

# **SYSTEM FOR PATCHING RIVER AND RESERVOIR WATER QUALITY DATA**

Report to the  
**WATER RESEARCH COMMISSION**

by

**CE HEROLD, BJJ HAASBROEK & RH STEYN**  
Umfula Wempilo Consulting & Hydrosol

**WRC Report No. 2327/1/16**  
**ISBN 978-1-4312-0802-9**

**June 2016**

**Obtainable from**

Water Research Commission  
Private Bag X03  
Gezina, 0031

[orders@wrc.org.za](mailto:orders@wrc.org.za) or download from [www.wrc.org.za](http://www.wrc.org.za)

**DISCLAIMER**

This report has been reviewed by the Water Research Commission (WRC) and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the WRC, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

# EXECUTIVE SUMMARY

## INTRODUCTION

Salinity modelling and associated system operating rules is an integral part of water resources planning. The validity of the ensuing management decisions is heavily dependent on the observed water quality data against which the models are calibrated. The MOVE software is aimed at making the most of the available relatively infrequent grab sample data to arrive at reasonable estimates of the observed river or dam water quality time series.

## RATIONALE

This project was aimed at developing a user-friendly interactive modelling system to patch daily water quality data and aggregate it into flow-weighted monthly values compatible with the monthly time step used in the system water resources planning models.

## OBJECTIVES AND AIMS

**Aim 1:** Develop a user-friendly integrated water quality patching system.

**Aim 2:** Incorporate enhancements of the existing MOVE program.

## METHODOLOGY

The MOVE patching system has been implemented in the Delphi programming language, and includes the following features:

- Sound data input management allowing more flexible data input
- Incorporation of user-friendly interactive plotting facilities
- Logging and reporting of outlier removal and the reinstatement of previously removed values
- Combination of daily flow data from different stations
- Regression between primary and secondary water quality variables
- Merging of associated water quality station datasets
- Partitioning of datasets, by flow range or by season
- Moving regression patching for river stations with or without replacement or anchoring
- Linear interpolation patching at reservoir stations

- Calculation of flow-weighted or arithmetic average concentrations, taking account of specified minimum numbers of observed and patched water quality values
- Export and reporting of observed and processed datasets and charts
- Provision of online help screens.

## **RESULTS AND DISCUSSION**

The first aim of developing a user-friendly integrated water quality patching system has been met. The second aim of incorporating enhancements to the existing MOVE program has been met and exceeded.

## **CONCLUSIONS**

MOVE represents a powerful new user-friendly tool for the patching and evaluation of water quality data. The moving regression patching technique is an innovation.

Benefits include the means to:

- Speed up water quality data patching and better utilise scarce professional skills
- Rapid in-depth processing that would otherwise have been neglected
- Access documentation via online help screens
- Flag, document and remove outliers.

MOVE makes it possible to explore a number of attractive possibilities outlined below.

## **RECOMMENDATIONS FOR FUTURE RESEARCH**

- Disseminate this powerful tool to practitioners
- Develop a national patched water quality database
- Online assessment of confidence intervals
- Optimise sampling frequency to meet modelling requirements
- Expand to non-conservative variables
- Patch point source inflow and abstraction datasets
- Adjust the minimum required number of observation points per month with flow.

## ACKNOWLEDGEMENTS

The authors would like to thank the Reference Group of the WRC Project K5/2327/1 for the assistance and the constructive discussions throughout the duration of the project. These comprise:

Dr Jennifer Molwantwa (Chairperson)	Water research Commission
Mr Trevor Coleman	Golder Associates
Mr Brendan Hohls	Department of Water and Sanitation
Dr Sebastian Jooste	Department of Water and Sanitation
Mr Simon Ngoepe	Department of Water and Sanitation
Mr Nico Rossouw	Aurecon
Dr Mike Silberbauer	Department of Water and Sanitation
Mr Pieter van Rooyen	WRP
Mr Jurgo van Wyk	Department of Water and Sanitation
Dr Christopher Herold (Project team)	Umfula Wempilo Consulting
Mr Bennie Haasbroek (Project team)	Hydrosol

This page was left blank intentionally

# TABLE OF CONTENTS

EXECUTIVE SUMMARY .....	I
ACKNOWLEDGEMENTS.....	III
TABLE OF CONTENTS.....	V
LIST OF FIGURES .....	IX
LIST OF TABLES.....	XI
LIST OF ABBREVIATIONS .....	XII
<b>1 INTRODUCTION AND OBJECTIVES.....</b>	<b>1</b>
1.1 Motivation .....	1
1.2 Outcomes and expected impacts .....	2
1.3 Objectives and aims .....	2
1.3.1 Aim 1 .....	2
1.3.2 Aim 2 .....	2
<b>2 METHODOLOGY .....</b>	<b>3</b>
2.1 Inception meeting .....	3
2.1.1 Existing model .....	3
2.1.2 Deficiencies of the old model .....	4
2.1.3 New features .....	4
2.2 Business Process Analysis.....	5
2.3 Model specifications .....	6
2.4 Adaptation and export .....	6
2.5 Model structures and user interface .....	7
2.6 Data input and conversion routines.....	7
2.7 Data patching routines .....	8
2.8 Monthly aggregation.....	8
2.9 Online help functionality .....	9
2.10 System testing.....	9
<b>3 THEORY .....</b>	<b>10</b>
3.1 Combination of daily flow datasets.....	10
3.1.1 Naturalisation .....	11
3.1.2 Mean runoffs for periods of overlap.....	11
3.1.3 Daily flows at the water quality station.....	12
3.2 Regress primary water quality variable from secondary variable.....	12
3.3 Patching by linear interpolation .....	13
3.3.1 Linear interpolation calculation .....	14
3.3.2 Extrapolation .....	14
3.4 Patching by moving regression .....	16
3.4.1 Criteria for selecting moving regression patching.....	16
3.4.2 Moving regression calculation .....	17
3.4.3 Linear regression.....	18
3.4.4 Power regression .....	19
3.5 Anchoring of patched values onto observed values .....	21

3.6	Aggregation theory .....	22
3.6.1	Flow-weighted monthly aggregation.....	22
3.6.2	Arithmetic aggregation .....	23
3.6.3	Flow aggregation.....	23
4	<b>USER DIALOG SCREENS.....</b>	<b>24</b>
4.1	Start up dialog .....	24
4.1.1	Version identification .....	24
4.1.2	Screen layout and navigation .....	25
4.1.3	Dataset legend .....	25
4.2	Project dialog.....	25
4.2.1	Opening a Project from scratch.....	25
4.2.2	Editing a Project .....	26
4.2.3	Adding a new Project .....	27
4.2.4	Cloning a new Project .....	27
4.2.5	Deleting a Project.....	27
4.3	Site dialog.....	28
4.3.1	Adding a Site .....	28
4.3.2	Editing a Site .....	29
4.4	Source dialog .....	29
4.4.1	Adding a Source.....	30
4.4.2	Editing a Source .....	30
4.5	Import flow dataset dialog (from a DWS hydro standard daily text file).....	31
4.5.1	Zooming .....	33
4.5.2	Panning .....	34
4.5.3	Insert place marker.....	34
4.5.4	Removal of outliers.....	34
4.5.5	Export of tabulated outliers.....	35
4.5.6	Export of Project data file .....	36
4.6	Import flow dataset dialog (from a user defined column format file).....	37
4.6.1	Daily river flow dataset .....	38
4.6.2	Importing a point source flow dataset.....	39
4.7	Import water quality dataset dialog.....	41
4.7.1	Importation from a DWS WMS standard .CSV output file .....	41
4.7.2	Importing a water quality dataset from a user defined file .....	45
4.8	Combine flow datasets dialog.....	46
4.8.1	Theoretical overview .....	47
4.9	Regress primary and secondary water quality datasets.....	50
4.10	Merging water quality datasets.....	53
4.11	Initialisation for data analysis .....	56
4.11.1	Patch Water Quality Dataset dialog .....	57
4.12	<b>PATCHING BY LINEAR INTERPOLATION .....</b>	<b>58</b>
4.12.1	Time series data tabulation .....	59
4.12.2	Initialisation tab.....	60



	4.12.3	Statistics tab .....	60
	4.12.4	Cancelling interpolation .....	61
	4.12.5	Finalisation of linear interpolation .....	61
	4.12.6	Tabulating linear interpolation results.....	62
	4.13	PATCHING BY MOVING REGRESSION.....	63
	4.13.1	Initialisation for patching by moving regression.....	63
	4.13.2	Moving regression plot .....	65
	4.13.3	Plotting patched daily time series .....	68
	4.13.4	Replacement of patched values with observed values.....	69
	4.13.5	Anchoring of patched values onto observed values .....	71
	4.13.6	Final removal of outliers .....	72
	4.13.7	Tabulation of final patched daily concentrations.....	74
	4.13.8	Display final moving regression initialisation data .....	75
	4.13.9	Plot final patched water quality data.....	75
	4.14	MONTHLY AGGREGATION OF DATASETS .....	76
	4.14.1	Selection of monthly aggregation .....	76
	4.14.2	Aggregate water quality dataset dialog .....	76
	4.14.3	Aggregate water quality dataset dialog .....	79
	4.14.4	Information displayed in Data tab.....	83
	4.14.5	Information displayed in Statistics tab .....	84
	4.14.6	Final aggregated monthly chart.....	85
	4.14.7	Export of aggregated monthly water quality file.....	86
5		DATA REQUIREMENTS AND PRE-PROCESSING .....	88
6		RESULTS, TREATMENT OF RESULTS AND DISCUSSION.....	96
	6.1	Aim 1: Develop a user-friendly integrated water quality patching system. ....	96
	6.2	Aim 2: Incorporate enhancements of the existing MOVE program. ....	96
7		CONCLUSIONS .....	97
	7.1	Patching tool available .....	97
	7.2	Unique features .....	97
	7.3	Increased efficiency.....	97
	7.4	Improved quality .....	97
	7.5	Documentation .....	97
	7.6	Flagging outliers .....	97
	7.7	New possibilities.....	98
8		RECOMMENDATIONS .....	99
	8.1	Dissemination .....	99
	8.2	Develop national patched database .....	99
	8.3	Confidence intervals.....	99
	8.4	Optimise sampling frequency .....	99
	8.5	Expand to non-conservative variables.....	100
	8.6	Improve patching at reservoir stations .....	100
	8.7	Patching point sources .....	101
	8.8	Handle gaps in new point source inflows .....	101

8.9	Adjusting the minimum number of observation points with flow .....	101
9	LIST OF REFERENCES .....	102
	APPENDIX A: USE CASE DIAGRAM .....	103
	APPENDIX B: USE CASE SEQUENCES .....	104
	APPENDIX C: OBJECT RESPONSIBILITIES .....	107
	APPENDIX D: DWS HYDROLOGICAL CHECK CODES.....	113

## LIST OF FIGURES

<b>Fig.</b>	<b>Description</b>	<b>Page</b>
3.1	Hypothetical catchment with multiple flow monitoring stations .....	10
3.2	Error adjustment across the gap between sampling points .....	22
4.1	Start-up dialog .....	24
4.2	Project creation dialog .....	26
4.3	Project editing dialog .....	26
4.4	Project manipulation dialog .....	27
4.5	Delete confirmation dialog .....	28
4.6	Creation of a new Site .....	28
4.7	Editing a Site.....	29
4.8	Creation of a new Site .....	30
4.9	Editing a Source .....	31
4.10	Importing a flow dataset from WMS database standard .CMS file .....	31
4.11	Designation of the raw flow dataset in the decision tree .....	32
4.12	Importing a flow dataset from WMS database standard .CMS file .....	33
4.13	Removal of outliers .....	35
4.14	Exporting tabulation of outliers .....	35
4.15	Outlier export file .....	36
4.16	Data export file .....	37
4.17	Exporting tabulation of outliers .....	38
4.18	Importing a point source dataset .....	40
4.19	Importing a water quality dataset from a DWS WMS standard .CSV output file ...	41
4.20	Interactive water quality chart .....	43
4.21	Water quality chart with corresponding flow .....	44
4.22	Selection of the Combine Dataset option .....	47
4.23	Combine water quality datasets dialog .....	48
4.24	Selecting option to regress water quality datasets .....	49
4.25	Water quality regression dialog .....	50
4.26	Display of regression equation information .....	51
4.27	Water quality regression dialog .....	51
4.28	Primary water quality variable after patching by regression with secondary variable .....	52
4.29	Selection of the Merge datasets dialog .....	53

4.30	Merge water quality datasets dialog (initial display) .....	54
4.31	Merge water quality datasets dialog .....	55
4.32	Selection of water quality patching option .....	56
4.33	Initialisation for water quality patching .....	56
4.34	Plot of patched water quality dataset .....	58
4.35	Tabulation of observed and interpolated water quality dataset .....	58
4.36	Tabulation of linear interpolation assumptions .....	59
4.37	Tabulation of linear interpolation statistical results .....	60
4.38	Plot of patched water quality dataset .....	61
4.39	Tabulation of final interpolated patched water quality dataset .....	61
4.40	Initialisation for patching by moving regression .....	62
4.41	Example of exported seasonal flow plot .....	64
4.42	Moving regression statistical plot with no partitioning .....	65
4.43	Moving regression statistical plot with seasonal partitioning .....	66
4.44	Selection of moving regression period .....	67
4.45	Daily moving regression plot .....	68
4.46	Daily moving regression with replacement with observed values .....	69
4.47	Daily moving regression with anchoring onto observed values .....	70
4.48	Interactive removal of outliers from Patch WQ dataset chart .....	71
4.49	Result after removal of outlier .....	72
4.50	Tabulation of moving regression data .....	74
4.51	Plot of patched daily concentrations and observed flows .....	75
4.52	Initialisation for monthly aggregation .....	76
4.53	Interactive Plot dialog .....	79
4.54	Interactive plot dialog after changing patching selection for January 1980 .....	81
4.55	Aggregated data displayed in Data tab for moving regression option .....	82
4.56	Monthly data shown in Statistics tab .....	84
4.57	Final average monthly water quality file .....	85
4.58	Selection of option to export monthly water quality data file .....	85
4.59	Monthly water quality export file in standard WQT / WRPM format .....	86

## LIST OF TABLES

<b>Tab.</b>	<b>Description</b>	<b>Page</b>
2.1	Summary of business Process analysis sub-deliverables .....	6
4.1	Water quality variable adjustment factors .....	45

## LIST OF ABBREVIATIONS

BPA	Business Process Analysis
Combine	Combine two or more flow datasets to patch the daily flows at a water quality monitoring point
.CSV	Comma delimited file
DMS	Dissolved Major Salts
DWA	Department of Water Affairs (now DWS)
DWAF	Department of Water Affairs and Forestry (now DWS)
DWS	Department of Water and Sanitation
EC	Electrical Conductivity
GIGO	Garbage In Garbage Out principle
MAR	Mean Annual Runoff
Merge	Merge two or more water quality datasets to patch the water quality dataset at a designated monitoring site
MOVE	Moving regression water quality patching and aggregation system
Project	Group of flow and water quality monitoring sites
RG	WRC Reference Group advising on the project
RMSE	Root Mean Square Error
Site	Flow and/or water quality monitoring site
Source	Source of flow and/or water quality data at a monitoring site
TDS	Total Dissolved Solids
UML	Unified Modelling Language
WMS	Water Management System – DWS water quality database
WQ	Water Quality
WQT	Water Quality TDS model
WRPM	DWS' Water Resources Planning Model

# 1 INTRODUCTION AND OBJECTIVES

## 1.1 Motivation

Salinity modelling, along with associated water quality operating rules, is an integral part of the Water Resources Planning Model (WRPM) that is used extensively in South Africa to plan water resources. Operation of the salinity modelling component requires calibration of the Water Quality TDS (WQT) hydro-salinity model. However, validity of the salinity modelling is dependent on the observed water quality data against which it is calibrated. The integrity of the model simulations and the management decisions based on them is thus dependent on obtaining a sound sequence of observed salinity data. The GIGO (Garbage In Garbage Out) principle applies and calibrating models against incorrect data can lead to misleading results that can seriously misdirect the planning decisions based on them. The same is true of any other form of water quality modelling.

Observed raw water quality data comprises relatively infrequent grab samples separated by long gaps, typically a week long, but sometimes stretching to month long gaps. Relationships established between flows (for which there are continuous measurements) and the water quality grab samples can be used to patch the missing water quality data and thereby establish a much better estimate of the aggregated flow-weighted monthly salt concentrations required to calibrate the WQT model. Various techniques are incorporated in the MOVE program to patch river and reservoir salinity data, both of which require different techniques. A moving regression, with or without anchoring, has proved to be particularly useful for patching water quality at river stations.

A low budget program incorporating the basic functions was initially developed to support a Department of Water Affairs and Forestry (DWAF)<sup>1</sup> project for the development of an integrated water resources management plan for the Olifants River catchment (Herold, 2007). This program was subsequently used in other studies. However, the original program lacks user-friendly features and is cumbersome to use, involving a complex multi-step process that requires much scarce specialist time to complete. This makes it extremely difficult to disseminate the technology to other users. Aside from the formidable problem of sustainable dissemination, time constraints also put pressure on users to forgo the benefits of the technological advances and take short cuts that can easily lead to misleading results. Moreover, the original program lacks useful features that have now been included.

---

<sup>1</sup> The Department of Water Affairs and Forestry (DWAF) later became the Department of Water Affairs (DWA) and is now called the Department of Water and Sanitation (DWS).

This project was aimed at developing a user-friendly interactive modelling system to patch daily water quality data and aggregate it into flow-weighted monthly values compatible with the monthly time step used in the system water resources planning models. This combines decades of experience in detecting outliers, patching and aggregating records into one software package that new water quality modellers can use. This will speed up the process, thereby improving the likelihood that sound principles will be employed. The availability of the package will make it easier to train other users and free more specialist time to attend to model calibration, options analysis and evaluation of the results.

While the primary motivation is to patch salinity data, MOVE can also be used to patch other types of water quality variables, including the conservative salts comprising the Total Dissolved Solids (TDS), such as sodium and sulphate, as well as non-conservative variables such as phosphate.

## **1.2 Outcomes and expected impacts**

The benefits to be derived include:

- Capture years of expert know-how and make it available to all users.
- Enable better training.
- Save time in undertaking salinity model calibrations and modelling.
- Provide a tool for researchers and water quality modelling practitioners to improve the extraction of valuable information from available raw data.
- Facilitate the alerting of water quality database custodians to likely outliers.
- Provide a tool for the possible generation of a national patched and improved water quality database.
- Provide a tool to help optimise the sampling frequency required at key monitoring stations to facilitate reliable load estimation to support water quality modelling.

Expand the software to better deal with non-conservative water quality variables.

## **1.3 Objectives and aims**

The aims of the project are as follows:

### **1.3.1 Aim 1**

Develop a user-friendly integrated water quality patching system.

### **1.3.2 Aim 2**

Incorporate enhancements of the existing MOVE program.



## **2 METHODOLOGY**

The MOVE patching system has been implemented in the Delphi programming language, making use of user-friendly data input, analysis routines and interactive screen plots.

The tasks have been divided into a series of steps described in Sections 2.1 to 2.10.

### **2.1 Inception meeting**

An inception Reference Group (RG) meeting was held to discuss and refine the methodology and project outcomes.

#### **2.1.1 Existing model**

The features of the existing FORTRAN program were demonstrated and its weaknesses and the proposed enhancements were discussed. The methodology and deliverables were also discussed and refined.

The earlier FORTAN model facilitates entry of a file of daily flow data and one containing grab sample monitoring data. This data can comprise a standard daily flow file derived from the Department of Water and Sanitation's (DWS) hydrology database and grab sample water quality extracted from the DWS's Water Management System (WMS). Alternatively the daily flow data and/or the water quality data can be entered from user-prepared files in a designated format.

A choice can be made between patching by linear interpolation (suitable for reservoir data) or by moving regression between flow and water quality (suitable for river stations). The former is suitable for reservoir stations where storage attenuation and manipulated water releases weaken the correlation between flow and water quality, but at the same time storage attenuation slows the rate of change of water quality and improves the representativeness of sparse water quality sampling renders the linear interpolation. The latter is highly relevant for river stations where water quality fluctuates rapidly in response to changes in river flow.

The model allows flow-range partitioning of the data.

The moving regression option also facilitates choice of regression equation (linear or power regression) and allows a range of moving regression periods to be iterated automatically and the statistics for each to be tabulated.

The option of anchoring the regressed daily water quality values onto the observed data points is also facilitated.

The patched daily water quality data is then aggregated into monthly values for use in the WQT calibration model, which in turn is used as input to the widely used DWS's Water Management Planning Model (WMPM) to carry out system analyses. This aggregation is most commonly expressed in the form of flow-weighted monthly water quality averages. This is the most important form since the product with monthly flow values then yields the salt load that is most critical to the calibration of the WQT model. The simple arithmetic mean can also be specified, since this is relevant when comparing the average quality of the water abstracted by run-of-river schemes.

### **2.1.2 Deficiencies of the old model**

The main gaps in the old patching system include:

- It is based in an old operating system that is not user friendly.
- It lacks plotting facilities, necessitating a tedious multi-step evaluation process to identify and eliminate outliers.
- The removal of outliers and the reasons for doing so is not documented.
- The regression of the primary water quality variable, such as Total Dissolved Solids (TDS), with a more frequently monitored secondary variable, such as electrical conductivity (EC), is not supported.
- A procedure for combining multiple flow station records is not available. This is desirable if the flow record at or near to the water monitoring station is broken and one or more adjacent stations are needed to fill the gaps.
- Merging of water quality datasets is not supported. This is important since some sampling points measure the same body of water, such as at a dam station where sampling can occur in the lake adjacent to the dam wall, at a downstream flow gauging weir and in left and right bank canal and pipe outlets. Split records also occur when an old flow gauging station (which is often also used for water quality sampling) is closed in favour of a new weir or cableway station located nearby. It is also not uncommon for a new data collector to open a new sampling point, not realising that there is already a sampling record at or near to the same locality.

### **2.1.3 New features**

The new water quality patching model incorporates the following features:

- Object orientated code designed for user friendliness and flexibility.
- Superior data input management allowing more flexible data input.

- Incorporation of user-friendly interactive plotting facilities.
- Logging and reporting of outlier removal and the reinstatement of previously removed values.
- Regression between primary and secondary water quality variables.
- Inclusion of combination of flow data from different stations.
- Merging of water quality station datasets.
- Improved partitioning of datasets, by flow range or by season. The new model also allows the automatic splicing together of partitioned patched data and the relevant statistics.
- Relaxation of the minimum number of patched water quality values required before monthly averages will be aggregated, to take account of days when there is no flow. Allowance is also made for specifying seasonal (monthly) minimums.
- Superior data and chart export and reporting.
- Provision of online help screens.

The Business Process Analysis (BPA) for the model development was presented and discussed at the inaugural meeting (see Section 2.2 below).

## **2.2 Business Process Analysis**

A Business Process Analysis (BPA) outlining the desirable features of the interactive MOVE patching system was developed using the standard Unified Modelling Language (UML) methodology.

The BPA was a high level first step aimed at describing what was required of the programmers.

This activity involved development of the following to guide programme development:

- Use Case Diagram
- Use Case Sequences
- Object responsibilities.

The BPA consists of the three sub-deliverables contained in Appendices A, B and C, summarised in Table 2.1.

**Table 2.1: Summary of Business Process Analysis sub-deliverables**

<b>Deliverable</b>	<b>Description</b>	<b>Appendix</b>
MOVE Use Case Diagram	Use Case Diagram displaying all use cases required for using MOVE to patch water quality data.	A
MOVE Use Case Sequences	A report that lists the steps in every use case.	B
MOVE Object Responsibilities	A report that lists every step in every use case under the logical modules responsible for controlling the step.	C

### **2.3 Model specifications**

Model specifications were developed to flesh out the high level BPA with particular emphasis on the required user interface screens and plots, again to communicate requirements to the programmers.

The model specifications provide information on the envisaged structure of the screens in sufficient detail to allow the programmers to proceed. These are often in the form of sketches. It must be recognised that this was a temporary step in the development path that has been superseded by the final product that has been implemented. Inevitable changes took place during the implementation stage to minimise programming complexity, eliminate unnecessary repetition, make subroutines more universally applicable for other purposes and to account for needs that became apparent as the coding was implemented and in response to user needs identified by Reference Group members and the project team.

In view of the transitory (although important) nature of the specifications, they have been omitted from the final report. Instead reporting has been focussed on the final product that supersedes them.

Delphi was chosen as the object orientated programming language since this is versatile enough to be incorporated into DWS systems, if so desired at a later stage. It is also the programming language used in all the other DWS planning models.

### **2.4 Adaptation and export**

Blocks of coding of the original FORTRAN program were adapted as independent elements that could be incorporated easily into the new Delphi program. Additional standalone blocks of coding were also developed to implement the identified enhancements required for the new water quality data patching system.

While this was an essential step along the way to developing the new system, it is now redundant since the relevant parts are now embedded in the new Delphi program.

However, the underlying theory is an integral part of the final report. It is not reported on in this section as it has been consolidated into Section 3, which deals with theory for both new and old components of the patching system.

## **2.5 Model structures and user interface**

This step comprised implementation in Delphi of the basic model skeleton and user interface screens. These were derived from the BPA and model specifications discussed in Sections 2.2 and 2.3.

The screens and their operation are not discussed here since they are only part of the overall system. Instead they are discussed along with the rest of the system in Section 4.

## **2.6 Data input and conversion routines**

Each **Project** is structured to contain two primary folders namely a **Data** folder and an **Analysis** folder. The interface dialog screens, processing routines, interactive screen plots and export facilities that were developed and are contained in the **Data** directory are outlined below.

Dialogs are provided to input and process flow and water quality data, convert units to a consistent internal project convention, store the data, plot the data, eliminate identified outliers and document the reasons for their removal.

If data for a primary (e.g. TDS) and secondary (e.g. EC) water quality variable have been stored, then a simple linear regression can be used to patch gaps in the primary water quality variable record using the secondary water quality variable values. The processed patched water quality dataset is then stored. Plots of the regression results can be examined to identify and eliminate further outliers and store the modified results.

Processed or raw datasets for the primary water quality variable from two or more adjacent monitoring stations can be merged to form one dataset, following a defined order of priority. The merged dataset is stored. Plots of the different water quality datasets can be compared to determine the validity of the merger and to eliminate outliers. The merged dataset is then recalculated and stored.

Daily flow data for two or more flow gauging stations can be combined to fill in gaps in the primary record at the station nearest to the water quality monitoring point, taking account of incremental net (i.e. inflow minus abstraction) point source inflows. The combined flow dataset is stored. Plots of observed flows at the different gauging stations can be compared to identify and eliminate outliers. The combined flow dataset is then recalculated and stored.

Comments regarding removed outliers can be exported, allowing the user to report them to the source organisation. Key statistics of the processed data can also be exported, as can time series datasets.

The user dialog screens are shown and discussed in Section 4.

## 2.7 Data patching routines

The data patching routines for each **Project** are contained in the **Analysis** folder. The following outlines the interface dialog screens, processing routines, interactive screen plots and export facilities that were developed and are contained in the **Patching** folder, under the **Analysis** folder.

Provision is made to patch the designated daily water quality time series by linear interpolation, which is appropriate for reservoir stations.

Alternatively, for river stations, a moving regression between water quality and flow can be used to patch the daily water quality time series. The regression equation type is selected. The moving regression allows flow range or seasonal partitioning and automatically iterates for a range of moving regression periods. A chart showing relevant goodness of fit results for the specified range of moving regression periods is displayed, from which the most appropriate is selected for further processing. Time series plots of observed flows and observed and patched water quality values is produced. Further outliers that are identified can be removed and documented. Time series of raw moving regression water quality values, with replacement of patched values with observed ones and with anchoring of the regressed values onto observed values can be displayed simultaneously or individually. The plots can be exported, as can tabulations of patched and observed time series and the key assumptions made in the patching process.

## 2.8 Monthly aggregation

The interface dialog screens, processing routines, interactive screen plots and export facilities that were developed and are contained in the **Aggregation** folder, under the **Analysis** folder are outlined below.

Criteria for the aggregation of patched daily water quality data into either flow-weighted monthly concentrations or arithmetic monthly averages are defined for each season. These include the minimum number of daily observed values required per month, below which the monthly average will not be calculated and will be defined as a gap. Similarly the minimum number of patched daily values per month is set.

An interactive plot displaying observed daily flows and water quality variable concentrations and patched daily concentrations derived from raw moving regression is generated. This includes a plot of daily water quality concentrations with patched values and a plot of daily concentrations with the patched values anchored onto the observed values. The corresponding aggregated monthly averages are also shown. The user selects the most appropriate patching method (regressed, with replacement or with anchoring) for the entire period. This general choice can be modified for individual months by clicking onto a different monthly average on the interactive plot. Inspection of the plot screen may also reveal further outliers that can be removed and logged.

The aggregation selection criteria, plot screen and tabulated time series can be exported. The aggregated monthly time series can also be exported in the standard format used as input to the WQT model.

## **2.9 Online help functionality**

Online help screens relevant to each dialog screen have been developed. The entire help document can also be viewed.

## **2.10 System testing**

The first level of checking occurred at the 2<sup>nd</sup> Reference Group Meeting when the system was demonstrated and comments were made. Subsequently the project team also detected various items that required attention. The system was also distributed to Reference Group members to carry out their own testing.

The project team assessed the desired changes according to their importance and the project scope. Those changes that do not impair the operation of the system but are desirable new features have been listed as “nice to have” for consideration for incorporation in future projects.

The project team identified and implemented 20 change requests during this stage. Others that have merit but are beyond the scope of the project were flagged for possible future implementation.

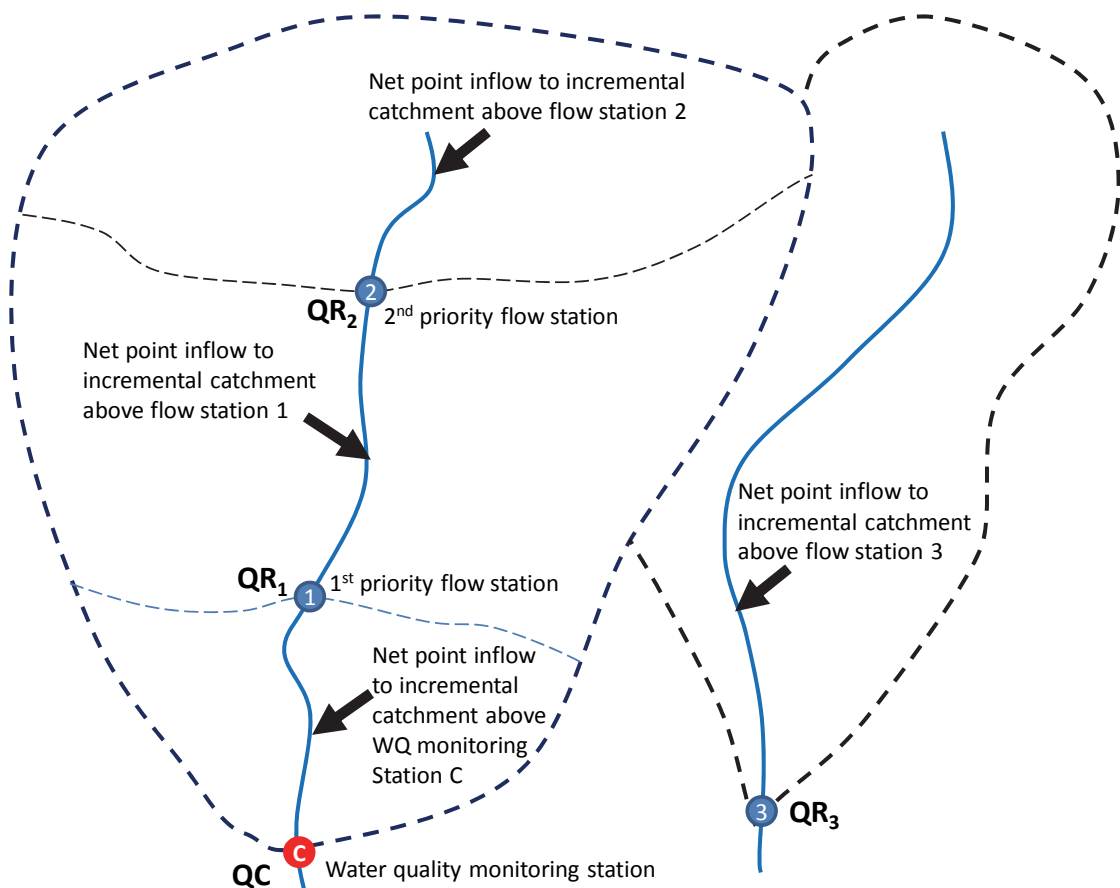
### 3 THEORY

The theory underlying the patching process is discussed in the following sections.

#### 3.1 Combination of daily flow datasets

The objective is to patch daily flow data at the water quality monitoring point using the flow records at each flow gauging station taking account of the priority of each station and the incremental net (inflow minus abstraction) point source inflows.

Figure 3.1 illustrates a hypothetical catchment showing the water quality sampling point at which patched water quality data is required (station C) and three flow gauging points (1, 2 and 3).



**Figure 3.1: Hypothetical catchment with multiple flow monitoring stations**

The criteria for choosing the order of priority include proximity to the water quality monitoring point, reliability of the flow record, climatic and hydrological representativeness, the sample size and the number of gaps in the record. In the absence of data in close proximity to the



monitoring station, the flow record in an adjacent river can even be used to approximate the flow at the monitoring point, provided that the climatic and hydrological conditions are similar. The user is required to choose the appropriate order of priority.

At least one flow station is required to facilitate regression between discharge and water quality to infill gaps between observed water quality samples. However, if no flow data is available it is still possible (although less accurate) to estimate monthly water quality values.

In the example flow gauging point 1 is the highest priority station to patch flows at the water quality station, while station 3 is the lowest priority for patching.

### 3.1.1 Naturalisation

The first step is to carry out a first order naturalisation of the daily flows at each of the specified flow gauging points, taking account of net point source inputs. (A full hydrological analysis taking account of urban areas, irrigation, wetlands and storage effects is beyond the scope of this study and is of secondary importance since the primary purpose of the flow data is to provide a pattern against which to correlate with observed concentrations.) Since the relative locations of the flow gauging stations and the water quality sampling station can differ markedly from the example used in Figure 3.1, it is necessary for the user to provide files of the cumulative net point source inflows upstream of each point.

The naturalised flow is calculated as:

$$QN_{i,j} = QR_{i,j} - QP_{i,j} \dots\dots\dots (3.1)$$

Where

$QN_{i,j}$  = Naturalised flow at gauging station i on day j

$QR_{i,j}$  = Observed flow at gauging station i on day j

$QP_{i,j}$  = Cumulative net point inflow upstream of gauging station i on day j

Negative  $QN_{i,j}$  values are acceptable since this reflects river losses that need to be taken into account.

If  $QR_{i,j} < 0$  (i.e. no river flow data), then the naturalised flow  $QN_{i,j}$  is set to -1. to denote missing data.

### 3.1.2 Mean runoffs for periods of overlap

The cumulative naturalised runoffs for the first priority flow gauging station,  $QN_1$ , and the second priority station,  $QN_2$ , are calculated for periods of overlap between the two stations.

The ratio between the two totals is then used to patch flows at station 1 for all days for which data is missing at station 1 but there is data for station 2.

This procedure is then repeated using the patched naturalised data at station 1 and the patched naturalised data at station 3, and so on until all of the stations have been used in their order of priority.

### 3.1.3 Daily flows at the water quality station

The patched daily flows at the 1<sup>st</sup> priority station after using the observed records at all of the lower priority flow stations are then used to estimate the flows at the water quality station:

$$QC_j = QN'_{1,j} \cdot (AC/A1) + QCP_j \dots\dots\dots (3.2)$$

Where:

- QC<sub>j</sub> = Patched flow at water quality monitoring station on day j (m<sup>3</sup>/s)
- QN'<sub>1,j</sub> = Patched naturalised flow at flow gauging station 1 on day j (m<sup>3</sup>/s)
- AC = Total catchment area upstream of water quality monitoring station (km<sup>2</sup>)
- A1 = Total catchment area upstream of flow gauging station 1 (km<sup>2</sup>)

If QC<sub>j</sub> > 0. then QC<sub>j</sub> is set equal to 0.

The above procedure is required to complete the graphs that are plotted in the **Combine flow datasets** dialog.

### 3.2 Regress primary water quality variable from secondary variable

Where appropriate a linear regression between the primary water quality variable (e.g. TDS) and the secondary water quality variable (e.g. EC) is carried out. This is used to extend the primary water quality record for days when the primary value is missing but there is a value for the secondary water quality variable.

