A Software Tool for the Estimation of Pollutant Loads in Rivers and Streams Using Time Series Data

Tennakoon, S.B.[1,4], N.Marsh[2,4] and S.Arene[3,4]

[1]Queensland Environmental Protection Agency No 80, Meiers Road, Indooroopilly
[2]CSIRO Land and Water, No 80, Meiers Road, Indooroopilly
[3]Griffith University, Nathan Campus, 170 Kessels Road, Nathan
[4]eWater Cooperative Research Centre, University of Canberra, Canberra
Email: sunil.tennakoon@epa.qld.gov.au

Keywords: *Catchments, River, Constituents, Loads, Time series data*

EXTENDED ABSTRACT

The determination of constituent loads or loading rates from rivers and streams is an essential element in environmental and ecological management. There has been a change in the focus of water quality target setting from the traditional focus on ambient conditions for in-stream biota to the development of load-based targets. The interest in load-based targets is particularly prevalent where the focus is on the receiving waters, such as the Great Barrier Reef catchments. Whilst there has been a general increase in urgency to set load-based targets, there has only been limited discussion to date as to the appropriate methods and procedures for assessing current conditions and developing loads-based targets.

The analysis of concentration and flow data for the computation of loads is not a trivial task. Due to the relatively sparse nature of concentration data, special load estimation techniques should be applied to reduce estimation errors. To date around 30 alternate methods have been developed under four major categories such as simple integration, ratio methods, regression techniques and complex models. These techniques have been developed based on some assumptions about the behaviour of pollutant concentrations in-stream during the times when the water quality was not sampled. The differences in these methods stems largely from the variation in the underlying data types and distributions used to calculate the loads. Selection of an appropriate load estimation method for a particular location and a data set is not an easy task as there are often no reference or true values available for comparison. However, some methods of load calculation can provide the level of uncertainty of estimates and to some extent these measures may be used to determine preferences among available techniques.

The National Action Plan for Salinity and Water Quality Program in Queensland supported by the Department of the Environment and Heritage (DEH) Coastal Catchments Initiative has developed the Loads Tool. The basic architecture of this tool is given in Figure 1. This tool presents nine of the most common methods for long-term load calculation and four methods for estimating loads from storm events. A further built-in function is available to calculate event mean concentration (EMC) values which are useful in catchment modelling exercises. The tool has a user friendly interface compatible with other catchment modelling software tools developed by the eWater CRC. This tool also contains an interactive help system to guide the user in selecting the most appropriate method for their circumstances. The beta version of this loads tool has been tested in Queensland catchments and received very positive feedback from the users. The revised version of the Loads Tool will be available nationally through the eWater CRC Integrated Monitoring and Assessment (IMAS) product development program.

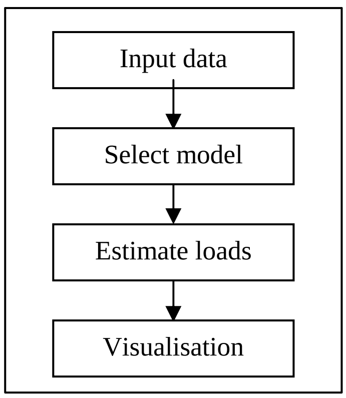

Figure 1. Basic structure of the Loads Tool.

1. INTRODUCTION

Accurate estimation of constituent loads of a river or stream is difficult due to the complex behaviour of constituents in natural streams and the variability of water flows. The constituent load is the weight of constituents which passes a cross-section of the river in a specific amount of time. Loads are expressed in mass units (e.g. tons, kilograms), per specific period of time and are sometimes known as loading rate or flux. Loading rate is the product of pollutant concentration and discharge rate. Therefore an essential component of loads estimation is flow rate, which is the volume of water that passes a cross-section of the river in a specific amount of time and has units of volume/time such as cubic meters per second. The other required component for estimating loads is the constituent concentration which has units of weight per unit volume of water (milligrams per litre). In natural river systems, both flow rate and concentration values vary drastically over time. In urban situations, discharge of effluents, construction activities and other human induced processes add complex variability to both flow and concentration components. In rural condition the main sources pollutants are diffuse sources and seasonal changes with different rainfall regimes, land use activities, and discharges of various chemicals from primary production systems cause spatial and temporal variability in flow rates, as well as concentrations.

