layer panel. To close an opened time series, click the **Time series** name in the chart layer panel then click the **Remove** button.

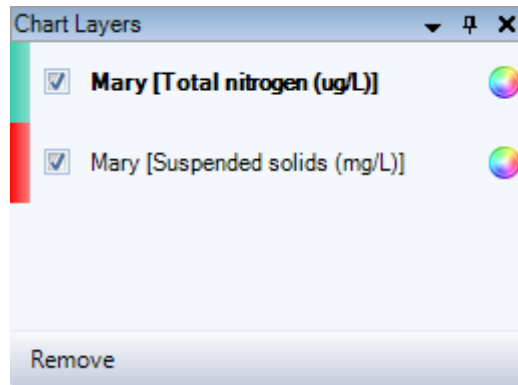


Figure 4.13: Chart Layers Panel

### 4.2.4 Creating an empty Time series

An empty **Time series** can be created by clicking the **File -> New -> Time Series...** button. The **Create New Time Series** form will appear. The **Time series** name, indicator, **Project** and set can be specified. Click the **Save** button to create the new time series. The new **Time series** will be stored in the specified **Project** and set in the **Projects panel**.

### 4.2.5 Editing Time series properties

To edit the properties of a time series, **left-click** the **Time series** in the **Projects panel**, then click the **Edit -> Properties** button. The **Edit Time Series** form will appear. The **Time series** name, indicator, **Project** and set can be modified. The **Time series** log can be viewed by clicking the **View Log** button. The log will not show any unsaved changes.

### 4.2.6 Deleting a Time series

To delete a time series, **left-click** the **Time series** in the **Projects panel**, then click the **Edit -> Delete** button. A prompt will be displayed to confirm the action. Click the **Yes** button to confirm and delete the time series. If the **Time series** is open in the **Chart layers panel**, it will be closed.

### 4.2.7 Viewing a Time series log

Each **Time series** contains a log of any processes performed on the time series. To view the log of an opened time series, click the **View -> Log** button. The **Log Viewer** displays the details of each process performed in descending chronological order.

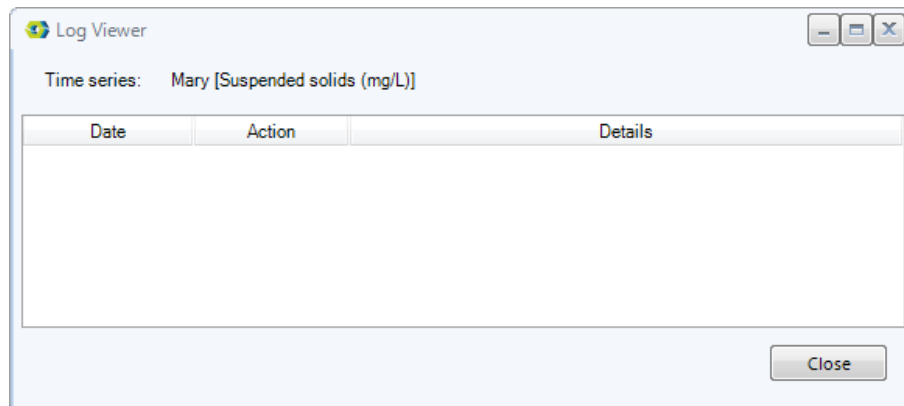


Figure 4.14: Log Viewer

### 4.2.8 Saving a time series

To save an edited time series, click the **File -> Save** button. The active layer will be saved. To save all opened time series, click the **File -> Save All** button. When a **Time series** is saved, a historical **Time series** is created which preserves the history of the time series.

### 4.2.9 Saving a Time series under a different name

To save an opened **Time series** under a different name, first activate the layer in the **Chart layers panel**. Click the **File -> Save As...** button and enter a name for the new **Time series** in the Name field. Click **Save** to save the new **Time series** using the specified name.

### 4.2.10 Viewing historical time series

When a **Time series** is modified and saved, a historical version of the **Time series** is created. A historical **Time series** contains a copy of the **Time series** before any changes were made. To view all historical **Time series** of a time series, expand the **Time series** name in the **Projects panel**. The historical **Time series** are displayed and are labelled by the date they were created. **Double-click** a historical **Time series** to view it. The processes performed in a historical **Time series** can be viewed by clicking the **View -> Log** button.

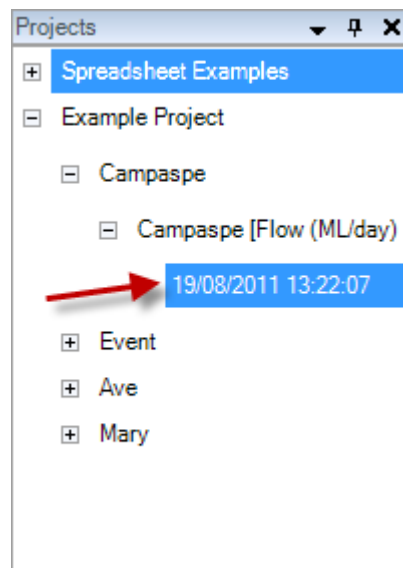


Figure 4.15: Historical Time series Expanded View

#### 4.2.11 Editing and saving a historical time series

Historical **Time series** can be edited and saved like a normal time series, but a historical **Time series** will be copied and converted to a normal **Time series** when it is first saved to preserve the history of the time series.

#### 4.2.12 Merging two or more time series

To merge two or more **Time series** into a single time series, first select the **Time series** you wish to merge in the **Projects panel** by holding the **control** or **shift** keys while **left-clicking** the **Time series** names. Click the **Edit -> Merge Time Series...** button to display the **Time Series Merge Wizard**.

The wizard displays a chart with the two **Time series** displayed as one. Any conflicting sample results are highlighted in the chart. To step back or forward through each conflict, click the **Previous conflict** or **Next conflict** button. The active conflict is highlighted with a red vertical stripe while all other conflicts are displayed with a grey vertical stripe.

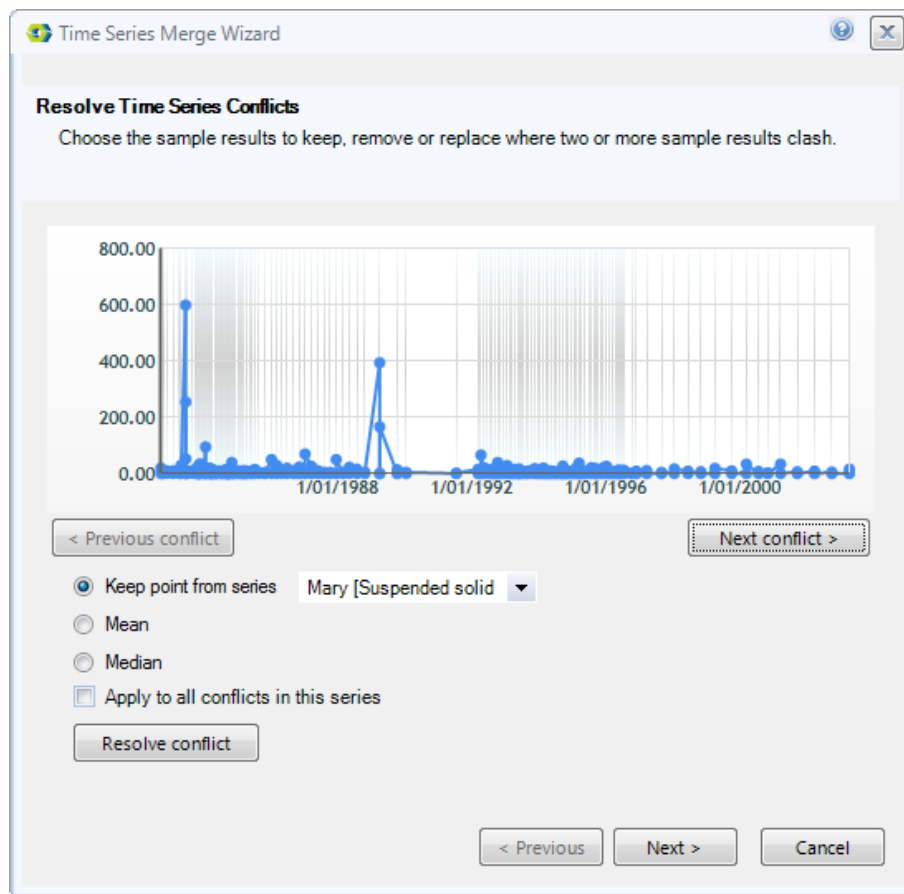


Figure 4.16: Time Series Merge Wizard

To resolve the active conflict, click the Keep point from series, mean or median radio button. The Keep point from series option provides a combo box to select a **Time series** to preserve. Any points from other conflicting **Time series** are discarded. Click the **Resolve conflict** button to resolve the active conflict using the chosen method.

To resolve a group of conflicts, check the **Apply to all conflicts** in this series checkbox, then click the **Resolve conflict** button to resolve the conflicts using the chosen method.

When all conflicts have been resolved, the chart is expanded to provide a preview of the merged time series. Click the **Next** button to confirm the action. The **Merge destination** box provides an option to overwrite one of the merged **Time series** while discarding all other **Time series** involved in the merge. Choose a **Time series** to retain and click the **Finish** button.

### 4.2.13 Copying and pasting sample results to and from external applications

To copy one or more sample results from the chart, select the required sample results and click the **Edit -> Copy** button. The sample results are copied in tab-delimited format. Copied sample results can be pasted directly into an external document or spreadsheet.

To copy a column of data from an external spreadsheet, first ensure a date column is directly adjacent to a data column. Select the data in both columns and click the **Edit -> Copy** button in the spreadsheet software. Return to the Water Quality Analyser and click the **Edit -> Paste** button after a **Time series** has been opened in the **Chart layers panel**. The data will be converted to sample results and presented in the chart.

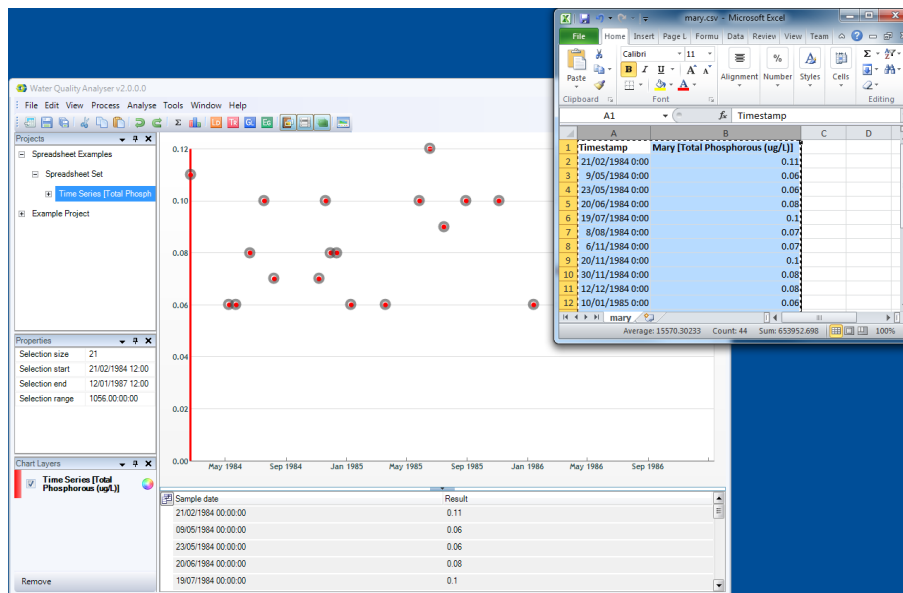


Figure 4.17: Copying From a Spreadsheet

**Time series** data can be directly pasted into an existing **Time series** from any source if it is in comma or tab delimited format.

### 4.2.14 Exporting a time series

To export a time series, the **Time series** must first have been opened in the **Chart layers panel**. Click the **File -> Export...** button to export the time series. The **Time series** will be saved in Microsoft Excel 2007 format.

### 4.2.15 Exporting a chart image

To export a chart image, a **Time series** must first have been opened in the **Chart layers panel**. Click the **File -> Export Image...** button to export an image of the chart. If the chart is zoomed or panned, the saved image will reflect those changes.

## 4.3 Time Series Visualisation

### 4.3.1 Layers

#### 4.3.1.1 Opening a time series

To open a **Time series** for viewing in the **Data management and visualisation module**, import the **Time series** into the **Projects pane**. Expand the **Project** and click the expansion icon next to the **Project** to set name. **Double-click** the **Time series** name to open it. The **Time series** will be displayed in the chart and in the **Chart layers panel**.

#### 4.3.1.2 Overlaying two or more time series

To display multiple **Time series** overlayed on the chart, open each time series. Each **Time series** will be displayed in the chart and in the **Chart layers panel**. The Y-axis of each **Time series** is adjusted to fit equally on the chart.



Figure 4.18: Overlaid Time Series

#### 4.3.1.3 Activating a Time series layer

To activate a layer when two or more **Time series** are opened, click the layer name. The active layer name appears bold while inactive layers appear without emphasis. Any actions performed in the chart (such as selecting or deleting sample results) only apply to the active layer.

#### 4.3.1.4 Temporarily hiding a time series

To temporarily show or hide an opened time series, click the checkbox in the **Chart layers panel** next to the layer name. The **Time series** will be removed from the chart.

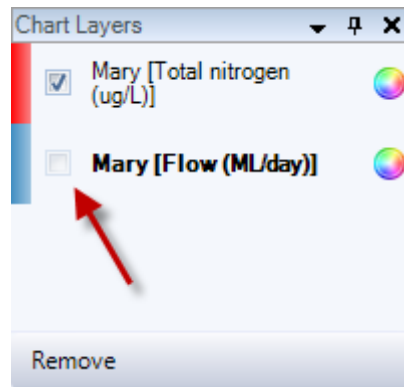


Figure 4.19: Hidden Time Series

#### 4.3.1.5 Removing a time series

To remove a **Time series** from the **Chart layers panel**, left-click the layer name, then click the **Remove** button. If there are any unsaved changes to the layer, a prompt to save or discard the changes will be displayed.

#### 4.3.1.6 Customising the appearance of a time series

To customise the colour, opacity and style of an opened time series, click the **Colour swatch** button next to the layer name. The **Layer Properties** form is displayed.



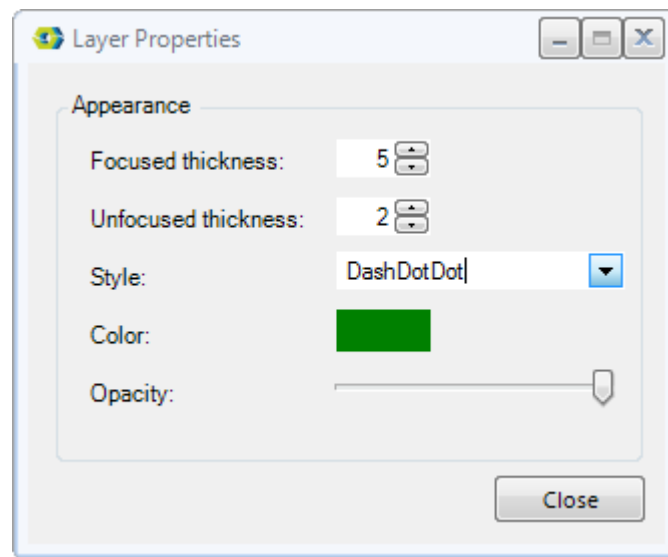


Figure 4.20: Layer Properties

## 4.3.2 Chart

### 4.3.2.1 Zooming with the mouse wheel

To zoom the chart with the mouse wheel, **left-click** on the chart once to give it focus. Position the cursor over an area of interest and scroll the mouse wheel forward to zoom to that area. To zoom out, scroll the mouse wheel in the opposite direction. To reset to the standard zoom level, click the **Edit -> Chart Zoom** Reset button.

### 4.3.2.2 Zooming with click and drag

To zoom using click and drag, **right-click** and **drag** over an area of interest in the chart. Release the mouse button to zoom to the area of interest.

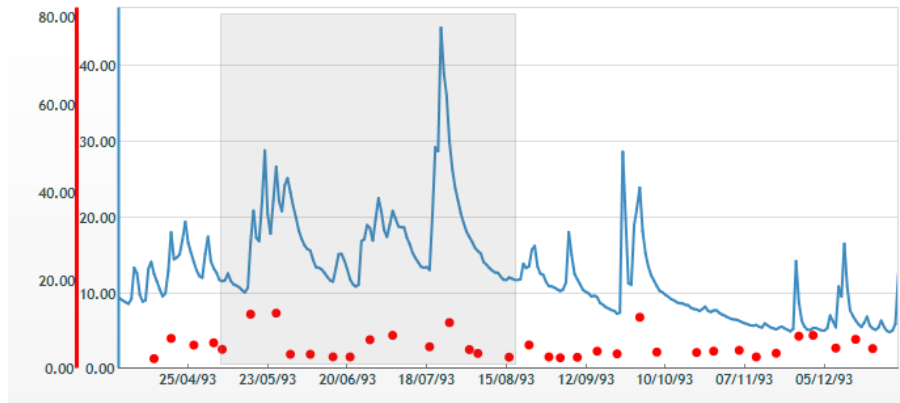


Figure 4.21: Zooming With Click and Drag

#### 4.3.2.3 Zooming to an existing selection

To zoom to an existing selection, first ensure that two or more points are selected in the chart. Click the **View -> Zoom to Selection** button. The chart zooms to the window of selected points.

#### 4.3.2.4 Resetting the zoom level

To reset the chart to the standard zoom level, click the **View -> Chart Zoom Reset** button. The chart resets to the default zoom level or the values specified in the **View -> Axis Options** menu.

#### 4.3.2.5 Panning through a zoomed time series

To pan through a zoomed chart, hold the **spacebar** then **left-click** and **drag** the chart. The chart pans in the direction of the mouse. Release the left mouse button or spacebar to stop panning.

#### 4.3.2.6 Setting the default range for an axis

Click the **View -> Axis Options** button to view the default minimum and maximum values for the X and Y axes. If the checkbox of a value is unchecked, the value is automatically calculated based on the range of the opened time series.

If an axis minimum and maximum is specified, the chart can still be zoomed. The specified values will be used when the chart zoom level is reset.

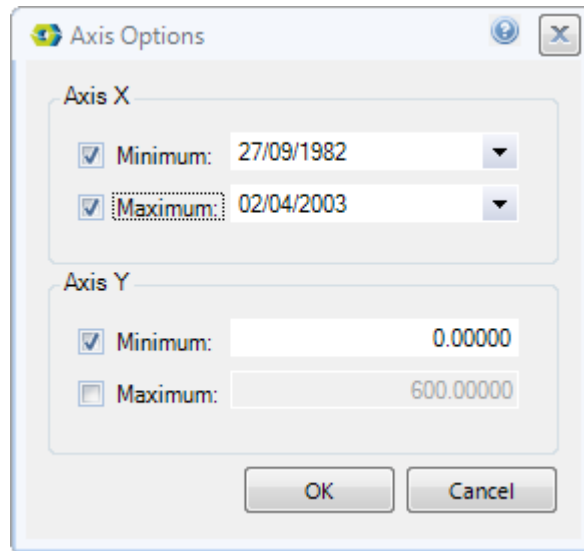


Figure 4.22: Axis Options

### 4.3.3 Interaction

#### 4.3.3.1 Displaying a Time series in XY plot mode

By default, flow **Time series** are shown in line plot mode and all other **Time series** are shown in XY plot mode. To toggle the display mode, click the View -> XY Plot button.

#### 4.3.3.2 Displaying scale breaks

Scale breaks collapse blank areas of the chart to preserve space. To enable or disable scale breaks, **right-click** the chart and click the **Scale break (non-interactive)** button. Scale breaks only appear if a vertical range of 10% or more can be collapsed. When scale breaks are enabled, the chart is not interactive (the chart cannot be zoomed or panned). To re-enable chart interactivity, disable scale breaks.

#### 4.3.3.3 Displaying the chart legend

To show or hide the chart legend, **right-click** the chart and click the **Toggle legends** button. The legend appears at the top right of the chart. If a chart image is exported, the legend is included in the exported image.

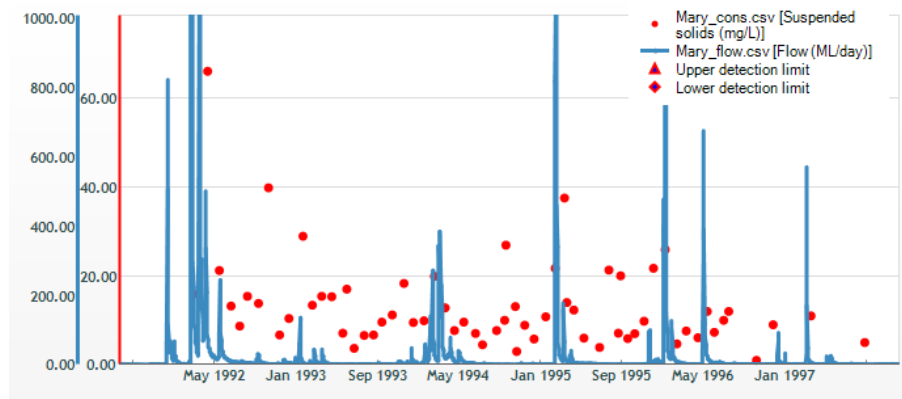


Figure 4.23: Chart Legend

#### 4.3.3.4 Selecting sample results on the chart

To select a single sample result, click the sample result in the chart. To select multiple sample results, **left-click** and **drag** on the chart. Any points within the selection box are selected when the mouse button is released.

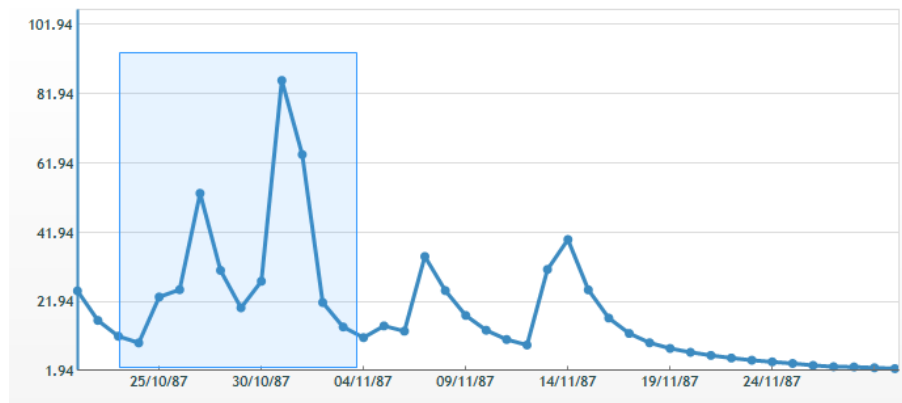


Figure 4.24: Selecting Sample Results on the Chart

#### 4.3.3.5 Selecting sample results using the grid

To select a sample result using the grid, click on any cell in a row. To select multiple sample results, click a row selector (the leftmost column) then hold the **shift** key and click another row selector. All intermediate rows will be selected. To select rows on an individual basis, hold the **control** key and click one or more row selectors.

#### 4.3.3.6 Inverting a selection

To invert an existing selection, press the **Select -> Invert Selection** button.

#### 4.3.3.7 Cancelling a selection

To cancel a selection, press the **escape** key or click in an empty area of the chart.

#### 4.3.3.8 Adding to an existing selection

To add one or more sample results to an existing selection, hold the **shift** key, then **left-click** and **drag** the sample results to be selected.

#### 4.3.3.9 Subtracting from an existing selection

To subtract one or more sample results from an existing selection, hold the **alt** key, then **left-click** and **drag** the sample results to be deselected.

#### 4.3.3.10 Adding or subtracting a single sample result from an existing selection

To add or subtract a single sample result from an existing selection, hold the **shift** key and **left-click** the sample result.

#### 4.3.3.11 Selecting baseflow

To select the baseflow of a flow time series, click the **Edit -> Select Baseflow** button. To select quickflow, first select baseflow then click the **Edit -> Invert Selection** button. The Baseflow level is calculated using a three pass Lyne-Hollick filter.

#### 4.3.3.12 Viewing properties of a selection

To view the properties of a selection of sample results, select two or more sample results in the chart or grid. The **Properties** panel will display the properties of the selection. The properties include the total number of sample results in the selection, the start and end dates of the selection and the total range of the selection.

#### 4.3.3.13 Viewing properties of a sample result

To view the properties of a sample result, click the sample result in the chart or grid. The **Properties** panel will display the properties of the sample result.

---

#### 4.3.3.14 Switching to fullscreen mode

To view the Water Quality Analyser in fullscreen mode, click the **View** -> **Fullscreen** button.

## 4.4 Time Series Processing

### 4.4.1 Editing the properties of a sample result

To edit the properties of a sample result, click the sample result in the chart or grid, then click the **Edit** -> **Properties** button. The sample date, result, depth, QA rating, data type and sub data type can be updated.

**Edit Sample Result**

**Edit existing sample result**  
Edit an existing sample result

**Properties**

Indicator: Total Suspended solids (mg/L)

Sample date: 18/01/1993 12:00:00 AM

Result: 128.8 mg/L

Depth: 0.0

QA rating: 10

**Data type**

Data type: Normal

Sub data type: Normal

Save Cancel

Figure 4.25: Edit Sample Result Form

#### 4.4.2 Deleting a sample result

To delete a sample result, click the sample result in the chart or grid, then click the **Edit** -> **Delete** button.

#### 4.4.3 Deleting a group of sample results

To delete a group of sample results, select two or more sample results in the chart or grid, then click the **Edit** -> **Delete** button. All selected points will be deleted.

#### 4.4.4 Trimming a time series

To trim a **Time series** to a specific event or area of interest, select all points to be kept. Then invert the selection (**Edit** -> **Invert Selection**) and press Delete. The unwanted areas of the **Time series** will be deleted. Alternatively, a range of results can be selected in the grid by left-clicking the row selector then shift + left-clicking another row. All intermediate rows will be selected.

#### 4.4.5 Infilling gaps in a time series

A **Time series** may contain missing data or "gaps" in an otherwise regular time series. To infill any gaps in a time series, use the **infill** processor (Process -> Infill). The infill processor detects the sample frequency of a **Time series** and finds any areas where sample results are missing.

The infill processor presents a summary of the action to be performed in the **Gap filling method** box. If the **Time series** contains gaps, the Gap filling method shows the size of the smallest gap. Gaps are also highlighted in the chart. If the **Time series** does not have a fixed time step, the greatest common divisor will be used and the summary information will indicate "(GCD)" rather than "(recommended)".

The **Curve fitting method** box provides a choice of three infilling methods (linear, cosine, mean).

---

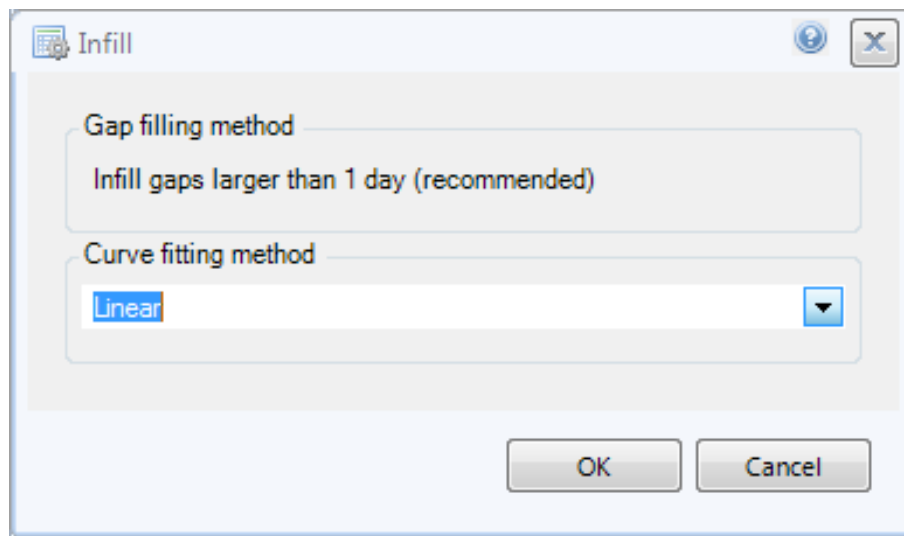


Figure 4.26: Infill Processor

#### 4.4.6 Resampling a time series

The *resample* processor can resample a *Time series* to a new sample frequency. If the *Time series* has several apparent sample frequencies, they will be presented in the *Resample using an existing gap* box. A custom frequency can be selected by clicking the *Resample using a custom gap* radio button and entering a time span in the time span editor. The *Curve fitting method* box provides a choice of three infilling methods (linear, cosine, mean).



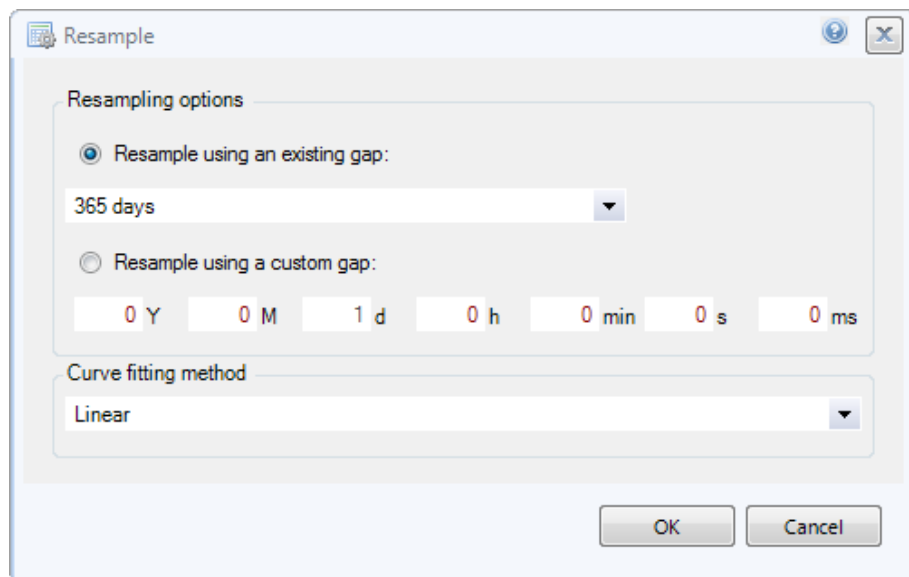


Figure 4.27: Resample Processor

#### 4.4.7 Aggregating a time series

The *aggregate* processor can aggregate a *Time series* using one of several aggregation modes (sum, mean, median, random). The random option takes a random sample result of each group and discards any other results. The time span selector can be used to specify the aggregated sample frequency.

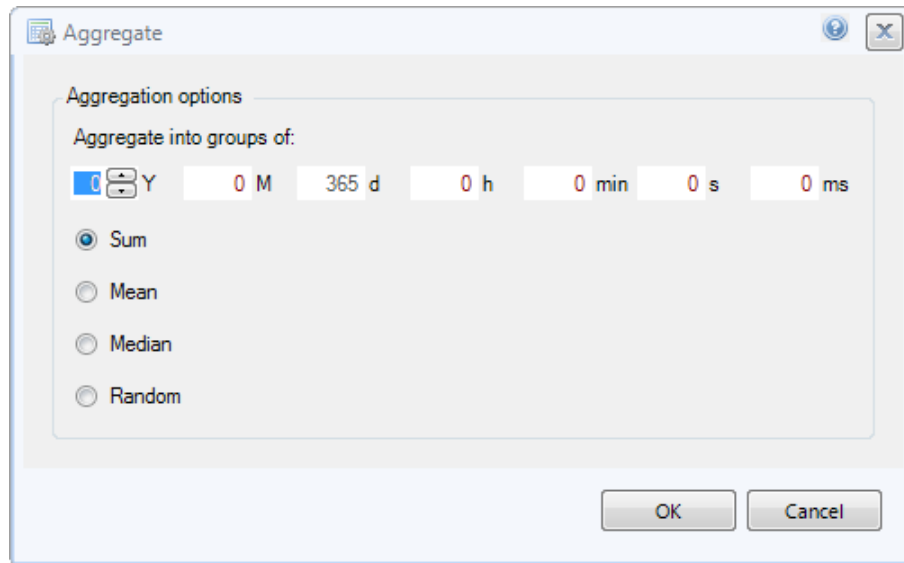


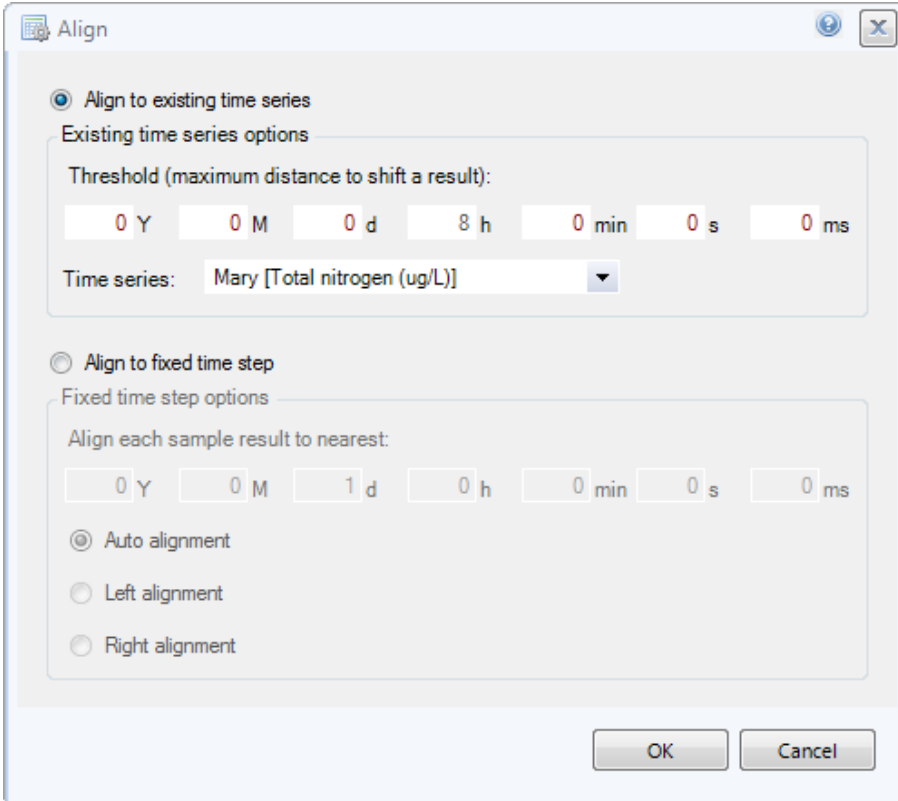
Figure 4.28: Aggregate Processor

#### 4.4.8 Aligning a Time series to another Time series or a fixed time step

The **align** processor is used to correct small errors in the sample dates of a time series. The processor can align the sample results to a specific time interval or to the time interval of another time series. This can be used when two **Time series** (such as Flow and TP or TN) are related but their sampling dates are slightly mismatched.

To align a **Time series** to another time series, ensure both **Time series** are in the same set, then click the **Align to existing time series** radio button. The align processor provides a threshold option (the maximum distance a sample result can be shifted) and a combo box to specify the **Time series** to be used as a reference.

To align a **Time series** to a fixed time step, click the **Align to fixed time step** radio button. The time span selector can be used to specify the time interval for alignment. If the **Left alignment** or **Right alignment** radio buttons are clicked, sample results will only be shifted to the nearest time interval on the left or right.



The image shows a software dialog box titled "Align". It contains two main sections. The first section, "Align to existing time series", is selected with a radio button. It includes a sub-section "Existing time series options" with a "Threshold (maximum distance to shift a result):" field set to 0 Y, 0 M, 0 d, 8 h, 0 min, 0 s, 0 ms, and a "Time series:" dropdown menu showing "Mary [Total nitrogen (ug/L)]". The second section, "Align to fixed time step", is unselected. It includes a sub-section "Fixed time step options" with an "Align each sample result to nearest:" field set to 0 Y, 0 M, 1 d, 0 h, 0 min, 0 s, 0 ms, and three radio buttons: "Auto alignment" (selected), "Left alignment", and "Right alignment". At the bottom right are "OK" and "Cancel" buttons.