The regression equation is in the form:

$$CP = A + B \cdot CS \dots\dots\dots (3.3)$$

Where:

- CP = Primary water quality variable
- CS = Secondary water quality variable
- A = Constant
- B = Constant

Constants B and A are calculated as:

$$B = (\sum(CS \cdot CP) - \sum CS \cdot \sum CP / N) / (\sum CS^2 - (\sum CS)^2 / N) \dots\dots\dots (3.4)$$

$$A = (\sum CP - B \cdot \sum CS) / N \dots\dots\dots (3.5)$$

Where:

N = Sample size

The linear correlation coefficient is given by:

$$R = (\sum(CS \cdot CP) - \sum CS \cdot \sum CP / N) / \sqrt{[(\sum CS^2 - (\sum CS)^2 / N) \cdot (\sum CP^2 - (\sum CP)^2 / N)]} \dots\dots\dots (3.6)$$

The ratio (variable RATIO) between the primary and secondary variable values on days when both are available is calculated as:

$$RATIO = CP / CS \dots\dots\dots (3.7)$$

The average ratio (AVERAT) between the two datasets is calculated as:

$$AVERAT = (\sum RATIO) / N \dots\dots\dots (3.8)$$

The variable arrays required to produce the plots in the **Regress dataset** dialog are derived from the equations discussed above.

### 3.3 Patching by linear interpolation

Patching by linear interpolation between sample points is most appropriate for reservoir stations where for all but the largest floods the water quality at the dam outlet is not correlated with the outflow rate. This is because for most of the time outflows are managed to meet downstream water requirements. Moreover, the relationship with the instantaneous inflow to the dam is also broken by the cumulative effect of several months or years of antecedent runoff accumulated in the reservoir. For the above reasons it is inadvisable to do a regression of water quality against outflow for reservoir stations with substantial storage relative to daily inflow.

Fortunately this same storage attenuation has the beneficial effect (in terms of patching the water quality data) of damping the rate of change in water quality at the dam outlet, thereby greatly increasing the accuracy of patching by linear interpolation than would be the case at a river station with the same sampling frequency. However, it must be cautioned that vertical stratification can result in the observed water quality at the surface at the dam wall differing from that at canal outlets and at sluice gate outlets.

In view of all of the above considerations, simple linear interpolation between the observed sampling points is considered a good means of estimating the water quality variable load leaving a dam, against which hydrological models need to be calibrated.

Exceptions can occur if the dam is relatively small, in which case the water quality behaviour could approximate that of a river station during flood conditions.

Weirs located downstream on regulated reaches in close proximity to a dam also need to be treated in the same manner, seeing as the water quality at the weir then reflects conditions in the Dam, rather than daily (or even monthly) catchment runoff.

### 3.3.1 Linear interpolation calculation

Linear interpolation takes place between pairs of observed concentration points and is given by the equation:

$$C_x = CO_{i-1} + (CO_i - CO_{i-1}) / (D_i - D_{i-1}) \cdot (D_x - D_{i-1}) \dots\dots\dots (3.9)$$

Where:

- $C_x$  = Patched concentration on day X
- $CO_i$  = Observed concentration on day i
- $CO_{i-1}$  = Observed concentration on day i-1
- $D_i$  = Day i
- $D_{i-1}$  = Day i-1
- $D_x$  = Day X, for which the concentration is being patched.

### 3.3.2 Extrapolation

The MOVE tool is not intended for extrapolating far beyond the limits of the observed concentration dataset. As such it does not attempt to take account of seasonal patterns and extrapolating over long periods of time is inadvisable.

When linear interpolation is selected, it is necessary to define the means of extrapolating beyond the ends of the data. Three options are provided in the **Patch WQ dataset** dialog (see Figure 4.33):

- Linear extrapolation
- Set to end point concentration
- Set to average interpolated concentration.

Extrapolation should only be used to extend the daily patching to the start (or end) of the first (or last) month for which observed data is available.

### 3.3.2.1 Linear extrapolation

Linear extrapolation simply extends the patching for the period preceding the first observed data point using the linear interpolation equation derived from the first and second observed concentrations. Similarly, extrapolation beyond the last observation point uses the interpolation equation derived from the second last and last observation points. Since this method of extrapolation can lead to negative values, a check is incorporated to set all negative extrapolated values to zero. This method is only appropriate for very short extrapolation periods since it is highly dependent on the relative values of only two observation points.

Backwards extrapolation is governed by the following equation, which is similar in form to equation (3.9):

$$C_j = CO_1 + ((CO_2 - CO_1) / (DO_2 - DO_1)) * (j - DO_1) \dots\dots\dots (3.10a)$$

Where:

- $C_j$  = Extrapolated concentration on day j, where  $j < DO_1$
- $CO_1$  = concentration of first observation point on day  $DO_1$
- $CO_2$  = concentration of second observation point on day  $DO_2$

The equation for forwards extrapolation is as follows:

$$C_j = CO_{N-1} + ((CO_N - CO_{N-1}) / (DO_N - DO_{N-1})) * (j - DO_{N-1}) \dots\dots\dots (3.10b)$$

Where:

- $C_j$  = Extrapolated concentration on day j, where  $J > DO_N$
- $CO_N$  = concentration of last observation point on day  $DO_N$
- $CO_{N-1}$  = concentration of second observation point on day  $DO_2$

### 3.3.2.2 End point concentration extrapolation

This extends the patching horizontally to the end of the first and last months using the concentration of the observed end point. This method can be applied to slightly longer extrapolation periods, but will lose validity as the cumulative inflow increases along with evaporation losses.

For backward extrapolation the concentrations for all days preceding the date of the first sample point are set equal to that of the first observation point. For forward extrapolation the concentration of all points is set equal to that of the last observation point.

$$C_j = CO_1 \dots\dots\dots (3.11a)$$

or

$$C_j = CO_N \dots\dots\dots (3.11b)$$

### 3.3.2.3 The average of interpolated concentrations

In this method of extrapolation the average of all of the observed and interpolated concentration values is use for both backward and forward extrapolation:

$$C_j = (\sum_{i=DO_1}^{i=DON} Ci)/NP \dots\dots\dots (3.12)$$

Where:

- Ci = patched concentration on day i
- DO<sub>1</sub> = first day for which there is patched or observed concentration data
- DO<sub>N</sub> = last day for which there is patched or observed concentration data
- NP = sample size after patching using equation (3.9)

Gaps in the patched record are excluded from the calculation.

All of the above extrapolation methods have the severe disadvantage of failing to accommodate temporal variation, seasonal fluctuation and trends. Hence it is advisable that the criteria used to aggregate the daily data into monthly averages (see Figure 4.52) should be set to ensure that extrapolation is not used to patch gaps longer than a month.

## 3.4 Patching by moving regression

### 3.4.1 Criteria for selecting moving regression patching

Patching by moving regression is most appropriate when there is a strong direct relationship between daily flow and concentration. This is eminently suitable for river sampling stations, where there is usually a strong relationship between daily flow and concentration. For example, the salinity of river water usually drops substantially during the passage of floods and rises steadily during dry periods. The relationship for non-conservative pollutants is more complex, since a first flush effect often causes an initial rise in concentration as accumulated detritus is washed off catchment surfaces and declines later when the pollution source has been depleted. Rapid depletion of salts stored on urban paved surfaces is particularly noticeable.

As a general rule a moving regression should be selected for river stations that are remote from the influence of upstream dams.

The rationale for using a moving regression, rather than a straight regression using the entire dataset is to take account of the influence of antecedent conditions. It has been observed that the store of readily available soluble salts increases during dry periods and is depleted during wet periods, the washoff efficiency being related to the rainfall intensity. This is particularly apparent in urban areas where much of the runoff concentration is dominated by runoff from paved surfaces (Metcalf and Eddy, Inc., 1971). But this is also relevant for pervious catchment areas where part of the runoff is derived from interflow that draws on the soil moisture storage (Herold, 1981). Typically, if two floods of equal magnitude occur with only a short period of zero rainfall between them, the salt concentration trough for the second flood will be lower than that for the first one, since the store of soluble salts readily available for the second flood will be lower. Conversely, if the two floods are separated by a prolonged drought period, the steady accumulation of soluble salts can result in the later flood exhibiting higher salt concentrations.

Non-conservative pollutants can display a more pronounced first flush effect, with a sharp rise in concentration on the rising limb of the flood, followed by a reduction in concentration later after the readily available surface pollutants are depleted. But the general trend is similar, also attributable to the antecedent flow effect.

### **3.4.2 Moving regression calculation**

The validity of water quality model calibration, and hence the ensuing management decisions that they support, is highly dependent on the validity of the observed samples on which the calibration is based. This is particularly important when the modelling time step is relatively long. Since system planning models are typically operated at a monthly time step, it becomes extremely important to arrive at good estimates of the observed monthly water quality time series. In this regard salinity modelling is particularly important in semi-arid environments, which is the prevalent condition over most of South Africa. This is especially relevant where industrial, mining and irrigation catchment development plays a prominent role. Moreover, since the salinity of downstream impoundments is determined by the salt loads entering dams, estimation of the flow-weighted salt concentrations during flood events is critically important. Flow-weighted salt concentrations are also of primary importance in determining the mixed salinity downstream of river junctions.

Reliable hydro-salinity modelling to support water quality driven system operating rules, such as those used in Vaal Barrage for the strategically important Vaal River system, depends on deriving reliable estimates of flow-weighted monthly salt concentrations. Salt concentrations usually drop dramatically during floods, during which most of the water is exported. Since infrequent grab samples seldom coincide with flood peaks and instead tend to measure the

high concentrations associated with low flow conditions, simple arithmetic averages can lead to a systematic over-estimation of catchment salt export. This in turn leads to over-estimation of the salt loads entering downstream reservoirs, with errors that can persist for many months, particularly in Southern African semi-arid conditions that necessitate the provision of large reservoirs, typically storing one or more times the catchment MAR (Mean Annual Runoff).

The pronounced inverse relationship between observed river concentration and flow results in a significant correlation between the two. This is particularly useful for patching gaps between sparse grab samples since flow records are typically continuous. Since the flow regime can change substantially during such gaps, calculation of the average monthly concentration from the few sparse observed monitoring points can be far from representative.

The simplest way to take advantage of this relationship is to correlate concentration with flow and use an appropriate regression equation to estimate the fluctuations in water quality variable concentration during the temporal gaps between the observed concentration / flow pairs.

Two types of regression have been considered:

- Linear regression
- Power regression.

### 3.4.3 Linear regression

The linear regression equation takes the form:

$$C = A + B \cdot Q \dots\dots\dots (3.13)$$

Where:

- C = Estimated concentration (appropriate units)
- A, B = Regression constants (intercept and slope)
- Q = Flow rate (standardised to m<sup>3</sup>/s for this study)

The slope, B, is calculated as:

$$B = (\sum Q \cdot C - \sum Q \cdot \sum C / N) / (\sum Q^2 - (\sum Q)^2 / N) \dots\dots\dots (3.14)$$

Where:

- N = Sample size (number of observed (Q,C) pairs)



The intercept, A, is given by:

$$A = (\sum C - B \cdot \sum Q) / N \dots\dots\dots (3.15)$$

The correlation coefficient, R, is calculated as:

$$R = \{ \sum(Q \cdot C) - (\sum Q \cdot \sum C) / N \} / \sqrt{(\sum Q^2 - (\sum Q)^2 / N) \cdot (\sum C^2 - (\sum C)^2 / N)} \dots\dots\dots (3.16)$$

While a linear regression can yield reasonable correlations, there is nothing to restrain the estimation of negative values. It also fails to mimic the semi-open ended higher concentrations that can develop. Hence, while this method of regression could occasionally yield useful results, it is not recommended.

### 3.4.4 Power regression

A power regression yields much more acceptable results since it cannot generate negative concentrations and maximum concentrations are unbounded:

$$C = A \cdot Q^B \dots\dots\dots (3.17)$$

The regression constants A and B and the correlation coefficient, R, are calculated as:

$$B = [\sum(\ln Q) \cdot (\ln C) - (\sum \ln Q) \cdot (\sum \ln C) / N] / [\sum(\ln Q)^2 - (\sum \ln Q)^2 / N] \dots\dots\dots (3.18)$$

$$A = \exp[(\sum \ln C) / N - B \cdot (\sum \ln Q) / N] \dots\dots\dots (3.19)$$

$$R = (\sum(\ln Q) \cdot (\ln C) - \sum(\ln Q) \cdot \sum(\ln C) / N) / \sqrt{(\sum(\ln Q)^2 - (\sum \ln Q)^2 / N) \cdot (\sum(\ln C)^2 - (\sum \ln C)^2 / N)} \dots\dots\dots (3.20)$$

Where:

$\ln Q$  = Natural logarithm of the observed flow

$\ln C$  = Natural logarithm of the observed concentration

While shorter time intervals can be used, this study is focussed on patching data to a monthly time step, which is the practical norm for water resources planning studies.

However, correlation for the full period does not take account of the antecedent effect. The moving regression concept has been introduced to take better account of the antecedent effect.

A moving regression with a period of M samples takes the first M pairs of (Q,C) values to calculate the first regression equation and its corresponding correlation coefficient. This

equation is then used to patch concentrations from the observed flow values from the 1<sup>st</sup> to the (M/2+1)<sup>th</sup> (integer arithmetic) data pairs.

The second regression equation and correlation coefficient is calculated using the 2<sup>nd</sup> to (M+1)<sup>th</sup> data pairs and is used to patch the gap from the (M/2+1)<sup>th</sup> (Q,C) data pair until the (M/2+2)<sup>th</sup> data pair. In this manner the selection of the M (Q,C) data pairs is rolled through the entire dataset, with each regression equation used to patch between the two most central points for which simultaneous flow and concentration data are available.

Once the last observed data pair is reached, the last regression equation is used to patch not only the gap between the central pair of observed values, but all the way to the date of the last pair observed (Q,C) values.

The average correlation coefficient for all of the observed (Q,C) pairs are then calculated as:

$$RA_M = [R_1 \cdot (M/2+1) + \sum_{i=2}^{K-1} R_i + R_K \cdot (K-M/2)] / K \dots\dots\dots (3.21)$$

Where:

- RA<sub>M</sub> = Average correlation coefficient for a moving mean period of M data pairs
- R<sub>1</sub> = Correlation coefficient for first M data pairs
- M = Moving regression period (number of valid (Q,C) data pairs)
- K = Total number of observed (Q,C) data pairs
- R<sub>K</sub> = Correlation coefficient for the last M observed (Q,C) data pairs.

In the general case each regression equation is used to patch only the central gap and hence the associated correlation coefficient has a weighting of 1 (half assigned to the observation point preceding the gap plus half assigned to the observation point at the end of the gap). Of necessity the first regression equation has to represent all of the points from the first observation pair until the end of the central gap (i.e. the first MOVE/2+1 data pairs and is weighted accordingly. Similarly, the correlation coefficient for the last regression equation also receives a higher weighting.

It would have been more rigorously exact to ignore the first and last M/2 data pairs, but this pedantic academic approach would have been unacceptable since water quality sampling data is too precious to waste. After all, the objective is to make the best of the available data, not to discard it for the sake of academic purity.

The Root Mean Square Errors (RMSE) for concentration and load are also calculated. Weightings for end effects do not have to be factored in since the RMSE is not calculated

from equations spanning several points. The RMSE for each moving regression period are calculated as:

$$\text{RMSC} = \sqrt{\sum\{C - A \cdot Q^B\}^2 / K} \dots\dots\dots (3.22)$$

$$\text{RMSL} = \sqrt{[\{C \cdot Q - (Q \cdot A \cdot Q^B)\}^2] / K} \dots\dots\dots (3.23)$$

Where:

RMSC = Standard error (RMSE) of concentration values

RMSL = Standard error (RMSE) of load values

### 3.5 Anchoring of patched values onto observed values

Anchoring is used to fit the patched concentration values onto the observed points and also improve the plausibility of the patched values between the observed points.

The anchoring method that was employed calculates the error between the regression line and the observed values at both end points of the gap and applies a linearly tapering adjustment between the two points. This is illustrated in Figure 3.2.

The error on day  $D_x$  is calculated as:

$$E_x = C_n - CO_n - (CO_{n+1} - C_{n+1}) \cdot (D_x - D_n) / (D_{n+1} - D_n) \dots\dots\dots (3.24)$$

And the anchored concentration is given as:

$$C_x' = C_x - E_x \dots\dots\dots (3.25)$$

Where:

$E_x$  = Error between first observed and regressed concentration values on day x

$C_n$  = Regressed concentration on day n at the start of the gap

$CO_n$  = Observed concentration on day n at the start of the gap

$CO_{n+1}$  = Observed concentration on day n+1 at the end of the gap

$C_{n+1}$  = Regressed concentration on day n+1 at the end of the gap

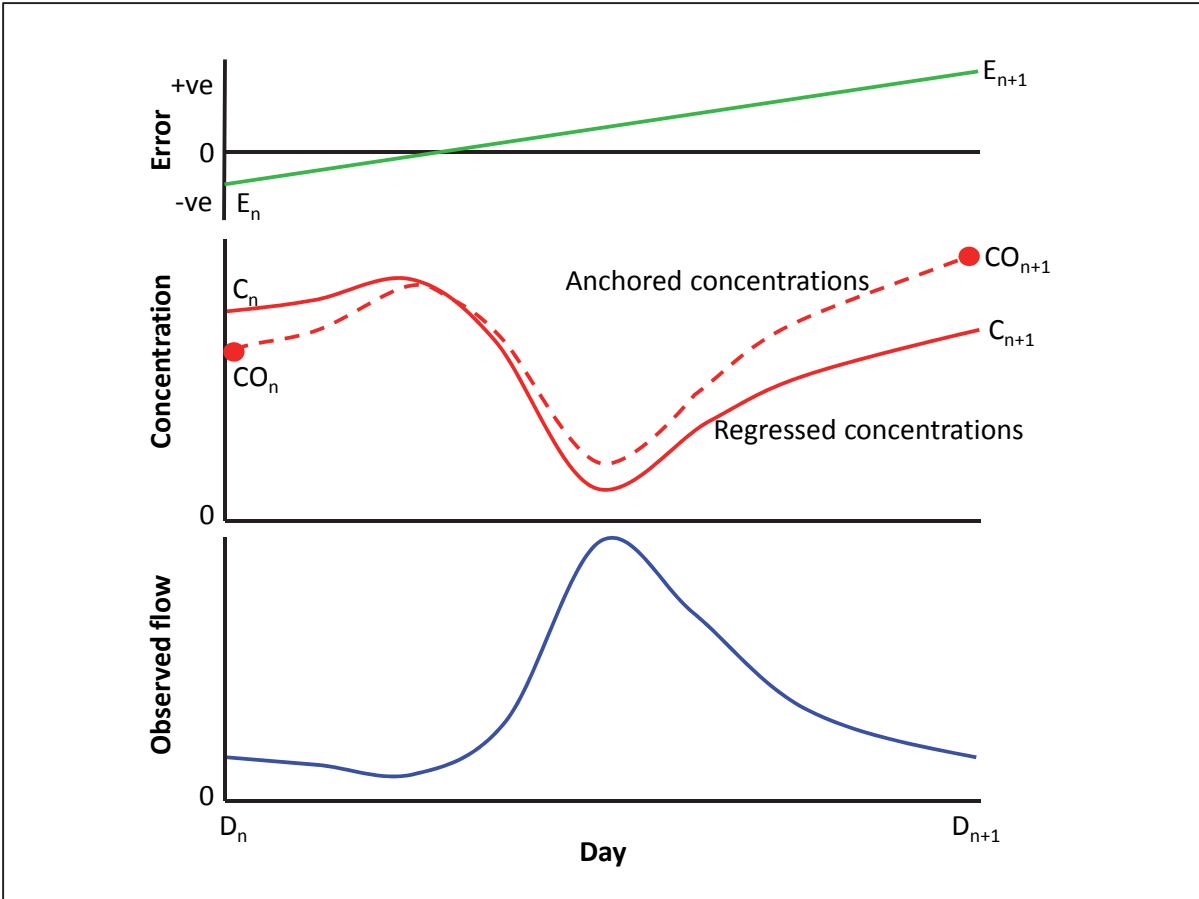
$D_x$  = Day being patched

$D_n$  = Day n at start of gap

$D_{n+1}$  = Day n+1 at end of gap

$C_x'$  = Patched concentration on day x after anchoring

$C_x$  = Regressed concentration on day x



**Figure 3.2: Error adjustment across the gap between sampling points**

This method preserves the shape of the moving regression between flow and concentration, but applies a smoothed adjustment between the end points to ensure that the patched line passes through the observation points.

**3.6 Aggregation theory**

**3.6.1 Flow-weighted monthly aggregation**

Flow-weighted monthly aggregation is estimated as:

$$C_w = (\sum_{i=1, N_c} C_i \cdot Q_i) / (\sum_{i=1, N_c} Q_i) \dots\dots\dots (3.26)$$

Where

- $C_w$  = Flow-weighted average monthly concentration
- $i$  =  $i^{th}$  day of month for which both flow and patched concentration data is available
- $N_c$  = Number of days in the month for which both flow and patched concentration data is available

$C_i$  = Patched concentration on day i

$Q_i$  = Patched flow on day I ( $m^3/s$ )

### 3.6.2 Arithmetic aggregation

The arithmetic monthly concentration is estimated as:

$$C_A = (\sum_{i=1, N_c} C_i) / N_c \dots\dots\dots (3.27)$$

Where

$C_A$  = Arithmetic average monthly concentration

i =  $i^{th}$  day of month for which both flow and patched concentration data is available

$C_i$  = Patched concentration on day i

### 3.6.3 Flow aggregation

The average monthly flow is estimated as:

$$Q_A = (\sum_{i=1, N_q} Q_i) \cdot 0.0864 / N_q \dots\dots\dots (3.28)$$

Where

$Q_A$  = Average monthly flow ( $10^6 m^3$ )

i =  $i^{th}$  day of the month for which both flow and patched concentration data is available

$N_q$  = number of days in the month for which flow data is available

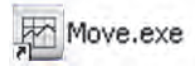
$C_i$  = Patched concentration on day i

## 4 USER DIALOG SCREENS

The MOVE patching system has been implemented in the Delphi programming language, making use of user-friendly data input and processing windows and interactive screen plots.

### 4.1 Start up dialog

The program is initiated by selecting the icon:



When **Move.exe** is invoked for the first time an empty **Project** screen is opened.

#### 4.1.1 Version identification

Left clicking on **File** (in the top left hand corner) opens the pop-up menu "Version". Selecting this displays the version of the program being used (Circle 1 in Figure 4.1). **File** can be left-clicked from any screen to display the program version number.

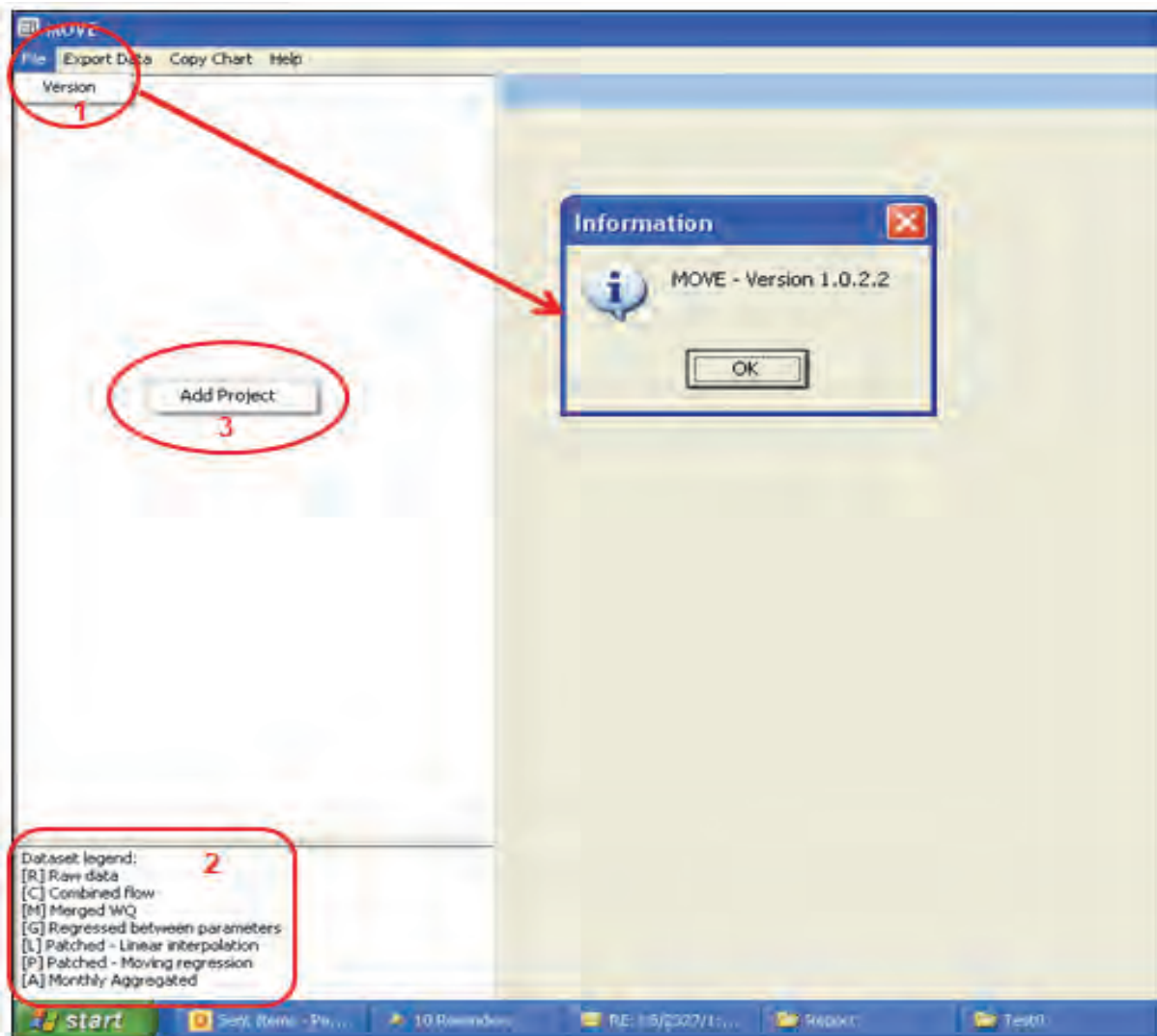


Figure 4.1: Start-up dialog

### 4.1.2 Screen layout and navigation

All of the screens show both a left hand and right hand panel.

Once populated, the left hand panel shows an expandable or collapsible tree that is used to navigate to different dialog screens. This is similar to the familiar folder trees used in many windows applications.

Definitions of the functional features of each dialog screen are displayed in the right hand panel.

### 4.1.3 Dataset legend

The bottom section of the left hand panel always contains a key to indicate the type of time series dataset being accessed (Ellipse 2 of Figure 4.1). Each dataset is prefixed by a letter enclosed by square brackets defined in the key as:

- [R] Raw data (A daily time series dataset of either a water quality or flow variable values.)
- [C] Combined flow (Processed daily flow time series resulting from the combination of more than one flow dataset.)
- [M] Merged WQ (Processed daily time series after merging more than one water quality time series)
- [G] Regressed between parameters (Processed time series for the primary water quality variable, say TDS, after patching missing daily entries with values regressed from the secondary water quality variable, say EC.)
- [P] Patched (Processed daily time series of water quality variable after patching using a moving regression against flow or linear interpolation between break points.)
- [A] Monthly aggregated (Processed monthly time series of water quality variable obtained from either flow-weighted or un-weighted aggregation of a daily water quality time series.)

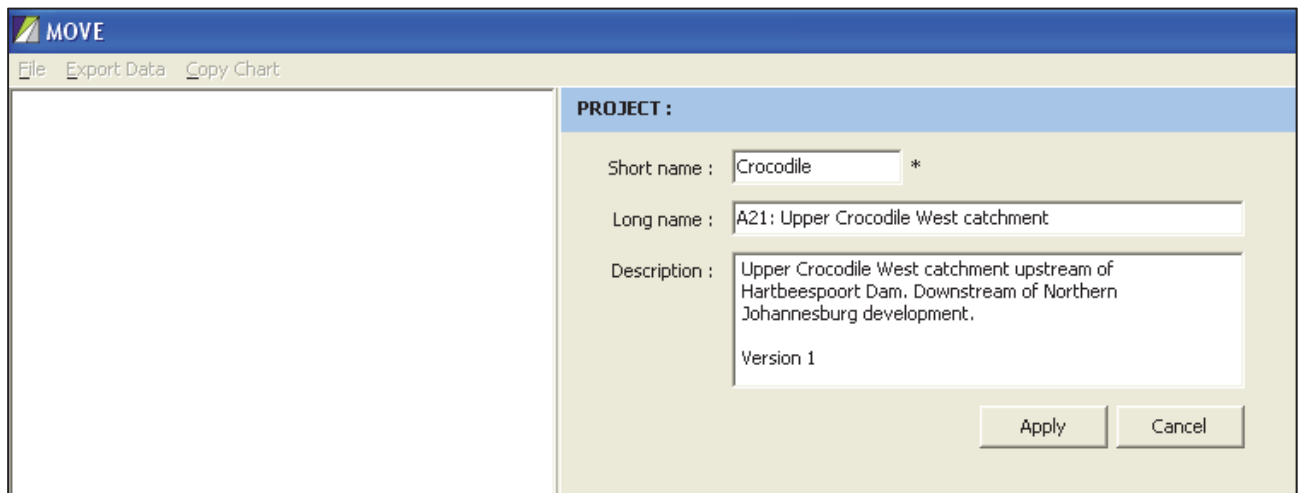
## 4.2 Project dialog

### 4.2.1 Opening a Project from scratch

A **Project** can be thought of as a catchment or group of catchments in a study area that can contain one or more flow and water quality monitoring stations.

Right clicking in the left hand panel of an empty initial screen opens the pop-up menu **Add Project** that facilitates creation of a new **Project** (Ellipse 3 of Figure 4.1).

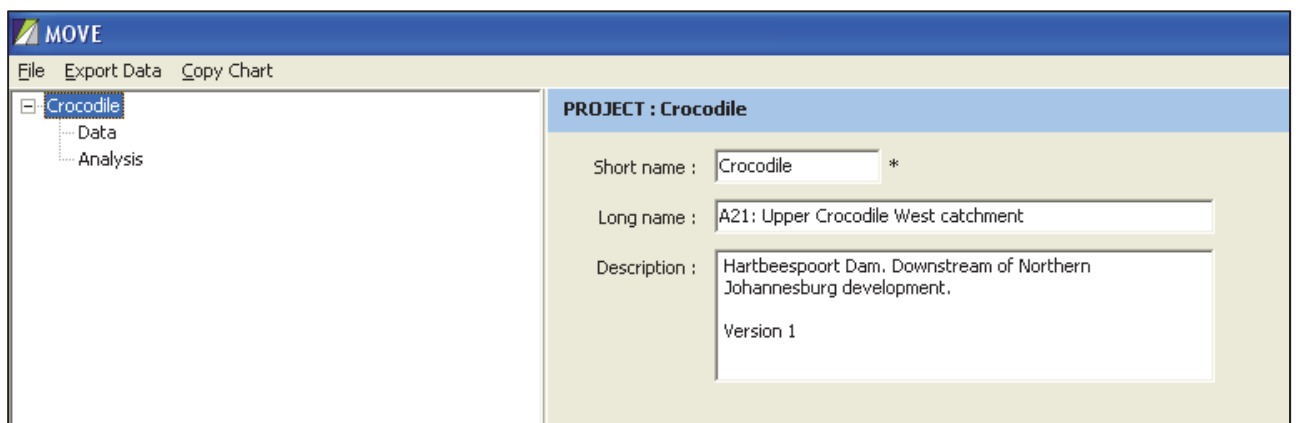
Left clicking on **Add Project** opens up a window in the right hand panel that enables entry of the **Project** description (see Figure 4.2).



**Figure 4.2: Project creation dialog**

Only the **Project** short name is compulsory, although it is advisable to also provide a fuller description to enable other users to identify the **Project**.

Selecting the **Apply** button consolidates the definition of the new **Project** and displays the **Project** name in the left hand panel, with two sub-folders: **Data** to store various datasets and **Analysis** to store patched datasets (see Figure 4.3).



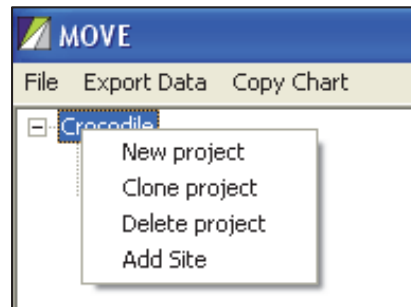
**Figure 4.3: Project editing dialog**

#### 4.2.2 Editing a Project

Left clicking on any existing **Project** in the left hand panel brings up the dialog shown in Figure 4.3, which facilitates editing of the **Project** description.



Right clicking on a **Project** name in the left hand panel opens the pop-up menu shown in Figure 4.4 that allows the following options:



**Figure 4.4: Project manipulation dialog**

- |                        |  |
|------------------------|--|
| <b>New Project:</b>    | Create a new <b>Project</b> from scratch   |
| <b>Clone Project:</b>  | Clone a new <b>Project</b> to look identical to the existing one that was clicked on |
| <b>Delete Project:</b> | Delete the existing <b>Project</b>   |
| <b>Add Site:</b>       | Add a new <b>Site</b> to the <b>Project</b> .  |

The first three options are directly **Project** related. The fourth has to do with building the tree below a **Project** (see Section 4.3).

#### 4.2.3 Adding a new Project

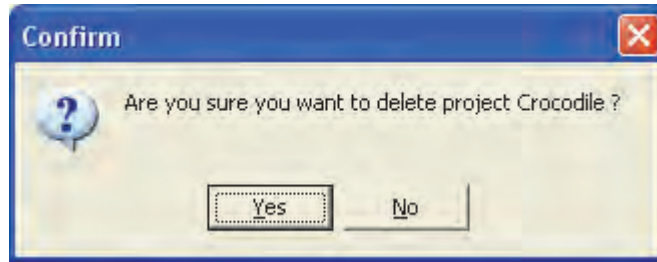
Selecting the **New project** option from the drop down menu facilitates creation of a new **Project** as described in Section 4.2.1.

#### 4.2.4 Cloning a new Project

Selecting the **Clone project** option creates a new project that has a tree structure and input files identical to that of the **Project** that was selected. The user can then make the changes required to define the details of the new **Project**.

#### 4.2.5 Deleting a Project

When the **Delete project** option is selected a pop-up dialog will first appear to confirm the deletion, after which the *entire* **Project** and all its subsidiary tree branches will be removed. This same warning is displayed before any of the subsidiary folders is deleted. Heeding this warning is very important since on confirmation the entire tree below the **Project** will be removed.



**Figure 4.5: Delete confirmation dialog**

### 4.3 Site dialog

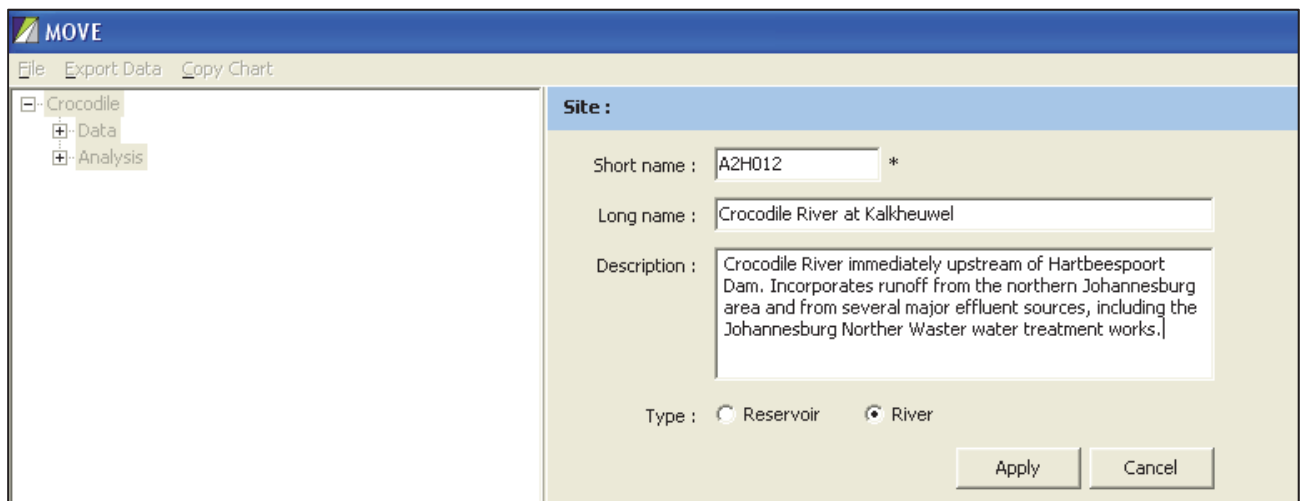
#### 4.3.1 Adding a Site

A **Site** is a flow and / or water quality monitoring station.