Flow is generally recorded on a regular basis by measuring the stage, or water height, and using a previously established rating curve to convert stage into flow. In many locations flow measurements are made inexpensively using automated equipment and many techniques for flow measurements are available (Water Measurement Manual 1997). The measurement of flow is a well-established science, and many excellent books and reports are available to describe the methods involved. In contrast, constituent concentration measurements are expensive and can range in cost from a few dollars to more than a thousand dollars per sample, depending on what parameter or parameters are being measured. Obtaining concentration measurements usually involves taking water samples to a laboratory for chemical analysis. Further, stringent quality control and quality assurances are required to obtain reliable concentration values. Because concentration measurements are much more expensive than flow measurements, usually constituent observations are made less frequently than flow observations. Frequency and methods of sampling water quality in streams depends on available resources, nature of constituent behaviour, statistical designs, accessibility and available technology. As a consequence there are many load calculation techniques to account for sampling frequency and alternate sampling methods. These techniques have been grouped into four major categories based on some assumptions about the behaviour of in-stream constituent concentrations in the times when water quality was not sampled.

A software tool has been designed and developed to estimate loads using time series of flow data and concentration data. The current version of the software contains nine different methods of load calculations for long-term estimates and four methods for event based assessments. The software consists of a comprehensive help system including guidance for selecting appropriate methods to counterpart quantity and quality of available data. A user-friendly user interface was developed for easy operation with the facility of visualising predicted and estimated outputs. The current version has the ability to calculate input data variability and the next version will include the feature of calculating uncertainties of estimated loads for each of the methods. The software tool was tested and evaluated for its usability and possible errors during the prototyping stage. The beta version was tested via a number of workshops held in Queensland and very positive responses were received from end-users. The latest version of this loads tool will be available through the eWaterCRC toolkit website.

2. METHODS OF LOAD CALCULATION

Many different approaches have been used to calculate loads from observed concentration and flow data (Preston *et al*. (1989); Letcher *et al*. (1999)). Some are more precise and/or more accurate than others; some are only appropriate under special circumstances (Huai-en *et al*. 2003). The available methods of load estimation have been categorised under four main types:

- Averaging techniques
- Ratio method
- Regression method
- Complex models

2.1. Averaging Techniques

Assumptions are made about how concentrations vary in time between samples. Averaging approaches use some form of average in the calculation of the loads. The simplest approach involves multiplying the average concentration for a period of time by the mean daily flow for each day in the time period to obtain a succession of estimated daily (unit) loads. Commonly, concentration values for un-sampled dates are derived by interpolation or extrapolation of concentration data. Typical interpolation techniques are linearly interpolate between concentrations.

2.2. Ratio Techniques

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

2.3. Regression or Rating Curve Techniques

Assuming a relationship between flow and concentration, the concentration of un-sampled periods are inferred from the flow data and known flow-concentration relationship. These techniques can also be extended to include relationships with other variables such as lagged concentrations, lagged flows, seasonality and long-term trend. These techniques can only be used where a relationship between variables is established and that relationship can reasonably be expected to hold during un-sampled periods.

2.4. Complex Models

Complex mathematical formulas are used to estimate flow rates as well as constituent concentrations. A number of mathematical models are available and the accuracy of estimates is heavily dependent on the type of model, input data and the model parameters used. Some of the models use EMC (Event Mean Concentration) and DWC (Dry Weather Concentration) for estimating annual loads. Generally, complex mathematical models are used to estimate loads from urban and intensive agricultural areas for the reason that there are additional sources of variations in concentration values. Complex mathematical models are not used in current version of the Loads tools.

3. PERFORMANCE OF THE METHODS

Averaging methods are generally biased, and the bias increases as the size of the averaging window increases. For example, a monthly load can be calculated by multiplying the average concentration for the month by the discharge for the month and an appropriate conversion factor to account for the change of units. A quarterly load can be similarly derived using the quarterly discharge and average concentration. In general, the annual load, which is the sum of the four quarterly loads, will be more biased than the annual load, which is the sum of the 12 monthly loads.

Regression approaches can perform well if the relationship between flow and concentration is sufficiently defined; linear throughout the range of flows; and constant throughout the year. Stratification may allow these requirements to be met piece-wise. However, the regression approach may lead to large errors in estimated loads if the available data contains unusual observations that fall away from the trend of the rest of the data, especially if these are associated with high flows.

In most studies, ratio approaches performed better than regression approaches, and both performed better than averaging approaches. In particular, ratio approaches which include a bias correction factor and are used in a stratified mode generally showed low to no bias, relatively high precision, and resistance to undue influence of unusual observations (Preston *et al*. 1989).