Figure 4.29: Align Processor

#### 4.4.9 Shifting the dates of a time series

The **shift** processor can shift the **Time series** back or forward in time by a fixed time span. To shift using a relative offset, click the **Relative** radio button and enter a time span in the time span selector. To shift using an absolute offset, click the **Absolute** radio button and enter a date into the date selector. The first sample result will be shifted to this date and all other points updated accordingly.

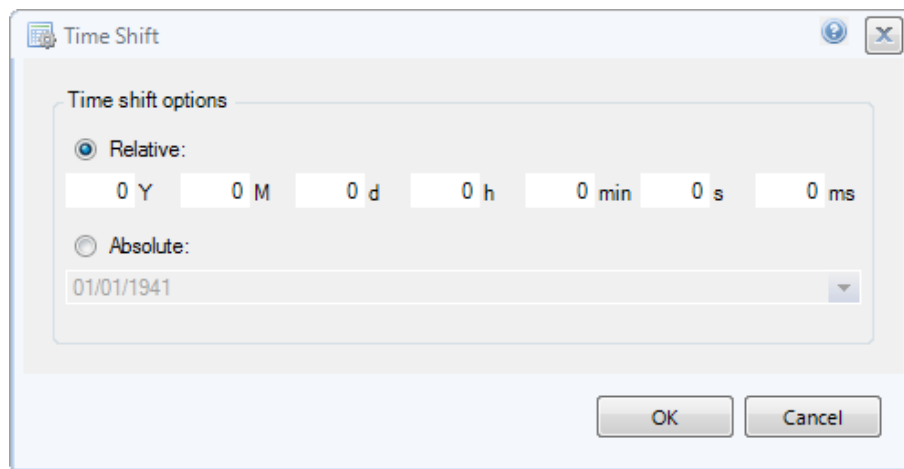


Figure 4.30: Shift Processor

#### 4.4.10 Transforming a time series

The **transform** processor can apply an expression to all sample results in a time series. To apply an expression, enter it into the **Expression** box. Click the **Quick reference** button for more information about the supported syntax. **x** represents the sample result value.

Sample results can be **deleted** by creating an expression that returns **false**. For example, `if(x >= 0.05, true, false)` would delete all sample results whose values are less than 0.05.

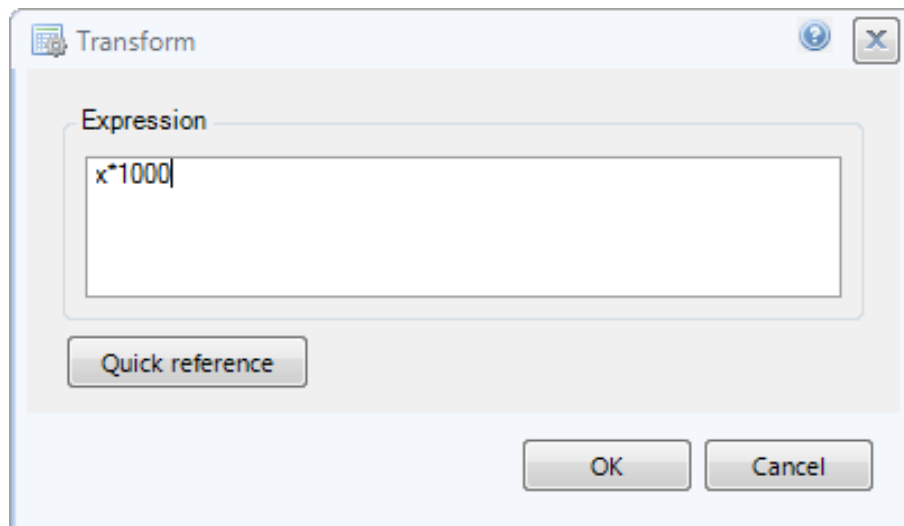


Figure 4.31: Transform Processor

#### 4.4.11 Processing outliers

If a **Time series** contains outliers, they can be detected and removed with the **outliers** processor which is suitable for non-normally distributed data sets. The processor will indicate if any outliers were detected using the Walsh method. The number and type of outliers found will be shown in the **Outlier test result** box. If any outliers are found, it is possible to delete them using the **All**, **Larger**, or **Smaller** radio buttons.

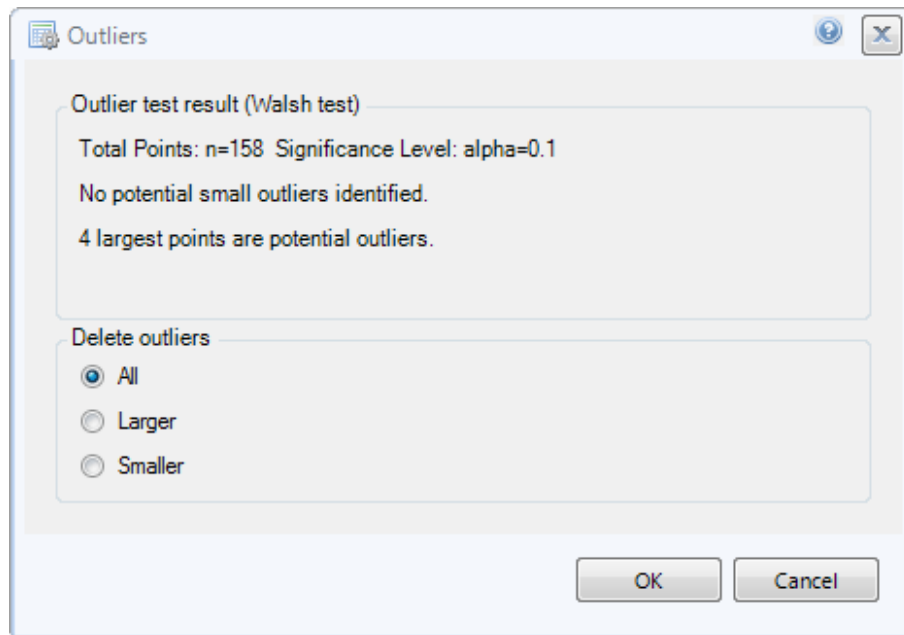


Figure 4.32: Outlier Processor

#### 4.4.12 Processing detection limits

If a **Time series** contains data with detection limits, they can be processed with the **detection limits** processor. The processor will indicate if any detection limits were found. Detection limits are identified by the presence of a less-than or greater-than symbol in the imported data (e.g. " $< 0.05$ " is a lower detection limit).

The **Replacement method** box provides several options for processing the detection limits. Changes detection limits may alter actual statistics and consulting statistician is recommended before making changes. The **Replace** option replaces all detection limits with the specified value. The **Percent** option multiplies all detection limits by the specified percentage. The **Keep same (DL)** option does not modify the detection limits. The **Halve (DL/2)** divides each detection limit in half.

Once the detection limits are processed they will be treated as normal data.

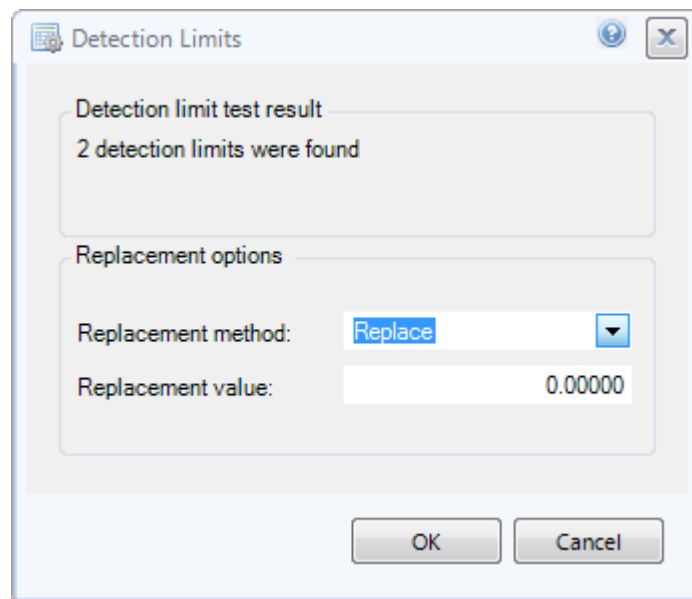


Figure 4.33: Detection Limit Processor

#### 4.4.13 Processing conflicting sample results

If a **Time series** contains conflicting sample results (results with an identical sample date), they can be processed with the **conflict** processor. The processor will indicate if any conflicting sample results were found. The sample results can be processed using one of several modes (sum, mean, median, random). The random option takes a random sample result of each conflicting group and discards any other results.

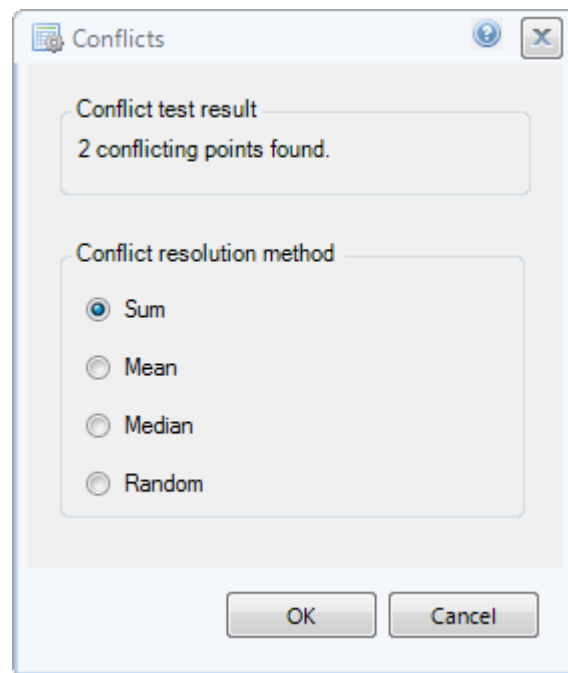


Figure 4.34: Conflict Processor

## 4.5 Statistical Analysis

The tool calculates basic statistics for the active time series. These statistics include mean, median, standard deviation, min, max and gaps.

### 4.5.1 Viewing basic statistics

To view basic statistics about a **Time series** such as mean, median and standard deviation, click the **Analyse -> Basic Statistics...** button. The **Summary statistics** form provides a numeric value for the Max, Min, Mean, Median, Standard deviation, Max gap and Min gap information. To display a value on the chart, check the associated checkbox.

Summary Statistics

Summary statistics

Number of points: 50

☐ Max: 20200

☐ Min: 10

☐ Mean: 8331.4

☐ Median: 8032

☐ Standard deviation: 5028.08

☐ Max gap: 366 days

☐ Min gap: 365 days

☐ LOWESS model Smoothing factor: 0.45

☐ Moving average

0 Y 0 M 730 d 0 h 0 min 0 s 0 ms

Box plots

☐ Show box plots

Box percentile: 25 / 75

Whisker percentile: 0 / 100

Grouping period:

0 Y 0 M 365 d 0 h 0 min 0 s 0 ms

OK Cancel

Figure 4.35: Summary Statistics Form

A LOWESS model or moving average line can also be displayed on the chart. The window of the moving average can be adjusted with the time span selector.

Box plots can also be displayed by checking the **Show box plots** checkbox. An overall box plot is displayed as well as box plots for the grouping period specified in the time span selector. The box percentile, whisker percentile and the grouping period can be adjusted.

#### 4.5.2 Viewing a histogram or frequency plot

To view a histogram or frequency plot of a time series, click the **Analyse -> Histogram** button. The number of bins can be adjusted using the **Intervals** numerical editor. The data can also be presented in Frequency plot or Cumulative frequency plot mode by



clicking the associated radio button.

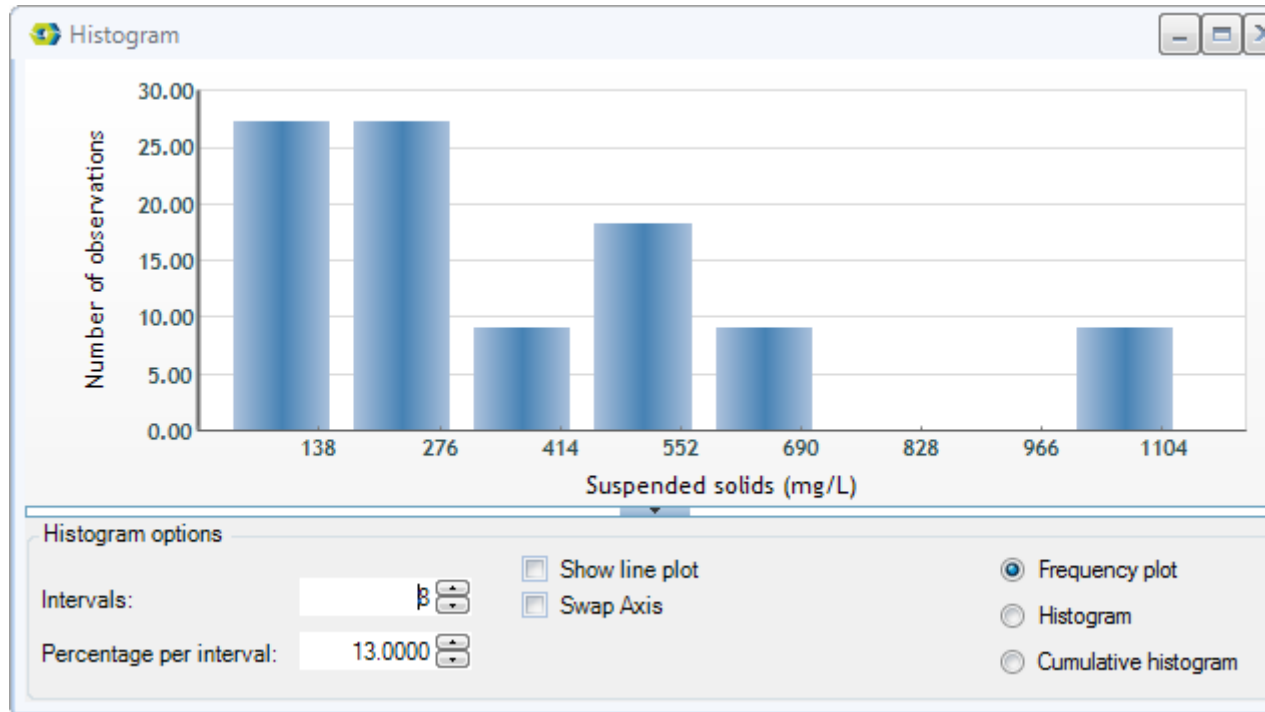


Figure 4.36: Histogram Form

## 4.6 Load Estimation

### 4.6.1 Introduction

The Loads tool has been designed and developed to estimate loads of constituents, using *Time series* of flow data and concentration data. The current version of the software contains nine different methods for making long-term load estimates, and four for achieving event-based assessments. The software comprises a comprehensive help system, including guidance in appropriate method selection. A user-friendly user interface simplifies operation and allows visualization of output estimates. The current version allows calculation of estimated load uncertainties for some selected methods.

### 4.6.2 Overview

The determination of constituent loads in streams is an essential element in environmental and ecological management. The constituent load is the mass or weight of a constituent which passes a cross-section of the stream during a specific amount of time. Load is expressed as mass per unit time (often tonnes per year). The load is de-

terminated from constituent concentration and discharge. Discharge has units of volume, usually cubic metres or megalitres.

Flow rate is usually measured on a routine basis by using flow meters, or measuring water height and applying a previously established rating curve to convert height into flow. Generally flow measurements are relatively inexpensive to make and are available on an almost continuous basis. Measurement of flow is a well-established science, and many techniques are available for different type of water flows.

In contrast to flow measurements, constituent concentration measurements are expensive: they can range in cost from a few dollars to more than a thousand dollars per sample, depending on the parameter or parameters being measured. Obtaining concentration measurements usually involves taking water samples to a laboratory for chemical analysis. Stringent quality control and quality assurance are required to obtain reliable concentration values. Because concentration measurements are much more expensive than flow measurements, it is almost always true that chemical observations are available less frequently than flow observations.

There are many methods of sampling concentrations and for determining the relationship between flow and concentrations. As a consequence there are many load-calculation techniques to account for the range of sampling methods, flow versus concentration relationships and the catchment types. Due to the relatively sparse and variable sampling frequencies of concentration data, a number of special load estimation techniques have been developed to minimise estimation errors. These techniques have been developed based on some assumptions about the behaviour of in-stream pollutant concentrations during the times when water quality was not sampled. Four main types of estimation techniques are available:

- Averaging techniques
  - Ratio method
  - Regression
  - Catchment models. There are many methods of sampling concentrations and for determining the relationship between flow and concentrations.
1. **Averaging techniques:** Assumptions are made about how concentrations vary in the time between samples. Averaging approaches use some form of average in the calculation of the loads. The simplest approach involves multiplying the average concentration for a period of time by the mean daily flow for each day in the time period, to obtain a succession of estimated daily (unit) loads. Commonly, concentration values for un-sampled dates are derived by interpolation of concentration data. Typical interpolation techniques merely linearly interpolate between concentrations. These techniques require an assumption that concentrations from individual samples represent the average daily concentration for the sampled day. The average daily concentration on un-sampled days is subsequently determined through linear interpolation between sampled concentrations.
  2. **Ratio Techniques:** Statistics derived from the available concentration samples and flow time-series are used to estimate loads over longer time spans. In this method, the daily load is calculated as the product of concentration and flow on days on
-

which samples are taken, and the mean of these loads is calculated. The mean daily load is then adjusted by multiplying it by a flow ratio, which is derived by dividing the average flow for the whole year by the average flow for the days on which chemical samples were taken. The adjusted mean daily load is multiplied by 365 to obtain the annual load. A bias correction factor can be included in the calculation, to compensate for the effects of correlation between discharge and load. Several different ratio techniques are available.

3. **Regression or Rating Curve Techniques:** A relationship is assumed to exist between flow and concentration for a particular time period, say daily, and the concentrations during unsampled periods are inferred from the flow data. These techniques can be extended to include relationships with other variables such as lagged concentrations, lagged flows, seasons of a year and long-term trend. These techniques can only be used where a relationship between variables is established and that relationship can reasonably be expected to hold in unsampled periods.
4. **Catchment Models:** Complex mathematical formulas are used to estimate flow rates as well as constituent concentrations. A number of mathematical models are available and the accuracy of estimates is heavily dependent on the type of model, input data, and the parameters the models use. Some of the models use EMC (Event Mean Concentration) and DWC (Dry Weather Concentration) for estimating annual loads. Generally complex mathematical models are used to estimate loads from urban and intensive agricultural areas.

There are a large number of loads estimation techniques available under the above-mentioned categories. This version of the Loads software contains the following methods for long-term loads calculations.

#### ***Averaging techniques***

- Flow x Concentration
- Average Load
- Flow weighted concentration
- Linear interpolation of concentration
- Flow stratified sampling.

#### ***Ratio method***

- Beale ratio

#### ***Regression***

- Concentration power curve
- USGS seven parameter method

Formulas and the descriptions of each of these methods are available in the Help system.

---

The Loads tools software has been designed to estimate loads on a long-term basis or for an event.

#### ***Event loads estimation***

Especially for particulate pollutants of non-point origin, the flux varies drastically over time, with fluxes during storm runoff events often several orders of magnitude greater than those during low flow periods. It is common for 80 to 90% or more of the annual load to be delivered during events that in total last only 10% of the year. Clearly it is critical to sample during these periods, if an accurate load estimate is to be obtained.

### **4.6.3 Selecting a flow and concentration Time series for load analysis**

To select a flow and concentration ***Time series*** for load analysis, first upload data to the WQA. Please refer to the data import section in the data module for instructions on how to do this.

In the projects area, select and click on the ***Time series*** of flow data. Data will then be visualized in the plot panel, for data processing if required. Please remember to save the data when data processing is complete. The corrected data will be saved in a new time series.

Then select the concentration data series. This will also be displayed in the chart for processing if required.

Names of these two data sets are displayed in the left bottom panel (Layers). Data layers can be added or removed here, and the appearance of the chart can be modified. You can also change the appearance of the chart using the button next to the name of the data series.

Once happy with the data click ***Analyse*** -> ***Loads...*** button to open the Loads tool.

---

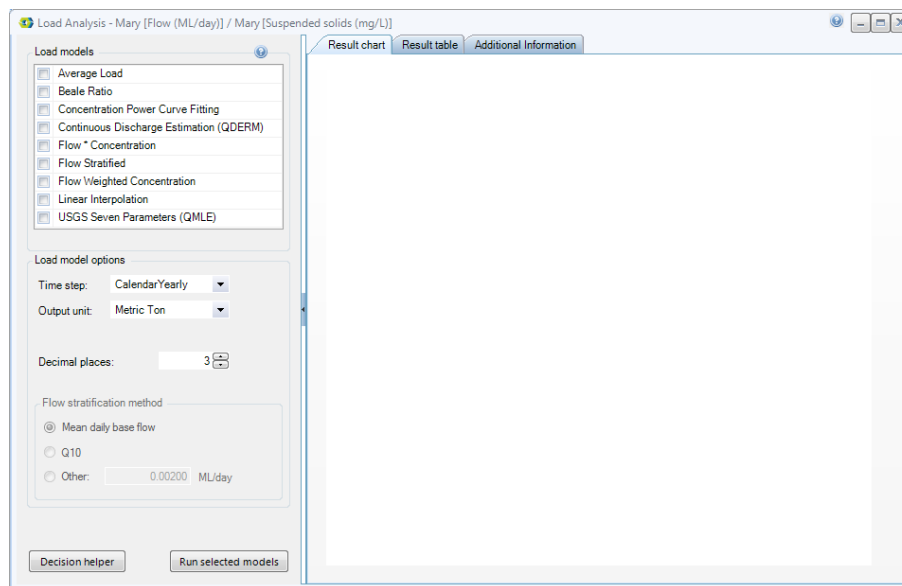


Figure 4.37: Loads Tool

#### 4.6.4 Selecting an analysis period

Constituent loads can be estimated for different time scales depending on requirements. The time steps include annual, monthly, weekly, daily or total load for the entire time duration between start and end date. Select the required time steps using the **Time Step** option available under **Load model options**.

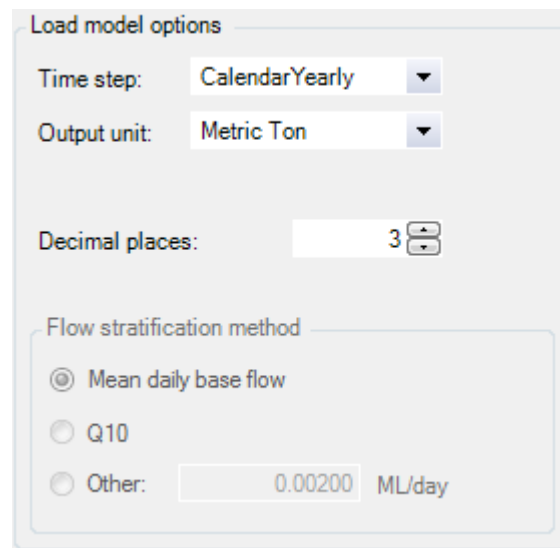


Figure 4.38: Load Time Step

Estimate loads for a specific time period by selecting the time period on the plot and use the mouse to click and drag. Click on the Load button to open the load module.

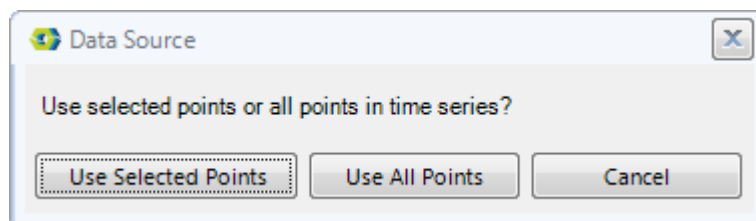


Figure 4.39: Load Analysis Range

#### 4.6.5 Manually selecting load methods

Nine different load estimation techniques are available in this version of the software for long-term basis load-estimations. You can select of all these methods, a few or one preferred method/s for estimating loads.

Click on the tick box for the methods that you want to select. When the selection is completed click on the **Run selected models** button to perform calculations.

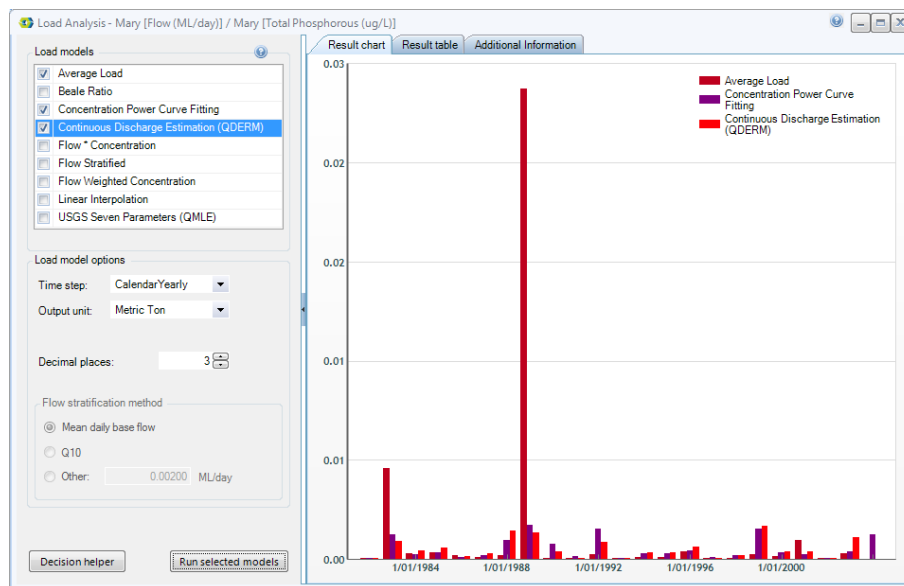


Figure 4.40: Load Results

If you have selected the **Flow Stratified Sampling** method, it is necessary to select a flow separation method to separate high flow from the base flow. There are three different methods available in the **Low Flow** panel.

**Q10:** Use up to 10<sup>th</sup> percentile of the flow rates as the base flow.

**MD base flow:** Use the mean daily base flow to separate the high flow and base flow. The base flow is calculated using the Lyn-Hollick digital filter as described in River Analysis Package ([www.ewatercra.com.au/toolkit/RAP](http://www.ewatercra.com.au/toolkit/RAP)).

**Value:** In this option, you can select a flow rate manually to separate high flow and base flow.

Selecting a better load estimation method for a given dataset is not a trivial exercise. Neither the input data variability nor the uncertainty measures, provide a reliable means of choosing between methods. Consequently, evaluation of load estimation methods must rely on comparative studies in which several methods are used to calculate loads from the same data, with results compared to the "true" load, which is independently known.

The best method can often be a matter of choice, since different methods incorporate different types of errors in estimations.

#### 4.6.6 Using the decision helper to select load methods

The **Decision Helper** can assist with selection of a load estimation methodology.

Number of samples, type of flows, and constituent characteristics such as strength, type, and consistency of the flow and concentration relationship can all play an important role

in estimator performances.

WQA uses a decision tree to assist with selection of appropriate load-estimation techniques (Appendix 2) based on past experience and reported results from many studies. The decision tree drives a decision helper which can guide selection of load-estimation methods for a particular location and dataset. Suitability of a choice is not guaranteed.

Click on the **Decision helper** button to activate this process. You need to have a fair idea of the input data; click on appropriate box/s in the Decision helper to describe the input dataset.



The image shows a 'Method Decision Helper' window with a vertical flowchart. It consists of five rounded rectangular boxes connected by downward-pointing blue arrows. Each box contains a title and a list of radio button options. The first box is 'Type of estimate' with 'Events' and 'Long-term' (selected). The second is 'Catchment type' with 'Urban' and 'Rural' (selected). The third is 'Rural catchments' with 'Intensive agriculture' and 'Other land uses' (selected), which includes a sub-note: 'Forest, pasture, broadacre agriculture etc.'. The fourth is 'Time scale of water body response' with four options: 'Steady state low flow', 'Steady state high flow (stratified)', 'Long term monthly/annual' (selected), and 'Short term dynamic hours/days/months'. The fifth is 'Long term monthly or annual' with 'High sample number' and 'Low sample number'.

Method Decision Helper

Type of estimate

- ☐ Events
- ☒ Long-term

Catchment type

- ☐ Urban
- ☒ Rural

Rural catchments

- ☐ Intensive agriculture
- ☒ Other land uses  
Forest, pasture, broadacre agriculture etc.

Time scale of water body response

- ☐ Steady state low flow
- ☐ Steady state high flow (stratified)
- ☒ Long term monthly/annual
- ☐ Short term dynamic hours/days/months

Long term monthly or annual

- ☐ High sample number
- ☐ Low sample number

Figure 4.41: Decision Helper

The system will select an appropriate method and you have the option to add more methods or change the method.

#### 4.6.7 Viewing load estimates

Estimated loads can be viewed using the tabs in the loads analysis panel. There are three major tabs in this panel. The left-hand tab shows you the histograms of calculated loads.

#### 4.6.8 Viewing calculated numbers

The next tab presents the calculated loads in a table. Loads have been estimated for the "USGS seven parameter method" and "Power Curve" methods for the entire period. In this approach, missing concentration values have been derived from the flow and concentration relationship obtained using the existing data.

For time steps that do not have at least three data points, the load cannot be calculated and blanks will appear in the output table.

Some methods of estimating loads have the additional desirable feature of providing a measure of the uncertainty of the load estimate. However, the uncertainty estimates for different load calculation methods cannot be directly compared, because they reflect different kinds of "error". It would be logical to confine the notion of error to the difference between estimated loads and the actual (but unknown) loads - a difference which is due only to sampling and analytical error. For these reasons, sometimes uncertainty measures do not provide a reliable way of choosing between methods.

This version of the Loads tool calculates uncertainty for only two methods: the simple integrated method of Flow \* Concentration; and the flow-weighted concentration method. More details of the uncertainty calculation are given in Appendix 4.

The estimated uncertainty values are given in the column next to the estimated loads.

##### ***Save outputs***

Click on "Save" on the menu bar to export the estimated loads output table to a spreadsheet file.

The contents of the summary table can be copied and pasted to MS word or other software, to save the results or to get a printout.

#### 4.6.9 Viewing additional information

The third tab contains information relating to the estimates. Scroll the bar vertically and horizontally if necessary to view all records. Information on data used and regression parameters is available here. Where regression methods have been used, the estimated R-squared provides a sense of estimation accuracy.

---

#### 4.6.10 Saving load estimates

Especially for particulate pollutants of non-point origin, the flux varies drastically over time, with fluxes during storm runoff events often several orders of magnitude greater than those during low flow periods. It is common for 80 to 90% or more of the annual load to be delivered during events that in total last only 10% of the year. Clearly it is critical to sample during these periods, if an accurate load estimate is to be obtained.

Load estimation for event-based data relies on a subset of the load-calculation techniques used for long-term estimates. Event Mean Concentration (EMC) values are also estimated. Details of these estimation methods are available in Appendix 3.

The techniques used for event load calculations are as follows:

1. Simple integration methods (Flow x Concentration)
2. Linear interpolation of concentration data
3. Concentration power curve
4. The Beale ratio estimator

#### 4.6.11 Input data preparation for events

Generally, an event will begin when water flow exceeds the base flow, and the event will come to an end when the flow rate is reduced to the base flow level. To calculate the load for the entire event, consider the event's exact time duration. It is typically difficult to sample concentrations at the precise times an event starts and finishes, and at the time of peak concentration for the storm runoff event. Therefore some assumptions and pre-data preparation are required to estimate loads using the linear interpolation method. You **must** include concentration values for the beginning and end of the event to avoid problems.

In regards to your event data, there are three options available in the drop down menu. These options are: Do not change, Apply, and Remove.

**Do not change:** Do not change concentration *Time series* data. Concentration *Time series* represent true values and should not be changed. If there are no concentration values before the "start" or the after the "end", then some methods will underestimate loads.

**Apply:** A concentration value is added at the "start" and "end" of the event to constrain or tie down the event load.

Tie down at the start of the event =  $0.5 \times$  minimum observed concentration in the dataset.

Tie down at the end of the event = minimum observed concentration in the dataset.

**Remove:** Use this option where you have manually added tie-down values to the input concentration time series. Remove these values if you are using the Beale Ratio or Power Curve methods, but keep them when applying other methods.

---

#### 4.6.12 Method selection for events

Select "Event" as the time step in the load output panel.

Alternatively you can use the Decision helper to select an event-load estimation method.

Four basic event-load estimation methods are available in this version of the Loads tool.

You can select or deselect the method using the tick box available next to the name of the method. Please refer to "Help" for details of these methods.

Select appropriate tie down option. (refer to "Input data preparation" section above).

##### Event Mean Concentration (EMC) values

A summary of event load estimates is presented in the additional information tab. It shows the start and end date of the **Time series** flow data, location of input files, units of loads, and the total flow volume.

Result chart   Result table <b>Additional Information</b>	
Load unit	Metric Ton
Total flow	3514813012.8 m <sup>3</sup>
Start date	23/01/2005 1:00:00 AM
End date	6/02/2005 2:00:00 AM
Time span	14 days and 1 hour
Project	Example Project
Set	Event
Sample location	Example Sample Location
Flow series	Event [Flow (m <sup>3</sup> /s)] Version 1
Concentration series	Event [Suspended solids (mg/L)] Version 1
Power curve fitting (a)	18.028
Power curve fitting (b)	0.27
Power curve fitting (r <sup>2</sup> )	1
EMC Beale Ratio	74.54 mg/L
EMC Concentration Power Curve Fitting	169.939 mg/L
EMC Flow * Concentration	6.531 mg/L
EMC Linear Interpolation	639.35 mg/L




Figure 4.42: Event Mean Concentration Values

Event mean concentration (EMC) values are calculated using the following equation:

$$\text{EMC} = \text{Total Load} / \text{Total volume.}$$

#### 4.6.13 Creating a new unit for load estimate output

Any SI mass units not raised to a power are automatically included in the load output unit options. To add a new mass unit, click the **Tools** -> **Edit Indicators...** button. The **Edit Indicators and Units** form appears. Click the **Units** tab, then click the **Create New...** button. Enter the short name of the unit, for example **kg**. The Name textbox is automatically filled when the short name is recognised. The mass unit will be available when the **loads tool** is opened.

### 4.7 Trend Tool

#### 4.7.1 Introduction

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. Trend tool has been designed to assist in these processes including data checking, data preparation, visualisation, exploratory data analysis and statistical analysis. This is 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 tool contains 13 statistical tests. Statistical techniques have been extracted from the WMO/UNESCO Expert Workshop on Trend/Change Detection and on the CRC for Catchment Hydrology publication *Hydrological Recipes*. This version consists of options for pre-processing of data and exploratory analysis of water-quality trends. Details of water quality trend assessment process are given in Appendix 7.