The **Add site** option (Figure 4.4) is used to create a monitoring station. It can only be accessed by right clicking on an existing **Project**. (Clicking on a **Data** folder has no effect since this is merely a generic receptacle to store subsidiary folders.)

The new **Site** is stored in the tree structure under the **Data** folder. The **Data** folder does not need to be created by the user as it appears automatically when a **Project** is created (see Figure 4.3.).

Figure 4.6 shows the example of a **Site** for monitoring station A2H012 is being created.



**Figure 4.6: Creation of a new site**

Aside from providing the **Site** short name (compulsory) and optional name and description, the user also selects one of the two radio buttons specifying if the **Site** is a river station or a reservoir station. This is necessary since the patching process for river stations differs

substantially for the two types of station. The different patching processes are dealt with later in the Sections dealing with **Analysis**.

When the **Apply** button is selected the new **Site** is consolidated and displayed in the left hand panel navigation tree.

#### 4.3.2 Editing a Site

Left clicking on a **Site** in the left hand panel opens the **Site** dialog in the right hand panel, allowing editing of any of the fields.

Right clicking on a **Site** produces the pop-up menu shown in Figure 4.7. Selecting the **Clone site** option is analogous to cloning a **Project**, with the new **Site** containing a sub-tree structure and subsidiary datasets identical to that of the existing **Site**.

The **Delete Site** option removes the existing **Site** along with all of its subsidiary folders and files. A warning is displayed to confirm this selection before the files are removed.



Figure 4.7: Editing a Site

#### 4.4 Source dialog

A **Source** can be conceptualised as the provider of flow and/or water quality datasets applicable to a **Site**. More than one such **Source** can be applicable to a **Site**. For example flow data might be obtained from one organisation and water quality data from another. It also sometimes occurs that two or more organisations monitor water quality at the same **Site** (For example at some locations in the Waterval River catchment data are available from Rand Water, Sasol, one or more mining companies, the DWS routine grab sampling programme and, until it was vandalised, from a DWS electrical conductivity data logger.) A

**Source** could also comprise data that have been derived from one or more of the above **Sources** and then pre-processed by the user.

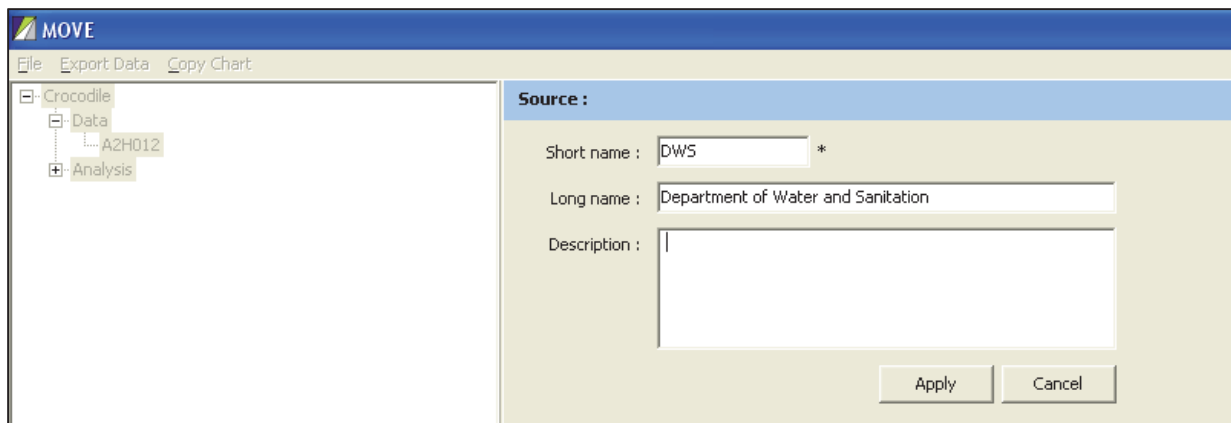
Only one flow and one water quality dataset may be imported per **Source**. If an organisation holds more than one water quality record (say from a normal DWS grab sampling system and continuous EC data logger records) then it is necessary to open two **Sources** (called, say, **DWS** and **DWS data logger**)

#### 4.4.1 Adding a Source

The **Add Source** option (see Figure 4.7) facilitates the addition of a new **Source** below the **Site**.

Figure 4.8 shows the creation of a new **Source**.

Selecting the **Apply** button consolidates the definition of the new **Source** and enters its name into the left hand panel navigation tree.

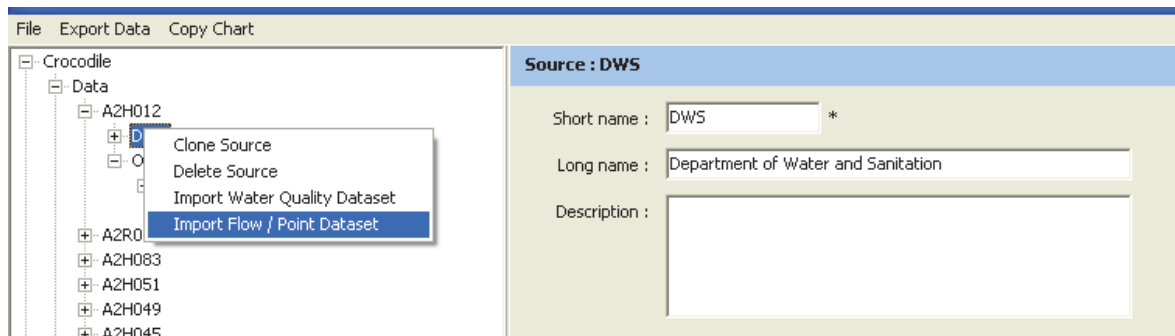


**Figure 4.8: Creation of a new Site**

#### 4.4.2 Editing a Source

Left clicking on a **Source** opens up the **Source** dialog screen (see Figure 4.8), the fields of which can then be edited.

Right clicking on a **Source** in the left hand navigation panel opens the pop-up menu shown in the left hand panel of Figure 4.9, which allows the user to clone a new **Source** from the existing one or delete the existing **Source**.

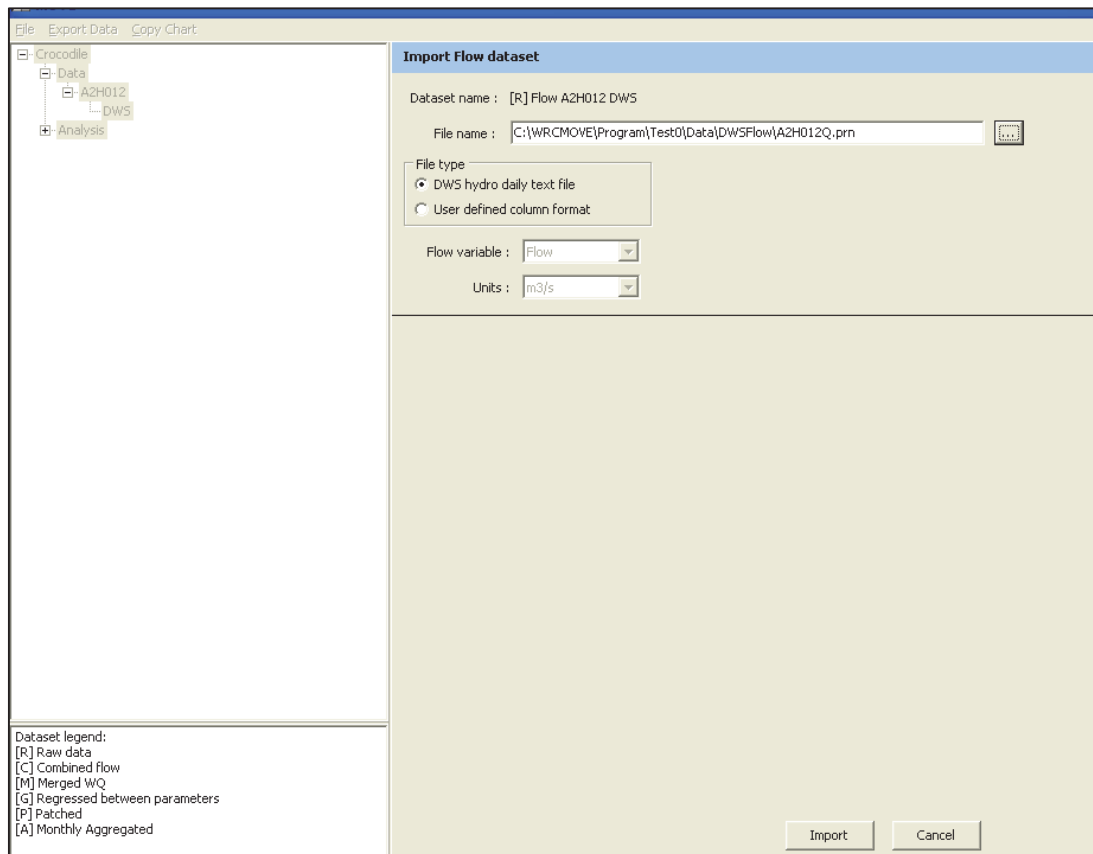


**Figure 4.9: Editing a source**

This pop up menu also facilitates the importation of a flow or water quality time series dataset.

#### 4.5 Import flow dataset dialog (from a DWS hydro standard daily text file)

Selection of the **Import Flow / Point Dataset** option from the pop up menu shown in Figure 4.9 opens the dialog in the right hand panel shown in Figure 4.10.



**Figure 4.10: Importing a flow dataset from WMS database standard .CMS file**

The user browses to select the desired DWS Hydro standard text file. The **DWS hydro daily text file** radio button is chosen.

A standardised daily flow file generated by the DWS hydrological database gives the flow units as m<sup>3</sup>/s, which is also the standard internal flow unit used in the MOVE model.

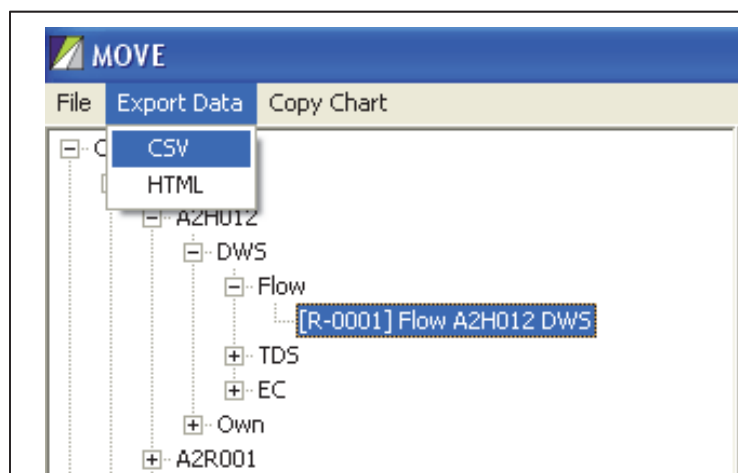
The format of a standardised DWS daily flow file is as follows:

Code	Date	Flow	Check
C8H036	140.00 19971231 0000	0.000	255
C8H036	140.00 19980101 0000	0.008	1
C8H036	140.00 19980102 0000	0.009	1

For our purposes the most important fields that have to be used are the 3<sup>rd</sup> (observation date), the 5<sup>th</sup> (average daily flow in m<sup>3</sup>/s) and 6<sup>th</sup> (check code).

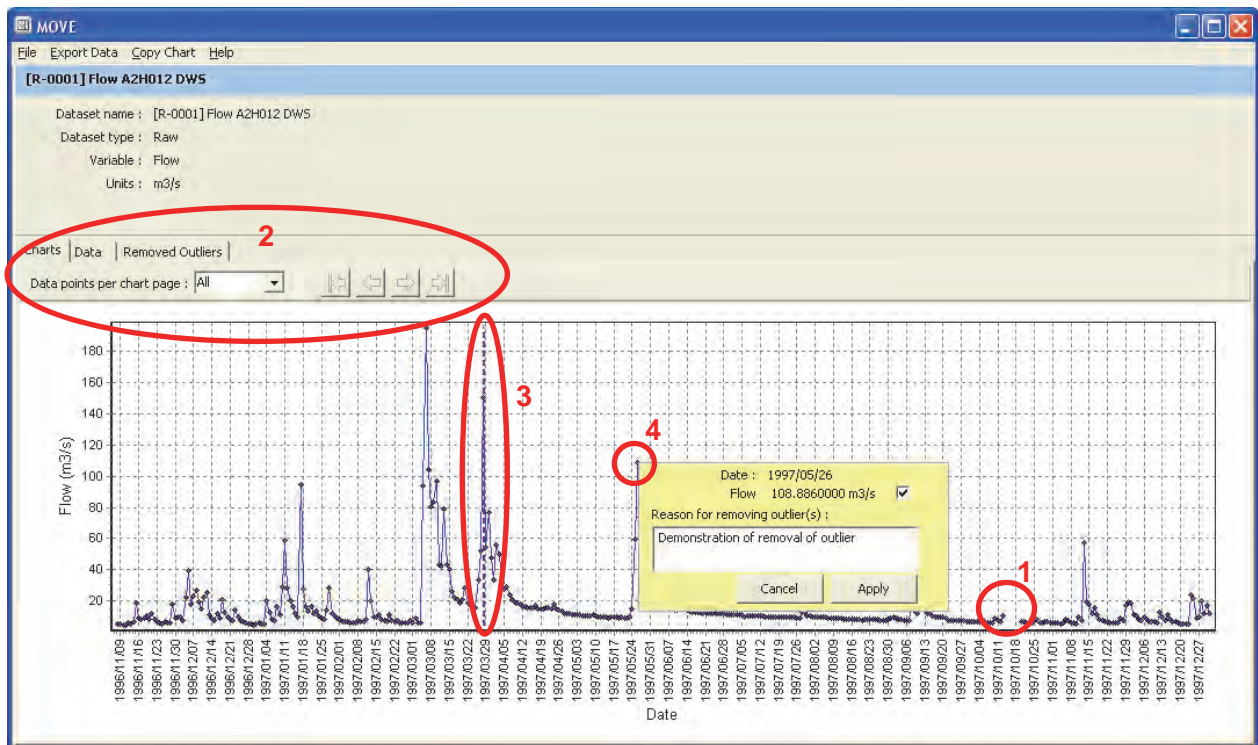
The check code is used to determine whether or not the flow for that day can be used, or should be recorded as missing data. A full list of the flow check codes is given in Appendix D.

Selecting the **Import** button opens a dialog, which in the example is called **Flow A2H012 DWS** (where “Flow” is the file type, “A2H012” is the **Site** name and “DWS” is the data **Source**). The **Project** flow file has a similar name, prefixed by **[R-0001]**, where “R” denotes “raw” data and 0001 is a unique dataset serial number. The left hand panel of this dialog is shown in Figure 4.11.



**Figure 4.11: Designation of the raw flow dataset in the decision tree**

The right hand panel of the dialog shown in Figure 4.12 displays an interactive chart of the imported daily flow dataset. A table to view daily flow values is also provided.



**Figure 4.12: Plot of observed daily flow dataset**

In the example shown the flow values for the period 14 to 19 October 1997 (see circle 1 in Figure 4.12) are shown as zeros, but the flow data for these days is flagged with the quality code “170”, which denotes a period of no record. The program automatically adjusts these values to the missing data code, hence the gap in the graph during this period. (A common error made by unwary users is to ignore the check code and incorrectly enter the data as zeros.)

The program also checks for a range of other codes and makes appropriate adjustments. For example, “Good continuous data” (code 1) is always accepted. Likewise “Good edited data” (code 2) is also accepted. However, “Accumulated Unreliable” (code 79) is treated as missing data, etc.

The underlying rationale is that bad data is much worse than no data since it leads to misleading results.

#### 4.5.1 Zooming

By default the chart starts by showing all data points (see ellipse 2 in Figure 4.12). This view allows zooming by left clicking on the top left hand corner of the desired plot area, holding the mouse down, navigating to the bottom right hand corner of the new chart area and then releasing the mouse. The zoomed chart area is then displayed. The zooming process can be repeated to zoom in further if necessary.

Zooming is helpful when clicking onto finely separated points on the graph that are difficult to differentiate. It is also valuable for examining low flows to identify suspect zero flow values.

Undoing the zooming is facilitated by clicking anywhere on the chart, holding the mouse down, navigating to the left and then releasing the mouse.

#### **4.5.2 Panning**

The drop down menu (see ellipse 2 in Figure 4.12) allows the chart display to be confined to a selected fixed number of days. The navigation arrows shown in ellipse 2 are then activated and can be used to pan to the left or right of the displayed chart area.

Zooming cannot be done when the panning option is active.

#### **4.5.3 Insert place marker**

Holding down the Ctrl key and clicking on the horizontal axis inserts a vertical place marker line (see purple dashed vertical line in ellipse 3 in Figure 4.12). This feature is particularly useful when more than one chart is displayed since the vertical line extends through all of the charts. Hence a hydrological feature, such as a flood peak, can be related to a change in concentration. The place marker is removed by holding down the Alt button and clicking anywhere on the horizontal axis.

#### **4.5.4 Removal of outliers**

The charts have also been made interactive, allowing outliers to be clicked on and deleted (see circle 4 in Figure 4.12). When this is done the user is required to enter reasons why the data point was removed. This information is then stored in a table showing which points have been removed and why.

Data points can be removed one by one from the same chart. All deletions are immediately shown as gaps in the plot. A number of entries can be removed simultaneously by clicking on the **Data** tab, highlighting the desired entries, right clicking in the highlighted area and selecting "Remove outliers" from the pop up menu (see Figure 4.13).

A record of the accumulated deletions can be tabulated by selecting the **Removed Outliers** tab. This record is stored in the **Project** database and is added to whenever new deletions are made. Deletions can also be reversed by right clicking on a tabulated outlier and selecting "Restore" from the pop up menu. The previously deleted values are immediately restored to the chart and the underlying **Project** dataset.



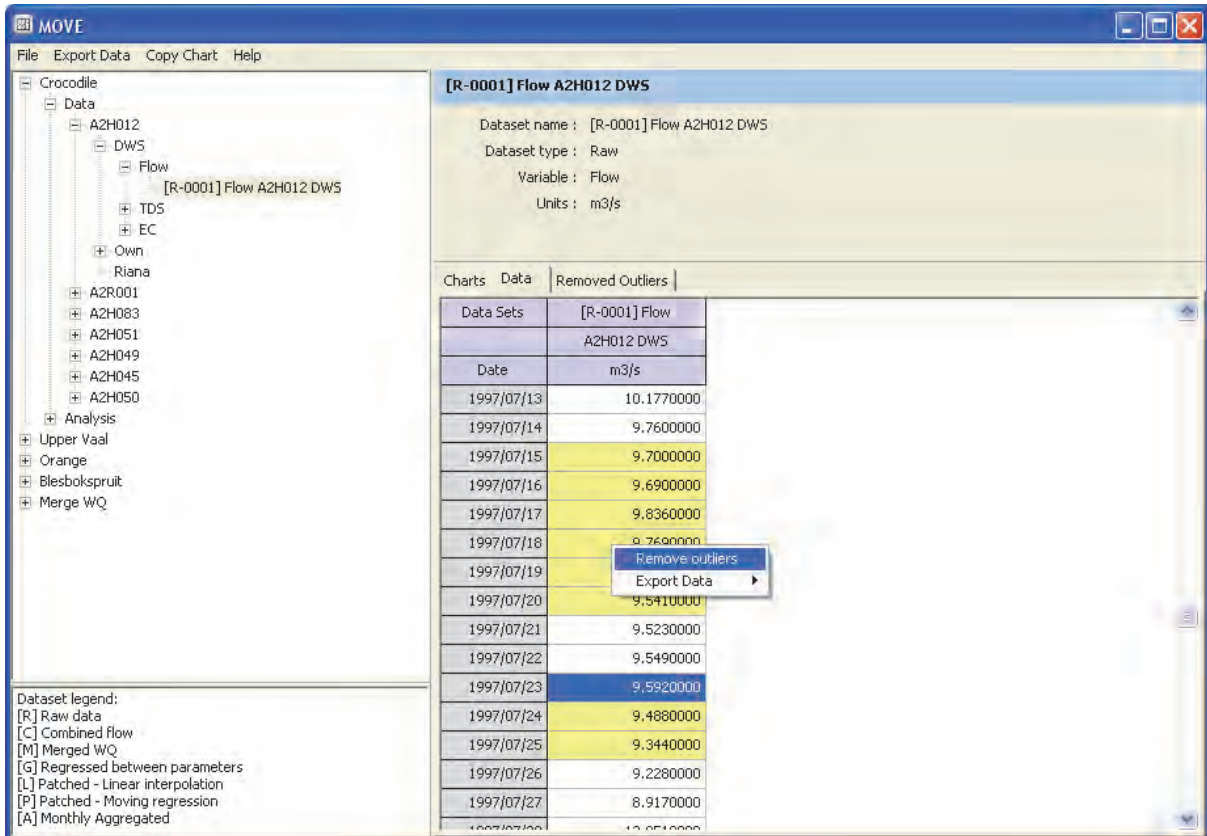


Figure 4.13: Bulk removal of outliers

#### 4.5.5 Export of tabulated outliers

A tabulation of outliers can be exported to a Comma Delimited (.CSV) or HTML format file by clicking on **Export Data** at the top right hand side of the screen (see Figure 4.14), while the **Removed Outliers** tab is open.

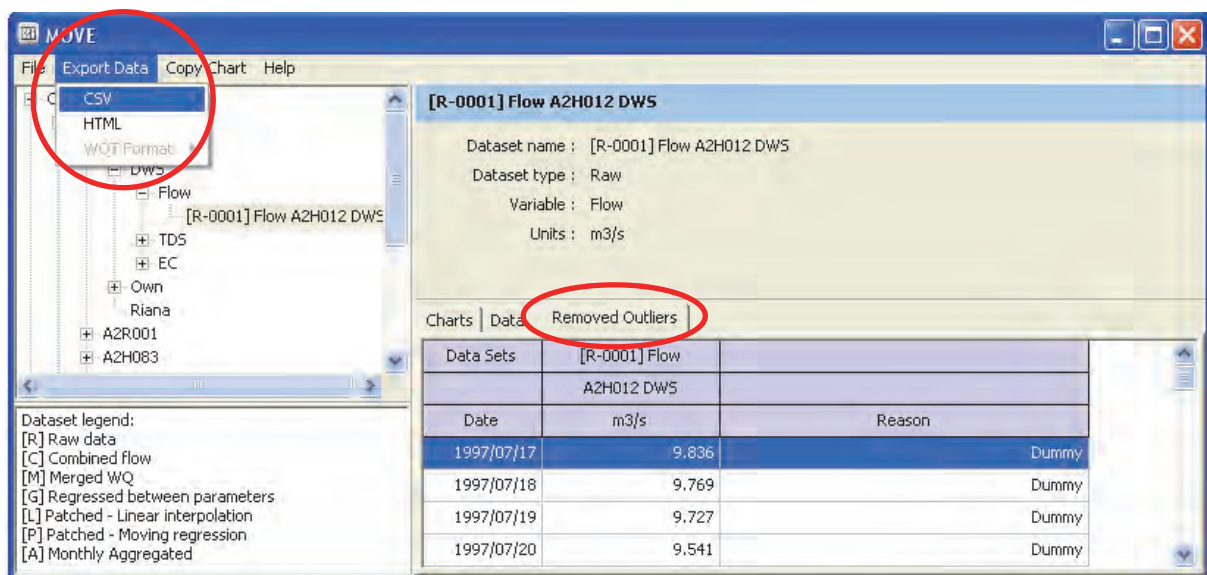


Figure 4.14: Exporting tabulation of outliers

Figure 4.15 shows an outlier report in .CSV format.

	A	B	C	D	E	F	G
1	Outliers : [R-0001] Flow A2H012 DWS						
2							
3	Date	Flow (m3/ Reason					
4	25/05/1997	59.681	Demonstration of the removal of an outlier				
5	26/05/1997	108.886	Demonstation.				
6							
7	Date Created : 2015-03-31 12:56:28						
8							
9							

**Figure 4.15: Outlier export file**

Such tabulations can be stored in **Project** reports and databases and sent to **Source** organisations to alert them to the identified outliers. After verifying the suspected outliers the **Source** organisation can then rectify their original database by making similar deletions or, at the very least, flagging the suspect data to avoid passing it on to other users or using incorrect data in internal studies. Once again, missing data are far preferable to misleading data.

All deletions of outliers are only made to the **Project** database and any exported files. The original raw data files remain unchanged.

#### 4.5.6 Export of Project data file

Clicking on **Export Data** while the **Data** tab is open exports the adjusted values to a .CSV file. Figure 4.16 shows such a data export file.

	A	B	C	D
1	DataSet : [R-0001] Flow A2H012 DWS			
2				
3	Date	Flow (m3/s)		
4	08/11/1996	4.918		
5	09/11/1996	4.922		
6	10/11/1996	4.62		
7	11/11/1996	4.386		
8	12/11/1996	5.684		
9	13/11/1996	5.252		
10	14/11/1996	6.457		
11	15/11/1996	18.802		
12	16/11/1996	9.123		

Figure 4.16: Data export file

#### 4.6 Import flow dataset dialog (from a user defined column format file)

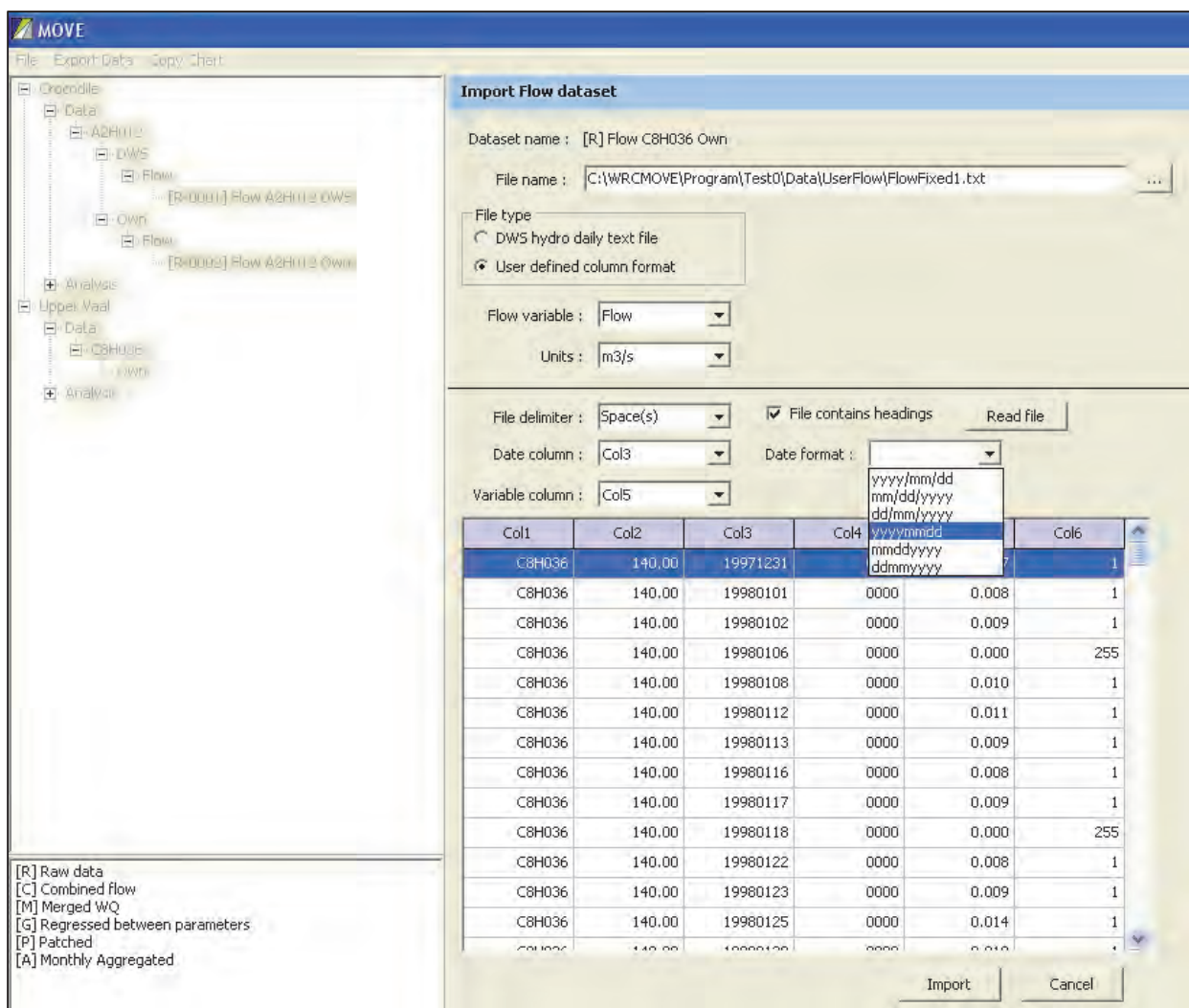
Provision is made for the importation of flow data from **Sources** other than the DWS's standard hydrological database, since other organisations have their own data storage conventions. This also accommodates the pre-processing of data that may have originated from the WMS system (or any other **Source**).

However, the stipulation is made that the data must be in column format with one date entry per line. Each line must contain at least two columns, one for the date and one for the corresponding flow value. Any number of columns in any order can be accommodated. The first row of the file may contain column headings.

Selecting the **User defined column format** radio button from the **Import Flow / Point dataset** dialog (Figure 4.10) opens the dialog shown in Figure 4.17.

There are two types of flow data that can be selected from the drop down list, namely **Flow** and **Point**.

When this option is selected the variable name is shown as **Flow** by default. This defines the variable as being daily river flow. (The alternative variable name that can be obtained from the drop-down menu is **Point**, which is described in Section 4.6.2.)



**Figure 4.17: Importing a user defined column format flow dataset**

#### 4.6.1 Daily river flow dataset

Gaps in a daily dataset time series are permissible. Such gaps can be filled using the standard missing data flow value (“-1”), or by simply omitting the missing data (resulting in a step in the chronological dates).

The onus is on the user to have already manipulated the data to remove unreliable data and present it as a file in column format.

Since external **Source** organisations can use a range of flow units, it is necessary to define the flow units using the drop down menu provided. Internally the flow units are converted to the **Project** standard of m<sup>3</sup>/s.

Valid (i.e. positive values) are automatically converted to the internal **Project** standard (m<sup>3</sup>/s). The appropriate conversion factors are as follows:

<u>Flow units</u>	<u>Factor</u>
m <sup>3</sup> /s	1.0000
l/s	0.0010
m <sup>3</sup> /day	1/86400
kl/d	1/86400
MI/d	1/86.4

The user defines the column delimiter used in the **Source** file and whether the file contains headings before selecting the **Read file** button. A tabulation of the file contents is then displayed in the bottom window, from which the user selects the column that contains the dates and defines the date format, and the column that contains the flow values.

Selection of the **Import** button then displays a chart of the flow data identical in form to Figure 4.12, which can be manipulated as before to remove outliers and export tables of outliers and **Project** data.

From here on the chart navigation and file export functions are identical to those described in sections 4.5.1 to 4.5.5. Similar capabilities apply to all of the remaining chart screens and export functions.

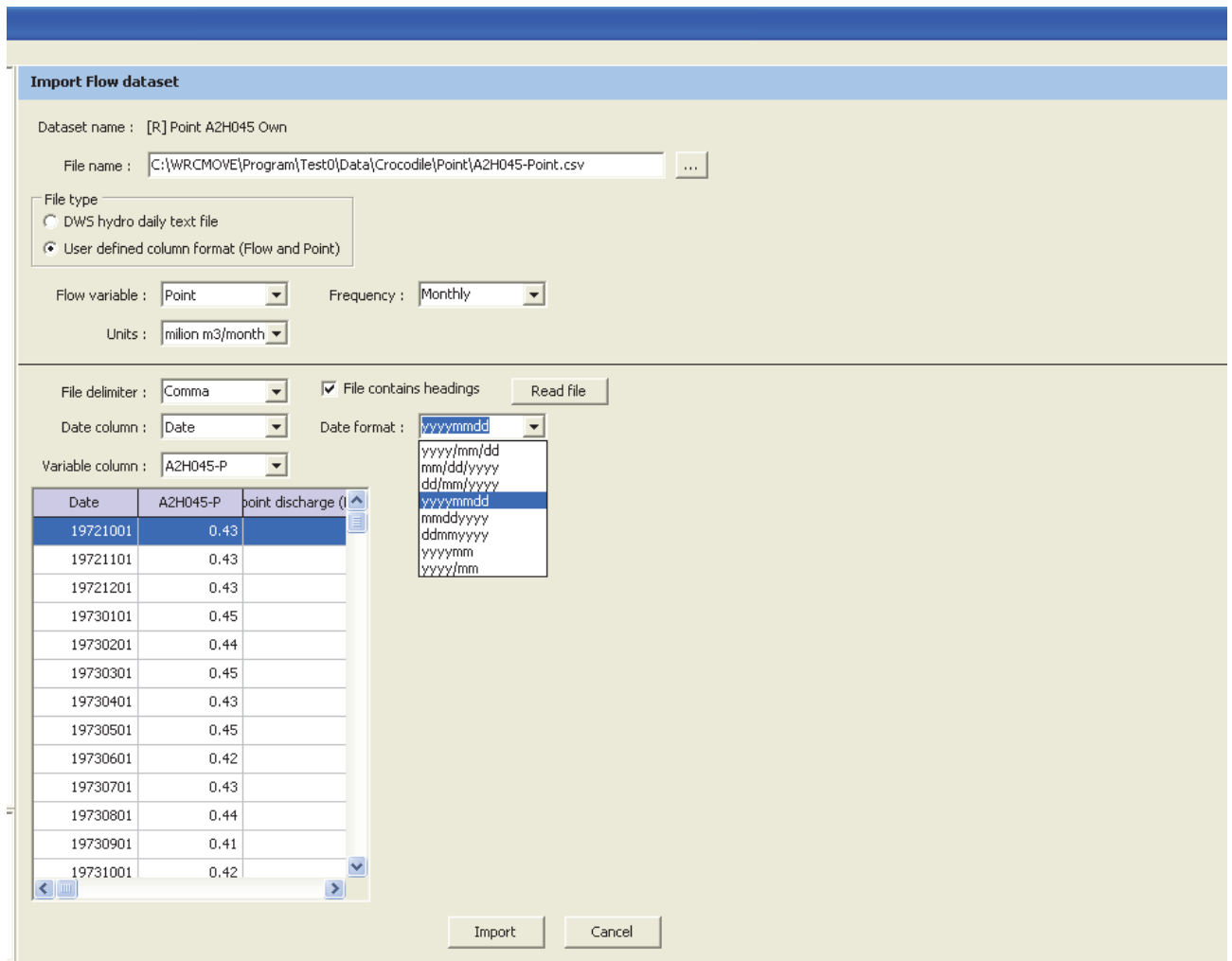
#### 4.6.2 Importing a point source flow dataset

Point source datasets are always in the form of user defined files. These differ from daily river flow data in that the input files are usually reported as monthly flows.

An important difference between river flow and point source datasets is that the latter comprises net cumulative values. As such negative values (representing net abstraction as opposed to net inflow) are perfectly valid. Another condition is that the prepared time series dataset must be complete with no gaps in the record for the entire patching period.)

After clicking on a **Source** in the navigation tree in the left hand panel, the **Import Flow Dataset** option is selected (Figure 4.9). A dialog the same as that for daily flows is opened (See Figure 4.17) and the **User defined column format** radio button is selected as before. Selection of **Point** from the **Flow variable** drop-down menu opens a new window from which a data frequency of either **Monthly** or **Daily** can be selected from the **Frequency** drop-down menu. Figure 4.18 shows the resulting dialog.

If the point flow data is provided at a daily frequency, the units can be in any of the standard daily flow units listed in Section 4.6.1.



**Figure 4.18: Importing a point source dataset**

If the data is supplied in the form of monthly data then the following standard units are available for selection from the monthly data drop down list:

Flow units	Factor
kl/month	$1/(86400 \cdot NM)$ - where NM is the number of days in the month
MI/month	$1/(86.4 \cdot NM)$
$10^6 m^3/month$	$1/(0.0864 \cdot NM)$

In such instances the corrected daily flow in  $m^3/s$  is assigned to every day in the month. This approximation is generally acceptable since point source flows tend to exhibit less variation than is the case for catchment runoff.

Selection of the **Import** button produces an interactive chart identical in form to Figure 4.13.

