SAMPLING MASS LOADS IN RIVERS

A review of approaches for identifying, evaluating and minimising estimation errors

Water and Rivers Commission
SAMPLING MASS LOADS IN RIVERS

A review of approaches for identifying, evaluating and minimising estimation errors

Prepared by
B P Degens and R D Donohue
Aquatic Science Branch
Resource Science Division
Water and Rivers Commission
Acknowledgments

The authors (Brad Degens and Rob Donohue) would like to acknowledge the sincere and continuing efforts of the staff in the Aquatic Science Branch (WRC), South Coast Regional Office (WRC) and Greenbase Environment Information Centre for data collection, management and advice. Specific thanks must be extended to Sam Nelson, (Greenbase Consulting), Brad Jakowyna (WRC), Tom Rose (WRC) and Malcolm Robb (WRC) for their valuable editorial comments and advice.

Recommended Reference


We welcome your feedback

A publication feedback form can be found at the back of this publication, or online at www.wrc.wa.gov.au/public/feedback/

ISSN 1327-8436

Printed on recycled stock

2002

Cover Photograph:
Ellen Brook Gauging Station (616189) during high flow
Photograph by Rob Donohue
# Contents

Acknowledgments................................................................................................................... ii

Recommended Reference........................................................................................................ ii

Contents .................................................................................................................................... iii

Summary .................................................................................................................................. 1

1 Introduction ............................................................................................................................ 2

2 Material fluxes in rivers, sampling and errors ................................................................. 4
   2.1 Sampling theory and components of mass load errors ......................................................... 4
   2.2 What use is knowledge of errors associated with load estimations? ................................. 5
   2.3 Where do errors occur during measurement? ..................................................................... 6
      2.3.1 Discharge measurement errors ...................................................................................... 6
      2.3.2 Collection/sampling errors ........................................................................................... 7
      2.3.3 Transport, storage, analysis and calculation errors ......................................................... 8
   2.4 Characteristics of fluxes from catchments ........................................................................ 8
      2.4.1 Influence of catchment properties .................................................................................. 8
      2.4.2 Flux patterns in discharge from catchments ................................................................. 10

3 Strategies that have been developed to sample mass loads ......................................... 13
   3.1 Sampling methodology .................................................................................................... 13
      3.1.1 Discrete sampling ......................................................................................................... 13
      3.1.2 Composite sampling .................................................................................................... 13
      3.1.3 Automated water sampling ........................................................................................ 14
   3.2 Temporal sampling patterns ............................................................................................ 15
      3.2.1 Fixed-interval sampling ............................................................................................... 15
      3.2.2 Stratified sampling ...................................................................................................... 16
      3.2.3 Time-stratified sampling ............................................................................................. 17
      3.2.4 Flow-stratified sampling ............................................................................................. 18
      3.2.5 Continuously updated flow-stratified sampling .......................................................... 19
      3.2.6 Automated probability sampling ................................................................................ 20

4 Mass load calculation methodologies and associated errors ....................................... 22
   4.1 Interpolation approaches ................................................................................................. 22
      4.1.1 Linear interpolation ...................................................................................................... 23
      4.1.2 Averaging estimators .................................................................................................. 23
      4.1.3 Ratio estimators ......................................................................................................... 24
   4.2 Extrapolation approaches ............................................................................................... 25
      4.2.1 Bias correction factors ................................................................................................ 26
      4.2.2 Transfer functions ..................................................................................................... 27
5 Estimation of mass load errors using simulation methods .............................................. 28
  5.1 Errors arising from interpolation and ratio algorithms ................................................... 28
  5.2 Errors introduced by the use of extrapolation algorithms (rating curves) ......................... 30

6 Conclusions ................................................................................................................. 33

7 References and Recommended Reading ........................................................................ 37

Glossary ......................................................................................................................... 42

Publication feedback form .............................................................................................. 43
Summary

A review of studies of mass load errors was undertaken to evaluate the methods used to estimate mass loads in rivers, approaches taken to determine magnitude of errors associated with manual and automated sampling approaches and the main sources of errors occurring during the estimation process. The main principles of relevance for mass load estimation in Western Australia can be summarised as:

- Understanding the pathways of constituent transport and the timing of these in influencing fluxes from catchments is fundamental for designing and managing monitoring programs to estimate the loads of specific nutrients or contaminants.

- Manipulation of sampling frequency can have immediate implications for the accuracy and precision of load estimates.

- Acceptable bias and accuracy for all nutrients and contaminant loads can rarely be achieved.

- Consistent year-to-year precision and accuracy is generally not obtainable by conventional grab sampling strategies.

- Load calculation algorithms can magnify errors if applied inappropriately and can not compensate for poorly collected data.

- Control of errors in sampling programs is fundamental to meet information objectives.

Measuring mass loads transported by rivers and drains has been one of the objectives of most water quality monitoring programs in southern Western Australia. The relative ease with which estimates of mass loads can be derived from flow and water quality data, without reference to errors, can deceive managers and rivers scientists into believing that the measurements are consistently reliable. A major part of managing water quality monitoring programs is ensuring that the reliability of the collected data sufficiently meets information needs. Managing programs to measure mass loads with consideration of the principles outlined above will achieve this broad objective by improving the reliability and usefulness of the final estimates.

The Water and Rivers Commission has developed a load measurement system that overcomes the problems of uncontrolled bias and precision errors highlighted in this report. This system has been successfully implemented on several rivers in southern WA to enable measurement of mass loads with known precision and accuracy. Each measurement unit consists of an automated water sampler controlled by data logger using a stage-initiated program to determine sampling. The parameters for the program are pre-determined using a statistical modelling package (PlaNet). This package enables simulation of different sampling strategies (obtained by manipulating the logger program parameters) and the effects of these on the accuracy and precision of load measurements are evaluated. Details and results of this approach will be provided in subsequent reports.
1 Introduction

The measurement of the mass of materials carried by rivers to receiving waters such as wetlands, estuaries and oceans has been the subject of considerable measurement and research efforts. The primary intention of these efforts has been to provide reliable information for catchment and estuarine management (Littlewood, 1992). Constituents of water quality attracting most interest are commonly nutrients such as nitrogen (N) and phosphorus (P), and suspended sediments (TSS) (Cohn et al., 1992; Littlewood, 1992). However, base cations (Walling and Webb, 1988), salt (Schofield and Ruprecht, 1989), heavy metals and organic compounds (Preston et al., 1989; Richards and Holloway, 1987) have also received scientific and management attention.

In Western Australia, there are a range of statutory management needs that drive demand for estimates of (usually annual) mass loads. Load estimates have been determined for rivers, streams and drains in the Peel-Harvey catchment to meet ministerial conditions so that evaluation of water quality trends attributable to catchment management initiatives can be determined (Kinhill Engineers, 1988; McComb and Humphries, 1992). Until recently, annual mass loads for rivers draining into Oyster and Princess Royal Harbours were also needed to meet Department of Environmental Protection reporting requirements, which are based on evaluating trends in nutrient loads to the water-bodies (EPA, 1990). There is a need to estimate loads to initiate and calibrate estuarine management and ecological models in Wilson Inlet (National Eutrophication Management Program; Cudleigh and Simpson, 2000) and for the Swan-Canning Estuary for the Swan-Canning Cleanup Program (Swan River Trust, 1999). Likewise, estimates of loads delivered from the Swan-Canning catchment are also used to validate and initiate models of catchment nutrient export (Kelsey, 2001). On sub-catchment scales, estimates of nutrient loads (input and output) are necessary to evaluate the performance of constructed wetlands (Rasin et al., 1997) or to evaluate the impacts of land management practice on discharges to receiving waters (for example, assessing the environmental impact of the Ord River Irrigation Area on the lower Ord River and Estuary; H.G Gardiner and Associates et al., 2000). Commonwealth initiatives such as National Land Water Resources Audit and National Pollutant Inventory for broad scale assessment and classification of impacts on receiving water bodies.

Estimation of mass loads transported from catchments by rivers fundamentally involves quantifying the total flux of water quality constituents transported by rivers past a point, usually at the bottom of the catchment, over a fixed time period (Littlewood, 1992). Ideally, this is obtained by summing the instantaneous flux of water quality constituents determined at infinitely small time intervals across the period of interest. Flux of constituents, in this context, consists of the rate of mass transported past a fixed point and is generally approximated by the product of instantaneous flow multiplied by concentration of the water quality constituents transported in the flowing water. The common approaches to estimating total flux, however, involve infrequent estimates of concentrations and frequent estimates of flow from automated gauging stations. These less than ideal conditions have led to the development of a variety of sampling schemes and computational algorithms to estimate mass loads, including averaging methods, ratio-based methods or regression estimators (Yaksich and Verhoff, 1983; Preston et al., 1989; Burn, 1990; Littlewood, 1992).
Understanding the magnitude and sources of errors associated with estimating mass loads is fundamental to understanding the limitations of the estimates in meeting particular information requirements (Ellis, 1989; Ward et al., 1990). Furthermore, understanding the implications of infrequent or inappropriate sampling of water quality in rivers and drains enables identification of management options to minimise the errors to within acceptable limits (Thomas, 1985; Littlewood, 1992). Good sampling design relies on knowledge of key transport pathways to predict the timing of mass transport from catchments, particularly in relation to hydrologic discharge. Despite significant efforts to quantify mass loads in WA rivers, the errors associated with these estimates remains uncertain. It is difficult to use mass load information without knowing the reliability of the estimates. The hidden nature of the estimation errors means that these are often ignored, to the detriment of any conclusions based on this information.

In this review we evaluate methods that have been used to derive estimates of mass load in rivers. We also describe the various methods used to evaluate the accuracy and precision errors of mass load estimates associated with various sampling and load calculation approaches. Practical limitations imposed by manual sampling of loads and the physical impossibility of sampling the entire flux passing a point over a period has resulted in the development of a range of sampling strategies and calculation methods to estimate mass loads. The errors arising from these approaches remain hidden, but must be known to effectively use the information for management decisions. Understanding the sources and potential magnitude of errors associated with load estimation strategies greatly assists management of these problems and improves understanding of the extent to which load information can be used to meet different information objectives. This review has an emphasis on the sources and magnitude of bias and precision errors of mass load measurements. In particular, emphasis is on the extent to which these errors vary between rivers and the level of uncertainty associated with information obtained by less than ideal sampling strategies. Although most examples in the review apply to total nitrogen, phosphorous and suspended sediment loads, the principles are equally relevant to estimation of other mass loads in rivers (salt, heavy metals, pesticides etc.).
2 Material fluxes in rivers, sampling and errors

2.1 Sampling theory and components of mass load errors

To evaluate errors associated with estimates of mass loads it is necessary to consider the influence of errors incurred due to sampling strategies and subsequent calculation methods. Usually the measurement of any environmental properties involves defining the “population” of interest and sampling to obtain the best possible unbiased, precise estimate of the true population. In the context of estimating mass load from catchments, the population of interest is the series of instantaneous chemical fluxes occurring in rivers over the period of measurement. Timing of sampling should essentially be conducted to provide best estimates of these fluxes. Applying this strategy to flows in rivers, however, is complicated by the variability of chemical fluxes from catchments, which arise because of the range of spatially and temporally variable processes influencing fluxes (discussed in detail in Section 2.4). These variations make defining and targeting the “population” of fluxes difficult to achieve before or even during sampling (Littlewood, 1992). Furthermore, patterns of stream chemistry are associated with auto-correlation and independence problems that violate estimating variance and sample size using standard statistical (parametric) approaches (Ellis 1989; Littlewood 1992; Richards and Holloway 1987).