---

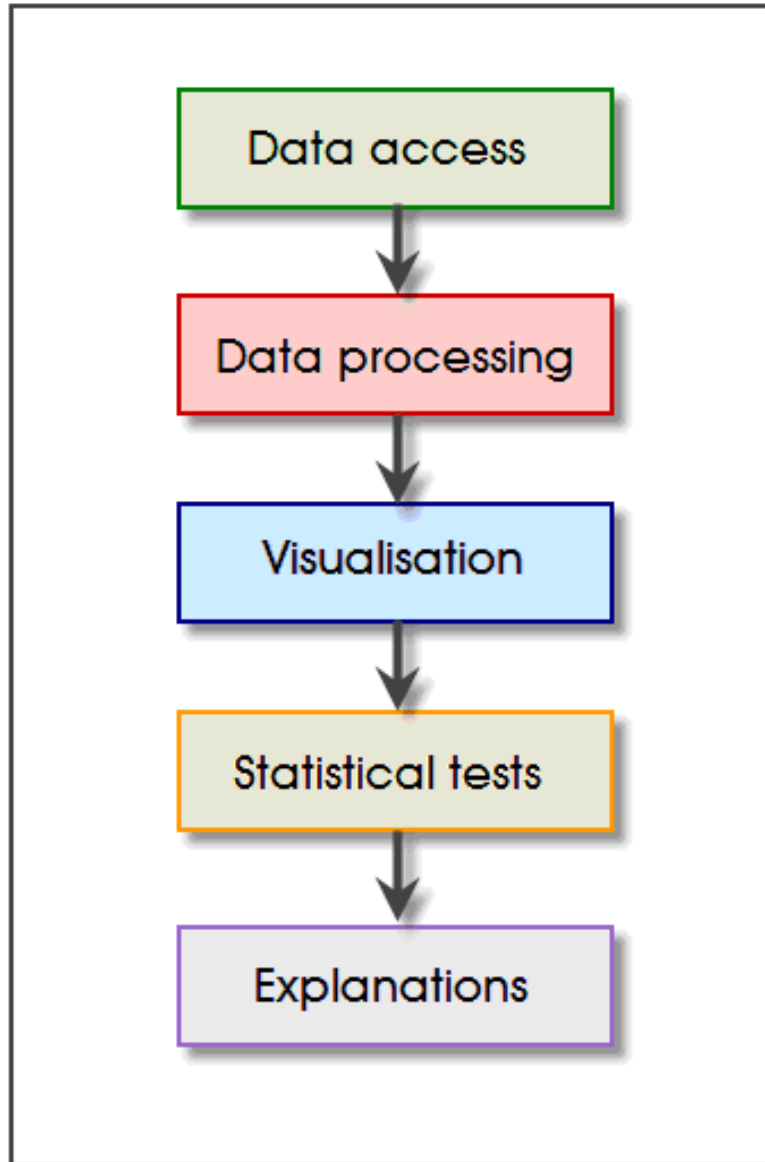


Figure 4.43: Basic logical structure of the trend tool

#### 4.7.2 Overview

TREND tool has following 13 statistical tests that can be used to test for trend, change and randomness in hydrological and other *Time series* data:

---

- Mann-Kendall (non-parametric test for trend)
- Spearman's Rho (non-parametric test for trend)
- Linear Regression (parametric test for trend)
- Distribution-Free CUSUM (non-parametric test for step jump in mean)
- Cumulative Deviation (parametric test for step jump in mean)
- Worsley Likelihood Ratio (parametric test for step jump in mean)
- Rank-Sum (non-parametric test for difference in median from two data periods)
- Student's t (parametric test for difference in mean from two data periods)
- Median Crossing (non-parametric test for randomness)
- Turning Points (non-parametric test for randomness)
- Rank Difference (non-parametric test for randomness)
- Autocorrelation (parametric test for randomness)
- Seasonal Kendall (non-parametric test for trend)

#### 4.7.3 Features

- Data access, processing and visualisations
- Supports various **Time series** data input formats
- Exploratory data analysis
- Allows easy statistical testing using different tests
- Provides simple statement of test result
- Displays test statistic and critical values for various statistical significance levels
- Performs re-sampling analysis to determine critical test statistic values
- Allows easy retrieval of test results

#### 4.7.4 Audience

Trend tool is designed for water managers, hydrologists, environmental scientists, consultants and researchers to facilitate statistical testing for trend, change and randomness in **Time series** data.

---

### 4.7.5 Limitations and cautionary notes for users

The statistical tests in this tool are only valid if the **Time series** data is not serially correlated. Most (but not all) time-series data with time steps shorter than the annual time step are serially correlated. The Autocorrelation test (one of the tests in TREND) can be used to test if the **Time series** data is serially correlated. Users are strongly encouraged to carry out an exploratory data analysis (EDA) before performing statistical analysis. The data can be explored using the visualisation and data smoothing features available in the data management module. EDA allows much greater appreciation of the features in the data than summary statistics or statistical significance level. Outliers, detection limits and obvious errors in the data can also be detected through EDA. A well-conducted EDA may eliminate the need for a formal statistical analysis. At the very least, the user should view a **Time series** plot of the data (with a trend line fitted to the data) before stepping into statistical analysis.

Users should have a good understanding of the statistical tests and assumptions. Users should also note that a statistical test provides evidence, not proof (if a trend/change is detected, the reason for the trend/change must be investigated). Statistical significance is not the same as importance (e.g., a change may be detected, but the size of the change may be so small that it is of little importance). Complex modelling process which requires explaining the detected trend is not supported in this tool.

### 4.7.6 Data requirements

Trend tool requires a continuous **Time series** as input data. Please refer to the Appendix 7 for specific data requirements for water quality trend assessment. Data processing and appropriate aggregation is required before performing the statistical tests.

#### Using the trends tool

##### Getting started

To run **TREND**, open the WQA and upload a **Time series** data set. To select a **Time series** for trend analysis, locate the **Time series** in the Projects panel then double-click to open it.

##### Data Input

A continuous **Time series** data is required to run TREND. Please refer to data management module for importing a dataset. Further details for the process of data collection and processing are given in Appendix 7.

Before performing statistical test, data processing, aggregation and exploratory data analysis should be carried out (refer to data processing section).

### 4.7.7 Exploratory analysis

A visual assessment can be performed using some data smoothing techniques such as LOWESS or moving average techniques. Click on **Summary statistics** under **Analysis** menu. Then click on **LOWESS** and choose appropriate smoothing factor to generate data smoothing curve.

---



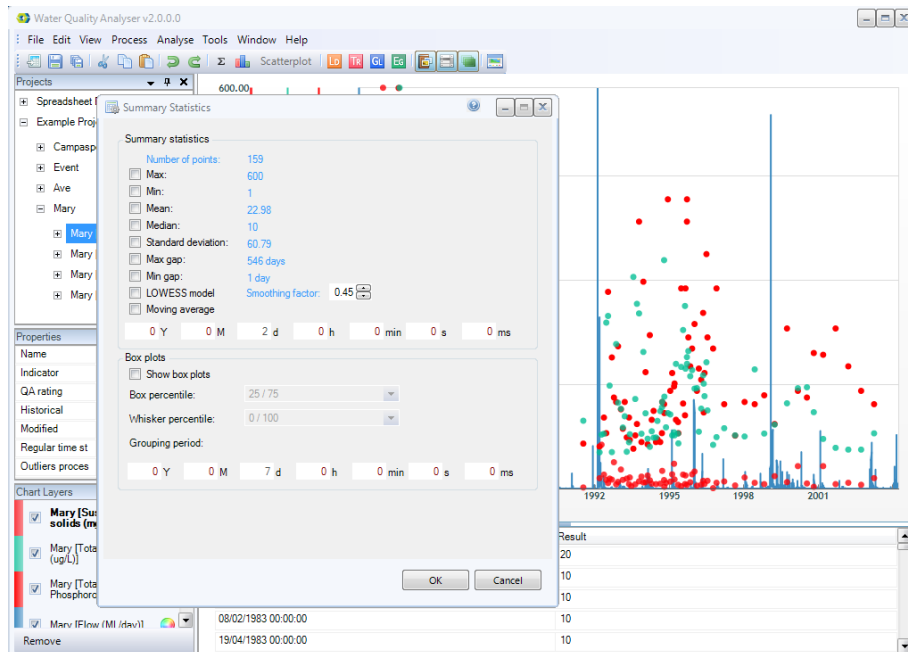


Figure 4.44: Data smoothing using LOWESS

You also you can select **Moving average** option to generate moving average curves.

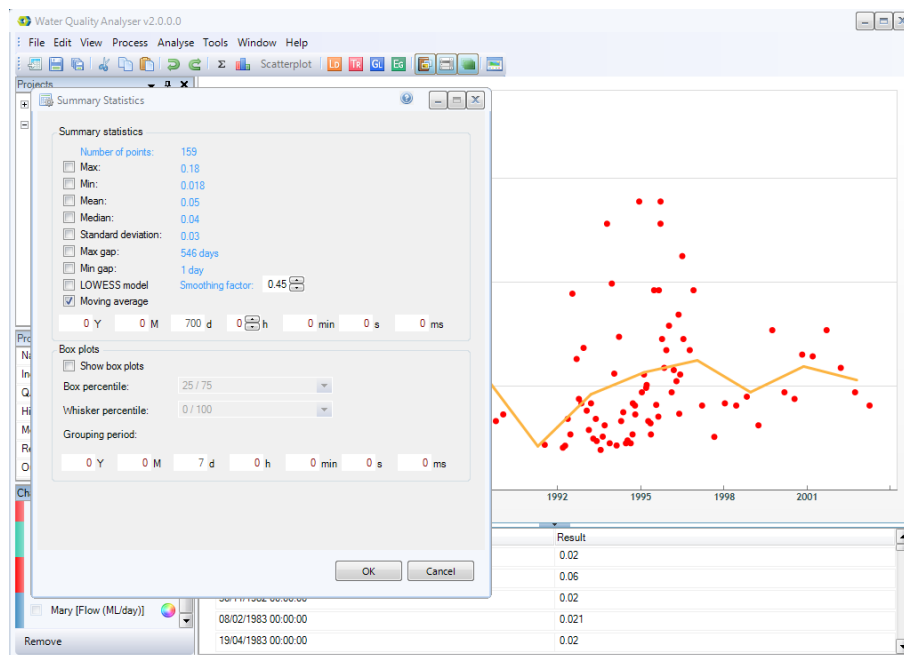


Figure 4.45: Moving averages

#### 4.7.8 Selecting an analysis period

To select an analysis period before launching the trends tool, open the **Time series** then Left-click and drag to select a range of points. Click the **Analyse -> Trends...** button. A dialog appears with options to Use Selected Points, Use All Points or Cancel.

Click the **Use Selected Points** button to perform the trend analysis on the selected points only.

Click the **Analyse -> Trends...** button to launch the trends tool.

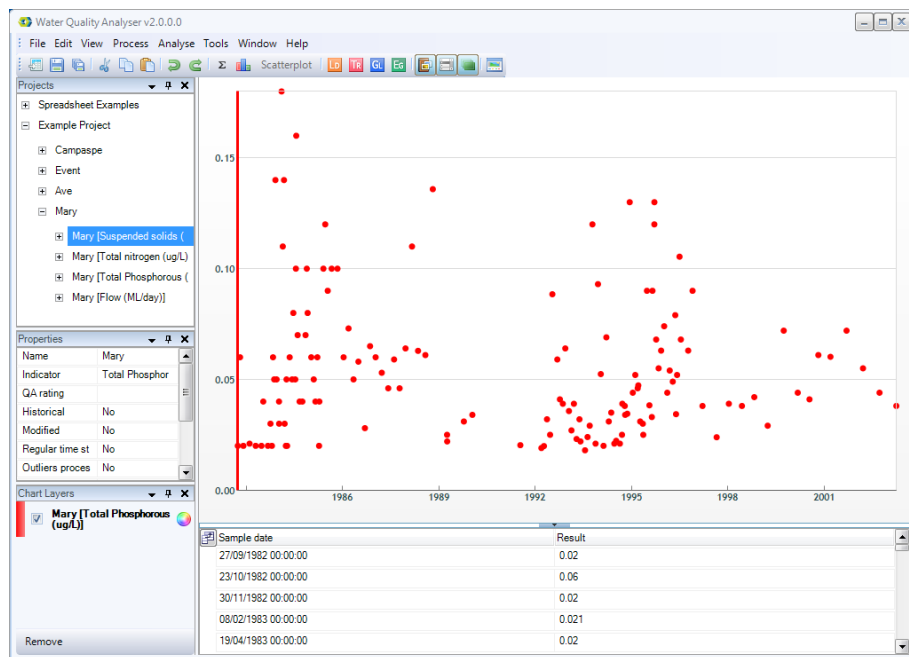


Figure 4.46: Open trend module

#### 4.7.9 Trend Tests selection

The trends tool provides trend tests categorised into four groups:

- Trend tests
- Step change tests
- Difference tests
- Randomness tests

To enable or disable a trend test, click the checkbox next to the trend test name.

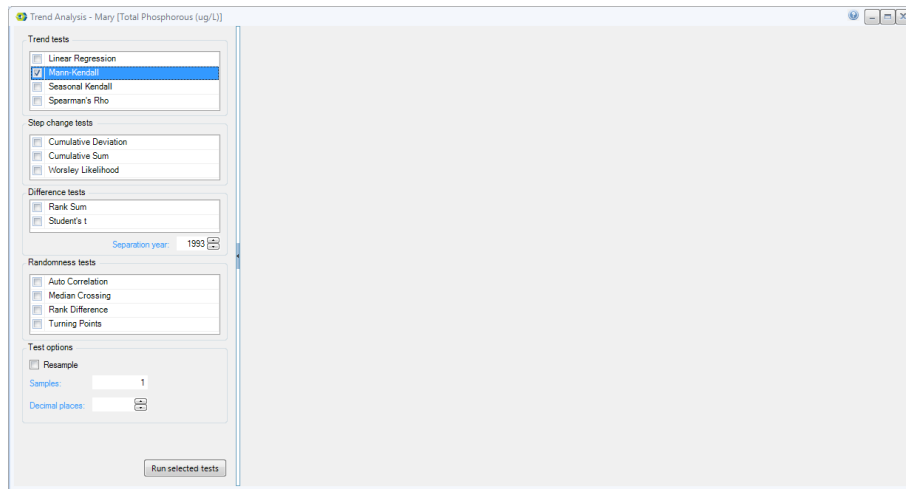


Figure 4.47: Available statistical tests

#### 4.7.10 Running the tests

Before running a trend analysis, ensure at least one trend test is enable. Click the **Run selected tests** button. The **Result chart** and **Result table** appears. When available, the result chart displays visual trend information provided by the selected trend tests.

#### 4.7.11 Results screen

##### Test graph display panel

The result from each statistical test is presented separately. Choose the "Result chart" to see the graph display.

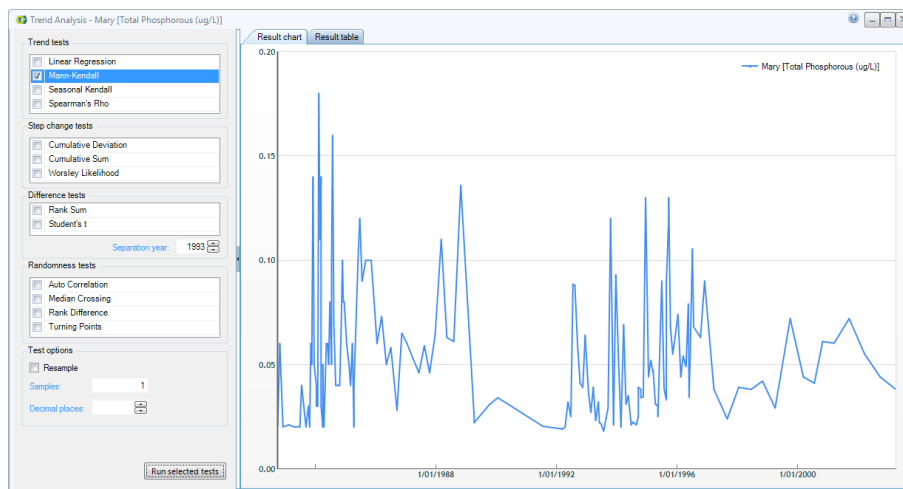


Figure 4.48: Graphical display

### Summary statistics panel

Click on **Results table** to view the statistical result of the analysis.

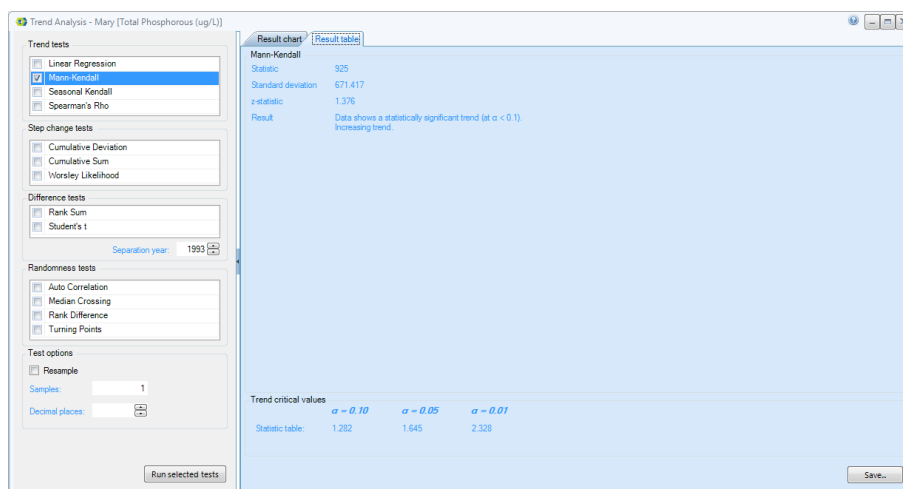


Figure 4.49: Result statistics

### Description of statistical tests

This section provides a succinct description of the basic concepts in statistical testing that are relevant to TREND. The reader can refer to Kundzewicz and Robson (2000) for a more detailed description of statistical testing for trend/change, and to standard text books on statistics for more detailed information.

### Basic concepts

*Hypothesis:* The starting point of a statistical test is to define a null hypothesis ( $H_0$ ) and an alternative hypothesis ( $H_1$ ). For example, to test for trend in a time series,  $H_0$  would be that there is no trend in the data, and  $H_1$  would be that there is an increasing or decreasing trend.

*Test statistic:* The test statistic is a means of comparing  $H_0$  and  $H_1$ . It is a numerical value calculated from the data series that is being tested. Significance level The significance level is a means of measuring whether the test statistic is very different from the (critical) values that would typically occur under  $H_0$ .

*Power and errors:* There are two possible types of errors. Type I error is when  $H_0$  is incorrectly rejected. Type II error is when  $H_0$  is accepted when  $H_1$  is true. A test with low Type II error is said to be powerful.

#### Significance level

The significance level ( $\alpha$ ) is a means of measuring whether the test statistic is very different from values that would typically occur under  $H_0$ . Specifically, the significance level is the probability of a test statistic value as extreme as, or more extreme than the observed value assuming no trend/change ( $H_0$ ).

For example, for  $\alpha = 0.05$ , the critical test statistic value is the value that would be exceeded by 5% of test statistic values obtained from randomly generated data. If the test statistic value is greater than the critical test statistic value,  $H_0$  is rejected.

The significance level is therefore the probability that a test detects a trend/change (reject  $H_0$  when none is present (Type I error)).

A possible interpretation of the significance level might be:

$\alpha > 0.1$  little evidence against  $H_0$

$0.05 < \alpha < 0.1$  possible evidence against  $H_0$

$0.01 < \alpha < 0.05$  strong evidence against  $H_0$

$\alpha < 0.01$  very strong evidence against  $H_0$

$0.05 < \hat{\alpha} < 0.1$  possible evidence against  $H_0$

$0.01 < \hat{\alpha} < 0.05$  strong evidence against  $H_0$

$\hat{\alpha} < 0.01$  very strong evidence against  $H_0$

For most traditional statistical methods, critical test statistic values for various significance levels can be looked up in statistical tables or calculated from simple formulas, provided that the test assumptions are satisfied. Where test assumptions are violated, re-sampling methods can be used to estimate the significance level of a test statistic. For detecting trend/change of any direction, the critical test statistic value at  $\alpha/2$  is used (two sided tail). For detecting trend/change in a pre-specified direction (e.g., an increasing trend), the critical test statistic value at  $\hat{\alpha}$  is used (one-sided tail). TREND gives results for a two-sided tail test.

#### Re-sampling analysis to estimate significance level

Re-sampling analysis is a robust method for estimating the significance level of a test statistic. It is particularly useful when the test assumptions are violated. In re-sampling analysis, the original *Time series* (input data) is re-sampled to provide many replicates

of *Time series* data of equal length as the original data. The *Time series* data for each replicate is obtained by randomly selecting data value from any year in the original *Time series* continuously until a *Time series* of equal length as the original data is constructed. In TREND, the data are re-sampled with replacement (bootstrapping method), i.e., a replicate series may contain more than one of some values in the original series and none of other values. The test statistic value of the original *Time series* data can then be compared with the test statistic values of the generated data (replicates) to estimate the significance level.

For example, if the test statistic value of the original data is greater than the 950th highest test statistic value from 1000 replicates,  $H_0$  is rejected at  $\alpha = 0.05$  (i.e., a trend/change is detected, with a 5% probability that this trend/change is incorrectly detected). Therefore, the critical test statistic values for significance levels of  $\alpha = 0.1$ ,  $\alpha = 0.05$  and  $\alpha = 0.01$ , are the 90th, 95th and 99th percentile values respectively of test statistic values from the generated (re-sampled) time series.

#### Parametric and non-parametric tests

Parametric tests assume that the *Time series* data and the errors (deviations from the trend) follow a particular distribution (usually normal distribution). Parametric tests are useful as they also quantify the change in the data (e.g., magnitude of change in the mean or gradient of the trend). Parametric tests are generally more powerful than non-parametric tests. Where the assumption of normally distributed data is violated, re-sampling analysis can be used to estimate the significance level or critical test statistic values for various significance levels. Non-parametric tests are generally distribution-free. They detect trend/change, but do not quantify the size of the trend/change. They are very useful because most hydrologic *Time series* data are not normally distributed.

#### 4.7.12 Description of statistical tests in TREND

TREND has 13 statistical tests that can be used to test for trend, change and randomness in hydrological and other *Time series* data:

- Mann-Kendall (non-parametric test for trend)
  - Spearman's Rho (non-parametric test for trend)
  - Linear Regression (parametric test for trend)
  - Distribution-Free CUSUM (non-parametric test for step jump in mean)
  - Cumulative Deviation (parametric test for step jump in mean)
  - Worsley Likelihood Ratio (parametric test for step jump in mean)
  - Rank-Sum (non-parametric test for difference in median from two data periods)
  - Student's t (parametric test for difference in mean from two data periods)
  - Median Crossing (non-parametric test for randomness)
  - Turning Points (non-parametric test for randomness)
  - Rank Difference (non-parametric test for randomness)
-

- Autocorrelation (parametric test for randomness)
- Seasonal Kendall (non-parametric test for trend)

The null hypothesis  $H_0$  for the tests for trend is that there is no trend in the data. The null hypothesis  $H_0$  for the tests for step jump in mean/median is that there is no step jump in the mean/median.

The null hypothesis  $H_0$  for the test for difference in means/medians is that there is no difference in the means/medians between two data periods.

The null hypothesis  $H_0$  for the tests for randomness is that the data come from a random process. TREND provides results for a two-sided tail test, therefore, where critical test statistic values are given in the following descriptions, they are for a two-tailed test. Note the statistical tests in TREND are relatively easy to understand and the user can gain a good appreciation of the tests by following the descriptions below.

#### Mann-Kendall Test

This method tests whether there is a trend in the **Time series** data. It is a non-parametric test. Then **Time series** values (  $X_1, X_2, X_3, \dots, X_n$  ) are replaced by their relative ranks (  $R_1, R_2, R_3, \dots, R_n$  ) (starting at 1 for the lowest up to n). The test statistic  $S$  is:

$$S = \sum_{i=1}^{n-1} \left[ \sum_{j=i+1}^n \text{sgn}(R_j - R_i) \right]$$

where

$$\text{sgn}(x) = 1 \text{ for } x > 0$$

$$\text{sgn}(x) = 0 \text{ for } x = 0$$

$$\text{sgn}(x) = -1 \text{ for } x < 0$$

If the null Hypothesis  $H_0$  is true, then  $S$  is approximately normally distributed with:

$$\mu = 0$$

$$\sigma = n(n-1)(2n+5)/18$$

The z-statistic is therefore (critical test statistic values for various significance level can be obtained from normal probability tables):

$$z = |S - 1| / \sigma^{0.5} \quad \text{if } S > 0$$

$$z = 0 \quad \text{if } S = 0$$

$$z = |S + 1| / \sigma^{0.5} \quad \text{if } S < 0$$

A positive value of  $S$  indicates that there is an increasing trend and vice versa.

#### Spearman's Rho Test

This is a rank-based test that determines whether the correlation between two variables is significant. In trend analysis, one variable is taken as the time itself (years) and the



other as the corresponding **Time series** data. Like the Mann-Kendall Test, the  $n$  **Time series** values are replaced by their ranks. The test statistic  $\rho_s$  is the correlation coefficient, which is obtained in the same way as the usual sample correlation coefficient, but using ranks:

$$\rho_s = S_{xy} / (S_x S_y)^{0.5}$$

where

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

$$S_y = \sum_{i=1}^n (y_i - \bar{y})^2$$

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

and  $x_i$ (time),  $y_i$ (variable of interest),  $\bar{x}$  and  $\bar{y}$  refer to the ranks ( $\bar{x}$ ,  $\bar{y}$ ,  $S_x$  and  $S_y$  have the same value in a trend analysis).

For large samples, the quantity  $\rho_s \sqrt{n-1}$  is approximately normally distributed with mean of 0 and variance of 1 (critical test statistic values for various significance levels can be obtained from normal probability tables).

#### Linear Regression Test

This is a parametric test that assumes that the data are normally distributed. It tests whether there is a linear trend by examining the relationship between time (x) and the variable of interest (y). The regression gradient is estimated by:

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

and the intercept is estimated as:

$$a = \bar{y} - b\bar{x}$$

The test statistic  $S$  is

$$S = b / \sigma$$

where

$$\sigma = \sqrt{\frac{12 \sum_{i=1}^n (y_i - a - bx_i)^2}{n(n-2)(n^2-1)}}$$

The test statistic  $S$  follows a Student-t distribution with  $n-2$  degrees of freedom under the null hypothesis (critical test statistic values for various significance levels can be obtained from Student's t statistic tables).

The linear regression test assumes that the data are normally distributed and that the errors (deviations from the trend) are independent and follows the same normal distribution with zero mean.

**Distribution Free CUSUM Test**

This method tests whether the means in two parts of a record are different (for an unknown time of change). It is a non-parametric test (distribution free). Given a **Time series** data ( $x_1, x_2, x_3, \dots, x_n$ ), the test statistic is defined as:

$$V_k = \sum_{i=1}^k \text{sgn}(x_i - x_{\text{median}}) \quad k = 1, 2, 3, \dots, n$$

where

$$\text{sgn}(x) = 1 \text{ for } x > 0$$

$$\text{sgn}(x) = 0 \text{ for } x = 0$$

$$\text{sgn}(x) = -1 \text{ for } x < 0$$

The distribution of  $V_k$  follows the Kolmogorov-Smirnov two-sample statistic ( $KS = (2/n) \max |V_k|$ ) with the critical values of  $\max |V_k|$  given by:

$$\alpha = 0.10 \quad 1.22\sqrt{n}$$

$$\alpha = 0.05 \quad 1.36\sqrt{n}$$

$$\alpha = 0.01 \quad 1.63\sqrt{n}$$

A negative value of  $V_k$  indicates that the latter part of the record has a higher mean than the earlier part and vice versa.

$$E(x_i) = \mu \quad i = 1, 2, 3, \dots, m$$

$$E(x_i) = \mu + \Delta \quad i = m+1, m+2, \dots, n$$

where  $\mu$  is the mean prior to the change and  $\Delta$  is the change in the mean. The cumulative deviations from the means are calculated as:

$$S'_0 = 0 \quad S'_k = \sum_{i=1}^k (x_i - \bar{x}) \quad k = 1, 2, 3, \dots, n$$

and the rescaled adjusted partial sums are obtained by dividing the  $S'_k$  values by the standard deviation:

---

$$S_k'' = S_k' / D_x$$

$$D_x^2 = \sum_{x=1}^n \frac{(x_i - \bar{x})^2}{n}$$

The test statistic Q is

$$Q = \max |S_k''|$$

and is calculated for each year, with the highest value indicating the change point. Critical values of  $Q/\sqrt{n}$  are given in the table below. A negative value of  $S_k'$  indicates that the latter part of the record has a higher mean than the earlier part and vice versa.

N	Q/ $\sqrt{n}$ at significance level		
	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
10	1.05	1.14	1.29
20	1.10	1.22	1.42
30	1.12	1.24	1.46
40	1.13	1.26	1.50
50	1.14	1.27	1.52
100	1.17	1.29	1.55
$\infty$	1.22	1.36	1.63

Figure 4.50: Critical values

#### Worsley Likelihood Ratio Test

This method tests whether the means in two parts of a record are different (for an unknown time of change). The test assumes that the data are normally distributed. It is similar to the Cumulative Deviation Test but weights the values of  $S_k'$  \* depending on their position in the time series.

$$Z_K' = [k(n-k)]^{-0.5} S_K'$$

$$Z_K'' = Z_K' / D_x$$

The test statistic W is:

$$W = \frac{(n-2)^{0.5} V}{(1-V^2)^{0.5}}$$

where  $V = \max|Z_K''|$

Critical values of  $W$  are given in the table below. A negative value of  $W$  indicates that the latter part of the record has a higher mean than the earlier part and vice versa.

N	W at significance level		
	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
10	3.14	3.66	4.93
15	2.97	3.36	4.32
20	2.90	3.28	4.13
25	2.89	3.23	3.94
30	2.86	3.19	3.86
35	2.88	3.21	3.87
40	2.88	3.17	3.77
45	2.86	3.18	3.79
50	2.87	3.16	3.79

Figure 4.51: Critical values of  $W$

#### Rank-Sum Test

This method tests whether the medians in two different periods are different. It is a nonparametric test. To compute the rank-sum test statistic: Rank all the data, from 1 (smallest) to  $N$  (largest). In the case of ties (equal data values), use the average of ranks;

Compute a statistic  $S$  as the sum of ranks of the observations in the smaller group (the number of observations in the smaller group is denoted as  $n$ , and the number of observations in the larger group is denoted as  $m$ ); and Compute the theoretical mean and standard deviation of  $S$  under  $H_0$  for the entire sample

$$\mu = n(N+1)/2$$

$$\sigma = [nm(N+1)/12]^{0.5}$$

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

$$Z_{rs} = (S - 0.5 - \mu)/\sigma \quad \text{if } S > \mu$$

$$Z_{rs} = 0 \quad \text{if } S = \mu$$

$$Z_{rs} = (S + 0.5 - \mu) / \sigma \quad \text{if } S < \mu$$

$Z_{rs}$  is approximately normally distributed, and the critical test statistic values for various significance levels can be obtained from normal probability tables.

#### Student's t Test

This method tests whether the means in two different periods are different. The test assumes that the data are normally distributed. The Student's t test statistic  $t$  is (critical test statistic values for various significance levels can be obtained from Student's t statistic tables):

$$t = \frac{(\bar{x} - \bar{y})}{S \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

where  $x$  and  $y$  are the means of the first and second periods respectively, and  $m$  and  $n$  are the number of observations in the first and second periods respectively, and  $S$  is the sample standard deviation (of the entire  $m$  and  $n$  observations).

#### Median Crossing Test

The  $n$  **Time series** values are replaced by 0 if  $x_i < x_{median}$  and by 1 if  $x_i > x_{median}$ . If the **Time series** data come from a random process, then  $m$  (the number of times 0 is followed by 1 or 1 is followed by 0) is approximately normally distributed with:

$$\mu = (n - 1) / 2$$

$$\sigma = (n - 1) / 4$$

The z-statistic is therefore (critical test statistic values for various significance levels can be obtained from normal probability tables):

$$z = |(m - \mu)| / \sigma^{0.5}$$

#### Turning Points Test

The  $n$  **Time series** values are assigned 1 if  $x_{i-1} < x_i > x_{i+1}$  or  $x_{i-1} > x_i < x_{i+1}$ , otherwise they are assigned as 0.

The number of times 1 appears ( $m^*$ ) is approximately normally distributed with:

$$\mu = 2(n - 2) / 3$$

$$\sigma = (16n - 29) / 90$$

The z-statistic is therefore (critical test statistic values for various significance levels can be obtained from normal probability tables):

$$z = |(m^* - \mu)|/\sigma^{0.5}$$

#### Rank Difference Test

The  $n$  **Time series** values are replaced by their relative ranks starting at 1 for the lowest up to  $n$ . The statistic  $U$  is the sum of the absolute rank differences between successive ranks:

$$U = \sum_{i=2}^n |R_i - R_{i-1}|$$

for large  $n$ ,  $U$  is normally distributed with:

$$\mu = (n+1)(n-1)/3$$

$$\sigma = (n-2)(n+1)(4n-7)/90$$

The  $z$ -statistic is therefore (critical test statistic values for various significance levels can be obtained from normal probability tables):

$$z = |(U - \mu)|/\sigma^{0.5}$$

#### Autocorrelation Test

The lag-one autocorrelation coefficient is calculated as:

$$r_1 = \frac{\sum_{i=1}^{n-1} (x_i - \bar{x})(x_{i+1} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

If the **Time series** data come from a random process, then the expected value and variance of  $r_1$  are:

$$E(r_1) = -1/n$$

$$Var(r_1) = (n^3 - 3n^2 + 4)/[n^2(n^2 - 1)]$$

The  $z$ -statistic is therefore (critical test statistic values for various significance levels can be obtained from normal probability tables):

$$z = |r_1 - E(r_1)|/Var(r_1)^{0.5}$$

#### Seasonal Kendall methods

---

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.

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.

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sgn}(x_j - x_k)$$

where

$$\text{sgn}(\theta) = \begin{cases} 1 & \text{if } \theta > 0 \\ 0 & \text{if } \theta = 0 \\ -1 & \text{if } \theta < 0 \end{cases}$$

#### 4.7.13 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.

Grayson, R.B., Argent, R.M., Nathan, R.J., McMahon, T.A. and Mein, R. (1996) Hydrological Recipes: Estimation Techniques in Australian Hydrology. Cooperative Research Centre for Catchment Hydrology, Australia, 125 pp. Trend User Guide

Kundzewicz, Z.W. and Robson, A. (Editors) (2000) Detecting Trend and Other Changes in Hydrological Data. World Climate Program - Water, WMO/UNESCO, WCDMP-45, WMO/TD 1013, Geneva, 157 pp.

Sunil Tennakoon, David Robinson and Shuangyang Shen (2009) Decision Support System for Temporal Trend Assessment of Water Quality Data. 18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009 <http://mssanz.org.au/modsim09>

#### 4.7.14 Acknowledgement

David Robinson at Environmental Protection Agency in Victoria, Australia provided technical information for upgrading the Trend Tool.

## 4.8 Dashboard

### 4.8.1 Introduction

The Dashboard provides an interactive display area that allows multiple charts to be presented on a single panel for analysis of statistical data. Users can also easily compare and contrast different charts at a glance.

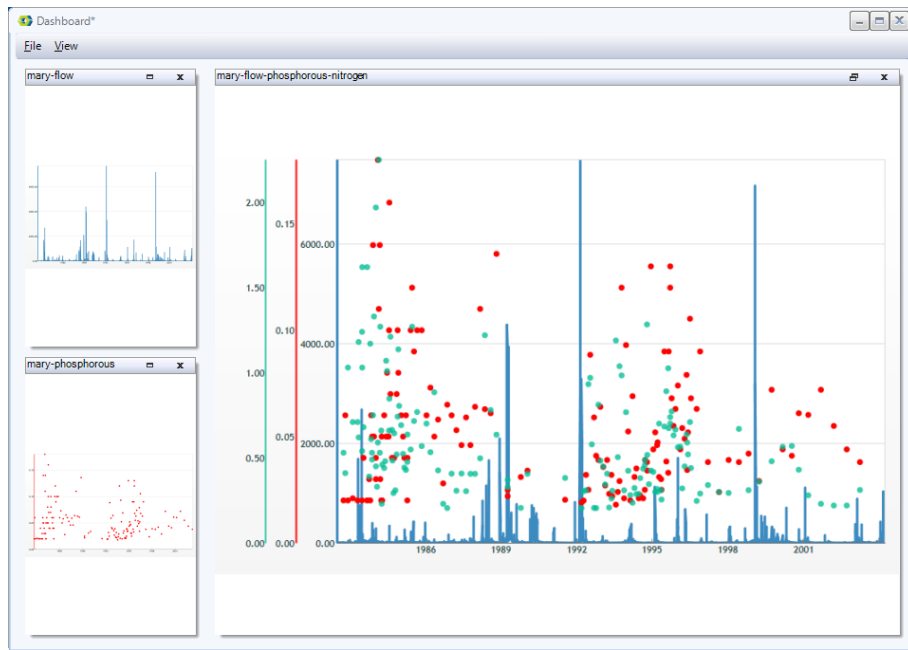


Figure 4.52: WQA Dashboard

### 4.8.2 Features

- View multiple charts generated from the tests at one glance
- Customise the view of the Dashboard
- Displayed charts can be saved as a collection of charts to a single file and can also be loaded back onto the Dashboard
- Able to save each individual chart as an image file
- Open an individual chart from an image file
- Allows the user to copy the charts from another source (e.g. Internet) and paste it onto the Dashboard and vice versa.
- Displayed charts on the Dashboard can be easily re-arranged and enlarged.