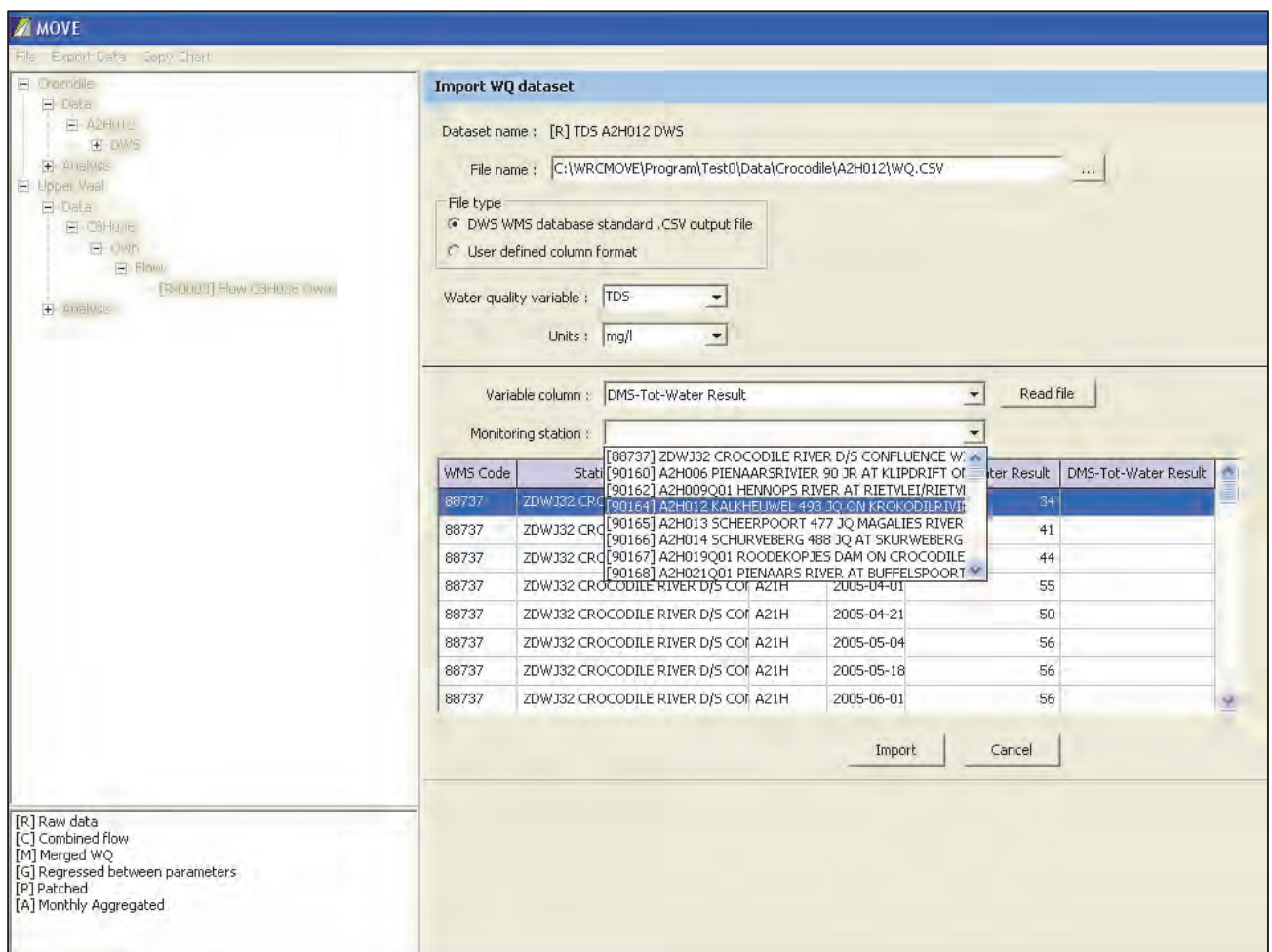


## 4.7 Import water quality dataset dialog

Water quality data can be imported from a standard comma delimited output file derived from the DWS's Water Management System (WMS) database, or from a user defined data file.

### 4.7.1 Importation from a DWS WMS standard .CSV output file

Selection of the **Import WQ Dataset** option (see Figure 4.9) opens the screen shown in Figure 4.19.



**Figure 4.19: Importing a water quality dataset from a DWS WMS standard .CSV output file**

First the browse button is used to navigate to the data file containing the desired standard water quality WMS database export file. In this case the file WQ.CSV was selected, which is a DWS file generated by the Water Management System (WMS) database. Such WMS files typically contain a large number of water quality variables and data for a number of monitoring stations. Several rows of grab sample data are stored for each monitoring station.

Each water quality variable has columns for the analysed variable concentration and the detection limit. If the concentration is set equal to the detection limit, then the program sets the value at half of the detection limit, based on the deduction that the concentration must lie somewhere at or between zero and the detection limit. This is especially important for nutrients and heavy metals where most sample analysis results are below the detection limit, with only occasional values rising to elevated concentrations. In such circumstances the mean and median concentration can be substantially exaggerated if all the values that are below the detection limit are interpreted as being equal to the detection limit. Setting them to half of the detection limit provides a more realistic estimate of the likely true sample mean.

Selection of the **DWS WMS database standard.CSV output file** radio button notifies the program to expect data in this standard format.

Since each **Source** organisation has its own nomenclature and system of units for water quality variables and data can be derived from several such **Sources** for different **Sites** used in a **Project**, it is essential to define a standard set of **Project** variable names and units. It is also desirable to use short standard names so as not to clutter chart and table captions. Drop down lists of already defined standard variable names and units are provided, from which the appropriate values can be selected. In this example, the variable name TDS was chosen to represent the Dissolved Major Salts (DMS) values provided by the WMS and the standard units of mg/l. New variable names can be added to the drop down lists as required. These remain in the drop down lists for future use in the **Project**.

Selecting the **Read File** button populates the model with the WMS database information to aid in visualisation of the key data. At the same time drop down lists of available water quality variables (**Variable column**) and monitoring station descriptions (**Monitoring station**) are populated. The user then selects the desired water quality variable (in this case entitled "DMS-Tot-Water Result") and the desired monitoring station ("[90164] A2H012 KALKHEUWEL 493 JQ ON KROKODILRIVIER").

Human intervention is required to determine the water quality variable type used in the WMS database (in this case "DMS" means dissolved major salts) that corresponds to the specified standard variable ("TDS" – total dissolved solids). TDS and DMS are not identical, but are closely aligned. TDS is derived from evaporation of the sample and weighing the resulting sediment, while DMS is calculated from the sum of significant ions. But the one is a good surrogate of the other.

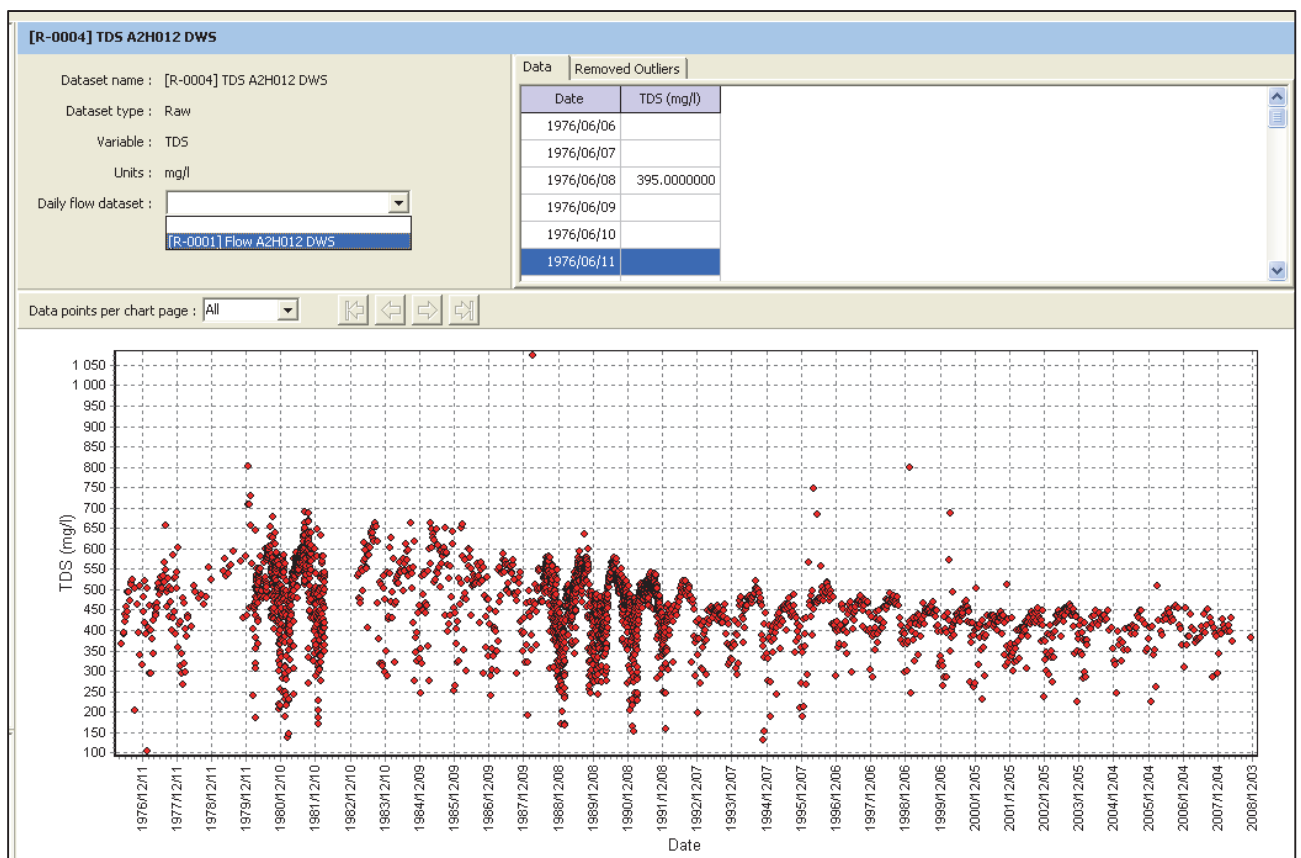
The model has an internal list defining the units used in the WMS system for various water quality variables. In this instance the units for DMS is mg/l. These are compared with the



units specified for the standard variable (TDS), which also has units of mg/l, so the conversion factor used to standardise the values in the WMS system is 1.0. (If the standard units were, say, µg/l then the WMS values would have had to be multiplied by 1000 to bring them to the same specified **Project** standard.)

Most of the water quality variable units used in the WMS database are in mg/l, although some, like electrical conductivity (mS/m) and coliforms (counts/100 ml), are uniquely different.

Selecting the **Import** button generates the water quality variable chart shown in Figure 4.20.

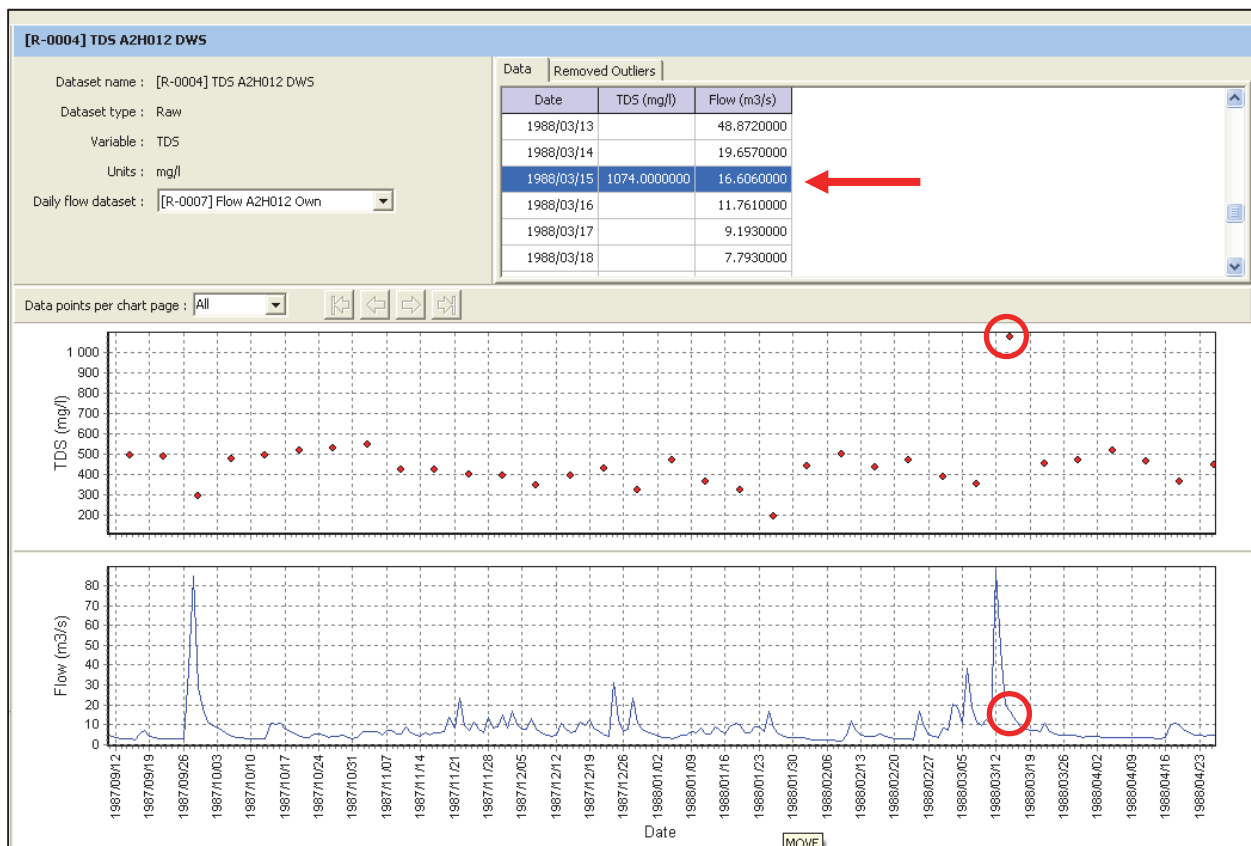


**Figure 4.20: Interactive water quality chart**

The daily flow plot of an existing **Project** dataset can also be added to a water quality chart via the drop-down menu, as shown in Figure 4.21.

Plotting both water quality concentration and flow provides an excellent means of identifying both water quality and flow outliers. For example, an unusually low salinity value during a low flow period or a period with a gradually changing flood recession limb is often indicative of an incorrect sample value. Similarly, a very high salinity value during a flood can also indicate a suspect water quality sample. This is the case on 15 March 1988 (ringed in red in Figure 4.21) when the sharp rise in TDS concentration to 1074 mg/l is inconsistent with the

high flow of 16.6 m<sup>3</sup>/s during a smooth flood recession limb (compare with the preceding and succeeding days seen in the plot and the tabulated values). This can cause substantial over estimation of the observed TDS load. If the rise in TDS was caused by a saline spillage it would be reasonable to expect a corresponding rise in flow, which did not occur. The conclusion is that the TDS sample on this day is highly suspect and should be removed to avoid misleading results.



**Figure 4.21: Water quality chart with corresponding flow**

Conversely, zero flows that are unsupported by high (or missing) salinity values indicate incorrect flows, especially if the drop in flow rate is sudden and occurs during the wet season. This is a common error most often attributable to **Source** organisations or users incorrectly representing missing data as zero values. This is a serious error when it occurs during wet periods because such gaps in flow records most often occur when the modular limit of the weir is exceeded during big flood events. Hence the zero values are diametrically opposed to the true change in flow. Retaining erroneous zero values is extremely misleading as it can lead to serious under-estimation of the salt loads entering downstream dams.

The relative heights of the water quality and flow chart panels can be adjusted by clicking on the horizontal splitter and dragging it up or down. The same applies to all other multi-chart screens.

#### 4.7.2 Importing a water quality dataset from a user defined file

As before, selection of the **Import WQ Dataset** option (see Figure 4.9) opens the right hand panel shown in Figure 4.19.

Selecting the **User defined column format** radio button opens a dialog the top half of which is identical to Figure 4.19 and allows browsing to select the **Source** data file, the standard **Project** variable name and units used by the **Source** organisation.

The bottom portion of the right hand panel is similar to that for importing user defined flow files (see Figure 4.17), which facilitates definition of the file column delimiter, the header line control, and reading in the defined input file. The date and variable columns and the date format are then selected.

The standard water quality units used internally in the system and the appropriate conversion factors are listed in Table 4.1.

**Table 4.1: Water quality variable adjustment factors**

Standard Units	User-specified Units <sup>1</sup>	Standard Code	Affected variables	Adjustment factor
mg/l	µg/l	NH4	Ammonia	0.001
		Ca	Calcium	
		Cl	Chloride	
		F	Fluoride	
		Fe	Iron	
		Mg	Magnesium	
		Mn	Manganese	
		NO3	Nitrate	
		NO2	Nitrite	
		K	Potassium	
		Na	Sodium	
		Si	Silica	
		SO4	Sulphate	
		SS	Suspended solids	
		TDS	Total dissolved solids / salts	
TAL	Total alkalinity			
TH	Total hardness			
µg/l	mg/l	various	various heavy metals	1000.
mS/m	µS/cm	EC	Electrical conductivity <sup>2</sup>	0.10
counts/100 ml	Counts/100 ml	<i>E. coli</i>	Escherichia coli	1.0
		FC	Faecal coliforms	
		TC	Total coliforms	
-	-	pH	pH	1.0
NTU	NTU	Turb	Turbidity	1.0
°C	°C	TEMP	Temperature	1.0

**Notes:**

- 1 In instances where the user-defined units differ from those listed, the user will need to carry out appropriate file manipulations to convert to recognised units.
- 2 In the DWS' WMS database electrical conductivity is corrected to 25°C. Some users like Rand Water correct to 20°C. Since electrical conductivity is sensitive to temperature, it is important to obtain appropriate temperature conversion tables to adjust the electrical conductivity values to a common base. This needs to be done during the pre-processing stage. (Generally this is not necessary if EC is a secondary variable used to patch TDS at a single **Site** since then the TDS and EC would be regressed against one another using analyses derived from common samples, rendering the units immaterial, provided they remain the same throughout the processing period.)

Selection of the **Import** button produces a water quality chart identical in form to Figure 4.20. A plot of daily flows can be added to the chart as described before in Section 4.7.1.

#### 4.8 Combine flow datasets dialog

The objective is to patch daily flow data at the water quality monitoring point using the flow records at each flow gauging station and taking account of the priority of each station.

This procedure is used when the water quality sampling point to be patched does not correspond to a flow gauging point. It is also used to patch deficient daily flow data using flow gauging records at adjacent **Sites**.

#### 4.8.1 Theoretical overview

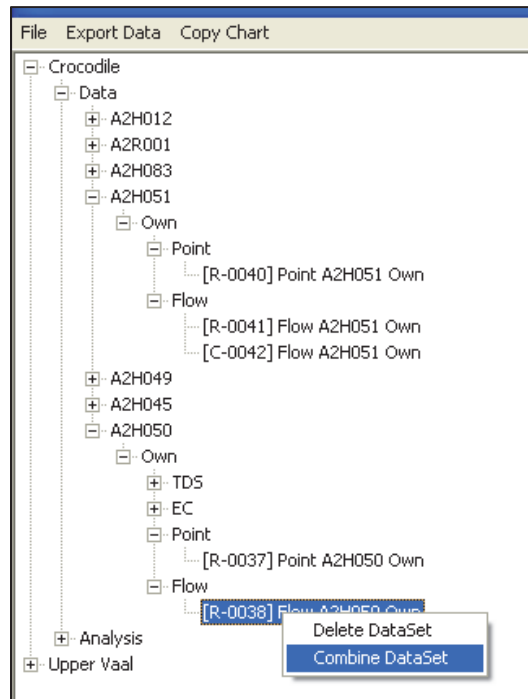
The theory is provided in Section 3.1 and comprises the following steps:

- Naturalise the flow record(s) by reversing the net effect of upstream point source inputs and abstractions.
- If there is more than one usable flow gauge, prioritise the flow **Sites** taking account of proximity to the water quality monitoring point, similarity of the intervening catchments and the reliability and length of flow records.
- The mean naturalised runoffs for the highest and second highest priority stations are calculated for periods of overlap between the two stations. The ratio between the two mean flows is used to scale the observed flows at the second priority station to patch the naturalised daily flows at the first priority station for all days when observations at the first priority station are missing but there are data at the second priority station. This patched record then becomes the new adjusted record at the first priority station.

This process is repeated to further patch the adjusted record at the first priority station using the record at the third priority station, and so on, until all of the stations have been used or until no further patching is required.

- The ratio of the catchment area upstream of the water quality sampling point to that of the highest priority flow gauging point is then used to scale the patched flows at the highest priority flow gauging station to estimate the naturalised daily flows at the water quality sampling point.
- The historical daily flows at the water quality sampling point are then estimated by adding the net point inflow upstream of the water quality sampling point.

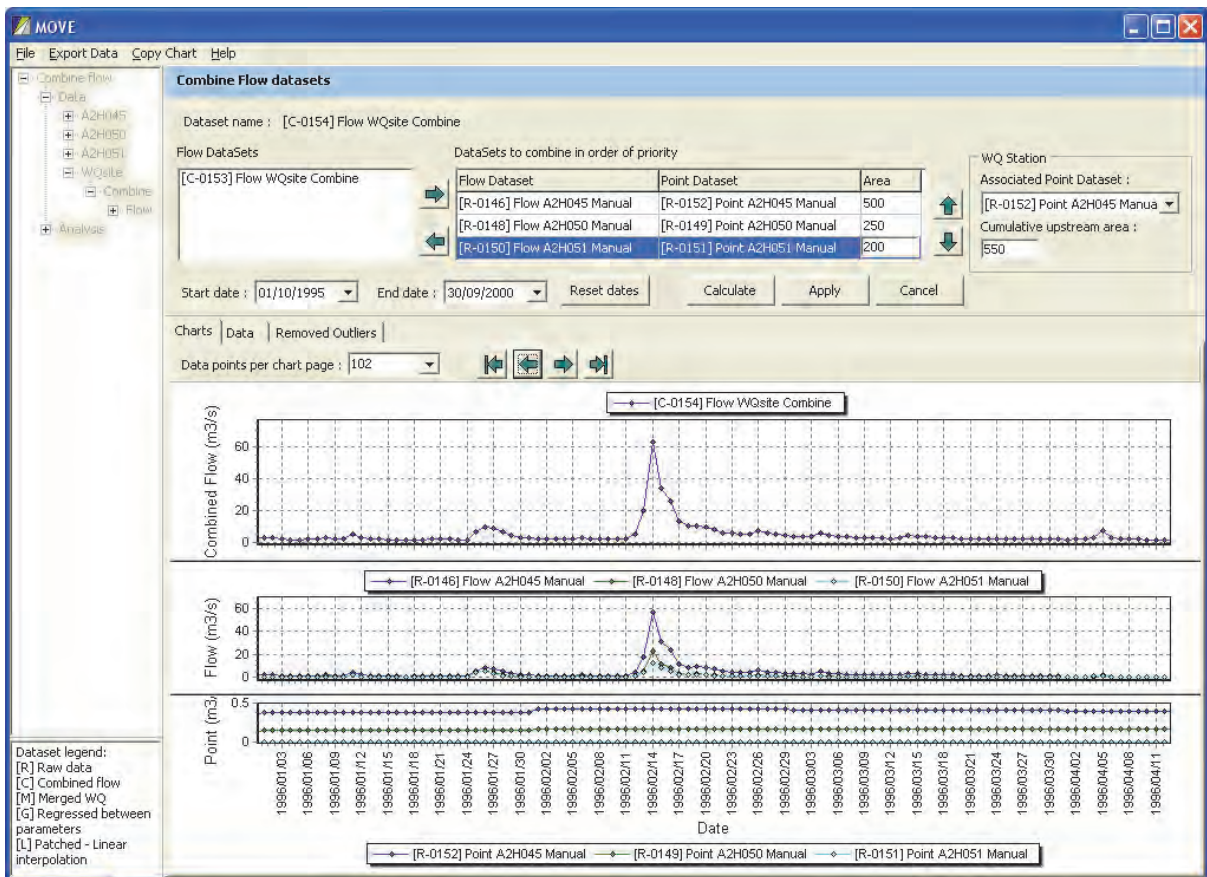
Clicking on a raw (prefixed by a “[R]”) daily flow (denoted by “Flow” in its name) dataset in the right hand panel opens a pop-up menu from which the **Combine Datasets** option is selected (see Figure 4.22).



**Figure 4.22: Selection of the Combine Dataset option**

The **Combine Datasets** dialog that opens in the right hand panel is shown in Figure 4.23.

The daily flow dataset from which the **Combine Datasets** dialog was opened automatically appears in the first column of the CheckBox labelled “Flow Dataset”. The remaining available **Project** flow datasets are listed in the left hand CheckBox. Highlighting one of the files in this window makes it available to be moved to the right hand window by using the green horizontal right-pointing navigation arrow. Similarly, highlighting a selected dataset in the Middle CheckBox allows it to be deselected and moved back to the available pool in the left hand CheckBox using the green left-pointing navigation arrow.



**Figure 4.23: Combine water quality datasets dialog**

As each daily flow dataset is added, a new daily flow time series is stacked onto the middle chart.

Highlighting a flow dataset in the middle CheckBox allows its priority to be increased or decreased using the up and down green navigation arrows to the right of the CheckBox.

The second column ("Point dataset") is initially populated with "No point input" entries. If this default is accepted then all the daily point inputs are set equal to zero. Clicking on one of the "Point dataset" entries in the second column of the middle CheckBox opens a drop-down menu showing available point source data sets. The one corresponding to the net point inflow to the cumulative catchment upstream of the "Flow dataset" should be selected. (The net point inflow is the sum of all point inflows less the sum of all abstractions. As such negative net point inflows are acceptable, when the abstraction exceeds the point inflow.)

As each point source dataset is added a new daily point source time series is stacked into the bottom chart panel. The point source time series is given the same colour coding as its corresponding daily river flow time series

The area column is initially populated with zero entries, which must be replaced with the total catchment areas upstream of each flow gauging station.

Next the dataset containing the net point inflow upstream of the water quality sampling station (“Associated Point Dataset”) is specified by either accepting the default or using the drop-down menu to select the appropriate dataset. The catchment area upstream of the water quality sampling point (“Cumulative upstream area”) is also defined.

Selecting the **Calculate** button calculates and displays the patched historical daily flows at the water quality sampling point.

Selecting the **Apply** button consolidates the calculation and places the patched flow dataset (prefixed by [C]) in the tree in the left hand panel.

#### 4.9 Regress primary and secondary water quality datasets

Infilling gaps in the primary water quality record by regression with a secondary water quality variable is undertaken when there is a strong relationship between the two variables. The most common application is the use of Electrical Conductivity (EC) as a surrogate for Total Dissolved Solids (TDS).

The theory is described in Section 3.2.

Right clicking on any of the water quality datasets under the **Data** folder in the left hand panel opens the pop up menu shown in Figure 4.24, which allows selection of the **Regress Dataset** option. This action defines the primary water quality variable (in this example **[R-0004] TDS A2H012 DWS**, a TDS dataset).

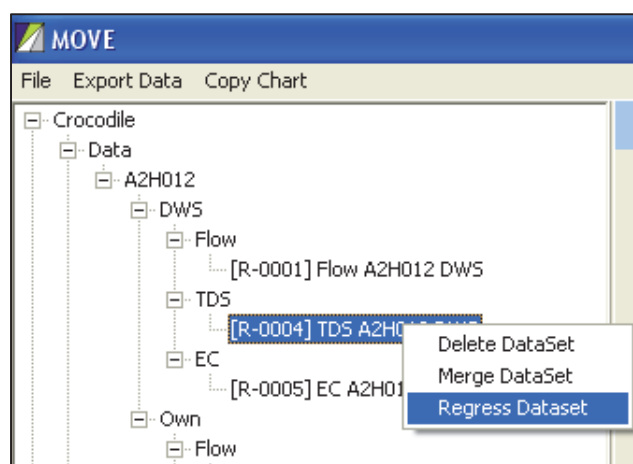
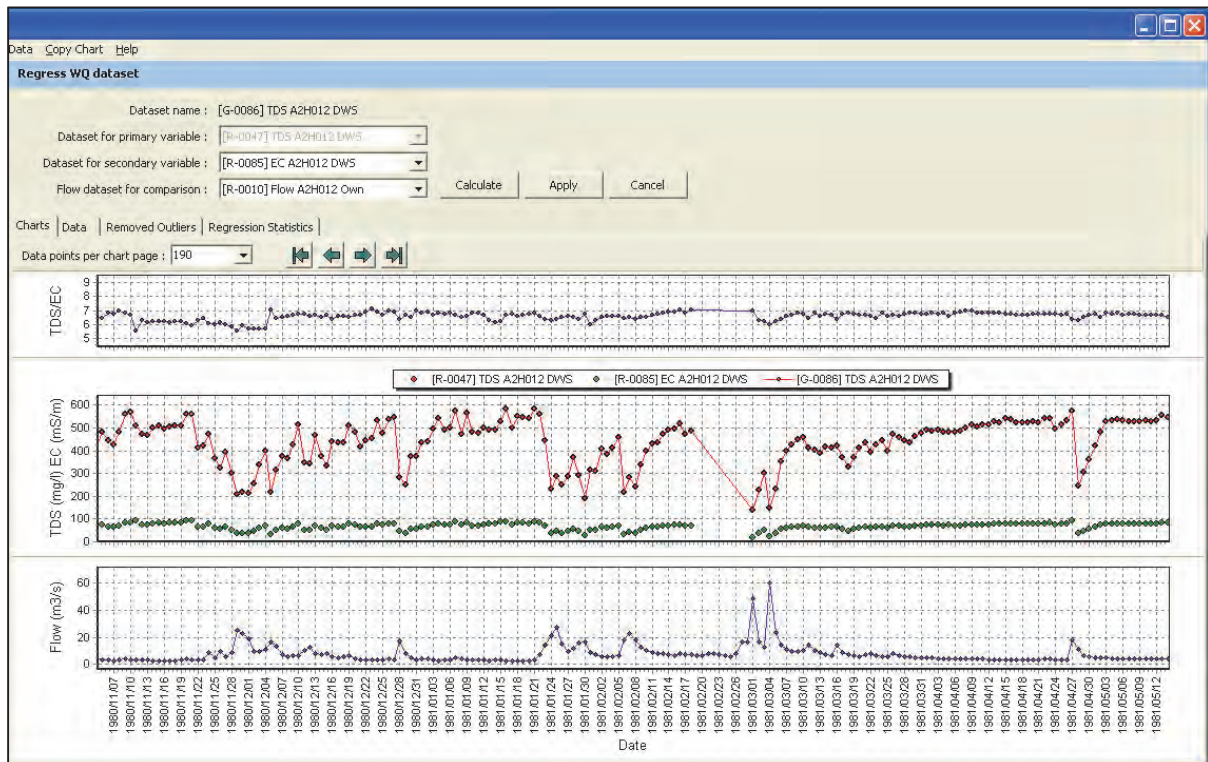


Figure 4.24: Selecting option to regress water quality datasets



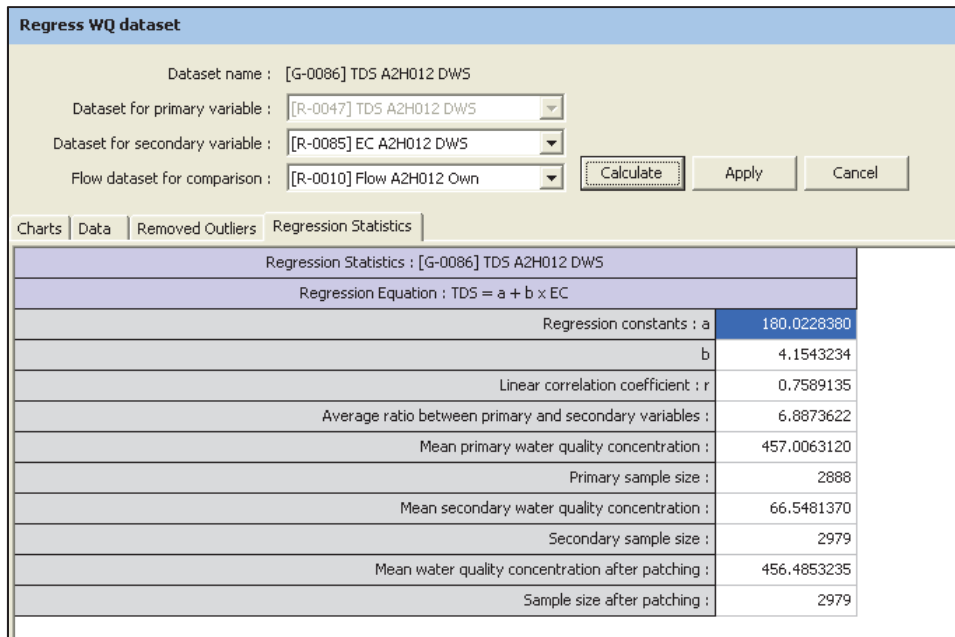
This selection opens the **Regress WQ dataset** dialog shown in Figure 4.25.

Initially this dialog shows a chart of only the primary (i.e. TDS) variable. As soon as the secondary variable (i.e. EC) is selected from the drop down menu a separate chart panel is shown underneath the primary variable and the ratio between the two is shown in the top chart panel. If available, it is always very useful to also include a plot of the daily flows (bottom chart panel). Although this plays no role in the regression, concurrent flow values are very useful in identifying and confirming outliers.



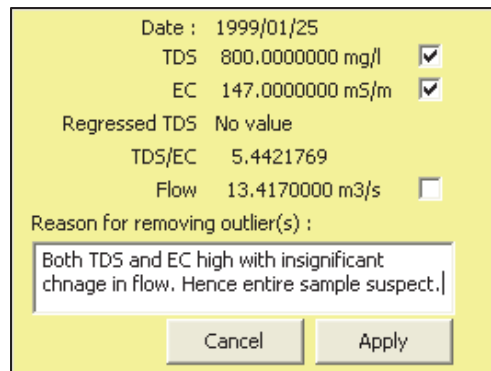
**Figure 4.25: Water quality regression dialog**

Selecting the **Calculate** button calculates the regression and displays the regression constants (intercept and slope) and correlation coefficient (see Figure 4.26).



**Figure 4.26: Display of regression equation information**

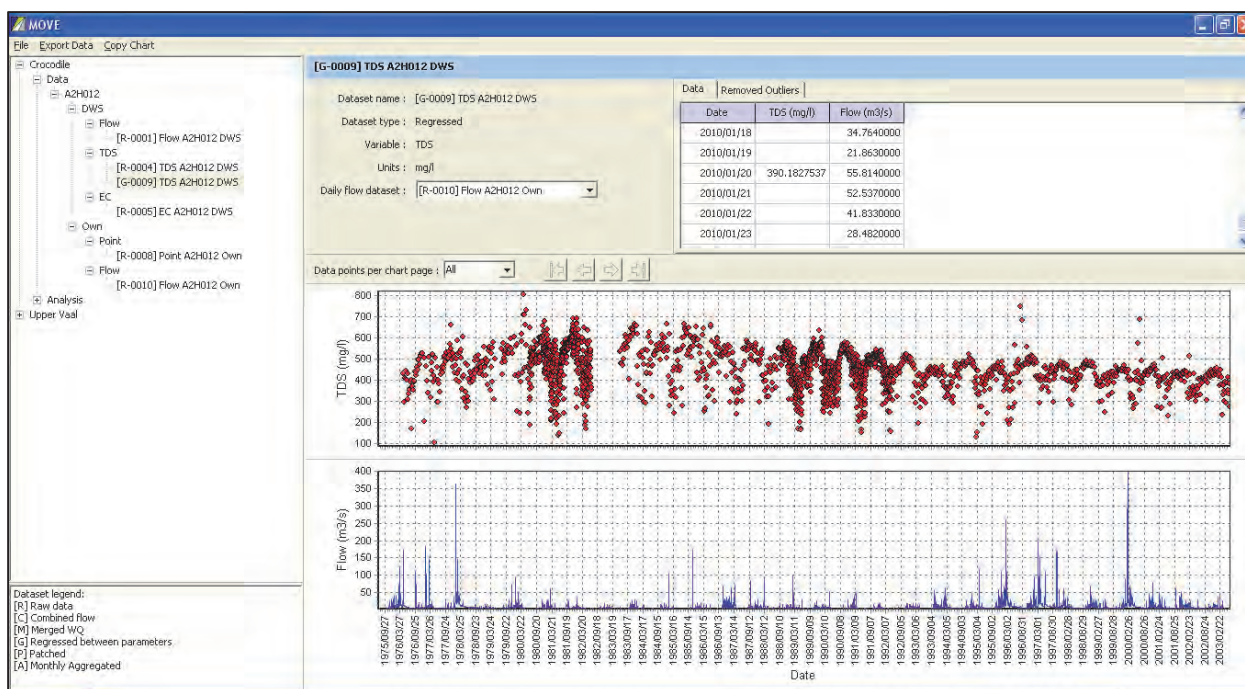
The outlier pop-up dialog can be accessed by clicking onto a point in any of the three chart panels (see Figure 4.27).



