The optimisation of water quality monitoring schemes

Jason Lessels, BLWSc (Hons 1)

A thesis submitted in fulfilment
of the requirements of the degree of
Doctor of Philosophy

Hydrology and Geo-Information Sciences Laboratory
Department of Environmental Sciences, Faculty of Agriculture and Environment
The University of Sydney

2014

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the University or other institute of higher learning, except where due acknowledgement has been made in the text.

Jason Lessels, BLWSc (Hons 1)

4 February 2013

Thesis Summary

The monitoring of water quality is necessary to ensure that the health of catchments is maintained. Water quality monitoring is often undertaken by government agencies to identify trends, assess management strategies and the state of catchments. Many water quality studies attempt to identify the quantity and timing of nutrients exported from a catchment. The accuracy of the monitoring scheme is largely controlled by the sampling scheme. Financial constraints are one of the limiting factors and under this constraint, sampling schemes often combine limited sample sizes with estimation methods. The use of statistical methods allows catchment managers to improve the information on expensive water quality properties based on the relationship with low cost properties. Many water quality monitoring programs have access to limited historical data, therefore there is a requirement for methods which can use this limited data to improve water quality monitoring schemes. This thesis aims at examining the effect of event-based sampling, using historical data to improve sample designs and the use of model-based geostatistical methods to improve the quantification of nutrient exports.

Rainfall events are often associated with the export of large quantities of sediment and nutrients. This relationship is exaggerated in catchments with long periods of base-flow conditions and short rainfall events. Monthly routine sampling is commonly used in Australia due to financial constraints. Increasingly, event-based sampling is being used in conjunction with routine sampling to improve load estimates. The use of event-based sampling is still limited as this form of sampling requires additional equipment and maintenance. Therefore the benefit of including event-based sampling on the accuracy of load estimates is of interest. However, without access to continuously sampled water quality data it is difficult to determine the effect of event-based sampling on load estimates. One approach to overcome the lack of data is to simulate continuous water quality data using stream discharge. Chapter 3 outlines a simulation procedure using a linear mixed model to characterise the relationship between total phosphorus and stream discharge. The benefit of including event-based sampling is evaluated based on the accuracy of annual load estimates. The results showed that event-based sampling improved the accuracy of load estimates for all catchments. Relating these findings to catchment characteristics found event-based sampling had

a greater effect on the accuracy of load estimates in catchments with a large relief and high annual rainfall.

Based on the information gained from chapter 3, it is clear that accurate load estimates require the combination of event-based and routine (base-flow) sampling. The next logical question which should be addressed is; *how many samples should be taken?* Evaluations of the required sample sizes are often based on the uncertainty of precision. The estimation of the uncertainty of the mean is easily obtained when probabilistic sampling schemes are used. However, in situations where non-probabilistic sampling schemes are used, a model-based approach is required to ensure the estimated uncertainty of the mean is unbiased. Generally there is little information on the form of sampling scheme used for the collection of historical water quality data. In addition, most water quality sampling schemes are non-probabilistic and generally auto-correlated in time. Therefore a model-based approach is required to estimate the uncertainty of the mean in relation to the estimated mean. As the samples are generally not equally spaced through time, a variogram model is used to characterise the temporal auto-correlation between samples. A simulation based approach is then used to estimate the standard error of the mean based on different sample sizes. The results of the analysis indicated that there is little improvement in precision for both annual base-flow and event-flow estimates above 12 samples. The improvement in precision for both base and event-flow conditions was most related to the urban cover in each catchment.

The results of chapters 3 and 4 show the importance of event-based sampling as these periods are associated with large nutrient and sediment exports and increased uncertainty. Of particular interest for several reasons is the event mean concentration. Equally spaced temporal sampling is one of the most commonly applied event-based sampling schemes. However using this form of sampling requires the use of a model to obtain unbiased estimates of the event mean concentration. Probabilistic based sampling schemes do not require a model-based approach. Several probabilistic methods have been evaluated for event-based estimates. These schemes have been shown to provide accurate estimates of event-mean concentrations, however these methods have not been widely implemented possibly due to the complexity of the methods. Therefore, chapter 5 evaluates the use of a simplified time stratified sampling scheme for event-based sampling. The proposed method stratifies the mean event hydrograph based on key hydrological components (e.g. the rising and falling limbs). As each event is not identical to the mean event hydrograph, re-stratification after each event is presented as a method to estimate the mean concentration of the key hydrological components. The results showed that the method can provide accurate event mean estimates of two water quality properties controlled by different hydrological transport processes.

Load estimation methods are commonly used in water quality studies to characterise the mass of nutrients or sediments which are exported from a catchment over a given period of time. Many prediction methods use information from other hydrological properties for these estimations. However, many of these load estimation methods should not be used as these methods assume the use of probabilistic sampling methods. With this restriction, there is a requirement for estimation methods which do not require probabilistic sampling. Linear mixed models are presented as an alternative estimation method. Linear mixed models provide the ability to account for the temporal auto-correlation between water quality samples and allow for covariates to improve predictions. Chapter 6 investigates the use of stream discharge, stream discharge derived variables and turbidity as covariates for predicting total nitrogen and total phosphorus. All fitted models found significant temporal auto-correlation between the samples. The inclusion of turbidity as a covariate improved the accuracy of the total nitrogen and total phosphorus predictions by 15%. In addition to improving predictions during base-flow conditions the inclusion of turbidity as a covariate improved the predictions of total phosphorus during storm events by 24%.

This research has applied various model-based geostatistical methods to examine the effect of different sampling schemes and sample sizes on the accuracy of water quality monitoring schemes. In regards to water quality sampling this research has found that event-based sampling is important in providing accurate load estimates. This research has lead to the following recommendations in relation to water quality sampling:

- There is little improvement in accuracy above 12 samples per year on the estimation of mean annual base-flow concentrations in several south-eastern Australian catchments.
- To simplify and improve the accuracy of event mean concentrations, a time-stratified sampling scheme should be applied.

In relation to water quality load estimation, this research does not recommend the use of traditional load estimation methods due to their assumptions. Instead of these methods this thesis recommends;

- The use of linear mixed models combining routine and event-based samples and low cost continuously monitored explanatory variables (e.g. stream discharge, turbidity) to estimate annual loads.

- This research recommends the use of affordable water quality surrogates to be incorporated in the linear mixed models to increase the accuracy of the load estimates.

Acknowledgements

I thought I would never be at the stage of writing these acknowledgements, however, the day has finally come. I have had a fun and exciting time undertaking this research and I'm thankful to all who have helped me along the way. During my thesis I have had the pleasure to work with many people who have made this such an enjoyable experience.

Firstly, I would like to thank my supervisor Dr Thomas Bishop. In a language he will understand; *Very wise, is Tom.* Without his advice, help, support and patience I would not have been able to learn the ways of the force. I'm extremely grateful for his time, effort and trust during the last 4 years.

Without planning it, I spent many days in the field and many more in the lab. During this time, I relied heavily on the support of Floris van Ogtrop. Floris was with me in the trenches of the now infamous "Afghanistan" and we spent several days cooking in the lab. For his help, I'm extremely grateful.

I would also like to thank Senani Karunaratne for his help in the field and for making the final stages of this thesis more enjoyable. In addition, Senani gave me the right information to get my new position and the possibility to go hunting for polar bears.

Most of this thesis has relied on data from the Sydney Catchment Authority. Both Dr Rob Mann and Dr Grant Tranter have patiently assisted me with my (at times constant) requests and questions over the last 4 years.

During the last few years I had the pleasure to work with several other PhD students. Joe Henry was there when I first started and struggled along side me; learning the ways of latex and R. Dr Michael Nelson was also there to support me, his help and knowledge was remarkable and I'm very grateful to him for all of his contributions along the way. I would like to also thank Dr Willem Vervoort for his assistance during my work.

During the first year of my PhD I was involved with FAPA. Being part of this group of people made the first year so much easier and fun. I wont forget the fun that was had planning and partying in the beloved Ross St building. My family have also helped me throughout this process. Thanks Troy for your help with the building of the equipment which lead to one of the following chapters. Brock for helping me with the programming. Thanks mum and Gill for being there over the last 4 years. Mum: *I told you it wasn't a crazy idea.*

Lastly, I would like to thank the one and only, Birgit. I couldn't have done this without you and I'm glad I got to do this with you by my side. I will be forever sorry for my broken promise, but I did eventually finish. I'm looking forward to our future in sunny Scotland.

To dad, I would be nothing without your suggestions.

Knowledge is hassle - Karl Pilkington, 2006

Published/submitted chapters

Chapter 3:

Lessels, J.S, Bishop, T.F.A, 2013. A simulation based approach to quantify the difference between event-based and routine water quality monitoring schemes. Hydrology Research.

Chapter 4:

Lessels, J.S, Bishop, T.F.A, 2014. Using the precision of the mean to estimate suitable sample sizes for monitoring total phosphorus in Australian catchments. Hydrological Processes, DOI: 10.1002/hyp.10205.

Chapter 5:

Lessels, J.S, Bishop, T.F.A, 2013. Characterisation of events using stratified random sampling. Water Resources Research.

Chapter 6:

Lessels, J.S, Bishop, T.F.A, 2013. Estimating water quality using linear mixed models with stream discharge and turbidity. Journal of Hydrology. 498 (13-22), DOI: 10.1016/j.jhydrol.2013.06.006

Contents

List of Figures

List of Tables

Chapter 1

General introduction

Water is one of the most important substances for human life. Therefore, the maintenance of water quality is vital to ensure continuous water supply for agricultural, recreational, ecological, industrial and drinking requirements. With growing pressure from land-use change and population growth, governments are under increasing pressure to maintain water quality. Water quality monitoring must be undertaken by government agencies to provide information on how to secure the quality of water into the future. The monitoring of water quality is complex and its accuracy and efficiency is determined by the sampling scheme and the statistical analysis used to quantify the state of the system. One of the largest constraints of water quality monitoring programs is cost which is largely due to; the number of sites, the number of water quality variables analysed, the travel time to each monitoring site, and the maintenance of each site. These high costs restrict the amount of samples and the number of sites which can be monitored.

Several studies have examined the effects of different water quality sampling schemes. These studies have often focused on the required frequencies of systematic sampling. Different sampling frequencies have been recommended; weekly (Johnes, 2007), daily or even sub-daily (Wade et al., 2012). These studies have provided great insight into the relationship between sample frequencies and water quality processes. However, the collection of data for these studies is extremely expensive and requires specialised equipment or high analytical costs. For example the study of Wade et al. (2012) involved the use of in-situ phosphorus digestion based equipment. The equipment required mains electricity and a large housing unit (Wade et al., 2012). In addition to these requirements large quantities of reagents were required which increased the maintenance of the equipment (Wade et al., 2012). The use of this equipment allows for advances in the understanding of water quality properties, however it is unrealistic for this technology to be widely implemented in the near future. In addition, as these studies are based in Western Europe it is difficult to draw direct comparisons to Australian catchments.

There is a distinct difference between the recommendations of scientific studies and the sampling schemes used by catchment managers. Due to financial constraints monthly sampling is the most commonly used sample frequency by catchment managers (Bartley et al., 2012; Wade et al., 2012). This form of sampling has been found to provide accurate estimates for catchments which do not have large rainfall events (Harris et al., 2007). However, this form of sampling will often miss key rainfall events associated with large proportions of sediment and nutrient exports. This is especially important in Australia where catchments are generally characterised by long durations of base-flow and large rainfall events (Drewry et al., 2009; Hopmans and Bren, 2007), which can contribute up to 60% of 6 years of sediment in one event (Drewry et al., 2009). Based on this, catchment managers are increasingly combining monthly sampling with event-based sampling. Event-based sampling is often undertaken using automatic pumping devices which use stream discharge to determine when to commence sampling. The inclusion of event-based sampling provides information on sediment and nutrient exports during these periods.

For many catchment managers it is a statutory requirement of government agencies to annually report the state of catchments. In Australia, catchment managers use the ANZECC guidelines (ANZECC, 2000) to assess the state of a catchment. The guidelines provide target concentration ranges for various chemical, physical, and biological properties. For example, the Sydney Catchment Authority reports the percentage of samples over the government guidelines (Sydney Catchment Authority, 2011). Percentage values are given for event and base-flow condition by dividing the samples based on the associated stream discharge. This form of reporting may not be a true representation of the state of the catchment due to the small sample size. In addition, this form of reporting does not provide a measure of uncertainty which is important for assessing the usefulness of the information (Harmel et al., 2009). Therefore there is a requirement for statistical methods which can provide a more useful analysis. As sampling, and analytical costs are often too high, alternative methods are required to improve estimates which can incorporate low cost variables which are related to water quality.

A common objective of water quality studies is to estimate the total quantity of sediment or nutrient exports during a period of time. There are many different load estimation methods (> 30), however there is no method which is clearly the best (Marsh and Waters, 2009). These methods are very beneficial in providing information on the state of a catchment water quality trends. However Australian catchment managers relate water quality to the ANZECC guidelines, which are based on concentration values. Therefore ratio and averaging based methods may not be the most appropriate methods for Australian catchment managers in regards to the ANZECC guidelines. Based on this catchment managers should use trend based estimation methods (e.g. regression models) which use explanatory variables to predict concentration. A major benefit of these estimation methods is the ability to incorporate information from low cost variables which are related to water quality.

There are several issues related to using trend estimation methods. These methods generally require normally distributed data, a form of probabilistic sampling, and use a linear trend between water quality and the explanatory variables. Water quality data is generally non-normally distributed and highly positively skewed. To overcome the non-normality, many studies use a log transformation (Kuhnert et al., 2012). In most cases the logarithm transformation meets the requirement of normally distributed data. However, in some situations there is a non-linear rather than linear trend between the transformed water quality and stream discharge (Kuhnert et al., 2012). Recently, generalised additive models (GAMs) have been used to overcome this problem. GAMs use a smoothing function instead of a linear relationship which allows for greater flexibility in the relationship (Kuhnert et al., 2012).

A further limitation of trend estimation methods is the assumption that the samples are independent and identically distributed (*iid*) (Lark and Cullis, 2004) i.e. the samples have no serial correlation. This can be achieved by using a probabilistic sampling scheme e.g simple random sampling. This assumption of *iid* samples is of significance in regards to water quality as many water quality sampling schemes do not use simple random sampling. Without the use of random sampling it can not be assumed that the data is independent and identically distributed (*iid*). Therefore the applied model must examine and account for the potential temporal auto-correlation between the samples. Without accounting for the temporal auto-correlation between the samples, the uncertainty of the estimates may be biased (Lark and Cullis, 2004). Due to the temporal nature of water quality, temporal auto-correlation may be present in situations where the data is *iid.* In these situations, the model will be valid, however it may be possible to improve the uncertainty of the estimates by accounting for this temporal auto-correlation.

The uncertainty of the estimates is an important component and should not be ignored. Uncertainty estimates provide valuable information which can be used to improve monitoring programs (Harmel et al., 2009). Using the uncertainty of the water quality estimates, it is possible for catchment managers to re-design the sampling scheme in order to minimise the amount of samples in relation to the uncertainty of the water quality estimates (de Gruijter et al., 2006). Another desire of catchment managers is to assess how many more samples are required, therefore the relationship between the sample size and the uncertainty of the water quality estimates can be used (de Gruijter et al., 2006). In addition, catchment managers need unbiased uncertainty estimates of predicted water quality to ensure correct management decisions are implemented (Horowitz, 2003).

Water quality monitoring is an expensive and complex task. Currently there are differences between the recommendations of scientific studies and The methods used by catchment managers. Catchment managers require methods which utilise information from historical data to improve the accuracy and efficiency of monitoring programs. In relation to Australian catchments, there is a requirement for methods which can provide continuous concentration estimates. Uncertainty estimates are also required to ensure appropriate management decisions are made. These estimation methods must incorporate additional information from variables which are less expensive to minimise the cost of the monitoring program.

The aims of this thesis are:

- Outline a method to evaluate different temporal sampling schemes using limited historical data. Using this method, quantify the reduction of uncertainty from the inclusion of event-based sampling.
- Examine the sample size requirements in relation to the uncertainty of the mean estimates based on historical temporal data. Outline the use of geostatistical methods to recommend sample size requirements for events and base-flow conditions.
- Design and outline a simple but flexible event-based sampling scheme. The desired sampling scheme must be flexible enough to account for different transport pathways.
- Examine the use of a geostatistical based method to estimate continuous water quality concentrations. The method must account for a biased sampling scheme and provides unbiased uncertainty estimates.

1.1 References

ANZECC, 2000. Australian and New Zealand guidelines for fresh and marine water quality. Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand, Canberra, 1–103.

Bartley, R., Speirs, W., Ellis, T., Waters, D., 2012. A review of sediment and nutrient concentration data from Australia for use in catchment water quality models. Marine Pollution Bulletin 65 (4), 101–116.

de Gruijter, J., Brus, D., Bierkens, M., Knotters, M., et al., 2006. Sampling for natural resource monitoring. Springer Verlag.

Drewry, J., Newham, L., Croke, B., 2009. Suspended sediment, nitrogen and phosphorus concentrations and exports during storm-events to the Tuross Estuary, Australia. Journal of Environmental Management 90 (2), 879–887.