### 4.8.3 Using dashboard

The test results are presented in charts. Dashboard provides an interactive viewing area for analysing multiple charts on one panel and arranged the charts in a grid fashion.

To launch the Dashboard, simply just click on the Dashboard icon which is found on the toolbar of the WQA.

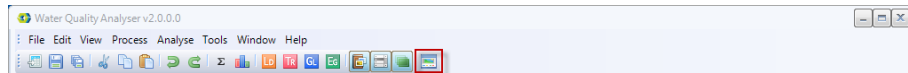


Figure 4.53: Toolbar of WQA: Dashboard Icon

### 4.8.4 Adding a chart to Dashboard

The Dashboard is initially empty when it is first launched. Charts can be added onto the Dashboard through a number of ways and will be described in detail below:

#### (i) Adding an empty tile (without a chart image) to Dashboard

To create an empty chart, click the **File** -> **Add Tile** on the menu bar. Enter a name for the tile and click the **OK** button. An empty tile will be added to the Dashboard.

Alternatively, you can right-click on any empty area on the Dashboard and select **Add Tile** to add an empty tile.

An empty tile does not have a chart as it just creates an area for a chart to be displayed. A chart can be displayed on the empty tile by importing the chart from an image file, or by pasting the image onto it.

#### (ii) Adding charts from Chart Layers Panel to Dashboard

The chart on the **Chart layers panel** can be added to the Dashboard.

Right-click on the chart on the Chart layers panel and select Send to Dashboard.

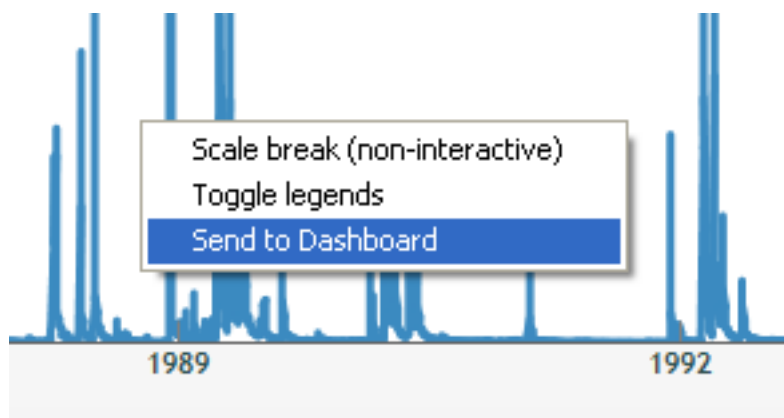


Figure 4.54: Chart layers panel: Right-click on the chart

Right-click on the chart on the **Chart layers panel** and select **Send to Dashboard**. The application will prompt you to enter a name for the chart. Enter a name and click the **OK** button. The chart will appear in the Dashboard with the specified name displayed at the top of the chart.

The chart can be saved as an image file.

**(iii) Pasting a chart image from other sources to Dashboard**

You can copy an image of a chart from a document or from an image editor and paste it onto the Dashboard for your analysis.

Copy the desired image of the chart. Focus on Dashboard and hold both **control** and **V** key to paste the image. The application will prompt you to enter a name for the new tile. Enter a name and click the **OK** button. A new tile with the chart image will be added to the Dashboard.

Alternatively, you can right-click on Dashboard and select **Paste Chart**. Enter a name and click the **OK** button. Similarly, a new tile with the copied chart image will be added to the Dashboard.

You can also paste the image onto an empty tile by right-clicking on the empty tile that you choose to paste the image in and select **Paste Image**. The copied image will be displayed in the tile.

#### 4.8.5 Removing the Chart from Dashboard

To remove any tile with/without a chart from the Dashboard, click **X** on the tile that you choose to remove.

If you would like to remove a chart from a tile, right-click on the chart and select **Delete Chart**. The chart will be removed from the tile and the tile will be emptied. This does not remove the tile from the Dashboard.

#### 4.8.6 Renaming a chart

Each chart has a name which appears at the top of the chart. The name of the chart can be changed. To rename the chart, right-click on the chart and select **Change Chart Name**. Enter a new name for the chart and click the **OK** button. The new name will be reflected at the top of the chart.

#### 4.8.7 Copying the chart and pasting onto a document

Right-clicking on a specific chart on the Dashboard will bring up a list of options that you can select. Click **Copy Chart** to copy the chart and you can paste it as an image onto any other document files or even in an image editor.

---

### 4.8.8 Copying a chart image and pasting onto Dashboard

You can copy an image from the web or other documents and paste it into the tile on the Dashboard. Right-click on the tile and select **Paste Chart** to paste the chart as an image.

Note that by holding both **control** and **V** keys paste the image onto a **new** tile.

### 4.8.9 Saving the chart image

The chart in the tile can be saved and stored onto the computer as an image file. Right-click on the chart and select **Save Chart As Image**. A **Save As** dialog box will appear and will prompt you to specify a location that you wish to save your chart. Enter a name for the image file and click the **OK** button. The image will be saved on your specified location with the specified image file format.

### 4.8.10 Saving the Dashboard

The Dashboard together with the existing collection of charts can be saved as a binary file. The binary file also stores the view of the Dashboard. To save the Dashboard in a new binary file, click **File** -> **Save As** on the menu bar and a **Save As** dialog box will appear. Enter a name for the Dashboard in the **Name** field and specify a location that you wish to save your dashboard. Click **Save** to save the Dashboard.

To save the Dashboard to the current existing binary file, either click the **File** -> **Save** on the menu bar or hold both **control** and **S** keys.

When the Dashboard is saved, a "Dashboard Saved" message box appears. Click **OK** to proceed.

When any changes are made to the Dashboard, for instance adding of new charts to Dashboard, changing the name of the chart, re-arranging the charts etc., the title bar of the Dashboard will be appended with an asterisk "\*". An asterisk indicates that changes have been made to the Dashboard. When the Dashboard is saved, the asterisk will be removed to show that the current Dashboard is updated.

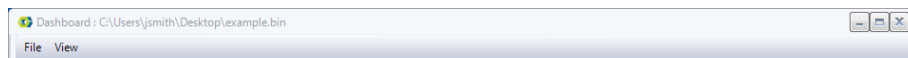


Figure 4.55: Tile bar of the Dashboard: An asterisk indicates that changes have been made to the Dashboard

### 4.8.11 Open previously saved binary file/Dashboard

To open a Dashboard that was previously saved in a binary file, click the **File** -> **Open** on the menu bar.

If there are charts that are currently displayed on the Dashboard, a "Do you like to save?" message box appears to prompt you to save the current Dashboard before loading

another binary file. Click **Yes** if you wish to save the current Dashboard and **No** if you wish to open another binary file without saving the current Dashboard.

An **Open File** dialog box will appear for you to select the binary file that you wish to load. Navigate to the saved location of the binary file, select the file and click **Open**. The Dashboard will load up with the set of charts that was saved in the binary file.

#### 4.8.12 Opening a new Dashboard

To open a new Dashboard, click **File** -> **New** on the menu bar. The application will prompt you to save the current Dashboard if the changes made to the dashboard have not been saved. Click **Yes** if you wish to save the current Dashboard and **No** if you wish to open a new Dashboard without saving the current Dashboard.

#### 4.8.13 Clearing the Dashboard

A shortest way of removing all the tiles on the Dashboard is by clicking **File** -> **Clear Tiles** on the menu bar.

#### 4.8.14 Re-arranging the tiles on the Dashboard

The tiles on the Dashboard are positioned in a grid fashion. The tiles on the Dashboard can be arranged by dragging and dropping the charts to the desired location on the Dashboard. All tiles must be in a normal state. If any of the charts are enlarged, the repositioning of charts is disabled.

#### 4.8.15 Configuring the view of the Dashboard

The view of the Dashboard can be configured to your preference. You are able to set the maximum number of large tiles can be viewed and to define the displaying position of the large tiles on the Dashboard.

The view settings of the Dashboard are found in **View** on the menu bar. Click **View** on the menu bar to change the position of the large tiles and to set the maximum number of large tiles allowed on the Dashboard.

The following diagram on the right shows that a maximum of 2 tiles are allowed to be enlarged and the large tiles are displayed on the left. The diagram on the right shows that the Dashboard has only 1 large tile and the large tile is positioned on the top of the normal-sized tiles.

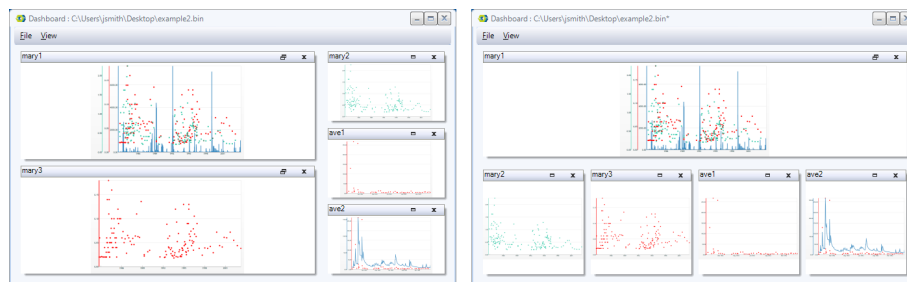


Figure 4.56: Configuring the View of Dashboard

#### 4.8.16 Closing the Dashboard

To close the Dashboard, click **File** -> **Close** on the menu bar. If there are changes made to the current Dashboard or the current Dashboard has not been saved, the application will prompt you to save before closing. Click **Yes** if you wish to save the current Dashboard and **No** if you wish to close the Dashboard without saving the current Dashboard.

If you have clicked **Yes**, a **Save As** dialog box will appear. Enter a name for the Dashboard in the **Name** field and specify a location that you wish to save your dashboard. Click **Save** to save the Dashboard. A "Dashboard Saved" message will appear to indicate that the Dashboard has successfully saved. Click the **OK** button to proceed to closing the Dashboard.

## 4.9 Guideline Tool

### 4.9.1 Introduction

The *Guidelines Tool* is software for the calculation, management and application of water quality guidelines using the procedures specified in ANZECC & ARMCANZ 2000 and the Queensland Water Quality Guidelines (QWQG 2006). This software includes a database containing default guideline values contained from the ANZECC & ARM-CANZ guidelines, and some state-specific values. Users can search the database for guidelines and results can be used for testing against time-series data collected from a water body. The tool can calculate locally relevant water quality guidelines using both biological effects data and reference data, as recommended by ANZECC & ARMCANZ 2000. The software can also be used to set water quality targets using a proportional improvement approach. The user interface was developed for easy operation and to let users visualise input data and estimated outputs.

### 4.9.2 Overview

#### Water quality guidelines

A water quality guideline is a recommended numerical concentration level or a narrative statement (e.g. visual appearance of a water body) that will support and maintain the designated use (environmental values) of a particular water body. Water quality guidelines are provided for chemical and physical parameters of water and sediment, as well as biological indicators. Nationally agreed processes of setting and applying water quality guidelines have been developed jointly by the Australian and New Zealand Environment and Conservation Council and the Agriculture and Resource Management Council of Australia and New Zealand (ANZECC & ARMCANZ 2000).

#### **The Guidelines tool**

The Water Quality Guidelines software tool accommodates the process described in ANZECC & ARMCANZ 2000 to assist various stakeholders involved in water quality testing and assessments. The statistical procedure used in this tool is based on the ANZECC & ARMCANZ 2000 method and is recommended by the Queensland Environmental Protection Agency. The current version of Water Quality Guidelines Tool contains following components:

- Water quality guideline database
- Database search facility
- Applying water quality guideline
- Derive guideline using biological effects data
- Derive guideline using reference site data
- Water quality target setting using proportional improvements
- Managing and updating database.

The software has been designed with a user-friendly interface so that minimum computer skills and training are required to use the tool.

#### **4.9.3 Water quality guideline database**

A database has been designed to provide a sensible and flexible basis for storing, searching and editing guideline values, water quality objectives and their relevant information. The database structure has been developed to provide maximum flexibility for entering, storing and editing data records.

The integrated database to this tool contains the water quality guideline values available in ANZECC & ARMCANZ 2000, and some State-specific values. You can update this default database by adding your locally specific values, or refine existing values.

#### **4.9.4 Applying water quality guidelines**

The Guidelines software tool can be used to test data from a water body against selected national or state-specific guideline values, or a user-specified guideline value. The test is performed by comparing the guideline against an appropriate-size confidence interval around the median of the test data. The size of the confidence interval is

---

approximately 95% when testing against a single upper or lower value, and 90% when testing against a guideline comprising an upper and lower value. These interval sizes have been chosen to keep the Type I error rate to 0.05.

#### **Procedure for deriving water quality guideline values**

In situations where available guideline values are not relevant or appropriate for a particular location or the issue, locally relevant values need to be derived. Procedures for deriving these guideline values have been explained in the ANZECC & ARMCANZ 2000 documents and several methods are available depending on the type of obtainable data and the assessment process. The preferred approaches for selecting these methods are listed in ANZECC & ARMCANZ 2000, in order of preference, as follows:

1. Use biological effects (toxicity) data
2. Use local reference data
3. Use regional reference data
4. Use available generic effect-based guideline.

#### **Use of biological effects data**

The Guidelines software tool is designed to estimate the protecting concentrations of chemicals (toxicants) such that a given percentage of species will survive. The preferred input data are those derived concentration values (preferably NOEC (No Observable Effect Concentrations)) from environmentally realistic and well conducted multiple species toxicity tests. Estimates of protecting concentrations are computed by fitting a certain distribution to the input data. A distribution called the Burr type III is used in this software as recommended by ANZECC & ARMCANZ 2000. The Burr III distribution is a very flexible 3-parameter distribution, which can provide good approximations to many commonly used distributions such as the log-normal, log-triangular and Weibull (see Appendix 1). The software computes the concentration value corresponding to a certain level of percentile of the species to be protected if the concentration of the chemical is less than the estimated concentration.

#### **Use of reference data (local or regional)**

Often toxicity data are unavailable, and there are a number of practical difficulties with the biological effects data approach. In such cases, water quality data collected from reference sites can be used to derive guideline values. A reference site is a site whose condition is considered to be a suitable baseline or benchmark for assessment and management of sites in similar water bodies (Queensland Water Quality Guidelines, 2006). The criteria for selecting a reference site and the required number of samples for deriving guidelines are available in ANZECC & ARMCANZ 2000. Generally, guideline values are derived for physical and chemical parameters for ecosystems protection in terms of the 80th and/or 20th percentile values estimated from the data collected from an appropriate reference site. This choice is arbitrary and some state or locally specific percentile values may be used. The reliability of the derived guideline value will depend on the quality and quantity of collected data and, like most statistical measures, errors in percentile estimates will reduce with increasing sample size. Based on these analyses, results presented in QWQG 2006 recommended that estimates of 20th or 80th percentiles at a reference site should be based on a minimum of 18 samples collected

---

over at least 12 months. For 50th percentiles a smaller minimum number of samples ( $\sim 10$ -12) would be adequate but in most situations it would be necessary to collect sufficient data for the 20th and 80th percentiles.

#### **Setting water quality targets**

Generally water quality targets are defined using a specific time scale considering the current condition of a water body and the desired water quality objectives. Defining water quality objectives for a water body is a lengthy and complex process which requires identification of environmental values and derivation of corresponding guideline values. There is also a serious economic and social concern in setting water quality targets in many places. Due to complexity in this process, water quality targets for an impaired water body can be set using a certain level of proportional improvements compared to the current condition. An example of a target would be: a 50% reduction in total suspended solids at the river mouth over 10 years. These targets could be any level of percentages agreed by the stakeholders. The Setting Water Quality Targets tool helps you to set targets using the proportional improvement method. You must input data from the selected site, and then the software estimates the current median and relevant confidence values. You can then enter the required percentage or the expected improvement to estimate the target value and its confidence intervals.

#### **4.9.5 Features**

The Guidelines tool currently has the following features:

- A searchable database of guideline values from ANZECC & ARMCANZ 2000 and some State specific water quality guidelines and water quality objectives.
- A function for calculating guidelines using biological effects data.
- A function for calculating percentile based guidelines in accordance with procedures and associated confidence intervals from ANZECC & ARMCANZ 2000.
- A function for testing waters against guidelines and targets.
- The capacity to store regionally specific guidelines and targets (and associated metadata) as calculated by the Guidelines tool or developed elsewhere.
- The capacity to set water quality targets using proportional improvements.

#### **4.9.6 Audience**

The Guidelines tool is intended for technical and operational staff of regional bodies, government agencies, environmental and engineering consultants and water managers involved in the setting and testing of water quality.

#### **4.9.7 Related documents**

- ANZECC & ARMCANZ 2000 Monitoring & Reporting Guidelines
-



- ANZECC & ARMCANZ 2000 Water Quality Guidelines
- NHMRC 2005 Recreational Water Quality Guidelines
- Queensland Water Quality Guidelines

#### 4.9.8 Data requirements

##### Input Data

The Guidelines tool accepts *Time series* data. It should have two or more columns, the first being the date and the subsequent columns the value of the measurements. Measurements with an exact common time step (e.g. daily, monthly, quarterly, etc.) will provide the best visualisations within Guidelines.

To derive biological effects guidelines or trigger values, toxicological data derived from laboratory experiments are required (refer to Appendix 1). According to ANZECC 2000, a minimum of two years of contiguous monthly data at the reference site is required to establish a valid guideline value.

#### 4.9.9 Product components

The current version of the Guideline tool has the following components:

- Water quality guideline/objectives database
- Database search facility
- Applying water quality guideline
- Derive guideline using biological effects data
- Derive guideline using reference site data
- Water Quality target setting using proportional improvements
- Managing and updating database.

#### 4.9.10 Limitations and caution notes for users

The Guidelines tool will calculate and test data against guideline values regardless of the amount and quality of the information that you enter. It's important to note that there are some recommended minimum data requirements for the calculation of a guideline value and for testing against a guideline (see ANZECC & ARMCANZ 2000 and QWQG 2006). Whilst every care has been taken in developing this tool, there may be circumstances that have not been adequately catered for. We urge you to explore the tool for your own use and report any bugs or desired enhancements to the development team as you encounter them.

---

#### 4.9.11 Using Water Quality Guideline Tool

Open the program from the **Start** menu; point the cursor to **Programs > Water Quality Tools > Water Quality Analyser**, or click on the following icon to open the **Water Quality Analyser** software:



Figure 4.57: Water Quality Analyser software icon

#### 4.9.12 Running the software

Once you upload (import) a data set using the options available in the data management module, the Guidelines tool can be navigated using the options available in the guideline module.

#### 4.9.13 Search guideline library (database)

The database in the Guidelines tool allows you to search for records that relate to a number of criteria. The following steps will get you started.

To search a guideline click the **Analysis -> Guidelines** button. The guideline search panel appears. The guideline search form appears and provides options to filter a search based on indicators, protection type, water types, and source or guideline type as well as using the keyword search.

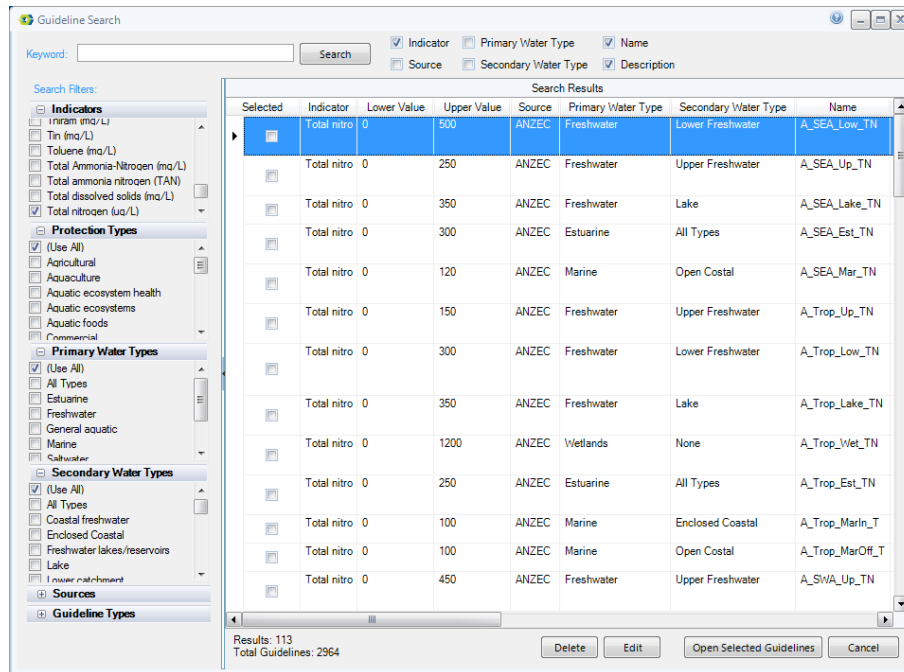


Figure 4.58: Guideline Search

Check a checkbox next to a filter value or enter a keyword to narrow down the search results. A list of relevant guidelines is shown in the **Search Results** box. If you need to test a data set which is already opened, check the **Selected** checkbox of one or more guidelines, and then click the **Open Selected Guidelines** button to open the guidelines for testing. The guidelines appear in the guidelines panel and in the chart. The result value in the guidelines panel indicates whether the active **Time series** passes or fails the guideline.

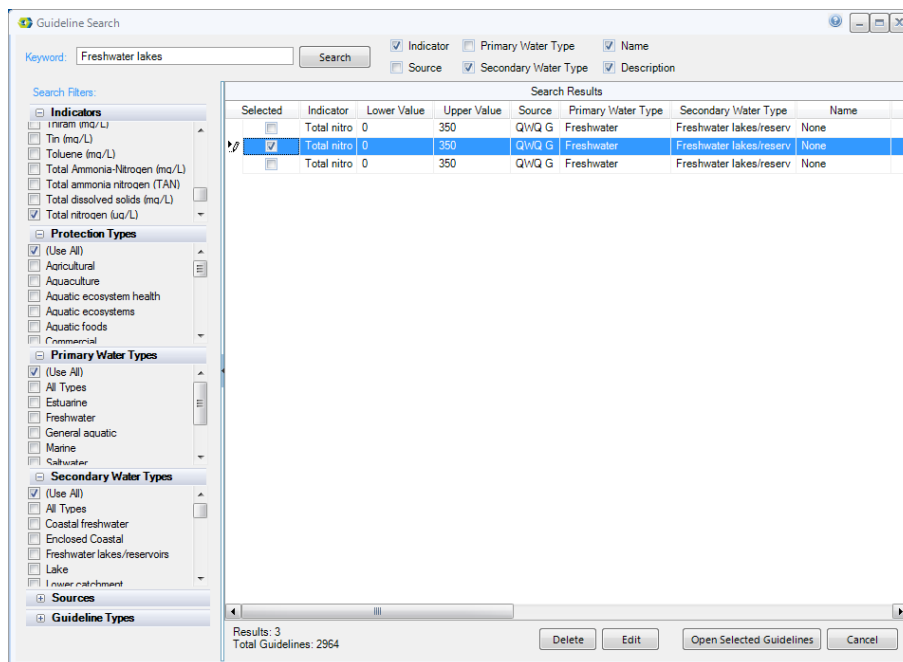


Figure 4.59: Narrow down Guideline Search result

#### 4.9.14 Testing (apply) water quality guidelines or targets

The water quality of a site can be assessed against a guideline or target. This can be achieved by comparing the median value of the water quality variable at a site over a period of time against the guideline or target. The following instructions provide the basics of how this can be achieved with the Guidelines tool.

Open the input data **File** using the data management module and make as active **Time series**

You need to select the type of guideline to be tested from the following options.

If you had selected values to test against from a search of the library, the result will be shown in the **Results** box and the screen should resemble the one below, with the data and the test value shown in the graph.

If you haven't already selected values to test against, you may now do so by either clicking **Search library** (which will take you to the **Search library** screen).

Select the guideline value you want for the test. Also you need to select the guideline type by pressing the **Edit** button on the search screen. You can select **Upper**, **Lower** or **Both** based on your requirements.

**Edit a guideline target**  
Edit an existing guideline target

General Locations **Guidelines**

C.I.: 0.00 Start Date: 22/06/2007  
 Sample Size: 0 End Date: 22/06/2007  
 Testing Type: Upper

	Percentile	Estimate	Lower C.I.	Upper C.I.
Upper Guideline:	80	10	0	0
Median:		0	0	0
Lower Guideline:	20	0	0	0

Save Cancel

Figure 4.60: Edit Guideline Target: Guideline tab

#### 4.9.15 Using a user defined value for testing

If you want to test the data set with your own guideline value, click on **Test Guideline** option in the menu. The following panel will appear

New Guideline Target

**Create a new guideline target**  
Create a new guideline target

General Locations Guidelines

Name: Test guideline

Source: User-defined

Indicator: Total Phosphorous (ug/L)

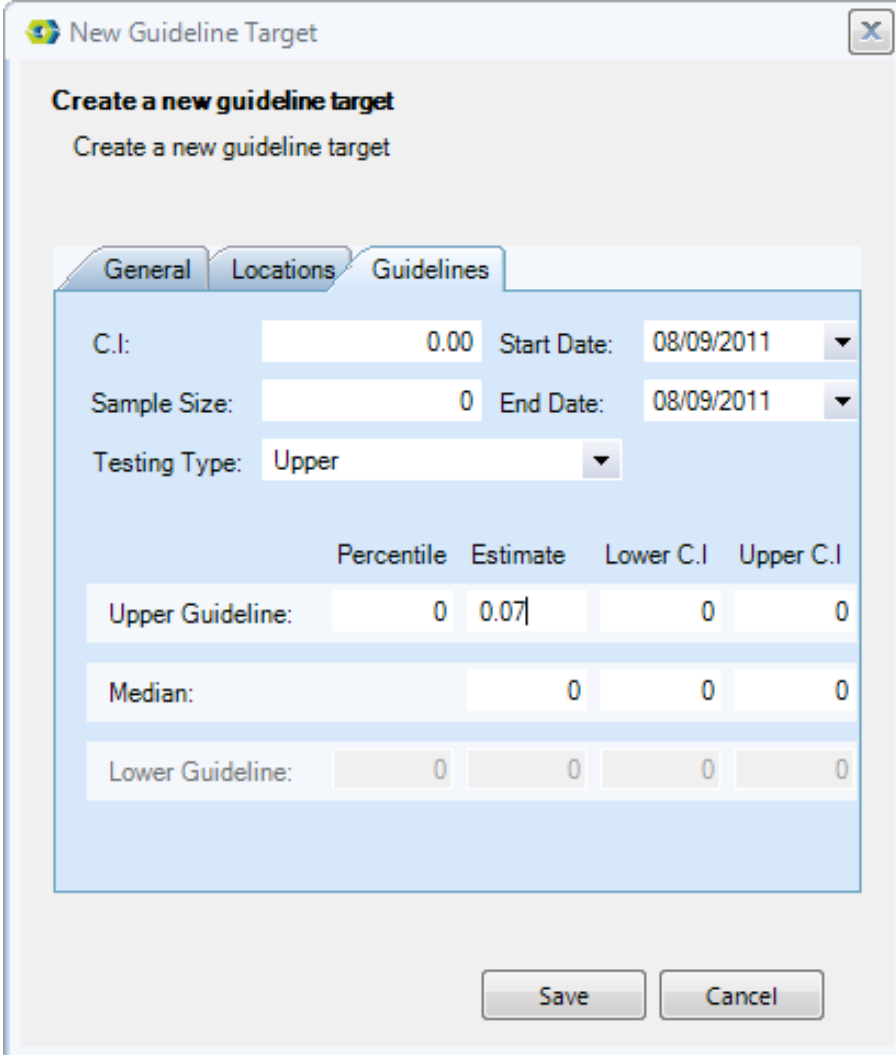
Protection type: Aquatic ecosystems

Type: Departure from reference

Save Cancel

Figure 4.61: Test with users own guideline values

You need to give a name for the guideline and select the indicator from the list. The completion of other fields in this panel is optional. The click on the **Guideline** tab.



**Create a new guideline target**  
Create a new guideline target

General Locations **Guidelines**

C.I.: 0.00 Start Date: 08/09/2011  
Sample Size: 0 End Date: 08/09/2011  
Testing Type: Upper

	Percentile	Estimate	Lower C.I.	Upper C.I.
Upper Guideline:	0	0.07	0	0
Median:		0	0	0
Lower Guideline:	0	0	0	0

Save Cancel

Figure 4.62: Test with users own guideline values

Select the test type upper or lower and enter the number in the **Estimate** box and click on **Save** button. The result will appear in the data panel. The test result will appear in the lower right-hand side of the window.

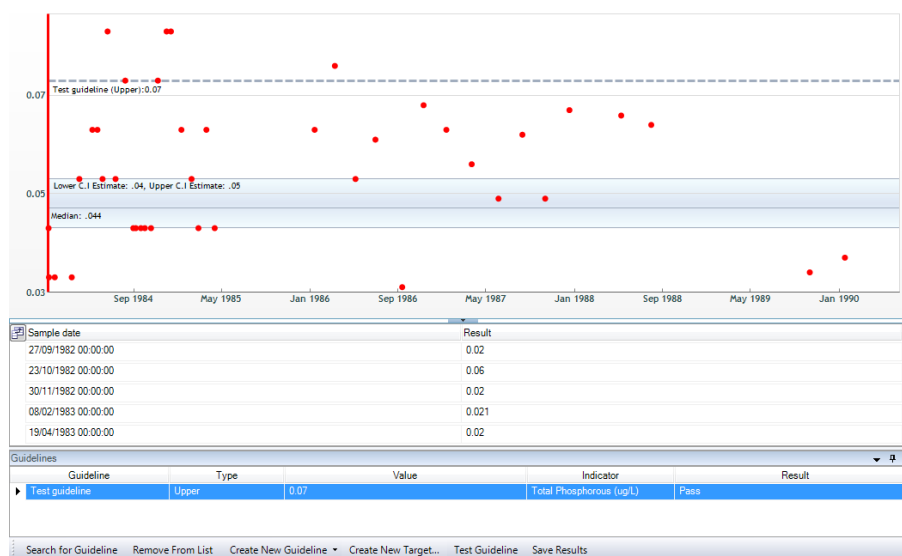


Figure 4.63: Test results

#### 4.9.16 Deriving water quality guidelines

This tool can derive both biological effects guidelines (trigger values) as well as guidelines from reference site data. Please refer to Appendix 1 for the details of deriving biological effect values and Appendix 2 for deriving guidelines from reference site data. If you have local guidelines values already derived, you can add those values into the built-in guideline database in this tool.

#### 4.9.17 Calculating guidelines using biological effects data

This software is designed to estimate the protecting concentrations of chemicals (toxics) such that a given percentage of species will survive. The preferred data are those derived concentration (preferably NOEC (No Observable Effect Concentration)) values from environmentally realistic and well conducted multiple species toxicity tests. NOEC is defined as the highest concentration of toxicant at which no statistically significant effect is observable compared to the controls (the statistical significance is measured at the 95% confidence level). At least three NOEC values are required from a field test.

To create a guideline from biological effects data, click the **File -> New -> Guideline -> From Biological Effects Data...** button. The **Biological Effects Guideline Creator** form appears. Select a toxicant data **File** by clicking the **Open...** button in the **Toxicant file** box. The toxicant data **File** should contain one numerical value per line. When the **File** is opened, the values appear in the chart. The **Model fitting** result box shows the name of the model and coefficients of the fitting result. The fitted line is also displayed in the chart. Adjust the **Percentile** or **Concentration** values in the **Protecting percentile vs concentration** box, then click the arrow to apply the change. Click the **Save...** button in the **Guideline** result box. The **New Guideline Target** form appears with pre-filled values for the upper guideline



percentile and estimate. Assign a name and location to the guideline then click the **Save** button to save the new guideline into the local database.

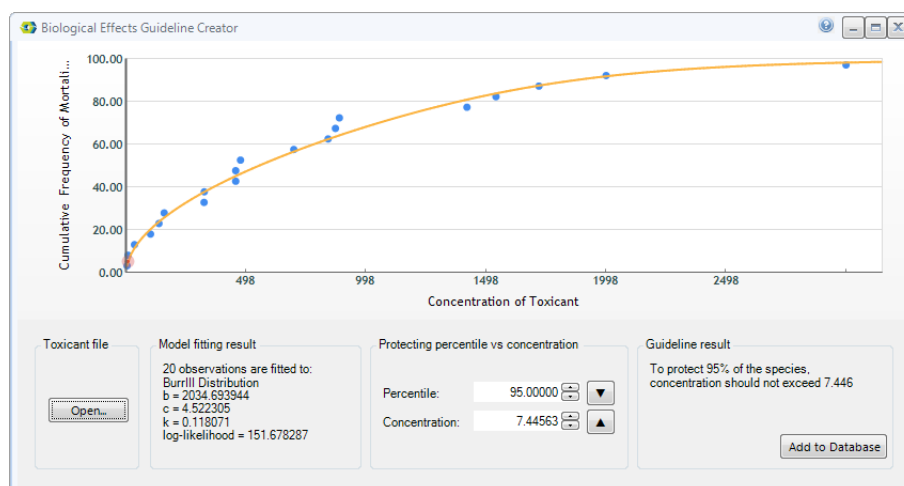


Figure 4.64: Biological Effects Guideline Creator

The estimations of the protecting concentrations are computed by fitting a certain distribution to the input data. The distribution used in this software is called the Burr III distribution, and it is that recommended by ANZECC & ARMCANZ 2000. Once the parameters are estimated they can be used to compute the cumulative distribution function of the appropriate distribution (Burr III or one of the limiting distributions), which is plotted with the input dataset in the results window. The software computes the concentration value corresponding to a certain level of percentile of the species to be protected if the concentration of the chemical is less than the estimated concentration. Further details on the distributions that are fitted and the estimation techniques used are given in Appendix 1.

#### Protected percentile and concentration values

There is a data entry box in the results window where a protecting percentile or concentration can be typed in. When the return/enter button is hit the percentile corresponding to the entered concentration is computed, or vice versa. The concentration can only be between 0 and the highest concentration shown on the plot.

Please note that the Y- axis is the cumulative frequency of mortality and the survival percentage should be 100 minus the percentage of mortality.

The bottom left-hand corner of the window contains the details of the model fitting (more details in Appendix 1).



Figure 4.65: Details of Biological Model Fitting

#### 4.9.18 Calculating guidelines using reference data

The calculation of percentile-based guidelines can be achieved by estimating the required percentiles from a reference dataset. The following instructions provide the basics on calculating guidelines using reference data with the Guidelines tool.

The details of statistics used in deriving these guidelines are given in Appendix 2.

To create a guideline using reference data, first ensure a **Time series** is open in the Chart layer view. Click the **File -> New -> Guideline -> From Reference Data...** button. The **Reference Data Guideline Creator** form appears. The active **Time series** is displayed in the chart along with the default upper and lower guideline values. Select a guideline testing type (**Upper limit**, **Lower limit** or **Both**) in the **Guideline Type** box. Specify an upper and/or lower percentile value for the guideline in the **Set Guideline Limits** box. The position of the guideline limits in the chart update to reflect the new values. The **Guideline result** box shows the estimated value and confidence interval for the upper and/or lower guideline.