**Figure 4.27: Water quality regression dialog**

This permits checking of one or more of the primary or secondary water quality parameters or the corresponding flow for deletion. When the Apply button is selected the outliers are removed from all of the checked datasets and the details are added to the list of outliers.

After removing outliers it is important to recalculate the regression equation (**Calculate regression** button). In the case of station A2H012 on the Crocodile River, removing just three of the most obvious outliers increased the correlation coefficient from 0.759 to 0.940 and radically changed the slope and intercept of the regression line (the slope increased from 4.15 to 6.38). This emphasises just how important it is to remove outliers before further processing. Figure 4.28 shows the result after selecting the Apply button.



**Figure 4.28: Primary water quality variable after patching by regression with secondary variable**

The regressed patched primary water quality values are stored in a **Project** dataset prefixed by [G] that is displayed in the left hand panel navigation tree, named **[G-0009] TDS A2012 DWS** in this example. The dialog screen shown in Figure 4.28 bears the same name.

The interactive plot can be used to remove and document identified outliers, which are removed from the primary and/or secondary [R] datasets. Outliers can be viewed and exported as necessary.

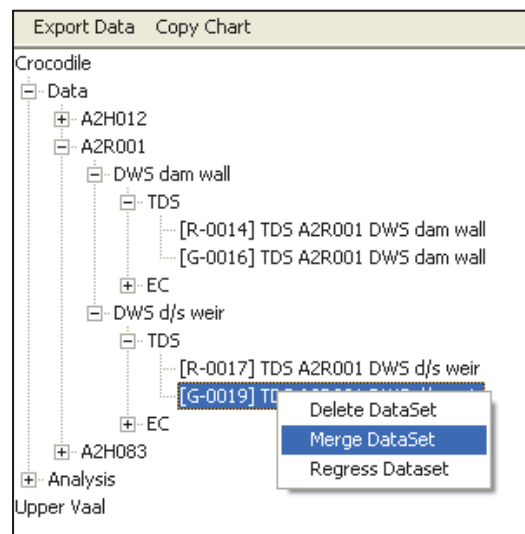
#### 4.10 Merging water quality datasets

Merging of datasets is carried out when two or more monitoring stations are close together, such as when an old station is closed and replaced by a new nearby one, or at reservoir stations that typically have sampling stations at a downstream weir component, in the lake at the spillway and in one or more outlet canals or pipe abstractions. Merging of water quality datasets is also required when the DWS's WMS database contains more than one sampling code for the same point. This often occurs when new DWS personnel take on pollution control activities and open "new" water quality monitoring stations that are actually identical to existing stations. As many as three different station codes have been identified for the same locality, each covering significant portions of the historical record. There are also places where more than one organisation takes water quality samples. For example at some

sites in the Waterval River samples are collected by Evander Gold Mines, Sasol, Rand Water and DWS, and at one time DWS also had continuous EC probes in operation.

When more than one component has been established at a dam, routine sampling is usually carried out at all sampling sites whenever the site is visited. Under such circumstances sampling at the highest priority site will usually suffice. However, all of the sampling stations are not established at the same time. Sampling in the lake near the dam wall often precedes sampling at other localities by several years. While the downstream weir is often the most representative sampling station (since it aggregates both bottom releases and top spillages, which can exhibit significant differences due to stratification) such a weir might only have been built some years after the construction of the dam. Moreover, during large floods access to the weir site may be restricted and only the dam wall water surface sampling is available. Hence there is a need to merge water quality datasets at most dam sites.

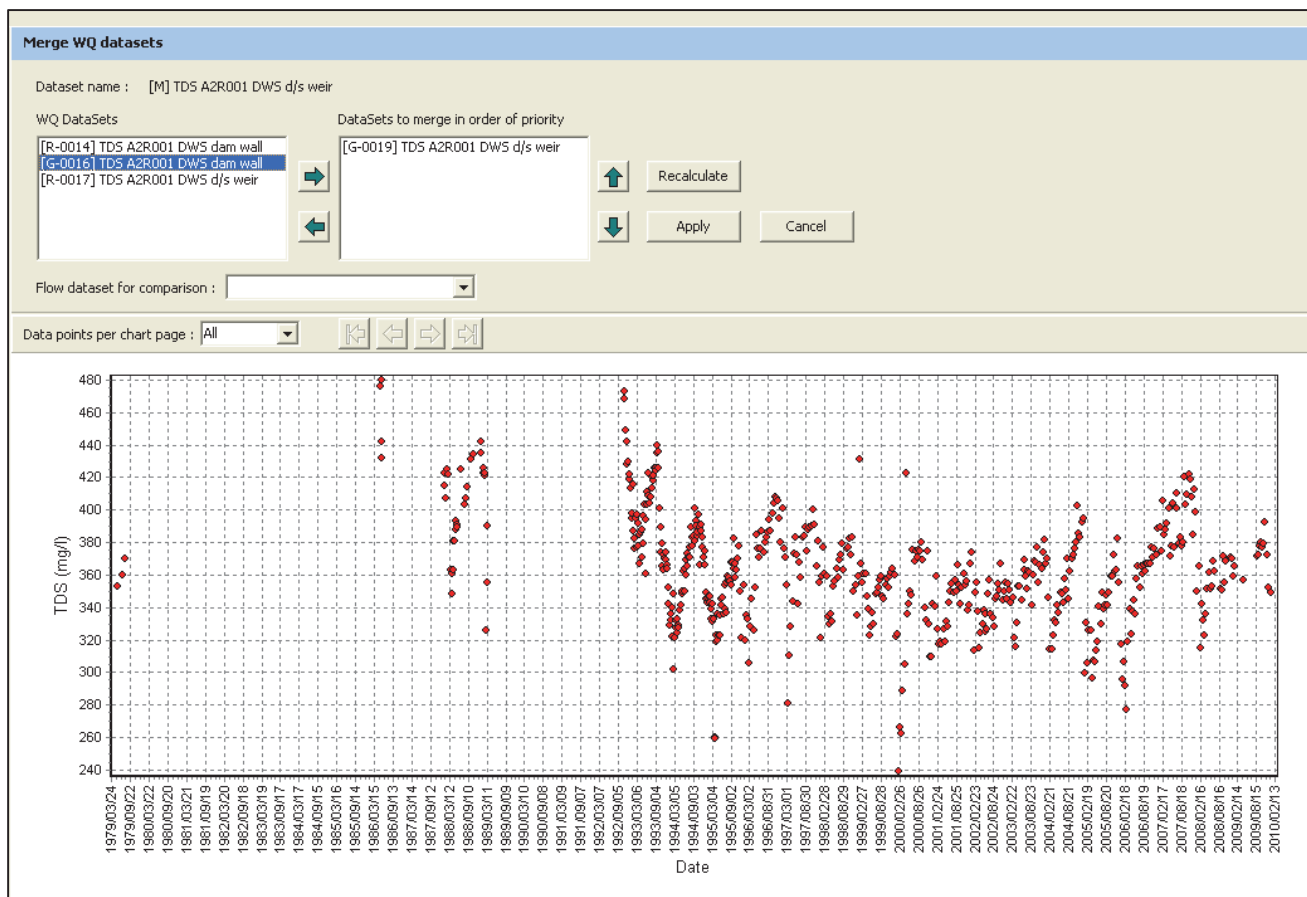
Clicking on a raw [R] or regressed [G] water quality dataset in the left hand panel opens a pop-up menu from which the **Merge datasets** option is selected (see Figure 4.29).



**Figure 4.29: Selection of the Merge datasets dialog**

The **Merge WQ datasets** dialog that opens in the right hand panel is shown in Figure 4.30.

The water quality dataset that was initially selected automatically appears in the top right hand ListBox. All raw [R] and merged [M] datasets for the entire **Project** are placed in the left hand ListBox ready for possible selection. All of them are made available for selection because of necessity merging must involve more than one **Site** and possibly **Source**. Which ones will be needed is only defined when datasets are selected and moved into the right hand ListBox.



**Figure 4.30: Merge water quality datasets dialog (initial display)**

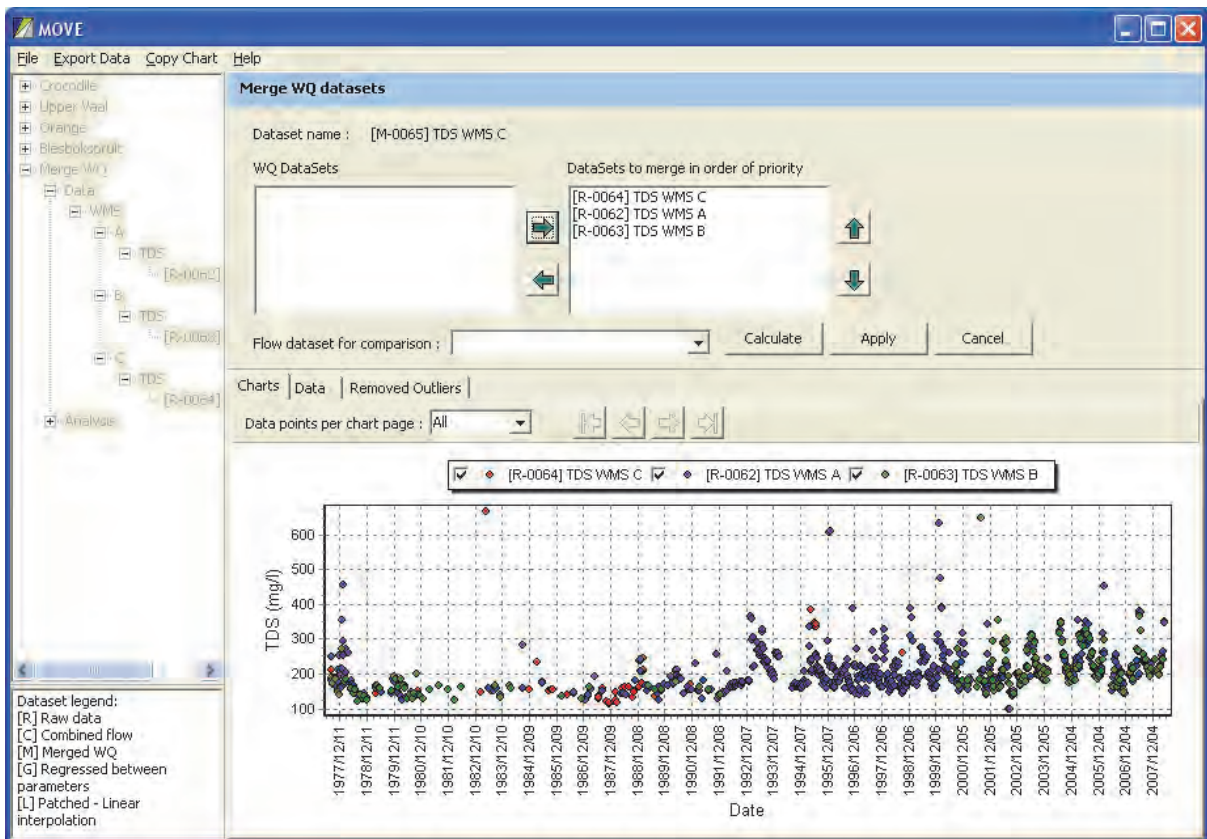
Highlighting one of the files in the left hand ListBox makes it available to be moved to the right hand ListBox containing the selected datasets by using the top green right-pointing navigation arrow. Similarly, highlighting a previously selected dataset in the right hand ListBox is de-selected and moved to the available pool in the left hand ListBox when the green left-arrow is clicked.

Highlighting an entry in the right hand ListBox also allows its priority to be increased or decreased using the green up and down navigation arrows.

The chart denotes every selected water quality dataset in a different colour to facilitate their identification (see Figure 4.31). This is very important since sometimes this comparison shows that one or more of the datasets is too dissimilar to rationally be merged with the others. This can only be discerned from visual inspection of the results.

The **Recalculate** button merges the datasets selected in the right hand top window in the specified order of priority. The extended merged dataset is immediately displayed in the water quality chart.





**Figure 4.31: Merge water quality datasets dialog**

Selecting the **Apply** button consolidates the changes into a merged water quality dataset, prefixed by **[M]** in the left hand navigation panel.

#### 4.11 Initialisation for data analysis

The data patching option is accessed by first opening the **Analysis** option from the decision tree in the LHS panel by clicking on the “+” sign to the left of **Analysis**, or double clicking on **Analysis**. This action opens up the **Sites** that have previously been set up from the **Data** option, and all the subsidiary imported flow and water quality datasets are available to be used in the **Analysis** option.

Right clicking on the appropriate **Site** opens the pop-up menu shown in Figure 4.32 from which the **Patch WQ dataset** option can be selected.

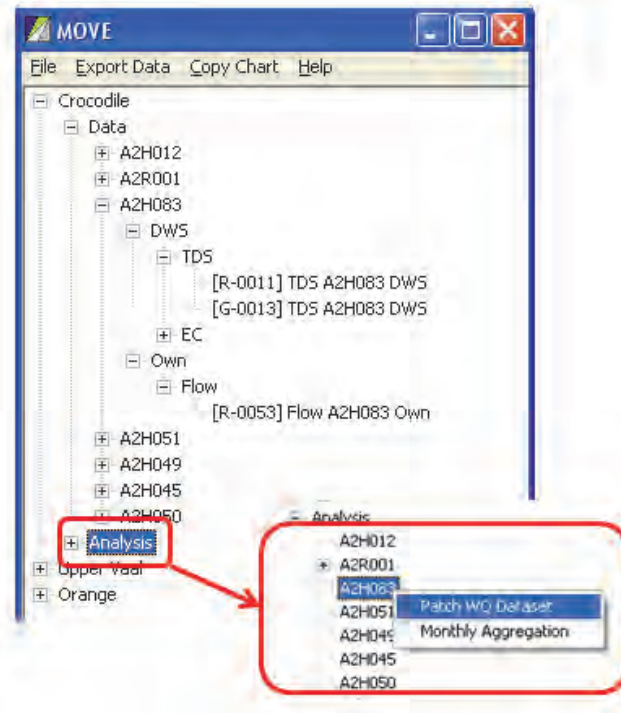


Figure 4.32: Selection of water quality patching option

#### 4.11.1 Patch Water Quality Dataset dialog

Clicking on the **Patch Water Quality Dataset** option opens up the **Patch Water Quality Dataset** initialisation dialog in the right hand side panel (see Figure 4.33).

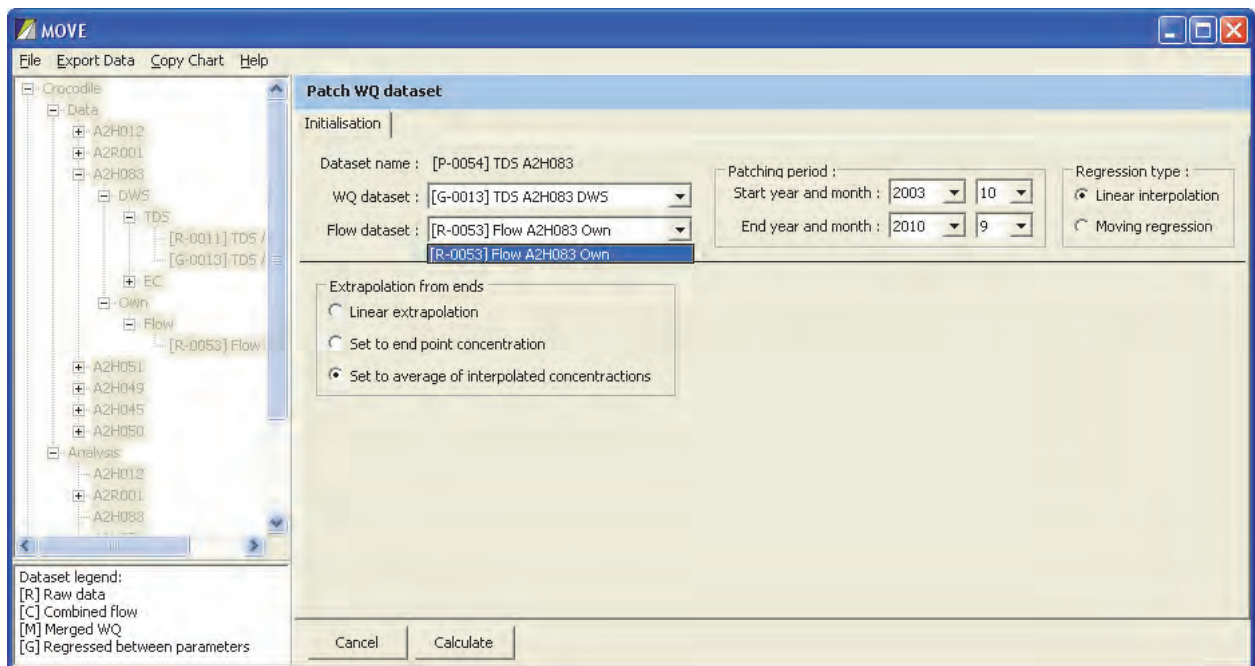


Figure 4.33: Initialisation for water quality patching

The top section of this panel requires selection of the following:

- A pre-prepared water quality dataset and a flow dataset, i.e. a [R] (raw), [C] Combined, [M] (merged) or [G] (regressed) type file.
- The start and end year and month of the patching period.
- The type of patching, which can be:
  - Linear interpolation, or
  - Moving regression.

When first opened from the left hand side panel decision tree, the linear interpolation initialisation tab is displayed.

#### 4.12 PATCHING BY LINEAR INTERPOLATION

Linear interpolation is appropriate for patching observed water quality data at reservoir stations where storage attenuation and managed releases obscure the correlation between water quality variable concentration and flow date.

The theory is provided in Section 3.3.

The Patch Water Quality Dataset dialog opens with the linear interpolation option checked (see figure 4.33). Before proceeding the most appropriate extrapolation method should be selected to define the method for patching the daily data from the first observed data point to the start of the month in which it is located. This also needed for the short period from the last observed data point to the end of the month in which it is located. (These extrapolations are extended much further, but patching over a gap longer than one month is not recommended.)

Three extrapolation options are provided:

- Linear interpolation beyond the end points
- Set to the end point concentration
- Set to the average interpolated concentration.

Selection of the **Calculate** button from the **Patch WQ dataset** dialog (Figure 4.34) calculates and plots the linear interpolation. The plot chart can be stored to the desktop by selecting **Copy Chart** from the top of the screen.



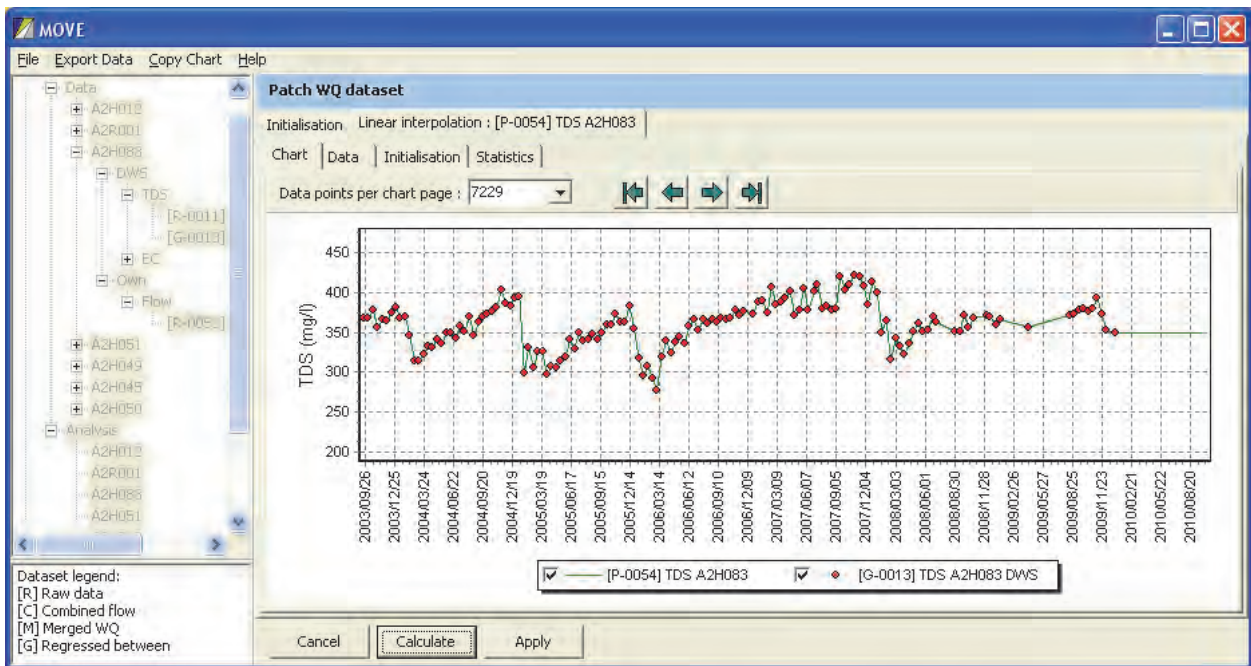


Figure 4.34: Plot of patched water quality dataset

#### 4.12.1 Time series data tabulation

Selecting the **Data** tab (see Figure 4.34) provides a tabulation of observed and interpolated concentrations.

The tabulation can be exported to the desktop by selecting the **Export Data** option (ringed in red in Figure 4.35). The table can be exported either as a HTML or .CSV file.

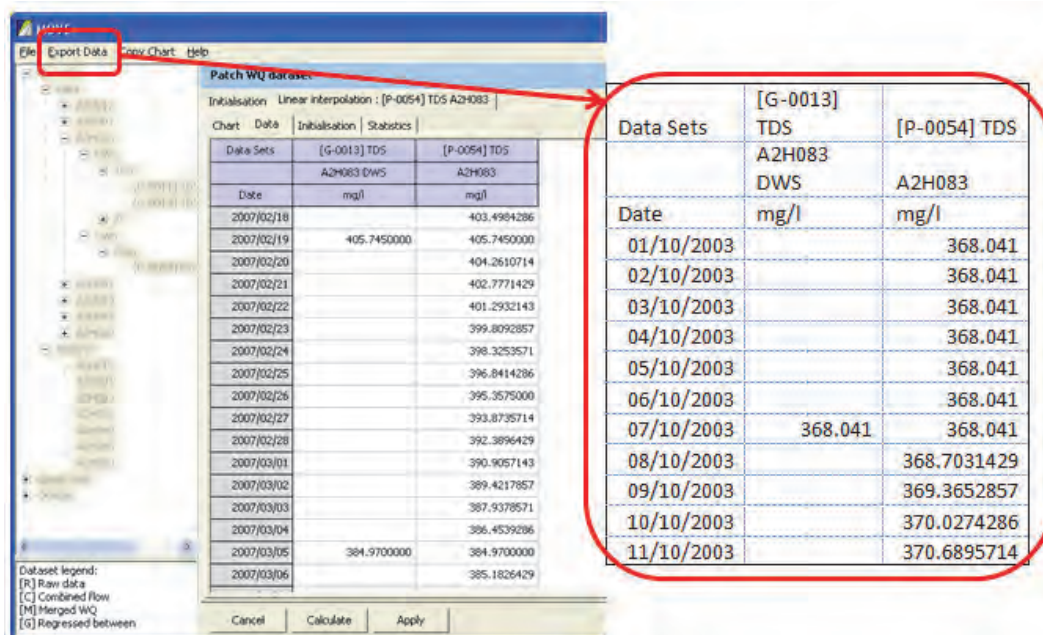


Figure 4.35: Tabulation of observed and interpolated water quality dataset

The purpose for which charts and data are exported is two-fold. In the first instance it is to provide material for reports. No less important is the need to preserve the background data and assumptions used to produce the reported material. This is necessary since large amounts of data are generated to carry out the patching and it is impractical to store every version that is tested. Instead the final results are stored along with the assumptions and raw data required to generate them. It is particularly important to store records of all changes made to remove outliers and the reasons for doing so.

#### 4.12.2 Initialisation tab

Selecting the **Initialisation** tab from any of the other tabs (e.g. from the **Chart** tab shown in Figure 4.34 or from the **Data** tab shown in Figure 4.35) provides a summary of the key assumptions used to generate the linear regression, as shown in Figure 4.36.

The initialisation data shown in Figure 4.36 can also be exported to a .CSV or HTML file, in a similar manner to exporting from the **Data** tab.

Linear Interpolation Initialisation		
WQ Dataset	[G-0013] TDS A2H083 DWS	
StartDate	1994/10/01	
EndDate	2013/09/30	
Extrapolation from ends	Set to end point concentration	

**Figure 4.36: Tabulation of linear interpolation assumptions**

#### 4.12.3 Statistics tab

Selecting the **Statistics** tab provides summary statistics that can be exported as a .CSV or HTML file shown in Figure 4.37. (Correlations and standard errors are not shown for the linear interpolation option since the patched interpolated line passes exactly through every observed data point yielding spurious results.)

Patch WQ dataset	
Initialisation	Linear interpolation : [P-0054] TDS A2H083
Chart	Data
Initialisation	Statistics
Linear Interpolation Statistics	
Number of observed points	407
Average concentration of observed points	354.66341
Number of patched and observed points from first to last observation points	5572
Average concentration of patched and observed points from first to last observation points	355.65338
Total number of patched and observed points	6940
Average concentration of all patched and observed points	354.27441

**Figure 4.37 Tabulation of linear interpolation statistical results**

#### 4.12.4 Cancelling interpolation

Selecting the **Cancel** button from the bottom of any of the tabs resets the results and initialisation data and restores control to the decision tree in the left hand side panel.

#### 4.12.5 Finalisation of linear interpolation

Selecting the **Apply** button from the bottom of any of the tabs consolidates and plots a chart of the linear interpolation.

Provision is made to include a daily flow file (see Figure 4.38). This was not required previously to generate the linear interpolation, since it is independent of the flow at the dam outlet. However, the daily flows are highly relevant for calculating the flow-weighted average monthly concentrations used at the observation route for model calibration. The flow data also provides extremely useful visualisation clues to identify flood events that may have contributed to long term fluctuations in reservoir concentrations.

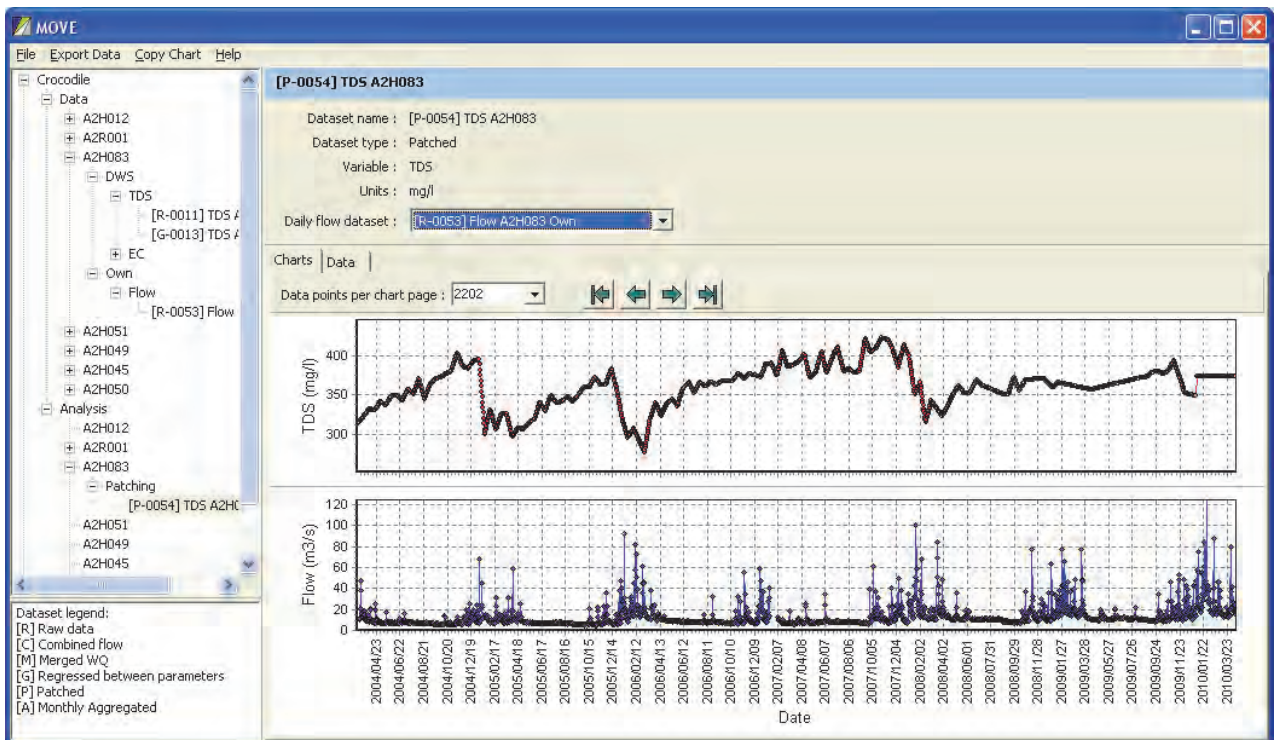


Figure 4.38: Plot of patched water quality dataset

#### 4.12.6 Tabulating linear interpolation results

Selecting the **Data** tab from the finalised linear interpolation patching plot shown in Figure 4.38 produces a tabulation of the daily time series of flow and patched water quality.

The resulting tabulation shown in Figure 4.39 can be exported as before.

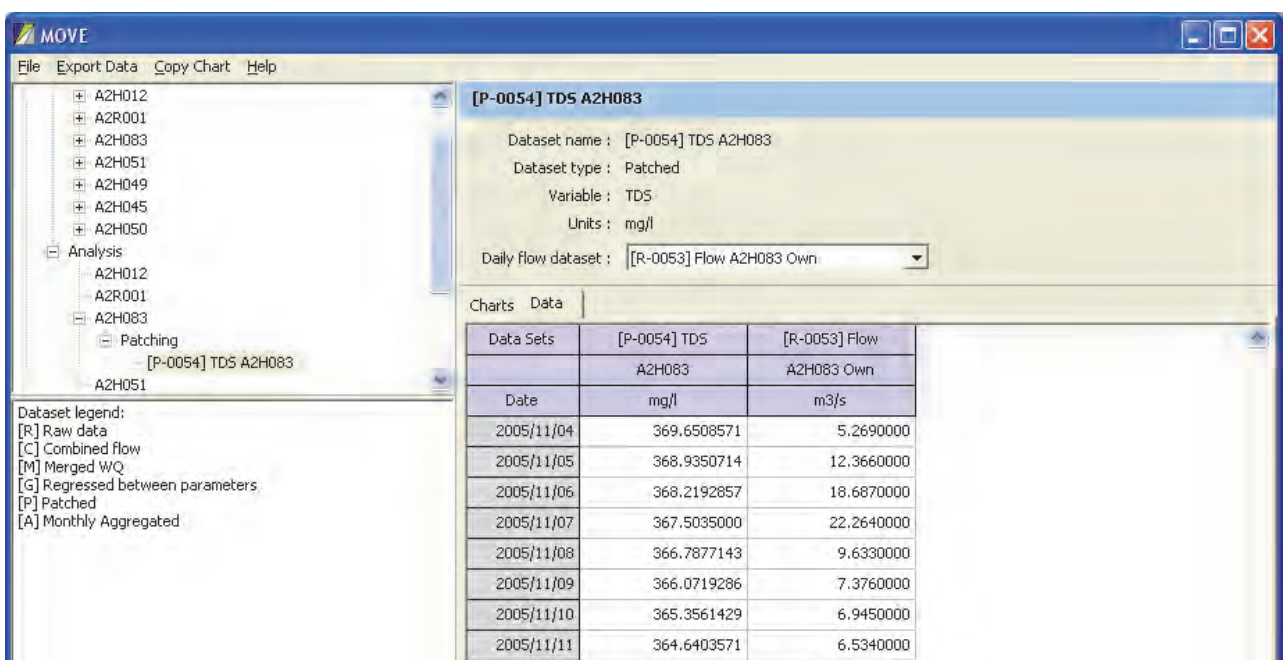


Figure 4.39: Tabulation of final interpolated patched water quality dataset



## 4.13 PATCHING BY MOVING REGRESSION

Patching by moving regression is most appropriate when there is a strong direct relationship between daily flow and concentration. This is eminently suitable for river sampling stations, where there is usually a strong relationship between daily flow and concentration.

The theory is provided in Section 3.4.

### 4.13.1 Initialisation for patching by moving regression

The patching system iterates to allow a number of moving regression results to be tested. Figure 4.40 shows the initialisation screen that appears after selection of the **Moving regression** radio button in the top section of the initialisation screen shown in Figure 4.33 and then selecting the **Calculate** button.

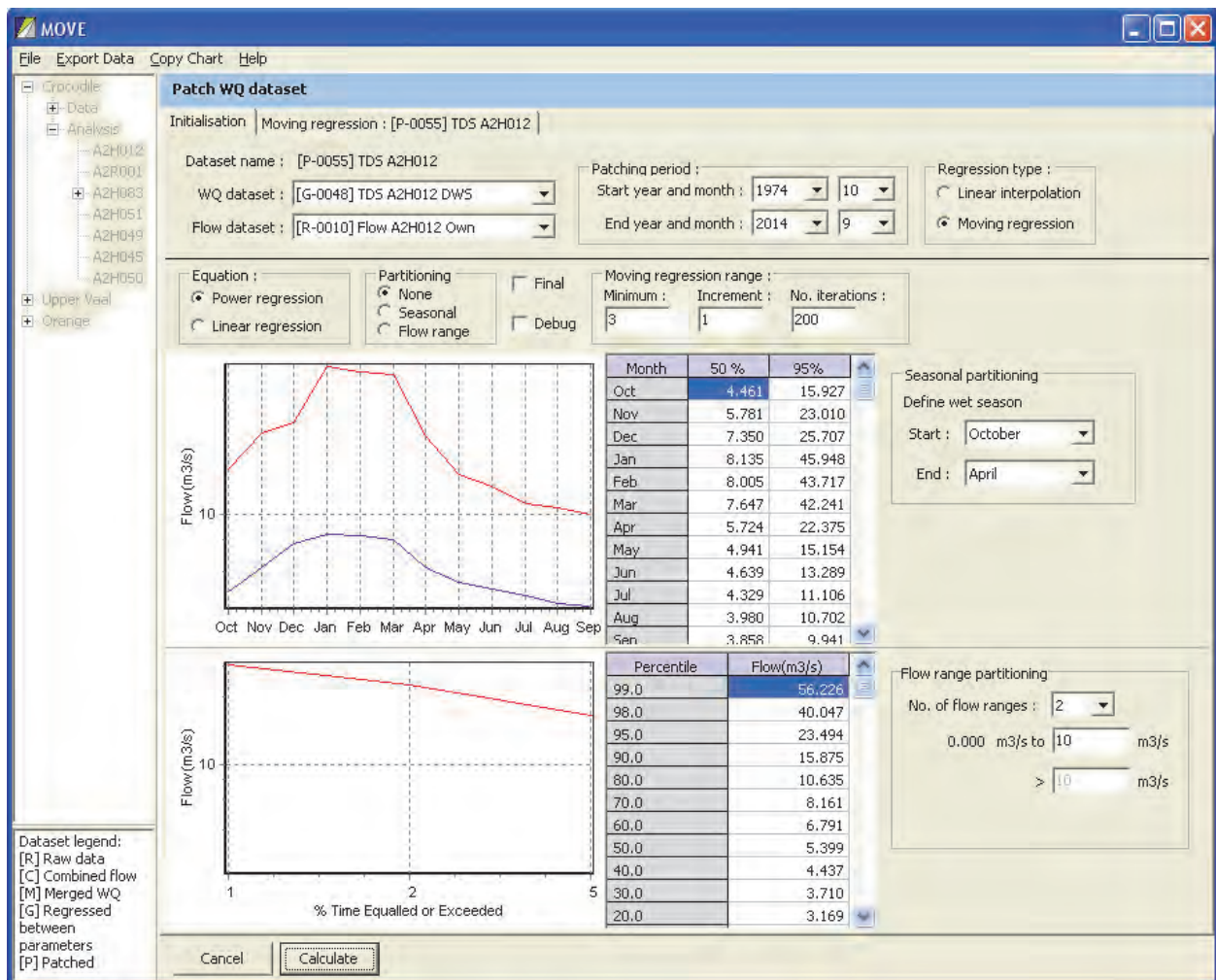
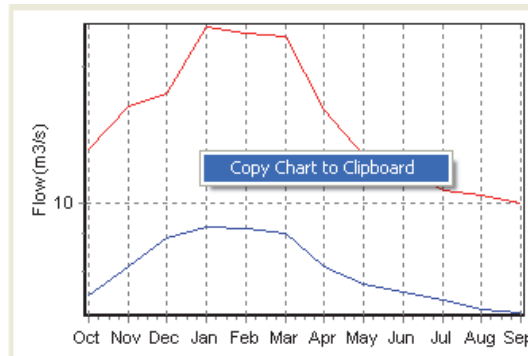


Figure 4.40: Initialisation for patching by moving regression

The bottom panel of the moving regression initialisation tab facilitates specification of the following:

- **Equation:** Choice of regression equation type (power or linear).
- **Partitioning:** Choice of partitioning. Available options include:
  - **None:** A moving regression based on the entire observed water quality dataset is used
  - **Seasonal:** Partition data into wet and dry seasons
  - **Flow range:** Up to three flow partition ranges.
- **Final:** This checkbox should only be selected once the desired moving regression period and the partitioning criteria have been identified. This is important since this selection generates a new screen and the data used in the initialisation screen is discarded.
- **Debug:** This checkbox opens a debug file that contains comprehensive observed daily flow and concentration data for every partition and for every moving regression period. This includes all regression equations, patched concentrations and associated statistics. The primary purpose is for model development and testing. It is recommended that this box is not checked for normal use since it generates large files and the required write operations significantly increase the computational time.
- **Moving regression range:** Define moving regression iteration parameters:
- **Minimum:** Starting moving regression sample size
- **Increment:** Amount by which each successive moving regression sample size is to be incremented
- **No. iterations:** Number of iterations.
- **Seasonal partitioning:** Drop down menus are provided to define the starting and ending months of the wet season. This is only required if seasonal partitioning is used. A pre-generated chart is provided showing the median and 95-percentile flows for each (monthly) season, along with a corresponding tabulation, to assist in making this choice. If flow range partitioning or no partitioning is later selected, these definitions are preserved for later reuse, if so required.
- **Flow range partitioning:** Up to three flow range partitions may be defined. This is only required if flow partitioning is selected. A pre-generated flow duration curve chart is provided, along with a corresponding tabulation, to assist the user in selecting appropriate partitions.