Harmel, R., Smith, D., King, K., Slade, R., 2009. Estimating storm discharge and water quality data uncertainty: A software tool for monitoring and modeling applications. Environmental Modelling & Software 24 (7), 832–842.

Harris, R., Sullivan, K., Cafferata, P., Munn, J., Faucher, K., 2007. Applications of turbidity monitoring to forest management in California. Environmental Management 40 (3), 531–543.

Hopmans, P., Bren, L., 2007. Long-term changes in water quality and solute exports in headwater streams of intensively managed radiata pine and natural eucalypt forest catchments in south-eastern Australia. Forest Ecology and Management 253 (1-3), 244–261.

Horowitz, A., 2003. An evaluation of sediment rating curves for estimating suspended sediment concentrations for subsequent flux calculations. Hydrological Processes 17 (17), 3387–3409.

Johnes, P., 2007. Uncertainties in annual riverine phosphorus load estimation: impact of load estimation methodology, sampling frequency, baseflow index and catchment population density. Journal of Hydrology 332 (1-2), 241–258.

Kuhnert, P., Henderson, B., Lewis, S., Bainbridge, Z., Wilkinson, S., Brodie, J., 2012. Quantifying total suspended sediment export from the Burdekin river catchment using the loads regression estimator tool. Water Resources Research 48 (4), W04533.

Lark, R., Cullis, B., 2004. Model-based analysis using REML for inference from systematically sampled data on soil. European Journal of Soil Science 55 (4), 799–813.

Marsh, N., Waters, D., 2009. Comparison of load estimation methods and their associated error. In: 18th IMACS World Congress MODSIM. Vol. 4. Cairns.

Sydney Catchment Authority, 2011. Annual water quality monitoring report annual water quality monitoring report 2010-11. Annual report, Sydney Catchment Authority, Sydney, Australia.

Wade, A., Palmer-Felgate, E., Halliday, S., Skeffington, R., Loewenthal, M., Jarvie, H., Bowes, M., Greenway, G., Haswell, S., Bell, I., et al., 2012. Hydrochemical processes in lowland rivers: insights from in situ, high-resolution monitoring. Hydrology and Earth System Sciences 16 (11), 4323.

Chapter 2

Literature review: an overview of water quality monitoring

2.1 Introduction

Within Australia, there is growing pressure on the supply of water for; agricultural, recreational, industrial and drinking requirements. With increasing population and a changing climate there is increasing difficulty in maintaining water quality across the country. Water quality monitoring is undertaken throughout the country by many different government agencies and is estimated to cost over $142 million annually (Bartley et al., 2012). At a smaller scale, monitoring of 10 sites in the upper Murrumbidgee catchment had an annual cost of $150,000 (Newham et al., 2001). In addition to growing pressure on the quality of the available water, severe droughts over the past decade have had a large effect on water supply (Palmer, 2012). This shortage is causing a greater pressure on the quality of the available water as it is becoming increasing difficult to supply potable water. Appropriate water quality monitoring schemes are therefore required to provide the best achievable assessment of the state of catchments.

Understanding and developing a clear and concise objectives is vital to ensure a successful water quality monitoring scheme is implemented. The aim of the monitoring will effect the two major questions of a water quality monitoring scheme; *What form of sampling should be used?* and *How many samples are required?* Having a clear aim for the monitoring scheme will assist in the development of an appropriate sampling scheme, as the use of incorrect sampling schemes will restrict the information attainable about the system. However, many monitoring schemes use a single sampling scheme to address several different aims which can reduce the efficiency of the applied sampling scheme.

The monitoring of water quality is expensive (Bartley et al., 2012). One of the most expensive components of the monitoring schemes is the analysis of the samples. This cost places a large restriction on water quality sampling and is exaggerated by the monitoring of several sites which results in additional costs due to travel. Many catchment management agencies rely on monthly sampling; however this form of sampling will often miss key rainfall events (Drewry et al., 2009). Therefore the addition of event based sampling is becoming more common. The combination of these two forms of sampling increases the complexity of the analysis required to provide clear and meaningful annual reports.

As many monitoring schemes have limited data and rely on a mixture of sampling methods, there is a need for statistical based methods which maximise the information gained from this data. This information is required in order to detect trends and assess the state of the catchment. One of the main objectives of these methods is to provide an estimate of the exported load of a particular water quality variable (e.g. suspended sediments) over a given period (e.g. a year). Annual loads are often used as these allow for the comparison between years and catchments. There are however, a large number of load estimation methods available. A single study by Marsh and Waters (2009) compared 34 different methods. Several studies have examined the ability of each method to provide accurate estimates (Cassidy and Jordan, 2011; Johnes, 2007; Jordan and Cassidy, 2011; Marsh and Waters, 2009), however there is still no single method which is universally recommended. In addition to statistical based methods, several mechanistic water quality methods exist (e.g. SWAT (Arnold et al. , 1998), QUAL2E (Brown and Barnwell, 1987)), however these methods are out of the scope of this review.

This literature review will focus on the two major components of water quality monitoring;

1. sampling schemes and
2. load estimation methods.

An analysis of commonly used sampling schemes will be given with particular attention given to the suitability of event-based sampling schemes. With an overview of common used sampling schemes, load estimation methods will be reviewed in regards to the effect of sampling schemes on the suitability and accuracy of these estimation methods.

2.2 Water quality aims

The first stage of the design of a monitoring scheme is the development of its objectives. The aims of the monitoring scheme are important as they will affect the decisions made about the sampling scheme and the most appropriate form of statistical analysis. de Gruijter et al. (2006) define three aims of monitoring schemes:

- *trend monitoring*,
- *status monitoring*, and
- *compliance monitoring.*

Trend monitoring aims at detecting changes through time and generally involves long term monitoring (de Gruijter et al., 2006). In regards to water quality monitoring, this form of monitoring is often used to detect changes in the land management/use (e.g. Lane et al. (2006)). Status monitoring is used to characterise the state of the system through time (de Gruijter et al., 2006). For example, this form of sampling is used by the Sydney Catchment Authority to assess the improvement from upgrading sewage treatment plants (Sydney Catchment Authority, 2012) and has been used by Hopmans and Bren (2007) to estimate suspended sediment export changes under forest plantations. Compliance monitoring is the the process of determining if the state of the monitored property complies with guidelines (de Gruijter et al., 2006). This is a common aim of water quality monitoring schemes in Australia which use the guidelines of the Australian and New Zealand Environmental Conservation Council (ANZECC) (ANZECC, 2000).

Water quality monitoring schemes often address all of these aims using a single sampling scheme. For example, the Sydney Catchment Authority (SCA) uses a combination of monthly sampling and rainfall event sampling downstream of the sewage treatment plant in Goulburn, New South Wales, Australia. Using these samples the SCA annually reports the number of samples which are above the recommendations of the ANZECC guidelines (Sydney Catchment Authority, 2011). For example, the sampling scheme used to estimate the frequency of samples which exceeded the ANZECC guidelines (Sydney Catchment Authority, 2008a) was also used to estimate the reduction in nutrient loads from sewage treatment plants (Sydney Catchment Authority, 2008b). Under these situations it is important that the selected sampling scheme is suitable for all aims.

2.3 Sampling schemes

For the purposes of this literature review it is important to clarify the terminology used to describe sampling schemes. A *sample* is an individual observation which is collected at a point in time during a certain period. However, the collection of one sample is not instantaneous and therefore the period of time over which the physical act of collecting the water sample is undertaken is separated into *sample units*. In regards to temporal water quality monitoring, sample units are of equal length and are generally based on the time required to collect a single sample. The *sample frequency* of a sampling scheme refers to the temporal interval at which regular sampling is applied. The *sample size* of a sampling scheme is the term given to the number of samples collected within a certain time period.

When designing a sampling scheme it is important to consider what nutrients are of interest (Johnes, 2007). In addition to quantifying nutrient exports, information about the controlling processes of nutrient exports are also of interest. There are two main sources which contribute to nutrient exports in a catchment, point and diffuse (Cassidy and Jordan, 2011; Chen et al., 2013; Lam et al., 2010; Johnes, 2007). When a nutrient enters a stream from a single point (e.g. a wastewater treatment plant) these sources are clearly defined spatially in a catchment. Alternatively, when a nutrient enters a stream via large scale pathways (e.g groundwater), the sources of these nutrients are more complex and are defined as diffuse. It is important to understanding the affect of these sources on nutrient loads and the what form of sampling is required to correctly quantify these loads (Cassidy and Jordan, 2011; Johnes, 2007).

There are two categories of sampling schemes; probabilistic (design-based) and non-probabilistic (model-based) (de Gruijter et al., 2006). Probabilistic sampling schemes use *inclusion probabilities* to determine sample statistics such as the mean and variance. This form of sampling allocates an inclusion probability to each sample based on the randomisation used in the sampling scheme. Non-probabilistic approaches do not require a well defined sampling scheme or known inclusion probabilities. These forms of sampling require the use of a model to describe the uncertainty of the estimates.

Non-probabilistic water quality sampling schemes are the most common. An example of this is the use of regular intervals through time (e.g. weekly, monthly, quarterly) for sampling. This form of sampling is commonly used in water quality studies due to the ease of implementation and the temporal nature of the sampling. However, as this sampling scheme is non-probabilistic, the use of this form of sampling requires the use of a model to account for the potential temporal correlation between the samples (de Gruijter et al., 2006; Lark and Cullis, 2004). In addition, it is important that the time between samples is short enough to adequately characterise short range variation. (Lohr, 2009; Madrid and Zayas, 2007; Thomas, 1985).

Water quality sampling is often undertaken using two forms of sampling schemes; routine and event-based sampling (Drewry et al., 2009; Marsh and Waters, 2009; Moosmann et al., 2005). These two schemes are often used focussing on two separate aims. Routine sampling is often used to monitor base-flow conditions in catchments and provide information on long term trends (Hopmans and Bren, 2007). This is the most common form of sampling scheme, and generally only used in conjunction with event-based sampling where finances allow. Event-based sampling is used to sample streams during periods of high flows associated with rainfall events.

2.3.1 Diffuse and point source nutrients

Point sources and diffuse sources have a large effect on the required sample sizes and the form of sampling required in a catchment (Johnes, 2007). Several studies have investigated the effect of both point an diffuse sources on the timing and the quantity of nutrients exported from a catchment (Cassidy and Jordan, 2011; Chang, 2008; Chen et al., 2013; Lam et al., 2010; Johnes, 2007). In addition, these papers have also investigated the required sampling frequencies in catchments with certain sources of pollutions (Cassidy and Jordan, 2011; Johnes, 2007). These sources have often been linked to dominant land types with many studies using land uses and types to reflect theses sources (e.g urbanisation cover to reflect the effect of waste water treatment plants) (Chang, 2008; Johnes, 2007).It is therefore important to consider the dominant land use and cover in a catchment

A common source of diffuse nutrients are agriculture activities which increase the amount of nitrogen in the catchment, either via the growth of livestock or by the addition of fertilisers (Cassidy and Jordan, 2011). In addition suspended sediment is a major source of nutrients in Australian catchments and is often closely related to particulate phosphorus (Drewry et al., 2009), which is often delivered into a stream bound to suspended sediments with the largest exports occurring during rainfall events. These events are often quite short in duration, but contribute large proportions to annual loads (Drewry et al., 2009; Hopmans and Bren, 2007).

A common point source in Australian waterways are waste water treatment plants which are often responsible for releasing water with high nutrient concentrations into streams (Cassidy and Jordan, 2011; Johnes, 2007). These plants often alter the stream discharge of the stream by releasing water a constant rate, which can have a positive affect on the health of the stream. However, during rainfall events, these facilities are often required to release excess and often untreated effluent into the streams as they reach maximum capacity. The effect of these point sources on streams are often complex, but in general, often require sub-daily sampling to accurately characterise these processes (Johnes, 2007).

As different sources of nutrients have different flow paths and different temporal dynamics it is important for the dominant sources of nutrients in a catchment to be considered when designing a sampling scheme. The complexity from the combination of different sources may also increase the complexity of the sampling scheme, by requiring sampling at different frequencies through time and the requirement of catchment specific sampling schemes.

2.3.2 Routine sampling

Non-probabilistic sampling schemes are the most commonly used sampling schemes for routine sampling. Several studies have used near continuous water quality data to evaluate the effect of sampling frequencies on the accuracy of load estimates (Wade et al., 2012; Cassidy and Jordan, 2011; Bowes et al., 2009; Johnes, 2007). These differences in the required sample frequencies between catchments and studies provides information on the effect of catchment characteristics on the uncertainties of estimates. Regular interval sampling schemes, are commonly used by government agencies, including agencies in Australia (Bartley et al., 2012). Probability based sampling schemes have received very little attention for base-flow sampling, however the results of these studies have shown these methods to provide accurate estimates (Arabkhedri et al., 2010; Cassidy and Jordan, 2011).

Linking the required sample size to catchment characteristics is important to provide information on the form of sampling in an unmonitored catchment. Catchment size has been shown to have an effect on the required sample frequency (Moatar et al., 2006). Moatar et al. (2006) found the size of the catchment was inversely related to the required sample frequency. The results of the study by Cassidy and Jordan (2011) were similar as they found near continuous sampling was required in small upland catchments. The recent study by Wade et al. (2012) found daily sampling was required for a catchment in which daily releases from a sewage treatment plant were within the catchment.

Table 2.1 provides a summary as an example of several recent studies which have compared the effect of different sampling frequencies on the accuracy of annual load estimates. The majority of these studies have been based in Europe in small catchments ($<$100 km^2). The general consensus of these studies is that the use of sub-daily, or at least daily, sampling is recommended to provide accurate annual load estimates. For long term trend analysis, Wade et al. (2012) and Bowes et al. (2009) recommend weekly sampling. The results of these studies are often based on the accuracy of annual load estimations. In contrast to these findings, Horowitz (2003) suggests that fewer than 12 samples per year are adequate to estimate annual suspended sediment exports. Horowitz (2003) shows how the combination of samples from multiple years can be used to provide accurate annual estimates using only 6 samples. In addition, Horowitz (2003) suggests the use of sampling of the stream discharge distribution may be more appropriate than sampling through time.

There is a distinct difference between the recommendations of the majority of papers shown in Table 2.1 and that of the practices of catchment management agencies (Bowes et al., 2009). For example, the environmental agency in the UK is responsible for collecting water quality samples at monthly or at best fortnightly frequencies at the majority of monitoring sites (Bowes et al., 2009). This practice is similar to that of the applied frequencies in Australia (Bartley et al., 2012). Table 2.2 summaries common routine sampling schemes used in Queensland, New South Wales and Victoria. Quarterly sampling is the most commonly used sampling frequency in Queensland. In New South Wales and Victoria monthly sampling is the most commonly used sam-

Table 2.1 – Recommended sampling schemes by several studies

Study location	Catchment size	Water quality property	Study findings	Reference
36 sites in Europe and USA	600 - 600000 km^2	Suspended particulate matter	Required sample size is related to catchment size	Moatar et al. (2006)
17 sites within the Thames River catchment, UK.	25 - 1282 km^2	Total phosphorus	Weekly with daily sampling on the highest 35 days of the year	Johnes (2007)
Frome River, Dorset, UK	414 km^2	Total phosphorus, soluble reactive phosphorus, total oxidisable nitrogen and dissolved reactive silicon	Weekly sampling for long term trend analysis and sub-daily for the characterisation of storms	Bowes et al. (2009)
3 sites in Northern Ireland	3 -5 km^2	Total phosphorus	Near continuous sampling is required	Cassidy and Jordan (2011)
3 sites within the Thames River Catchment, UK.	85 - 1400 km^2	Total Reactive Phosphorus and nitrate	Daily or sub-daily sampling	Wade et al. (2012)

ple frequency. As these sampling frequencies often miss key rainfall events (Cassidy and Jordan, 2011; Drewry et al., 2009; Wade et al., 2012), many catchment agencies are using or currently in the process of combining these base-flow samples with event-based sampling.

Few studies have examined the use of probabilistic sampling schemes for base-flow conditions. The use of simple random sampling for annual load estimates was evaluated by Cassidy and Jordan (2011). The results indicated that sample sizes of 365 provided accurate base-flow estimates. However this form of sampling did not capture the exports during events throughout the year (Cassidy and Jordan, 2011). By using other probabilistic sampling schemes which use information on the hydrological processes, improvements in accuracy and reductions in the sample size are possible (Thomas, 1985). In this regard, Arabkhedri et al. (2010) evaluated adaptive cluster sampling to provide unbiased estimates of suspended sediments. Adaptive cluster sampling is commonly used in ecological studies investigating rare populations (de Gruijter et al., 2006; Lohr, 2009). In regard to suspended sediment exports, rainfall events are treated equivalent to rare populations where large quantities of suspended sediment are exported (Arabkhedri et al., 2010). Arabkhedri et al. (2010) explains how this form of sampling uses systematic sampling at regular intervals, and a cluster of samples is collected when stream discharge is found to be above a pre-determined threshold. The results showed adaptive cluster sampling was more accurate than model based methods (Arabkhedri et al., 2010).