For this review, accuracy is defined as the closeness of the load estimates to the actual load in the river in the period of interest. Inaccuracy is commonly reported as bias (used interchangeably with accuracy in this report), which is the difference between the estimates obtained by sampling in a particular way and the true load (frequently reported as the bias of the median value of estimates from the true median value). Precision is the variation of the measured values relative to the true variation of the population (Fig. 1). It is possible for estimates to be precise, but biased and vice versa, however, precision and bias errors are not necessarily independent. A consequence of less than perfect sampling is that estimates exhibit variation which is composed of that due to the true variation of the population as well as errors due to the act of sampling and measurement, commonly termed systematic errors (Fig. 1). These errors contribute to reducing accuracy or reducing precision relative to the target (or true) population, as illustrated by measurements relative to target or actual values (Fig. 2).

The main cause of precision and bias errors in mass load estimates is due to concentrations sampling poorly corresponding with flux patterns. Failure to conduct sampling according to the temporal variations in flux from catchments results in over-sampling of some of part of the flux and under-sampling other parts of the flux (Meybeck et al., 1996; George et al., 1996). To generalise, bias errors in arise when sampling under or over-estimates the true flux that occurs, whereas precision errors arise when sampling under- or over-estimates the true variation in flux (Donohue and Nelson, 2000). For example, sampling that consistently misses periods of high flux will produce underestimates of mass loads (ie negative bias) and sampling that only occurs during times of greatest flux, but misses the variation in flux leading up to and following these events will produce imprecise estimates of mass loads.
2.2 What use is knowledge of errors associated with load estimations?

All sampling strategies basically aim to control measurement errors so that the information obtained is relevant and reliable for scientific and management needs. These needs determine the accuracy and precision of measurements that are acceptable (Ellis, 1989; Ward et al., 1990). It is frequently impractical to conduct sampling to completely minimise systematic errors in estimates of mass loads. However, sampling can be managed to produce estimates with known and acceptable levels of bias and precision (Cohn et al., 1989). Knowledge of the errors associated with less optimal, often more practical, sampling strategies can greatly assist managers to conduct minimum cost monitoring programs that still provide useful information.

The requirement for estimates of nutrient loads in Western Australia is driven by a range of objectives including nutrient budgets of catchments or estuaries, evaluation of management strategies in catchments or identification and modelling of key nutrient transport pathways from catchments (see introduction). Meeting these requirements may demand mass load information with differing degrees of accuracy and precision.

Accuracy (minimal bias) in load estimates is important for long-term estimates of chemical loading into lakes, wetlands, estuaries and coastal zones. Bias errors are additive over time and across catchments, whereas precision errors can cancel out (Cohn et al., 1989). The consequences of bias errors accumulating over time or space are that long-term nutrient balances and export rates for catchments and receiving water-bodies can be substantially over or under-estimated (Cohn et al., 1989).

Precision errors are of interest for evaluation of long-term trends in river loads or whether loading targets have been met, particularly where bias errors are consistent from year to year. This is critical where the rate of change in mass loading is the information objective to evaluate trends in loads due to land or stream management initiatives. If the precision of the estimates varies widely between years or
catchments it becomes uncertain whether observed trends reflect the desired improvement in environmental conditions or imprecision in the estimates (Cohn et al., 1989). Similarly, precision errors in mass load estimates can also compromise the capacity to detect trends or determine whether targets are met if this error varies from year to year (Millard, 1996).

In some situations knowledge of both precision and bias errors is of importance, particularly when using mass loading information to calibrate catchment or estuarine models (Cohn et al., 1989; de Wit and Pebesma, 2001). The long-term predictive power of these models can be compromised where limited loading information relative to the time-scale of predictions determined using the model.

An important consideration in designing any monitoring program where mass load information is considered a useful indicator of changes is that the cost and reliability of estimates obtained by load estimation should be compared against alternative strategies to meet information objectives. These can include detection of trends at fixed points or snap-shots of water-quality across catchments. The relative costs of each strategy must be evaluated against the management objectives that the resulting water quality data are required to evaluate. Monitoring of mass loads is unlikely to meet all information requirements.

2.3 Where do errors occur during measurement?

Errors in load estimates can occur at all stages from sampling to the final calculation of results. As outlined in the introduction, mass loads are broadly approximated as the sum of instantaneous fluxes, which are the products of concentration and flow at points in time. Errors can arise during estimation of both concentration and flow. These errors are additive in that the errors occurring during each stage of measurement contribute to the overall error of the final estimate. Errors can accumulate during measurement as a result of:

- discharge measurement,
- collection/sampling,
- storage and transport,
- laboratory analysis,
- transcription/recording, and
- calculations.

2.3.1 Discharge measurement errors

The role of discharge errors in the estimation of loads is often ignored. Errors in discharge derive from type of control, frequency of measurement, hydraulic jump, dam backwater, tidal and river confluence transitions, equipment measurement error, completeness of available ratings, measurements and calculation methodology for integrating the velocity cross section, and in the derivation of the rating curve. These and other factors relevant to the measurement of flow are discussed by Chow (1959) and Bos (1989).
Discharge error can also vary over the course of long-term river flow records due to problems such as deterioration in weir structures, sedimentation or deepening of still ponds (Fenton and Keller, 2001). Dean and Marks (1995) have published a generic approach for estimating errors in discharge. Additional sources of flow errors can occur when extrapolating flow from gauging stations either upstream or downstream to estimate flow at the site where water quality measurements have been made.

2.3.2 Collection/sampling errors

Sampling error is often the largest source of error and is the focus of the current review. Errors in estimating nutrient concentrations contain both spatial and temporal components. Spatial errors commonly arise in estimating the flux of nutrients in the cross-section of a river using a single water sample at a point in time (Fig. 3). Individual samples are intended to provide an estimate of the average concentration in the river at the time of sampling. However, the precision and bias of nutrient concentrations in samples taken at a particular point on the surface of the river (even if taken at exactly the same geo-referenced point each time) is likely vary depending on stage height. The profile of nutrient concentrations in rivers varies both with depth and in a transverse plane across rivers (Martin et al., 1992; Meybeck et al., 1996; Hicks and Duncan, 1997). The non-uniform nature of river channels frequently results in different transverse concentration patterns depending on stage height. This arises because of changes in upstream flow patterns changes resulting in changes in mixing regimes, cross-sectional flow velocity profiles and, consequently, the cross-sectional profiles of entrained sediment (Horowitz et al., 1990; Meybeck et al., 1996). The variations in catchment properties and fluxes of constituents from source areas in catchments within and between storms will also contribute to variations in the transverse patterns of water quality constituents. Furthermore, single point sampling below the surface in the middle of streams can fail to capture nutrient transport associated with surface materials or bed loads (Fig. 3).
From the perspective of estimating mass loads, spatial sampling errors can also occur when the water quality-monitoring site is not providing a representative estimate of the mass load for the whole catchment of interest. This commonly occurs when monitoring sites are located at flow gauging stations that are located upstream of the river mouth and effectively miss some of the nutrient sources in the lower part of the catchment. The extent of this error depends upon the nature of the catchment that has been missed. Naturally, where the un-sampled catchment area has more intensive land uses compared with the monitored part of the catchment or different transport-pathways, measured loads may significantly underestimate the true loads. This has occurred in assessment of nutrient loading to the Peel-Harvey Estuary, where sampling one of the major catchments (the Serpentine River) was conducted at a point above the areas of the catchment where substantial amounts of nutrient transport are likely to have occurred (Peel Harvey Management Group, 1998).

Temporal error occurs when sampling to estimate changes in concentrations over time. Much of this error can be due to the failure of interval estimates of nutrient concentrations to reflect the true pattern of nutrient concentrations occurring in rivers. These errors can be further compounded by use of inappropriate calculation methodologies. Load calculation methods produce estimates by making inferences about the likely variation in concentrations between sampling times that ranging from none to highly flow-related changes in concentrations. Where the assumed changes implicit within these mathematical functions do not match the actual patterns, errors can be introduced to the final load estimate.

2.3.3 Transport, storage, analysis and calculation errors

Although not discussed in great detail in this review, other often overlooked sources of error should not be discounted when in considering load estimation errors. The magnitude of errors occurring due to sample deterioration during storage and transport, laboratory analytical measurement errors, transcription of flow and concentration data and calculation mistakes are commonly thought to be small, but can be up to 5% in some instances (Ellis, 1989; Keith, 1996).

2.4 Characteristics of fluxes from catchments

2.4.1 Influence of catchment properties

Patterns of any ion, nutrient or pollutant fluxes from catchments are dominated by the relative rates of transport by various pathways in catchments during and between rainfall events. Detailed understanding of these pathways is the first step in determining the best times to sample rivers to ensure reliable estimates of mass loads. Transport pathways through Western Australian catchments can be generalised as either surface flow across soils, lateral sub-surface pathways through the unsaturated zone of soil profiles, or vertical movement through soils to ground-water which generally discharges at low points in the landscape (Gerritse and Schofield, 1989; Turner et al., 1990). Differences between land-form, soil chemical reactivity and structure/permeability and processes of nutrient release influence the timing and magnitude of fluxes occurring by each pathway through effects on hydrological flow and movement of analytes to river channels (Walling and Webb, 1986; Kendall et al., 1995; Buttle 1998). Shifts in fluxes through these pathways during storms and between rainfall events cumulatively produce characteristic temporal patterns in chemical transport from catchments. Spatial variations in the occurrence of these pathways, due to variations in rainfall, soils and land-uses across catchments can also have a significant influence on flux patterns over time (Walling and Webb, 1986).
Catchment hydrology, land-form, soil chemical reactivity and importance of point and non-point (diffuse) sources influence flux patterns in rivers (Walling and Webb, 1986; Littlewood, 1992). When diffuse sources dominate, patterns of flux are controlled by release of leachable forms of nutrients/water quality constituents from landscape stores, soil reactivity and transport pathways in catchments (Walling and Webb, 1989; Kendall, et. al., 1995; Buttle 1998). Sampling to estimate mass loads from diffuse sources can greatly benefit from understanding how differences between the properties of catchments influence the timing, magnitude and frequency of mass loads in rivers. This is particularly important for programs where operational realities and budget constraints severely limit the frequency of sampling. For water quality constituents arising from point source discharges to streams/rivers, the effects of these on flux at the bottom of catchments generally depends on the size and timing of discharge from these sources, rather than the activation of various hydrological transport pathways. Increases in discharge during storms will generally dilute in-stream concentrations from distinct point sources discharges. The exception occurs for point sources characterised by nutrient sources seasonally isolated from surface- and ground-water such as land-based effluent treatment systems and intensive land-uses (eg. horticultural activities). The effects of these point sources on catchment nutrient fluxes can be more dependent on the occurrence of water flow pathways resulting in connection of the sources with waterways. In WA, point sources can include sewage treatment plants, industrial discharge points, effluent from piggeries dairies and feedlots and un-sewered urban developments.

Factors such as relief, soil permeability and vegetation cover influence discharge and can similarly influence chemical fluxes in rivers. With smaller or high relief catchments, flow and constituent flux generally exhibits greater peaks and flows per unit area, subsiding to lower fluxes and flow per unit area between storms (Fig. 4; Richards and Holloway, 1987). Streams draining small catchments generally tend to respond rapidly to rainfall (that is flashy), compared with streams draining larger catchments (Walling, 1977; Richards and Holloway, 1987; Richards and Baker, 1993). This occurs because in smaller catchments there is less time for in-stream mixing, dilution, retention or re-suspension of constituents during flows, which tend to moderate peak concentrations in rivers draining large catchments (Richards and Holloway, 1987). Furthermore, merging of discharge peaks often occurs in larger catchments, which masks individual transport events occurring in different parts of catchments (Richards and Holloway, 1987; Richards and Baker, 1993). Such straightforward interrogation of basic catchment properties, however, can misrepresent the complex underlying processes involved in chemical/material transport through catchments to waterways (discussed below).