The seasonal and flow range partitioning plots can be exported individually by double clicking on the appropriate chart, which displays the pop-up window shown in Figure 4.41. Clicking on this sends a copy of the chart to the desktop, from where it can be pasted into a document.



**Figure 4.41: Example of exported seasonal flow plot**

Selecting the **Cancel** button should be used with care since it returns control to the decision tree in the LHS panel. This resets the initialisation screen.

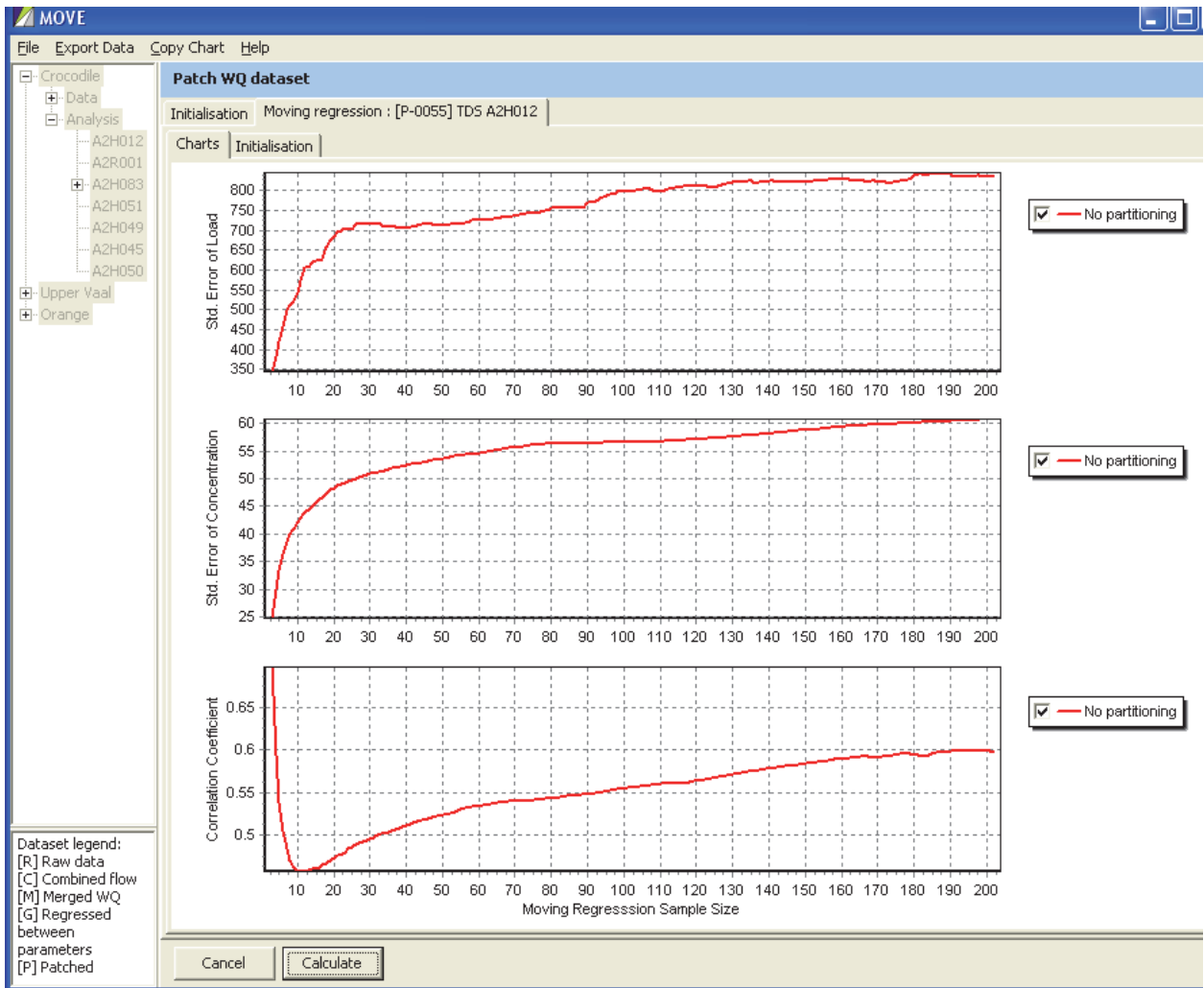
Selecting the **Calculate** button causes the model to iterate through the specified range of moving regressions and plot the moving regression statistical results.

#### 4.13.2 Moving regression plot

Clicking on the **Calculate** button generates the moving regression statistics plot shown in Figure 4.42.

The minimum moving regression value for this example was chosen as 3 to illustrate the initial sharp fall in correlation and the corresponding sharp rises in the standard errors. This is as expected since a moving regression with a period of 2 must give a perfect fit, since the regression line has two parameter values and therefore the regression line must pass exactly through both observed points. However, this fit is spurious since the sample size is far too small and does not incorporate enough variety of high and low flows to generate a meaningful regression equation.

It is evident from the example that the correlation coefficient reaches a minimum and then begins to climb again after a moving regression period of about 12 and continues to rise until reaching a new peak of 0.6 at a moving regression sample size of about 196 (Q,C) pairs. This is due to the increasing number of observed observation pairs incorporating more of the relatively rare high flow events.



**Figure 4.42: Moving regression statistical plot with no partitioning**

However, at the same time the concentration and load Root Mean Squared Error (RMSE) values continue to increase. This is influenced by the increasing variability of concentrations with increasing flow, leading to increased RMSE values. It is hypothesised that the magnitude of the error in concentration during floods increases because the change in concentration in the rising limb of the hydrograph differs from that in the recessional limb. Sometimes the observed concentration trough is also lagged after the flood peak, thereby weakening the inverse correlation between flow and concentration. Finally, after some big floods the catchment soils can wet up sufficiently to substantially increase the interflow and base flow component of the catchment runoff, resulting in higher than expected salinity during smaller floods occurring soon afterwards. Since the relatively infrequent grab samples (often taken at weekly intervals or even less frequently) can coincide with any part of the flood hydrograph, the RMSE value tends to increase as more flood events are included in the moving regression sample.



In the case of the calculated loads, the relative magnitude of the RMSE is magnified by the fact that the increasing concentration error during floods is weighted more heavily by the corresponding substantial rise in flow rate, since load is the product of flow and concentration. (This is evident when the RMSE values for concentration and load are expressed as percentages of the respective mean concentration and load. In the example for a 195 point moving regression the RMSE is 13% of the average concentration but the RMSE for loads is 37% of the average load.) Even a small lag in the concentration response during floods can lead to a large error, which is then squared in the RMSE calculation, thereby giving it a very much larger weighting than is the case for samples taken during dry weather, when after multiplication by the much smaller base flow translates into a much smaller squares of the error that go into the calculation of the RMSE.

The squaring of the errors in the calculation of the RMSE and in calculating the regression constants makes it all the more important to identify and eliminate sampled data outliers. Just one abnormally high value can have a profound effect on the goodness of fit and the associated RMSE.

The result of partitioning the data into wet and dry seasons is shown in Figure 4.43.

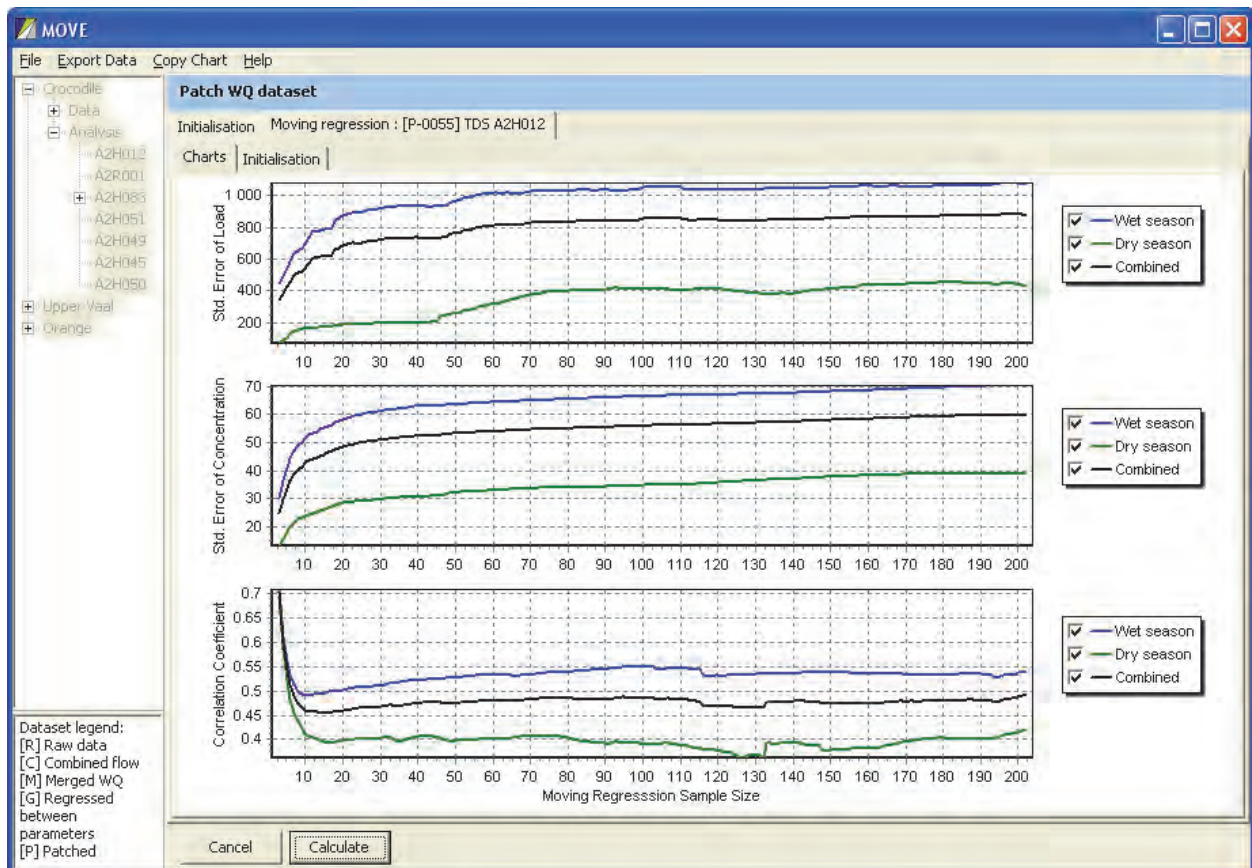


Figure 4.43: Moving regression statistical plot with seasonal partitioning

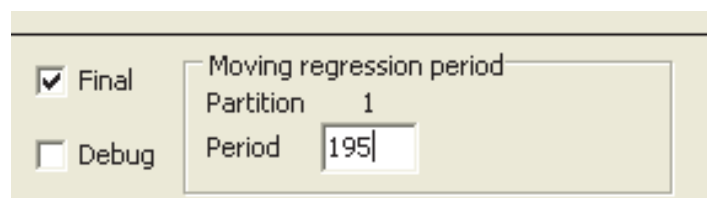
In this example seasonal partitioning leads to deterioration in the combined correlation coefficient to 0.49 while the RMSE values remain about the same, indicating that for this monitoring station not partitioning yields better results.

This attributable to the partitioning reducing the sample sizes and also reducing the variance between observed high and low concentration events, thereby weakening the correlation with flow.

A tabulation of the input files and parameter values specified can be viewed by selecting the **Initialisation** tab. This information can be stored in .CSV or HTML format.

### 4.13.3 Plotting patched daily time series

After choosing a moving regression sample size, the patched daily time series is generated by selecting the main initialisation tab in Figure 4.40 and selecting the **Final** checkbox (see Figure 4.44).



**Figure 4.44: Selection of moving regression period**

A window entitled **Period** is then opened to the right of this checkbox into which the chosen moving regression sample size is entered.

Selecting the **Calculate** button then displays a chart showing the daily flows, observed water quality variables and the daily concentrations patched by moving regression (see Figure 4.45).

Examination of Figure 4.45 reveals that there is still more that can be done to improve the fit between the observed and patched values. This is obvious from the fact that the regressed values seldom if ever pass exactly through any of the observed points. In fact, by their very nature regression lines tend to attenuate the amplitude of natural fluctuations. As a result they often under-estimate the peak concentrations observed in the dry season (as is apparent in the last two months of the plot) and at the other extreme over-estimate concentrations of the troughs that occur during the wet season.

It is important to observe that regression lines are blunt mathematical approximations of reality; they do not define reality itself. They are not based on physical cause-effect relationships and are therefore not able to explain all of the fluctuations observed in a real

catchment system. The objective here is not to produce an elegant statistical model, but to rather derive a good estimate of the fluctuations in water quality that took place during the gaps between sampling points.

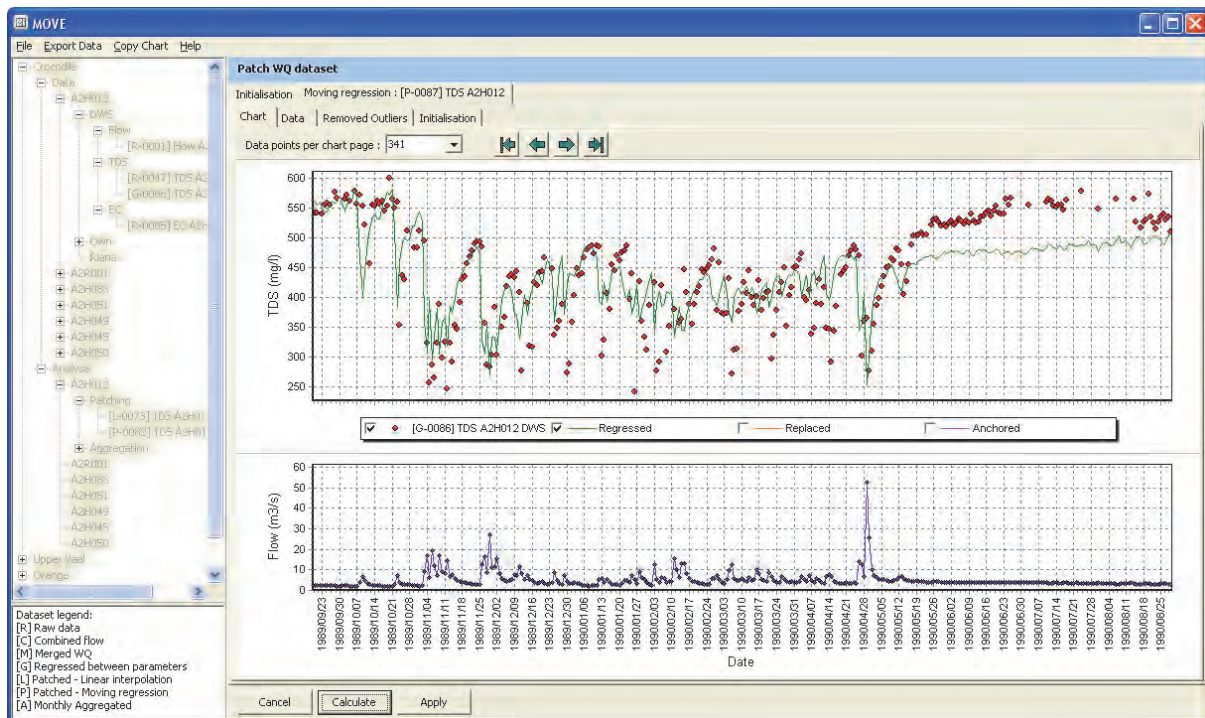


Figure 4.45: Daily moving regression plot

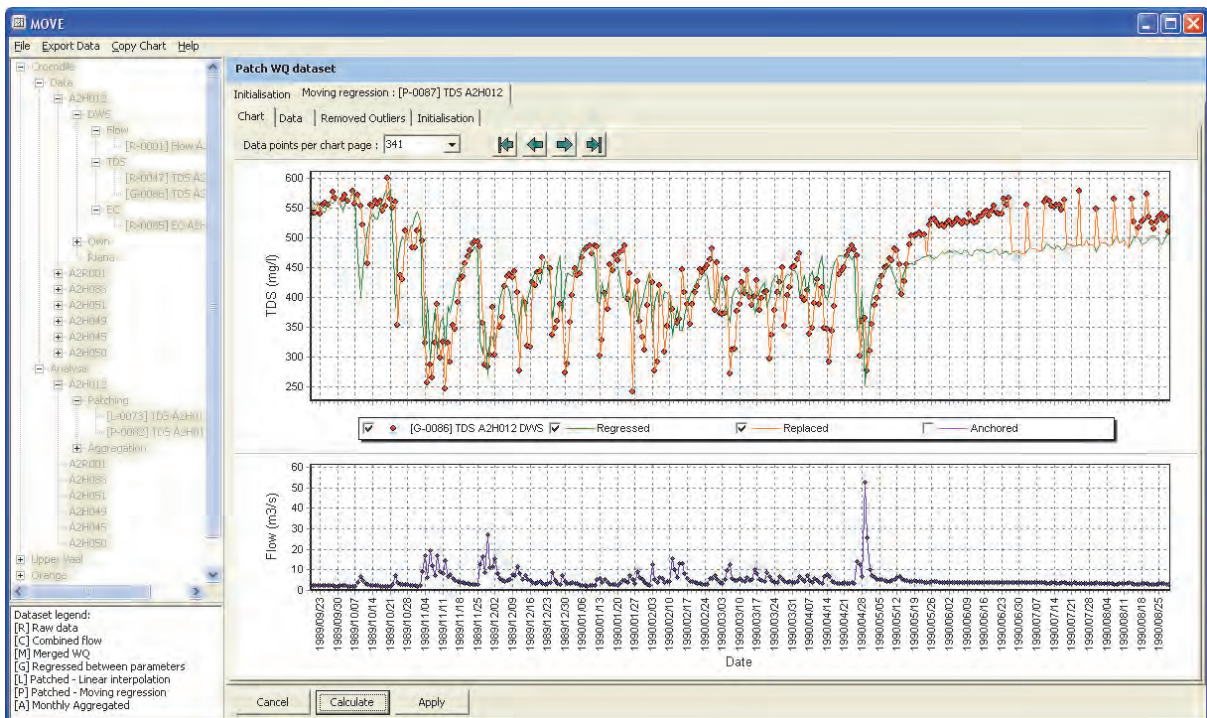
#### 4.13.4 Replacement of patched values with observed values

It is clear from Figure 4.45 that during the winter period the regressed concentrations are significantly lower than the observed values occurring during this low flow period. Moreover, during most of the high flow events the regressed concentration plots are significantly higher than the observed concentrations.

Figure 4.46 shows the results of replacing patched daily values with observed values.

This method brings about a significant improvement, since it makes use of all of the observed values, thereby improving the amplitude of the patched values. At this stage the standard errors discussed in Section 4.13.2 and shown in Figure 4.42 and Figure 4.43 have been reduced to zero and the correlation coefficients set to 1.0 since the patched lines now exactly pass through all of the observed points.





**Figure 4.46: Daily moving regression with replacement with observed values**

However, while superficially this may appear to be the case, we know that although the fit is better than the original regression line, it is certainly not perfect. In the first place, we know that the instantaneous grab samples themselves do not fully represent changes in flow and concentration that occurred during the course of the day on which they were sampled. We also know that some sample outliers may still remain undetected. More importantly, we also know that we have patched through 6-day long gaps between each of the observation points (for weekly sampling) and that the regression used to do so is likely to have introduced errors of similar magnitude to those between the original regression points and the observation points. As such, about six-sevenths of the original errors have not yet been adjusted. Finally, while the changes during the wet period look plausible, those during the last two months certainly do not. During this winter low flow period the uniform base flows do not give any plausible reason why the patched values should consistently fall 16% below the observed values. It is clearly inadequate to adjust every 7<sup>th</sup> point up to the observed values, but leave the other six patched values back down at the level of the regression line, which is clearly not the best vehicle to represent reality during this period.

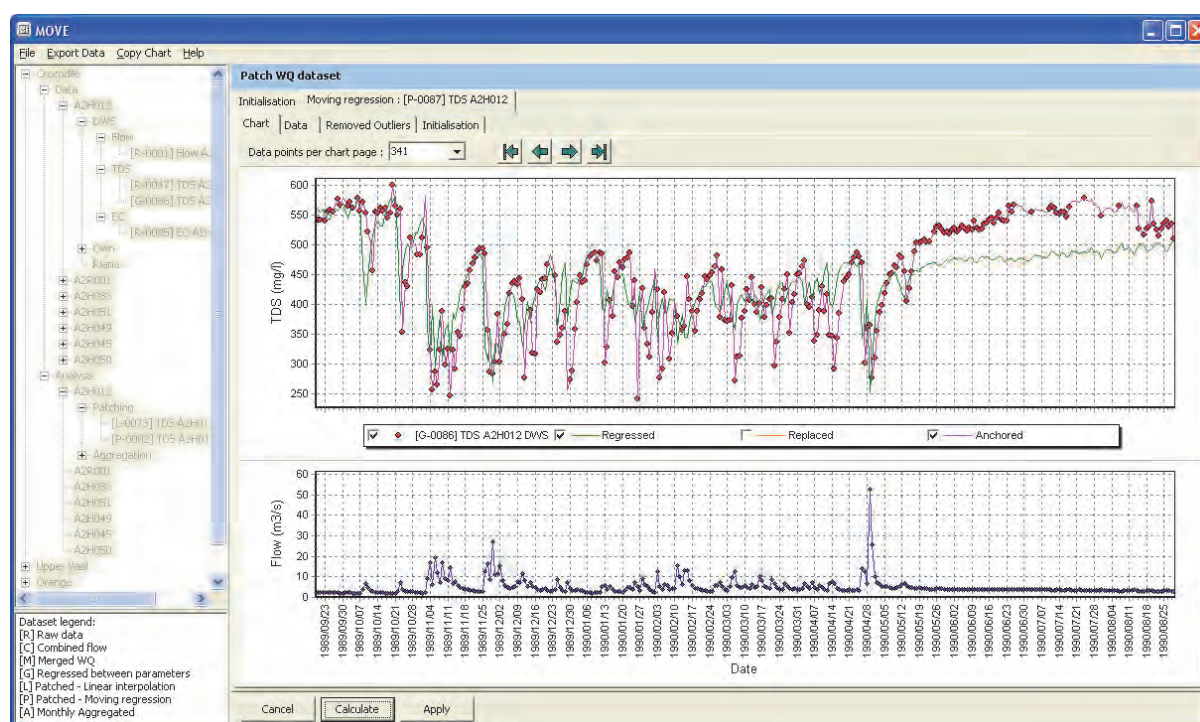
The above implies that there is still more that can be done to improve the representativeness of the patched values.

#### 4.13.5 Anchoring of patched values onto observed values

The next step was to not just replace one in seven regressed points with observed values, but to anchor them onto the observed points. The anchoring method that was employed preserves the shape of the moving regression between flow and concentration, but applies a smoothed adjustment between pairs of observed points to ensure that the patched line passes through the observed points.

The theory is provided in Section 3.5.

It is noticeable from Figure 4.47 that this method eliminates the identified problems presented by the inaccurate regression line during the low flow months.



**Figure 4.47: Daily moving regression with anchoring onto observed values**

It could be argued that these adjustments partially invalidate the statistical principles inherent in the regression equation. However, we must reiterate that the objective here is to make the best use of the observed data to patch the daily concentrations in the intervening gaps. To do this properly we must aim to comply with actual observed processes, rather than slavishly forcing the patching to conform to blunt statistical tools.

Examination of Figure 4.47 shows clearly that the regression fails hopelessly to mimic the observed concentrations during the last 3 months and that the anchoring provides a much more plausible result. The absurdity of slavishly sticking to the regression line is made even more clear by examination of the 3<sup>rd</sup> last month, when daily sampling took place. Clearly the

actual observed daily data provides a far more logical estimate of the daily concentrations during this period than the regression line that is obviously in error.

Close examination of the more complex concentration variations during the wet periods shows that the ensuing patching is more plausible, since the regression line predicts rises in concentration during some smaller flow events whereas the observations show the opposite. The main justification for giving preference to the regression line in these instances would be if there is good reason to suppose that the sampled water quality is wrong. However, many of the events are corroborated by multiple samples. Moreover, this argument is baseless, since the regression itself is much more seriously affected by outliers (occurring at any point in the regression period), whereas application of the anchoring method confines the error to only the bad sample itself and the gaps immediately adjacent to it.

#### 4.13.6 Final removal of outliers

This stage of the **Patch WQ dataset** dialog, before selection of the **Apply** button, is the final opportunity to remove outliers since this is the stage when the daily patching is completed.

Figure 4.48 illustrates the identification and removal of an outlier on 2008/10/08.

However, after all of the outliers have been removed from the chart, the patched daily concentrations remain the same as before (see Figure 4.49). This is because the new moving regression lines still need to be re-generated.

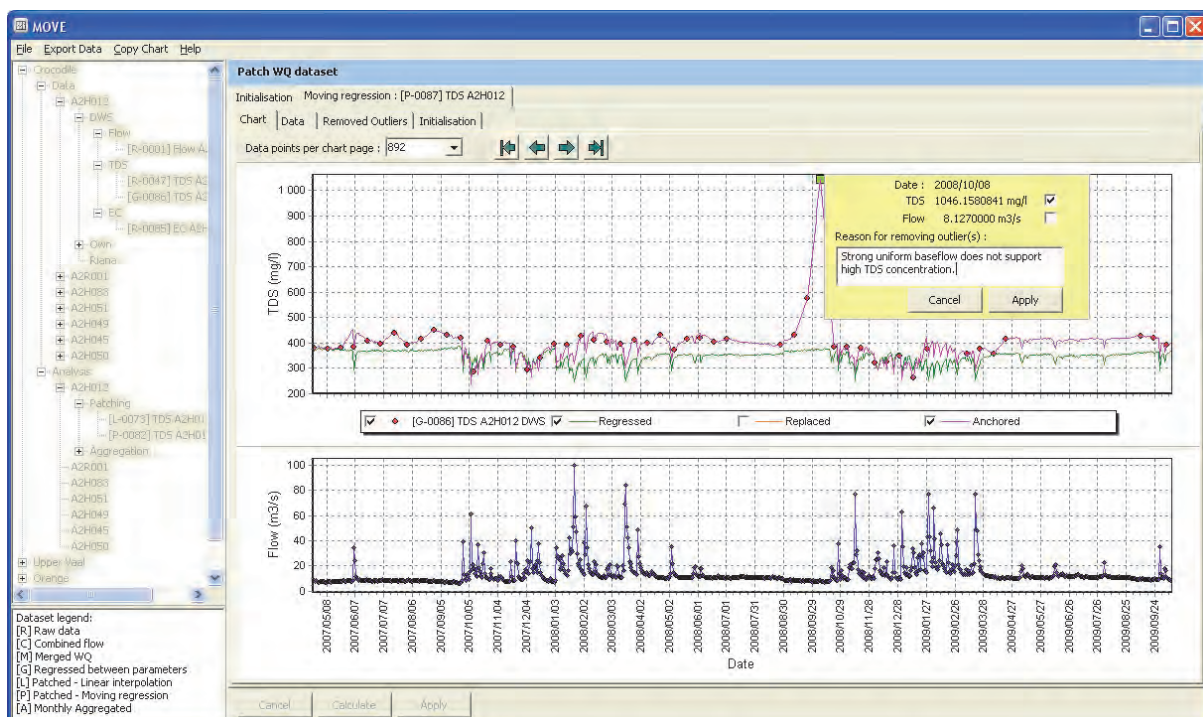
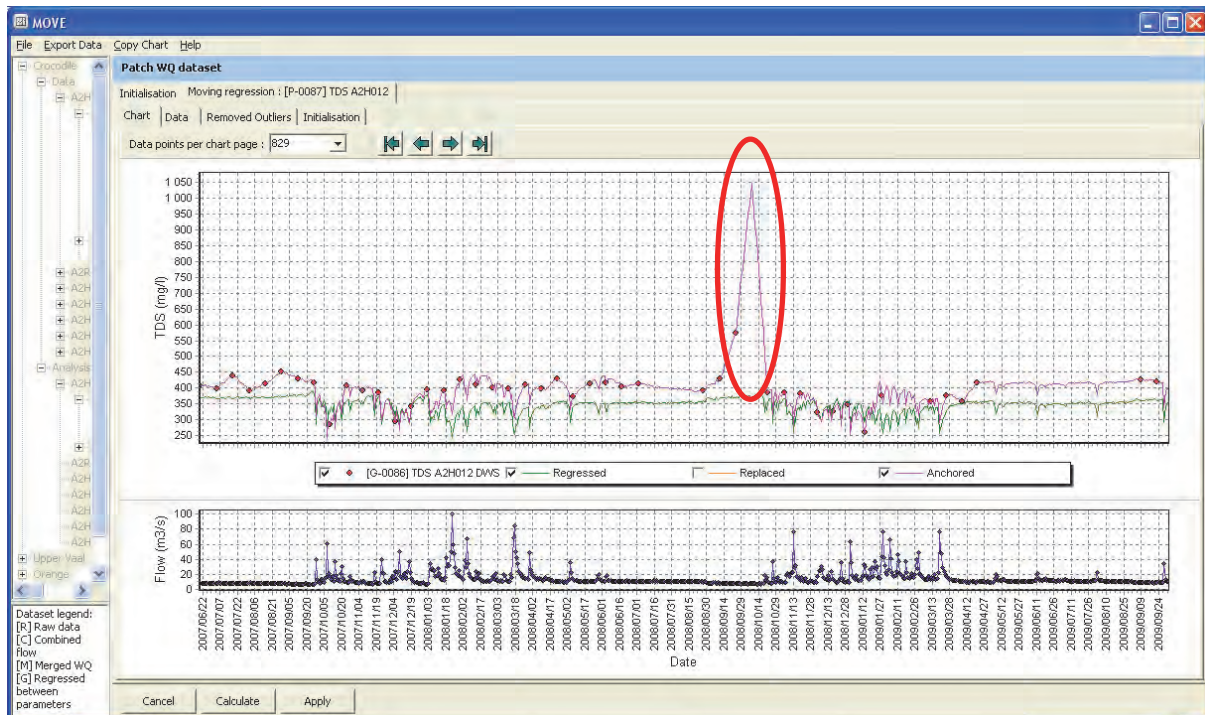


Figure 4.48: Interactive removal of outliers from Patch WQ dataset chart



Selection of the **Calculate** button causes the model to re-run the preceding moving regression step and also recalculate the two dependent daily patched datasets using replacement with observed values and anchoring onto observed values.

Selection of the **Apply** button then consolidates the daily patching data that is then used in the subsequent **Aggregation** analysis.



**Figure 4.49: Result after removal of outlier**

It is important to observe that the changes made at this late stage in the process cannot be made automatically to the original raw datasets. This is because the water quality dataset used in the moving regression (or linear interpolation) may have been derived from a [G-nnnn] type dataset that is the product of a regression between the primary water quality variable (TDS) and the secondary variable (EC). Human intervention is required here since when the TDS and EC datasets are compared it may be found that both TDS and EC values were available on that day. If that is the case, then if the TDS value is the outlier that would have been removed, but an underlying EC data value that may still be valid may still be available for that day. A decision then has to be made whether or not to retain the EC value and use the new regression between EC and TDS to calculate a new more acceptable TDS value, or whether the EC value also has to be rejected. It is also possible that the underlying water quality dataset used in the patching process might be the result of prior merging of two or more monitoring stations, again necessitating human intervention to make key decisions and unravel the cause effect chain.

After regenerating the moving regression datasets it is feasible to proceed with further processing in the aggregation stage. However, after removing outliers at so late a stage it is advisable that the user returns to the earlier TDS / EC regression stage and/or the dataset merging stage to determine which raw datasets merit adjustment to facilitate reporting back to the **Source** organisation and possible restoration of the now missing data from one of the other datasets.

The same would apply to the flow data, if two or more flow station records have been combined

The removed outliers list for the patched dataset provides a ready record of the dates (in this example 2008/10/07) that need to be examined to determine which raw datasets need to be amended.

In instances when datasets used in the moving regression (or linear interpolation) are still raw datasets [R-nnnn], then there is no need to make any changes prior to the moving regression stage since they would already have been amended when the outliers were removed from the moving regression **Source** file. Thus, in the example used in this section there is no need to alter the flow dataset since it is a [R-nnnn] type dataset and would have been changed right when outliers are removed from the moving regression **Source** flow dataset.

This may sound like a long process, but it actually goes quite quickly since all of the key processes are automated and the external datasets have already been imported.

The effort required to amend raw datasets at the end of the patching process emphasises the need to pay attention to each of the earlier data manipulation steps to trap as many (if not all) outliers as early in the process as possible.

#### **4.13.7 Tabulation of final patched daily concentrations**

A table showing the time series of observed daily flows and concentrations and the daily concentrations that were patched using the moving regression technique can be displayed by selecting the **Data** tab (see Figure 4.50). This tabulation can be exported as a .CSV or HTML file.



MOVE

File Export Data Copy Chart Help

**Patch WQ dataset**

Initialisation Moving regression : [P-0094] TDS A2H012

Chart Data Removed Outliers Initialisation

Data Sets	[G-0086] TDS	[R-0010] Flow	[P-0094] TDS	[P-0094] TDS	[P-0094] TDS
	A2H012 DWS	A2H012 Own	A2H012 (Regressed)	A2H012 (Replaced)	A2H012 (Anchored)
Date	mg/l	m3/s	mg/l	mg/l	mg/l
1980/02/15	-1	12.2230000	357.4634893	357.4634893	505.3657518
1980/02/16	-1	11.5070000	362.8616397	362.8616397	502.6380356
1980/02/17	-1	11.9690000	359.3322330	359.3322330	490.9827624
1980/02/18	459.0000000	15.7840000	335.4753370	459.0000000	459.0000000
1980/02/19	-1	16.2990000	332.8115004	332.8115004	427.5253413
1980/02/20	-1	23.8970000	302.6461567	302.6461567	368.5491756
1980/02/21	-1	13.2330000	350.4855911	350.4855911	387.5777879
1980/02/22	-1	10.7210000	369.2925442	369.2925442	377.5739189

**Figure 4.50: Tabulation of moving regression data**

#### 4.13.8 Display final moving regression initialisation data

Selection of the **Initialisation** tab gives a summary of the input files used and the assumptions made to generate the patched daily time series.