2.3.3 Event based sampling

With large proportions of nutrients and sediments being exported during rainfall events, a form of sampling is required to monitor these exports. A study from the UK by Johnes (2007) found that the 5 largest rainfall events contributed 42% of the annual exported total phosphorus. This requirement is more pronounced in Australian catchments which are generally characterised by long base-flow conditions and short sporadic rainfall events (Drewry et al., 2009; Hopmans and Bren, 2007).

Table 2.2 – Example sampling schemes used by some Australian government agencies

State/region	Reporting Agency	Routine sampling frequency	event based sampling	reference
New South Wales	Sydney Catchment Authority	Monthly sampling at most of the 44 sites.	Event-based sampling at a subset of sites	Sydney Catchment Authority (2011)
	Hornsby council	Monthly sampling at most of the 35 sites	Several event sampling sites under construction	Hornsby Council (2011)
	Murrimbidgee Irrigation	Routine monthly sampling at 23 sites	Grab samples when rainfall > 25mm	Murrimbidgee Irrigation Ltd (2012)
Victoria	*Multiple agencies*	Monthly sampling at 280 sites	Event-based sampling at a subset of sites	Smith and Nathan (2012)
Queensland	*Multiple agencies*	Monthly or less frequent samples at 595 sites	Event-based sampling at 13.3%	Department of Environment and Resource Management (2012)

A study based in south eastern Australia found 70% of 6 years of sediments were exported during a single rainfall event (Hopmans and Bren, 2007). Similar to routine sampling, event-based sampling is limited by financial constraints. In addition, the available equipment to collect samples also limits the available sampling schemes. Cassidy and Jordan (2011); Jordan and Cassidy (2011); Wade et al. (2012) used continuously measured total phosphorus at a sub-hourly basis. These experiments provided great insight into the controlling processes in catchments, but they are not typical, are expensive and require access to electricity (Burt et al., 2011; Wade et al., 2012). Catchment managers therefore require the use of less dense sampling schemes which requires consideration of which sampling scheme to use. Both probabilistic and non-probabilistic sampling schemes have been evaluated in terms of the accuracy of event load and event mean estimates. However, similar to that of base-flow, non-probabilistic sampling methods are most commonly used for event-based sampling (Harmel et al., 2006; King and Harmel, 2003). Several probabilistic sampling schemes have been designed in an attempt to improve the accuracy of event load and event mean estimates.

Many catchments are located in remote locations with limited access and no power supply Drewry et al. (2009); Wade et al. (2012). For example, the monitoring sites of the Sydney Catchment Authority have limited access and due to safety reasons after 10-15 mm of rainfall, access to sites is restricted to helicopters (Sydney Catchment Authority, 2011). Without a power supply, monitoring schemes must rely on battery powered sampling devices which use a pump to collect and store samples. Once collected, the samples are often stored in an open container until a field technician can retrieve the samples from the field. This equipment is expensive which has restricted its use by catchment managers (Burt et al., 2011; Drewry et al., 2009; Wade et al., 2012). By using these devices, the sampling scheme is restricted to the options of the software controlling the sampler. This is easily programable without any additional hardware with the commonly used ISCO 3700 automatic sampler (Teledyne ISCO, USA). An alternative method is the use of passive siphon sampling bottles described by Graczyk et al. (2000). These bottles have a one way valve attachment which allows the bottles to be filled once the stream reaches a pre-determined height and restricts the mixing of the sample. By using siphon sampling bottles it is only possible to sample during the rising stage of the hydrograph (Graczyk et al., 2000) at pre-determined heights. This form of sampling is often combined with manual samples during the receding limb of the hydrograph (Drewry et al., 2009). The use of siphon sampling devices requires a model based approach to provide unbiased estimates of the mean and its uncertainity (de Gruijter et al., 2006).

Non-probabilistic sampling schemes are the most common forms of event sampling schemes (Harmel et al., 2006; King and Harmel, 2003). Equally spaced samples in regards to stream discharge and time are the two most common sampling schemes (Harmel et al., 2006; King and Harmel, 2003). Many of these sampling schemes use a pre-determined stream height to commence the sampling, and have a length of time or stream height to reflect the end of the event where sampling stops (Gall et al., 2010; Harmel et al., 2002). Equally spaced temporal samples were used in the study by Hopmans and Bren (2007) to estimate the export of suspended sediments. The same method was used by Birkel et al. (2011) to separate the event and base flow contributions during an event using isotopes. Equally spaced stream discharge sampling has been used by Romeis et al. (2011); Harmel et al. (2009) to sample rainfall events. Two studies examining the effect of bush-fires on sediment and nutrient exports have collected samples based on the stream height levels, during the rising and falling limbs of the event (Smith et al., 2010; Lane et al., 2006). Lopez et al. (2000) outlines how the 1st derivative (*flux*) of stream discharge is a more accurate predictor variable for suspended sediments in small flashy streams. What is often overlooked with these sampling schemes is the requirement of statistical methods which account for the auto-correlation between the samples. Due to the temporal nature of the samples, it should be assumed that there is auto-correlation in the observations. Drewry et al. (2009) presented the auto and cross correlation between stream discharge and total phosphorus, total nitrogen and suspended sediments. Therefore models which are used with these sampling schemes must account for this correlation within the samples (Thomas, 1985, 1988; Crawford, 1991; Cohn et al., 1992; Cooper and Watts, 2002; Cohn, 2005).

Two alternative methods incorporate the acquisition of real time data to improve sampling schemes. Investigating sediment exports, Lewis (1996) used a turbidity sensor to monitor real-time changes and collected samples as pre-determined threshold levels were reached on both the rising and falling limbs of the hydrograph. This sampling scheme is based on the strong correlation between suspended sediments and turbidity (Lewis, 1996). The purpose of this sampling scheme is to optimise sampling to improve the relationship between sediments and turbidity (Lewis, 1996) and is designed to improve predictions of suspended sediments by using continuously measured turbidity (Lewis, 1996). More recently Gall et al. (2010) presented a sampling scheme which uses historical data to predict the receding limb of the hydrograph at the peak of each hydrograph in real time. Gall et al. (2010) outlined how this can be combined with equally spaced samples in relation to stream discharge. This sampling, combined with the sampling at pre-determined stream heights, ensures the rising limb of each event is captured (Gall et al., 2010). These two sampling schemes outline how the incorporation of real time data for variables related to water quality can be used to improve the sampling schemes. These methods are model-based and require appropriate models to provide unbiased estimates.

Several studies have investigated probabilistic sampling schemes for event-based sampling (Thomas, 1985, 1988; Thomas and Lewis, 1993, 1995), however these methods often require equipment in addition to automatic samplers (Thomas and Lewis, 1993). Such methods allow for unbiased estimates without the requirement of models to account for biased sampling schemes and auto-correlation. Probabilistic methods which have been assessed include; selection-at-list-time (Thomas, 1985), time stratified sampling (Thomas and Lewis, 1993) and flow stratified sampling (Thomas and Lewis, 1995). Selection-at-list-time is an improvement of probability proportional to size sampling which uses a list of random numbers to determine when a sample should be taken in relation to stream discharge in real time (Thomas, 1985). The time stratified sampling method outlined by Thomas and Lewis (1993) divides the hydrograph into strata of different lengths designed to concentrate sampling during large changes in stream discharge. Flow-stratified sampling stratifies the event based on the stream height and direction of the stream hydrograph (Thomas and Lewis, 1995). An evaluation of the three methods found time-stratified sampling was the most accurate of the three methods (Thomas and Lewis, 1995). The results of these studies have shown that it is possible to sample key areas of event hydrographs and obtain accurate event means (Thomas and Lewis, 1995). These sampling schemes have received little attention which may be due to the focus on sediment exports and not nutrients exports and the perceived complexity of the methods.

2.4 Common load estimation methods

As continuous sampling is not feasible for most monitoring schemes, catchment studies require methods to estimate water quality during unsampled periods. Several studies have examined the accuracy of these methods for different water quality variables. These estimation methods are generally assessed on the accuracy of load estimates over a given period (e.g. annually). However, there has been no clear outstanding method from these studies, and Marsh and Waters (2009) evaluated 34 different methods in a single study. There are three main types of estimation methods; averaging, ratio and regression based methods (Marsh and Waters, 2009). The assumption of simple random sampling is a requirement of these commonly used estimation methods (Cohn et al., 1992; Thomas, 1985), which is generally overlooked. Recently, studies have examined alternative flexible estimation methods which do not require a linear trend between water quality and the explanatory variables. Several studies have examined and compared the three load estimation methods (Aulenbach and Hooper, 2006; Cassidy and Jordan, 2011; Horowitz, 2003; Johnes, 2007; Quilbé et al., 2006; Salles et al., 2008; Webb et al., 1997). The results of these studies have provided insights into the effects of the different methods on the precision and bias of annual load estimates. These studies have found similar levels of accuracy between the methods and a general underestimation of annual loads. For example, Cassidy and Jordan (2011) found that annual total phosphorus was underestimated by up to 60% and

Webb et al. (1997) found annual suspended sediment was underestimated by up to 57.2% by load estimation methods.

Retrospective load calculations are based on calculating the mass of nutrients/sediments which are exported over a given period (Aulenbach and Hooper, 2006). The relationship between stream discharge and nutrient/sediment loading is widely used in the literature, and is represented by the equation;

$$L = \int \mathbf{C}_t \mathbf{Q}_t dt \tag{2.1}$$

where the total load (L) is the product of the solute concentration ($\mathbf{C}$) and the discharge ($\mathbf{Q}$) over time (t) (Aulenbach and Hooper, 2006). Eq. 2.1, assumes that there is a constant ($\sim$15 minutes) record of both discharge and the concentration. Due to the expensive nature of water quality monitoring this equation is rarely used without the prediction of water quality concentration.

Many load estimation methods assume the data was collected using simple random sampling (Cohn et al., 1992; Thomas, 1985). If the sampling was not random, the model must examine the data for the potential of temporal auto-correlation (de Gruijter et al., 2006; Lark and Cullis, 2004). If the model does not account for the temporal auto-correlation the uncertainty estimates may be biased (de Gruijter et al., 2006; Lark and Cullis, 2004). However, due to the temporal nature of the water quality data, the presence of temporal auto-correlation is likely.

Ideally, a load estimation would not require probabilistic sampling and would provide the ability to account for the auto-correlation between observations.

2.4.1 Averaging estimation methods

Averaging methods are the simplest form of load estimation methods. These methods use the average concentrations multiplied by the average discharge over the same period (Aulenbach and Hooper, 2006). An example of this method is;

$$\hat{L} = K\bar{c}\bar{q} \tag{2.2}$$

where $\bar{c}$ and $\bar{q}$ is the average concentration and average discharge respectively for a certain period of time and K is a unit-conversion factor (Cooper and Watts, 2002). Salles et al. (2008) found averaging methods gave accurate load estimates when sampling frequencies were sub-daily. Load estimates using these methods tend to underestimate annual loads where sampling frequencies are large as rainfall events are often missed (Cooper and Watts, 2002; Quilbé et al., 2006). It is important to note that these methods assume the use of simple random sampling (Cooper and Watts, 2002; de Gruijter et al., 2006). If this assumption is not met, which is often the case then the estimators can be highly biased (Cooper and Watts, 2002; de Gruijter et al., 2006).

2.4.2 Ratio estimation methods

Ratio estimators are designed to improve estimations by including all observed discharge observations. Therefore this method does not only rely on the discharge coincident with WQ samples but all collected discharge values. The most simplified form of this calculation is:

$$\hat{L} = \frac{\bar{l}}{\bar{q}}Q \tag{2.3}$$

where $\bar{l}$ and $\bar{q}$ are the average of concentration and discharge respectively, and Q is the total discharge (Cooper and Watts, 2002). This simple form of ratio-estimator will lead to biased load estimates (Cooper and Watts, 2002). Several correction methods have been developed to minimise this bias (Cooper and Watts, 2002). One of the most commonly used ratio-estimators for load estimates is the Beale-ratio estimator equation (Beale, 1962)

$$\hat{L} = \frac{\bar{l}}{\bar{q}}Q\left(\frac{1+\frac{1}{n}\frac{cov(l,q)}{\bar{l}\bar{q}}}{1+\frac{1}{n}\frac{var(q)}{\bar{q}^2}}\right). \tag{2.4}$$

The Beale ratio estimator has been found to provide almost unbiased estimates (Cooper and Watts, 2002). Ratio estimators have been recommended in situations where no linear relationship between the concentration and stream discharge is present (Quilbé et al., 2006). Evaluating several estimation methods based on annual total phosphorus exports in the UK, Johnes (2007) found ratio estimators to be the most accurate methods in catchments with low population density and a high base flow index, based on weekly sampling frequencies. However, these methods performed poorly with less frequent monthly observations Johnes (2007). However, ratio-estimators assume a form of probabilistic sampling has been used (de Gruijter et al., 2006). Citing Royall and Cumberland (1981); Cooper and Watts (2002) suggest that probabilistic sampling is not required if the model is a straight line through the origin. In addition to this, ratio-estimators require the concentration to be proportional to that of the discharge and that the concentration variance increases as the discharge variance increases (Cooper and Watts, 2002; de Gruijter et al., 2006).

2.4.3 Regression estimation methods

Regression estimation methods (rating curves) are often used to model the linear relationship between log concentration and log stream discharge. Using the continuous stream discharge, concentration is predicted (Cassidy and Jordan, 2011). The regression method is;

$$c(\mathbf{t}) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2.5}$$

where $c(\mathbf{t})$ is the log transformed concentration, $\mathbf{X}$ is the trend matrix based on the log transformed stream discharge and $\boldsymbol{\beta}$ is a vector of the model coefficients. The error term of the regression model $\boldsymbol{\epsilon}$; is a vector of independent, normally distributed errors. Several studies have investigated the use of regression based methods (Aulenbach and Hooper, 2006; Cooper and Watts, 2002; Horowitz, 2003; Johnes, 2007; Quilbé et al., 2006). These methods are known to underestimate suspended sediment during events and overestimate suspended sediment during base-flow periods (Horowitz, 2003). Cooper and Watts (2002) found poor estimates based on a single year of monthly observations, but accurate estimates using several years of observations. Regression methods have been suggested to be used in situations where strong correlations are present (i.e. $r^2 \geq 0.5$)(Quilbé et al., 2006). In addition Cooper and Watts (2002) found the estimates of these regression methods are unbiased and accurate where the model was correct.

Regression models are often fitted using ordinary-least-squares, however this assumes the errors are *iid* with the distribution; $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2)$ (Lark and Cullis, 2004). If simple random sampling has not been used the estimates will be biased (Lark and Cullis, 2004) and this requires the model to examine and account for temporal autocorrelation between the samples (de Gruijter et al., 2006; Lark and Cullis, 2004; Thomas, 1985).

2.4.4 Generalised additive models

Recently several studies have used Generalised additive models (GAMs) to model water quality (Kuhnert et al., 2012; Morton and Henderson, 2008; Smith et al., 2013; Wang et al., 2011). These studies have shown how GAMs offer the ability to account for the complexity of the non-linear nature of water quality data (Morton and Henderson, 2008). Morton and Henderson (2008) also note that the complexity can be further accentuated by errors from data collection or laboratory analysis. GAMs use smoothing functions to model the relationship between the explanatory variables and water quality (Wood, 2006), this provides a flexible method to account for the complexity of many non-linear water quality processes (Morton and Henderson, 2008).

GAMs have been used for a variety of applications. For example, Smith et al. (2013) used GAMs to examine long term trends of nitrogen and phosphorus in three catchments with dairy farming. Using stream discharge, year, and month as covariates the study found a trend between milk production and nitrogen and phosphorus exports (Smith et al., 2013). Wang et al. (2011) used a GAM with several stream discharge derived variables to estimate suspended sediments and nitrogen in north eastern Australia. Using this approach Kuhnert et al. (2012) included additional covariates to improve the estimates. In addition, Kuhnert et al. (2012) also outlined how to provide unbiased confidence intervals of the predictions.

2.4.5 Accounting for auto-correlation within observations

The form of statistics used to analyse the sampled observations is determined by the aim of the monitoring scheme and sample design adopted. If the water quality monitoring scheme used a systematic sampling scheme with regular temporal intervals then a standard auto-regressive time series analysis would suffice. However, since many sampling schemes use a combination of systematic sampling schemes and event-based sampling, standard time series statistical methods are not appropriate. As the samples are sampled through time, the potential of auto-correlation between the samples cannot be ignored,as it can be highly likely that the current state is in some way affected by the previous state of the system through time. Therefore the samples must be examined for auto-correlation. There are two main approaches used to account for the auto-correlation between samples. The first method includes a temporal component in the estimation method (Cohn et al., 1992; Diebel et al., 2009; Guo et al., 2002; Kuhnert et al., 2012; Smith et al., 2013; Wang and Tian, 2013). The second method uses traditional time series based methods in the form of auto-regressive models are also used to account for the auto-correlation between observations (Kuhnert et al., 2012; Morton and Henderson, 2008; Ömer Faruk, 2010; Wang et al., 2011).