The chemical form of nutrient/pollutant stores in catchments frequently determines rates of supply from diffuse sources. In the case of nutrients, diffuse sources can consist of broad-scale slow-release sources of nutrients from soils as well as releases of nutrients attributable to seasonal additions to soils across large areas (eg fertilizers or animal wastes). This was exemplified in studies of the Peel-Harvey, where fluxes of phosphorus were attributable to seasonal fertiliser applications as well as release of phosphorus from...
slow- and fast-dissolving stores in soils (Schofield et al., 1985; Ritchie and Weaver, 1993). As a consequence, the contributions of these sources to catchment fluxes can vary depending on the rate, frequency and spatial distribution (due to rainfall) of water flows through landscapes. Exhaustion of supplies of materials able to be transported from catchments can occur during storms and across seasons if rates of hydrological flux exceed rates of supply from stores in soils. Examples of this include reduced transport of salt from saline seeps (dilution by seasonal ground-water flux), declines in phosphorous transport as soil stores of soluble phosphorus are seasonally depleted (Ritchie and Weaver, 1993) and reduced nitrate transport as a result leaching exceeding nitrate production (Turner et al., 1990).

The reactivity of different pollutant/nutrient forms has a major influence on the flux through each pathway (Gerritse and Scofield 1989, Heathwaite et al., 1996; Richards and Baker, 1993). Transport of chemicals by each of the generalised hydrological pathways outlined above is determined by the relative reactivity of the chemicals in soils and sub-soils through which water flows and subsequent transformations that may alter the transport rates. These interactions can slow or deplete mass transport by various pathways, thus modifying which of the pathways through which most flux occurs.

Transport of non-reactive chemicals (such as nitrate and salts) can occur through both surface and sub-surface or ground-water pathways depending on the dominant flow pathway (Heathwaite et al., 1996; Richards and Baker, 1993). Transport of reactive chemicals, such as phosphorus, heavy metals or pesticides, by shallow sub-surface and deep ground-water water flows, however, is generally minimal in catchments where soils are highly reactive (contain high clay and iron/aluminium oxide contents; Richards and Baker, 1993; Kirkby et al., 1997; Stevens et al., 1999). Similarly, redox and iron availability can also minimise sub-surface and ground-water transport of reactive chemicals (Gerritse et al., 1990). In these situations, surface transport pathways are likely to dominate fluxes of reactive chemicals and generally involve movement of more particulate and soil-bound forms, rather than water-soluble forms. An additional influencing factor for the transport of some chemicals is the rate of transformations occurring during transport by different pathways. This mainly applies to nitrogen where transport via sub-surface and ground-water pathways can be regulated by microbial oxidation of ammonia to nitrate (ammonification), retention of nitrogen by terrestrial or aquatic organisms or loss to the atmosphere through denitrification processes (Kendall, et al., 1995). Similar transformation processes can modify pathways of pesticide movement where degradation occurring during movement through various sub-surface pathways can influence the importance these pathways in transport from catchments (Leonard et al., 1987).

### 2.4.2 Flux patterns in discharge from catchments

The combined effects of chemical fluxes moving through each transport pathway during and between rainfall events produces characteristic temporal flux patterns from catchments (Walling and Webb, 1986; Richards and Baker, 1993; Evans and Davis 1998). These patterns are characterised as peak fluxes occurring either before, at the same time or after peak discharge during storm events with combinations of these resulting in multiple or merged peaks (Richards and Baker, 1993). The timing of the dominant flux pathways during storms can produce characteristic cyclical shapes in the relationships between discharge and concentration (Q-C relationships; Evans and Davis 1998; Johnson and East 1982; Sklash and Farvolden 1979; Gillham 1984). For example, in a single storm occurring in the Ellen Brook catchment, the flux of TN and TP in discharge water is delayed relative to peak discharge resulting in a positive, but anti-clockwise hysteresis response in the Q-C relationship (Fig. 5).

Both the shape (ovate, concave, convex) and the direction (clockwise/anticlockwise) of the hysteresis pattern can be diagnostic of the flow pathways dominating during individual storm events (Evans and Davis 1998; Johnson and East 1982). These changes reflect changes in the quality of water from
discharges of ground-water, antecedent soil water and surface event water (run-off) during storms (Evans and Davies 1998). For example, where concentrations of chemicals in surface water are greater than in sub-surface pathways, concentrations will increase with increasing flow (Littlewood 1992), resulting in peak flux occurring early during storms with a positive, clockwise Q-C response (Williams, 1989; Evans and Davis 1998). However, if concentrations of chemicals in ground-water discharges are greatest (often dominating between storms), surface run-off during storms will dilute in-stream concentrations from ground-water and produce negative, often anticlockwise Q-C responses (Evans and Davis 1998).

Both the shape and direction of the hysteresis pattern can change from storm to storm and across seasons (Walling and Webb, 1986). This occurs particularly when the shape (and the direction) of the hysteresis loop reflects changes in the size of the readily dissolvable soil store during high rainfall periods. As the readily soluble store is exhausted the relative concentrations in surface run-off, soil water (sub-surface) and ground water change (Johnson and East, 1982; Evans and Davies, 1998; Walling and Webb, 1986). Increases in ground-water levels during winter and spring can also influence hysteresis patterns across seasons as the sub-surface and ground-water contributions to flows increases. Consequently, although hysteresis patterns can be generalised, the Q-C relationship may not follow exactly the same pattern between storms, seasons or flow periods, even in the same river (Walling and Webb, 1986). Variations in hysteresis from storm to storm can also be related to the pattern of rainfall occurring in different source areas. This is particularly evident in large catchments such as the Avon River catchment. For example, during the January 2000 flood event, there were source areas in the Lockhart catchment activated by
summer rainfall that had not been contributing to fluxes from the catchment for more than 20 years (Muirden, 200). These were likely to have influenced the flux pattern from the catchment during this event. Regardless of the nature of the Q-C relationship, once the characteristic patterns are known these can be exploited to maximise the effectiveness of sampling water quality to estimate nutrient loads (Thomas, 1985; Littlewood, 1992).

Variation in the Q-C relationship between storms often results in the failure to detect generalised hysteresis patterns over seasons, although the direction of the dominant responses (negative, positive or no response) is still evident. Interpretation of these generalised seasonal responses (the only aspect of hysteresis captured using fixed-interval sampling) is limited to the general water quality of the dominant pathways between events as compared to during events. For example, a positive Q-C response in total phosphorus concentrations in Sleeman Creek (Fig. 6a) indicates storm-activated phosphorus pathways, whereas in Mills St MD, the negative Q-C response indicated that storm flows activated transport pathways were associated with better water quality than those occurring between events (Fig. 6b).

It is important to note that not all substances exhibit a dependence on flow or cyclic patterns of change in concentration (Whitfield and Whitley, 1986; Ellis, 1989). For example, variation in the concentration of ammonia (Ellis, 1989) and some base cations (Preston et al., 1989) are can be independent of changes in stream discharge. In these situations, there is little benefit to sampling based on patterns of hydrological discharge.
3 Strategies that have been developed to sample mass loads

Practical limitations are a major factor constraining effective sampling of concentrations of water quality constituents in rivers to estimate mass loads. Water quality sampling is commonly conducted using fixed-interval manual sampling techniques because of easy implementation and low operation costs. However, these approaches may not reliably measure the temporal patterns in concentrations of chemicals in rivers needed to reasonably estimate mass loads. In practice, many of the difficulties in producing accurate and precise loading estimates can only be solved either by very frequent continuous sampling of rivers or by increasing flexibility in sampling by allocating more of the available sampling budget to periods of greatest flux (stratified sampling). This is most practicably achieved using automated water sampling, although is not impossible with manual sampling. The following sections describe the variety of sampling methods and patterns that have been used to sample mass loads in rivers and drains.

3.1 Sampling methodology

Samples can be collected as either discrete or composite samples (time- or flow-composite) by manual or automated means. Discrete samples provide information on the water quality occurring in a river cross-section at a point in time, whereas composite samples can provide aggregated information about the average water quality occurring over time or flow periods. These sampling strategies are described in this section, ending with a discussion of automated sampling.

3.1.1 Discrete sampling

Discrete samples commonly involve “grab” sampling where a sample of water is obtained from the fastest flowing point in a river, generally in the centre of the flow, just below the surface. Automated water samplers can also be used to obtain discrete samples, however, reliance on fixed-intakes can mean that samples are not be consistently taken from the fastest flowing point of rivers or drains. The positioning of intake ports in relation to the main cross-sectional flow can result in biases in sample concentrations (Newburn, 1996). Positioning too deep or too shallow can result in imprecision that varies with stage height. These problems are greatest when samples are taken from a point where bed-load or bank turbulence can influence constituent concentrations and the effect varies depending on flow (see discussion in section 2.5).

3.1.2 Composite sampling

Composite sampling involves combining samples taken at either fixed time or fixed flow intervals to provide a series (or even only one) of representative measures of the water quality occurring in river over time. Time-composite sampling involves sampling at fixed intervals over a period to provide a single sample that represents the nutrient concentrations occurring in the period. The nutrient concentrations of the final composite samples are a time-weighted average over the period of sampling. Although reducing the number of analyses and costs of estimating loads (Shih et al., 1994; Izuno et al., 1998), this strategy should only be undertaken when the variation of concentrations occurs over periods of times greater than the sampling interval. Where this does not occur time-composite sampling may result in biasing of measurements towards the conditions that prevail for most of the time (Ellis, 1989). This strategy has
been used with many auto-samplers in WA where, for example, 250 mL samples were collected every 8 hours and combined into a single 750-L sample that is representative of quality in a 24 hour period (Peel Harvey catchment, Gerritse and Schofield, 1989). Analysis costs for such sampling may not be minimised since one sample per day is collected irrespective of whether any changes in flux have occurred.

Flow-composite sampling differs from time-composite sampling by combining samples taken on the basis of flow volumes rather than time intervals. This sampling is often undertaken using automated water samplers to combine samples taken at specified amounts of flow occurring. Samples can be combined to represent small or large flow volumes depending on the extent of expected changes in constituent concentrations with flow volumes. The final samples contain concentrations of water quality constituents weighted according to the volume of flows that have occurred (that is with more of the concentration made up of samples from high flows than low flows).

Composite samples can reduce analysis costs for rivers where frequent flow events occur, while theoretically providing an unbiased measure of overall load (Shih et al., 1994; Nicholson and Clark, 1994). A significant disadvantage of the approach, however, is that the water quality information is only useful for mass load estimation and results in a loss of information about the exact temporal nature of water quality required for process studies or trend analyses (Ellis, 1989).

3.1.3 Automated water sampling

Automated water sampling provides flexibility in sampling times that can overcome biases and errors introduced by the more inflexible manual sampling strategies. Although dependent on initially expensive sampling equipment and accurate flow measurements, auto-samplers can provide significant improvements in accuracy and precision of water quality sampling without needing to increase sample analyses or personnel time involved in sampling. This technology, however, is not completely devoid of systematic sampling errors and only reduces temporal sampling errors (which is often the largest component of systematic errors).

Auto-samplers can enable sampling to be controlled by programmable loggers using information on catchment and river conditions to determine sampling. Sampling can be triggered depending on rainfall in catchments (measured by tipping bucket rain gauges) or turbidity, pH, electrical conductivity, or even specific ions in rivers directly measured by probes. Such instruments, however, can require rigorous servicing to ensure reliable functioning (Lewis, 1996; Ross-Rakesh et al., 1999). It is important to acknowledge, however, that automated sampling will only dramatically improve estimation of loads when mass transport fluxes are storm driven and associated with large variation from storm to storm.