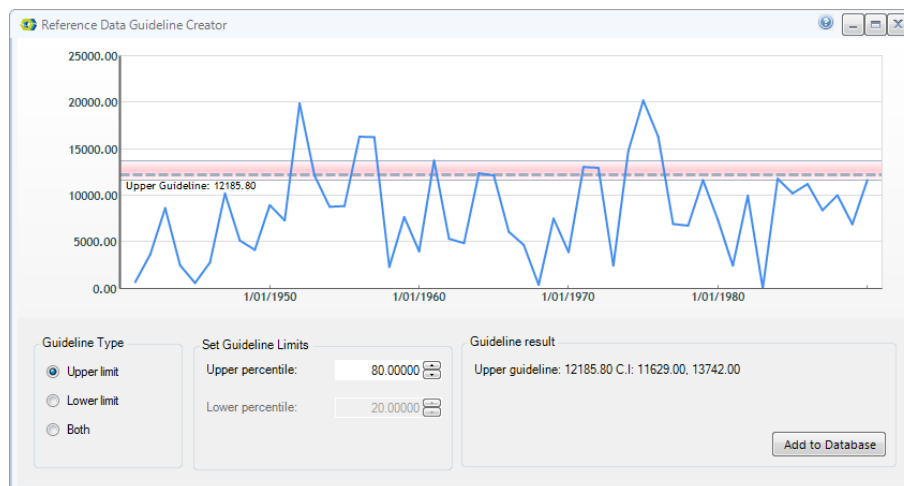


Figure 4.66: Reference Data Guideline Creator

Click the **Add to Database** button. The **New Guideline Target** form appears with pre-filled

values for the guideline percentile and estimate. Assign a name and location to the guideline then click the **Save** button to save the new guideline into the local database.

**Edit a guideline target**  
Edit an existing guideline target

General Locations **Guidelines**

C.I.: 0.00 Start Date: 22/06/2007  
 Sample Size: 40 End Date: 22/06/2007  
 Testing Type: Upper

	Percentile	Estimate	Lower C.I.	Upper C.I.
Upper Guideline:	80	12	12	14.000000
Median:		50	7	9
Lower Guideline:	20	0	0	0

Save Cancel

Figure 4.67: Guideline pre-filled reference data

Before committing these calculated values and statistics into the database, you need add additional information for these calculated guideline values. Please follow the instructions given in the next section to input required information.

#### Adding other required information

Click on the **General** tab and enter following information, or select items from the drop-down list available next to the text box.

**Edit Guideline Target**

Edit an existing guideline target

**General** | Locations | Guidelines

Name:

Source:

Indicator:

Protection type:

Type:

Figure 4.68: New Guideline Target : General tab

**Name:** Type name to identify the guideline value

**Source:** Type the source of the guideline value (e.g. User defined).

**Indicator:** Enter indicator name or select one from drop down list

**Environmental value:** Enter the environmental value for this guideline value

**Guideline type:** Indicate whether it is upper or lower

Note that the **Name**, **Parameter** and **Protection** fields are <u>compulsory</u>.

After completing this section click the **Location** tab to enter location details and water types.

The screenshot shows a software window titled "Edit Guideline Target" with a close button in the top right corner. Below the title bar, the text "Edit a guideline target" and "Edit an existing guideline target" is displayed. The window has three tabs: "General", "Locations", and "Guidelines". The "Locations" tab is currently selected and highlighted. Inside the "Locations" tab, there is a section labeled "Locations:" containing a list of four items: "General" (checked with a blue checkmark), "NSW", "QLD", and "TAS". Each item has a small square icon to its left. Below the list, there are two dropdown menus: "Primary water type:" set to "Marine" and "Secondary water type:" set to "Open Costal". At the bottom right of the window are two buttons: "Save" and "Cancel".

Figure 4.69: New Guideline Target : Location tab

Note that, **Primary** and **Secondary** water type fields are compulsory.

Click on the **Guideline** tab and view already populated values.

Once you complete all the fields, click **Save** to commit information into the database

#### 4.9.19 Add user defined values to the database

To add a user-defined guideline, click the **File -> New -> Guideline -> User Defined...** button. The **New Guideline Target** form appears. The form is split into three sections: **General**, **Locations** and **Guidelines**. The general section contains fields for the guideline name, source, indicator, protection type and guideline type. The locations section contains a

tree view of locations as well as fields to enter the primary and secondary water type. To associate the guideline with one or more locations, check the checkbox next to the location name. A guideline can be associated with one or more states, divisions, basins or specific locations. The guidelines section contains fields to enter the guideline values. For an upper guideline, set the **Testing Type** field to **Upper**. For a lower guideline, set the **Testing Type** to **Lower**. For a guideline with both upper and lower lifimits, set the testing type to both. Once the testing type is specified, fill in the percentile, estimate, lower C.I and upper C.I for the guideline values. Click the Save button to save the new guideline into the local database.



The screenshot shows a dialog box titled "New Guideline Target" with a close button (X) in the top right corner. Below the title bar, the text "Create a new guideline target" is displayed twice. The dialog has three tabs: "General", "Locations", and "Guidelines". The "General" tab is selected and highlighted. Inside the "General" tab, there are five labeled input fields, each with a dropdown arrow on the right:

- Name: Example Guideline
- Source: User-defined
- Indicator: Total nitrogen (ug/L)
- Protection type: Aquatic ecosystems
- Type: Guideline

At the bottom of the dialog, there are two buttons: "Save" and "Cancel".

Figure 4.70: User defined Guideline: General tab

#### Editing the guideline library (database)

To edit an existing guideline, click the **Analyse -> Guidelines...** button. The **Guidelines**

*tool* appears. Click the **Search for Guideline** button on the guidelines tool. Search for the guideline of interest, then left-click the guideline in the **Search Results** list. Click the **Edit** button to open the **Edit Guideline Target** form. Click the **Save** button to save any changes.

#### 4.9.20 Set water quality targets using proportional improvement

The Guidelines tool can be used to set water quality targets using a proportional-improvements approach. These targets could be any level of percentages agreed by the stakeholders. This tool assists users to set targets using the proportional improvement method. Users must input data from the selected site and then the software estimates the current median and relevant confidence values. The user can then enter the required percentage or the expected improvement, to estimate the target value and its confidence intervals.

To create a target using the proportional improvement method, first ensure a **Time series** is open in the Chart layer view. Click the **File -> New -> Guideline-> From Proportional Improvement...** button. The **Proportional Improvement Target Creator** form appears. The active **Time series** is displayed in the chart along with the median value. Set the desired median change by entering a percentage value in the **Median change (%)** field. The target median in the chart updates to reflect the new value. The target is also displayed in the **Guideline result** box, and the confidence intervals in the **Confidence Intervals** box.

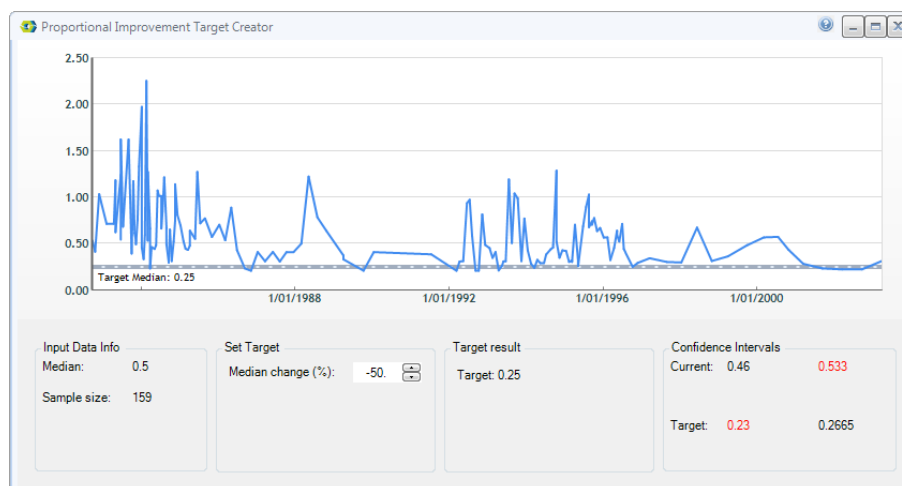


Figure 4.71: Set Targets

#### 4.9.21 References

ANZECC & ARMCANZ (2000). Australia and New Zealand Guidelines for Fresh and Marine Water Quality. Paper 4. Australian and New Zealand Environment and Conservation Council and the Agriculture and Resource Management Council of Australia and New Zealand.

Cambell, E., Palmer, M., Shao, Q. and Wilson, D. (2000) BurriOZ Software User Manual, CSIRO Mathematical and Information Sciences.

Lagarias, J.C., J.A. Reeds, M.H. Wright and P.E. Wright (1998) Convergence properties of the Nelder-Mead Simplex method in low dimensions. *SIAM J. Optim.* Vol 9, No 1. pp 112-147.

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 (2006) Queensland Water Quality Guidelines, 2006, Environmental Sciences Division. Environmental Protection Agency, No. 160, Ann Street, Brisbane, Queensland.

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

Tennakoon, S.B, I. Ramsay, N. Marsh and R. O'Connor (2007) The Integrated Monitoring and Assessment System (IMAS): A Decision Support System for Water Quality Monitoring and Assessment Programs. MODSIM 2007, International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2007, Christchurch New Zealand.

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.

## 4.10 eGuides

### 4.10.1 procedures-eguides-introductionIntroduction

This is an electronic document which consists of a number of documents commonly referred to for water-quality monitoring, modeling, assessment and guidelines for protecting environmental values and aquatic eco-systems. You can manually browse, or select the 'search' tab to search individual or all documents using key words, and you can view the searched items or copy them to another document or print them out for later reference. This provides efficient and quick access to water-quality-related information and knowledge.

### 4.10.2 Overview

eGuide is an electronic document which consists of a number of water-quality guideline documents that are commonly referred to. The current version of eGuides contains the following documents.

- ANZECC/ARMCANZ 2000 Monitoring & Reporting Guidelines
  - ANZECC/ARMCANZ 2000 Water Quality Guidelines
  - NHMRC 2005 Recreational Water Quality Guidelines
-



- Queensland Water Quality Guidelines

These documents have been compiled into a standard "HTML" version of Windows help systems. You can select the document that you would like to manually browse, or select the 'search' tab to search all the guides for some key words. The searched items can be viewed, copied to another document or printed out for later reference.

#### 4.10.3 Features

You can open and search individual document or the entire set of documents together, depending on your requirement.

With its standard MS Windows help format, the set of documents will provide a valuable knowledge base for water-quality data analysis and assessment.

#### 4.10.4 Audience

eGuides is useful for anyone who is involved in water-quality monitoring, data analysis, assessments, water-quality management or regulation activities.

#### 4.10.5 Limitations and caution notes for users

This version of eGuides contains national level documents relevant nationwide, and some that are specific to Queensland. Water-quality guidelines and other related documents for states other than Queensland will be included in a later version.

##### Using eGuides

##### *Getting started*

To open the eGuides tool, click the **Analyse -> eGuides...** button.

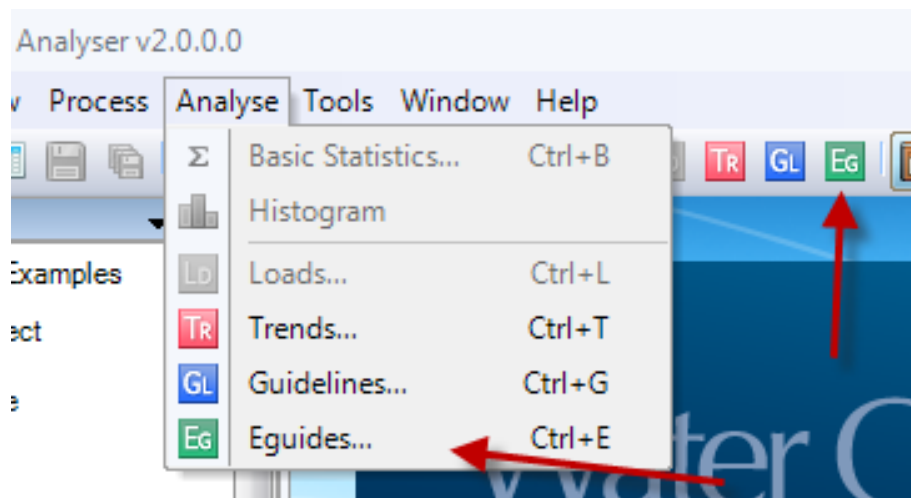


Figure 4.72: Opening eGuide

The eGuides tool appears. Click the name of an eGuides document to expand the cover view of that document. The document can be opened singly or as a subsection in the entire library of eGuides documents. To open the document as a subsection, click the **Open entire library** and navigate to selected document radio button. To open the document as a single file, click the **Open selected document only** button. Click the document cover image or the **Open this document** button to open the document. The document appears in a **Microsoft Windows Help window**.

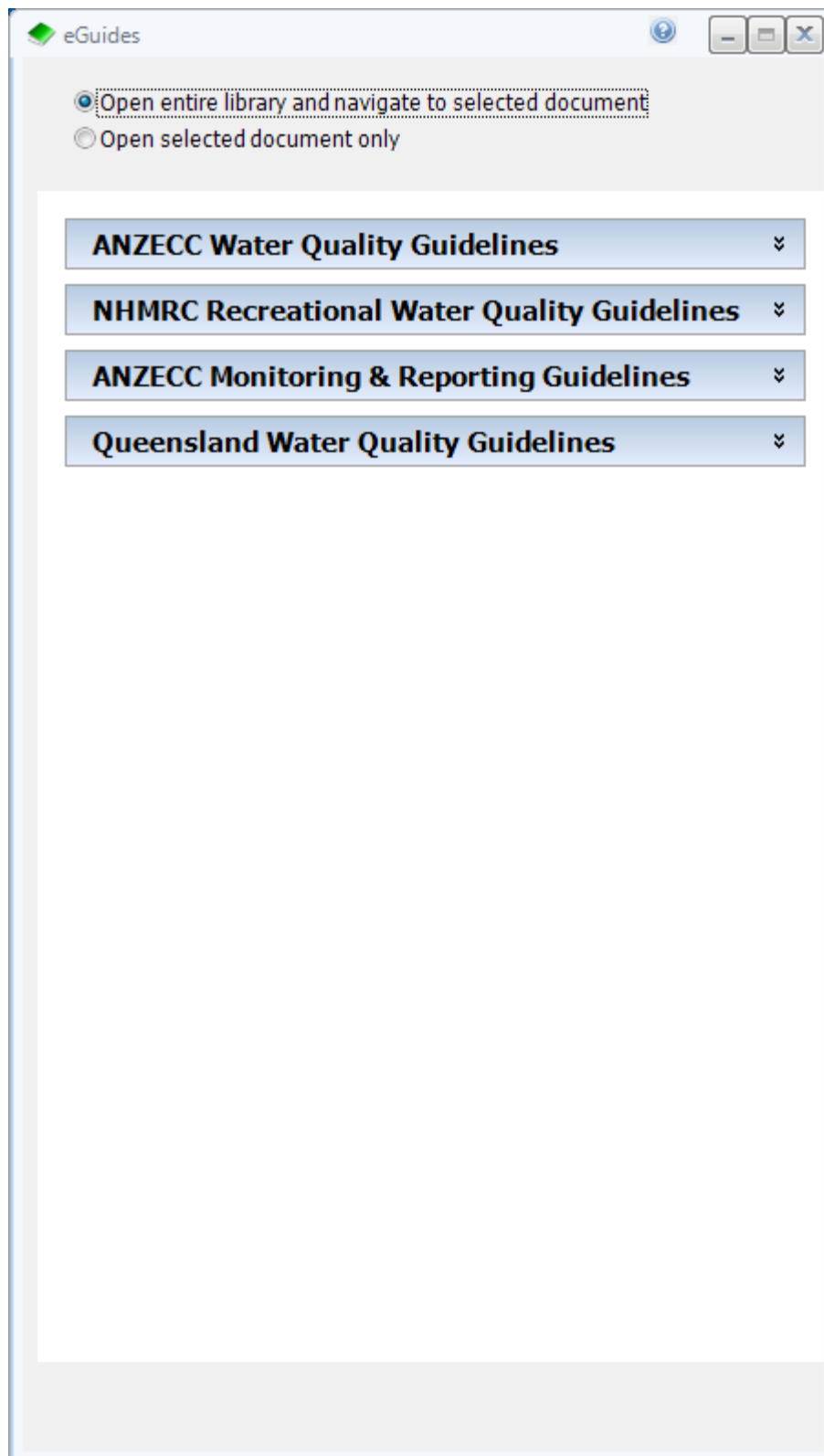


Figure 4.73: eGuides

***Open document/s:***

Click on the ***Open selected document only*** option and select a document that you want to open.

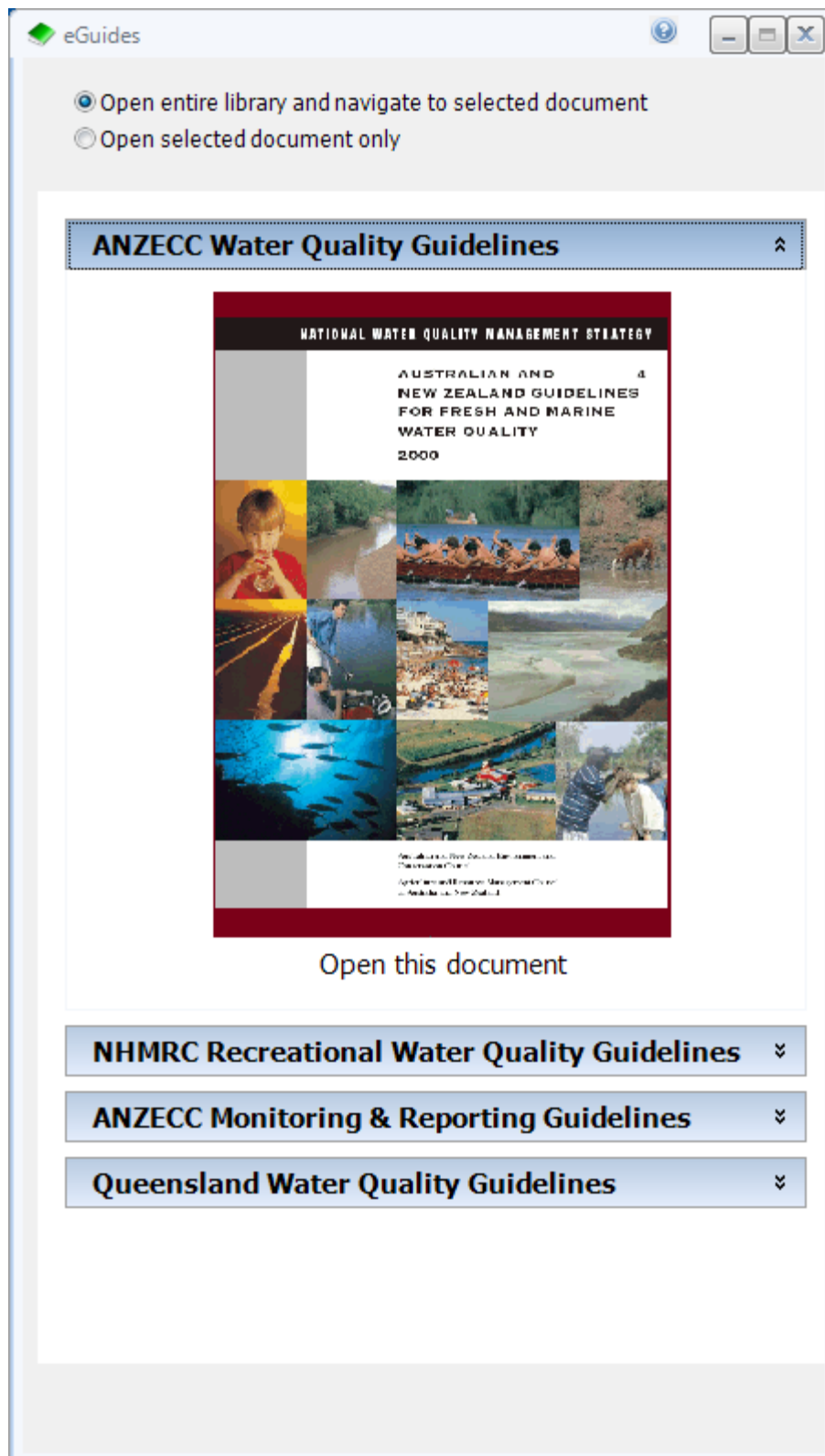


Figure 4.74: eGuides

The user help document is a standard "HTML" help document and you can search or navigate through the contents page to find required information.

This software can be navigated using the menu structure in the left hand side of the screen. The user help available with the software also provides assistance to users running the software.

Select **Help** on eGuide option for more information on using this document.

If you have opened all the documents, you will see the following screen.

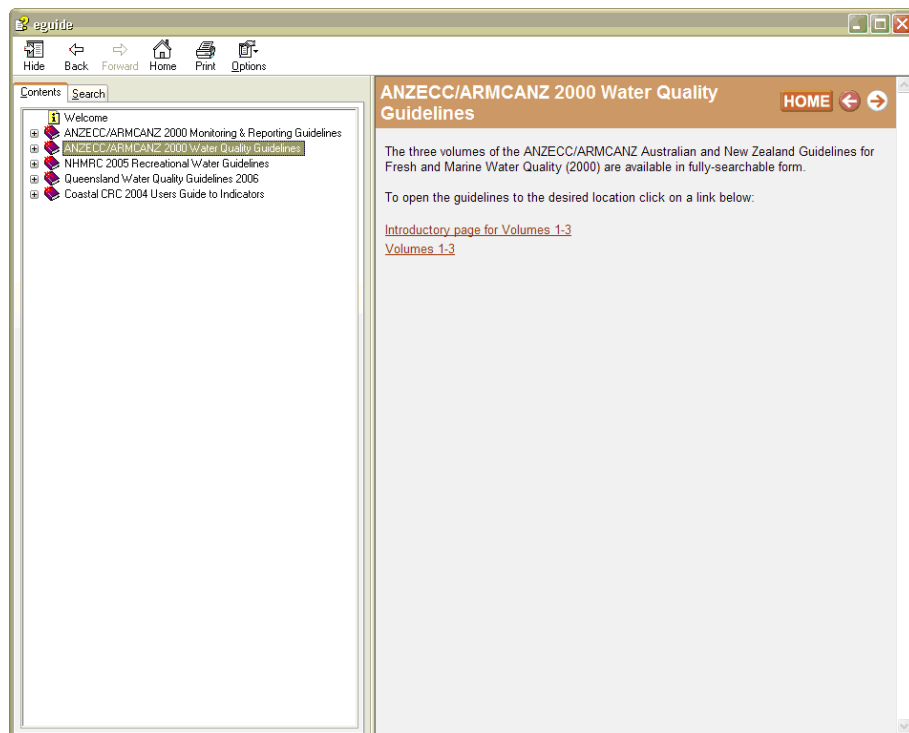


Figure 4.75: eGuides: Content or Search Options

You can navigate the system using either **Contents** or **Search** options.

#### 4.10.6 The Contents Tab

The Contents tab located on the far left of the help system contains the table of contents tree, which is a hierarchical system showing the structure of the document. You can click on a topic in the table of contents and jump straight to that topic at any time. You will always know where you are, because the topic you are on is highlighted.

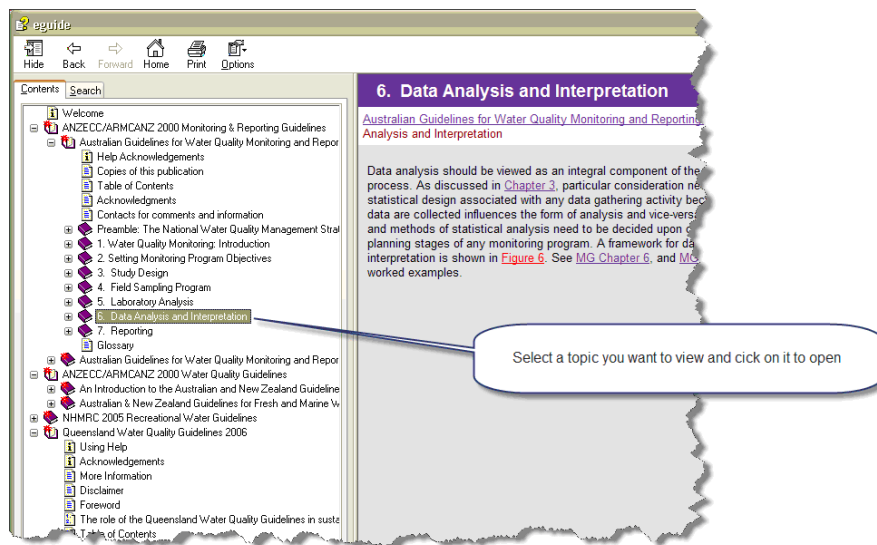


Figure 4.76: eGuides Content

#### 4.10.7 The Breadcrumb Trail

At the top of each topic you will see a list of topics followed by a >. Each topic name is an active link enabling you to jump to higher levels in the hierarchy of the current topic.

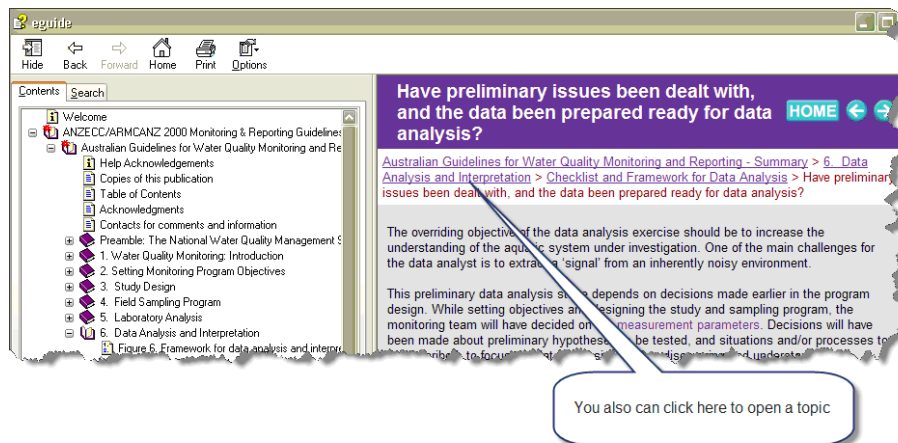


Figure 4.77: Breadcrumb

### 4.10.8 The Navigation Toolbar

The Navigation toolbar to the right of the header text allows you to navigate between topics in the table of contents.

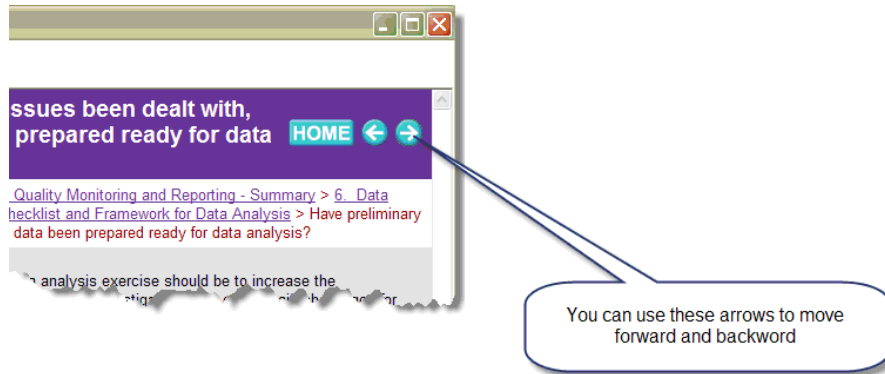


Figure 4.78: Navigation

Clicking **Home** will take you to the first topic in the section.

Clicking the **Previous** and **Next** buttons will move you to the previous and next topic, respectively. It follows the order the topics appear in the table of contents tree, but operation is limited to within each help file.

### 4.10.9 The HTML Help Toolbar

You can navigate using the HTML Toolbar (standard Windows buttons along the top of the Help system) to a limited extent only.

Clicking **Back** will display the last topic you visited

Clicking **Forward** will take you forward to the next page. (This only works if you have clicked the back button and want to move forward again.)

Clicking **Options** will display a popup menu listing the options available in the viewer.



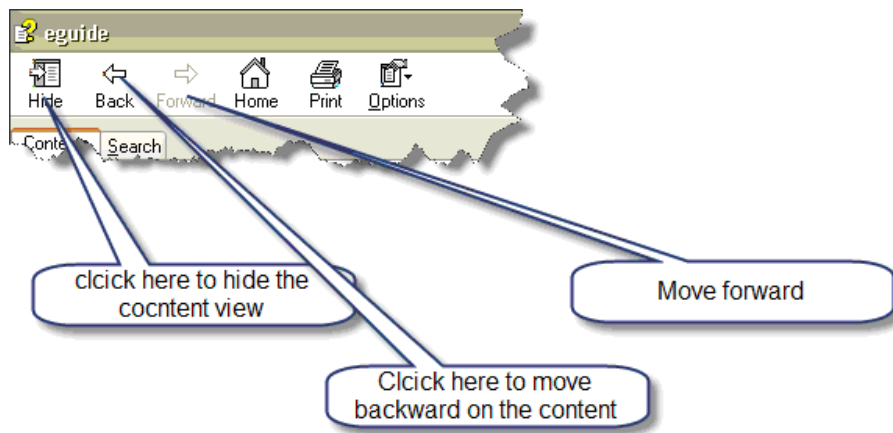


Figure 4.79: Help tool bar

The **Options** button available in the menu bar also offers options similar to those described earlier. You also can turn on and off the search highlight feature using the options available here.

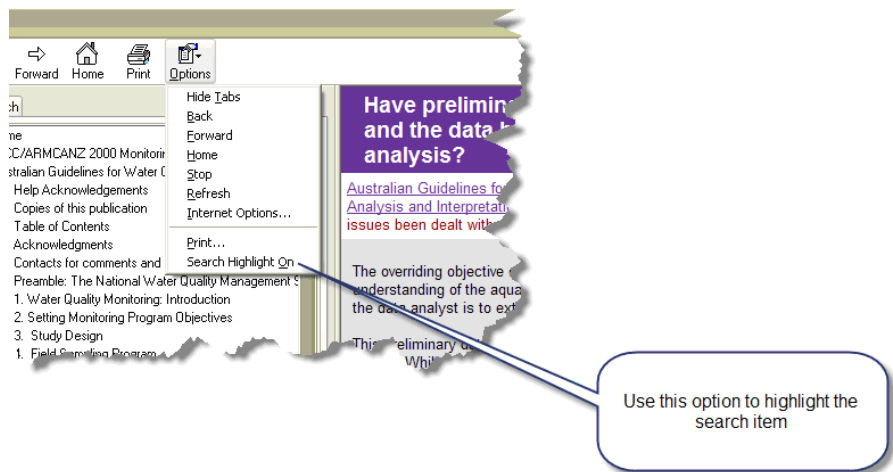


Figure 4.80: Help tool bar: Options

#### 4.10.10 Print a Help topic

1. Find and display the Help topic you want.
2. In the **Help** window, click **Print**.
3. Select the printing options you want.

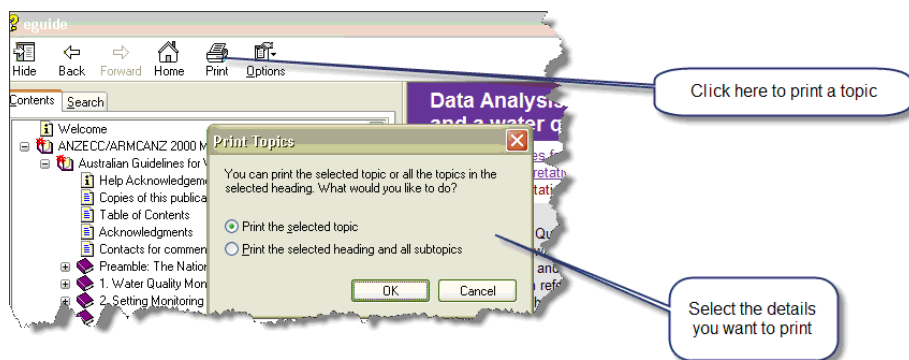


Figure 4.81: Printing a topic

Use search option to find information

Click on the **Search** tab to select a search option.

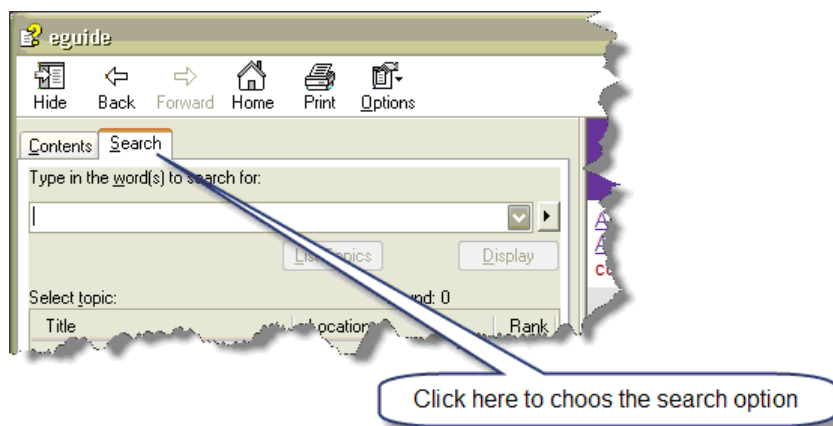


Figure 4.82: Search

You can type the search words in the text box and hit the **List Topic** button to perform the search. Note: If you type words within " " it will search for the entire phrase; otherwise it will search for all available words among those you typed in the box.

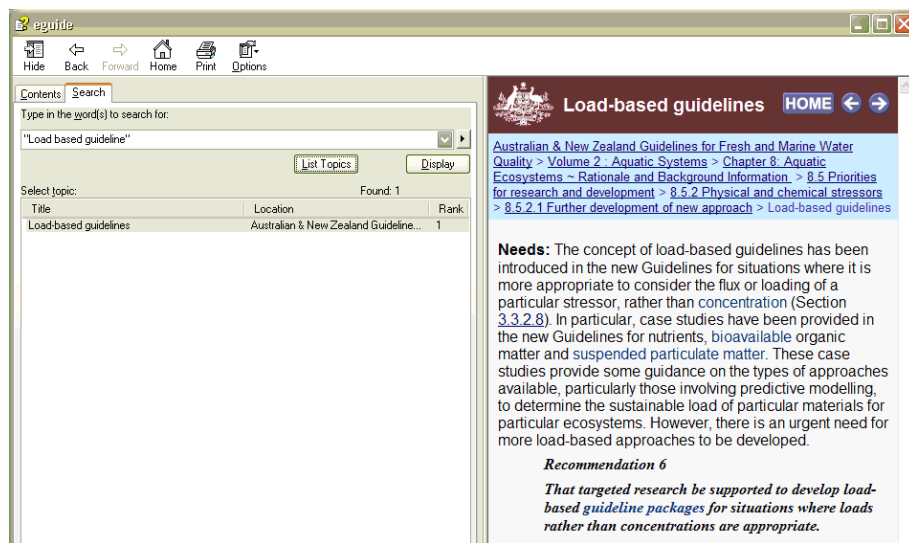


Figure 4.83: Search results

You can also use **AND**, **OR**, **NEAR**, **NOT** operators to refine your search.

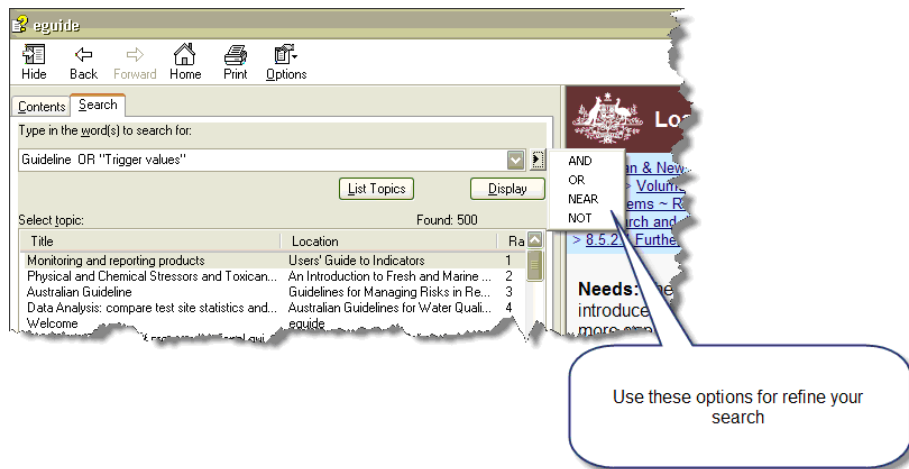


Figure 4.84: Search options

#### 4.10.11 Sort searched items

The search results can be sorted by documents using the **Location** as show below.