Several studies have included temporal components such as seasonal components in models to account for temporal changes (Rode and Suhr, 2007). It is important to consider these seasonal terms in areas which have seasonal controlling factors (e.g. snowmelt, tropical climates) (Kuhnert et al., 2012). This approach generally includes an additional term to the estimation model which accounts for the seasonality of the variables. Cohn et al. (1992) introduced a sinusoidal component in a simple regression model to account for seasonality in the form of two terms:

$$\beta_5 \sin(2\pi T) + \beta_6 \cos(2\pi T), \tag{2.6}$$

where T is the time of year in decimal years since the origin. These two additional terms were used in a linear regression model, similar to equation 2.5 to predict the log transformed concentration (Cohn et al., 1992). An example of a time series produced using these two terms is shown in figure 2.1. This seasonal term can be thought of as a surrogate to represent a variety of different seasonal processes in the catchment. Cohn et al. (1992) also explains how the amplitude of the peak and the day of the year that the peak day is can be set using β_5 and β_6. The inclusion of this term can improve the estimation of the seasonality in the observed data if this exists. This approach has been undertaken by several other studies (Diebel et al., 2009; Guo et al., 2002; Wang and Tian, 2013). Another method is to include decimal days into the model (Diebel et al., 2009; Smith et al., 2013) and a term to represent the month of the year (Smith et al., 2013). The study by Smith et al. (2013) found the inclusion of a term representing the month of the year increased the adjusted r^2 from 79 to 84.3%. These studies have shown how the inclusion of terms representing the time of year or seasonality can increase the accuracy of the estimates. However, it is important to distinguish that these methods rely on additional model terms to account for temporal correlation, and do not attempt to use a model to directly account for the auto-correlation between the observations.

Correlation between observations is based on the linear association between the two samples at different points in time. Samples collected through time are often serially correlated, which implies samples which are closer to each other in time are more correlated than samples further apart (Heuvelink et al., 2010). An example of a temporally correlated process is provided by Diggle (1990). Figure 2.2.a. presents hourly luteinizing hormone blood samples. An auto-correlation function is often used to examine the auto-correlation between samples at differing temporal lags. The lag k auto-correlation is defined by

$$\rho_k = \frac{(z_t - \mu_z)(z_{t+k} - \mu_z)}{\sigma_z^2}, \tag{2.7}$$

where z_t is the observation at time t, μ_z and σ_z^2 are the mean and variance of z. Figure 2.2.b. shows the auto-correlation of the process shown in 2.2.a.. The par-

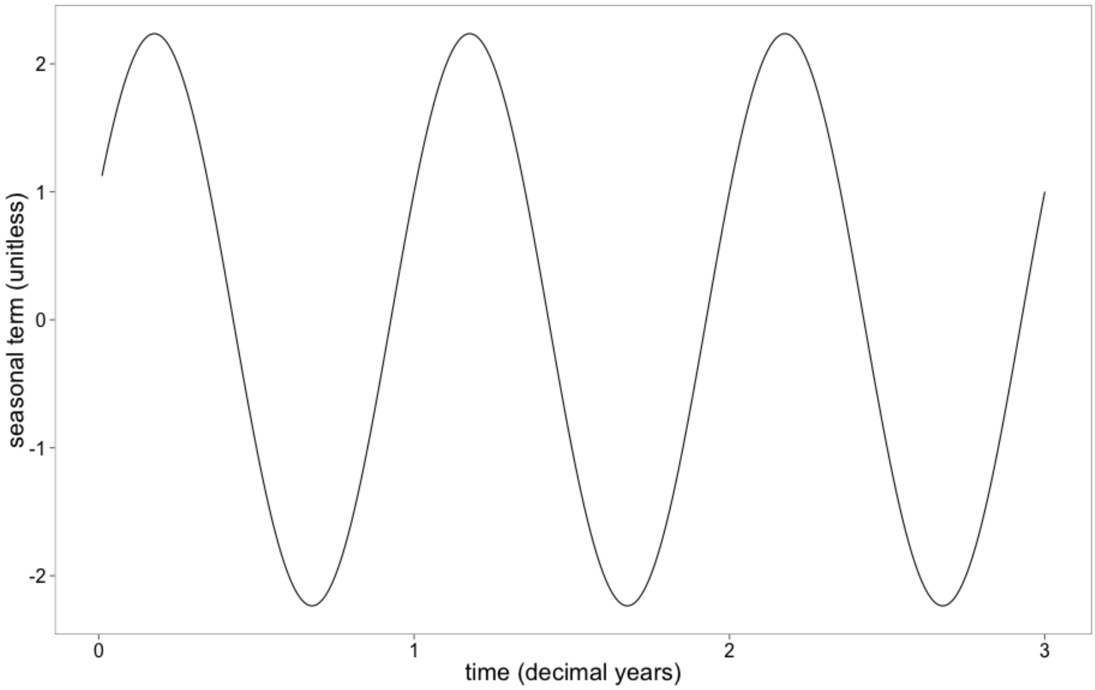

Figure 2.1 – An example of the seasonal term over three years based on equation 2.6

tial auto-correlation function is similar to the auto-correlation function yet examines the correlations between observations, k, time intervals apart while removing the impact of intervening observations and is shown in figure 2.2.c.. Based on this auto-correlation between the samples through time is the use of an auto-regressive (AR) model. These models make estimates against those observations that are k time intervals apart. An example of this is represented by an AR-1 model:

$$Z_{(t+1)} = \mu + a\{Z_{(t)} - \mu\} + \epsilon_{(t)}, \tag{2.8}$$

where a is the AR parameter, $Z_{(t)}$ is a value of a process at a discrete time, t and μ is the mean of the process, with an absolute value < 1 describing how $Z_{(t)}$ effects $Z_{(t+1)}$ and $\epsilon_{(t)}$ is an uncorrelated residual with zero mean and variance.

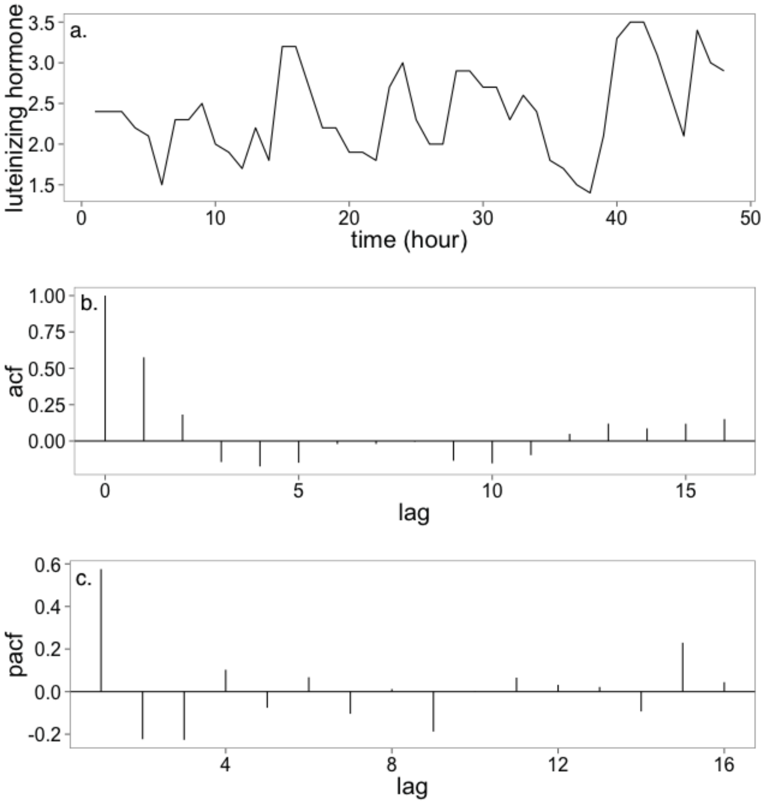

Figure 2.2 – An example of an auto-correlated time series (a) and the auto-correlation between samples based on an auto-correlation function.

For most studies it is not possible to use auto-regressive based methods to model water quality as the samples are often collected using a combination of event-based and systematic sampling. Under this limitation, studies must average samples into larger time intervals (Morton and Henderson, 2008). Using this method, the averaging of observations to a monthly interval would suffice for long term trends, but short term trends would require sub-daily to daily sampling (Morton and Henderson, 2008). An alternative method, which several studies have applied, is the use of auto-regressive models to account for the auto-correlation in the residuals of the water quality predictions (Kuhnert et al., 2012; Morton and Henderson, 2008; Wang et al., 2011). The study of Smith et al. (2013) found auto-correlation in the residuals, however noted the reduction of auto-correlation by the inclusion of a temporal term reduced the auto-correlation. Another study by Cohn et al. (1992) found temporal auto-correlation in the residuals of 23 out of 24 models. These findings re-enforce the importance of examining the data for temporal auto-correlation.

2.5 Explanatory variables other than stream discharge

Most estimation based models use stream discharge as a predictor of water quality concentration, due to the availability and the general log linear relationship. However, many water quality variables are not always directly related to stream discharge (Lewis, 1996). One major limitation of using stream discharge as an explanatory variable is the inability to account for hysteresis during events (Eder et al., 2010). In addition, short term diurnal (Cassidy and Jordan, 2011), seasonal changes (Cohn et al., 1992) and different transport flow-paths increase the complexity of the relationship between stream discharge and water quality. Under these restrictions, several studies have investigated the use and potential improvement of other predictors for estimating water quality variables (Cohn et al., 1992; Kim and Furumai, 2012; Krueger et al., 2009; Lewis, 1996; Mano et al., 2009; Wang et al., 2011). These additional explanatory variables fall into two broad categories; stream discharge derived and indirect measures of water quality.

In an attempt to account for different hydrological controlling processes, several papers have derived explanatory variables from stream discharge (Cohn et al., 1992; Krueger et al., 2009; Mano et al., 2009; Wang et al., 2011). Cohn et al. (1992) introduced a quadratic term to model the relationship between water quality and stream discharge. More recently additional variables have been included to account for the rising and falling limbs during events (Krueger et al., 2009) and to account for the hysteresis within each event. This was also combined with an additional variable to represent each individual event. Daily stream discharge characteristics such as mean and maximum stream discharge were included in a regression model to reflect daily stream variations (Mano et al., 2009). The term *first-flush* refers to the higher levels of nutrient and sediment exports during the first rainfall event after a long dry season (Wang et al., 2011). To account for this phenomenon, Wang et al. (2011) introduces an additional term which uses a weighted average of recent discharge measurements to reflect the state of the catchment. Although the inclusion of these terms has had mixed results, the significance of these terms to estimate water quality provides insight into the controlling hydrological processes.

Turbidity is often used instead of stream discharge for improving estimates of suspended sediments (Lewis, 1996; Sun et al., 2001; Jastram et al., 2010; Jones et al., 2011). Strong correlations between turbidity and total phosphorus (Jones et al., 2011) and total nitrogen (Kim and Furumai, 2012) have also been identified. The use of *in situ* turbidity sensors is becoming more popular due to the relatively low cost and limited maintenance requirements, with several sensors using wipers to ensure a clean surface. The use of *in situ* turbidity measuring devices should be strongly considered to improve estimates of sediment related variables until alternative direct measuring devices become affordable (Jones et al., 2011). Based on this strong relationship, *in situ* turbidity monitoring is used throughout the state of California to monitor the impacts of forestry activities (Harris et al., 2007). It is important to consider the differences between the commercial sensors which will lead to differences in turbidity readings (Harris et al., 2007; Sun et al., 2001). Under this constraint it may be necessary to develop site, instrument and storm specific relationships for estimation (Harris et al., 2007; Sun et al., 2001).

Seasonality and related processes have also beeen included as explanatory variables (Cohn et al., 1992; Kuhnert et al., 2012; Wang et al., 2011). Cohn et al. (1992) included a component to account for seasonal trends. In a catchment in tropical Australia, the study of Wang et al. (2011) found no signifiant effect of a seasonal term. Following the work of Wang et al. (2011), Kuhnert et al. (2012) found the inclusion of a term reflecting the amount of bare ground was a significant explanatory variable. The results indicated a 2.1% decrease in suspended sediment exports with each percentage increase in vegetation cover Kuhnert et al. (2012).

2.6 Review summary and project outline

This literature review has summarised commonly applied water quality monitoring schemes. The accuracy of these monitoring schemes are affected by the form of sampling and statistical analysis. Sample size recommendations of several studies have suggested the use of daily/sub-daily sampling. However, the sampling frequencies of many government agencies is much lower due to financial constraints. Compensating for this low sample frequency, government agencies often include event-based sampling to capture periods of large exports. The combination of routine and event samples restricts the types of statistical analysis which can be used. The three common load estimation methods assume a form of probabilistic sampling. In addition, these estimation methods do not account for the temporal auto-correlation between samples.

Systematic sampling is the most widely used water quality sampling scheme. This form of sampling often uses a monthly interval, however it often misses key rainfall events, which are associated with large nutrient and sediment exports. Increasingly, event-based sampling is being used to sample these periods. However the combination of routine and event-based samples increases the complexity of the statistical analysis required. As event-based exports contribute large proportions of annual load exports it is important to use an appropriate sampling scheme. Several probabilistic sampling schemes have shown how these methods provide an easy method to derive unbiased event-based statistics. These methods have received little attention with the majority of event-based sampling schemes being systematic.

One of the largest financial constraints on water quality monitoring programs are laboratory costs. Due to these costs, most government agencies use monthly sampling. This form of sampling does not reflect the current literature which suggests the use of daily/sub-daily sampling. With this difference between scientific literature and government practices, future scientific research must consider this difference to provide meaningful results to catchment managers. Many catchment monitoring programmes have access to limited historical data, however there are some tools available which can provide meaningful information from this data.

As many of the applied sampling schemes use systematic sampling and often combine this with event-based sampling the load estimation method must accommodate these forms of sampling. Unfortunately the common sampling schemes do not meet the assumptions of the three most commonly used load estimation methods. These estimation methods assume the use of probabilistic sampling. Recently, studies have investigated the use of GAMs to estimate water quality loads using explanatory variables. A major benefit of GAMs is the ability to use smoothing functions to represent the relationship between water quality and explanatory variables. The inclusion of more affordable variables which reflect water quality has been shown to reduce the uncertainty of estimates without the requirement of additional samples.

Many studies have investigated the impact of sample frequencies and estimation methods on the effect of load estimates. However, these studies have not taken into account the use of additional low cost explanatory variables. Several studies have examined the inclusion of various explanatory variables which reflect the hydrological processes of water quality. These studies have shown how the inclusion of these variables in estimation methods increases the accuracy of water quality estimates. Therefore the use of explanatory variables increases the accuracy of the estimates without increasing sample sizes.

As water quality samples are collected through time, the potential of temporal autocorrelation between samples cannot be ignored. Very few studies have accounted for the auto-correlation between samples. By combining routine and event-based sampling it is not possible to use auto-regressive methods, therefore several studies have included temporal components in regression models and GAMs. Studies have examined the residuals of these models for auto-correlation. These studies have identified auto-correlation in the residuals and recommended that load estimation methods examine the residuals for this correlation.

The accuracy of water quality monitoring programmes are heavily affected by the form of sampling and the statistics used to summarise the data. This review has highlighted the current methods of water quality monitoring, and the limitations of these methods. The objectives are to develop and investigate;

1. Methods to evaluate the importance of event-based sampling using historical data is required (chapter 3).
2. Catchment managers require a method to estimate the effect of sample sizes on the uncertainty of mean estimates using historical data (chapter 4).
3. A new stratified sampling scheme to provide unbiased water quality event means without the use of load estimation models (chapter 5).
4. Using linear mixed models to incorporate additional explanatory variables and account for the temporal auto-correlation between samples (chapter 6).

2.7 References

ANZECC, 2000. Australian and New Zealand guidelines for fresh and marine water quality. Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand, Canberra, 1–103.

Arabkhedri, M., Lai, F., Noor-Akma, I., Mohamad-Roslan, M., 2010. An application of adaptive cluster sampling for estimating total suspended sediment load. Hydrology Research 41 (1), 63–73.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., Williams, J. R. 1998. Large area hydrologic modeling and assessment part I: Model development1. JAWRA Journal of the American Water Resources Association, 34(1), 73-89.

Aulenbach, B., Hooper, R., 2006. The composite method: an improved method for stream-water solute load estimation. Hydrological Processes 20 (14), 3029–3047.

Bartley, R., Speirs, W., Ellis, T., Waters, D., 2012. A review of sediment and nutrient concentration data from Australia for use in catchment water quality models. Marine pollution bulletin 65 (4), 101–116.

Beale, E., 1962. Some uses of computers in operational research. Industrielle Organisation 31 (1), 27–28.

Birkel, C., Soulsby, C., Tetzlaff, D., Dunn, S., Spezia, L., 2011. High-frequency storm event isotope sampling reveals time-variant transit time distributions and influence of diurnal cycles. Hydrological Processes 26 (2), 308–316.

Bowes, M., Smith, J., Neal, C., 2009. The value of high-resolution nutrient monitoring: A case study of the River Frome, Dorset, UK. Journal of Hydrology 378 (1), 82–96.

Burt, T., Howden, N., Worrall, F., McDonnell, J., 2011. On the value of long-term, low-frequency water quality sampling: avoiding throwing the baby out with the bathwater. Hydrological Processes 25 (5), 828–830.