Despite the sampling control afforded by auto-samplers, problems can arise due to location and maintenance of intake ports and deterioration of sample chemistry during storage. As discussed above, inappropriate location of fixed intake ports for auto-samplers risks introduction of additional biases to water quality measurements (Fig 7). Detection
of these problems is possible by comparisons of water quality in grab samples with that of samples obtained by an automated sampler for a range of flow events. Additional problems can also occur when intakes are exposed to sunlight and become overgrown with algae (Fig. 8) or become entrapped with debris during storm events, thus resulting in contaminated samples.

On-site storage of samples collected by auto-samplers for long periods of time before analysis (over a month in some cases) can result in increased errors associated with measurement and analysis (Nicholson and Clark, 1994; Kotlash and Chessman, 1998). There might be substantial changes in the nutrient forms as a result of storage at field temperatures before analysis (especially transformation of dissolved inorganic and organic nutrients). However, for total suspended sediments and total phosphorus, this is generally minimal (Kotlash and Chessman, 1998). On-site preservation of samples using refrigeration can improve sample preservation, but the effectiveness might vary between rivers (Kotlash and Chessman, 1998).

The potential problems that can occur with auto-samplers highlight that correct installation and strict maintenance is critical to the operation of the units to ensure collection of highly reliable water quality information.

3.2 Temporal sampling patterns

Numerous sampling strategies have been developed to measure mass loads in rivers and drains ranging from simple fixed-interval or storm-based stratified sampling to elaborate sampling protocols incorporating stratification and probabilistic elements using programmable automated sampling units. All strategies were initially developed for quantifying sediment loads, but are equally applicable to mass loads of other nutrients and water quality constituents.

3.2.1 Fixed-interval sampling

The simplest sampling pattern involves fixed-interval sampling with no consideration of flow (also termed systematic sampling). Sampling frequencies normally range from daily to monthly, with the longer intervals being increasingly recognised as being suitable only for instantaneous evaluation of river conditions (Ward et. al., 1990; Ellis, 1989). Fixed-interval sampling is a common feature of many water quality monitoring programs in WA.
3.2.2 Stratified sampling

Stratification of sampling involves allocating a higher proportion of the total sampling effort to periods of highest flux. This does not necessarily result in sampling occurring exactly in phase with hydrological discharge, but can mean that sampling effort is greater earlier or later during storms (depending on when constituent flux occurs in relation to discharge; see Q-C relationships discussed earlier). As for all sampling strategies, load estimations are improved by any sampling practices enabling measurement water quality more closely matching the true patterns of water quality than can be obtained by fixed-interval sampling. In practice, stratification is undertaken by dividing measurement periods into different flow events and varying sampling intensity within these according to the expected flux likely to occur during these times. Determining which flows to allocate sampling for best stratification requires an understanding of the timing and magnitude of nutrient transport from catchments, particularly in relation to the hydrologic flux (as discussed earlier).

For nutrients, sediments and sediment-bound contaminants mobilised primarily during storm events, simple stratification may be achieved by dividing the hydrograph into two strata (storms and base-flow events or high-flow and low-flow events). This enables more intensive sampling within peak flow periods than between these events (Littlewood, 1992; Richards and Holloway, 1987). Since greatest variation in concentrations (and therefore flux) occurs within storm events for many water quality constituents, increased stratification of storms has greatest benefit to load estimation (Thomas and Lewis, 1995). Variation in concentration with increasing flow is often caused by variations in discharges via the various transport pathways in catchments. In turn this is related to variation in the occurrence and intensity of rainfall, the rate of infiltration, soil chemistry and the location of source areas for chemicals in catchments.

Identification of flow characteristics in which the concentration of the measured contaminant is highest and most variable is generally the most immediate way to improve monitoring of flux (Richards and Holloway, 1987; Preston et al., 1989; Rekolainen et al., 1991). Obviously, when less storm-activated transport pathways dominate constituent fluxes, such as via ground-water discharge, it is important to sample during flows between storm events. Since flow is a continuous variable there is the temptation to improve predictions of quality by using increasing smaller or complex definitions of strata. For example storm flow quality can be broken down into rising limb and falling limb, and the rising limb into discharge contributions from surface event water, discharges of vadose water and groundwater. However, studies have shown that the largest gains in precision and accuracy for some rivers can be achieved with relatively simple stratification strategies (high-flow vs low-flow) with little further gains achieved by additional optimisation of strata definitions (Richards and Holloway, 1987; Rekolainen et al., 1991; Littlewood, 1992; Thomas and Lewis, 1995). Ultimately, the number of strata classes is frequently constrained by the analyses that can be afforded.

Stratification of sampling does not improve the errors of mass load estimates for all constituents carried by rivers. For some less reactive constituents such as salts, heavy metals and pesticides, stratification according to flows can result in little gain in precision and accuracy (Richards and Holloway, 1987; Preston et al., 1989) because of the less flow dependent behaviour of these nutrients in some catchments. In some streams draining smaller catchments, stratification may also not improve precision if there is considerable variation in water quality between streams (Kronvang and Bruhn, 1996). Furthermore, overtly simple stratification approaches, for example where increased sampling is allocated exclusively to high flows, can result in highly biased and imprecise loading estimates (Rekolainen et al., 1991). In such situations, it is the failure to measure any unexpected high flow events occurring during between sample observations of quality that results in bias.
Nevertheless, more complex stratification of flux from catchments (using flow) potentially provides estimates of mass loads that can be more useful for evaluating the impacts of catchment management as well as better data for models of natural phenomena (Thomas and Lewis, 1995). In many cases, establishing strict definitions of where flow should be stratified is only practical for automated sampling where a programmable logger or computer can be used to achieve greater real-time control over stratification (Thomas and Lewis, 1995). Two such techniques, time-stratified and flow-stratified sampling, are described in the following sections.

### 3.2.3 Time-stratified sampling

Time-stratified sampling programs involve randomly timing samples within a series of time windows, which are intermittently selected during sampling depending on stage heights (Thomas and Lewis, 1993). Sampling is computer controlled using stage height information to intermittently determine which time window is appropriate and what sampling occurs within each window (Fig. 9). The time windows, representing strata, are set before sampling by analysis of previous stage-duration data for a river. Common stage height intervals and directions of change are grouped and allocated time windows in which random sampling occurs (Thomas and Lewis, 1993). The size of the time window (strata duration) is varied to control sampling frequencies depending on the nature of the flows (defined by stage height and direction of change information). Generally, shorter strata are used when stage is high and rising, indicating the likelihood of high flows or rapidly increasing flow. In contrast, longer strata are used when stage is low and declining, since these conditions generally indicate that lower flows or inter-event periods are likely (Thomas and Lewis, 1993). Often, only a single sample will be taken when each of the strata occurs.

![Figure 9: Illustration of time-stratified sampling where stage data are periodically used to set strata duration based on pre-selected strata criteria.](image)

This strategy can provide unbiased estimates of the load and variance of individual nutrients or sediments (Thomas and Lewis, 1993). Selection of the time window for each strata can greatly influence the variation in estimated loads using this strategy (Thomas and Lewis, 1993; Fig. 9). Varying the duration of
sampling intervals in relation to stage characteristics directly influences the probability of sampling chemical fluxes and therefore the subsequent reliability of load estimates.

As with all constituents, setting of strata definitions demands consideration of the expected concentrations of chemicals occurring at different times during storm events. Nutrients and other water quality constituents with delayed fluxes (relative to hydrographic peaks) will require intervals that are shorter during falling limbs of hydrographs. In contrast, nutrients and pollutants that dilute with flow require more evenly spaced intervals. As a consequence, all strata need to be set at short intervals when measuring loads of different chemicals by time-stratified sampling to ensure consistent errors for each load estimate.

3.2.4 Flow-stratified sampling

Flow-stratified sampling is similar to time-stratified sampling except that sampling is within flow-based strata and regulated by pre-defined probability rather than the duration of the strata (Thomas and Lewis, 1995). Real-time stage information is used to continuously determine which pre-defined stratum (based on stage height and direction of change) is occurring at any time (Fig. 10). A probability routine determines whether samples are taken each time a strata occurs, thus enabling variations in sampling to occur in relation to the occurrence of particular river flows. In contrast with time-stratified sampling, flow-stratified sampling varies the numbers of samples taken between flow events and results in the total duration of strata over the whole flow period being unequal (Thomas and Lewis, 1995). This strategy results in sampling being distributed across flows with the frequency varying as flows rise and fall as different strata are recognised and sampling within each is regulated at different frequencies (Fig. 10).

Figure 10: Illustration of flow-stratified sampling strategy where regular updates of stage information are used to dynamically control strata allocation and sampling within strata is dependent on Bernoulli probability outcomes.
Randomised sampling within each flow-stratum is conducted using a routine where, when each strata occurs, a randomly generated value is compared with a pre-set value to determine whether a sample is taken. This approach essentially uses Bernoulli sampling probabilities, such that sampling is determined by pre-set probability values. Altering these values for each stratum modifies the chance that samples will be taken within strata (Thomas and Lewis 1995). For example, increasing the probability will increase the likelihood of sampling while the stratum occurs, so that more samples over the whole year will be taken in this stratum. A problem with this approach, however, is that there can be poor control over sampling individual storms (since not all storms are automatically sampled). There can be excessive sampling in large storms, with no samples taken in some short duration storms (Thomas and Lewis, 1995). Furthermore, the number of samples taken is not fixed, since this varies from flow-period to flow-period depending on the nature of storms (Thomas and Lewis, 1995).

The precision of load estimates using flow-stratified sampling depends upon the number of samples taken and pattern of individual storms (Thomas and Lewis, 1995). Consequently, the criteria used to define strata and the probabilities weighting sampling within strata need to be manipulated to achieve best precision. Over long flow periods, the strategy offers opportunities to minimise load errors by enabling splitting strata associated with large storms. Load errors can sometimes be increased if larger flow events are classified into the same stratum resulting in samples either missing higher points in the flows or oversampling more of the larger flows (Thomas and Lewis, 1995).

### 3.2.5 Continuously updated flow-stratified sampling

This type of automated sampling relies on the use of an algorithm that controls sampling and continuously adjusts the frequency of sampling based on the sampling and flows that have already occurred (Burn, 1990). In essence, the patterns of flows that emerge during a sampling year are repeatedly used as feedback information to regulate the allocation of the remaining samples taken in the year. Unlike the approaches discussed above, this added flexibility enables sampling intensity and timing to vary in according to the nature of flows, thereby avoiding under or over sampling (and exceeding the number of available sample bottles). This strategy is suited to programs where financial constraints limit the maximum number of samples that can be analysed.

The basic outline of the sampling protocol summarised from Burn (1990) is outlined below using the example of 15-minute evaluation/update intervals.

1. Real time measurement of flow is taken across a 15-minute interval and assigned to a stratum according to the current strata boundaries (a continually varying set of limits, see below).

2. The total time the flow stratum has occurred since the stratum was last sampled is determined. This value is compared with the trigger variable, which sets the time between sampling within each strata (constantly updated as flow occurs).

3. If the trigger is equalled or exceeded, a sample is taken otherwise, sampling is delayed for the current stratum.

4. The flow values defining each stratum are revised by evaluating whether the probability of total flow for the year will reach the expected total flow. If there is a significant chance that total flow will be less than expected, strata boundaries are re-allocated to capture smaller changes in flow (i.e. re-allocation of strata boundaries depending on flow progress during the year).
The times between samples within strata are also modified according to changes in strata boundaries to ensure that the expected numbers of samples taken over the flow period are near the target number (limited by the budget at the beginning of the year).