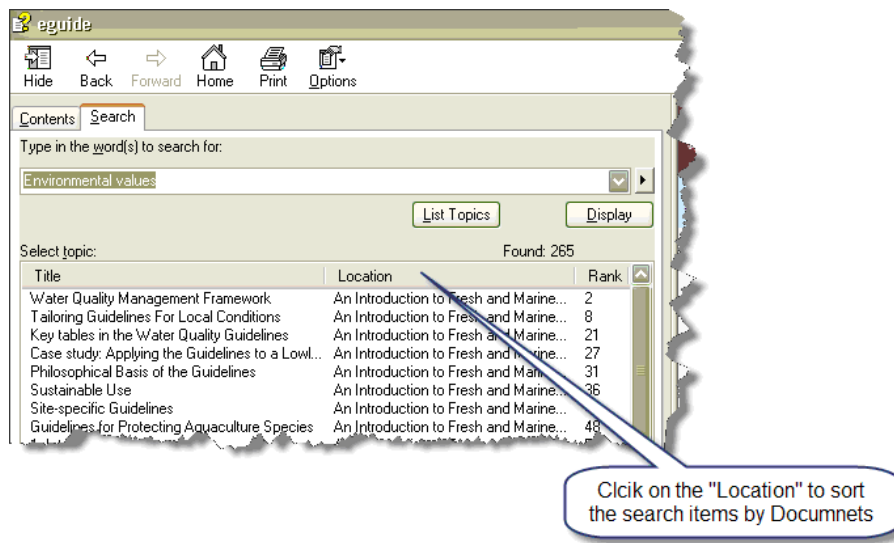


Figure 4.85: Sorting search items

#### 4.10.12 Find definitions for technical terms

Some definitions of technical terms can be found by clicking on the highlighted terms in the search results

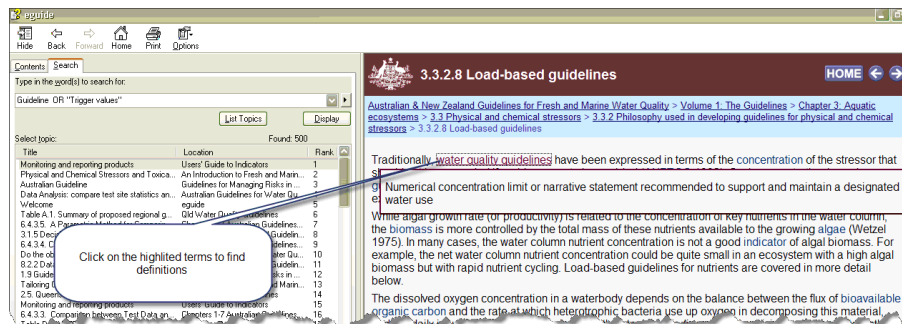


Figure 4.86: Find definitions

You also can use some hyperlinks in the search results to find required information.

#### 4.10.13 Copy and paste selected information into a document

You can copy the searched information into a document by simply selecting the information you want to copy. Use the mouse to select the required information and right click and choose the **copy** option.

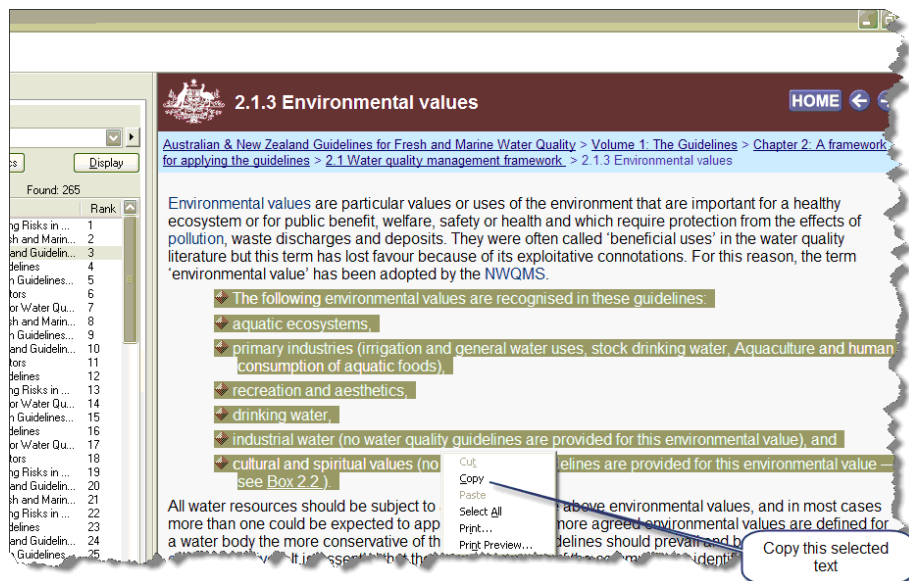


Figure 4.87: Copy and paste

You also can use **Ctrl + C** and **Ctrl + V** key strokes for copy and paste

#### 4.10.14 Exit eGuides

To close the eGuides click on the cross at the top right-hand corner of the window.

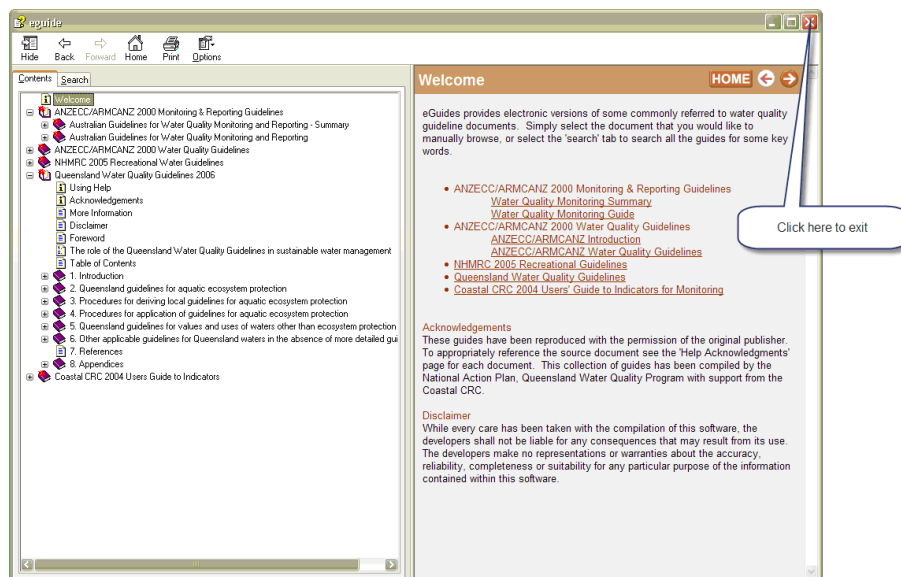


Figure 4.88: Exit eGuide

## Chapter 5

# Troubleshooting

### 5.1 Loads tool shows zero total load

Check whether the flow and concentration time series are aligned (have matching sample dates). Use the ***Align processor*** to prepare the time series if necessary.

### 5.2 Application does not start

Ensure the Microsoft .NET Framework 4 is installed by checking the ***Add*** or ***Remove Programs*** list in the ***Control Panel***.

The installer can be downloaded at

<http://www.microsoft.com/downloads/en/details.aspx?FamilyID=9cfb2d51-5ff4-4491->

## Chapter 6

# Appendices

[Appendix 1 - Biological approach of guideline setting](#)

[Appendix 2 - Procedures for deriving guidelines using reference data \(local or regional\)](#)

[Appendix 3 - Pollutant loads estimation in rivers or streams](#)

[Appendix 4 - Decision tree with preferred load estimation methods](#)

[Appendix 5 - Events load estimation](#)

[Appendix 6 - Protocols for the optimal measurement and estimation of nutrient loads](#)

[Appendix 7 - Assessment of Water Quality Data For Temporal Trends](#)

### 6.1 Appendix 1 - Biological approach of guideline setting

For deriving toxicity guidelines, aquatic toxicity data obtained from laboratory studies with single or multiple species in a controlled environment, are the most preferred type of data. The Guidelines software tool is designed to estimate the protecting concentrations of chemicals (toxicants) such that a given percentage of species will survive; and for it, too, the preferred input data are those derived concentration values (preferably NOEC) from environmentally realistic and well conducted multiple-species toxicity tests. NOEC is defined as the highest concentration of toxicant at which no statistically significant effect (at the 95% confidence level) is observable compared to the controls (ANZECC & ARMCANZ 2000). The local taxa list with existing sources of tolerance data can be matched with available national or international ecotoxicology data. This information is available at the USEPA ecotox database (<http://cfpub.epa.gov/ecotox>) or at the Australian society for ecotoxicology database.

A number of non linear models can be used, but, as recommended by ANZECC & ARM-CANZ 2000, in this version of the Guidelines software the estimates of the protecting concentrations are computed by fitting a Burr type III distribution to the input data.

#### **Statistical methodology**

The Burr III distribution is the distribution that is required by the (Queensland) Environ-



mental Protection Agency. There are other distributions fitted to the data: for example, the normal distribution and the log-logistic distribution, which are distributions that the users of this software may be familiar with. However, these latter two distributions are provided only as a reference guide and are not used for the estimation of the protecting concentrations.

### Fitting the distributions

#### *Burr III fitting*

The Burr III distribution is a very flexible 3-parameter distribution, which can provide good approximations to many commonly used distributions such as the log-normal, log-triangular and Weibull. The cumulative distribution function (CDF) for the Burr III distribution is

$$F_{\text{BurrIII}}(x) = \frac{1}{\left(1 + \left(\frac{b}{x}\right)^c\right)^2}$$

where  $x$  is concentrations

$b$ ,  $c$  and  $k$  are model parameters.

The 3 parameters of the Burr III distribution,  $b$ ,  $c$ , and  $k$ , are estimated by maximising the log-likelihood function (which is based on the probability distribution function). This maximisation is performed using the Nelder-Mead simplex algorithm (Lagarias et al. 1998), an optimisation technique that is not reliant on derivative information.

A complication of the Burr III distribution is that at limits of some of the parameters the Burr III distribution tends to a limiting distribution. As  $k \rightarrow \infty$  the Burr III distribution tends to the reciprocal Weibull distribution. As  $c \rightarrow \infty$  the Burr III distribution tends to the reciprocal Pareto distribution. In practical terms, if the Burr III distribution is fitted and  $k$  is estimated to be greater than 100, the estimation procedure is carried out again with a reciprocal Weibull distribution fitted. Similarly for the reciprocal Pareto distribution, if  $c$  is greater than 80.

Once the parameters are estimated they can be used to compute the CDF of the appropriate distribution (Burr III or one of the limiting distributions), which is plotted with the input dataset in the results window.

After the Burr III distribution has been fitted to the data, the protecting concentration (for preserving, for example, 90% of the species) is estimated using the estimated distribution parameters to compute the concentration such that the probability of there being a greater concentration (according to the fitted distribution) is 90%.

#### Estimating the protecting concentration

The protecting concentration is only calculated from the Burr III distribution (or an associated limiting distribution) fitted to the data. The software user must compute the concentration corresponding to the statement that 'q% of the species should be protected if the concentration of the chemical is less than the estimated concentration'. The value of  $q$  should be between 0 and 100 and, for the expected use of the software, will be somewhere close to 100, such as 80, 85, 90 or 95.

For a given value for  $q$ , the protecting concentration is estimated from the Burr III distribution as:

$$PC(q) = \frac{b}{[(1/(1-q))^{1/k} - 1]^{1/c}}$$

If the limiting distribution of reciprocal Weibull is used then the protecting concentration is estimated as:

$$PC(q) = (-\alpha/\ln(1-q))^{1/\beta}$$

where  $\alpha$  and  $\beta$  are the two parameters of the reciprocal Weibull distribution that were estimated in the fitting step. Similarly, if the reciprocal Pareto distribution has been necessarily fitted, the protecting concentration is estimated as:

$$PC(q) = x_0(1-q)^{1/\theta}$$

where  $x_0$  and  $\theta$  are the two parameters of the reciprocal Pareto distribution that were estimated in the fitting step.

#### References

- ANZECC & ARMCANZ (2000). Australian and New Zealand Guidelines for Fresh and Marine Water Quality. Paper 4. Australian and New Zealand Environment and Conservation Council and the Agriculture and Resource Management Council of Australia and New Zealand.
- Cambell, E., M. Palmer, Q. Shao and D. Wilson (2000) BurrliOZ Software User Manual. CSIRO Mathematical and Information Sciences.
- Lagarias, J.C., J.A. Reeds, M.H. Wright and P.E. Wright (1998) Convergence properties of the Nelder-Mead Simplex Method in low dimensions. *SIAM J. Optim.* 9(1):112-147.
- Shao, Q. (2000) Estimation for hazardous concentrations based on NOEC toxicity data: An alternative approach. *Environmetrics* 11:583-595.

## 6.2 Appendix 2 - Procedures for deriving guidelines using reference data (local or regional)

#### The statistical basis of the software

The software uses a non-parametric procedure to calculate percentile-based trigger values. This is a three-step procedure as follows:

1. The sample data are put into ascending order:  $Y_1, Y_2, \dots, Y_n$ .
  2. The rank ( $r$ ) of the required percentile is calculated using the formula  $1 + p(n-1)$ , where  $p$  is the proportion corresponding to the required percentile and  $n$  is the sample size.
  3. The percentile,  $X_p$ , is calculated from the linear interpolation formula  $(1 - rfrac)Y_{rint} + rfracY_{rint+1}$ , where  $rfrac$  and  $rint$  represent the fractional and integer parts of  $r$ .
-

The software defaults to the 20th and 80th percentiles as per ANZECC (2000).

In order to ensure consistent rigour into the calculation of the target values, diagnostics are calculated. These diagnostics represent a confidence interval that is approximately 80% wide. This interval has been chosen to keep the Type I error rate to 0.05 when testing the equality of two parameters with approximately equal standard errors (Payton, Greenstone & Schenker 2003). It is calculated using the binomial distribution as follows:

1. The sample data are put into ascending order:  $Y_1, Y_2, \dots, Y_n$ .
2. Calculate the rank (*rlow*) of the lower value of the interval by determining the  $\frac{\alpha}{2}$  quartile of a binomial distribution with size  $n$  and proportion corresponding to the required percentile.
3. The lower value of the interval is  $Y_{rlow}$ .
4. Calculate the rank (*rupp*) of the upper value of the interval by determining the  $1 - \frac{\alpha}{2}$  quartile of the binomial distribution described in 2 above.
5. The upper value of the interval is  $Y_{rupp}$ .

#### **Testing against guidelines**

The software allows you to test data against a selected or specified guideline value. This is achieved by comparing the guideline against an appropriate-size confidence interval about the median of the test data. The size of the interval is approximately 95% when testing against a single upper or lower value and 90% when testing against a guideline comprising an upper and lower value. These interval sizes have been chosen to maintain the Type I error rate at 0.05. The median and confidence interval are calculated in the same manner as detailed above.

#### **References**

- ANZECC & ARMCANZ 2000, *Australian and New Zealand Guidelines for Fresh and Marine Water Quality*, **Australian and New Zealand Environment and Conservation Council, and Agriculture and Resource Management Council of Australia and New Zealand**, Canberra.
- EPA 2006, *Queensland Water Quality Guidelines 2006*, Environmental Protection Agency, Brisbane.
- Payton, M. E., Greenstone, M. H., & Schenker, N. 2003, 'Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance?' 6pp. *Journal of Insect Science*, 3:34.

### **6.3 Appendix 3 - Pollutant loads estimation in rivers or streams**

Measures of the mass of materials (e.g. suspended sediment, total nitrogen, total phosphorus) carried by rivers and tributaries to receiving waters such as wetlands, estuaries and oceans are now an important performance benchmark in natural resources management initiatives and environmental management plans being implemented in Queensland and throughout Australia. The primary intention of these measurements has been to provide reliable information, which can be used for catchment and estuarine management.

---

While there is broad guidance on the various calculations of load that can be used, it is commonly acknowledged that for any one catchment the most appropriate method will depend on:

- The hydrological features of the catchment,
- The accuracy in load estimation required, and
- Data (flow and concentration) availability.

This software tool is designed to capture all the commonly used load calculation methods, and offer guidance on how they should be applied.

#### Background information

The pollutant load is the mass or weight of a pollutant of interest that passes a particular point of the river within a specified time (estimated using the loading rate and the time concerned).

The loading rate, or flux, is the instantaneous rate at which the load passes a point of reference on a river (e.g. a sampling station); it has units such as grams/second or tonnes/day. This is the product of pollutant concentration (usually in mg/L) and discharge rate.

The discharge rate, or flow, is the instantaneous rate at which water is passing the reference point, and it has units of volume/time such as cubic metres/second.

If we could directly and continuously measure the flux of a pollutant in a typical river or stream, we would see that the flux changes continuously with time. A hypothetical example is shown in Figure 1. The load for the period of time in the graph would be equal to the area shaded under the curve.

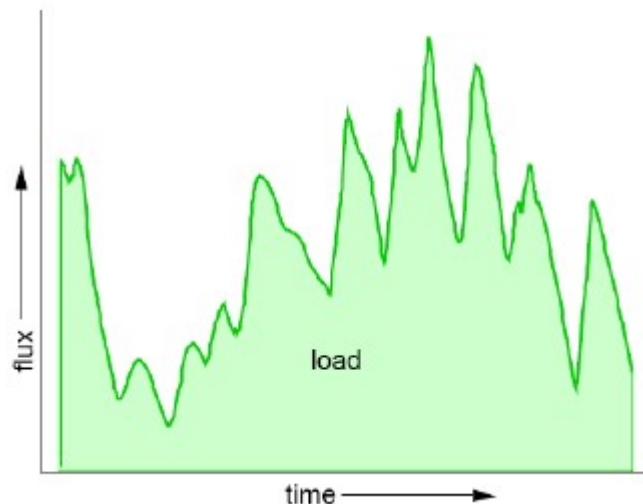


Figure 6.1: Load Flux Time

Mathematically, the load is the integral over time of the flux:

---

$$\text{Load} = \int_a flux(t) dt$$

There are several problems with the practical application of this concept of the load. The first is that we cannot measure flux directly. Instead, we measure flux as the product of concentration and flow. The second problem is that, while concentration and flow are both continuous functions of time, we cannot measure them continuously. Thus the integral which is the load must be estimated by summing the products of a sequence of discrete measurements of concentration and flow:

$$\text{Load} = k \sum_{i=1}^n c_i q_i t_i$$

where  $c_i$  is the  $i^{\text{th}}$  observation of concentration,  $q_i$  is the corresponding observation of flow, and  $t_i$  is the time interval represented by the  $i^{\text{th}}$  sample.

For constituents derived from storm events, the flux varies drastically over time, with fluxes during wet season or storm runoff events often several orders of magnitude greater than those during low flow periods. It is possible that 80 to 90% or more of the annual load is delivered during 10% or less of the time. Therefore, it is critical to sample during these periods of high concentration or flow, if an accurate load estimate is to be obtained.

Water-flow data are generally recorded at a regular time steps through automated gauge height reading instruments. Concentration data are more difficult to automate and as a consequence the data are generally less frequent and less regular than flow data. There are three basic approaches for overcoming this discrepancy in the flow and concentration data.

1. Ignore most of the flow data and calculate the load using the concentration data and the flows, which were observed at the same time the samples were taken.
2. Find a way to estimate "missing" concentrations: i.e. estimate concentrations during the flows when concentration samples were not taken.
3. Do something in between - find a way to use the more detailed knowledge of flow to adjust the load estimated from matched pairs of concentration and flow.

Based on the above approach there are three major categories of load estimation:

- Simple numeric integration method,
- Regression method,
- Ratio method.

The accuracy of the load estimate will depend on the proper application of the most appropriate method. Little can be done to compensate for data sets which contain an insufficient number of observations collected using an inappropriate sampling design.

---

Many load estimation programs choose monthly or quarterly sampling with no better rationale than that it's convenient. To avoid this, the sampling, which will be needed for load estimation, must be established in the initial planning process, based on quantitative statements of the precision required for the load estimate. If load estimates are to be made, determine the precision needed, based on the uses to which they will be put. For example, "I want the load estimates to be within 5% of the true loads in 90% of the years for which calculations will be made". The resources necessary to carry out the sampling program must be known, and budgeted for, from the beginning.

#### **Methods of load estimation**

##### ***Numeric integration***

The simplest approach to estimate loads is direct numeric integration, where:

$$\text{Load} = \sum_{i=1}^n c_i q_i t_i$$

where

$c_i$  is concentration in the  $i^{\text{th}}$  sample

$q_i$  is corresponding flow

$t_i$  is the time interval represented by the  $i^{\text{th}}$  sample

Numeric integration is only satisfactory if the sampling frequency is high - often on the order of 100 samples per year or more, and sufficiently frequent that all major runoff events are well sampled. This method, and particularly the sampling strategy, assumes that flows are highly variable and that concentrations increase with flow. The sampling strategy is based on the assumption that most of the load occurs in a short period of time during storm runoff events, and that accurate loads can be obtained by sampling primarily during that period of time. If a pattern can be identified which will allow sampling to be allocated more efficiently by concentrating sampling at certain times, the schedule can be adjusted.

##### **Period-weighted method:**

For the period-weighted approach, measured concentrations are linearly interpolated between sample times.

##### **Numeric-integration method:**

As a numeric integration method, "averaging approaches" use some form of average in the calculation of the loads. The simplest approach involves multiplying the average concentration for a period of time by the mean daily flow for each day in the time period to obtain a succession of estimated daily (unit) loads. Generally, averaging approaches tend to be biased if concentration is correlated with flow (the calculated load is too low if the correlation is positive and too high if the correlation is negative).

The flow interval technique is a semi-graphical technique, which begins with a plot of the year's observed instantaneous fluxes as a function of instantaneous flows at the time the samples were taken. The plot is divided into several intervals of uniform size covering the range of mean daily flows for all days of the year. For each interval, the average flux is calculated and the number of days with mean daily flows in the interval is determined. The interval load is calculated as the product of the average flux, the

---

number of days in the interval, and the appropriate units conversion factor. The annual load is calculated by summing the interval loads.

In this method, the daily load is calculated as the product of concentration and flow on days on which samples are taken, and the mean of these loads is calculated. The mean daily load is then adjusted by multiplying it by a flow ratio, which is derived by dividing the average flow for the year as a whole by the average flow for the days on which chemical samples were taken. The adjusted mean daily load is multiplied by 365 to obtain the annual load. A bias correction factor can be included in the calculation, to compensate for the effects of correlation between discharge and load.

The basis of the ratio method is the assumption that the ratio of load to flow for the entire year should be the same as the ratio of load to flow on the days concentration was measured. Thus

$$L_a/Q_a = L_o/Q_o$$

where

$L_a$  is Average daily load for the year

$Q_a$  is Average daily flow for the year,

$L_o$  is Observed daily load

$Q_o$  is Observed daily flow

$$L_a = L + o \cdot Q_a/Q_o$$

$$\text{Annual Load} = L_a \cdot 365$$

Ratio estimators assume that there is a positive linear relationship between concentration and flow. In most applications of ratio estimators to pollutant load estimation, the calculations and sometimes the sampling program have been stratified, usually by flow and/or season.

#### **Regression method**

Regression approaches develop a relationship between concentration and flow, based on the samples taken. These data are used to establish a regression relationship of the form

$$c = mq + b$$

where

$c$  is concentration

$m$  is slope of the linear relationship

$q$  is flow

$b$  is intercept.

Then we use the relationship to estimate a representative concentration for days not sampled, usually using the mean daily flow as input to the regression equation. This

relationship may involve simple or multiple regression, and concentration or flux may be used as the dependent variable. In many applications, both concentration (or flux) and flow are log-transformed to create a dataset that is better suited for regression analysis. Regression relationships between log-transformed concentration or flux and flow are often called rating curves.

Finally, the annual load is calculated as the sum of the daily fluxes based on the estimated daily loads and mean daily flows

$$\text{Load} = \sum_{i=1}^3 65c_iq_i$$

Regression approaches assume only that there is a linear relationship between a dependent variable (concentration or flux) and one or more independent variables (typically flow but sometimes also higher powers of flow, time, seasonality, and other variables).

The aim of sampling is to thoroughly characterise the relationship between flow and concentration (or flux). The sampling of concentration and flow should be distributed over the flow regime.

#### ***Performance of the methods***

Some methods of estimating loads provide a measure of the uncertainty of the load estimate. Unfortunately, the uncertainty estimates of different load calculation methods cannot be directly compared, because they reflect different kinds of "error", and the estimated "error" may be different from the error we are interested in (sampling and analytical error). Generally evaluation of load estimation methods must rely on comparative studies in which several methods are used to calculate loads from the same data, and the results are compared with the "true" load, which is independently known.

- Averaging methods are generally biased, and the bias increases as the size of the averaging window increases. In general, the annual load which is the sum of the four quarterly loads will be more biased than the annual load which is the sum of the 12 monthly loads.
- Regression approaches can perform well if the relationship between flow and concentration is sufficiently well-defined, linear throughout the range of flows, and constant throughout the year.
- In most studies, ratio approaches performed better than regression approaches, and both perform better than averaging approaches. In particular, ratio approaches that include a bias correction factor and are used in a stratified mode generally show low to no bias, relatively high precision, and resistance to undue influence from unusual observations.

#### ***Simple integration***

##### **Data requirements**

Generally 100 samples per year or more are recommended. They should be sufficiently frequent that all major runoff events are well sampled.

##### **Biases**

---



Simple integration usually underestimates if events are not adequately sampled.

$$k \sum_{i=1}^n \frac{c_i}{n} \sum_{i=1}^n \frac{q_i}{n} = k \bar{c} \bar{q}$$

where

$c_i$  is the  $i^{th}$  sampled concentration

$q_i$  is the  $i^{th}$  sampled discharge (flow)

$\bar{c}$  is average of n concentration measurements

$\bar{q}$  is average of n discharge measurements

$k$  is number of time intervals in the period (e.g.  $k=365$ )

#### Description

Numeric integration is only satisfactory if the sampling frequency is high, - often on the order of 100 samples per year or more, and sufficiently frequent that all major runoff events are well sampled. This method, and particularly the sampling strategy, assumes that flows are highly variable and that concentrations increase with flow. The sampling strategy is based on the assumption that most of the load occurs in a short period of time during storm runoff events, and that accurate loads can be obtained by sampling primarily during that period of time. If a pattern can be identified which will allow sampling to be allocated more efficiently by concentrating sampling at certain times, the schedule can be adjusted. For the period-weighted approach, measured concentrations are linearly interpolated through time between samples.

Express the time of each sample and each flow observation as decimal days of the year. For example, noon on January 2 would be day 1.5, 6:00 p.m. would be day 1.75. For each concentration sample, establish a time window which starts halfway between the sample and the previous sample, and ends halfway between the sample and the next sample. The time window for the first sample during a storm should include only the interval based on the time to the next sample (i.e. should start at the beginning of the storm); the time window for the last sample during a storm should include only the interval based on the time from the preceding sample (i.e. should end at the end of the storm). Multiply the flow at the time of the sample by the concentration at the time of the sample and the time interval for the sample. The result is the load for the time interval. The annual load is calculated by adding all these loads together.

#### Scale averaging load estimates

##### Data requirements

Optimum sample number depends on the variability of the flow.

##### Biases

Scale averaging load estimates are biased if concentration is correlated with flow

##### Equation

$$\text{Load} = k \sum_{i=1}^n \frac{c_i q_i}{n}$$

where

$c_i$  is  $i^{th}$  sampled concentration

$q_i$  is  $i^{th}$  sampled discharge (flow)

$k$  is number of time intervals in period (e.g.  $k=365$ )

#### **Description**

Averaging approaches use some form of average in the calculation of the loads. This approach involves multiplying the average observed concentration by the average flow based on all days of the year to obtain an "average" daily load, which is then converted to the total load. Other variants are the monthly average concentration multiplied by the average flow for the month; the quarterly average concentration multiplied by the average flow for the quarter, etc.

Averaging methods are generally considered to be the simplest available techniques for pollutant load estimation, and are often applied because of a lack of more appropriate techniques. Estimates of load over a time period are made by using averages of discharge, concentration or load for a given subinterval and then summing these over the entire period. These averages may be over different time periods, such as monthly, quarterly or yearly, and can combine discharge and concentration in a number of different ways. Whilst these methods are easy to apply, the assumptions implicit behind such calculations, including independent and identically distributed data, are rarely met. This leads to bias in the estimation of loads, especially if the sampling program does not collect data from the entire range of discharge and concentration variability. Where a positive relationship occurs between concentration and discharge, loads will tend to be underestimated by time averaging, and where a negative relationship exists, loads will usually be overestimated. The magnitude of over- or underestimation will depend on the range in variation in concentration. This effect is likely to be more severe for suspended sediments which tend to show stronger positive relationships with discharge than nutrients which are transported in both dissolved and particulate form.

However, some averaging approaches have shown relatively high precision in some studies, and might be useful in special situations, for example if the goal is to detect a change in the load, and if detecting the change is more important than knowing the actual magnitude of the load.

#### ***Flow-weighted concentration method***

**Data requirements** Significant increases in nutrient inputs may occur during wet-weather flows. Therefore, sufficient dry weather flow and concentration data should be collected to avoid overestimating load contributions.

#### **Biases**

Generally, averaging approaches tend to be biased if concentration is correlated with flow: the calculated load is too low if the correlation is positive and too high if the correlation is negative. However, some averaging approaches have shown relatively high precision in some studies, and might be useful in special situations, for example if the goal is to detect a change in the load, and detecting the change is more important than knowing the actual magnitude of the load.

#### **Equation**

---

$$Q \frac{\sum_{i=1}^n c_i q_i}{\sum_{i=1}^n q_i} = Q \sum_{i=1}^n w_i c_i$$

where

$c_i$  is  $i^{th}$  sampled concentration

$q_i$  is  $i^{th}$  sampled discharge (flow)

$Q$  is total discharge for period

#### Description

Averaging approaches use some form of average in the calculation of the loads. This approach involves multiplying the average observed concentration by the average flow based on all days of the year to obtain an "average" daily load, which is then converted to the total load. Other variants are the monthly average concentration multiplied by the average flow for the month; the quarterly average concentration multiplied by the average flow for the quarter, etc.

Averaging methods are generally considered to be the simplest available techniques for pollutant load estimation, and are often applied because of a lack of more appropriate techniques. Estimates of load over a time period are made by using averages of discharge, concentration or load for a given subinterval and then summing these over the entire period. These averages may be over different time periods, such as monthly, quarterly or yearly, and can combine discharge and concentration in a number of different ways. Whilst these methods are easy to apply, the assumptions implicit behind such calculations, including independent and identically distributed data, are rarely met. This leads to bias in the estimation of loads, especially if the sampling program does not collect data from the entire range of discharge and concentration variability. Where a positive relationship occurs between concentration and discharge, loads will tend to be underestimated by time averaging, and where a negative relationship exists, loads will usually be overestimated. The magnitude of over- or underestimation will depend on the range in variation in concentration. This effect is likely to be more severe for suspended sediments which tend to show stronger positive relationships with discharge than nutrients which are transported in both dissolved and particulate form.

However, some averaging approaches have shown relatively high precision in some studies, and might be useful in special situations, for example if the goal is to detect a change in the load, and if detecting the change is more important than knowing the actual magnitude of the load.

#### *Linear interpolation of concentration data*

##### Data requirements

For linear interpolation of concentration data, we require relatively more samples per year, taken sufficiently frequently that all major runoff events are included.

##### Biases

There is bias if sample number is low, and the method generally gives an underestimate.

##### Equation

---

$$\sum_{j=1}^n \frac{c_j + c_{j+1}}{2} q_j$$

where

$c_j$  is  $j^{th}$  sample concentration

$q_j$  is inter-sample mean flow.

#### Description

Concentration values are linearly interpolated to represent un-sampled days. In this technique, assumptions are made about how concentrations vary in time between samples. Typically we linearly interpolate between concentrations or apply cubic splines to a time series of concentrations. These techniques assume that concentrations from individual samples represent the average daily concentration for the sampled day, and their average daily concentration on non-sampled days can be determined by linearly interpolating between fortnightly or monthly sampled concentrations.

#### Flow stratified sampling

##### Data requirements

Stratification is performed by dividing the population into homogeneous sub-units called strata. Subunits should be sampled separately and the estimates should be combined to obtain an estimate over the entire population.

##### Biases

For load estimation substantial improvements in estimation efficiency can be expected from stratification. However, a number of studies have demonstrated that the most significant gains in precision and accuracy are obtained by relatively simple stratification strategies (e.g. high-flow / low-flow dichotomy).

##### Equation

$$\sum_{j=1}^{n_s} \frac{N_j}{n_j} \left[ \sum_{i=1}^{n_j} q_{ij} c_{ij} \right]$$

where

$c_i$  is  $i^{th}$  sampled concentration

$q_i$  is  $i^{th}$  sampled discharge (flow)

$N_j$  is the number of time intervals in  $j^{th}$  stratum

#### Description

Whereas systematic sampling can give misleading results particularly in the presence of autocorrelation and periodicity, the stratified sampling method can overcome some of these difficulties. This strategy assigns sampling effort in proportion to flux - periods of high flux are sampled more intensively than periods of low flux. One of the main advantages of stratified sampling is the potential reduction in estimation variance.

#### Simple ratio

---

**Data requirements**

Variance estimates associated with ratio estimators are only reliable if the sample size exceeds 30 and is also large enough that the coefficients of variation of mean discharge and load are both less than 10%.

**Biases**

Ratio estimators assume that there is a positive linear relationship between dependent and independent variables, which passes through the origin.

**Equation**

$$\frac{\bar{l}}{\bar{q}}Q$$

where

$\bar{l}$  is average load for sample

$\bar{q}$  is average of n discharge

Q is total flow

**Description**

On days on which samples are taken, the daily load is calculated as the product of concentration and flow, and the mean of these loads is also calculated. The mean daily load is then adjusted by multiplying it by a flow ratio, which is derived by dividing the average flow for the year as a whole by the average flow for the days on which concentration samples were taken.

For example, if the mean of daily load calculations was 10 t/day and the mean daily flow for the whole year was 100 mL/d, but for those days where concentration data was collected, the mean daily flow was 50 mL/d then the total load would be 10 t/day over 365 days, then scaled by the flow proportion

$$(10/50) \cdot 100 \cdot 365 = 7300\text{t/year}$$

Ratio estimators assume that there is a positive linear relationship between dependent and independent variables which passes through the origin. Also, if the variance of the dependent variable is proportional to the magnitude of the independent variable, the ratio estimator is known to be the best linear unbiased estimator, i.e. the most precise among the class of unbiased estimators which assume a linear relationship. Both of these conditions are often satisfied, at least approximately, by relationships between load and discharge.

***The Beale ratio estimator*****Data requirements**

Needs a sample size that exceeds 30 or more to be reliable. This limitation also applies to each stratum if ratio estimation is applied within a stratified sampling scheme.

**Biases**

In particular, ratio approaches which include a bias correction factor and are used in a stratified mode generally show low to no bias, and relatively high precision, In most

studies, ratio approaches perform better than regression approaches, and both perform better than averaging approaches.