Brown, L. C., and Barnwell, B.O., 1987. The enhanced stream water quality models QUAL2E and QUAL2E-UNCAS: documentation and user manual. US Environmental Protection Agency. Office of Research and Development. Environmental Research Laboratory.

Cassidy, R., Jordan, P., 2011. Limitations of instantaneous water quality sampling in surface-water catchments: Comparison with near-continuous phosphorus time-series data. Journal of Hydrology 405 (1), 182–193.

Chen, D., Dahlgren, R. A., & Lu, J. 2013. A modified load apportionment model for identifying point and diffuse source nutrient inputs to rivers from stream monitoring data. Journal of Hydrology, 501, 25–34.

Chang, H. 2008. Spatial Analysis of Water Quality Trends in the Han River Basin, South Korea. Water Research 42 (13) (July): 3285–304.

Cohn, T., 2005. Estimating contaminant loads in rivers: An application of adjusted maximum likelihood to type 1 censored data. Water Resources Research 41 (7), W07003.

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

Cooper, D., Watts, C., 2002. A comparison of river load estimation techniques: application to dissolved organic carbon. Environmetrics 13 (7), 733–750.

Crawford, C., 1991. Estimation of suspended-sediment rating curves and mean suspended-sediment loads. Journal of Hydrology 129 (1-4), 331–348.

de Gruijter, J., Brus, D., Bierkens, M., Knotters, M., et al., 2006. Sampling for natural resource monitoring. Springer Verlag.

Department of Environment and Resource Management, 2012. Surface water ambient water quality network.
http://www.derm.qld.gov.au/water/monitoring/pdf/surfacewater_ambient2012.pdf

Diebel, M., Maxted, J., Robertson, D., Han, S., Vander Zanden, M., 2009. Landscape planning for agricultural nonpoint source pollution reduction iii: assessing phosphorus and sediment reduction potential. Environmental Management 43 (1), 69–83.

Diggle, PJ. 1990. Time series: a biostatistical introduction. Oxford University Press.

Drewry, J., Newham, L., Croke, B., 2009. Suspended sediment, nitrogen and phosphorus concentrations and exports during storm-events to the Tuross Estuary, Australia. Journal of Environmental Management 90 (2), 879–887.

Eder, A., Strauss, P., Krueger, T., Quinton, J., 2010. Comparative calculation of suspended sediment loads with respect to hysteresis effects (in the Petzenkirchen catchment, Austria). Journal of Hydrology 389 (1), 168–176.

Gall, H., Jafvert, C., Jenkinson, B., 2010. Integrating hydrograph modeling with real-time flow monitoring to generate hydrograph-specific sampling schemes. Journal of Hydrology 393 (3), 331–340.

Graczyk, D., Robertson, D., Rose, W., Steuer, J., 2000. Comparison of water-quality samples collected by siphon samplers and automatic samplers in Wisconsin. USGS Fact Sheet FS-067-00. USGS, Washington, DC.

Guo, Y., Markus, M., Demissie, M., 2002. Uncertainty of nitrate-N load computations for agricultural watersheds. Water Resources Research 38 (10), 1185.

Harmel, R., Cooper, R., Slade, R., Haney, R., Arnold, J., 2006. Cumulative uncertainty in measured streamflow and water quality data for small watersheds. Transactions of the ASAE (American Society of Agricultural Engineers) 49 (3), 689.

Harmel, R., King, K., Wolfe, J., Torbert, H., 2002. Minimum flow considerations for automated storm sampling on small watersheds. Texas Journal of Science 54 (2), 177–188.

Harmel, R., Smith, D., King, K., Slade, R., 2009. Estimating storm discharge and water quality data uncertainty: A software tool for monitoring and modeling applications. Environmental Modelling & Software 24 (7), 832–842.

Harris, R., Sullivan, K., Cafferata, P., Munn, J., Faucher, K., 2007. Applications of turbidity monitoring to forest management in California. Environmental Management 40 (3), 531–543.

Hopmans, P., Bren, L., 2007. Long-term changes in water quality and solute exports in headwater streams of intensively managed radiata pine and natural eucalypt forest catchments in south-eastern Australia. Forest Ecology and Management 253 (1-3), 244–261.

Hornsby Council, 2011. Annual water quality report – 2010-2011. http://www.hornsby.nsw.gov.au/__data/assets/pdf_file/0016/33343/2010-2011-water-quality-report-FINAL-19Dec2011.pdf

Horowitz, A., 2003. An evaluation of sediment rating curves for estimating suspended sediment concentrations for subsequent flux calculations. Hydrological Processes 17 (17), 3387–3409.

Heuvelink, G.B.M., Hoogland, T., Walvoort, D.J.J. 2010. A disposition of interpolation techniques. Wettelijke Onderzoekstaken Natuur & Milieu.

Jastram, J., Zipper, C., Zelazny, L., Hyer, K., 2010. Increasing precision of turbidity-based suspended sediment concentration and load estimates. Journal of Environmental Quality 39 (4), 1306–1316.

Johnes, P., 2007. Uncertainties in annual riverine phosphorus load estimation: impact of load estimation methodology, sampling frequency, baseflow index and catchment population density. Journal of Hydrology 332 (1-2), 241–258.

Jones, A., Stevens, D., Horsburgh, J., Mesner, N., 2011. Surrogate measures for providing high frequency estimates of total suspended solids and total phosphorus concentrations. JAWRA Journal of the American Water Resources Association 47 (2), 239–253.

Jordan, P., Cassidy, R., 2011. Technical note: Assessing a 24/7 solution for monitoring water quality loads in small river catchments. Hydrology and Earth System Sciences. 15, 3093–3100.

Kim, J., Furumai, H., 2012. Improved calibration of a rainfall-pollutant-runoff model using turbidity and electrical conductivity as surrogate parameters for total nitrogen. Water and Environment Journal (27) 79 – 85.

King, K., Harmel, R., 2003. Considerations in selecting a water quality sampling strategy. Transactions of the ASAE 46 (1), 63–73.

Krueger, T., Quinton, J., Freer, J., Macleod, C., Bilotta, G., Brazier, R., Butler, P., Haygarth, P., 2009. Uncertainties in data and models to describe event dynamics of agricultural sediment and phosphorus transfer. Journal of Environmental Quality 38 (3), 1137–1148.

Kuhnert, P., Henderson, B., Lewis, S., Bainbridge, Z., Wilkinson, S., Brodie, J., 2012. Quantifying total suspended sediment export from the Burdekin river catchment using the loads regression estimator tool. Water Resources Research 48 (4), W04533.

Lane, P., Sheridan, G., Noske, P., 2006. Changes in sediment loads and discharge from small mountain catchments following wildfire in South Eastern Australia. Journal of Hydrology 331 (3), 495–510.

Lam, Q. D., Schmalz, B., & Fohrer, N. 2010. Modelling point and diffuse source pollution of nitrate in a rural lowland catchment using the SWAT model. Agricultural Water Management, 97(2), 317–325.

Lark, R., Cullis, B., 2004. Model-based analysis using reml for inference from systematically sampled data on soil. European Journal of Soil Science 55 (4), 799–813.

Lewis, J., 1996. Turbidity-controlled suspended sediment sampling for runoff-event load estimation. Water Resources Research 32 (7), 2299–2310.

Lohr, S., 2009. Sampling: design and analysis. Cengage Learning.

Lopez, E., Soto, B., Rubinos, D., Diaz-Fierros, F., 2000. Flow-variation-paced sampling: a method for automatic sampling of streamflow during peak runoff periods. Journal of Hydrology 229 (3), 255–264.

Madrid, Y., Zayas, Z., 2007. Water sampling: traditional methods and new approaches in water sampling strategy. TrAC Trends in Analytical Chemistry 26 (4), 293–299.

Mano, V., Nemery, J., Belleudy, P., Poirel, A., 2009. Assessment of suspended sediment transport in four alpine watersheds (France): influence of the climatic regime. Hydrological Processes 23 (5), 777–792.

Marsh, N., Waters, D., 2009. Comparison of load estimation methods and their associated error. In: 18th IMACS World Congress MODSIM. Vol. 4. Cairns.

Moatar, F., Person, G., Meybeck, M., Coynel, A., Etcheber, H., Crouzet, P., 2006. The influence of contrasting suspended particulate matter transport regimes on the bias and precision of flux estimates. Science of the Total Environment 370 (2), 515–531.

Moosmann, L., Müller, B., Gächter, R., Wüest, A., Butscher, E., Herzog, P., 2005. Trend-oriented sampling strategy and estimation of soluble reactive phosphorus loads in streams. Water Resources Research 41 (1), W01020.

Morton, R., Henderson, B., 2008. Estimation of nonlinear trends in water quality: An improved approach using generalized additive models. Water Resources Research 44 (7), W07420.

Murrimbidgee Irrigation Ltd, 2012. Murrumbidgee irrigation Ltd. water quality framework.
http://www.mirrigation.com.au/ArticleDocuments/165/Water%20Quality%20EPA%20Reporting%2

Newham, L., Croke, B., Jakeman, A., 2001. Design of water quality monitoring programs and automatic sampling techniques. In: Proceedings of the third Australian stream management conference. Cooperative Research Centre for Catchment Hydrology. pp. 455–460.

Ömer Faruk, D., 2010. A hybrid neural network and ARIMA model for water quality time series prediction. Engineering Applications of Artificial Intelligence 23 (4), 586–594.

Palmer, N., 2012. Changing perception of the value of urban water in Australia following investment in seawater desalination. Desalination and Water Treatment 43 (1-3), 298–307.

Quilbé, R., Rousseau, A., Duchemin, M., Poulin, A., Gangbazo, G., Villeneuve, J., 2006. Selecting a calculation method to estimate sediment and nutrient loads in streams: Application to the Beaurivage River (Quebec, Canada). Journal of Hydrology 326 (1), 295–310.

Rode, M., Suhr, U. 2007. Uncertainties in selected river water quality data. Hydrology and Earth System Sciences Discussions, 11(2), 863-874.

Romeis, J., Jackson, C., Risse, L., Sharpley, A., Radcliffe, D., 2011. Hydrologic and phosphorus export behaviour of small streams in commercial poultry-pasture watersheds1. JAWRA Journal of the American Water Resources Association 47 (2), 367–385.

Royall, R., Cumberland, W., 1981. An empirical study of the ratio estimator and estimators of its variance. Journal of the American Statistical Association 76 (373), 66–77.

Salles, C., Tournoud, M., Chu, Y., 2008. Estimating nutrient and sediment flood loads in a small Mediterranean river. Hydrological Processes 22 (2), 242–253.

Smith, A. P., Western, A. W., Hannah, M. C., 2013. Linking water quality trends with land use intensification in dairy farming catchments. Journal of Hydrology 476 (0), 1 – 12.
http://www.sciencedirect.com/science/article/pii/S0022169412008967

Smith, H., Sheridan, G., Lane, P., Sherwin, C., 2010. Paired eucalyptus forest catchment study of prescribed fire effects on suspended sediment and nutrient exports in south-eastern Australia. International Journal of Wildland Fire 19 (5), 624–636.

Smith, W., Nathan, R., 2012. Victorian water quality monitoring network trend analysis. http://www.vicwaterdata.net/vicwaterdata/publishedDocs/docs/East%20Gippsland.pdf

Sun, H., Cornish, P., Daniell, T., 2001. Turbidity-based erosion estimation in a catchment in south Australia. Journal of Hydrology 253 (1), 227–238.

Sydney Catchment Authority, 2008a. Annual water quality monitoring report annual water quality monitoring report 2007-08. Annual report, Sydney Catchment Authority, Sydney, Australia.

Sydney Catchment Authority, 2008b. Sca annual catchment management report 2007-08. Annual report, Sydney Catchment Authority, Sydney, Australia.

Sydney Catchment Authority, 2011. Annual water quality monitoring report annual water quality monitoring report 2010-11. Annual report, Sydney Catchment Authority, Sydney, Australia.

Sydney Catchment Authority, 2012. Healthy catchments strategy 2009 -2012. Tech. rep., Sydney Catchment Authority, Sydney, Australia.

Thomas, R., 1985. Estimating total suspended sediment yield with probability sampling. Water Resources Research 21 (9), 1381–1388.

Thomas, R., 1988. Monitoring baseline suspended sediment in forested basins: the effects of sampling. Hydrological Sciences Journal 33 (5), 499–514.

Thomas, R., Lewis, J., 1993. A comparison of selection at list time and time-stratified sampling for estimating suspended sediment loads. Water Resources Research 29 (4), 1247–1256.

Thomas, R., Lewis, J., 1995. An evaluation of flow-stratified sampling for estimating suspended sediment loads. Journal of Hydrology 170 (1-4), 27–45.

Wade, A., Palmer-Felgate, E., Halliday, S., Skeffington, R., Loewenthal, M., Jarvie, H., Bowes, M., Greenway, G., Haswell, S., Bell, I., et al., 2012. Hydrochemical processes in lowland rivers: insights from in situ, high-resolution monitoring. Hydrology and Earth System Sciences 16 (11), 4323.

Wang, Y., Kuhnert, P., Henderson, B., 2011. Load estimation with uncertainties from opportunistic sampling data–a semiparametric approach. Journal of Hydrology 396 (1), 148–157.

Wang, Y., Tian, T., 2013. Sediment concentration prediction and statistical evaluation for annual load estimation. Journal of Hydrology (482) 69 – 78.

Webb, B., Phillips, J., Walling, D., Littlewood, I., Watts, C., Leeks, G., 1997. Load estimation methodologies for British rivers and their relevance to the lois racs (r) programme. Science of the Total Environment 194, 379–389.

Wood, S., 2006. Generalized additive models: an introduction with R. Vol. 66. Chapman & Hall/CRC.

Chapter 3

A simulation based approach to quantify the difference between event-based and routine water quality monitoring schemes

3.1 Summary

Rainfall events are often associated with large exports of nutrients and sediments. Many water quality monitoring schemes use a form of event-based sampling to quantify these exports. Water quality studies which have evaluated different sampling schemes often rely on continuously monitored water quality data. However, many catchment authorities only have access to limited historical data which consists of event-based and monthly routine samples. Therefore there is a requirement for a method to assess the importance of sampling events using information from limited historical data. This work presents an approach using unconditional simulation based on historical stream discharge. Such an approach offers site-specific information on optimal sampling schemes. A linear mixed model is used to model the relationship between total phosphorus and stream discharge and the auto-correlation of total phos-

phorus. The inclusion of event-based sampling improved annual load estimates for all sites with a maximum RMSE difference of 16.11 tonnes between event-based and routine sampling. Based on the accuracy of annual loads, event-based sampling was found to be more important in catchments with a large relief and high annual rainfall. Using this approach different sampling schemes can be compared based on limited historical data.

3.2 Introduction

The accuracy of water quality load estimates is directly related to the monitoring design used to collect the water samples. Ideally, the load of a water quality variable would be derived using continuously sampled data. However, it is not financially viable for most water quality monitoring programmes to collect continuous data (Bartley et al., 2012; Burt et al., 2011; Drewry et al., 2009). Therefore water quality sampling schemes must be designed to provide accurate load estimates with limited samples. Monthly sampling is commonly used throughout Australia (Bartley et al., 2012). However, monthly sampling often misses key rainfall events (Drewry et al., 2009), therefore many sampling schemes include a form of event sampling. Most studies examining the effect of different sampling schemes have been limited to catchments with access to continuous sampled data. Therefore there is a need for methods which can use site-specific historical water quality data to assess the effect of different sampling schemes on load estimation (e.g. event-based sampling).

Based on studies with access to sufficient data (Drewry et al., 2009; Johnes, 2007; Hopmans and Bren, 2007), it is generally accepted that a form of sampling is required during rainfall events to capture periods of high nutrient and sediment exports. This is especially important in catchments with long periods of base-flow as large exports of nutrients and sediments occur during short rainfall events (Jones et al., 2011; Drewry et al., 2009; Gao, 2008; Hopmans and Bren, 2007). A study by Hopmans and Bren (2007) observed large sediment exports during events, with 70% of 6 years of sediment load being exported during a single event in south east Australia. One particular water quality property of interest is total phosphorus (TP), as in large concentrations TP can cause algal blooms (Davis and Koop, 2006; Kristiana et al., 2011). A study focusing on 17 streams in the UK, found 20% of annual TP was exported during a single day and the largest 5 events within a year contributed to 42% of the annual TP load (Johnes, 2007).

To improve monitoring schemes it is important to understand the relationship between catchment characteristics and water quality variable exports. This information is required to help the design of monitoring schemes in unmonitored catchments. The relationship between catchment characteristics (catchment size, slope, rainfall, stream discharge, land use and land cover) has been investigated by several studies (Ahearn et al., 2005; Banner et al., 2009; Sliva and Williams, 2001). Relationships between phosphorus and agriculture land uses (Ahearn et al., 2005) and catchment size (Ahearn et al., 2005; Johnes, 2007) have been found. Focussing on phosphorus, Banner et al. (2009) found a relationship between phosphorus, catchment topography and land use for 2 sites in Kansas, USA. In addition, urbanisation (Sliva and Williams, 2001) and population density (Johnes, 2007) have been found to have an effect on water quality. Johnes (2007) also found a relationship between the base-flow index and the uncertainty of TP load estimates, with catchments with a low base-flow index having larger uncertainty than streams with a high base-flow index.