#### 4.13.9 Plot final patched water quality data

Selection of the **Apply** button from the **Chart** tab of the **Patch WQ dataset** dialog (see Figure 4.48) consolidates the moving regression and displays the plot shown in Figure 4.51.

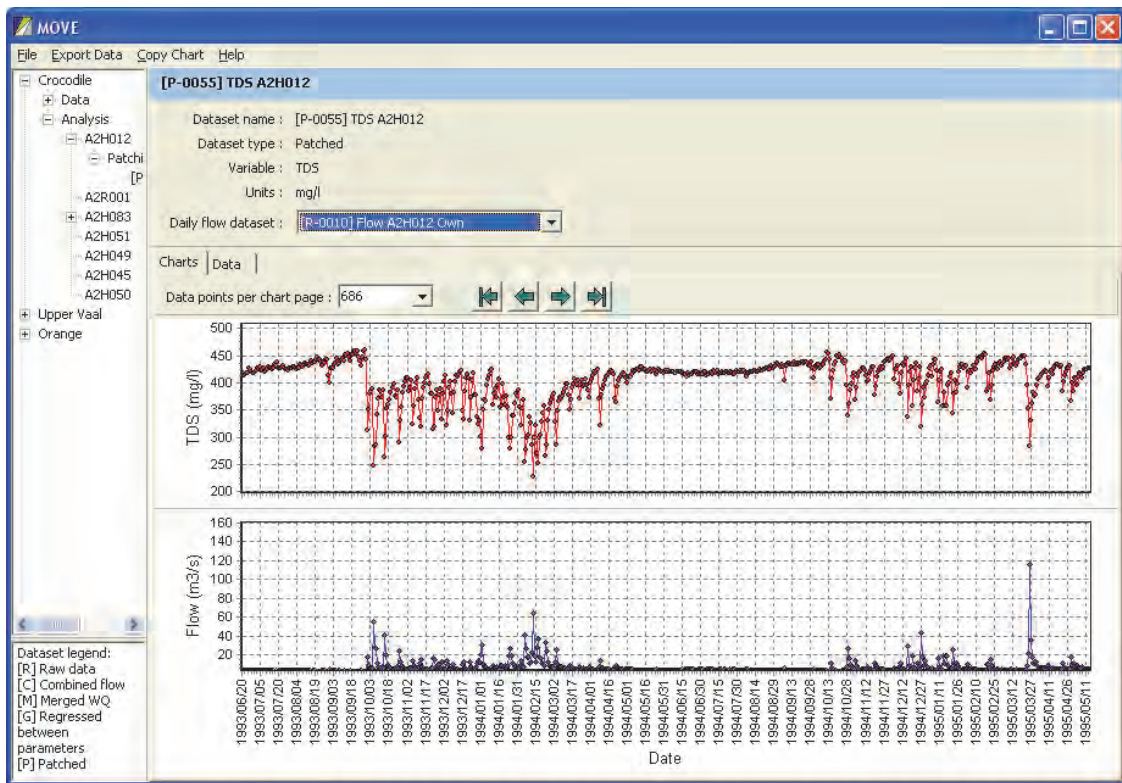


Figure 4.51 Plot of patched daily concentrations and observed flows

#### 4.14 MONTHLY AGGREGATION OF DATASETS

This is the final Analysis step to aggregate the patched daily data into flow-weighted or arithmetic monthly water quality data and average monthly flows for export as files in the standard WQT and WRPM models.

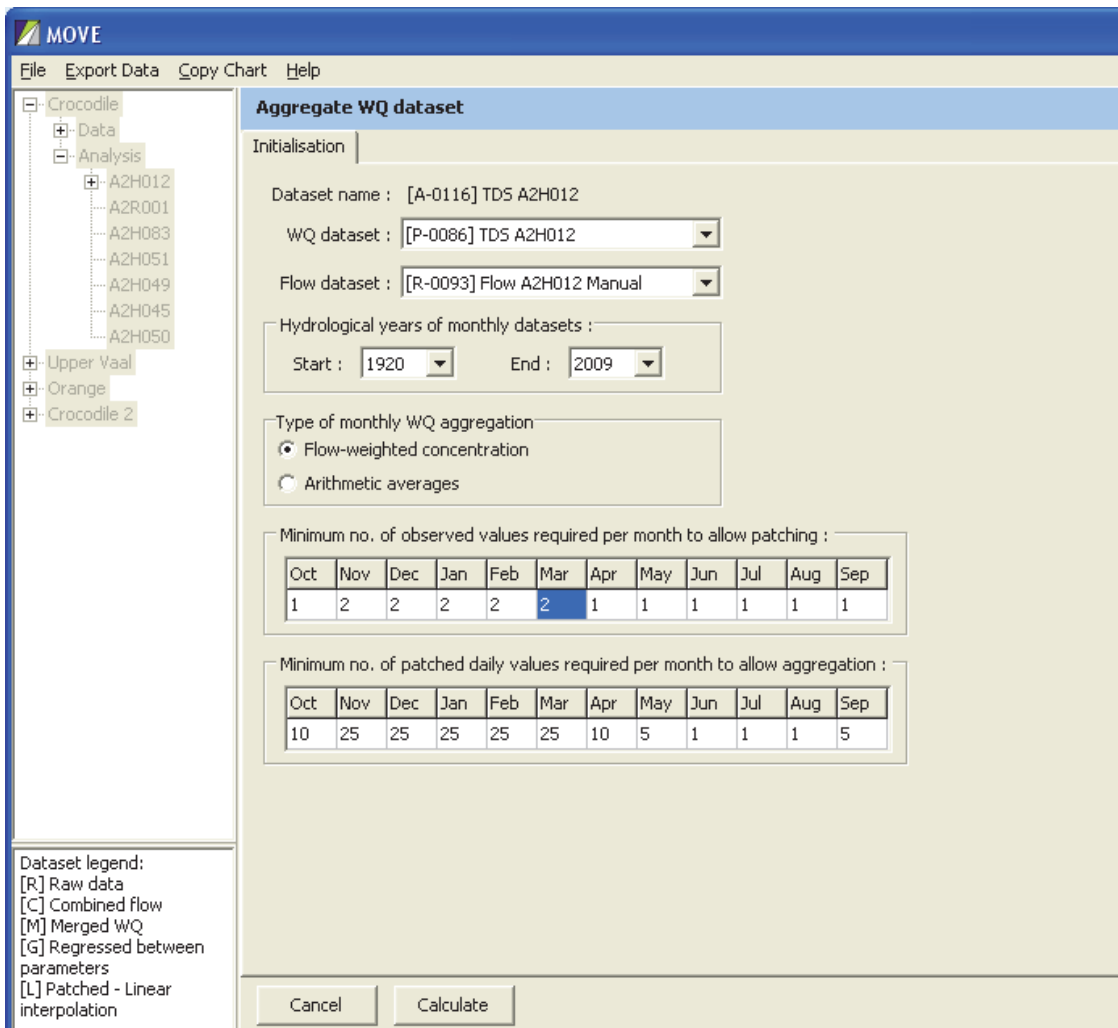
##### 4.14.1 Selection of monthly aggregation

**Monthly Aggregation** is the second and final **Analysis** option. Right clicking on the desired **Site** opens the pop-up menu (see Figure 4.32) from which the **Monthly Aggregation** option can be selected. This option draws on the patched datasets already prepared for the selected **Site** in the preceding **Patch WQ Dataset** step.

##### 4.14.2 Aggregate water quality dataset dialog

Clicking on the **Monthly Aggregation** option opens up the **Aggregate WQ dataset** Initialisation dialog in the right hand side panel shown in Figure 4.52.

The initialisation data that the user is required to provide is described in the following sections.



**Figure 4.52: Initialisation for monthly aggregation**

#### 4.14.2.1 Select patched water quality and flow data

A pre-prepared patched daily water quality dataset is selected from a drop down menu for the **Site**. An appropriate daily flow dataset is also selected.

#### 4.14.2.2 Set aggregation period

The user is required to specify the starting and ending years of the aggregation period. Hydrological years are specified since this corresponds to the standard WQT and WRPM model format for monthly data. It will be noticed that the date range specified in the example (1920) starts long before the date of the first water quality analysis and ends a few years after the last observation point. This has been done since the models require input data files that cover the full period to be simulated. The preceding and ending years can be padded with missing data entries manually after the aggregation is complete, but it is much more efficient to allow the program to do this automatically.

Long simulation periods are particularly important in modelling conservative water quality variables, such as TDS, since storage in catchment soils results in decades long lag times that have to be accounted for in model calibration.

#### 4.14.2.3 Specify aggregation type

Flow-weighted or arithmetic aggregation can be specified.

Flow-weighted aggregation is strongly recommended for water quality modelling since the load and flow passed downstream determines the mixing at downstream river junctions and the water quality occurring in downstream reservoirs.

However, when preparing datasets for a point of abstraction from a run of river scheme the arithmetic average would be more appropriate since this represents the average concentration of the water abstracted by users.

#### 4.14.2.4 Specify minimum number of observed water quality values

The minimum numbers of observed water quality data values required in a month before the monthly aggregation will be accepted are specified.

The program will fill in the missing data indicator value of -1 for every month where the number of observed analyses is less than the minimum that is specified.

In instances when minimums of zero are specified, the program will use the patched values even for months when there is no observed data. This is not recommended for observation points since too long a gap between observed points can lead to unreliable patched data. No data at all (with a gap in the data) is generally far preferable to misleading data.

However, a complete record is required for all point input data, such as effluent discharges and inter-basin water transfers, to enable the models to run. Although there are other methods for doing such patching, this model can be used for this purpose. Such extrapolation has little bearing on model calibration since calibration comparisons are only made against observed data. This is especially true when there is little storage in dams to carry over the effect of earlier patching errors.

The choice of the minimum number of samples required per month is strongly influenced by the historical sampling frequency. If too high a number is specified (for example 10, when the sampling frequency is weekly) it will eliminate the entire record, leaving nothing to calibrate against. Too small a target can result in poorer estimates of the flow-weighted monthly concentration. In the example a larger number of observation points is called for during the summer rainfall season, with fewer required during dry months. In general fewer

samples are required during dry months since there is little change in flow to drive a change in concentration. The consequences for load estimation are also less serious since the flow rate is low.

The relatively low minimum number of observed values per month used in the example is appropriate for weekly sampling. A small number of samples per month can be tolerated since the moving regression and anchoring patching process does provide a means of estimating the variation in daily concentrations during the month.

#### 4.14.2.5 Set minimum number of patched daily values

During the rain season a large number of patched daily values is required to obtain a reliable estimate of the flow-weighted concentration. This is because the concentration fluctuates markedly with flow rate and a number of floods can typically occur during a single month. Moreover, too few patched values increases the probability of entirely missing the flood peaks that convey most of the month's flow.

This is particularly important because gaps in the flow record often coincide exactly with high flow events when the modular limit of the gauging weir is exceeded. For this reason it is important to demand a high number of patched values per month during wet months. If this is not possible because the modular limit of the weir is too low, then the minimum number of patched daily values has to be relaxed. But then it is necessary to observe that the reliability of the aggregated monthly record is low and during the subsequent model calibration process greater emphasis needs to be placed on the records at more reliable flow gauging **Sites**. This is also a good reason, where possible, to establish water quality sampling points at reliable flow gauging **Sites**.

#### 4.14.3 **Aggregate water quality dataset dialog**

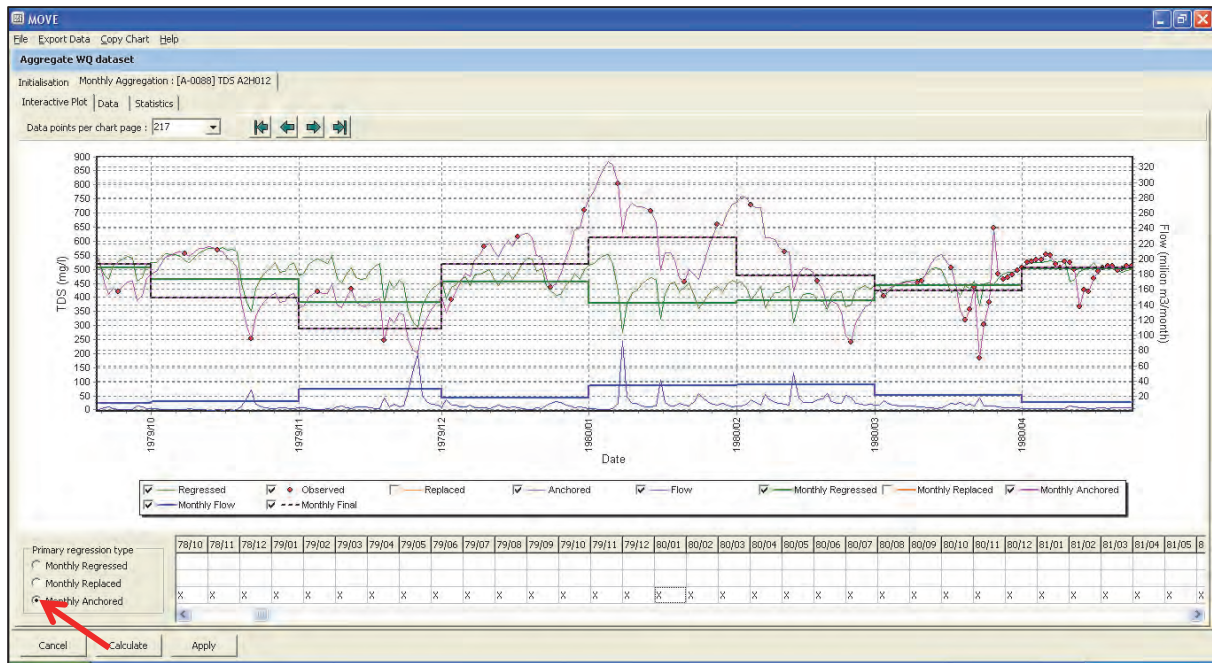
Selecting the **Calculate** button from the **Initialisation** tab opens the **Monthly Aggregation** dialog on the **Interactive Plot** tab.

A zoomed-in view is shown in Figure 4.53.

This dialog contains the following features:

##### 4.14.3.1 Interactive chart

When the patched water quality data is derived from a moving regression, i.e. where the **Source** water quality dataset is prefixed by [P-nnnn], this chart can include up to ten lines with the following legend definitions:



**Figure 4.53: Interactive Plot dialog**

- **Observed:** Discrete observed daily observed water quality values
- **Regressed:** Patched daily water quality values obtained by the Moving regression method
- **Replaced:** Patched daily water quality values replaced with Available observed values
- **Anchored:** Patched daily water quality values anchored on to Observed values
- **Flow:** Daily flow values
- **Monthly Regressed:** Flow-weighted (or arithmetic) monthly water quality Averages based on daily values patched using the moving regression method
- **Monthly Replaced:** Flow-weighted (or arithmetic) monthly water quality Averages based on the replaced patched daily values
- **Monthly Anchored:** Flow-weighted (or arithmetic) monthly water quality Averages based on the anchored patched daily values
- **Monthly Flow:** Average monthly flow calculated from the daily flow values
- **Monthly Values:** The selected monthly values.

Any of the lines can be removed from the chart by clearing check boxes from the legend.

In cases where patching was conducted using linear interpolation, i.e. where the **Source** water quality dataset is prefixed by [L-nnnn], this chart can include up to six lines with the following legend descriptions:

- Regressed: Patched daily water quality values obtained by the Linear interpolation method
- Observed: Discrete observed daily water quality values
- Flow: Daily flow values
- Monthly Regressed: Flow-weighted (or arithmetic) monthly water quality Averages based on daily values patched using the moving regression method
- Monthly Flow: Average monthly flow calculated from the daily flow values
- Monthly Values: The selected average monthly flow.

#### 4.14.3.2 Primary regression type selection panel

This panel (on the left hand bottom of the screen) allows selection of one of three primary regression types:

- Monthly Regressed: This selects the moving regression patching method for all months and superimposes a black dashed line over the **Monthly Patched** line
- Monthly Replaced: This selects the moving regression patching method with values replaced by the observed values for all months and superimposes a black dashed line over the **Monthly Replaced** line
- Monthly Anchored: This selects the moving regression patching method with values anchored on the observed values for all months and superimposes a black dashed line over the **Monthly Anchored** line.

Figure 4.53 shows the result of choosing the **Monthly Anchored** method as the primary regression type.

The interactive plot allows the user to change the method of selection for any chosen month. Changing the patching method for January 1980 to the **Monthly Patched** method is



achieved simply by clicking anywhere on the thick green horizontal line representing the average for January 1980 for this method (see the red arrow).

This action immediately removes the black dashed line from the **Monthly Anchored** method and places it over the green Monthly Patched line. At the same time the selection in the table at the bottom for the month 80/1 is moved to the **Monthly Patched** row. This is shown in Figure 4.54.



**Figure 4.54: Interactive plot dialog after changing patching selection for January 1980**

The new selection can also be made by navigating through the table shown at the bottom of Figure 4.54 until the desired date column and regression type row is reached and then select the appropriate cell. However, it is much quicker and more intuitive to make the selection directly from the chart since this is where the information required to make the selection judgment is observed and therefore the chart will already have been zoomed and panned to the appropriate place.

In this example this change in patching method was made because the two sparse observed samples straddle the highest peak flow during January 1980 and occur during low flow conditions and therefore exhibit high concentrations. This has the effect of lifting the concentration trough that occurs during the flood peak too high. Under these circumstances the patched line derived for the unadjusted moving regression gives a more plausible result.



## 4.14.4 Information displayed in Data tab

### 4.14.4.1 Source dataset derived from patching by moving regression

In instances where the patched daily values are derived from moving regression (i.e. the processed **Source** water quality dataset name is prefixed by a [P-nnnn]), selection of the **Data** tab displays a tabulation (see Figure 4.55).

Date	[A-0094] TDS	[A-0094] TDS	[A-0094] TDS	[A-0094] TDS	[A-0094] TDS	[A-0094] TDS	[A-0094] TDS	[A-0094] TDS	[A-0094] TDS	[A-0094] TDS
	A2H012 (Observed)	A2H012 (Regressed)	A2H012 (Replaced)	A2H012 (Anchored)	A2H012 (Flow)	A2H012 (Mnth. Val)	A2H012 (Mnth. Reg)	A2H012 (Mnth. Rep)	A2H012 (Mnth. Anc)	A2H012 (Mnth. Flow)
	mg/l	mg/l	mg/l	mg/l	million m3/month	mg/l	mg/l	mg/l	mg/l	million m3/month
1976/10/01	-1	487.9651344	487.9651344	458.1251717	4.8610000	338.5653793	338.5653793	331.0613218	305.9598750	38.7544608
1976/10/02	-1	366.9150762	366.9150762	331.1197990	15.6200000	-1	-1	-1	-1	-1
1976/10/03	-1	227.6266784	227.6266784	185.8760867	110.3010000	-1	-1	-1	-1	-1
1976/10/04	-1	291.5520663	291.5520663	243.8461601	40.0390000	-1	-1	-1	-1	-1
1976/10/05	205.0000000	258.6612207	205.0000000	205.0000000	65.3610000	-1	-1	-1	-1	-1
1976/10/06	-1	334.5461753	334.5461753	284.3832210	22.7980000	-1	-1	-1	-1	-1
1976/10/07	-1	375.5503746	375.5503746	328.8856867	14.2010000	-1	-1	-1	-1	-1
1976/10/08	-1	392.2172398	392.2172398	349.0508184	11.8880000	-1	-1	-1	-1	-1
1976/10/09	-1	402.1134018	402.1134018	362.4452468	10.7350000	-1	-1	-1	-1	-1
1976/10/10	-1	411.5954057	411.5954057	375.4255171	9.7580000	-1	-1	-1	-1	-1
1976/10/11	-1	420.4804389	420.4804389	387.8088167	8.9410000	-1	-1	-1	-1	-1
1976/10/12	-1	421.9756579	421.9756579	392.8023021	8.8120000	-1	-1	-1	-1	-1
1976/10/13	-1	428.4807742	428.4807742	402.8056848	8.2770000	-1	-1	-1	-1	-1
1976/10/14	-1	434.3848527	434.3848527	412.2080297	7.8260000	-1	-1	-1	-1	-1
1976/10/15	-1	442.4679435	442.4679435	423.7893869	7.2570000	-1	-1	-1	-1	-1
1976/10/16	-1	437.2803490	437.2803490	422.1000598	7.6160000	-1	-1	-1	-1	-1
1976/10/17	-1	435.4626418	435.4626418	423.7806180	7.7470000	-1	-1	-1	-1	-1
1976/10/18	-1	441.2846763	441.2846763	433.1009189	7.3370000	-1	-1	-1	-1	-1

**Figure 4.55: Aggregated data displayed in Data tab for moving regression option**

This tabulation contains the following columns of daily data:

- Date (Year/Month/Day)
- Observed daily water quality concentration values
- Daily water quality values patched by moving regression
- Daily water quality values patched by moving regression and replacement with observed values
- Daily water quality values patched by moving regression and anchored onto observed values
- Daily flow ( $m^3/s$ )
- Selected monthly flow-weighted (or arithmetic) average monthly water quality concentration
- Flow-weighted (or arithmetic) average monthly water quality concentrations calculated from observed daily flows and daily concentrations derived by moving regression
- Flow-weighted (or arithmetic) average monthly water quality concentrations calculated from observed daily flows and daily concentrations derived by moving regression with replacement by observed values

- Flow-weighted (or arithmetic) average monthly water quality concentrations calculated from observed daily flows and daily concentrations derived by moving regression with anchoring onto observed values
- Average monthly flow ( $10^6 \text{ m}^3$ )

#### 4.14.4.2 Source dataset derived from patching by linear interpolation

In instances where the patched daily values are derived from linear interpolation, i.e. the processed **Source** water quality dataset name is prefixed by a [L-nnnn]. The dataset contains the following columns:

- Date (Year/Month/Day)
- Observed daily water quality concentration values
- Daily water quality values patched by linear regression
- Daily flow ( $\text{m}^3/\text{s}$ )
- Selected monthly flow-weighted (or arithmetic) average monthly water quality concentration
- Flow-weighted (or arithmetic) average monthly water quality concentration calculated by linear regression
- Average monthly flow ( $10^6 \text{ m}^3$ )

#### 4.14.5 Information displayed in Statistics tab

##### 4.14.5.1 Source dataset derived from patching by moving regression

Selection of the **Statistics** tab displays columns of monthly water quality (CMON values) and monthly flow (QMON values -  $10^6 \text{ m}^3$ ) arranged in hydrological years (i.e. October to September) with rows for each year.

This monthly data is given for the following options:

- Water quality values patched by moving regression
- Water quality values patched by moving regression followed by replacement with observed values
- Water quality values patched by moving regression followed by anchoring onto observed values.

An example of such a file is shown in Figure 4.56.

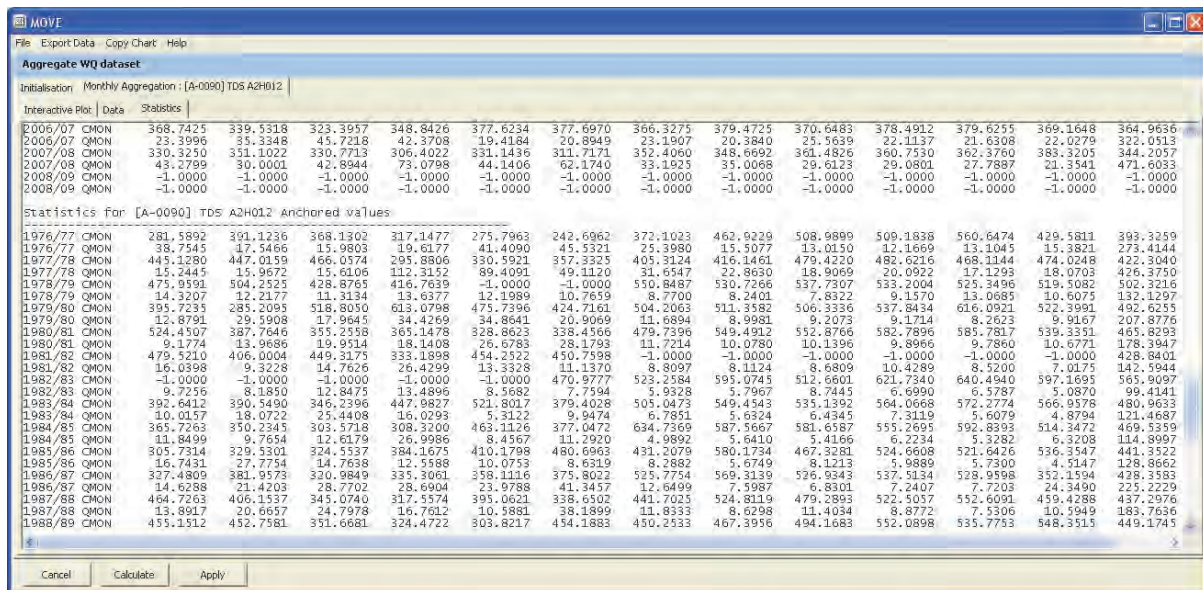


Figure 4.56: Monthly data shown in Statistics tab

#### 4.14.5.2 Source dataset derived from patching by linear interpolation

The form of the monthly flow and water quality dataset is the same as before, except that in this instance there is only one time series, derived from the linearly interpolated daily water quality values.

#### 4.14.6 Final aggregated monthly chart

Selection of the **Apply** button consolidates the final aggregated average monthly concentration chart (see Figure 4.57). Monthly flows are shown in the bottom chart and patched monthly average concentrations are shown in the top chart. The dates are hydrological years (i.e. 1985/1 represents October 1985 and 1985/12 represents September 1986).

In the case where patching was carried out by moving regression up to four types of flow-weighted (or arithmetic) average monthly concentration can be displayed:

- Patching by unadjusted moving regression (green line labelled Monthly Regressed)
- Patching by moving regression followed by replacement with observed values (Orange line labelled Monthly Replaced)
- Patching by moving regression followed by anchoring onto observed values (Pink line labelled Monthly Anchored)

- Selected patched monthly values, which can comprise a combination of the above three forms of patching (dashed black line labelled Monthly Final). In the example (Figure 4.57) the dashed line is superimposed on the pink Monthly Anchored line, but for 1990/6 (March 1990) the dashed line diverts to the green Monthly regressed line to reflect the previously made choices.

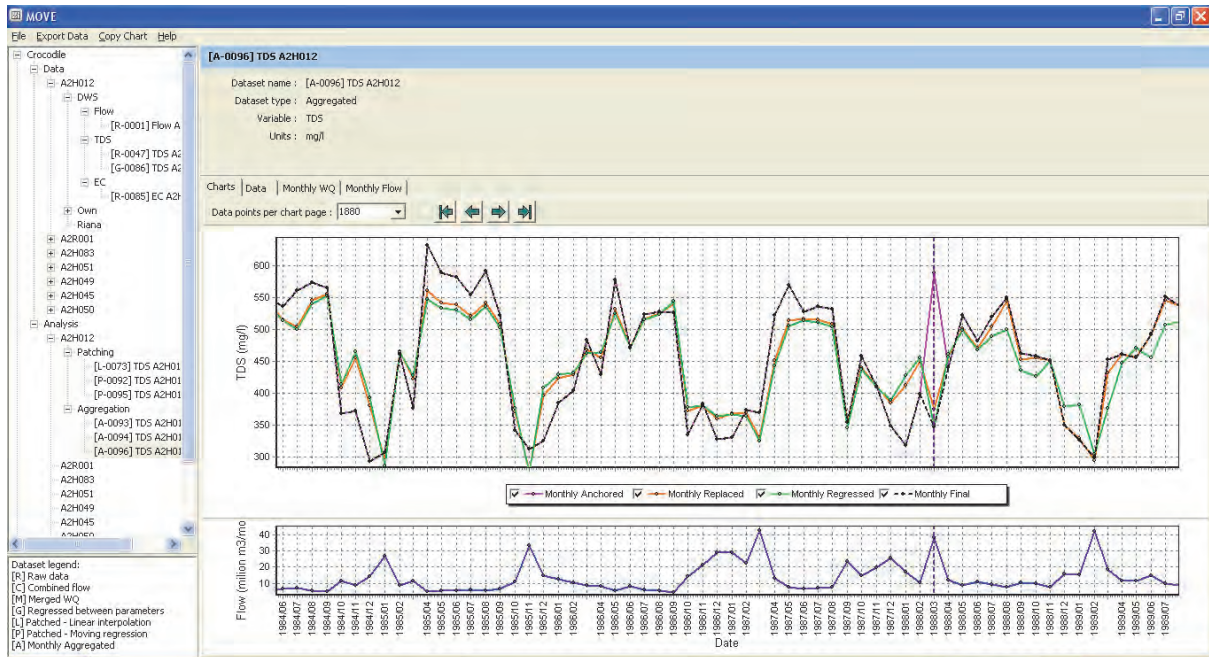


Figure 4.57: Final average monthly water quality file

#### 4.14.7 Export of aggregated monthly water quality file

A file containing aggregated monthly water quality data in the standard text format used as input to the WQT and WRPM models can now be generated by selecting **Export Data**, which now contains an extra drop down option for the export of such a file (see Figure 4.58).

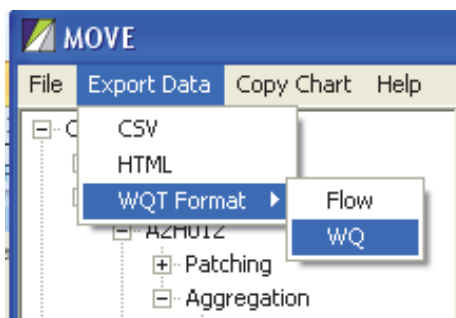


Figure 4.58: Selection of option to export monthly water quality data file

Figure 4.59 shows the resulting exported text file in the standard monthly format used as input to the WQT and WRPM models.

[A-0118] TDS AZH012 Monthly Values.TXT - Notepad

File Edit Format View Help

1971	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
1972	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
1973	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
1974	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
1975	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	331.50	419.24	468.19	505.90	511.45	185.77
1976	305.94	398.81	368.03	317.20	275.92	242.00	371.96	462.93	509.00	509.18	560.66	429.21	395.90
1977	445.05	446.83	465.88	295.43	330.32	357.39	407.56	431.99	504.56	485.61	487.77	473.22	427.63
1978	480.82	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	559.65	548.67	560.82	513.83	-1.00
1979	423.57	-1.00	523.99	604.18	474.90	426.03	504.21	511.25	505.99	537.71	616.01	522.25	470.76
1980	524.45	387.76	355.26	365.15	328.90	338.46	479.74	549.49	552.88	582.79	585.78	538.21	465.74
1981	479.54	406.11	449.32	333.26	454.26	450.54	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	213.92
1982	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	471.38	523.70	595.21	512.55	621.82	640.54	597.12
1983	390.49	390.49	346.53	447.41	521.80	379.82	505.45	549.45	535.11	563.71	572.36	566.83	480.79
1984	365.88	350.51	303.56	309.94	462.83	378.91	634.78	587.34	581.82	555.21	592.67	514.24	469.81
1985	310.10	332.80	324.71	384.08	409.71	481.22	431.25	579.94	469.86	524.41	522.17	534.27	442.04
1986	337.37	387.33	324.05	334.41	360.05	379.12	525.60	569.10	526.67	537.23	530.76	353.91	430.47
1987	466.18	407.74	347.98	318.01	401.49	588.70	440.88	524.87	480.08	522.26	552.63	462.14	459.41
1988	455.11	452.79	351.68	324.49	303.81	454.28	450.24	467.37	494.20	552.09	535.78	548.37	449.18
1989	526.02	344.85	370.40	411.00	396.53	385.72	373.90	469.02	542.18	562.78	543.79	515.00	453.43
1990	482.83	464.72	396.88	330.81	313.16	330.30	479.27	490.17	497.41	499.89	509.50	510.40	442.11
1991	434.73	430.71	367.10	360.06	398.20	422.61	402.31	452.90	484.40	506.52	492.67	500.15	437.70
1992	428.75	363.56	298.88	334.36	362.95	373.29	446.70	446.08	444.67	442.41	429.15	418.70	399.12
1993	330.86	319.44	320.83	354.76	327.17	446.25	399.60	465.64	477.93	476.95	503.41	486.22	409.09
1994	399.84	336.89	288.64	298.95	372.31	288.16	418.28	423.66	454.16	445.82	455.88	457.05	386.64
1995	379.16	283.95	283.35	292.22	431.69	469.32	503.82	527.55	474.97	504.34	490.37	504.26	428.75
1996	439.79	411.96	356.92	360.28	417.05	347.50	413.11	375.47	427.04	468.11	478.09	450.55	412.16
1997	446.99	392.28	354.92	360.95	389.85	423.81	419.69	460.40	470.40	471.40	472.24	465.25	427.35
1998	401.81	334.67	341.15	468.73	376.05	372.19	379.33	398.94	433.67	445.76	445.44	449.06	403.90
1999	436.14	384.37	339.26	304.68	292.28	504.14	409.11	409.19	432.74	437.20	453.96	432.98	403.01
2000	345.65	340.43	363.29	384.38	316.02	336.76	416.66	396.82	425.64	418.02	426.84	389.62	380.01
2001	344.32	368.86	357.73	322.43	340.01	364.23	394.23	379.50	382.76	426.36	424.02	435.82	378.35
2002	410.08	406.72	295.64	310.19	333.65	345.41	424.10	431.69	417.29	444.32	447.03	449.95	393.01
2003	409.88	-1.00	-1.00	-1.00	-1.00	-1.00	405.12	430.15	420.66	426.32	431.37	429.19	245.64
2004	419.44	376.75	312.54	295.32	348.17	363.82	364.08	399.13	404.77	400.37	425.46	444.29	379.51
2005	412.57	340.49	357.89	319.84	319.10	405.72	399.74	426.28	436.99	445.27	424.66	454.00	395.21
2006	417.87	366.69	-1.00	-1.00	-1.00	404.31	369.80	384.28	407.60	416.41	420.78	409.05	299.48
2007	332.91	366.23	325.33	372.75	411.52	376.71	407.03	374.47	-1.00	-1.00	-1.00	-1.00	246.91
2008	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
2009	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00

Ln 1, Col 1

Figure 4.59: Monthly water quality export file in standard WQT / WRPM format



## 5 DATA REQUIREMENTS AND PRE-PROCESSING

### OVERVIEW

This chapter provides an overview of the pre-prepared data required to run the MOVE water quality patching model.

The minimum requirement is very modest, comprising only two readily available files. One must contain time series water quality data at the river or dam sampling point at which the patching is to take place. The other is a file of observed daily flows at or near to the chosen water quality monitoring point.

The most convenient sources of this data are the relevant DWS databases.

### STANDARDISED WATER QUALITY DATA

The Water Management System (WMS) database is the most common source of water quality data. The data is obtainable from the DWS's Resource Quality Services based at the Roodeplaats Dam research centre, which is where the central laboratory is located. Data collected in various DWS monitoring programmes throughout South Africa is analysed in this laboratory and the results entered into the WMS water quality database. The WMS also stores water quality data provided by other organisations.

The WMS database stores long records of water quality analyses for thousands of monitoring stations located throughout South Africa for a large number of different water quality variables. Data is readily made available to users in the form of comma delimited (.CSV) files exported from the database.

The first row of such a file gives header records describing the data contained in each column.

The subsequent rows are grouped together according to uniquely labelled and sequentially ordered monitoring stations. Hence all the rows of data for a station are located next to one another and these in turn are ordered in ascending date sequence.

Data can be requested for one or more individually stated monitoring stations, or a request can be made for all the stations contained within a quaternary, tertiary or secondary catchment.

The required number of water quality variables requested can also vary from one upwards. Hence one file can contain a wealth of data for an entire **Project** area in a predictable stable format.

The MOVE patching system takes advantage of this, relieving the user of the need to know the intricacies of the input file structure. All the information required to locate monitoring stations and water quality variables is extracted from the file and displayed by the patching system for easy access. This can be done for as many monitoring stations and variables as required.

WMS water quality data can be obtained electronically from the DWS's RQS centre at Roodeplaat Dam. The current contact person is Marica Erasmus. Requests should be emailed and contact details are as follows:

Email: [maricae@dwa.gov.za](mailto:maricae@dwa.gov.za)  
Telephone: 012 808 9610  
Cell phone 082 908 2895

Specify the following in the request:

- Monitoring stations required (Individual station codes, or all stations within a quaternary, tertiary or secondary catchment. The latter will require more than one file to keep the emailed data down to manageable size.)
- List of water quality variables (e.g. DMS, EC, Na, Cl, SO<sub>4</sub>, PO<sub>4</sub>, etc.)
- Date period (usually the full record)
- Comma delimited format (i.e. .CSV)

## **STANDARDISED DAILY FLOW DATA**

Daily flow data is also available from DWS in standard format. In view of the size of such files, the data for each gauging station is supplied in an individual file in column format, one row per day in chronological order. Again there is no need for the user to be familiar with the file structure since this data is extracted automatically by the program, check codes are assessed to eliminate unreliable data and the information relevant to the user is displayed.