#### Equation

$$Q \left( \frac{\bar{l}}{\bar{q}} \right) \left\{ \frac{1 + \frac{1}{N} \frac{\rho \sigma_L \sigma_Q}{\bar{l} \bar{q}}}{1 + \frac{1}{N} \frac{\sigma_Q^2}{\bar{q}^2}} \right\}$$

where

$Q$  is total discharge for period

$\bar{l}$  is average load for sample

$\bar{q}$  is average of  $n$  discharge measurements

The term in curly brackets is the bias correction term.  $N$  is the expected population size (this is included in the calculation, to compensate for the effects of correlation between discharge and load).

#### Description

On days on which samples are taken, the daily load is calculated as the product of concentration and flow, and the mean of these loads is also calculated. The mean daily load is then adjusted by multiplying it by a flow ratio, which is derived by dividing the average flow for the year as a whole by the average flow for the days on which concentration samples were taken. A bias correction factor is included in the calculation, to compensate for the effects of correlation between discharge and load. The adjusted mean daily load is multiplied by 365 to obtain the annual load. When used in a stratified mode, the same process is applied within each stratum, and the stratum load is calculated by multiplying the mean daily load for the stratum by the number of days in the stratum. The stratum loads are then summed to obtain the total annual load.

Ratio estimators assume that there is a positive linear relationship between dependent and independent variables which passes through the origin. Also, if the variance of the dependent variable is proportional to the magnitude of the independent variable, the ratio estimator is known to be the best linear unbiased estimator, i.e. the most precise among the class of unbiased estimators which assume a linear relationship. Both of these conditions are often satisfied, at least approximately, by relationships between load and discharge.

#### Concentration power curve

#### Data requirements

The goal of sampling is to thoroughly characterise the relationship between flow and concentration. Site-specific monitoring data should be used whenever possible to check the accuracy of the predictions.

#### Biases

Regression approaches can perform well if the relationship between flow and concentration is well-defined, linear throughout the range of flows, and constant throughout the year.

#### Equation

$$c = aq^b$$

where

$c$  is concentration

$q$  is flow

$b$  is power

#### Description

Regression approaches develop a relationship between concentration and flow, based on the samples taken. These data are used to establish a regression relationship and coefficients "a" and "b".

This method is only valid if there is high relationship between flow and concentration

#### *The USGS seven-parameter model*

Previous studies have proven that loads of nutrient transport in rivers can be estimated accurately using a more complex regression model. The regression model given by Cohn et al.(1992) has been used by the United State Geological Survey (USGS) to estimate nutrients loads in may rivers. This model is also known as USGS seven parameter methods and presented in the following equation.

$$\ln [C] = \beta_0 + \beta_1 \ln \left[ \frac{Q}{\bar{Q}} \right] + \beta_2 \ln \left[ \frac{Q}{\bar{Q}} \right]^2 + \beta_3 [T - \bar{T}] + \beta_4 [T - \bar{T}]^2 + \beta_5 \sin[2\pi T] + \beta_6 \cos[2\pi T] + \varepsilon$$

The notations are as where  $\ln[ ]$  means the natural logarithm of the parameter,  $C$  is concentration,  $Q$  is flow rate, and  $T$  is time measured in years. The errors, denoted by  $\varepsilon$  are assumed to be independent and normally distributed with a mean value of 0 and variance  $\sigma\varepsilon$

$\beta_0$  through  $\beta_6$  are the seven parameters which must be estimated by multiple regression, and  $\bar{Q}$  and  $\bar{T}$  are "centring variables", which simplify the mathematics but have no effect on the load estimate. These centring variables are calculated using the following equations.

$$\tilde{W} = \bar{W} + \frac{\sum_{i=1}^n (W_i - \bar{W})^3}{2 \sum_{i=1}^n (W_i - \bar{W})^2}$$

where  $W = \ln(Q)$  and  $\tilde{Q}e^{\tilde{W}}$

$$\tilde{T} = \bar{T} + \frac{\sum_{i=1}^n (T_i - \bar{T})^3}{2 \sum_{i=1}^n (T_i - \bar{T})^2}$$

Note that if  $\beta_2$  through  $\beta_6$  are zero, this complex formula is comparable to the simple regression relationship, except that log-transformed variables are used and discharge is "centered".  $\beta_0$  is the intercept term and  $\beta_1$  is the slope term. This model includes the capability of accounting for curvilinear relationships between concentration and flow,

through the Q2 term, for trends over time (T and T2 terms), and for seasonality (sin and cos terms).

#### Transformation, back-transformation and bias avoidance

In order for concentrations estimated from the regression model to be reliable, the residuals (the differences between the predicted and observed concentrations used to calculate the regression model) must be normally distributed. In addition, it is desirable for the data to be well spread out over the range of observations. For these and several other reasons, regression models relating concentration to flow usually use log-transformed values. In order to be of much use, however, the resulting data must be back-transformed before calculating the loads. The obvious way to do this is by taking the anti-logs of the estimated concentrations. Statistical theory tells us, however, that when these back-transformed values are used to calculate average daily loads or total annual loads, the results will be biased low (Ferguson, 1986a, 1986b; Koch and Smillie, 1986a, 1986b; Cohn et al., 1989, 1992). In order to avoid this bias, a value must be added to each estimated log-concentration before it is back-transformed. According to Ferguson (1986a), under the assumption that the residuals are normally distributed, the appropriate procedure is to estimate the concentrations using

$$\hat{c} = e^{(\hat{y}^2/2)}$$

where  $\hat{y}$  is the log-concentration estimated from the regression model, and  $\sigma^2$  is the variance of the residuals of the regression model.

This is referred to by Cohn et al. (1989), as the quasi-maximum likelihood estimate, or QMLE. If the original transformation used common logs (base 10) rather than natural logs, the equivalent would be

$$\hat{c} = 10^{\hat{y} \cdot 2.65 \cdot \sigma^2}$$

Koch and Smillie (1986a, b) pointed out that if the assumption of normally-distributed residuals is violated by the actual data, the application of 12 or 12a can actually lead to overestimates which are farther from the true value than those provided by the uncorrected back-transformation. They applied a nonparametric correction factor called the smearing estimate (Duan, 1983), with even worse results, and concluded that there may be no single approach to bias correction which will work reliably in all cases.

However, as Ferguson (1986b) pointed out, Koch and Smillie failed to distinguish between systematic error (bias) and random error (precision), and the smearing estimate cannot be generally condemned on the basis of their results. The smearing estimate is a constant by which the estimated concentration is multiplied after exponentiation. The smearing estimate constant is

$$ksm = 1$$

$$n$$

$$e\epsilon i$$

$$i = 1$$

$$n \sum (15)$$



where  $\varepsilon_i$  is the  $i^{\text{th}}$  residual from the regression model; this is equivalent to the mean of the exponentiated residuals.

Cohn et al. (1989) proposed a minimum variance unbiased estimator (MVUE)

where  $\hat{L}$  is the estimated load,  $\hat{C}$  the estimated concentration,  $RC$  refers to the rating curve or regression relationship between log-load and log-flow, and everything to the right of  $\hat{LRC}$  is the MVUE bias correction factor.  $gm$  is a Bessel function described in their paper,  $m$  is the number of observations used to establish the regression relationship minus the number of parameters in it,  $s^2$  is the variance of the residuals from the regression relationship between load and flow, and  $V$  is a "leverage" term which is a function of the values of the independent variables from which a specific load (or concentration) estimate is to be calculated. As a consequence, the MVUE bias correction factor is not a constant.

The authors show that this correction factor performs better than the alternatives presented above.

Unfortunately, it is also quite cumbersome to calculate, though they offer computer code to evaluate it. Furthermore, Thomas (1988) points out that bias correction methods, including this one, depend on the validity of generally untested hypotheses, particularly that of normally distributed residuals (in log space).

In a later simulation study based on data from several rivers tributary to Chesapeake Bay, Cohn et al. (1992) found that the MVUE estimator generally gave fairly accurate results. They concluded that it was fairly insensitive to modest violations of the assumptions of the regression approach.

Cohn (1995) states that the three bias correction methods give nearly identical results if:

1. the assumed linear model is approximately correct,
2. the regression model is based on 30 or more observations, and
3. the model is not being used to extrapolate beyond the range of data used to calibrate it.

He says further that if only the first condition is satisfied, the MVUE estimate is the best. If all conditions are satisfied, the QMLE is a good choice because it is relatively easy to calculate. If the regression residuals are not normally distributed, the smearing estimate may be preferred, but in this case one must verify that the use of the regression model is appropriate.

#### Selecting a load estimate method (quick tips)

We will not make a comprehensive assessment of the utility of the various estimating methods. However, the following rules of thumb are useful in selecting a method.

- Averaging techniques are simple and relatively robust although biases will result from inadequate representation of storm events. In instances where the correlation between discharge and concentration is positive (negative), loads will tend to be underestimated (overestimated). Weighted averaging is a simple device

that, to some extent, overcomes this drawback. A substantial decline in precision associated with averaging methods has been reported as the sample size  $n$  decreases (i.e. the time interval over which samples are obtained increases).

- Stratification can potentially improve both the accuracy and precision of load estimates.
- Ratio estimators work best when discharge and concentration are linearly related (passing through the origin) with non-constant variance. They are considered more suited to situations where there is less intensively sampled concentration data. The Beale ratio estimator is widely used.

The selection of an appropriate load estimation technique depends not only on the availability of concentration and discharge data, but also on the hydrological characteristics of the catchment being considered, time-scale of waterbody response to loads, the desired accuracy of estimates and the preferred complexity of the load estimation technique. No single technique has been found to be optimal in the literature, with all techniques having some disadvantages associated with their use. The choice of technique will depend on the characteristics of the catchment being considered, and the availability of data for that catchment.

#### ***Uncertainty***

Some methods of estimating loads have the additional desirable feature of providing a measure of the uncertainty of the load estimate. Furthermore, the uncertainty estimates of different load calculation methods cannot be directly compared, because they reflect different kinds of "error". Further still, the estimated "error" may be different from the error we are interested in.

For example, the uncertainty estimate for the Beale Ratio Estimator includes a contribution which is due to differences between individual daily loads and the mean daily load in each stratum. If we are interested in the annual load for that year, we would not consider this to be a source of error, but rather a part of the natural variation of the system we are studying. We would prefer to confine our notion of error to the difference between our estimated mean daily loads and the actual (but unknown) mean daily loads, a difference which is due only to sampling and analytical error.

For these reasons, uncertainty measures do not provide a reliable way of choosing between methods. Consequently, evaluation of load estimation methods must rely on comparative studies in which several methods are used to calculate loads from the same data, and the results are compared with the "true" load which is independently known.

There are two basic approaches to simulation: systematic subsampling and Monte Carlo simulation. Systematic subsampling involves taking a dense dataset and splitting it into subsamples of the same size. The "true" load is the load based on the entire dense dataset.

Monte Carlo simulation involves random sampling of an empirical distribution of observations, or sampling of parametric distributions generalised from such observations, to produce any desired number of simulated sample sets. Typically 500, 1000, or 2000 sample sets are generated and evaluated. The "true" load is the load based on the entire dataset or calculated from the parametric distributions used.

---

There are advantages and disadvantages to each method.

Not surprisingly, accuracy and precision increase with increased frequency of sampling.

***Types of uncertainty affecting load estimation***

Type of Uncertainty	Definition	Load Examples
Epistemic		
Measurement Error	Occurs because measuring equipment and observers are imperfect resulting in random variation in a quantity.	
Systematic Error	Occurs when measurements are biased.	Wrong instrument calibration.
Natural Variation	Environmental change with respect to time, space or other variables that is difficult to predict.	Variation in flow seasonality. Landforms.
Inherent Randomness		
Model Uncertainty	Occurs because models are simple abstractions of reality and sometimes significant factors causing variation are missed or factors are not properly represented or parameterised.	Wrong load algorithm. Not accounting for estuaries or groundwater.
Subjective Judgement		Samples at wrong time or location resulting in unrepresentative sampling

The accuracy of the load estimate will depend on the proper application of the most appropriate method, little can be done to compensate for data sets which contain an insufficient number of observations collected using an inappropriate sampling design. Many load estimation programs choose monthly or quarterly sampling with no better rationale than 'it's convenient'. To avoid this, the sampling pattern, which will be needed for load estimation, must be established in the initial planning process, based on quantitative statements of the precision required for the load estimate. If load estimates are to be made, determine the precision needed, based on the uses to which they will be put. For example, "I want the load estimates to be within 5% of the true loads in 90% of the years for which calculations will be made". The resources necessary to carry out the sampling program must be known, and budgeted for, from the beginning.

***Types of uncertainty***

Overall, three sources of uncertainty contribute significantly to overall uncertainty in estimating sediment and nutrient load. These are 'knowledge uncertainty', 'stochastic uncertainty' and 'data uncertainty'.

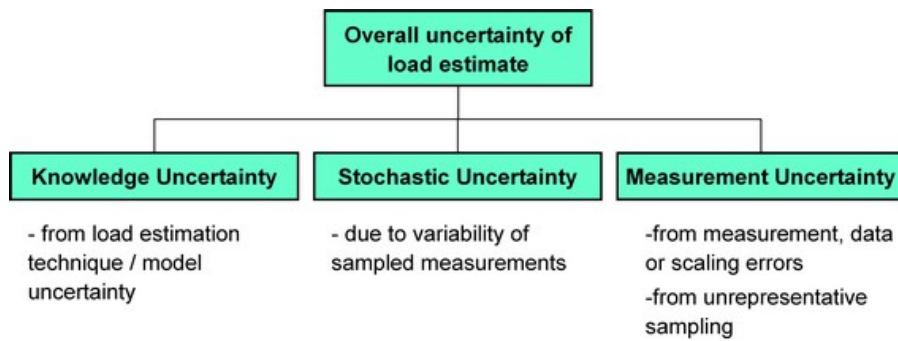


Figure 6.2: Types of Uncertainty

The annual load estimates could show significant variation due to the estimation technique selected. Unless other overriding factors are considered to be relevant, any of the estimates are equally legitimate as estimates of the 'true' load. However, the 'true' load is not known, and therefore, the selection of estimation technique is one source of uncertainty in load estimation. This source of uncertainty has been labelled 'knowledge uncertainty'. Knowledge uncertainty can be reduced through an increased understanding of the pollutant wash-off and transport processes. In general, for sites with limited high flow samples, methods that do not account for flow stratification will be downwardly biased. Likewise, for sites with highly seasonal concentration variation (e.g. irrigation areas), methods that do not account for time stratification will lead to imprecise and biased results. Such knowledge uncertainty can be incorporated by the specification of user-defined weights reflecting the user's preferences or beliefs in each of the methods.

In addition to knowledge uncertainty, 'stochastic uncertainty' also needs to be considered. Stochastic uncertainty is described by the deviation of water quality concentrations from any assumed value (e.g. a mean) and is usually represented by estimates of variance. This stochastic uncertainty increases when the variances of sampled concentrations are increased (scaled) to correspond to population variances.

## 6.4 Appendix 4 - Decision tree with preferred load estimation methods

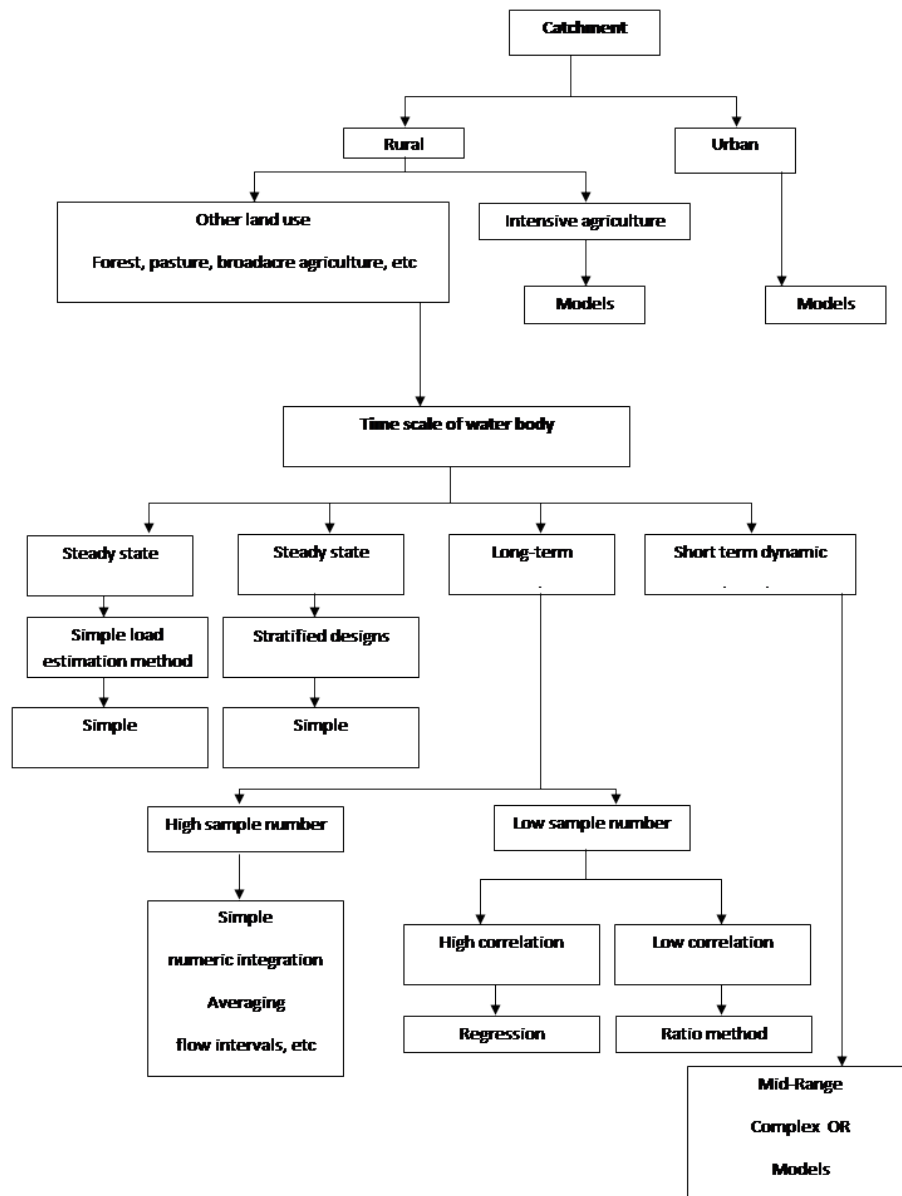


Figure 6.3: Decision Tree

## 6.5 Appendix 5 - Events load estimation

From an environmental perspective, storm-driven events are the most significant factor affecting the fate and dispersal of pollutants discharged to the riverine or marine environment. Most pollutants are generated and moved during events. Therefore estimation of loads in an event is important, and the techniques used are slightly different from the techniques used for estimating loads from long-term time series data.

The sampling strategy is based on the assumption that most of the load occurs in a short period of time during storm runoff events, and that accurate loads can be obtained by frequent sampling primarily during that period of time. The strategy also assumes that if samples are sufficiently close together in time, the concentration and flow between samples will change with time.

Details of events load estimation techniques available in the Loads tool are given below.

### Simple integration methods

A simple numeric integration technique can be used to calculate the load in an event.

This approach assumes that sampling during high flow periods is frequent enough that the sampled fluxes (concentrations and flows) closely match the continuous pattern of the actual fluxes, and in particular that the peak flux for each storm is not too badly underestimated. The likely validity of this assumption should be tested by comparing the flow pattern based on the flows at the times of the samples with that based on all of the flow data.

Express the time of each pollutant sample and each flow observation as decimal days of the year. For each pollutant sample, establish a time window, which starts halfway between the sample and the previous sample, and ends halfway between the sample and the next sample. The time window for the first sample during a storm should include only the interval based on the time to the next sample (i.e. should start at the beginning of the storm); the time window for the last sample during a storm should include only the interval based on the time from the preceding sample (i.e. should end at the end of the storm); see Figure 1. Multiply the flow at the time of the sample by the concentration at the time of the sample and the time interval for the sample, and use a related conversion factor as appropriate to convert loads unit into metric tonnes. Add all these loads calculated for each time interval together for the event load estimation.

Numeric integration is only satisfactory if the sampling frequency is high. This method, and particularly the sampling strategy, assumes that flows are highly variable and that concentrations increase with flow.

---

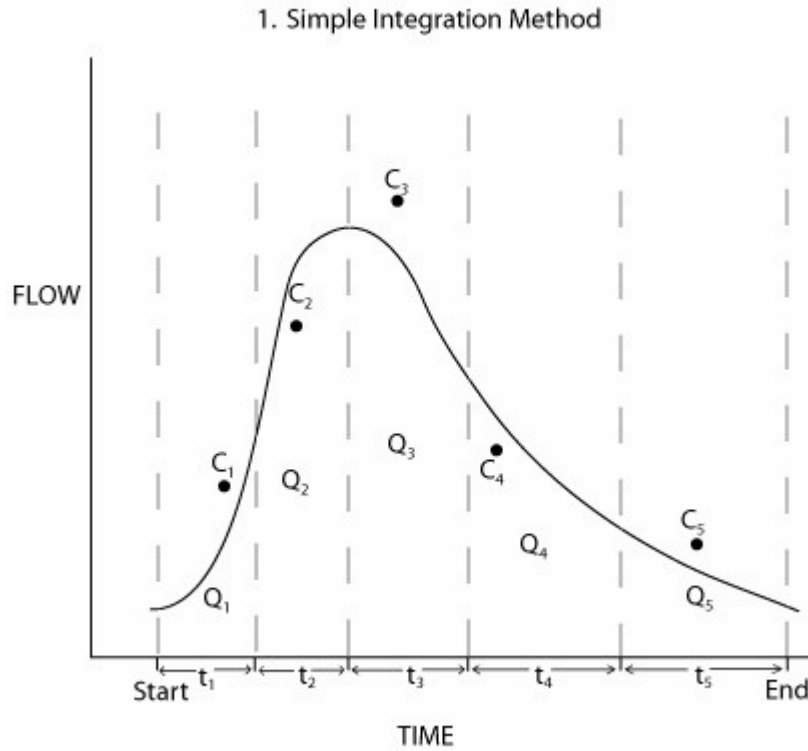


Figure 6.4: Simple Integration Method

Flow volume ( $Q$ ) for each time step can be estimated by integrating flow rate ( $q$ ) over time,

$$Q1 = q \cdot t1$$

$$\text{Load} = \{C1 \cdot Q1 + C2 \cdot Q2 + C3 \cdot Q3 + C4 \cdot Q4 + C5 \cdot Q5\} \cdot \text{conversion factor}$$

Use a conversion factor to convert the load to a mass unit. If the concentration is in mg/L and flow volume is in cubic metres, use the following conversion factor to convert load into metric tonnes:

$$\text{conversion factor} = 0.000001$$

$$k \sum_{i=1}^n \frac{c_i}{n} \sum_{i=1}^n \frac{q_i}{n} = k \bar{c} \bar{q}$$

where

$c_i$  is the  $i^{\text{th}}$  sampled concentration

$q_i$  is the  $i^{\text{th}}$  sampled discharge (flow)

$\bar{c}$  is  $n$  concentration measurements

$\bar{q}$  is  $n$  discharge measurements

$k$  is number of time intervals in period

#### **Linear interpolation of concentration data**

Loads for an event can be generated using linear interpolation between samples. Measured concentration values are linearly extrapolated to represent flow measurements. In this technique, assumptions are made about how concentrations vary in time between samples and how we can linearly interpolate between two consecutive concentrations to obtain a time series of concentrations (e.g. one-hour time step.)

However in this linear interpolation a problem will occur if you do not have the concentration values for the very start of the event and the end of the event. Generally, an event will begin when water flow exceeds the base flow, and the event will come to an end when the flow rate is reduced to the base flow level. In order to calculate the load for entire event you need to consider the entire time duration of the event.

Most of the time, it would be difficult to sample the concentration right at the point where the event starts and finishes, and at the peak in concentration, which must occur sometime during the storm runoff event. Therefore some assumptions and pre-data preparation would be required to estimate loads during the event using a linear interpolation method, as flows and concentrations can never be sampled at exactly the right moment.

Use the first and last-sampled concentration points in the event and tie down the values between these two points for calculation of loads (Figure 2a). In this option, you will underestimate the event load, as you have missed the pollutant loads before the first sample and after the last sample in the event.

---



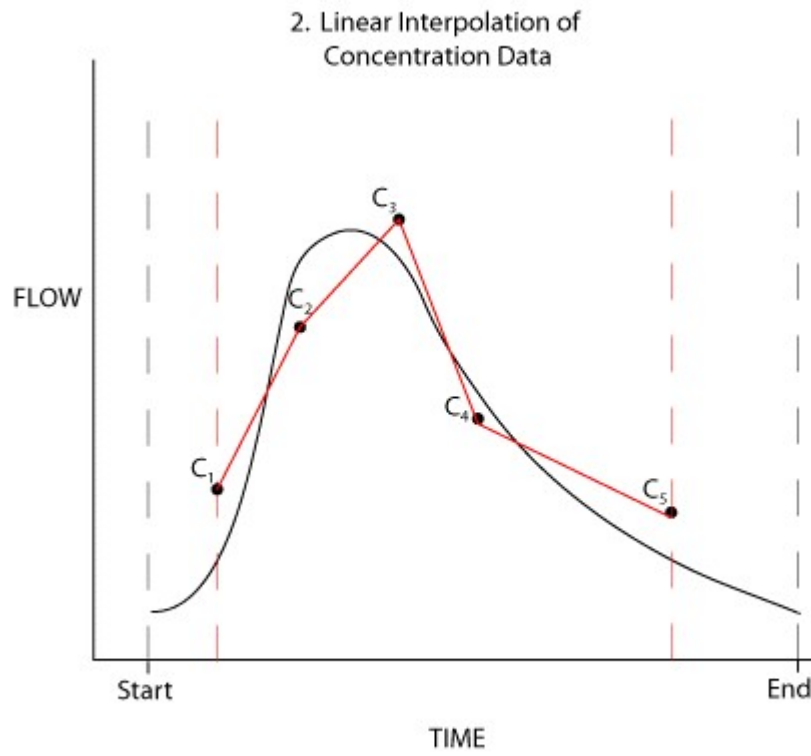


Figure 6.5: Linear Interpolation of Concentration Data

Add two new concentration points at the beginning and end of the event. This can reduce the errors in event-load estimation. As shown in Figure 2b, you can add an appropriate concentration value to the start of the event and the end of the event. Perhaps if you have a DWC (dry weather concentration) value for the stream, it would be appropriate use the DWC value as the concentration at the start and end of the event. In this case you need to prepare your input dataset by adding two new data points.

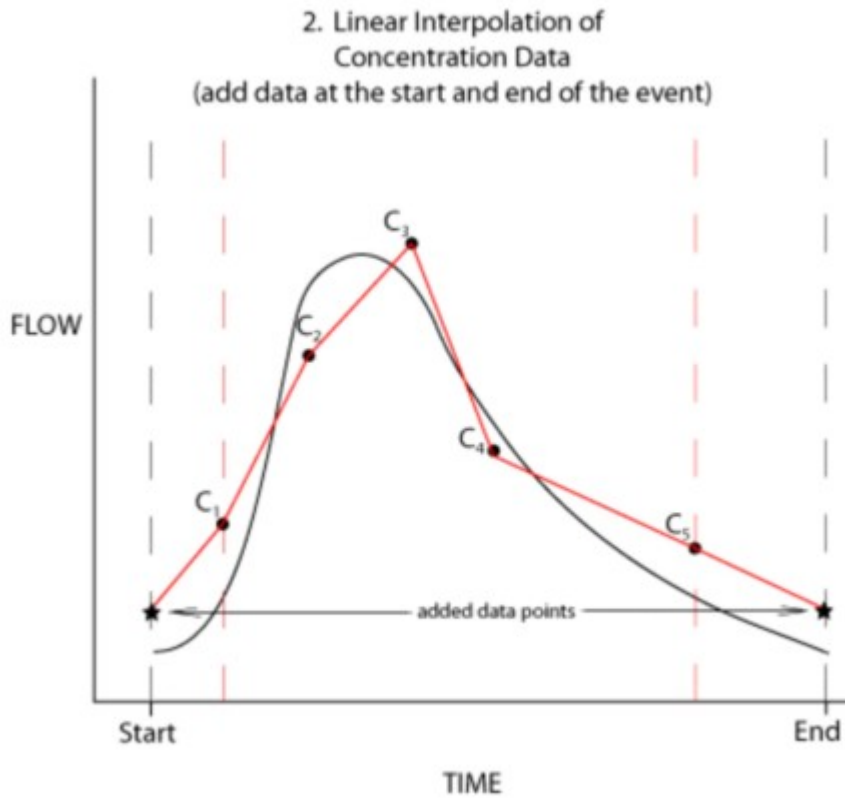


Figure 6.6: Linear Interpolation of Concentration Data Option Two

$$\sum_{j=1}^n \frac{c_j + c_{j+1}}{2} \cdot q_j$$

where

$c_j$  is  $j^{th}$  sample concentration

$q_j$  is inter-sample mean flow

#### Beale ratio estimator

The Beale ratio method has been designed for estimating pollutant loads on an annual or long-term basis. This method has been recognised as one of the best methods to estimate loads with a reasonable accuracy. A modified version of the Beale ratio method can be used to estimate loads in an event.

For times in which samples are taken, the hourly load is calculated as the product of concentration and flow, and the mean of these loads during the event can also be calculated (Figure 3). The mean hourly load is then adjusted by multiplying it by a

flow ratio, which is derived by dividing the average flow for the event as a whole by the average flow for the hours in which pollutant samples were taken. A bias correction factor is included in the calculation, to compensate for the effects of correlation between discharge and load. The adjusted mean hourly load is multiplied by the event duration in hours to obtain the event load.

Ratio estimators assume that there is a positive linear relationship between dependent and independent variables, which passes through the origin. Also, if the variance of the dependent variable is proportional to the magnitude of the independent variable, the ratio estimator is known to be the best linear unbiased estimator, i.e. the most precise among the class of unbiased estimators that assume a linear relationship. A bias correction factor is included in the calculation, to compensate for the effects of correlation between discharge and load.

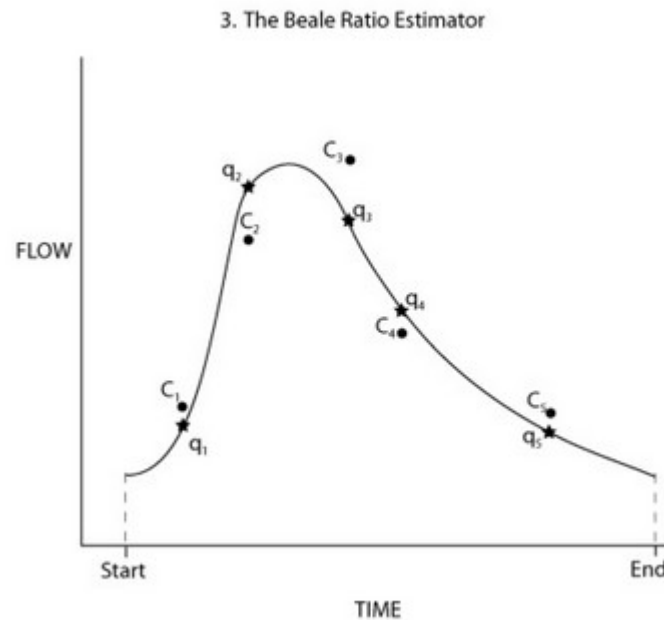


Figure 6.7: The Beale Ratio Estimator

Using the measured concentration values calculate hourly loads for each sample.

For sample C1 calculate the load by multiplying the flow rate in cubic metres per hour. Similarly estimate the loads for each sample for the entire event. Then calculate the average hourly load by adding all loads and dividing by the number of samples.

$$\text{Average hourly load} = (C1 \cdot q1 + C2 \cdot q2 + C3 \cdot q3 + C4 \cdot q4 + C5 \cdot q5) / 5$$

Then you need to calculate average hourly flow rates used in each sample point and the average hourly flow for the entire event.

$$\text{Flow ratio} = \frac{\text{Average flow for the entire event}}{\text{Average flow for the hours in which pollutant samples were taken}}$$

$$\text{Adjusted hourly load} = \text{Average hourly load} \cdot \text{Flow ratio}$$

A bias correction factor, estimated using the number of samples and flow variability, is used to improve the accuracy of the estimates. However, if the number of samples is high, the bias correction factor will become 1 or close to 1.

The total load for the entire event is estimated by multiplying adjusted hourly load by flow duration in hours for the entire event.

$$Q \left( \frac{\bar{l}}{\bar{q}} \right) \left\{ \frac{1 + \frac{1}{N} \frac{\rho \sigma_L \sigma_Q}{l \bar{q}}}{1 + \frac{1}{N} \frac{\sigma_Q^2}{\bar{q}^2}} \right\}$$

#### Concentration power curve

You can use a regression approach to develop a relationship between concentration and flow, based on the samples taken (see Figure 4). The measured (sampled) data are used to establish a regression relationship of the power curve. A number of regression techniques such as simple linear regression, power curve, quadratic function, or the USGS seven parameter techniques, are being used in load estimations. However, in this tool we use only a simple power function to develop the relationship between flow rate and concentrations. Site-specific monitoring data should be used whenever possible to check the accuracy of the predictions.

$$c = aq^b$$

where

$c$  is concentration

$a$  is coefficient

$q$  is flow

$b$  is power

A regression relationship is developed between concentration and flow, based on the times at which samples are obtained. The regression relationship may be based entirely on the event samples. Once the regression relationship is established, it can be used to estimate concentrations for each flow period in which a sample was not taken. The total load is calculated as the sum of the loads obtained by multiplying the measured or estimated concentrations by the flows.

Note: The time intervals for concentration data have to be equal to or greater than the flow data intervals.

## 6.6 Appendix 6 - Protocols for the optimal measurement and estimation of nutrient loads

Error Approximations

Technical Report

Prof. David R. Fox

Australian Centre for Environmetrics

Report 03/05

April 2005

### **Acknowledgements**

Financial support for this work has been provided under the Victorian Government's Regional Catchment Investment Plan (Project No. WG-0405-10.45); Queensland EPA and the Department of Environment and Heritage (Coastal Catchment Initiative); the University of Melbourne (Centre for Water and Landscape Management); Southern Rural Water; and Goulburn-Murray Water. Additional support has been provided by the University of Melbourne (Department of Civil and Environmental Engineering and the School of Anthropology, Geography, and Environmental Science), CSIRO Land and Water, and Melbourne University Private.

The author is particularly grateful to Mr. Geoff Gordon of the NSW EPA who made many valuable contributions to the derivation of equation 6 and the material in Appendix A. The contributions of Prof. Barry Hart, Dr. K.S. Tan and Dr. Teri Etchells who reviewed an earlier version of this report are also acknowledged.

### **Introduction**

The accurate estimation of total loads of sediments and nutrients is a problem that is attracting considerable attention among natural resource managers, environmental protection agencies, governments, landowners, and the general community. The delivery of sediments from Queensland catchments has been identified as a threat to the ecosystem of the Great Barrier Reef, while point and diffuse sources of land-based nutrients are implicated in the increased frequency and severity of algal blooms in water bodies around the country. Accordingly, there has been a growing trend towards the expression of aspirational and compliance targets for nutrients and sediments in terms of either a relative or absolute reduction in total load. For example, a 20% nutrient reduction target has been imposed on Queensland catchments impacting the Great Barrier Reef while the Victorian EPA has required a 40% reduction in the total phosphorous load from the McAlister Irrigation District by 2005 and a commensurate 40% reduction in total nutrient loads to the Gippsland Lakes by 2022. As noted by Henderson and Bui (2004), the quantification of errors and uncertainty is particularly important in the context of ecological risk assessments as a failure to do so may lead to risks being significantly under or over-estimated.

This report focuses on the quantification of errors associated with a number of common load estimation techniques. We also point out the duality between simple mean-based load estimators and ratio estimation techniques.