Sampling schemes can be divided into two main categories; probability and non-probability based (de Gruijter et al., 2006). Probability based methods rely on known inclusion probabilities to provide unbiased estimates of the mean and its uncertainty. Non-probability sampling should use a model-based approach as the inclusion probabilities are unknown (de Gruijter et al., 2006; Lark and Cullis, 2004). Several probability based sampling schemes have been shown to provide accurate estimates of suspended sediments (Lewis, 1996; Thomas, 1985, 1988; Thomas and Lewis, 1993, 1995). However, non-probabilistic sampling schemes are more commonly used. One of the most commonly used schemes is to sample at equal intervals in time (for examples see Salles et al. (2008); Birkel et al. (2011)). This is one of the most common event-based sampling schemes due to ease of implementation with available automatic sampling equipment.

Load estimation methods offer the ability to provide estimates over different time intervals (e.g. event-based or annually). Many estimation methods exist with Marsh and Waters (2009) comparing 34 different load estimation methods. Several studies have compared the various load estimation methods in relation to annual load estimates (Kronvang and Bruhn, 1996; Cooper and Watts, 2002; Cassidy and Jordan, 2011; Johnes, 2007; Marsh and Waters, 2009) with the majority of these methods falling into three main categories; average, ratio and regression methods. However the majority of these cannot be used with commonly applied sampling schemes (e.g. monthly or a combination of monthly and event-based sampling) as the sampling schemes are non-probabilistic, i.e. it is not possible to determine the probability of taking a sample at a specific time. Therefore average, ratio and regression based load estimation methods should not be used as they require a form of probabilistic sampling, a requirement that has been noted by several water quality based studies (Cohn et al., 1992; Cohn, 2005; Cooper and Watts, 2002; Crawford, 1991; Thomas, 1985, 1988).

Linear mixed models (LMM) provide unbiased water quality load estimates without the assumption of probabilistic sampling schemes. Differing from simple linear models, LMM account for auto-correlation between samples within the error term of the model. Linear mixed models are regularly applied in soil science to account for the spatial auto-correlation between samples (Lark and Cullis, 2004). Temporal water quality data is similar to spatial soil data as water quality data has been shown to be auto-correlated through time (Kuhnert et al., 2012; Wang et al., 2011). In addition to removing the need to meet the assumption of probabilistic sampling, LMMs offer the ability to incorporate additional covariates (e.g. stream discharge and turbidity) to improve predictions.

A lack of data is a major limitation of water quality studies. Several studies have found strong linear trends between water quality variables and low cost continuously measured surrogates (e.g. stream discharge and turbidity) and have used these relationships to provide continuous predictions of water quality variables (Webb et al., 2000; Kim and Furumai, 2012; Lewis, 1996; Wang et al., 2011; Kuhnert et al., 2012). Webb et al. (2000) simulated numerous water quality variables using a linear relationship with continuously monitored stream discharge data. These simulations provided a method to explore the accuracy of various load estimation methods based on limited historical water quality data (Webb et al., 2000), However, the simulation method of Webb et al. (2000) did not examine the potential for temporal auto-correlation between the observed water quality samples and used a single realisation of the relationship with stream discharge. An alternative approach is to use unconditional Gaussian simulation to simulate data based on a model fitted using a LMM. The advantage of this approach is that the model used to simulated the TP data is based on a valid statistical model describing the (co-)variation between water quality and discharge. The simulated data is based on the linear relationship with stream discharge while respecting the temporal auto-correlation of the observed water quality variables (Gebbers and Bruin, 2010).

In this work we simulate continuous TP using its relationship with stream discharge based on the LMM. Using these simulations we investigate the effect of using event-based sampling (in addition to routine sampling) in terms of the accuracy of annual load estimates. In addition, the effect of including event-based sampling is related to catchment characteristics as the information from this may help improve the design of monitoring schemes in unmonitored catchments. Therefore the aims of this work are to:

1. Illustrate a general approach to examine the effect of different sampling schemes on load estimates based on limited historical data.

2. Examine the effect of event-based sampling on estimates of the annual load of TP.

3. Investigate the relationship between catchment characteristics and improvements on load estimates using event-based sampling.

3.3 Materials and methods

3.3.1 Catchment description

The 9 monitoring sites considered in this work are located upstream of several storage dams which contribute the majority of Sydney's drinking water. The study sites are located west of Sydney with their location provided in figure 3.1. A summary of the topographical and hydrological features of each catchment is provided in table 3.1. The number of each monitoring site in table 3.1 is indicated in figure 3.1. The catchment size of the monitoring sites ranges from 20.2 to 4825.1 km^2. The annual rainfall of each catchment was estimated using thessian polygons based on data from nearby Bureau of Meteorology stations. Rainfall is highest in the east near the coast and to the north of the region with the lowest rainfall in the south west of the region. Forest is the main land-cover across the region, however grassland is the dominant land-cover in the Berrima Weir and the Jooriland catchments. Berrima Weir and Kedumba Crossing catchments both have more than 10% urban cover, while all other catchments have less than 3% urban cover. In addition to percentage of urban cover, population density was also estimated for each catchment based on data obtained from the Australian Bureau of Statistics (2011). The Australian Bureau of Statistics (2011) publishes population density data, based on defined statistical local areas which are sub-divisions of local government areas. Estimates of population density were made based on the coverage of these areas within each catchment.

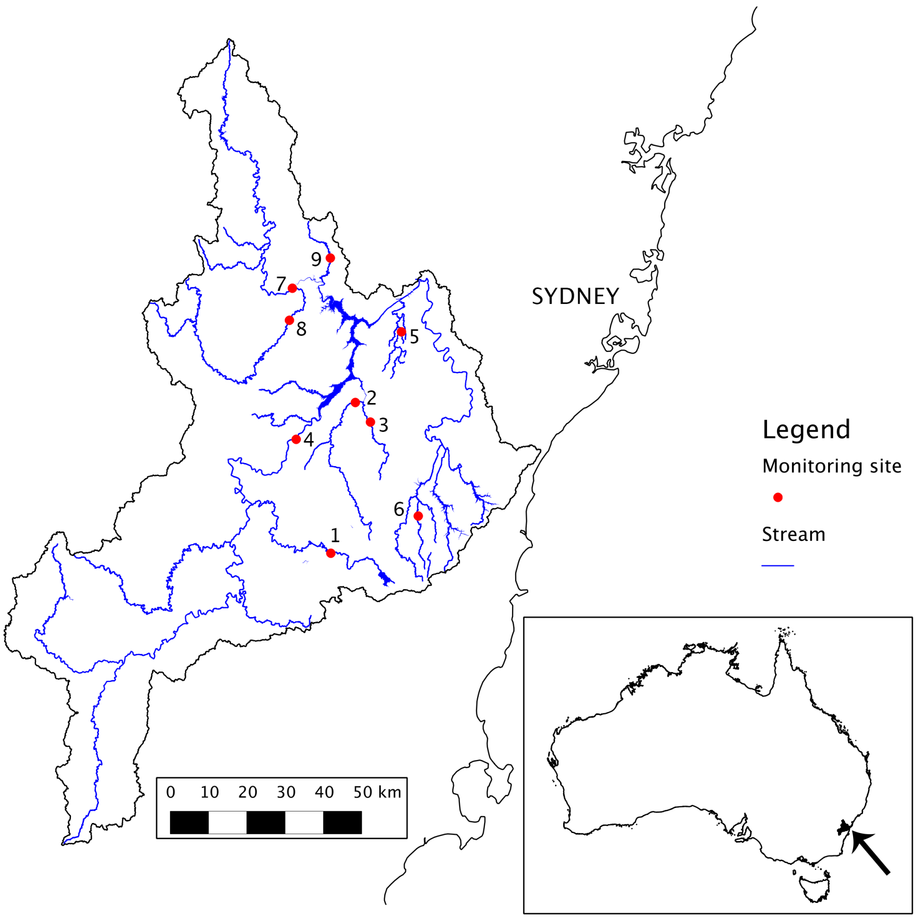

Figure 3.1 – Map of greater catchment area and location of monitoring sites and the location of the catchment within Australia.

Table 3.1 – Key features of each sub-catchment.

site	site number	area (km^2)	minimum elevation (m)	maximum elevation (m)	annual rainfall (mm)	annual discharge (ML)	dominant land cover	urban cover (%)	population density (km^2)
Berrima Weir	1	201.4	620	868	959.3	54790.1	*grassland*	10	80
Smallwoods Crossing	2	436.9	105	867	857.2	24419.21	*forest*	2	17.6
Little River	3	104	175	632	824.7	8801.2	*forest*	0	8.9
Jooriland	4	4825.1	108	1179	708.2	203643	*grassland*	1	11.1
Werombi	5	56.4	203	519	695	5923.2	*forest*	1	25
Burke River	6	88.3	328	777	1182.5	17230.4	*forest*	1	3.7
Kelpie Point	7	1447	116	1342	886.7	113734.3	*forest*	0	15.2
Cedar Ford	8	719.1	169	1382	932.4	75405.5	*forest*	0	0.8
Kedumba Crossing	9	72.5	152	1061	1263.6	14588.7	*forest*	13	52.9

3.3.2 Data description

Table 3.2 summarises the total phosphorus data for each site. The data was obtained from the Sydney Catchment Authority. Routine monthly sampling was used at all sites and this was combined with event-based sampling at most sites after 2001. Stream discharge was recorded every 15 minutes at each location. Total phosphorus was analysed using acid digestion methods at contracted laboratories meeting the standards outlined by Standard Methods (1995). The minimum total phosphorus included in this analysis for each site was restarted to 0.005 mg L^{-1}. Summary statistics are provided for each site with the mean total phosphorus ranging from 0.02 to 0.08 mg L^{-1}. Sample sizes of the sites ranges from 153 to 553 over the monitoring period of each site.

Table 3.2 – Water quality data summary.

site	start date	end date	mean	min	max	sd	skewness	n
Berrima Weir	Dec 1993	Dec 2010	0.08	0.006	0.87	0.09	4.75	392
Smallwoods Crossing	Jan 1991	Mar 2010	0.06	0.006	0.70	0.09	4.21	323
Little River	Aug 1991	Apr 2007	0.02	0.006	0.46	0.05	6.35	153
Jooriland	Feb 1994	Dec 2010	0.06	0.006	0.45	0.07	2.38	527
Werombi	Jan 1991	Dec 2010	0.04	0.006	0.73	0.05	6.99	535
Burke River	Oct 1991	Dec 2010	0.02	0.006	0.17	0.01	4.94	454
Kelpie Point	Jan 1991	Dec 2010	0.03	0.006	0.92	0.07	7.50	392
Cedar Ford	Jan 1991	Dec 2010	0.04	0.006	0.87	0.10	5.33	553
Kedumba Crossing	Jan 2002	Dec 2010	0.04	0.006	0.67	0.06	5.52	241

3.3.3 Statistical analysis

The simulation procedure used to compare the two sampling schemes is provided as a flowchart in figure 3.2. The procedure comprises 5 steps which are described below.

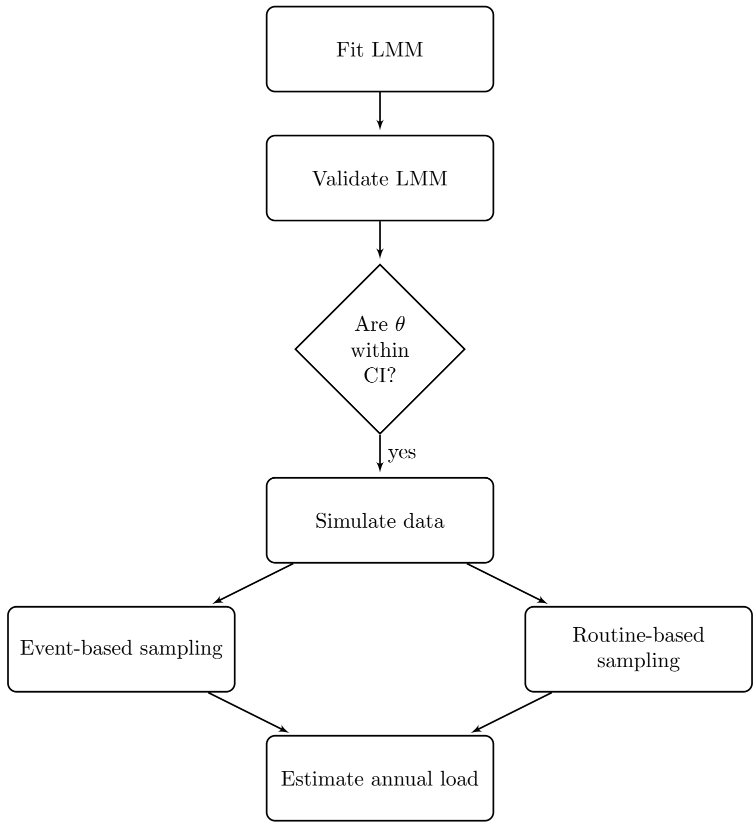

Figure 3.2 – Flow chart of the simulation process for each site.

3.3.3.1 Fitting a linear mixed model

In this work a LMM is used to model the relationship between TP and stream discharge. The LMM structure allows for the inclusion of quadratic terms to allow for non linear trends between TP and stream discharge, however this was not investigated in this study. In addition, the LMM accounts for the temporal auto-correlation between samples. The LMM used here has the form:

$$z(\mathbf{t}) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}, \tag{3.1}$$

where the concentration of TP ($z(\mathbf{t})$) is treated as a random process through time t, $\mathbf{X}$ is a n x p matrix of the explanatory variables and $\boldsymbol{\beta}$ is a n x 1 vector of model coefficients (Lark and Cullis, 2004). The LMM does not require independently and identically distributed (*iid*) samples because the $\boldsymbol{\eta}$ has a correlation structure $\boldsymbol{\eta} \sim \mathcal{N}(0, \mathbf{V})$, where $\mathbf{V}$ is a positive definite matrix of the variance (σ^2) and the covariance. However the LMM does require $z(\mathbf{t})$ to be Gaussian, therefore a Box-Cox power transformation (Box and Cox, 1964) is used conditional on the value of lambda (λ),

$$z(\mathbf{t})^* = \begin{cases} \log z(\mathbf{t}) & \text{if } \lambda = 0 \\ \frac{z(\mathbf{t})^\lambda - 1}{\lambda} & \text{otherwise.} \end{cases} \tag{3.2}$$

The LMM fitted in this work uses residual error maximum likelihood (REML) to fit the model using the geoR package (Ribeiro Jr and Diggle, 2001) in the R environment R Core Team (2012). By using REML the model coefficients ($\boldsymbol{\beta}$) are fitted conditional on the parameters of the variance-covariance matrix ($\mathbf{V}$) and the lambda (λ) value of the Box-Cox transformation (for a detailed description see Lark and Cullis (2004)). Several different covariance structures are available to model the auto-correlation (Lark and Cullis, 2004). In this study an exponential function is used to account for the auto-correlation between samples where the correlation matrix is defined as;

$$\begin{aligned} \mathbf{V}_{i,j} = \quad & \sigma^2 s \exp\left(-\frac{|\mathbf{x_i} - \mathbf{x_j}|}{a}\right), \quad & i \neq j \\ & \sigma^2, & i = j, \end{aligned} \tag{3.3}$$

where the variance σ^2 is the diagonal and $|\mathbf{x_i} - \mathbf{x_j}|$ is the temporal distance between two samples and a is the distance parameter of the exponential function. The temporal auto-correlation is described by s which is defined as;

$$s = \frac{c}{c_0 + c}, \tag{3.4}$$

where c_0 is the nugget variance or unexplained variance often referred to as the sampling error and $c_0 + c$ describes the maximum variance between two variables. The temporal structure of the fitted model is dependent on the observed samples, therefore it is important that samples are collected nearby in time to enable accurate estimation of the distance parameter and nugget semivariance.

3.3.3.2 Evaluation of the LMM

Each LMM is assessed for significance of the linear relationship with discharge and the temporal auto-correlation model. In addition to this, model validation is used to ensure the model (both the predictions, prediction variance) is a *true* representation of the observed total phosphorus. In this work, three steps are used to evaluate each model:

1. test the significance of the linear relationship with stream discharge,
2. test the significance of the temporal auto-correlation, and
3. validation of the prediction variance using leave-one-out-cross-validation (LOOCV).

Wald tests are used to test the significance of the relationship with stream discharge, see Lark and Cullis (2004) for a detailed description. The Akaike information criteria (AIC) (Akaike, 1974) is used to test the significance of the temporal auto-correlation of each model, by comparing the LMM with and without the temporal auto-correlation model. Cross validation methods, such as LOOCV are required to validate models in studies with limited data. LOOCV uses the LMM to predict each observation using all other observations for each observation used in fitting the LMM. Using the predictions the following is derived;

$$\theta_i = \frac{\{c(\mathbf{t}_i)^* - \tilde{C}^*_{(-i)}\}^2}{\sigma^2_{(-i)}}, \tag{3.5}$$

where $c(\mathbf{t}_i)^*$ is the i^{th} transformed concentration observation, $\tilde{C}^*_{(-i)}$ is the corresponding estimated concentration based on all other samples and $\sigma^2_{(-i)}$ is the estimated kriging variance. If the models a *true* representation of the process, θ has a mean $\bar{\theta} = 1$ and median $\breve{\theta} = 0.455$ (Marchant et al., 2010). Based on the methods of Marchant et al. (2010) it is possible to estimate 95% confidence intervals (CI) for the median and mean θ. If the θ statistics are within the CI, it is assumed the models are valid.