Daily flow data can be obtained by contacting Francinah Sibonyani

Email: [csp@dwaf.gov.za](mailto:csp@dwaf.gov.za)  
Telephone: 012 336 7912

Specify the following in the request:

- Daily data in electronic processing format is required (not in print format)
- Hydro codes for the flow stations required

- Period required

These two files are generally all that is required to run the model.

## **OTHER DATA SOURCES**

Water quality and/or flow files can also be pre-processed for use in MOVE. The easiest file type to use is a comma delimited (.CSV) file since it is easy to export from a spread sheet, which is the most convenient vehicle for pre-processing the data in a spread sheet. Space delimited and tab delimited files can also be used.

Other data sources include Water Boards, municipalities, municipal waste water treatment service providers (such as ERWAT, who also take sanitary samples upstream and downstream of effluent discharge points), mines and industries. Some DWS regional offices also keep their own pollution control data, which is often stored in *ad hoc* spread sheets or in paper files pending entry into the WMS system.

## **MANUALLY PREPARED WATER QUALITY DATA FILE**

The raw data used in manually prepared water quality input files can originate from any source and any number of formats. The only requirement is that the pre-processed product used as input to the MOVE patching system must conform to the following format:

Row 1: Headings for each column containing the desired data (optional, but advisable).

Row 2 to end: Any number of columns in any order, two of which must contain the following data:

### **Sampling date**

This is the day on which the corresponding sample was taken.

The sampling date can be in one of the following recognised formats (the same format for every row):

- yyyy/mm/dd
- mm/dd/yyyy
- dd/mm/yyyy
- yyyymmdd
- mmddyyyy
- ddmmyyyy



## Water quality variable(s)

The heading for the column containing the desired water quality variable can be defined by any code in the first row (or even none at all), provided that the user can recognise what they are. The units must be known so that the user can choose the correct one from the drop down list in the **Import WQ dataset** dialog.

The following standard water quality units are available:

- mg/l
- µg/l
- mS/m
- µS/cm
- counts /100 ml
- pH
- NTU
- °C

The required date and water quality columns can appear in any order and can be interspersed with columns containing intermediate calculations, or even unrelated data.

The **import WQ dataset** dialog permits the user to select a standard water quality code to be used for all data sources and stations in the **Project** area. It is necessary to standardise on this short code since each source organisation uses different nomenclature. For example, the WMS database denotes electrical conductivity as “**EC-Phys-Water result**”, other organisations might call it “**Conductivity**”, include units in the heading, etc. The computer model cannot recognise that all these names refer to the same variable. Hence it is necessary to define a standard variable name for the input data (for example “**EC**”, for which the standard SI units are **mS/m**). Using a standard code that is both short and recognisable is also desirable to obviate the generation of cumbersome and difficult to read chart captions and overly long decision tree names in the left hand panel of the dialog screens

The user can add new standard variable codes to the drop-down list for the **Project** as required. It is suggested that these be kept short so as not to clutter the dialog screens.

The model automatically carries out conversions for the most common units to bring them all to the same standard. For example, if the standard units for a variable are in mg/l, but a **Source** organisation uses µg/l, then the specified values will all be divided by 1000 to bring them to the same project standard. However, this is not done for all possible units that might be introduced from time to time. It is therefore advisable to pre-process all variable values to

conform to a recognisable standard of units, preferably the same as the standard used in the DWS's WMS database, to obviate possible confusion.

The standard water quality units applicable to different variables are listed below:

**Table 1: Water quality variable adjustment factors**

Standard Units	User-specified Units <sup>1</sup>	Standard Code	Affected variables	Adjustment factor
mg/l	µg/l	NH4	Ammonia	0.001
		Ca	Calcium	
		Cl	Chloride	
		F	Fluoride	
		Fe	Iron	
		Mg	Magnesium	
		Mn	Manganese	
		NO3	Nitrate	
		NO2	Nitrite	
		K	Potassium	
		Na	Sodium	
		Si	Silica	
		SO4	Sulphate	
		SS	Suspended solids	
		TDS	Total dissolved solids / salts	
		TAL	Total alkalinity	
TH	Total hardness			
µg/l	mg/l		Heavy metals	1000.
mS/m	µS/cm	EC	Electrical conductivity <sup>2</sup>	0.10
counts/100 ml	Counts/100 ml	Ecoli	Escherichia coli	1.0
		FC	Faecal coliforms	
		TC	Total coliforms	
-	-	pH	pH	1.0
NTU	NTU	Turb	Turbidity	1.0
°C	°C	TEMP	Temperature	1.0

## Notes:

- 1 In instances where the user-defined units differ from those listed, the user will need to carry out appropriate file manipulations to convert to recognised units.
- 2 In the DWS' WMS database electrical conductivity is corrected to 25°C. Some users like Rand Water correct to 20°C. Since electrical conductivity is sensitive to temperature, it is important to obtain appropriate temperature conversion tables to adjust the electrical conductivity values to a common base. This needs to be done during the pre-processing stage. (Generally this is not necessary if EC is a secondary variable used to patch TDS at a single **Site** since then the TDS and EC would be regressed against one another using analyses derived from common samples, rendering the units immaterial, provided they remain the same throughout the processing period.)

## MANUALLY PREPARED DAILY RIVER FLOW DATA FILE

The raw data used in manually prepared daily river flow data files can originate from any **Source**, such as Water Boards, municipalities, mines, industries, CMAs, DWS Regional Offices, etc. The only requirement is that the pre-processed product must conform to the following format:

Row 1: Headings for each column containing the desired data (optional, but recommended).

Row 2 to end: Any number of columns in any order that must include the following two data columns:

### Date

This should be the day on which the daily flow on the same row is applicable.

The daily flow sampling date can be in one of the following formats (the same format for every row):

- yyyy/mm/dd
- mm/dd/yyyy
- dd/mm/yyyy
- yyyymmdd
- mmddyyyy
- ddmmyyyy

## Daily flow value

The flow variable can have any of the following daily flow units, which can be selected from the drop down list in the **Import Flow dataset** dialog:

- m<sup>3</sup>/s
- l/s
- m<sup>3</sup>/day
- kl/day
- Ml/d

There is no provision to add to this list of standard flow units. If the units supplied by a **Source** organisation differ from these, then the pre-processing must include conversion to one of the above unit types.

The required columns can appear in any order and can be interspersed with columns containing intermediate calculations, or even unrelated data.

## Point source data

In addition to importing river flow data, the **Import Flow dataset** dialog is also used to import point source flow data.

This data is only required when the river flow records at two or more stations are to be combined to patch the daily flows at a water quality sampling station, or if the location of the river flow gauging station differs significantly from that of the water quality monitoring station and there are significant net effluent discharges upstream of one or both of the monitoring stations. In these cases it is necessary to take into account the cumulative net point source inflow upstream of each flow gauging point.

This point source data can be expressed as daily or monthly flows. The former uses exactly the same date formats and flow units as are available in the drop-down lists for daily river flow data.

Effluent flow data is often made available by different **Sources** as monthly totals. If the raw data is obtained in this form, then the following drop-down list of valid date formats is applicable:

- yyyy/mm
- yyyy/mm

The valid flow units become:

- kl/month
- MI/month
- million m<sup>3</sup>/month

As before, the required date and flow columns can appear in any order and can be interspersed with columns containing intermediate calculations, or even unrelated data.

More than one flow column can be included in the same file.

It is important to observe that only one dataset of total net point source inflow is permissible for each river flow station and the water quality monitoring station that it is to represent.

Since the combination of flow **Sites** is not a common eventuality, rather than unnecessarily burdening the programming task, the onus has been placed on the user to ensure that the pre-processing fills in all of the gaps in the historical point source flow records.

Some smaller municipalities also use very strange formats in their spread sheets, which cannot be generalised and have to be pre-processed by the user.

In every case MOVE processes the imported flow data into the internal project standard of m<sup>3</sup>/s using the following conversion factors:

<u>Flow units</u>	<u>Factor</u>
m <sup>3</sup> /s	1.0000
l/s	0.0010
m <sup>3</sup> /day	1/86400
kl/d	1/86400
MI/d	1/86.4
kl/month	1/(86400•NM) - where NM is the number of days in the month
MI/month	1/(86.4•NM)
10 <sup>6</sup> m <sup>3</sup> /month	1/(0.0864•NM)

## **6 RESULTS, TREATMENT OF RESULTS AND DISCUSSION**

### **6.1 Aim 1: Develop a user-friendly integrated water quality patching system.**

This aim has been met.

### **6.2 Aim 2: Incorporate enhancements of the existing MOVE program.**

This aim has been met and exceeded with additional enhancements that were identified as the project progressed,

## **7 CONCLUSIONS**

### **7.1 Patching tool available**

The aims of the project have been met in full. MOVE represents a powerful new user-friendly tool for the patching and evaluation of water quality data that encapsulates several years of expert know-how.

### **7.2 Unique features**

The moving regression patching technique is an innovation.

### **7.3 Increased efficiency**

A substantial benefit will accrue from the speeding up the water quality data patching process. This time saving should reduce project costs and free up scarce professional skills to focus on salinity model calibration, modelling and options analysis.

### **7.4 Improved quality**

The time and budget consuming complexity has often led to superficial handling of water quality data patching. Moreover, some practitioners are not even aware of what can be achieved or of the consequences of applying inferior techniques. MOVE provides a framework that draws attention to what is required and the means to rapidly carry out the required processing.

### **7.5 Documentation**

Although Dr Herold developed a precursor to MOVE in 2002 and it has been used in studies since then, this report is the first time it has been documented. Moreover, the online help screens made this documentation more accessible. The documentation goes a long way towards facilitating dissemination of the products.

### **7.6 Flagging outliers**

Bad data can hold serious consequences; especially since many commonly used statistical techniques make use of the squares of values. As a result the importance of outliers is greatly magnified, especially in water quality analyses where high values often occur. The consequences in terms of salinity model calibration are not as serious when a high concentration coincides with low dry weather base flow conditions since the ensuing loads remain low. However, a high concentration occurring during a flood can greatly distort the supposed salt load entering downstream dams, as can an exceptionally low value, with misleading results and decisions based on them.

Consequently it is extremely important for the outcome of modelling projects to identify and eliminate such outliers.

However, it is also important to flag outliers and draw them to the attention of water quality database custodians. If this is not done, then the same outliers will remain as pitfalls for future users of the data, with the number of outliers increasing steadily as the size of the database grows.

MOVE provides a means of both identifying and documenting outliers.

## **7.7 New possibilities**

MOVE makes it possible to explore a number of attractive possibilities outlined in Chapter 8.



## **8 RECOMMENDATIONS**

### **8.1 Dissemination**

The MOVE patching system provides water quality modellers with a powerful tool to improve both the quality and the efficiency of the patching of water quality records. It is therefore important to find ways to make this tool available to practitioners.

Dissemination can best be achieved through:

- Publicity via Water Wheel
- Training courses.

### **8.2 Develop national patched database**

MOVE provides a tool to assist in developing a national patched and improved water quality database. This would involve using MOVE to patch gaps in the daily water quality records at national and regional monitoring stations. A parallel database of aggregated monthly water quality records would be compiled. This would greatly simplify the task of water quality modelling practitioners and researchers.

An important by-product would be a database of suspect outliers and cleaned up datasets with the outliers removed. This would assist greatly in reducing the incidence of misleading results due to the use of suspect data.

### **8.3 Confidence intervals**

It is desirable to test for errors based on a range of conditions of sampling frequency, river type (dam and river reach), catchment size, state of catchment development and proportional effluent inflow. The water quality data available at each site could be sampled at lower frequencies to establish significant relationships with sampling frequency. Suitable algorithms could then be developed taking account to the most appropriate catchment and river characteristics. These could be used to provide the user with confidence intervals for each site of interest and even for different sections of the historical record. This may be important since historical sampling frequencies and catchment conditions are subject to change.

### **8.4 Optimise sampling frequency**

Modelling for the planning and operation of water resource systems is one of the most important uses of water quality data. However, it is often found that water quality data is too sparse to support reliable modelling. This can hold large economic consequences due to

either over- or under-estimating water quality impacts. On the other hand, water quality sampling is expensive and too high a sampling frequency can be wasteful.

Selecting monitoring stations with higher sampling frequencies over part of their record makes it possible to calculate the load and other key statistics using the complete (say) daily record. Then remove six out of every seven daily values and use MOVE to calculate the same key statistics. The exercise would then be repeated for two-weekly and monthly sampling frequencies. The results can then be used to calculate the change in the error with changing sampling frequency.

Optimum monitoring frequencies can then be derived for each condition and appropriate strategies devised.

This would be repeated for a range of key conditions, such as catchment size, upstream effluent discharge and nature of monitoring **Site** (river or dam).

MOVE is the ideal vehicle for cleaning up the primary datasets for each selected monitoring **Site**. A simple program or spreadsheet can be written to read in each checked data file and generate the necessary statistics.

## **8.5 Expand to non-conservative variables**

MOVE can be tested for its applicability to patch key non-conservative pollutants. Where necessary the software can be expanded to better deal with non-conservative water quality variables.

## **8.6 Improve patching at reservoir stations**

At present linear interpolation is used to patch water quality records at dam outlets. This is valid for most of the time where the storage capacity is large compared to the Mean Annual Runoff (MAR). However, smaller dams or larger weirs (such as Douglas) behave more like river stations during larger floods. It would therefore be useful to include the option of also carrying out a moving regression at dam stations and allowing the user to view the two patching options simultaneously and choose to use a moving regression during bigger floods, as appropriate.

Vertical stratification can also lead to significant systematic differences between the water spilling at the surface and that released via can outlets and low level sluice gates. A means of taking this into account should be investigated and implemented.

## **8.7 Patching point sources**

Gaps in the patched and aggregated water quality datasets at river stations do not affect the running of water quality models, where the dataset is used for calibration against. However, point source input datasets have to be complete since they are essential inputs to the model. Hence it is necessary to completely patch effluent inflow records. What helps here is that at a monthly scale effluent discharges are greatly attenuated compared with catchment runoff. However, they are subject to trends and seasonal variation.

MOVE can be adapted to include this process. Dr Herold has developed methodology for this purpose, taking account of trends and seasonal variation. Embracing this functionality within MOVE would be very advantageous for modelling practitioners.

## **8.8 Handle gaps in new point source inflows**

At present the stipulation has been made that the net point source inflow time series datasets entered into MOVE have to have been pre-processed and may not contain gaps. Modifying this to handle gaps in the net point source inflow record would be a useful feature. (Implementing 8.6 would be even better.)

## **8.9 Adjusting the minimum number of observation points with flow**

The aggregation option calls for the specification of seasonal monthly minimum numbers of observed water quality values per month. It would be useful to automatically adjust this requirement as the daily flow rate varies.

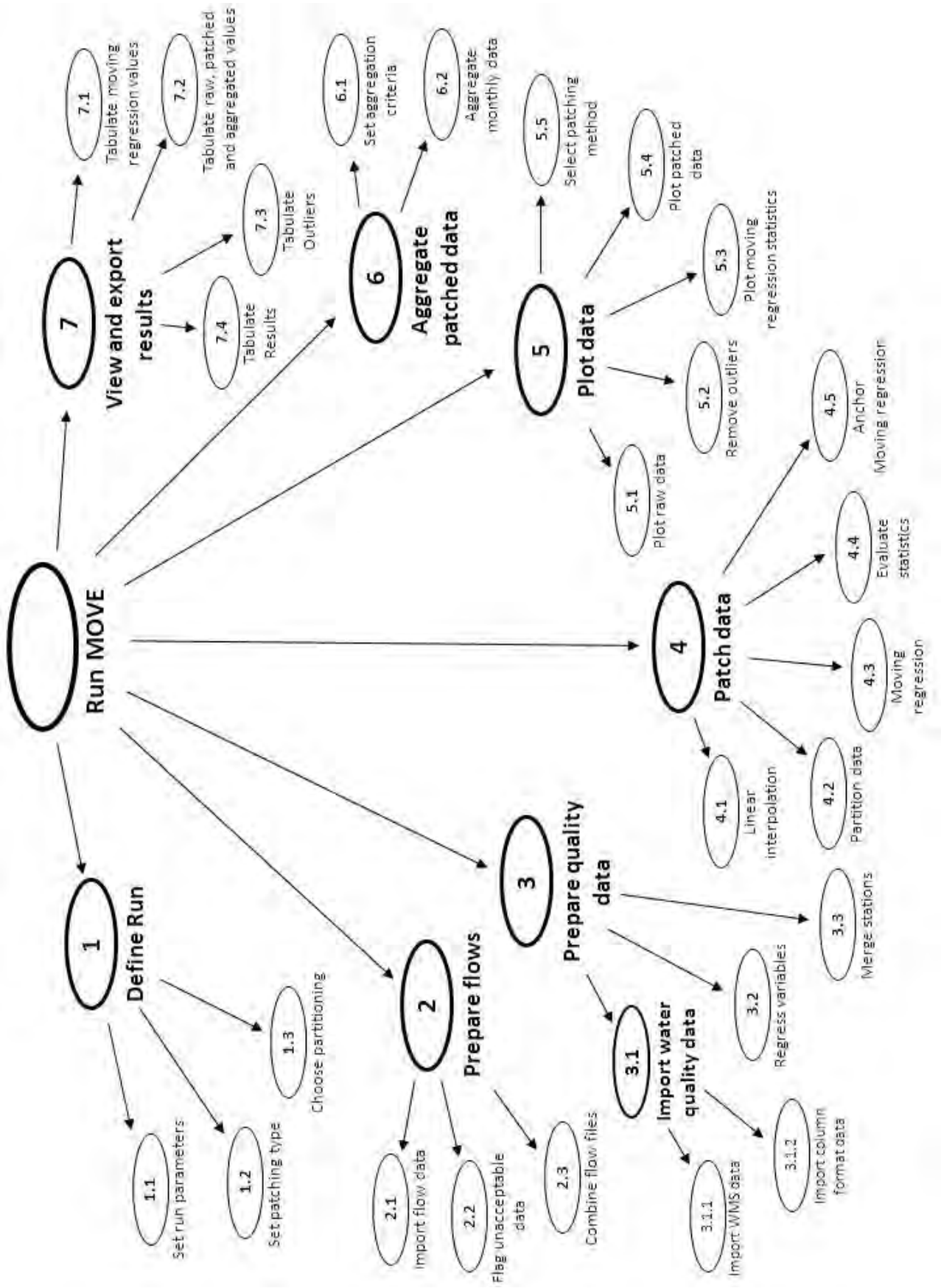
## 9 LIST OF REFERENCES

Herold, CE (1981) A model to simulate daily river flows and associated diffuse-source conservative pollutants. Report No. 3/81, Hydrological Research Unit, University of the Witwatersrand, Johannesburg, March 1981

Herold, CE (2007) The Upper and Middle Olifants River catchment: The development of an Integrated Water Resources Management Plan: Middle Olifants catchment: WQT model calibration. Umfula Wempilo Consulting report to Golder and Associates. Directorate of Project Planning, Department of Water Affairs and Forestry, Pretoria.

Metcalf and Eddy, Inc., Storm Water Management Model, Volume 1, Final Report, Environmental Protection Agency, Washington, D.C., 1971.

# APPENDIX A: Use Case Diagram



# APPENDIX B: Use Case Sequences

## 1. Define run

### 1.1) Set run parameters

Main Goal: Define run parameters

- |   |   |                       |
|---|---|-----------------------|
| 1 | Set date range and define primary and secondary water quality codes to be used in reporting the results for each water quality variable and appropriate units for each. (The primary water quality variable is the one to be used in the analyses (e.g. TDS). The secondary variable is the one to be regressed against the primary variable and used to estimate the primary variable (e.g. Electrical Conductivity, EC). This is necessary since EC is often measured at a higher frequency than is TDS. The primary variable could also be one of the major constituents of the TDS such as Na, Ca, Mg, K, Cl, SO4 or TAL with EC or TDS being the secondary value.) | Run definition Dialog |
|---|---|-----------------------|

### 1.2) Select Patching type

Main Goal: Determine if reservoir or river patching is appropriate

- |   |   |                       |
|---|---|-----------------------|
| 1 | Determine if reservoir or river reach patching is appropriate. Simple question and answer expert system provided to aid the user. | Run definition Dialog |
|---|---|-----------------------|

### 1.3) Choose partitioning criteria

Main Goal: Define flow, concentration or seasonal partitioning ranges.

- |   |   |                       |
|---|---|-----------------------|
| 1 | Interrogate type of partitioning required: by flow range, by concentration range or by season.  | Run definition Dialog |
| 2 | Define partitioning ranges. For flow or concentration define ranges as flow or concentration percentiles. Otherwise define the seasonal partition boundaries as the month when the high flow season starts and the month when the low flow season starts. | Run definition Dialog |

## 2 Prepare flows

### 2.1) Import flow data

Main goal: Read daily flow data

- |   |   |                    |
|---|---|--------------------|
| 1 | Read daily date and flow data in standard DWS Hydrology format. Store data in arrays if the date is within the range specified in 1.1) step 1. Flag days with missing data. | Flow Import Dialog |
|---|---|--------------------|

## APPENDIX B: Use Case Sequences

- |   |  |                    |
|---|--|--------------------|
| 2 | If the data file is in another column format, specify the number of header records and determine which columns contain the date and flow data. Set missing data code ("-1") for all array positions within the specified date range. | Flow Import Dialog |
| 3 | Read the daily date and flow data. Store data in arrays if the date is within the range specified in 1.1) step 1. Provide an option to cut and paste data.   | Flow Import Dialog |
| 4 | Read simulated monthly data in standard WRSM2000 format and disaggregate into daily values and store in array. Flag days with missing data.  | Flow Import Dialog |

### 2.2) Flag unacceptable data

- |   |  |                    |
|---|--|--------------------|
| 1 | Display type definitions that are accepted by default. Allow user to modify the acceptable list. | Flow Import Dialog |
| 2 | Modify array of data that was input in 2.1 by flagging unaccepted data as missing.               | Flow Import Dialog |

### 2.3) Combine flow files

Main goal: Combine two or more flow files

- |   |   |                              |
|---|---|------------------------------|
| 1 | Select flow files to be combined in order of preference                 | Flow File Combination Dialog |
| 2 | Allow for priority order to be changed.                                 | Flow File Combination Dialog |
| 3 | Define MARS of alternative flow files to facilitate normalisation.      | Flow File Combination Dialog |
| 4 | Combine files into a single daily flow file to be used in the analysis. | Flow File Combination Dialog |

### 3.1.1) Import water quality data in WMS format

Main goal: Read grab sample water quality data in DWS's standard WMS database output file format

- |   |  |                             |
|---|--|-----------------------------|
| 1 | Open the file containing the WMS data, read and display the water quality parameter codes contained in the first line and select the appropriate primary and secondary water quality parameters. | Water quality Import Dialog |
| 2 | Display station codes and descriptions contained in the WMS file and select the desired station.   | Water quality Import Dialog |
| 3 | Set missing data code ("-1") for all array positions within the specified date range.  | Water quality Import Dialog |

## APPENDIX B: Use Case Sequences

- |   |   |                             |
|---|---|-----------------------------|
| 4 | Read sample date and quality data for the selected primary and secondary (if applicable) water quality variable, store in arrays if the date is within the specified range. | Water quality Import Dialog |
|---|---|-----------------------------|

### 3.1.2) Import water quality data in column format

Main goal: Read grab sample water quality data in column format

- |   |  |                             |
|---|--|-----------------------------|
| 1 | Define columns containing date, primary and secondary (if applicable) water quality variables. Define number of header records.  | Water quality Import Dialog |
| 2 | Set missing data code ("-1") for all array positions within the specified date range.  | Water quality Import Dialog |
| 3 | Read sample date and quality data for the selected primary and secondary (if applicable) water quality variable and store in arrays if the date is within the specified range. | Water quality Import Dialog |

### 3.2) Regress variables

Main goal: Regress primary and secondary water quality variables

- |   |  |                             |
|---|--|-----------------------------|
| 1 | Calculate linear regression constants between primary and secondary variables  | Variables Regression Dialog |
| 2 | Use regression to calculate primary variable from secondary one and store in different array.  | Variables Regression Dialog |
| 3 | After using plot routine to identify and eliminate outliers, merge primary and regressed values to obtain a single file of primary values, with the observed primary data having first priority. | Variables Regression Dialog |



## APPENDIX C: Object Responsibilities

### Aggregation Dialog = Class(Dialog)

- |         |  |         |
|---------|--|---------|
| 6.1 - 1 | Set minimum number of observation points required for a month before an aggregated average will be included. Otherwise a missing data code will be used. 12 monthly seasonal values to be specified. | Not now |
| 6.1 - 2 | Set minimum number of patched daily values required per month before an aggregated average will be included. Otherwise the month is flagged as missing. 12 monthly seasonal values to be specified.  | Not now |
| 6.1 - 3 | Selection of flow-weighted or arithmetic aggregation,  | Not now |
| 6.2 - 1 | Carry out flow-weighted aggregation using the defined criteria (6.1).  | Not now |
| 6.2 - 2 | Carry out arithmetic aggregation using the defined criteria (6.1).   | Not now |

### Anchoring Dialog = Class(Dialog)

- |         |  |         |
|---------|--|---------|
| 4.5 - 1 | Anchor moving regression values against observed data.     | Not Now |
| 4.5 - 2 | Warn of instances when anchoring produces negative values. | Not Now |

### Data Partition Dialog - Class(Dialog)

- |         |   |         |
|---------|---|---------|
| 4.2 - 1 | Spilt data into partitions prior to carrying out moving regression. (This is not applicable to linear interpolation.) | Not now |
| 4.2 - 2 | Join partitions to form complete dataset.   | Not now |

### File Merge Dialog = Class(Dialog)

- |         |   |         |
|---------|---|---------|
| 3.3 - 1 | Define water quality stations to be merged in order of priority.  | Not now |
| 3.3 - 2 | Allow for change in order of priority.  | Not now |
| 3.3 - 3 | Merge files into single array. (But retain the individual station arrays to facilitate plotting results and to accommodate changes in order of priority.) | Not now |

### Flow File Combination Dialog = Class(Dialog)

## APPENDIX C: Object Responsibilities

2.3 - 1	Select flow files to be combined in order of preference	Not now
2.3 - 2	Allow for priority order to be changed.	Not now
2.3 - 3	Define MARs of alternative flow files to facilitate normalisation.	Not now
2.3 - 4	Combine files into a single daily flow file to be used in the analysis.	Not now

### Flow Import Dialog = Class(Dialog)

2.1 - 1	Read daily date and flow data in standard DWS Hydrology format. Store data in arrays if the date is within the range specified in 1.1) step 1. Flag days with missing data.	Not now
2.1 - 2	If the data file is in another column format, specify the number of header records and determine which columns contain the date and flow data. Set missing data code ("-1") for all array positions within the specified date range.	Not now
2.1 - 3	Read the daily date and flow data. Store data in arrays if the date is within the range specified in 1.1) step 1. Provide an option to cut and paste data.	Not now
2.1 - 4	Read simulated monthly data in standard WRSM2000 format and disaggregate into daily values and store in array. Flag days with missing data.	Not now
2.2 - 1	Display type definitions that are accepted by default. Allow user to modify the acceptable list.	Not now
2.2 - 2	Modify array of data that was input in 2.1 by flagging unaccepted data as missing.	Not now

### Linear Interpolation Dialog - Class(Dialog)

4.1 - 1	Linear interpolation of daily values between regressed and merged water quality monitoring points. This is appropriate for reservoir stations and river reaches that are heavily regulated by adjacent upstream reservoirs.	Not now
---------	---	---------

### Moving Regression Dialog = Class(Dialog)

4.3 - 1	Select range of moving regression periods.	Not now
4.3 - 2	Carry out moving regression of the primary water quality variable against daily flow. Use regression constants to infill water quality on days when flow data is available. Calculate moving	Not now

## APPENDIX C: Object Responsibilities

regression statistics (4.4). Repeat for each partition of the data (see 4.2). Repeat for each iterated moving regression period.

- |         |   |         |
|---------|---|---------|
| 4.3 - 3 | Plot regression statistics (5.4).   | Not now |
| 4.4 - 1 | Calculate the moving regression statistics for each moving regression period and the composite total, as well as the regression constants applicable between each pair of observed data points. | Not now |

### **Moving Regression Report Dialog = Class(Dialog)**

- |         |   |         |
|---------|---|---------|
| 7.1 - 1 | Prepare arrays containing moving regression constants and goodness of fit for each regression equation and showing the time range for which each is applicable. | Not now |
| 7.1 - 2 | Display the tabulated results.  | Not now |
| 7.1 - 3 | Print the tabulated results.  | Not now |
| 7.1 - 4 | Store the results as text file or excel spreadsheet.  | Not now |

### **Outliers Report Dialog = Class(Dialog)**

- |         |   |         |
|---------|---|---------|
| 7.3 - 1 | Prepare report on outliers that have been removed from each file during the patching procedure. | Not now |
| 7.3 - 2 | Provide metadata report metadata giving the reasons for removal of the outliers.                | Not now |

## APPENDIX C: Object Responsibilities

### Plot Dialog = Class(Dialog)

5.1 - 1	Plot time series on three axes. Bottom axis for observed daily flows, second axis for secondary variable (if any), top axis with two lines: one for primary variable regressed from secondary variable, second line for observed primary variable. Adjacent daily flow values to be joined symbols to be used only to show isolated points. All observed primary and regressed primary values to be denoted by different centred symbols and colours. Observed secondary values to be denoted by symbols. Choice of natural and logarithmic vertical scales. Option to drag window for date range and vertical axis range.	Not now
5.2 - 1	Click on points to select and then remove outliers.	Not now
5.2 - 2	Invoke (7.3) to document outliers.	Not now
5.2 - 3	After outliers removed, invoke 3.2 to re-regress variables and re-plot the screen (5.1).	Not now
5.3 - 1	Plot moving regression constants and goodness of fit characteristics for each moving regression period.	Not now
5.4 - 1	Plot two sets of axes. The lower showing observed daily flows, the upper the observed and patched water quality.	Not now
5.4 - 2	For reservoir patching the upper set of axes is to contain three plotted lines: First aggregated monthly values derived from patching by linear interpolation (derived from 6.1). Second daily values patched by linear interpolation using the same colour. Third symbols for each of the observed primary data points.	Not now
5.4 - 3	For River stations the upper set of axes is to contain seven sets of plotted lines, two for each patching method comprising monthly aggregated and patched daily values sharing the same colours. The three patching methods include linear interpolation, moving regression without anchoring and moving regression with anchoring. The last line is to show symbols only (no line) for the observed data points.	Not now
5.4 - 4	The user should be allowed to click on and remove any further values that are identified as outliers (see 5.2).	Not now
5.5 - 1	This is applicable to river patching. First enable selection of one patching method for the entire period. This should darken or otherwise highlight all aggregated monthly values for that patching method. Clicking on another aggregated monthly value should reassign the selection for that month to the new patching method. The option of removing the aggregated value for a month should be allowed.	Not now
5.5 - 2	Selected monthly values should be assigned to the final patching array.	Not now

## APPENDIX C: Object Responsibilities

### Results Report Dialog = Class(Dialog)

- |         |  |         |
|---------|--|---------|
| 7.2 - 1 | Prepare arrays containing raw, patched and aggregated water quality values. Indicate the files that have been merged and the periods for which different patching methods were used. | Not now |
| 7.2 - 2 | Display the tabulated results.   | Not now |
| 7.2 - 3 | Print the tabulated results.   | Not now |
| 7.2 - 4 | Store the results as text file suitable as input to WQT model or in excel spreadsheet.   | Not now |

### Run Definition Dialog = Class(Dialog)

- |         |   |         |
|---------|---|---------|
| 1.1 - 1 | Set date range and define primary and secondary water quality codes to be used in reporting the results for each water quality variable and appropriate units for each. (The primary water quality variable is the one to be used in the analyses (e.g. TDS). The secondary variable is the one to be regressed against the primary variable and used to estimate the primary variable (e.g. Electrical Conductivity, EC). This is necessary since EC is often measured at a higher frequency than is TDS. The primary variable could also be one of the major constituents of the TDS such as Na, Ca, Mg, K, Cl, SO <sub>4</sub> or TAL with EC or TDS being the secondary value.) | Not now |
| 1.2 - 1 | Determine if reservoir or river reach patching is appropriate. Simple question and answer expert system provided to aid the user.   | Not now |
| 1.3 - 1 | Interrogate type of partitioning required: by flow range, by concentration range or by season.  | Not now |
| 1.3 - 2 | Define partitioning ranges. For flow or concentration define ranges as flow or concentration percentiles. Otherwise define the seasonal partition boundaries as the month when the high flow season starts and the month when the low flow season starts.   | Not now |

### Variables Regression Dialog = Class(Dialog)

- |         |  |         |
|---------|--|---------|
| 3.2 - 1 | Calculate linear regression constants between primary and secondary variables  | Not now |
| 3.2 - 2 | Use regression to calculate primary variable from secondary one and store in different array.  | Not now |
| 3.2 - 3 | After using plot routine to identify and eliminate outliers, merge primary and regressed values to obtain a single file of primary values, with the observed primary data having first priority. | Not now |

## APPENDIX C: Object Responsibilities

### Water Quality Import Dialog = Class(Dialog)

3.1.1 - 1	Open the file containing the WMS data, read and display the water quality parameter codes contained in the first line and select the appropriate primary and secondary water quality parameters.	Not now
3.1.1 - 2	Display station codes and descriptions contained in the WMS file and select the desired station.	Not now
3.1.1 - 3	Set missing data code ("-1") for all array positions within the specified date range.	Not now
3.1.1 - 4	Read sample date and quality data for the selected primary and secondary (if applicable) water quality variable, store in arrays if the date is within the specified range.	Not now
3.1.2 - 1	Define columns containing date, primary and secondary (if applicable) water quality variables. Define number of header records.	Not now
3.1.2 - 2	Set missing data code ("-1") for all array positions within the specified date range.	Not now
3.1.2 - 3	Read sample date and quality data for the selected primary and secondary (if applicable) water quality variable and store in arrays if the date is within the specified range.	Not now

## APPENDIX D: DWS Hydrological check codes

Code	Print Quality	Text
1		Good continuous data
2		Good edited data
3		Preserved historical data
4	Q	Unaudited
5		Height derived from flow
6	D	Drops
7	Q	Good edited unaudited
8	@	Good Weekly Reading
25	Q	Unaudited Gauge Plate Readings / dip level readings
26	\$	Audited Gauge Plate Readings / dip level readings
27	&	Good monthly reading
50	S	Gap filled data
58	H	Hb not available\044 assume no submergence
59	V	Static or reverse flow due to backwater submergence conditions
60	A	Above Rating
64	E	Audited Estimate
65	E	Unaudited Estimate
66	*	Program Estimate
70	?	Unknown
78	U	Not Accumulated Unreliable
79	%	Accumulated Unreliable

## APPENDIX D: DWS Hydrological check codes

80	+	Accumulated Reliable
81	#	Wet day within accumulated rainfall period
90	<	Water level below instrument
91	>	Minimum Value
92	<	Maximum Value
93	<	Dry borehole
94	>	Artesian borehole level
95	<	Borehole Seepage
100	?	Flag Boshielo Dam under construction. New FSL not yet implemented.
130	E	Used previous week's level as an estimate for this week
140	!	Data not yet checked
150	^	Rating table extrapolated - flows estimated
151	M	Data Missing
152	~	Negative
153	F	No height data/ Flow data only
154	[	Reversal start
155	]	Reversal end
160	Z	No info for stage/discharge determination (zero dt loaded)
161	T	Rating missing
162	R	Rating unreliable
163	G	Gate(s) in operation - no spillway discharge
164	B	Continuously variable submergence flow derivation DT in operation
165	P	Estuarine water level recording only - no flow calculated



## APPENDIX D: DWS Hydrological check codes

170	M	Period of No Record (PNR)
171	M	Data Exists but Unreliable
172	M	Temporary Gap
173	?	Data Unreliable
201	[	Data not recorded or incomplete
245	V	Undefined submergence flow calc program exception
246	M	No cross-sectional area upstream of notch/structure
247	V	ha > rating table limit - no calculation performed
248	V	No hb data
249	V	No ha data
250	V	Structural submergence > 97.7%
251	V	Static or reverse flow possibilities
252	V	Froude number > 0.8 at inlet section
253	V	Flow not converging to constant val after max # iterations
254	A	Rating Table Exceeded
255	M	Data Missing