### **Load estimation**

A list of some 24 computational techniques for estimating a load was provided in Letcher et al. (2002). Most of these formulae can be classified as belonging to one of the groupings: mean-based estimators; ratio estimators; and regression estimators. In this

---

paper we consider a class of load estimators given by equation 1.

$$\hat{L} = K \left( \sum_{i=1}^{n_c} w_i c_i \right) \left( \sum_{j=1}^{n_q} v_j q_j \right) \quad (1)$$

where

$q_i$  is a measured concentration on the  $i^{th}$  occasion

$c_i$  is a measured flow on the  $i^{th}$  occasion

$w_j$  and  $v_j$  are weights

$K$  is a constant that reconciles the sampling time-step with the period of interest (eg. if concentrations and flows represent daily values and an annual load estimate is required, then  $K=365$ ).

#### Theoretical mean and variance

Before turning our attention to the properties of load estimators, it will be useful to develop some theoretical results for the expected value and variance of a load under certain distributional assumptions. In what follows we assume (not unreasonably), that the distribution of concentration ( $C$ ) and flow  $q_i$  are well described by the bivariate lognormal distribution given by equation 2 and that load,  $L = CQ$ .

$$f_{C,Q}(c, q) = \frac{1}{2\pi c q \sigma_C \sigma_Q \sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \left( \frac{\ln(c) - \mu_C}{\sigma_C} \right)^2 - 2\rho \left( \frac{\ln(c) - \mu_C}{\sigma_C} \right) \left( \frac{\ln(q) - \mu_Q}{\sigma_Q} \right) + \left( \frac{\ln(q) - \mu_Q}{\sigma_Q} \right)^2 \right] \right\}$$

where  $\mu$  and

$\sigma$  are the mean and standard deviation of the log-transformed data and

$\rho$  is the correlation between log concentration and log flow

Fox (2004) showed that the expected load is given by equation 3.

$$E[L] = \exp \left\{ (\mu_C + \mu_Q) + \frac{1}{2(1-\rho^2)} [(\rho\sigma_Q + \sigma_C)^2 + (\rho\sigma_C + \sigma_Q)^2 - 2\rho(\rho\sigma_Q + \sigma_C)(\rho\sigma_C + \sigma_Q)] \right\} \quad (3)$$

Furthermore, it can be established that the second (uncorrected) moment is:

$$E[L^2] = \exp \left\{ 2(\mu_C + \mu_Q) + \frac{2}{(1-\rho^2)} [(\rho\sigma_Q + \sigma_C)^2 + (\rho\sigma_C + \sigma_Q)^2 - 2\rho(\rho\sigma_Q + \sigma_C)(\rho\sigma_C + \sigma_Q)] \right\} \quad (4)$$

And so the variance is given as:

$$\text{Var}[L] = E[L^2] - (E[L])^2 \quad (5)$$

#### Uncertainty in load estimates

We next turn our attention to sampling properties of the estimator given by equation 1. In particular, it can be shown that an approximation to the variance is:

$$\text{Var}[\hat{L}] = K^2 \left( \sum_{i=1}^{n_c} w_i^2 \right) \left( \sum_{j=1}^{n_q} v_j^2 \right) \text{Var}[L] \quad (6)$$

For suitable choice of the weights  $w_i$  and  $q_i$  we can obtain variance approximations for a number of common load estimators. Furthermore, the duality between a ratio estimator of load and one obtained using flow-weighted mean concentrations can be established. These issues are covered under special cases 1-3.

**Special case #1: The naive estimator (average flow x average concentration)**

The simplest of all load estimators is a scaled product of the mean concentration and the mean discharge (flow). We refer to this as the "naive" estimator - its attractiveness lies in its computational simplicity, although serious biases (typically > 30%) result (Fox, 2004). The naive estimator is readily seen to be obtained by letting

$$w_i = \frac{1}{n_c}$$

and

$$v_j = \frac{1}{n_q}$$

and

$$\hat{L}_I = K \bar{C} \bar{Q} \quad (7)$$

and

$$\text{Var}[\hat{L}_I] = \frac{K^2 \text{Var}[L]}{n_c n_q} \quad (8)$$

**Special case #2: Load estimator using flow-weighted mean concentrations and unknown total discharge**

Unlike the naive estimator which assigns equal weight to each observed concentration, the flow-weighted mean concentration (fwmc) uses weights that are proportional to the magnitude of the associated flow. In this sense, the naive estimator may be thought of as a time-based average whereas the fwmc is a flow-based average. It is implicit in flow-weighted averaging that the flow and concentration data are contemporaneous whereas no such assumption was previously made. Thus,

$$n_c = n_q = n$$


---

and the weights for fwmc are

$$w_i = \frac{q_i}{\sum_{i=1}^n q_i}$$

$$v_i = 1 \quad \forall i$$

thus,

$$\hat{L}_2 = K' \frac{1}{\sum_{i=1}^n q_i} \left( \sum_{i=1}^n c_i q_i \right) \left( \sum_{i=1}^n q_i \right)$$

$$\hat{L}_2 = K' \sum_{i=1}^n c_i q_i \quad (9)$$

where

$K' = \frac{K}{n}$  (eg. if one month of daily concentration data are available to estimate an annual load using equation 9, then  $K=365$  and  $n=30$ ). The  $K'$  factor is needed in this case because the total discharge,  $\sum_{i=1}^n q_i$  is only known for the sample and not the entire period of interest. Furthermore,

$$\text{Var}[\hat{L}_2] = K'^2 \sum_{i=1}^n \left( \frac{q_i}{\sum_{i=1}^n q_i^2} \right) n \text{Var}[L]$$

$$= \frac{K^2 \text{Var}[L]}{n (\sum_{i=1}^n q_i)^2} \sum_{i=1}^n q_i^2 \quad (10)$$

It can be readily established that in the case  $n_c = n_q = n$ , the variance of  $\hat{L}_2$   $\hat{L}_1$ . To see this, we look at  $\text{Var}[\hat{L}_2] - \text{Var}[\hat{L}_1]$ .

$$\text{Var}[\hat{L}_2] - \text{Var}[\hat{L}_1] = \frac{K^2}{n (\sum_{i=1}^n q_i)^2} \sum_{i=1}^n q_i^2 \text{Var}[L] - \left\{ \frac{K^2 \text{Var}[L]}{n^2} \right\}$$

$$= K^2 \text{Var}[L] \left\{ \frac{\sum_{i=1}^n q_i^2}{n (\sum_{i=1}^n q_i)^2} - \frac{1}{n^2} \right\}$$

Hence,



$$\text{Var}[\hat{L}_2] > \text{Var}[\hat{L}_1]$$

$$\frac{\sum_{i=1}^n q_i^2}{n (\sum_{i=1}^n q_i)^2} - \frac{1}{n^2} > 0$$

$$\Rightarrow \sum_{i=1}^n q_i^2 - \frac{(\sum_{i=1}^n q_i)^2}{n} > 0$$

**Special Case #3: Load estimator using flow-weighted mean concentrations and known total discharge**

This case is identical to special case #2 with the exception that the fwmc is applied to the total (annual) discharge,  $\sum_{i=1}^K q_i$ . This, the weights are as before except that the  $v_j$  weights span the period of interest ( $j=1, \dots, K$ ) rather than the sample ( $j=1, \dots, n$ ). Thus,

$$\hat{L}_3 = \frac{\sum_{i=1}^n c_i q_i}{(\sum_{i=1}^n q_i)} Q \quad (11)$$

Where Q is the total (annual) discharge. Furthermore,

$$\text{Var}[\hat{L}_3] = \frac{K \sum_{i=1}^n q_i^2}{(\sum_{i=1}^n q_i)^2} \text{Var}[L] \quad (12)$$

Note, if we have sampling fraction  $f = \frac{n}{K}$ ;  $0 < f < 1$  then equation 10 can be written as

$$\text{Var}[\hat{L}_2] = \frac{\text{Var}[\hat{L}_3]}{f} \text{ and}$$

$$\text{Var}[\hat{L}_3] < \text{Var}[\hat{L}_2]$$

**The duality of the flow-weighted mean concentration load estimator and a ratio estimator**

Ratio estimation is a well known technique for potentially reducing the error (increasing the precision) of the estimate when an auxiliary variable that is correlated with the variable of interest is available. A full treatment of ratio estimators is given in Cochran (1977). In the present context, a ratio estimator is formed by assuming the ratio of the total load for the sample to the total discharge for the sample is the same as the corresponding quantities over the period of interest. That is

$$\frac{l}{q} = \frac{L}{Q}$$

Where  $l$  ( $L$ ) is the sample (population) load and  $q$  ( $Q$ ) is the sample (population) discharge. The ratio estimator is then

---

$$\hat{L}_{\text{ratio}} = \left( \frac{l}{q} \right) Q \quad (13)$$

Expanding equation 13, we have

$$\hat{L}_{\text{ratio}} = \frac{(\sum_{i=1}^n w_i c_i) (\sum_{i=1}^n v_i q_i)}{\sum_{i=1}^n q_i} \cdot \sum_{j=1}^K q_j \quad (14)$$

and letting

$$v_i = 1 \quad \forall i \text{ and}$$

$$w_i = \frac{q_i}{\sum_{i=1}^n q_i} \text{ we see that}$$

$$\hat{L}_{\text{ratio}} = \hat{L}_3$$

#### An example

We consider the estimation of the total phosphorous (TP) load in a drain (designated CG3) in Gippsland, Victoria during the 2004 irrigation season. The availability of daily flow and TP measurements enables us to compute the 'true' load as 5,517.10 kg. A random sample of  $n=29$  observations were taken and the results used to demonstrate the methods outlined in this paper. The parameters given in table 1 were estimated from the log-transformed flow and concentration data.

	Log-flow	Log-concentration
$\mu$	2.5561	-0.02834
$\sigma$	0.6706	0.8008
$\rho$		0.482

By substituting the parameter estimates in table 1 into equations (3) and (4) we obtain (using equation (4)) estimate the load variance to be  $\text{Var}[L] = 3132.863$ . We next obtain the load estimates using methods 1 - 3.

Our data yield:

$$n = 29,$$

$$\bar{c} = \frac{1}{n} \sum_{i=1}^n c_i = 1.17225, \text{ and}$$

$$\bar{q} = \frac{1}{n} \sum_{i=1}^n q_i = 11.4015$$


---

The duration of the irrigation is such that  $K=279$  days. Thus,

$$\hat{L}_1 = 279 \cdot 1.17225 \cdot 11.4015 = 3728.95\text{kg}$$

Compared to the 'true' load of 5517.10kg,  $n = 29$  is seen to underestimate the true load by 33%. This overestimation is a consequence of the high (positive) correlation between log-concentration and log-flow. A bias correction factor (Fox 2004) can be applied in attempt to reduce this effect. In this case an improved estimate is obtained by multiplying  $\hat{L}_2$  by

$\exp\{\text{Cov}[\ln C, \ln Q]\} = \exp(\rho \sigma_c \sigma_q) = 1.2954$ . This gives a modified total load of 4830.5kg which has reduced the bias to 13%.

From equation (8) we have

$$\text{Var}[\hat{L}_1] = \frac{279^2}{29 \cdot 29} \text{Var}[L] = 350220.28$$

and hence

$$SE[\hat{L}_1] = \sqrt{\text{Var}[\hat{L}_1]} = 591.8$$

From equation (9)

$$\hat{L}_2 = K' \sum_{i=1}^{29} c_i q_i$$

$$= \frac{279}{29} \cdot 434.410 = 4179.32\text{kg}$$

Compared to the 'true' load of 5517.10kg,  $n = 29$  is seen to underestimate the true load by 24%.

From equation (12) we have

$$\begin{aligned} \text{Var}[\hat{L}_2] &= \frac{279^2 \text{Var}[L]}{(29) \left( \sum_{i=1}^{29} q_i \right)^2} \sum_{i=1}^{29} q_i^2 \\ &= \frac{279^2}{29} \frac{(3132.863)(4553.11)}{330.643^2} = 350220.28 \end{aligned}$$

and hence

---

$$SE[\hat{L}_2] = \sqrt{\text{Var}[\hat{L}_2]} = 591.8$$

From equation (11)

$$\begin{aligned}\hat{L}_3 &= \frac{\sum_{i=1}^{29} c_i q_i}{\sum_{i=1}^{29} q_i} Q \\ &= \frac{(434.410) \cdot (5244.32)}{330.643} = 6890.17\text{kg}\end{aligned}$$

Compared to the 'true' load of 5517.10kg,  $\hat{L}_1$  is seen to overestimate the true load by 25%. From equation (12) we have

$$\begin{aligned}\text{Var}[\hat{L}_3] &= \frac{279 \sum_{i=1}^{29} q_i^2}{(\sum_{i=1}^{29} q_i)^2} \text{Var}[L] \\ &= \frac{279 \cdot (4553.11) \cdot (3132.863)}{330.643^2} = 36402.82\end{aligned}$$

and hence

$$SE[\hat{L}_3] = \sqrt{\text{Var}[\hat{L}_3]} = 190.8$$

## 6.7 Appendix 7 - Assessment of Water Quality Data For Temporal Trends

### 6.7.1 Introduction

The identification of trends in water quality can 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. These networks are also useful for other purposes eg. describing water quality, assessing compliance with regulatory objectives, comparing sites/regions, investigating processes or factors affecting water quality and estimating loads.

Water quality can vary spatially and temporally for many quite natural reasons. For the data from a water quality sampling program to be useful it must adequately describe the range and the norm of this natural variation. This is a fundamental requirement that 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 of the above purposes.

**To successfully investigate water quality trends one needs to first of all assess the data set being used to make sure it is suitable for trends analysis.**

## 6.7.2 Water quality variability

### 6.7.2.1 Spatial Variation

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.

Spatial variability can operate on a number of scales eg.

#### *Within a water body*

- along the length or width of a river.
- different parts or arms of a lake, or strata within a lake.

Rivers are generally assumed to be uniformly mixed, however there are circumstances when this might not be true, eg large pools may be different to riffles, especially during drought conditions when pools could stratify and current induced mixing could be very poor. Lakes are often poorly mixed especially if they are deep, dendritic in nature, stratified or have a number of inputs of varying character.

#### *Regional*

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. Other factors such as altitude, gradient, stream size, geology and topography also have an important influence on water quality. Human influences (landuse, impoundments, discharges) are superimposed upon and often correlate with many of the above factors.

All of these factors can interact to produce a broad range in waterbody types and water quality. Consideration of the above factors and the scale over which the monitoring program is operating (within a catchment, regional, statewide or nationwide) is an important factor affecting site selection and the density of sampling sites. Sites could be stratified to concentrate on certain aspects of the range in spatial variation or could attempt to represent the full range of variation.

Adequate coverage of spatial variation including natural and affected sites is essential not only for descriptive purposes, but such information as, for example, geographic patterns in trends, which can give important clues to their causes.

### 6.7.2.2 Temporal Variation

Cyclic variation in water quality can operate on a number of scales eg. daily, seasonally and over a number of years, decades or even longer. Superimposed upon this cyclic behaviour is a strong element of randomness or "chaotic behaviour". All of these sources of temporal variation can potentially be confounded with true temporal trends.

The daily cycle in light and temperature can have a strong affect on many water quality parameters (including biotic factors which can affect water quality). Therefore one needs

---

to ensure that the time of day of sampling is not a significant factor that could be producing consistent differences between sites or at the same site over time. This means that sampling time would need to be standardised or better still randomised throughout all of or part of the day, or at very least recorded.

To adequately define seasonal cycles or within year variations requires samples to be taken frequently over a number of years. 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, and the longer the interval the more important it is to pay attention to randomisation of the sampling time, and the longer it will take to accumulate sufficient data to describe seasonal and inter year variation and detect trends.

***How much data does one need to detect a trend?***

With three years of data it would be possible to obtain statistically significant results but such trends would mean little because of the high inter year variation found for many water quality indicators and the tendency for dry years and wet years to occur in runs. The frequency of sampling is the single most critical aspect of water quality network design affecting the power of a program to detect trends. An analysis of the effects of sampling frequency on the ability to detect trends undertaken for the Victorian Water Quality Monitoring Network (VWQMN) (EWQMC 1996) concluded that quarterly sampling had very little power to detect trends requiring at least 20 years of data to detect trends of high magnitude. Monthly sampling had a significantly increased power to detect water quality trends with a further significant increase in power occurring from monthly to fortnightly. Higher sampling frequencies provided little additional power but significantly increased the risk of the data being seriously affected by serial correlation. Lettenmaier (1978) recommends that the best sampling frequency for a trend detection program would be at least monthly (for sufficient statistical power) and not more than fortnightly (from considerations of serial correlation).

Investigations done for the review of the VWQMN (EWQMC 1996) indicated that at least 8 years of monthly data is needed to detect a change in the mean of one standard deviation. With five years of data only trends of high magnitude would be detectable. Ten years of data is however generally considered as a minimal requirement for trend analysis. As data sets extend into decades it becomes possible to detect trends of even small magnitude with a high degree of statistical certainty.

Ten years is still a short data set for trend analysis and it cannot be assumed that trends seen (or not seen) through that window will extend beyond that window. The longer the data set the more clearly we will be able to see trends and long term cycles and hopefully distinguish one from the other.

### **6.7.3 MAIN STEPS IN ASSESSMENT OF WATER QUALITY DATA FOR TRENDS**

A systematic approach to assessing water quality data for temporal trends is recommended so as to avoid or control 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

#### 6.7.3.1 Data Validation

Before assessing data for trends in water quality it is important to ensure that the data is sound otherwise all one may be doing is detecting trends in the procedures and methodology employed in collecting data. Ideally a data set for trend testing should be collected in a consistent manner, with few missing values. Well documented quality control procedures need to be in place at all stages from sample collection to analysis to ensure the quality of the data and that the effects of procedural changes can be tracked.

##### *Detection limits*

One needs to assess the data to see if detection limits exist to any great extent in the data set. This will vary for different indicators with some indicators being little or not affected by detection limits (eg pH and temperature) whereas others (eg metal toxicants) could have the majority of data points below the detection limit. If a large proportion of the data is below the detection limit it may not be possible to assess for trends. Applying statistical tests for trend to a data set with many detection limits, in an uncritical manner, will result in false trends being detected or real trends being obscured. If the detection limit varies this can create problems for trend detection, particularly if the change in detection limit is systematic rather than random. Ideally the analytical method chosen should be of sufficient sensitivity to ensure that the majority of the data points are above the detection limit, and the detection limit should not vary.

Choosing between various formal treatments of detection limit data is difficult because the data have been tampered with in only in one part of the range, i.e. near zero. Uncertainty applies to all observations regardless of where they lie along the range of the data. If the uncertainty estimate used to calculate the detection limit is used to round the data along its full range, then the values below detect get rounded to zero. Values above the detection limit are placed into categories with a bandwidth of twice the uncertainty estimate and are rounded to the mid point of the band (as are values below detect which are rounded to the mid point of their band i.e. zero). Use of such a rounding approach does not affect the mean but the variance is a little larger. If the differences between rounded and unrounded data are assumed to have a uniform distribution then that variance can be subtracted from the variance of rounded data to give back the variance of the original data (Rob Goudey, VIC EPA statistician pers. Comm.).

Non-parametric statistical tests for trend can be used for datasets that have detection limits. Such tests assume that the detection limit does not vary, if it does vary then the highest detection limit has to be applied to all values, and all values below this limit are recorded as below detect and are treated as ties (Hirsch and Slack, 1984). If a data set has a very small number of values with a detection limit different to the usual detection limit (particularly if it is much higher than normal) it may be better to simply exclude those values from the analysis, that is treat them as missing values. If there has been a systematic change in a detection limit such as might happen if there has been a method change or a progressive improvement in detection limits then that data set will

---

need careful treatment before it can be used for trend analysis. Such changes are often difficult to detect without careful examination of the data.

#### ***Consistency of units***

The units in which the data are reported need to be checked for consistency throughout the data set. A common error in the use of units exists with units for Electrical Conductivity (EC) where the two commonly used units differ by a factor of ten eg 1 mS/m = 10 uS/cm. The muddling of the two sets of units can often go undetected when data is entered and can produce sharp jumps or what appear to be extreme outliers in the data.

#### ***Consistency of data collection and methodology***

All aspects of sample collection and analysis have the potential to affect the measured values of indicators. Ideally a sampling program will have well defined protocols that have been designed to ensure that the results obtained are consistently correct, or at very least errors and biases introduced through sampling are known and controlled. For example indicators that are normally measured in the field are best measured in situ rather than in a sample bucket. Temperature can change very rapidly in a bucket (and hence any indicator affected by temperature) and in very dilute waters the carbon dioxide in exhaled air can significantly change the pH of water in a small container, thus the preference of individual operators regarding the way they take samples and in situ measurements could produce trends in the data. Any change in procedures need to be well documented so at the very least their impact upon the data can be assessed, if such documentation is not in place then interpretation of trends could be affected by uncertainty.

#### ***Time of day of sampling***

The normal daily variation of light and temperature can affect many water quality variables. Water temperature is an obvious indicator showing a strong diurnal trend. Dissolved Oxygen also shows a clear diurnal trend responding to changes in temperature as well as to changes to in-stream photosynthesis, which is strongly controlled by light levels. Photosynthetic activity also has a strong influence on pH. Other variables such as turbidity (affected by temperature) and stream flow (affected by water use of terrestrial plants) can also show clear diurnal trends.

Ideally, for a water quality monitoring program designed to detect trends over time, the time of day of sampling should not vary for a specific site. The time of day of sampling can vary between sites, but if it does it could confound comparisons between sites. To allow for trend testing and comparisons between sites the sampling time at a site should be randomised either over a 24 hour period, or within a given sampling time window. This is difficult to achieve and generally monitoring networks do not randomise the time of day of sampling. It is highly desirable to check how the time of day of sampling has varied for specific sites. This can be done simply by plotting time of day of sampling against date. Hopefully the time of day of sampling will not vary or if it does the variation is random. If variation in time of day of sampling is non- random and significant it may be necessary to use values for the indicator of interest that have been corrected for the time of day of sampling, or to at least be aware of how changes in the time of day of sampling could affect the value of an indicator.

#### ***Changes in sampling frequency, missing data and extra data***

---



The frequency of sampling should not vary and all seasons or periods of the year should be adequately sampled at the same intensity. Analysis of the distribution of sampling events between months, seasons and years should be undertaken to ensure an even distribution. Over sampling is as much of a concern as under sampling.

Large amounts of missing data can be problematic for trend detection; it can greatly weaken the power of statistical tests or prevent the use of many statistical tests, and can produce false trends or obscure trends that do exist. Ideally there should be no or very little missing data and what values are missing should be distributed randomly throughout the data set. If there is a long period of missing data between two periods of data collection then it would be best to treat the data set as comprising two discrete time periods of data collection and compare the means of indicator values between the two periods rather than attempt any trend analysis spanning the two time periods. The two data periods should be of similar length and cover complete years that is have no biases due to particular seasons not being adequately sampled. If there are many missing values then one will need to check if there are any biases or trends in the distribution of missing values. For example there may be a tendency to not sample in a particular season in which case it may be better to restrict the analysis to the seasons that are well sampled.

Sometimes it may not be a case of missing values but of a deliberate change in sampling frequency eg weekly to monthly, or there may be periods when extra samples have been included eg high flow event samples. In such cases one will need to randomly prune back the data set so as to produce a data set with a uniform sampling frequency. It is preferable to produce a data set of uniform frequency by pruning the data rather than aggregating the data. For example if part of a data set, that generally has a monthly sampling frequency, has been sampled weekly, then it would not be valid to produce a monthly sampling frequency by averaging four weekly samples as this would reduce the variance in the part of the data set so treated. One would need to randomly select a sample from each of the months that has been over sampled. If one leaves the extra samples in the data set and there is a significant portion of a data set that has been sampled at a higher frequency then it is likely that a greater number of extreme events will have been sampled in that part of the data set. This can create the illusion that the behaviour of the indicator has changed.

If there are extra data points in the data set related to events of interest then one really has two data sets mixed together and they need to be segregated to avoid biases.

#### ***Changes to analytical methods***

Different analytical methods have different strengths and weaknesses and are subject to different interferences. Therefore it is unusual for two different analytical methods to provide exactly the same answer in all circumstances. To control the potential impacts of changes in analytical method it is essential that all changes to methods are well documented and there are adequate periods of cross comparison. Ideally changes to analytical methods should be avoided; however this is not always possible. If there has been a change in analytical method then one would need to use the period of cross comparison to calibrate one method against another. Sometimes it is not possible to derive a satisfactory calibration and one simply has to accept that there is a break in the data set.

---

### 6.7.3.2 Visual Assessment and Exploratory Data Analysis

Visual Assessment and exploratory data analysis 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 and visual assessment involves using graphs to explore, understand and present data. It is an iterative process where graphs are plotted and refined so that important features of the data can be seen clearly. Visualising the data can readily identify outliers due to poor data quality, strong data independence/autocorrelation, step changes in methodology and recording technique and much more.

The time series plot is the most useful visual tool for analysing trend/change. The variable of interest is plotted against time as a scatter or line plot. A trend line can be fitted to the data. Techniques for fitting trend lines include moving average, linear regression, quadratic regression, and LOWESS smoothing. 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 eg. linear, monotonic, curvilinear, step trend and trend reversals.

#### 6.7.3.2.1 Things to watch for in time series plots:

##### ***Concordance between the smoothed curve and the data points***

Anomalies in the data can sometimes cause the smoothed curve to track through a region of the plot where there are few data points. For example if data are bi-modally distributed the smoothed curve may track between the two modes of data points and give a misleading impression of the trend. It may be necessary to investigate the reasons for data anomalies and perhaps partition or remove some of the data. For example base flow and high flow data could be plotted separately or values could be flow adjusted if it appears that flow may be having a strong influence upon the indicator of interest.

##### ***Log-normally distributed data or data with extreme outliers***

If the indicator of interest is log-normally distributed or has a small number of values of much greater magnitude than the majority of values then the majority of values will be plotted close to the baseline (zero on the y axis). This will make it very difficult to see what is happening with the majority of data points. This can usually be remedied by applying a log transformation to the data or by imposing an upper limit on the y (indicator) axis.

##### ***Tests for normality***

The trends tool offers a number of tests for normality. These can be useful as a part of the initial examination of the data. If data is not normally distributed a transformation could be applied to the data to normalise it. This can assist in viewing the data and when employing parametric tests for testing the significance of trends. Transformation of the data is not needed if non-parametric tests are used.

---

#### 6.7.3.2.2 Smoothing Techniques

The trends tool offers a number of basic smoothing techniques, which can be used to reduce the variability in the data and possibly reveal underlying trends or cycles in the data. In the context of temporal trend assessment data smoothing aims to produce a summary of how the indicator has changed over time. The noise in the data is smoothed over so as we are not distracted by it.

##### ***Simple Linear Regression***

Simple Linear Regression (SLR) can be used to fit a trend line to time series data that indicates the general movement in the data over time. A straight line could be drawn through the data points by eye but SLR is a more rigorous way of fitting a line to the data. Fitting a straight line to the data can be useful to indicate the overall tendency of the data (increasing or decreasing) over time, but does not reveal any other structure to the data eg. seasonal cycles, longer term cycles or trend reversals. It has the advantage that it is a simple technique, but should not be interpreted as implying that the indicator of interest is in some way dependant upon time only that it has a tendency to increase or decrease with time. One needs to keep in mind that other factors are probably affecting the indicator. For regression to be used the data must be normally distributed, if not a transformation will need to be applied to the data to normalise it. If the data is strongly non-normal the fitted line could miss the majority of the data points.

##### ***Moving Average***

The moving average can be used with time series data to smooth out short-term fluctuations or cycles and reveal longer-term cycles and trends. For example with time series data sampled on a monthly basis a moving average or total of 12 data points would remove any seasonality in the data and highlight longer-term trends and cycles.

##### ***LOWESS***

LOWESS/ LOESS (Locally Weighted Scatterplot Smoothing) is a useful technique for producing a smooth set of values from a time series or scatterplot where there is a lot of noise in the relationship. LOWESS is a robust technique when data are non-normally distributed. It makes no assumptions about the nature of the relationship between the indicator and time (eg linear or non linear) but lets the relationship emerge from the data itself. The model being fitted to the data gives more weight to points close to the point whose response is being estimated and less weight to those further away. The size of the smoothing window or tension value (proportion of the data being used expressed as a decimal from 0-1) can be selected with low values tending to produce a very non-smooth fit (tends to follow the data points closely) and high values producing a very smoothed fit. The value for the tension value is at the discretion of the analyst and depends on what features of the data that one is interested in. For example one could choose a low tension value that would track the within year variations in an indicator or one could choose a higher value that smooths out the within year variations but responds more to inter annual trends. A disadvantage of LOWESS compared to regression techniques is that it does not provide you with an equation to describe the fitted model.

##### ***Polynomial regression***

It may be noticed from a time series plot that an indicator seems to change in a non-linear or curvi-linear fashion with time. Polynomial regression is a technique that can

---

be used to fit curves to the data. In the context of temporal trend analysis polynomial regression fits data to the following equation.

$$Y = A + B.T + C.T^2 + D.T^3 + E.T^4$$

where Y is the value of the indicator and T is time. A,B,C,D,E ... are constants. Stopping at the T term would produce a straight line (simple linear regression). Including a  $T^2$  term would give a curved line and would allow you to model one turning point in the data. Including a  $T^3$  term would allow modelling of two turning points etc. The number of terms that one includes is largely a subjective choice. Lower order relationships such as  $T^2$  or  $T^3$  tend to produce very smoothed curves eg Fig X. High order relationships will follow movements in the data more closely. It is, however, generally recommended to avoid the temptation to over fit the data using high order polynomials as this can lead to results that are numerically unstable. If a more closely fitting relationship is desired for data that appears to change in a complex way with time it is recommended that a technique such as LOWESS be used.

#### 6.7.3.3 Statistical Tests For Testing The Significance Of A Trend

Once a possible trend has been identified from visual analysis of the data, one may want to test the statistical significance of a trend. A statistical test can be useful to sort out a trend where one may be uncertain of what the data is doing as the eye can be misled by extreme values in the data. Trend assessment should never rely on an uncritical application of statistical tests to large amounts of data without some visual assessment of the data and checking of the data. Otherwise all one is likely to do is to detect anomalies in the data and call them trends.

If the background assumptions of the particular statistical test you are using are not met eg using a parametric test when the data is clearly not normally distributed then erroneous results and conclusions can result. Serial correlation in the data also violates the assumption of independence of the data - that there is no short-term correlation between samples. When it occurs the p values calculated using statistical tests will be too low and trends may be falsely identified (Helsel, Mueller and Slack, 2006). Some basic but powerful approaches to testing the statistical significance of a trend are discussed below. The trends module has available a much larger range of tests, some of which may be useful or more appropriate in special circumstances.

##### ***Simple Linear regression***

SLR can be used to assess the statistical significance of the trendline fitted to the data, that is it can be used to test if the slope of the trend line is significantly different to zero. The outputs of a regression analysis also include a term known as R<sup>2</sup> or the coefficient of determination. R<sup>2</sup> is a measure of the proportion of the variation in the data that is explained by the regression. It varies from 0-1, with 1 indicating a perfect fit of the regression to the data.

##### ***Kendall and seasonal Kendall***

Kendall's Tau (Kendall 1975) is a basic non-parametric test for trend testing. Its only background assumption is that the random variable is independent and identically distributed (Smith, Hirsch and Slack 1982). 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 this test all possible pairs of data values are compared; if the later value in time is higher then a plus is scored; if lower then a minus is scored. If there is no trend in the data then the number of pluses and minuses should be about equal. If there is an increasing trend then values later in the time series will more likely be higher and many more pluses will be scored; if a decreasing trend then later values will more likely be lower and many more minuses will be scored. In the Seasonal Kendall test only data pairs from the same season are compared. This has the advantage in that the background assumption can now be significantly relaxed in that the random variable need only be identically distributed in like seasons (Hirsch, Slack and Smith 1982). Seasons can be defined by the investigator, but is usually individual months when monthly time series data are used, and thus only data pairs from the same month are compared. 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 (that is they change direction) or there are opposing trends in different seasons then the power of the test will be greatly weakened as such opposing trends will cancel out in the testing procedure.

#### ***Testing for a difference between two periods: Student's T test and the Rank Sum Test***

There are two situations when one would test for a step change rather than a linear or monotonic trend (Hirsch et al 1991). They are: 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 e.g. commissioning of a sewage treatment plant. In the latter case the decision about where in time the step has occurred should be based on prior knowledge of possible reasons for a step change rather than from examination of the data (i.e. the analyst notices a an apparent step but has no prior hypothesis that it should have occurred). To do otherwise would bias the significance level of a statistical test. If one wants to test for a step trend then a test that 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. The non-parametric tests can be used with normally distributed or non-normally distributed data. When comparing two data periods it is important that the data periods are multiples of complete years so as not to confound seasonal variability with differences between the two periods, and it is preferable that the two periods are balanced, i.e. about the same length.

#### **6.7.3.4 EXPLAINING THE TRENDS**

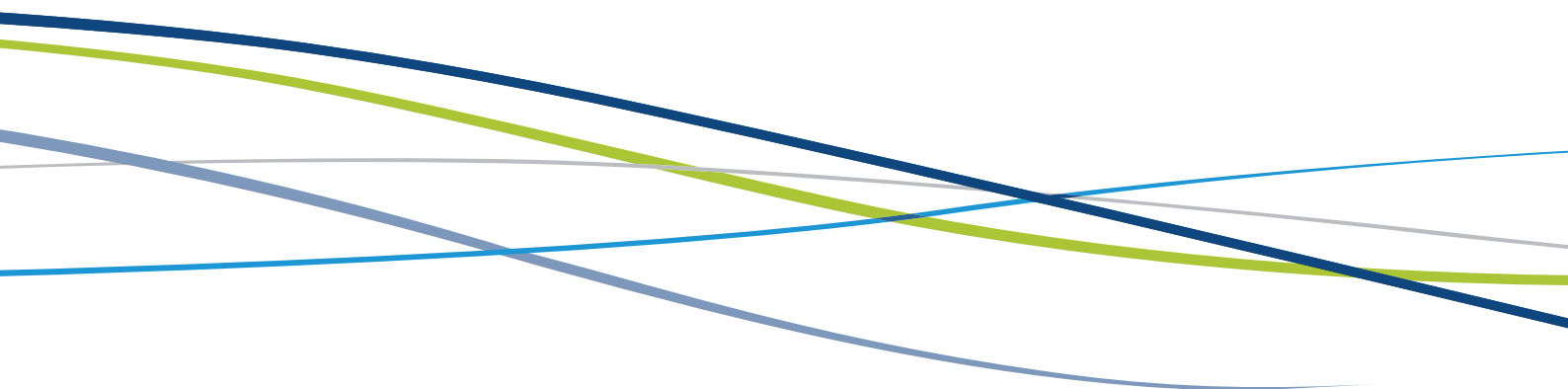
Many factors can give rise to trends in water quality. Identification of a trend is a first step, but probably of more importance is explaining a trend. Being able to explain a trend means that one can properly assess its significance and suggest appropriate manage-

---

ment 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 maybe 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 (eg streamflows, rainfall, other indicators and environmental information) but the ability to link in with other data sets eg through GIS. Capabilities in environmental modelling are becoming well developed but rely to a high degree on having the basic environmental information to work with, for which an adequate monitoring network is essential. Modelling of trends using potential explanatory variables is a complex process considered beyond the scope of this module.

#### 6.7.4 References

- EWQMC (1996), Environmental Water Quality Monitoring Committee "Testing The Waters" Review of Victorian Environmental Water Quality Monitoring: Draft Technical Discussion Paper, Melbourne, Victoria, 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.
- 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*, 27,803-813.
- Hirsch, R.M., and Slack, J.R. (1984), A nonparametric trend test for seasonal data with serial dependence: *Water Resources Research* 20, 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* 18, 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*, 14(4), 804-902.
-



eWater Cooperative Research Centre  
eWater Limited ABN 47 115 422 903

Innovation Centre, Building 22  
University Drive South  
Bruce ACT 2617  
Phone: 1300 5 WATER  
[support@ewater.com.au](mailto:support@ewater.com.au)  
[www.ewater.com.au](http://www.ewater.com.au)

© eWater CRC 2011