If the strata are changed, the length of time between sampling in each stratum is also adjusted based on estimation of the expected number of samples taken for each stratum. This step also ensures that the total number of samples at the end of the flow period is near the initial target value and that these have been allocated across the strata.

Repeat all steps at 15-minute intervals over which stage (and flow) measurements are reported.

The strategy has had limited evaluation. Potential problems of the strategy might arise when sampling Western Australia rivers where climate variability results in poor predictability of seasonal flows or where seasonal patterns can vary between uni-modal and bi-modal distributions between years (due to cyclonic summer rainfall). These problems will influence allocation of sampling at times during the year based on expected flows for the remainder of a flow year.

### 3.2.6 Automated probability sampling

This sampling strategy does not rely on stratification to divide and distribute samples in discrete units, rather that continuous measurements of flow are used to derive expected flux of specific nutrients or pollutants at any time which then influences the whether samples are taken. The simplest form of probability sampling involves assuming that the probability of taking a sample is equal with time (simple randomised sampling). This is only appropriate for streams where concentrations of compounds are evenly distributed in time, although not necessarily in unvarying amounts (Ellis, 1989). More commonly, sampling probability (and therefore sampling intensity) must be varied with time, since concentrations of water quality constituents can frequently vary with time. As discussed previously, precise and accurate estimates of mass loads are best achieved by regulating sampling effort in line with patterns of mass transport. In statistical terms, this can be described as sampling guided by probability proportional to the size of the flux (or PPS).

PPS involves the probability of a sample being taken at any point in time being weighted by the magnitude of the flux (size of the population) likely to be occurring (Thomas, 1985). Of course, information on the actual amounts of mass transport occurring at a point in time is not often immediately available. However, auxiliary variables such as flow, conductivity or turbidity can be used to predict the amounts of material being transported thereby guiding sampling probability. To apply PPS sampling correctly, it is necessary to determine the total amounts of material transported and assign the probability distribution of the population based on flow \textit{a priori}. It is obvious that this is impractical for measurement of material loads in rivers, however, the principles of the PPS approach were adapted into the strategy entitled ‘Selection at List Time’ or SALT (Thomas, 1985).

The SALT sampling strategy uses flow as an auxiliary variable to predict the likely concentration of water quality constituents and allocation of sampling probabilities as flow occurs, thus achieving dynamic allocation of sampling effort in proportion to mass load in real-time. In the SALT strategy, sampling decisions are conducted at set time intervals where the flow information at the mid-point of each interval is used to generate expected nutrient concentrations using rating curves (Thomas, 1985; Fig. 11). An estimate of the mass load occurring over the interval is calculated from this information and combined with all mass load estimates for previous intervals (Fig. 11). Decisions to sample are based on random allocation of sampling at different instants during accumulation of the expected total mass load likely to occur (which is established before sampling begins). For each interval, if the change in cumulative mass load includes these pre-set instants a sample is taken (see Fig. 11).
The precision, but not bias, of loads estimated by SALT sampling is strongly dependent on the reliability of rating curves to estimate mass loads for intervals (Thomas, 1985). These rating curves are derived before flow events using previous data collected for the river of interest or from rivers with similar catchment characteristics. SALT sampling may not ensure greatest precision in load estimates, particularly since sampling is biased towards periods of greatest loads rather than periods when loads are most variable.

SALT sampling enables inclusion of the sampling probability at the time of sampling, hence the error associated with load estimates can be readily calculated for flow periods. This overcomes the need to collect large numbers of samples to obtain estimates of load errors, as is necessary with other sampling strategies. Although flow is commonly used as an auxiliary variable, other variables may be equally suitable and greatly improve the poor predictions of nutrient concentrations that commonly occur using flow. Such alternative auxiliary variables could include turbidity (Lewis, 1996), pH or EC, but must be available in real-time through continuous analysis to influence the probability of taking samples.

![Diagram showing the protocol for SALT sampling](image)

Figure 11: Illustration of the computer protocol used in Selection at List Time (SALT) sampling where probability-controlled sampling occurs.
4 Mass load calculation methodologies and associated errors

All algorithms used to calculate mass loads \( L \) attempt to provide an approximation of integrated flux across a time period of interest (designated as \( t_1 \) to \( t_2 \)):

\[
L = \int_{t_1}^{t_2} F(t) dt
\]  

(1)

This algorithm is essentially the sum of the instantaneous flux at small intervals \( F(t) \) occurring over a set period (Yaksich and Verhoff, 1983; Littlewood, 1992) where the flux for each sub-interval is given by:

\[
F(t) = \int_{t_1}^{t_2} K Q(t) C(t) dt
\]  

(2)

Where \( Q(t) \) is the time series function describing flow patterns, \( C(t) \) is the time series function describing concentration patterns and \( K \) is a unit conversion factor.

Approximations of this function are achieved by summing the load calculated for small time intervals across the flow period. To overcome the problem of infrequently available concentration data obtained by sampling programs, algorithms have been developed that involve interpolation or extrapolation of collected data to determine concentrations between sampling points. Interpolation approaches involve estimating loads using instantaneous flow and concentration data taken at fixed points, whereas extrapolation approaches combines data collected over representative flow periods to generate relationships from which continuous concentration data is derived for available flow data. The derived concentration data (extrapolated from flow) is combined with collected concentrations to estimate overall load. It is important to note that all approaches implicitly assume that the error of estimating flows are small or is at least consistent at all sampling times and that the point of sample collection is representative of average concentrations.

4.1 Interpolation approaches

Interpolation approaches estimate loads by assuming that the fluxes estimated by sampling at points in time are representative of the fluxes in the un-sampled time period between samples. The more simplistic calculation algorithms essentially use averages of interval flow and concentration data or interval load data to provide estimates of continuous load. A variety of these calculation algorithms have been developed to deal with flow and concentration data available at different frequencies (Yaksich and Verhoff, 1983; Littlewood, 1992). Alternatives to the simplistic averaging approaches incorporate adjustments of mass load estimates based on ratios accounting for flow-dependent changes in concentration between sampling points. In the present review, only a brief account of the various computational approaches is presented. The reader is directed to Yaksich and Verhoff (1983); Littlewood (1992) and Preston et al. (1989) for more detail.
4.1.1 Linear interpolation

When sampling produces nutrient concentration data that represents the true temporal patterns of nutrient concentrations for the sampling interval, linear interpolation can be used to estimate equation (1). This approach is based on determining time-weighted concentration for flow measurements between times when concentrations were taken (Kronvang and Bruhn, 1996). Linear interpolation is widely used to estimate loads in Western Australian rivers (e.g., Forbes and Birch, 1987; Deely et al., 1993; Donohue et al., 1994).

\[
\text{Load} = \sum_{i=1}^{n+1} \sum_{t_i \leq t < t_{i+1}} q_t \frac{C_{t_i} (t_{i+1} - t) + C_{t_{i+1}} (t - t_i)}{t_{i+1} - t_i}
\]  

(3)

Where: \(C_{t_i}\) = concentration measured at time \(t_i\), \(C_{t_{i+1}}\) = concentration measured at the next time, \(t_i\) and \(t_{i+1}\) = are the times at the beginning and end of each interval when concentrations were measured (from times 0 to \(n\)) and \(t\) is the time at any time between \(i\) and \(i+1\) when

concentrations were sampled and flow data was recorded. \(q_t\) is the flow recorded at time \(t\) for intervals between the times \(t_i\) to \(t_{i+1}\), when concentration measurements were taken. In cases where concentration data is infrequent, the formula has the effect of weighting concentration information between sampling times.

HydSys uses a modified version of equation (3) where the flow volumes for time periods are directly integrated from stage heights and rating curve information.

4.1.2 Averaging estimators

An alternative, simpler algorithm based on the linear interpolation principle uses the concentration at the start (or the end) of the interval to estimate flux across intervals. Load is estimated as the sum of loads for each sampled interval weighted by the number of samples taken across the whole period (equation (4)). As with equation (3), this approach performs best with temporally intensive concentration data or where sampling best detects true patterns of concentrations occurring in rivers.

\[
\text{Load} = K \left[ \sum_{i=1}^{n} \frac{C_i Q_i}{n} \right]
\]  

(4)

Where: \(K\) = a conversion factor to account for differences in units and periods of time, \(Q_i\) = flow for interval \(i\), \(C_i\) = concentration for interval \(i\) (usually at the beginning or end of each interval).

Other variations on this algorithm estimate loads from the product of average flow across all intervals and average concentration (equation (5)). Averages can be calculated for daily, weekly, monthly or annual intervals, depending on the frequency of collected data (Littlewood, 1992). Annual load derived from equation (5) (mean concentration x mean flow) can be similar to annual load from equation (4) (mean of a series of concentration x flow products) when concentration is independent of flow (Ellis, 1989). Equation (4) can also be modified to deal with situations where the interval at which concentration data were collected differ from that for which average flow data are available. For example equation (6) calculates load from the product of monthly averages in concentration values and daily flow.
These approaches provide reliable estimates of loads where there is little variation in nutrient concentrations. Evidently, inadequate sampling resulting in too many samples from events between storms or too few samples during the part of flows where variation in flux is greatest can result in under or over-estimates of average concentrations determined within the calculation process. These ultimately carry through as errors in final estimates of mass loads.

\[
\text{Load} = K \left[ \frac{n}{n} \left( \sum_{i=1}^{n} \frac{Q_i}{n} \right) \right] \left( \sum_{i=1}^{n} \frac{C_i}{n} \right) \tag{5}
\]

\[
\text{Load} = \sum_{m=1}^{12} \sum_{j=1}^{N_m} Q_{jm} \left[ \sum_{i=1}^{n_m} \frac{C_{ijm}}{n_m} \right] \tag{6}
\]

For equation (6): \(C_{ijm}\) is monthly average concentration for month \(m\), \(n_m\) is the number of days in month \(m\); and \(Q_{jm}\) is the daily flow for day \(j\) in month \(m\).

Where sampling has been stratified according to flow, load can be calculated for each stratum (product of concentration and flow) then summed across all strata adjusted according to the proportion of time occupied by each stratum (equation (7)).

\[
\text{Load} = \sum_{s=1}^{2} \frac{N_s}{n_s} \left[ \sum_{i=1}^{n_s} Q_{is} C_{is} \right] \tag{7}
\]

Where \(s\) = the number of strata types (two in this case, base- and storm-flow), \(Q_{is}\) = flow for sample \(i\) in stratum \(s\), \(C_{is}\) = concentration for sample \(i\) in stratum \(s\), \(N_s\) = number of days over which each stratum occurs and \(n_s\) = number of days on which instantaneous load was sampled.

### 4.1.3 Ratio estimators

Ratio estimators were developed to improve on averaging-estimators essentially by including information regarding inter-relationships between average flows and loads at sampling times to derive annual load estimates. In this way, more use is made of the available information and underlying data patterns when calculating load estimates (Young et al., 1988). The optimal conditions for use of this ratio are when instantaneous fluxes and instantaneous flows are linearly related (with origin at zero) and variance in instantaneous fluxes increases with variance in instantaneous flows (Preston et al., 1989). In the basic form, a ratio estimate is usually calculated as:

\[
\text{Load} = \left( \frac{y}{x} \right) X \tag{8}
\]

where \(\bar{y}\) and \(\bar{x}\) are the sample means of two variables, \(y_i\) and \(x_i\), respectively. If the ratio of \(y_i\) to \(x_i\) is nearly the same for all samples, then the ratio estimate is of high precision. Ratio estimators are considered more suited to less frequently available concentration data than averaging estimators (Preston et al., 1989), hence are probably more appropriate for WA situations. Although a range of ratio estimators is available (Kronvang and Bruin, 1989; Preston et al., 1989), the Beale Ratio estimator (Beale, 1962) has been most widely applied in Europe and the USA and will be the only estimator discussed in this report.