3.1. Selection of Methods

Some methods of estimating loads have the additional desirable feature of providing a measure of the uncertainty of the load estimate (Fox 2005). However, the uncertainty estimates of different load calculation methods cannot be directly compared, because they reflect different kinds of "error". For example, the uncertainty estimate for the Beale Ratio Estimator includes a contribution which is due to differences between individual daily loads and the mean daily load in each stratum. It would be logical to confine the notion of error to the difference between estimated loads and the actual (but unknown) loads; a difference which is due only to sampling and analytical error. For these reasons, sometimes uncertainty measures do not provide a reliable means to choose between methods. Consequently, evaluation of load estimation methods must rely on comparative studies in which several methods are used to calculate loads from the same data. The results are compared with the "true" load which is independently known.

Obtaining a true value of constituent loads for a river or stream is almost impracticable in the real world. Therefore, selection of a better load estimation method for a given data set is not an easy task. It is difficult to say which method is better than others as different methods have different type of errors in the estimations. Number of samples, type of flows and constituent characteristics such as strength, type, and consistency of the flow and concentration relationship appear to play an important role in estimator performances. As guidance for selecting appropriate load estimation techniques, a decision tree was developed (Fig 2) using past experiences and reported results from many studies. This decision tree can be used to guide selection of load estimation methods for a particular location and data set, but does not guarantee the suitability of choice.

4. SYSTEM DESIGNS AND DEVELOPMENT

The Loads Tool software has been designed to estimate constituent loads of a river or stream for long-term data (annual) or for an event using time series of flow and concentration data. Among more than 30 different techniques available in the literature, we have used only nine different methods in this version of the software, based on recommendations received from a study conducted at Melbourne University (Fox 2004). These methods are listed below.

Averaging techniques
- Flow x concentration
- Average load
- Flow weighted concentration
- Inter sample mean concentration
- Inter sample mean concentration (mean flow)
- Linear interpolation of concentration
- Flow stratified sampling

Ratio method
- Beale ratio

Regression method
- Concentration power curve

Details of each of these methods together with algorithms used have been included in the software help system.

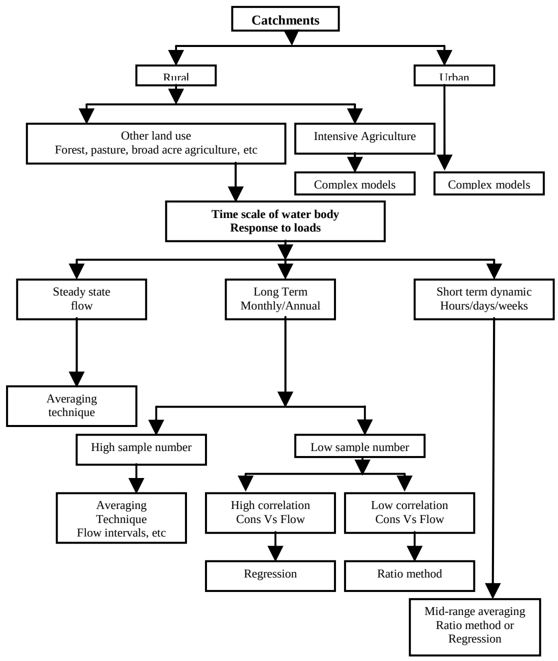

Figure 2. Decision tree for the selection of load calculation methods.

The targeted end-users for this loads tool are people who need to estimate pollutant loads in rivers or streams at local or regional levels, for managing and conserving water quality. These users include research institutes, environmental groups, state and local governments, regional natural resource management organisations and industry groups. Having focused on end-user requirements and capabilities, the interface was developed for easy operation by selecting available options on the screen. The program has been designed to be user-friendly. In particular the structure of the required input is easy to follow and the output is easy to comprehend with some visualisation capabilities. However, a thorough user manual and help system is provided that describes the required input data structures and the handling of outputs in detail.

4.1. User Interface

The Loads Tool interface utilises a clean windows style interface with navigation via a left tree-view and menu. All commonly accessed items are available through the tree view so that running the software requires minimum training. The interface was designed to be compatible (look and feel) with other eWater Cooperative Research Centre toolkit (www.toolkit.net.au) software. A number of interfaces have been designed for visualising estimated loads, predicted concentration values and the variance of input data. A standard help system has been built-in to the system to assist users, explaining basic concepts of load calculations and the details of various methods used in load calculations.

4.2. Code Development

The Loads Tool was written in the C# language using the Microsoft .NET Framework 2.0. Microsoft .NET Framework Version 2.0 Redistributable Package (x86) is required to install and run this software. The .NET Framework 2.0 allows creating applications within TIME (The Invisible Modelling Environment) (Rahman *et al*. 2005), making the codes reusable in other eWater CRC software projects.