3.3.3.3 Simulation of total phosphorus.

The purpose of the simulation approach here is to generate comparable realisations of continuous water quality based on the fitted LMM conditional on the observed discharge data. From this the continuous water quality data can be sampled, and load estimation predictions can be based on these samples and compared with the continuous time series. Geostatistical based simulations provide the ability to simulate a random process while preserving the auto-correlation of the random field (Gebbers and Bruin, 2010). Simulation is a commonly applied tool in geostatistics to generate spatially correlated fields, however as stated by Gebbers and Bruin (2010) there is no reason why it cannot be used to simulate temporally correlated fields. There are two forms of geostatistical simulation; conditional and unconditional (Gebbers and Bruin, 2010). Conditional simulation preserves the observed values while unconditional simulation uses a random starting point to simulate a random field with the same temporal auto-correlation as that of the observed correlation (Gebbers and Bruin, 2010). In this work unconditional Gaussian simulation is performed to generate 2000 realisations of TP at each site based on the LMM based on the correlation with observed discharge. A total of 2000 realisations were used as it was believed that this would be a sufficient amount for comparisons and would not take too many computation hours to complete. The simulation is based on the code of the gstat package (Pebesma, 2004) in the R environment. The simulation is performed using the coefficients and auto-correlation structure of the LMM (eq. 3.1). The simulations for each site are determined using;

$$r(\mathbf{t})_j = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{s}_j, \tag{3.6}$$

where $r(\mathbf{t})_j$ represents a realisation, $\boldsymbol{s}_j$ is a temporally correlated random field based on the temporal auto-correlation of the fitted LMM with mean 0 and j is the realisation and the trend described by ($\boldsymbol{\beta}$) with stream discharge in the design matrix ($\mathbf{X}$). The simulations are based on the box-cox transformed TP models and therefore the simulations are back-transformed using the inverse of the Box-Cox power transformation.

3.3.3.4 Sampling schemes.

For each realisation event-based and routine sampling are performed. Routine sampling is defined as monthly sampling with the sample being taken on the first Wednesday of each month. The event-based sampling scheme is a combination of the routine samples and event-based samples. The event-based samples are defined in a way to reflect commonly applied sampling schemes under the restrictions of commonly used auto-samplers. A common approach used for event-based sampling schemes is to use a trigger level to start event sampling, and continue sampling until all 24 samples are collected (the capacity of most auto-samplers). Therefore event-based sampling schemes were determined using the following steps;

1. Define the trigger level as the upper 5th percentile of stream discharge (Q_{95}).

2. Find all events using this trigger level.

3. Find the mean duration of all events which are longer than 6 hours.

4. Equally space 24 samples for each event over a period of time equal to the mean event length.

3.3.3.5 Load estimation.

As the two sampling schemes are non-probabilistic, LMM are used to estimate TP using a log linear relationship with stream discharge. For each realisation at each site a LMM is fitted for both routine and a combination of routine and event-based sampling. A major limitation of fitting LMM using REML is the computational power and time required to optimise the likelihood function of the model (Minasny and McBratney, 2007). To overcome this limitation and to maximise the efficiency of the model fitting the following procedure was used;

1. Fit a simple linear model to the sampled data.
2. Fit a variogram to the residuals of the linear model.
3. Use the values of this variogram as initial values to fit a LMM to a subsample of the observations.
4. Fit a LMM to all observations using the correlation structure of the previous LMM as the initial values.

Each LMM is fitted using a optimised implementation of the LMM likelihood function based on the geoR package using the RcppArmadillo package (Francois et al., 2011; Sanderson, 2010) in the R environment.

The AIC was used to determine if the temporal auto-correlation was significant, based on this, either universal kriging or simple linear regression is used to predict the TP (see Bivand et al. (2008, p 209) for a detailed description of universal kriging). Universal kriging is used where the temporal auto-correlation of the LMM is found to be significant and simple linear regression is used where there is no temporal auto-correlation found in the sampled data. Using the predicted TP the annual load of each realisation is estimated for years with $< 10\%$ stream discharge missing. Comparisons of the load estimates are made based on the root-mean-squre error (RMSE) between the observed and predicted annual load of each sampling method.

3.3.4 Relating annual load uncertainties to catchment characteristics

The effect of catchment characteristics on the accuracy of the load estimations is examined using the mean difference of the RMSE between the two sampling schemes, over all realisations. The catchment characteristics used are defined in table 3.1. The correlation between the mean difference of the RMSE and the catchment characteristics is examined. In addition, backwards elimination of a linear model is used to the find the most parsimonious model to explain the variation in the difference in RMSE, between catchments.

3.4 Results

General results are presented for all catchments but more detail is presented for Kelpie Point catchment to clearly illustrate all the steps presented in the flowchart (Figure 3.2).

3.4.1 LMM used for simulation

Table 3.3 summarises the model coefficients for each site. The temporal auto-correlation of all models was found to be significant. The range of the temporal auto-correlation was less than 7 days for 7 of the 9 catchments. This finding highlights the importance of accounting for temporal auto-correlation when modelling water quality. Stream discharge was found to be a significant predictor for all catchments.

Figure 3.3 shows the fitted linear trend of the LMM for the Keplie Point catchment. It may appear that the trend of the LMM is not the best fit of the data, however it is

Table 3.3 – Model coefficients of each site (*indicates significance).

site	β_0 (model intercept)	β_1 (stream discharge)	range (days) (a)	sampling error (c_0)	variance ($c_0 + c$)	lambda (λ)
Berrima Weir	-6.45	0.26*	30.00	0.92	2.45	-0.30
Smallwoods Crossing	-14.59	1.07*	5.56	5.33	14.47	-0.50
Little River	-71.82	5.65*	19.44	391.54	629.82	-0.86
Jooriland	-7.44	0.34*	2.96	0.60	2.16	-0.17
Werombi	-12.03	0.89*	2.33	1.49	6.27	-0.35
Burke River	-44.76	1.69*	2.25	31.71	166.71	-0.73
Kelpie Point	-28.6	1.96*	2.25	4.92	31.95	-0.54
Cedar Ford	-25.06	1.3*	4.20	3.52	45.54	-0.54
Kedumba Crossing	-15.08	1.01*	2.33	3.62	19.14	-0.44

important to note that the fitted linear trend is conditional on the Box-Cox transform and the temporal auto-correlation model. Figure 3.4 shows the fitted temporal auto-correlation with the experimental variogram for the Kelpie Point catchment. As the exponential structure of the temporal auto-correlation describes the variance between samples of different temporal distances, samples greater than 10 days apart are not correlated.

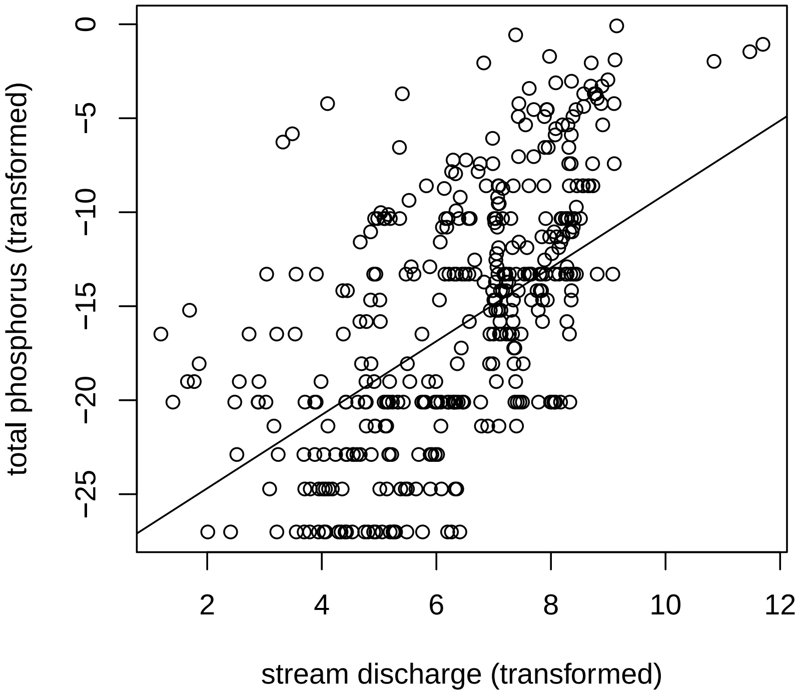

Figure 3.3 – Relationship between TP and stream discharge for the Kelpie Point catchment.

The LOOCV procedure predicts each observation using all surrounding observations where information from observations within the practical range ($3a$) of the temporal auto-correlation is included in the predictions. Figure 3.5 presents the transformed observed TP against the LOOCV predictions for Kelpie Point catchment. It appears that smaller TP observations are overestimated by the LMM. Using the estimated LMMs of each site, LOOCV was performed and is summarised in table 3.4. The

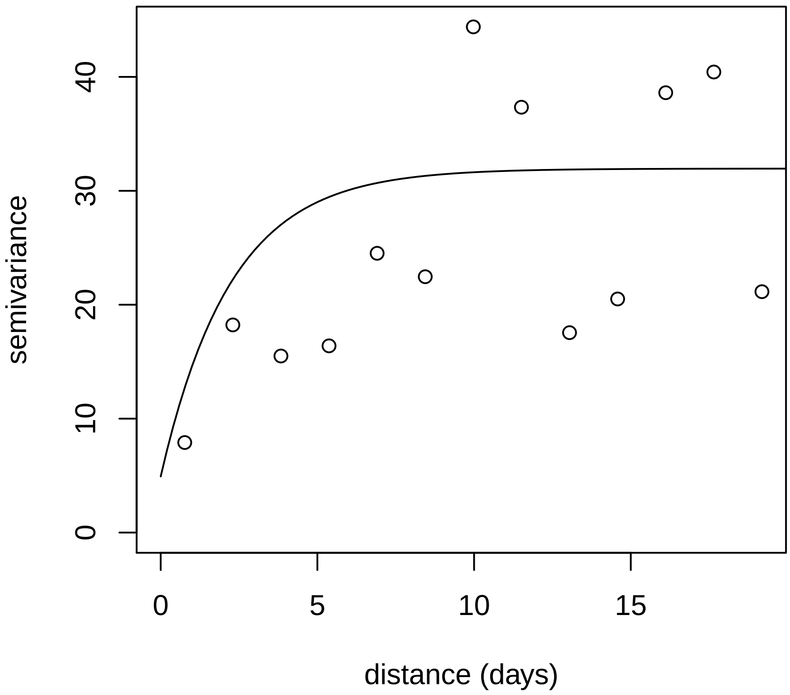

Figure 3.4 – Temporal structure with experimental variogram for the Kelpie Point catchment.

table reports the estimated mean ($\bar{\theta}$)and median ($\hat{\theta}$) (eqn. 4.8), and the estimated 95% CI for both $\bar{\theta}$ and $\hat{\theta}$. In addition, Lin's correlation coefficient (Lawrence and Lin, 1989) between the observed and LOOCV predicted values is provided. Lin's correlation coefficient is a measure of how closely the points fit to a 45 degree line. All sites except the Little River catchment had a correlation value greater than 0.6. The estimated $\bar{\theta}$ of each site was close to 1 and within the CI. Based on the $\hat{\theta}$ of the LOOCV only 5 sites fell within the estimated CI. This is important as Marchant et al. (2010) showed the $\hat{\theta}$ is a better measure of how close the estimated prediction variance represents the observed error. Therefore, simulations of TP were only conducted at the following 5 sites:

a. Smallwoods Crossing,

b. Little River,

c. Burke River,

d. Kelpie Point,

e. Kedumba Crossing.

Table 3.4 – Cross validation statistics of LMMs.

site	$\bar{\theta}$	$\hat{\theta}$	$\bar{\theta}\ CI_{95\%}$	$\hat{\theta}\ CI_{95\%}$	correlation (Lin's)
Berrima Weir	1.01	0.25	(0.84 - 1.15)	(0.36 - 0.57)	0.6
Smallwoods Crossing	0.91	0.36	(0.86 - 1.17)	(0.35 - 0.58)	0.7
Little River	0.97	0.55	(0.78 - 1.23)	(0.3 - 0.64)	0.49
Jooriland	0.95	0.25	(0.8 - 1.21)	(0.31 - 0.63)	0.76
Werombi	0.99	0.30	(0.88 - 1.14)	(0.37 - 0.56)	0.77
Burke River	1.03	0.40	(0.87 - 1.14)	(0.36 - 0.56)	0.6
Kelpie Point	1.02	0.44	(0.86 - 1.16)	(0.36 - 0.58)	0.68
Cedar Ford	0.99	0.33	(0.88 - 1.14)	(0.37 - 0.56)	0.76
Kedumba Crossing	1.02	0.46	(0.86 - 1.14)	(0.36 - 0.56)	0.63

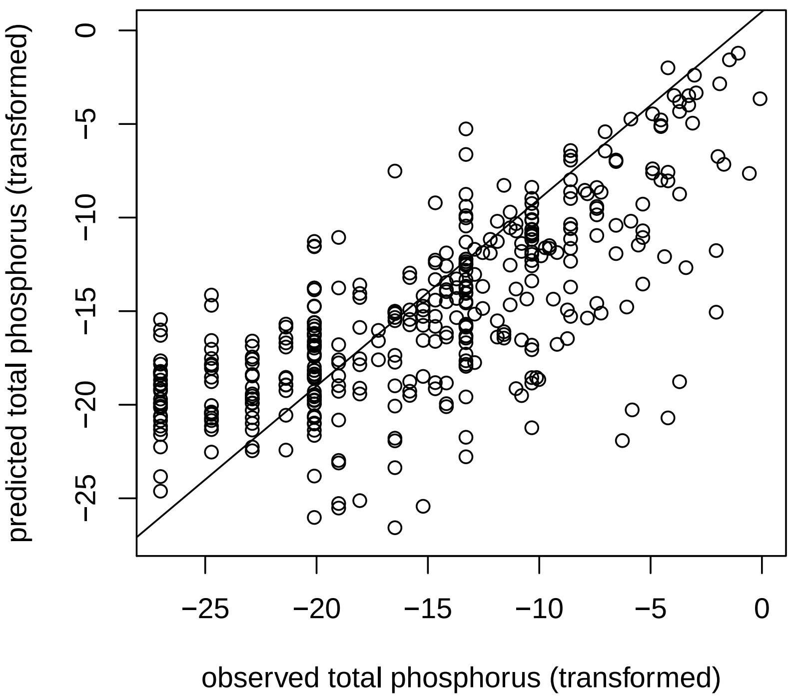

Figure 3.5 – TP model LOOCV predictions for the Kelpie Point catchment.

3.4.2 Simulation of total phosphorus

By using a LMM for each site and continuously monitored stream discharge it is possible to create auto-correlated simulations of total phosphorus. Figure 3.6 provides a box-plot of the simulated load of all realisations separated by each year for the Kelpie Point catchment. The figure highlights the differences in annual load between each year. Years 1992 and 1998 have the largest annual TP and within these years the annual load is highly varied. There is little variation in annual load in years with low annual loads (e.g. 2001 and 2002). These simulations provide a method to use historical data to compare the effect of different sampling schemes.

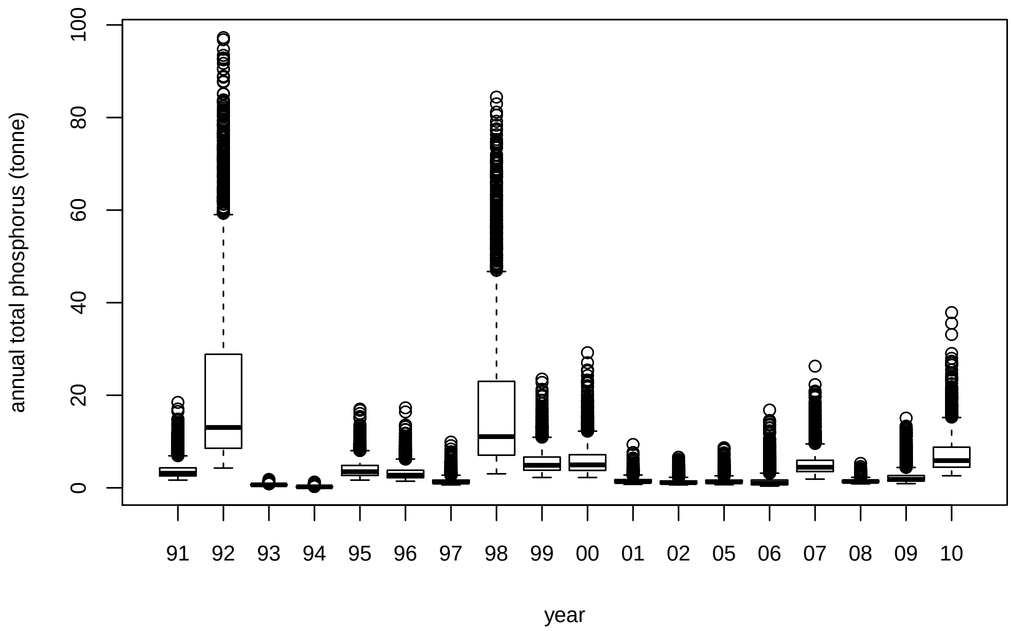

Figure 3.6 – Simulated annual load for the Kelpie Point catchment.