The Beale ratio estimator estimates annual loads using estimates of instantaneous loads at sampling times, annual flows and a ratio factor accounting for co-variance structure between instantaneous flux (load) and flow values (as described in equation 9). The mid-part of the equation essentially provides a mean flow-
weighted concentration for sampling times based on fluxes when samples were taken, whereas the ratio factor adjusts for co-variance between fluxes and flows at sampling times standardised by the variance in flows at sampling times. In effect, the ratio factor increases the influence of flow in estimating annual load when flow and load values are strongly co-variant.

\[
\text{Load} = Q_a \left( \frac{\bar{I}}{\bar{q}} \right) \left( 1 + \frac{1}{N} \frac{\text{cov}(l,q)}{\bar{q}} \right) \left( 1 + \frac{1}{N} \frac{\text{var}(q)}{q^2} \right)
\]

(9)

Where \( Q_a \) = annual flow, \( \bar{l} \) = average load for times when samples were collected (usually using average flows at time of collection), \( \bar{q} \) = average flow for the times when samples were collected, \( N \) = number of samples collected over year, \( \text{cov}(l,q) \) = co-variance between sampled loads and flow at time of sampling, and \( \text{var}(q) \) = variance of the flows at the time of sampling.

4.2 Extrapolation approaches

Load calculations using extrapolation methods rely on the use of mathematical functions to “fill-in” concentration data sets over flow intervals where samples were not collected which are then used to estimate mass loads often using the linear interpolation algorithm. The mathematical functions, commonly termed rating curves, describe the deterministic relationship between flow and concentrations of water quality constituents using historical data.

Rating curves are frequently in the form of simple log-linear or other parametric equations. Parametric equations are the simplest approach and permit the use of classical least-squares regression techniques. These relationships are often in the form of the log-linear equation:

\[
\ln C_t = \beta_0 + \beta_1 \ln Q_t
\]

(10)

Where \( C_t \) and \( Q_t \) are concentration and flow at time \( t \), respectively, and the terms \( \beta_0 \) and \( \beta_1 \) are slope and interceptor coefficients. The log fit enables use of linear regression techniques to solve for the terms \( \beta_0 \) and \( \beta_1 \).

Nonparametric curves such as LOWESS (Locally Weighted Scatterplot Smoothing) combined with probability distributions to describe the scatter about the curves (Donohue and Nelson, 1999) may fit flow-concentration relationship better. These are preferred in analyses of water quality time series (Helshel and Hirsch, 1992).

The reliability of extrapolation approaches strictly commonly depends on the extent to which single functions can be used to explain the flow-concentration relationships. Poor description of changes in concentrations of water quality constituents in relation to flow using deterministic functions introduces substantial bias and error to load estimations (Cohn, 1995). A well-recognised source of bias (inaccuracy) in mass load estimates using the extrapolation approach arises due to the back-transformation of the product of equation (10) to obtain estimates of constituent concentrations using flow data (Miller, 1984). Several approaches can be used to correct for this bias, however, the errors in precision due to rating curves are more difficult to quantify, and therefore not readily corrected. The following sections also cover transfer curves, which are a type of rating curve that attempts to capture the more dynamic aspects of flow-concentration relationships.
4.2.1 Bias correction factors

Substantial bias can be introduced to estimates determined from linear rating curves fitted to logarithmically transformed Q and C (Miller, 1984; Ferguson, 1986; Cohn et al., 1989). This error is a mathematical error in addition to the errors due to poor fitting or inappropriate rating curves. Bias is introduced to the concentration values when the solution of re-arranged rating curve is back-transformed to determine C (equation 11) (Miller, 1984). The equation for predicting constituent concentrations at any point in time (Ct) from instantaneous flow (Qt) is obtained by re-arrangement of equation (10):

\[ C_t = \exp(\tilde{\beta}_0 + \tilde{\beta}_1 \ln Q_t) \]

where: \( \tilde{\beta}_0 \) is the estimator for \( \beta_0 \) and \( \tilde{\beta}_1 \) is the estimator for \( \beta_1 \) in equation (10).

The back-transformation bias can be corrected using various adjustment factors, with many assuming that the regression errors of log-transformed data are normally distributed. Explanation of the theoretical background to these approaches is beyond the scope of this review and the reader is directed towards Miller (1984), Koch and Smillie (1986) and Cohn et al. (1989) for more detailed discussion. Briefly, there are three main regression-bias correction techniques:

1. Quasi Maximum Likelihood Estimator (QMLE; Cohn et al., 1989): simplest correction using result of \( \exp(s^2/2) \), where \( s^2 \) is variance of the regression residuals (Ferguson, 1986 see Cohn et al., 1989; Thomas, 1985).

2. Minimum Variance Unbiased Estimator (MVUE): Derived from statistical theory (Bradu and Mandlak, 1970) and can potentially provide unbiased estimates of nutrient concentration strictly when errors are normally distributed (Cohn et al., 1989).

3. Estimator for Non-parametric Error Distributions: Where the distribution of regression errors of log-transformed data is not normal, non-parametric correction factors have been proposed (Duan, 1983) and demonstrated by Thomas (1985) and Koch and Smillie (1986). As above, these basically rely on inclusion of a correction factor in the back-transformation equation.

Mathematical analyses of the correction factors has shown that most factors generally result in slight positive bias in load estimates (Thomas, 1985; Koch and Smillie, 1986; Cohn et al., 1989). In the absence of bias correction, back transformation can result in substantial negative biases for estimates of constituent concentrations over most flows increasing to positive biases for very large or small flows (Cohn et al., 1989). Corrections using the QMLE result in slightly positive biases for moderate flows, but can result in highly positive biases for very large or small flows. Use of the MVUE was claimed to produce no bias in load estimates (Cohn et al., 1989).

Uncertainty still surrounds when it is best to use different factors to correct for back-transformation bias. Protocols for rigorous selection of suitable methods are yet to be published particularly when determining whether significant improvement in bias is achieved by use of correction factors. There are numerous studies citing inconclusive results on the selection of appropriate correction factors (eg Thomas, 1985; Koch and Smillie, 1986; Walling and Webb, 1988). Koch and Smillie (1986) reported that the correction for back-transformation bias often resulted in over-estimates of sediment loads and questioned the general use of these corrections citing that the statistical distribution of regression residuals is often not constant. Other work showed that while some correction factors could reduce bias from greater than -90% to better than -70% on one river and from greater -70% to better than -40% on a second river, factors other than back-transformation bias were frequently responsible for load bias (Walling and Webb, 1988). A universal condition is that corrections for back-transformation bias can substantially reduce the bias of the
load estimates provided by rating curves only when the curves provide a good description of flow-concentration relationships (Cohn, 1995).

### 4.2.2 Transfer functions

Transfer functions provide an alternative extrapolation approach that attempt to overcome temporal variations in flow-concentration relationships related to Q-C hysteresis processes. Hysteresis in nutrient concentrations during individual storms, seasonal and/or annual flow periods can be a common cause of poor fitting rating curves. Transfer functions partially overcome this problem by incorporating mathematical time lags to account for predictable delay patterns between constituent concentrations and flow, particularly during suspended sediment transport (Gurnell and Fenn, 1984; Lemke, 1991; Littlewood, 1995). As with parametric functions, concentration data extrapolated using transfer functions is incorporated with collected data to estimate nutrient loads often using simple integration equations (equation (2)). The main statistical objective of transfer functions is to overcome serial correlation in the residuals of rating curves, which violates the assumption of independent errors needed to fit parametric regression functions (Cohn, 1995).

Transfer functions can use either single or multiple time series variables to estimate another time-series variable (Box and Jenkins, 1976). The behaviour of time-series concentration data can be related to the present or past values of other time-series variables such as flow, rainfall or temperature (Lemke, 1991). This is basically achieved using functions that provide estimates of concentration at points in time based on the flow, rainfall or temperature at different times ahead or behind in time. For example, flow values at each 15 minute time intervals in the past 24 hours can be weighted (depending on the correlation between flow and concentration at different time lags) to predict the concentration occurring when lagged responses in concentration occur in relation to flows. Likewise, for sediment transport the time series flow data ahead of time often provides best prediction of sediment concentrations, which generally proceed flow peaks.

The general form of a transfer equation with input of single time-series data such as flow is:

\[
C(t) = F(z)Q_{t-b} + E_t
\]

Where \(C(t)\) is the concentration of the time series at time \(t\); \(F(z)\) is the transfer function filter (enabling selective influence of flow \((Q)\) at various times \(t-b\) on concentration \((C)\)), \(Q_{t-b}\) is the flow at time \(t-b\) and \(E_t\) is an error term. The transfer function filter \((F(z))\) is determined by \(a_1 + a_2B + a_2B^2 + \ldots \ldots \) where the values for \(a_1, a_2\) etc are the transfer weights and \(B\) is the back-shift operator (Box and Jenkins, 1976).

While transfer functions represent a useful approach to capturing the dynamic nature of flow concentration relationships for load estimation, some dynamic aspects of these relationships still remain difficult to model. Transfer functions only provide an approximation of hysteresis and do not take into account variations in the hysteresis response with consecutive events or seasonal patterns. Furthermore, variance in hysteresis patterns can frequently be attributed to essentially random elements of chemical transport from catchments, such as the spatial variance in rainfall and in its intensity, run-off pathways and contributions to drainage from partial sources in catchments. These are difficult to capture using models with simple environmental input variables. Transfer functions are also less able to model exhaustion processes (Littlewood, 1992).
5 Estimation of mass load errors using simulation methods

A fundamental limitation in estimating and partitioning errors in measuring mass loads is the lack of near-continuous concentration data. This data is necessary to provide estimates of true loads against which the effects of different sampling and calculation strategies can be tested. There are only a few extensive data sets covering extended flow periods that containing daily (and more frequent) measurements of nutrient or pollutant concentrations that can be used to obtain low bias, high precision measures of loads (Richards and Holloway, 1987; Preston et al., 1989; Young et al., 1988).

Monte Carlo simulations have been used to overcome data limitations by generating simulated time-series concentrations for extended periods from short periods of intensive data (Richards and Holloway, 1987). These time-series data sets have been used to provide estimates errors between true mass loads and loads obtained by different sampling and calculation methods. Sampling strategies are simulated by generation of individual concentration values at points in time that are consistent with a particular frequency or pattern of sampling (Richards and Holloway, 1987; Preston et al., 1989; Young et al., 1988). Repetition of this process provides sets of data essentially representing the complete ranges of concentrations and flow data likely to have been collected during the flow period for each sampling strategy (reflecting different starting times of sampling etc). The mass loads calculated using this information is compared directly with the true loads obtained using the baseline data.

Alternative simulation approaches rely less on large data sets and use transfer functions to predict constituent concentrations, thereby avoiding the problems resulting from re-sampling intensive data sets (Littlewood, 1995). The Simulation and Methods Investigation of Load Estimates for Rivers (SMILER) program estimates nutrient concentration data for historical flow data using a transfer function relationship between flow and concentration (Littlewood, 1995). This function is derived from smaller flow and concentration data sets (Littlewood, 1995). The use of transfer functions in the SMILER approach avoids problems with back-transformation bias (see above) and can re-create hysteresis in flow-concentration relationships. However, selection of parameters for the transfer function is less statistically rigorous and is guided only by visual assessment of the model against plotted data (Littlewood, 1995). Mass load errors may still be slightly under-estimated by simulations of concentrations using transfer equations particularly since the full probabilistic nature of nutrient concentrations for flow events may not be completely captured by the equations.