4.3. Using the Tool

This software can be navigated using the tree-view structure in the left-hand side of the screen (Fig 3). The user help available with the software also provides assistance to users running the software. The user help document is a standard "HTML" help document allowing the user to search or navigate through the contents page to find required information. The Loads Tool can be used for long-term loads calculations as well and event-based load estimation.

Loads estimation from long-term data (e.g. annual loads): Required input data for load estimations are time series of flow data with regular intervals and concentration data. At least three concentration data points per defined time period (e.g. year) required for some methods and the time interval for concentration data should be equal to or greater than the flow measurement time intervals. Input concentration data files can have multiple constituents in separate columns and the input file specifications are available in the help system. As the first step, it is required to import the time series flow data into the loads software before uploading the concentration data file. Once input data is loaded into the tool the user can view flow and concentration data plots (Fig 3).

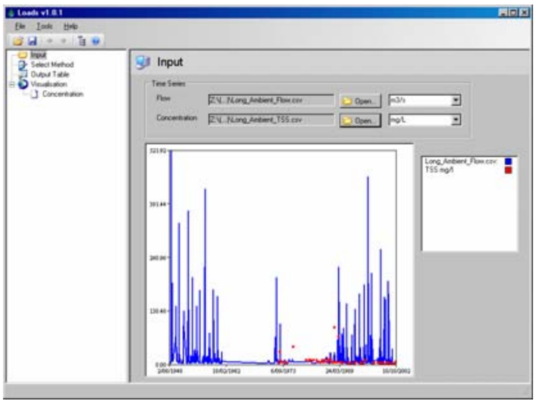

Figure 3. Example of visualised time series flow and concentration data.

Constituent loads can be estimated for different time scales depending on user requirements. These time steps include annual, monthly, weekly daily or total loads for the entire time duration between the start and end date of the time series data. Preferred estimation methods can be selected using the options available in the "Select Method" window (Fig 4).

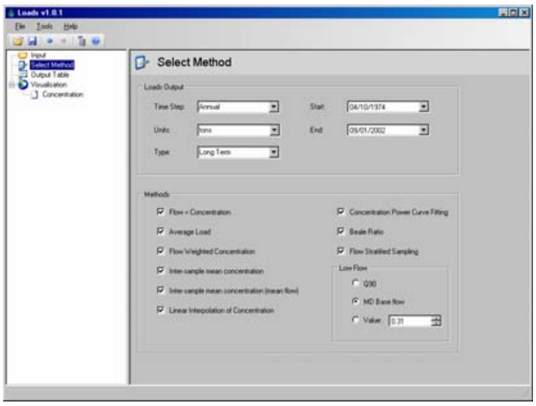

Figure 4. Selection of calculation methods window.

In the event of selecting the "Flow Stratified Sampling" method, the user is required to select a flow separation method, to separate high flow from the base flow. These options are available in the "Low Flow" panel and brief explanations for each of these methods are given below.

Q90: Use up to the 90th percentile of the flow rates as the base flow. About 10% of the flow is considered as high flow.

MD Base flow: Use the mean daily base flow to separate the high flow and base flow. The base flow is calculated using the Lyn-Hollick digital filter as described in the River Analysis Package (www.toolkit.net.au).

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

Estimated loads can be viewed using the "Output Table" option in the tree view. There are two major windows in the output screen. The left-hand window contains the estimated loads for different methods and for different time periods. The variance estimated in the last column is the variance of input concentration data and does not indicate the accuracy of load estimates. It is expected to include the feature of estimating uncertainty of the load estimates in the next release of Loads Tool. Plotted XY graphs for estimated loads using various load estimation methods can be viewed with the "Visualisation" option available in the tree-view. Visualisation of predicted concentration values using different techniques for missing data points is also available.

Event-based load estimation: Especially for particulate pollutants of non-point origin, the flux varies drastically over time, with fluxes during storm runoff events often several orders of magnitude greater than those during low flow periods. It is common that 80 to 90% or more of the annual load will be delivered during 10% of the time with events. Clearly, it is critical to sample during these periods if an accurate load estimate is to be obtained. A subset of load calculation techniques used for long-term estimates are used for estimating loads for event-based data. Event mean concentration (EMC) values are also estimated.

5. SOFTWARE EVALUATION AND TESTING

The Loads Tool was verified to ensure that all of the algorithms were properly represented in computer codes so that it estimates as intended. The prototype was validated using approximately 10 different independent data sets collected from Queensland catchments and the "GUMLEAF" (Tan *et al.* 2005) spreadsheet version of loads estimation software developed by Melbourne University. The Loads Tool was verified to ensure that all of the algorithms were properly represented in computer codes so that it estimates as intended. The prototype was validated using approximately 10 different independent data sets collected from Queensland catchments and the "GUMLEAF" (Tan *et al.* 2005) spreadsheet version of loads estimation software developed by Melbourne University. As an example, estimates of the annual total suspended solids loads using Beale Ratio method for Daintree catchment in Queensland are given in Table 1. Figure 5 shows the comparison of two estimates from GUMLEAF and Loads Tool. Some of the estimates differences in these outputs may be due to rounding off effects from two different tools.

Table 1. Estimated annual loads of total suspended solids using Beale Ratio method, estimates are in tonnes per year.

Year	GUMLEAF	Loads Tool	Difference
1974	40769	40952	-183.1
1976	8407	8261	145.9
1983	4645	4650	-4.8
1984	3866	3856	9.9
1985	7870	7874	-3.3
1986	4989	4990	-0.2
1987	10676	10765	-89.7
1988	3427	3393	34.2
1990	5468	5476	-8.4
1991	18877	18899	-21.7
1992	2506	2515	-8.3
1994	1056	1055	0.6
1997	2908	2911	-2.2
1998	14826	14888	-61.7

The field evaluation was conducted by the Queensland National Action Plan (NAP) regional officers to determine whether the system could achieve a satisfactory level of performance with respect to regional body requirements. Further, a number of workshops were conducted in Queensland to demonstrate the software, in which more than 100 end-users participated. Very positive feedback on the use and benefits of the tool was received.

6. CONCLUSIONS

The Loads Tool will be a valuable planning and operating tool for catchments groups, project leaders and government agencies, for estimating pollutant loads for different time scales. The Loads Tool offers a number of techniques for constituent load estimations in rivers or streams to match available quantity and quality of data. Potential applications for the tool include calculation of load-based guidelines, setting targets and comparison of best management practices or any other management activities for reduction in loads at the end of the valley or river mouths. The feature of event loads estimation and the calculation of event mean concentration (EMC) values will be useful for catchment modellers to parameterise some catchment hydrology models. The Loads Tool will be released at national level under the eWater CRC Integrated Monitoring and Assessment Systems (IMAS) product program.

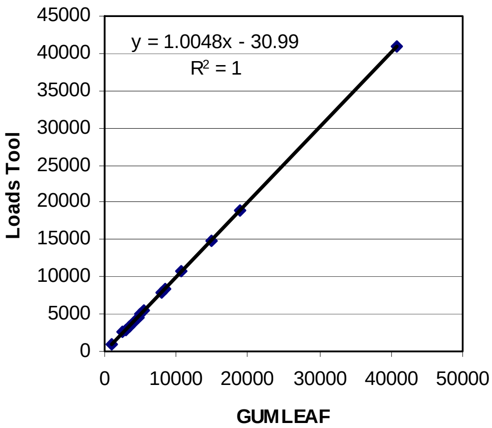

Figure 5. Comparison of GUMLEAF and Loads Tool outputs.

7. ACKNOWLEDGMENTS

The authors would like to acknowledge the contribution from Dr Andy Steven and Brendan Farthing at CSIRO Land and Water and Prof David Fox at the University of Melbourne. The Loads software was developed under National Action Plan for Salinity and Water Quality Program.

8. REFERENCES

Fox, D.R. (2004). Statistical Considerations for the Modelling and Analysis of Flows and Loads - Components of Load. Technical Report 02/04, Australian Centre for Environmetrics, Univ. of Melbourne, Victoria.

Fox, D.R. (2005). Protocols for the Optimum Measurements and Estimation of Nutrient Loads - Error Approximations, Technical Report 03/05, Australian Centre for Environmetrics, Univ. of Melbourne. Victoria.

Huai-en LI, Joseph Hun-Weil LEE and Ming CAI (2003). Nutrient Load Estimation Methods for Rivers. *International Journal of Sediment Research*, 18(4), pp 346-351.

Letcher, R.A., A.J.Akeman, W.S.Merritt, L.J. McKee, B.D. Eyre, and B. Baginska, (1999). Review of Techniques to Estimate Catchment Exports. NSW EPA Technical Report 99/73.

Preston, S.D., V.J, Bierman, and S.E. Silliman, (1989). An Evaluation of Methods for the Estimation of Tributary Mass Loads. *Water Resources Research*, 25 (6) 1379-1389.

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

Tan, K. S., T. Etchells, and D.R. Fox, (2005). GUMLEAF v0.1 alpha: Generator for Uncertainty Measures and Load Estimates using Alternative Formulae, Australian Centre for Environmentric, Univ. of Melbourne. Victoria.

Water Measurement Manual (1997). A water resources technical publication, U.S. Department of the interior bureau of reclamation. US Government printing office, Washington DC 20402.