5.1 Errors arising from interpolation and ratio algorithms

Simulations have generally been applied to evaluate the performance of interpolation, ratio estimator and rating curve methodologies in combination with sampling frequencies to estimate the loading of individual nutrients or contaminants (eg Richards and Holloway, 1987). There are limited examples of using simulation approaches to cover a wide range of water quality constituents (Richards and Holloway, 1987; Preston et al. 1989; Littlewood, 1995; Kronvang and Bruhn, 1996). Little is understood about the errors associated with load estimates in WA rivers, but some benefit can be obtained by examining studies of load errors in rivers elsewhere.

There is a complex array of interactions that occur between sampling strategies and calculation algorithms that influence the precision and accuracy of mass load estimates. A consistent outcome of all
studies is the influence of sampling frequency, where increases in sampling frequency invariably improve bias and precision, although the effect is often non-linear and can differ between rivers (Richards and Holloway, 1987; Yaksich and Verhoff, 1983; Cohn et al., 1989; Kronvang and Bruhn, 1996). It is not possible to generalise the errors reported in the literature because many are specific to the water quality constituent or river under investigation (eg. Table 1). For example in a large river such as the Sundusky River in the USA, the bias and precision errors associated with weekly sampling can vary between 0% to ~5% (under-estimation) and 3% to 64% respectively for total phosphorus loads depending on the calculation algorithm or sampling strategy (Table 1). In contrast, bias and precision errors for total phosphorus loads in smaller streams can be much different, for example, bias errors were up to 14% for total phosphorus loads in Savijoki stream (Finland) whereas precision errors were only between 5 to 34%. (Table 1). A similar pattern existed in a mid-sized stream such as the Gjern Å in Denmark, except that the estimates of mass load were associated with small negative biases (Table 1). Clearly, discussion of individual studies would not provide a lucid overview of the literature. However, there remain several conclusions relevant to Western Australia that can be summarised from the literature and are provided below.

Different averaging algorithms can be useful for interval concentration data, although this depends upon the frequency of collected data. For some rivers, some averaging algorithms can magnify both precision and/or bias errors for estimates based on low sampling frequencies (Preston et al., 1989; Rekolainen et al., 1991; Littlewood, 1992). Sampling at greater frequencies could reduce these problems (Preston et al., 1989; Rekolainen et al., 1991; Littlewood, 1992). In some cases, achieving minimal bias can be at the cost of magnifying precision errors (Littlewood, 1995; Dolan et al., 1981). For suspended sediment loads, bias could be completely reduced using particular interpolation algorithms, however these also increased precision error by more than 25% (Walling and Webb, 1985; Ferguson, 1987; Littlewood, 1992). The errors introduced by use of different algorithms also differs depending upon the water quality constituent of interest, with some algorithms resulting in much larger errors with loads of suspended particulate constituents than with loads of soluble constituents (Richards and Holloway, 1987; Preston et al., 1989).

Comparison of averaging and ratio estimators with and without stratification found that year to year bias could be substantial for all methodologies and that no approach performed consistently well across years (Young et al., 1988; Dolan et al., 1981; Rekolainen et al., 1991; Kronvang and Bruhn, 1996). Overall, the Beale Ratio Estimator generally provided estimates with generally minimal bias (Young et al., 1988; Preston et al., 1989; Rekolainen et al., 1991; Littlewood, 1992). However, use of this estimator with some unstratified data can reduce precision (Preson et al., 1989; Rekolainen et al., 1991; Kronvang and Bruhn, 1996) or have little effect on it (Littlewood, 1995). For unstratified sampling programs, achieving a bias of less than 10% for many nutrient loads can only be obtained by sampling at greater than weekly frequencies (Richards and Holloway, 1987; Rekolainen et al., 1991; Kronvang and Bruhn, 1996). Up to an order of magnitude of further improvements in bias and precision can be achieved using stratification and the Beale Ratio Estimator within each strata (Dolan et al., 1981; Richards and Holloway, 1987; Young et al., 1988), although not in all streams (Kronvang and Bruhn, 1996).

Simulations have shown that stratification designed to increase sampling effort during high flow periods can reduce bias and improve precision, but further refinement of the parameters defining strata can have little effect on load errors for some nutrients in larger rivers (Young et al., 1988). This may not apply to smaller rivers, where setting of stratification criteria sufficient to detect small flow events in rivers draining small catchments has greater benefits for bias and precision of load estimates than the increasing the sampling frequency within these events (Rekolainen et al., 1991). Similarly, for some small streams, stratification of a fixed number of samples when there are significant variations in constituent transport between events can result in greater imprecision in mass load estimates (Kronvang and Bruhn, 1996). There are no calculation methods that can overcome problems of unstratified or inadequate sample
collection. Artificial stratification of data after sampling followed by calculation of loads within these nominal strata has not been found to improve bias and precision (Richards and Holloway, 1987).

5.2 Errors introduced by the use of extrapolation algorithms (rating curves)

Estimation of the precision of mass load estimates determined using rating curves to “fill-in” missing constituent concentrations is inherently more difficult than accounting for errors introduced through interpolation calculation methods. Recognition and correction of bias introduced by back-transformation was achieved largely through field observations and statistical theory (see earlier sections). However, evaluating the bias and reduced precision introduced through use of rating curves has largely been achieved using simulation approaches (Thomas, 1985; Ferguson, 1987; Young et al., 1988; Preston et al., 1989). These have been most successful because the source of rating curve error is frequently difficult to quantify and is often due to the degree to which the rating curves explain the true relationship between nutrient concentrations and flow.

Simulation studies have shown that rating curve estimates of loads perform well when flow and concentration relationships are well described using parametric and non-parametric relationships (Ferguson, 1987) and remain consistent across flow years (Young et al., 1988; Preston et al., 1989). These conditions are generally more likely for soluble water quality constituents such as dissolved phosphate, nitrate and salts, than for sediments or particulate forms of nutrients. Where tight fitting flow-concentration relationships exist, errors associated with nutrient loads estimated by using rating curves can be comparable with those obtained using ratio estimators, however, often perform better than averaging estimators (Dolan et al., 1981; Ferguson, 1987; Preston et al., 1989).

Description of the flow-concentration relationship for some water quality constituents can be difficult to achieve using simple linear relationships. Significant scatter of values about fitted curves frequently occurs even when using non-parametric relationships and is generally due to sampling variability, hysteresis, exhaustion processes or simply the seasonal differences in flow-concentration relationships (Whitfield and Schreir, 1981; Walling and Webb, 1988). The multiplicity of factors controlling the transport of some water quality constituents from catchments, particularly particulate forms of these, can result in the rating curves of individual storms often differing greatly from that of rating curves established using long-term relationships (Walling and Webb, 1988; Kronvang and Bruhn, 1996). This will mean that relationships developed using only a few weeks of data within a year should not be extrapolated to an annual period or longer.

Numerous studies have reported the problems arising from poorly fitted rating curves. Application of a single rating curve across flow periods can result in substantial error in load estimates primarily because the single relationship fails to capture dynamic processes such as hysteresis (Thomas, 1985; Walling and Webb, 1988; Littlewood, 1992; Kronvang and Bruhn, 1996). For sediment loads, errors ranging between −80% to +900% can be introduced by the inappropriate use of rating curves, although these errors would include back-transformation bias (Walling, 1977). Similarly, bias and imprecision of some mass load estimates such as total nitrogen were magnified by more than double when using rating curves, compared with interpolation approaches (Kronvang and Bruhn, 1996). Even when the transformation errors are corrected, simulation studies have shown that the reduced bias can sometimes be offset by declines in precision (Thomas, 1985; Ferguson, 1987). Rating curves can also inadvertently introduce error when developed from flow periods that do not cover the majority of expected flow-concentration patterns for the river (Walling, 1977; Preston et al., 1989). In these cases the rating curves may work well for some
years, but result in substantial bias and poor precision in other years, which can result in large errors for studies of cumulative loads (Young et al., 1988; Preston et al., 1989).

Analyses of back-transformation bias indicated that the use of log-linear rating curves to estimate concentrations of water quality constituents from flow also magnified precision errors. At very large or small flows, precision errors in estimating suspended sediment concentrations could be magnified by more than 100% when using back-transformed estimates (Cohn et al., 1989). Consequently, the frequency of occurrence of these flow conditions can greatly influence the overall error of final sediment load estimates using extrapolation approaches. Most parametric bias correction factors can improve precision across a range of flows relative to the use of uncorrected rating curves, although the precision errors can still be magnified as much as 150% at large or small flows (Cohn et al., 1989).

It is universally apparent from all studies that the performance of rating curves to estimate mass loads by extrapolation of nutrient concentrations is best when:

(1) the linear (or parametric) model provides the best fit to the data,

(2) the model is developed using sufficient sample data (more than 30 samples), and

(3) the model is not used to extrapolate beyond the original data range used to calibrate the model (Cohn, 1995).

As with all calculation methods, rating curves cannot overcome poor quality concentration and flow data or sampling strategies that result in unreliable information about true concentrations of constituents in rivers and drains.
Table 1: Precision and bias errors (% of mean loads) estimated for total phosphorus loads using combinations of sampling frequencies, strategies and calculation algorithms for rivers of different sizes.

<table>
<thead>
<tr>
<th>Sampling Strategy</th>
<th>Calc. Algorithm</th>
<th>Sundusky River, USA (3240 km²)¹</th>
<th>Honey Creek, USA (386 km²)²</th>
<th>Gjern Å Stream, Denmark (103 km²)³</th>
<th>Savijoki stream, Finland (15 km²)⁴</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Weekly Samples</td>
<td>Fortnightly Samples</td>
<td>Weekly Samples</td>
<td>Fortnightly Samples</td>
</tr>
<tr>
<td>Unstratified</td>
<td>Averaging Estimator</td>
<td>-5</td>
<td>64</td>
<td>-12</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>Beale Ratio</td>
<td>-5</td>
<td>38</td>
<td>-13</td>
<td>55</td>
</tr>
<tr>
<td>Stratified</td>
<td>Averaging Estimator</td>
<td>0</td>
<td>4</td>
<td>-0.4</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Beale Ratio</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

NB: Differences between the magnitude of errors for different studies can partly be due to different statistical methods used to calculate bias and precision errors.

¹,²: Results from Richards and Holloway (1987)
³: Results from Kronvang and Bruhn (1996)
⁴: Results from Rekolainen et al. (1999)
6 Conclusions

A major part of managing water quality monitoring programs is ensuring that the reliability of the collected data sufficiently meets the information needs of the programs. In many cases, one of the main objectives of monitoring programs in Western Australia is to provide information on the mass loads from catchments to lakes, wetlands, estuaries and coastal waters. A major problem with using mass loads is that the errors associated with the estimates are often not immediately obvious and therefore can be overlooked or ignored which weakens any conclusions drawn using the information. The practice of estimating mass loads involves estimating fluxes of water quality constituents by combining often near-continuous flow data with concentration data determined from sampling rivers and drains at discrete intervals. Data from sampling at intervals unavoidably introduces errors into mass load estimates when used to estimate what the true time-series concentration might have been. Minimising these errors can only be achieved by judicious sampling of rivers to obtain best approximation of the true pattern of temporal variation in constituent concentrations. The best approach for guiding such sampling is to identify the main pathways of constituent transport through catchments to determine the timing and magnitude of constituent fluxes in relation to hydrological discharge from catchments.

The main cause of errors associated with sampling to estimate mass loads is commonly attributable to estimating concentrations of chemicals between sampling times as part of determining flux over a period. Bias errors generally occur when sampling either consistently under-estimates true concentration patterns (Fig. 12). These generally occur as a result of missing large events where concentrations are greatest. Precision errors occur when sampling misses the variations occurring between sampling times, although not necessarily the high and low points. The precision and accuracy errors of individual concentration estimates additively contribute to errors in estimating fluxes and subsequently the overall estimates of loads over longer time periods (the sum of instantaneous fluxes). While these errors are generally more obvious, it should not be overlooked that flow gauging errors can also be significant for poorly maintained, unstable gauging stations.

This review has shown that overseas studies of errors in measuring mass loads yield several main conclusions relevant to estimation of mass loads in rivers and streams draining south-western Western Australia. These can be summarised as a series of principles relating to the selection of appropriate sampling and calculation methods, measurement of specific nutrient loads, and management actions to ensure that the measurements meet information needs.

- **Control of errors in sampling programs is fundamental to meet information objectives.**

Management of sampling programs is effectively conducted by managing sources of error in sampling and calculation strategies to produce load estimates that are within acceptable reliability limits for the intended use of the information (Ellis, 1989). Ultimately, it is the size of the load errors acceptable by managers that determines what strategies need to be taken to ensure that load data is useful. Some objectives can be met with less accurate and precise estimates than other objectives. For detection of trends in loads, bias error is acceptable, but only if consistent with time. However, bias error can be unacceptable for catchment or estuarine nutrient budgets, particularly since this is additive between streams and over long periods (Cohn et al., 1989). Large precision errors associated with load estimates can effectively mask detection of any short-term (5-10 year) trends. These errors can be acceptable for long-term estimates of nutrient loads to an estuary (Cohn et al., 1989), but may not be appropriate for data to be used in models of seasonal estuarine processes requiring daily simulation steps. It will be necessary to review sampling programs where the errors in load estimates using some fixed interval-sampling strategies invalidate the use of the data to meet original information objectives or can be obtained by cheaper strategies.
Understanding the pathways of constituent transport and the timing of these in influencing fluxes from catchments is fundamental for designing and managing monitoring programs to estimate the loads of specific nutrients or contaminants.

The design of monitoring programs to effectively sample loads of specific nutrients must consider the main mass transport pathways in catchments and the timing of delivery in relation to river flows. Understanding the dominant temporal and spatial patterns in the transport of specific water quality constituents from diffuse sources can contribute greatly to optimising sampling of fluxes since these sources frequently cause greatest variation in river loads. As a general rule, bias in load estimates can be
minimised by sampling during periods of maximum flux (Richards and Holloway, 1987) while precision can be increased by sampling during periods where flux varies greatest over time. For many water quality constituents, high flow regimes can be targeted to obtain maximum fluxes and minimise bias in loads (Rekolainen et al., 1991), particularly in catchments where surface transport pathways dominate. Similarly, flow regimes associated with greatest variation in mass fluxes occur (often where concentrations vary greatest during flow events) can be targeted to improve precision in load estimates. Logically, such approaches demand consideration of whether greatest variation in fluxes occurs during base-flows or various phases of storm flow events. Greater sampling will also often be necessary in rivers draining smaller catchments as compared to rivers draining larger catchments because of the more variable and flashy nature of loads from smaller catchments (Walling, 1977; Richards and Holloway, 1987; Robertson and Roerish, 1999). Consequently, a sampling program designed for smaller catchments might not be appropriate for larger catchments and could incur unnecessary sample collection for desired accuracy and precision.

• Manipulation of sampling frequency can have immediate implications for the accuracy and precision of load estimates.

For all fixed sampling strategies, increasing the sampling frequency can have substantial effects on the accuracy and precision of load estimates, although this effect is often non-linear (Richards and Holloway, 1987; Rekolainen et al., 1991; Littlewood, 1992; Kronvang and Bruhn, 1996). For a given number of samples taken over a flow period, significant additional benefits to estimation of reliable mass loads can be gained by stratifying sampling rather than relying on fixed-interval sampling (Richards and Holloway, 1987; Rekolainen et al, 1991; Thomas and Lewis, 1995). Clearly for some sampling programs, further manipulations of sampling strategies to achieve more acceptable load estimates may be outside of available funds, in which case it might be necessary to revise the information objectives of these programs. Evaluation of the errors associated with different sampling strategies for specific rivers is only possible using simulation models developed using water quality information collected intensively over a period to detect true patterns of mass transport (Littlewood, 1992).

• Acceptable bias and accuracy for all nutrients and contaminant loads can rarely be achieved.

There are no universal grab sampling strategies that can be routinely applied to all rivers and all constituents to provide guaranteed precise and accurate estimates of a range of constituent loads. Grab sampling programs should therefore be designed to evaluate key water quality constituents of interest within acceptable degrees of accuracy and precision. Generally, a program designed to produce acceptable estimates of mass load (within precision and bias limits) for the most variable constituent is also likely to achieve acceptable estimates for less variable constituents (Richards and Holloway, 1987; Rekolainen et al, 1991). In many cases, total suspended sediments are subject to greatest errors in load estimation, thus programs designed to estimate these loads should also provide reasonable estimates of less variable constituents such as dissolved nutrients and ions. However, financial constraints often constrain sampling programs, such that the emphasis is often on minimising the greatest sources of load errors for only high priority constituents, rather than all monitored water quality constituents.

• Consistent year-to-year precision and accuracy is generally not obtainable by conventional grab sampling strategies

Precision of many non-automated mass load estimation strategies can vary between years to the extent that for some rivers there can be no combination of grab sampling strategy and calculation method that consistently provides high precision (Young et al., 1988; Littlewood, 1995; Kronvang and Bruhn, 1996). The bias and error of methodologies can vary substantially within and between flow years as a result of variations in concentrations of water quality constituents and flow conditions (Rekolainen et al., 1991; Kronvang and Bruhn, 1996; Robertson and Roerish, 1999). This highlights the fact that evaluation of errors associated with any mass load estimation program must be conducted over at least several years to
capture the true long-term error distributions. Bias poses particular problems in that small bias error associated with annual load estimates is additive resulting in significant errors occurring for load estimates covering extended periods of time. This problem was illustrated in simulation work conducted by Littlewood (1995) which showed that the bias of nitrate loads sampled weekly could vary from less than 5% over 15 year periods to up to between 2% to 20% over less than 2 year periods. Similarly, precision could also vary, although to a lesser extent, from better than 2% to greater than 15% depending on the time period (Littlewood, 1995). The shorter the time period over which the load is aggregated the greater the imprecision.

- **Load calculation algorithms can magnify errors if applied inappropriately and can not compensate for poorly collected data**

Many studies have found that calculation methodologies can not overcome the inaccuracies of inadequate or poorly collected data. This emphasises the importance of ensuring that design of water quality sampling strategies is the most effective way of obtaining information to meet information objectives. In general, the Beale Ratio estimator (or similar ratio estimator approaches) appears to provide the least biased estimates for monitoring data collected at fixed intervals, although can magnify precision errors in some cases (Richards and Holloway, 1987; Littlewood, 1992; Kronvang and Bruhn, 1996). Use of extrapolation approaches enables determination of loads for periods where water quality data is not extensive or unavailable, but can magnify bias and precision errors (Kronvang and Bruhn, 1996). Applying interpolation algorithms to estimate loads from some types of interval data can also magnify the errors of load estimates (Walling and Webb, 1985; Littlewood, 1992). Commonly, it is the inappropriate use of calculation algorithms without considering the sampling frequency and timing that introduces bias and reduced precision to load estimates (Walling, 1977; Richards and Holloway, 1987; Preston et al., 1989; Rekolainen et al., 1991).

There are a variety of sampling and stratification approaches available to achieve highly reliable estimates of mass loads suitable for different river types in south-western WA. Automated sampling using various computer control programs can achieve load estimates with high accuracy and precision by employing greater stratification or estimation of the probabilistic patterns of nutrient concentrations in flows. Water quality information provided by auto-samplers is necessary for the development of simulation models that provide the best basis for evaluating the likely long-term performance of various sampling strategies in terms of providing information that meets information objectives. However, whether auto-samplers necessarily provide the most cost-effective information over the long-term depends upon whether the required bias and accuracy of load estimates can be obtained by generally cheaper manual sampling strategies. A clear, often over-looked benefit of employing many automated sampling strategies, is that a greater understanding of mass transport dynamics is achieved, which can be immediately extended to refine long-term manual sampling strategies.

The Water and Rivers Commission has developed a load measurement system that overcomes the problems of uncontrolled bias and precision errors highlighted in this report. This system has been successfully implemented on several rivers in southern WA to enable measurement of mass loads with known precision and accuracy. Each measurement unit consists of an automated water sampler controlled by data logger using a stage-initiated program to determine sampling. The parameters for the program are pre-determined using a statistical modelling package (PlaNet). This package enables simulation of different sampling strategies (obtained by manipulating the logger program parameters) and the effects of these on the accuracy and precision of load measurements are evaluated. Details and results of this approach will be provided in subsequent reports.
References and Recommended Reading


Borg, H.; King, P. D. & Loh, I. C. 1987, Stream and ground-water response to logging and subsequent regeneration in the southern forest of Western Australia: Interim results from paired catchment studies. Water Authority of Western Australia. Report no. WH34.


Gerritse, R., Barber, C. & Adency, J. 1990, The impact of residential urban areas on groundwater quality: Swan Coastal Plain of Western Australia., Water Resources Series No. 3,


# Glossary

**Milligrams (mg)**
1000th of a gram

**Milligrams per litre (mg L⁻¹)**
1000th of a gram in one litre of water. Equivalent to 1 part per million.

**Water Quality Constituents**
Includes any materials or chemicals transported by river (soluble and insoluble). Can range from sediment and nutrients to salts, pesticides and heavy metals.

**Mass Load**
Sum of instantaneous flux occurring past a point on a river over a defined period of time (eg tonnes over a year)

**Flux**
Rate of mass transport occurring per unit time (eg. Kilograms per second). Estimated as the product of flow (volume) and concentration of constituents in the flowing water.

**Error**
Extent to which the median of a population obtained by sampling differs from the “true” population that was sampled.

**Bias or Inaccuracy Error**
Extent to which the median of a population obtained by sampling is shifted from the median of the “true” population that was sampled.

**Precision Error**
Extent to which the variation of a population obtained by sampling is shifted from the variation of the “true” population that was sampled.
Publication feedback form

The Water and Rivers Commission welcomes feedback to help us to improve the quality and effectiveness of our publications. Your assistance in completing this form would be greatly appreciated.

Please consider each question carefully and rate them on a 1 to 5 scale, where 1 is poor and 5 is excellent (please circle the appropriate number).

How did you rate the quality of information?

1 2 3 4 5

How did you rate the design and presentation of this publication?

1 2 3 4 5

How can it be improved?

1 2 3 4 5

How effective did you find the tables and figures in communicating the data?

1 2 3 4 5

How can they be improved?

How did you rate this publication overall?

1 2 3 4 5

If you would like to see this publication in other formats, please specify. (Eg. CD)

Please cut along the dotted line on the left and return your completed response to:

Publications Coordinator
Water and Rivers Commission
Level 2, Hyatt Centre
3 Plain Street
East Perth WA 6004
Fax: (08) 9278 0639
WATER AND RIVERS COMMISSION

Hyatt Centre
3 Plain Street
East Perth
Western Australia 6004

Telephone (08) 9278 0300
Facsimile (08) 9278 0301
Website: www.wrc.wa.gov.au