3.4.3 Estimating annual load

Table 3.5 summarises the key features of the routine and event-based sampling schemes of the simulated data. The trigger level is different for each catchment, however the event length of the catchments is similar, with the event length of the catchments being between 3 and 4 days for 4 of the 5 sites and less than 4 days for all sites. The sample size of the event-based sampling scheme combines the event samples with the routine monthly samples. The event-based sample size was at least twice as large as the routine sampling scheme, and more than 10 times as large for the Kedumba Crossing catchment.

Table 3.5 – Sample scheme characteristics of each site.

site	trigger level (ML day^{-1})	event length (days)	event samples	routine samples
a. Smallwoods Crossing	114.10	3.39	521	213
b. Little River	24.14	3.31	736	181
c. Burke River	80.14	2.64	994	233
d. Kelpie Point	787.07	3.89	784	237
e. Kedumba Crossing	76.64	1.48	1050	102

An important aspect of the LMMs is the modelling of the temporal auto-correlation. Table 3.6 summarises the mean temporal range of the models fitted to the two sampling schemes. The table also indicates the percentage of realisations where the temporal auto-correlation is significant. For all catchments except the Little River catchment, most realisations did not find significant temporal auto-correlation for routine sampling, with less than 6% of realisations finding temporal auto-correlation with a mean of over 100 days. For the routine sampling based models of the Little River catchment significant temporal auto-correlation was found for 18.75 % of realisations with a mean of 72.67 days. In contrast to routine sampling based models, event-based sampling models found significant temporal auto-correlation for over 90% of realisations with a mean correlation of less than 3 days for all catchments except the Little River catchment. The event-based sampling models of the Little River catchment found significant temporal auto-correlation in 52.05% of the realisations with a mean of 2.08 days. A potential reason for the differences in the Little River catchment is due to the temporal range of 19.44 days, where all other catchments used in the simulation procedure had a temporal range of less than 6 days. The difference in temporal significance and temporal range between the routine and event-based sampling models is due to the distances between the samples of the respective sampling schemes. The inclusion of the temporal structure is beneficial as the prediction procedure uses WQ information from nearby observations to improve predictions.

Table 3.6 – Temporal auto-correlation statistics of LMMs of simulated data.

site	event-based		routine-based	
	mean range (days)	temporally significant (%)	mean range (days)	temporally significant (%)
a. Smallwoods Crossing	2.50	90.30	112.62	5.35
b. Little River	2.08	52.05	72.67	18.75
c. Burke River	1.43	99.65	126.93	5.75
d. Kelpie Point	1.52	98.75	116.36	5.00
e. Kedumba Crossing	1.76	100.00	116.68	4.50

Figure 3.7 presents the estimated annual load against simulated annual load for both the event-based and routine sampling schemes for all realisations for the Kelpie Point catchment. Annual load estimates using the event-based models (Fig. 3.7a.) shows how the variance of the predicted annual load increases as the simulated annual load increases. Figure 3.7b. shows the routine based annual load estimates. The routine based estimates underestimate annual loads over 20 tonnes. The under estimation of the routine based sampling indicates that event-based sampling will improve estimates of years with large annual loads.

The RMSE is used to evaluate the accuracy of each realisation. As an aim of the work is to determine the benefit of including event-based sampling on load estimation, the difference between the RMSE of event-based and routine sampling is used. Figure 3.8 is a box-plot of the difference between the routine RMSE and the event-based RMSE for the Kelpie Point catchment. For each realisation if the difference between the RMSE is positive, the event-based estimate is more accurate than the routine estimate. Table 3.7 provides a summary of the differences between the RMSE of the catchments. The table also includes the improvement of event-based sampling (IES) which is the percentage of realisations where the inclusion of event-based sampling did not improve the accuracy of load estimates. The IES of all sites, except the Little River catchment is less than 10% and 20.9% for the Little River catchment. This indicates that the inclusion of event-based sampling increases the accuracy of the

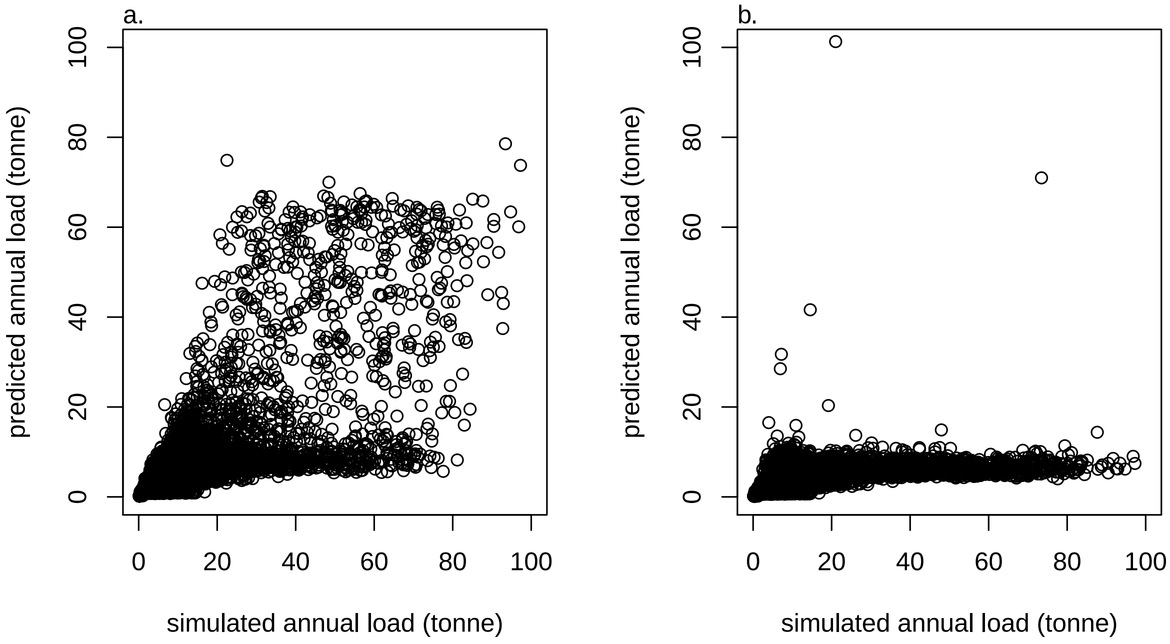

Figure 3.7 – Event-based (a) and routine (b) annual load estimates for the Kelpie Point catchment.

load estimations for the majority of the realisations. Within the realisations where the inclusion of event-based sampling did not improve the estimates, the correlation between TP and stream discharge was stronger for the routine models. In addition, there was a higher percentage of event-based models without significant temporal auto-correlation for these realisations therefore there was a less of a benefit to be gained from event-based samples.

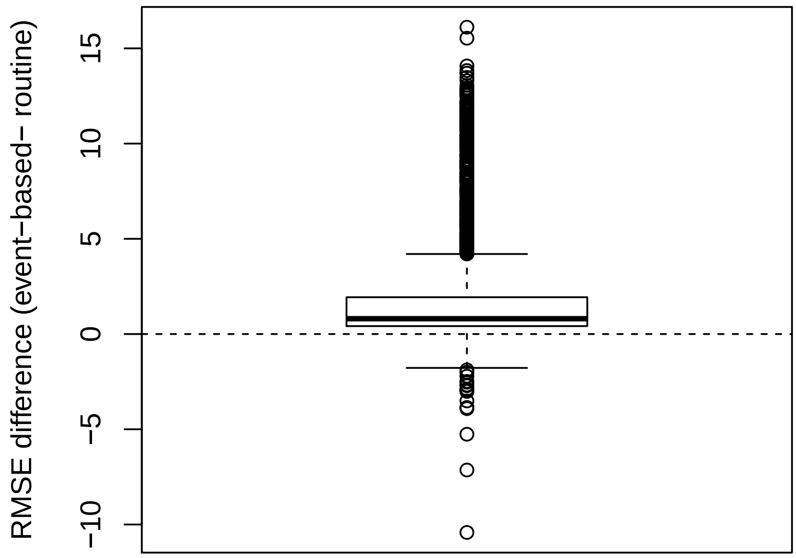

Figure 3.8 – RMSE difference of all realisations for the Kelpie Point catchment.

Table 3.7 – Summary of differences between routine and event

site	min RMSE difference (tonne)	max RMSE difference (tonne)	mean RMSE difference (tonne)	IES (%)
a. Smallwoods Crossing	-1.77	5.94	0.55	8.40
b. Little River	-0.74	1.54	0.11	20.90
c. Burke River	-0.12	0.39	0.05	6.90
d. Kelpie Point	-10.42	16.11	1.93	5.60
e. Kedumba Crossing	-0.14	0.74	0.17	3.35

3.4.4 Relating event-based sampling improvement with catchment characteristics

Finding relationships between the IES and catchment characteristics is important for the improvement of sampling schemes at sites without historical data. The correlation between catchment characteristics and the IES at each site is provided in table 3.8. Annual rainfall, elevation range, urban cover and population density are the most correlated with the IES. All catchment characteristics except elevation range are negatively correlated with the IES. In addition the percentage of urban land cover and population density had similar correlation values of -0.51 and -0.5 respectively. Backwards elimination resulted in a final model including annual rainfall and elevation range. Figure 3.9 shows the predicted values against the IES. It is important to note that this model is not intended to be used for prediction for other sites, but to provide a guide to the relationship between the IES and catchment characteristics. The results of the backwards elimination indicate that annual rainfall and elevation range decreases the improvement of accuracy from the inclusion of event-based samples decreases. This indicates that event-based sampling is more important in small flashy upland catchments as opposed to large lowland catchments. This response may be due to the faster hydrological responses in upland catchments.

Table 3.8 – Correlation between catchment characteristics and mean RMSE difference.

catchment characteristic	correlation
Catchment area	-0.28
Annual rainfall	-0.62
Annual stream discharge	-0.35
Urban cover	-0.51
Population density	-0.5
Elevation range	0.6

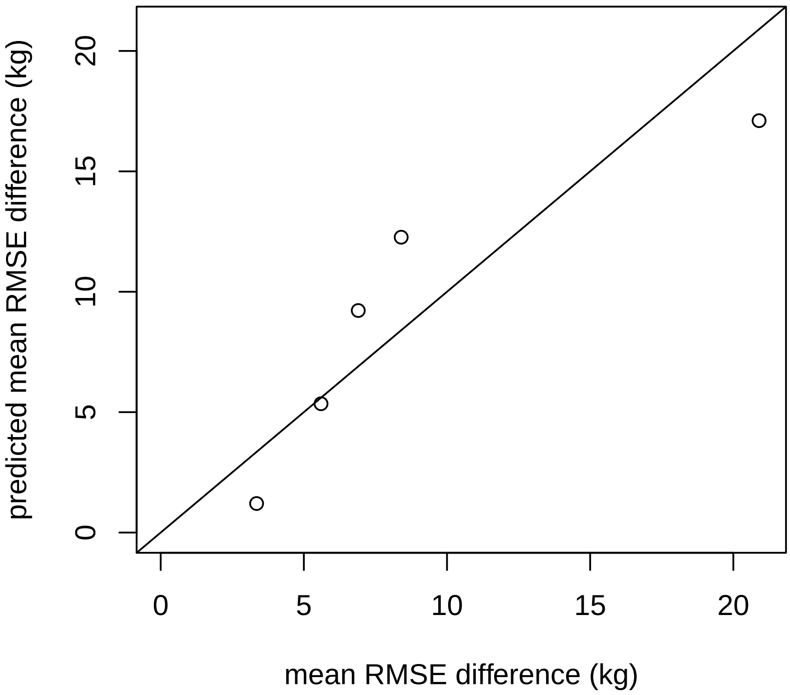

Figure 3.9 – Relationship between mean RMSE difference and predicted mean RMSE difference.

3.5 Discussion

It is important to consider the temporal auto-correlation of water quality observations (Wang et al., 2011). The LMMs used in this work combine fixed effects (e.g. stream discharge) with the temporal auto-correlation (random effects) to provide unbiased TP parameters. By accounting for the auto-correlation between the samples, the LMM does not require the use of probabilistic sampling. Stream discharge was found to be a significant predictor of total phosphorus for all catchments. In addition, temporal auto-correlation was found for all catchments. Of the 9 catchments it was only possible to fit a valid statistical model for 5 catchments in terms of how well the prediction variance matches the errors. One reason for this may be that some of the catchments had unusual observations which may or not have been erroneous. These were difficult to model. The use of additional covariates which characterise hydrological processes (e.g. hysteresis) may improve how the prediction variance matches the errors. Several studies have investigated robust methods to increase the flexibility of the models and account for the statistical outliers (Marchant et al., 2010). Recent work of Papritz et al. (2012) may be beneficial for water quality modelling, as it allows the use of robust LMM. Another option is to use additional covariates to account for other hydrological processes which improve the ability of the model to fit the unusual observations.

The characterisation of the effect of event-based sampling on load estimation is often performed using continuously sampled water quality. However, as many water quality monitoring studies only have access to limited historical data, simulation based methods offer the ability to assess different water quality sampling schemes based on site-specific models describing the variation in WQ. The geostatistical simulation approach used in this work allowed for the simulation of auto-correlated TP using a LMM to describe the relationship with stream discharge TP. Using simulated TP data it is possible to compare different sampling schemes, and load estimation methods without the requirement of continuously sampled data. Results of this study also support the findings of other studies such as Marsh and Waters (2009), as models which included event-based sampling were more accurate. The procedure outlined in this work provides the ability for site specific analysis without the need of continuous WQ data. In addition, these simulations could be used by catchment managers to compare other forms of sampling schemes and load estimation methods. Using the same procedure outlined in this work, it would be easy to simulate other water quality properties using cheap covariates such as; electrical conductivity and turbidity.

Linking the IES to the catchment characteristics is important to identify the controlling hydrological processes. The results of this study indicate that the combined effect of the range of elevation and annual rainfall have negative correlation with importance of event-based sampling. This indicates that as the elevation range (the catchment becomes flatter) and rainfall in a catchment decrease the improvement of accuracy by including event-based sampling also decreases. This trend indicates that event-based sampling is not as important in catchments with little relief and small annual rainfall, however the inclusion of event-based sampling improved the annual load accuracy of all catchments. The negative relationship between the IES and population density was similar to that of urban cover. This relationship reflects the findings of Sliva and Williams (2001); Johnes (2007) where population density and urban cover has been shown to increase nutrient exports.

3.6 Conclusions

Simulation based methods are necessary for catchment managers and researchers to investigate the impact of different monitoring practices on the accuracy of load estimations. These simulation based methods remove the requirement of access to continuously sampled water quality data to assess different sampling schemes. Using this simulation based approach this work has outlined how;

- The use of event-based sampling improved the accuracy of annual loads for all study catchments.
- The improvement of accuracy was less for catchments with a smaller relief and lower annual rainfall.

- Simulation of TP provided the ability to compare different sampling schemes using limited historical data.

The LMMs used in this work provide the ability to model the temporal auto-correlation in the TP data, however some catchment models were not found to be valid representations of the observed data. The LMM allowed for the simulation of TP based on the LMM for catchments where the LMM was found to be a valid representation. Methods such as this are beneficial in situations with limited data.

3.7 References

Ahearn, D., Sheibley, R., Dahlgren, R., Anderson, M., Johnson, J., Tate, K., 2005. Land use and land cover influence on water quality in the last free-flowing river draining the western sierra Nevada, California. Journal of Hydrology 313 (3), 234–247.

Akaike, H. 1974 A new look at the statistical model identification. Automatic Control, IEEE Transactions on, 19(6), 716-723.

Australian Bureau of Statistics, 2011. 3218.0 population estimates by statistical area level 2, 2001 to 2011.

Banner, E., Stahl, A., Dodds, W., 2009. Stream discharge and riparian land use influence in-stream concentrations and loads of phosphorus from central plains watersheds. Environmental Management 44 (3), 552–565.

Bartley, R., Speirs, W., Ellis, T., Waters, D., 2012. A review of sediment and nutrient concentration data from Australia for use in catchment water quality models. Marine Pollution Bulletin 65 (4), 101–116.

Birkel, C., Soulsby, C., Tetzlaff, D., Dunn, S., Spezia, L., 2011. High-frequency storm event isotope sampling reveals time-variant transit time distributions and influence of diurnal cycles. Hydrological Processes 26 (2), 308–316.

Bivand, R., Pebesma, E., Gómez-Rubio, V., 2008. Applied spatial data analysis with R. Springer Verlag.

Box, G., Cox, D., 1964. An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological), 211–252.

Burt, T., Howden, N., Worrall, F., McDonnell, J., 2011. On the value of long-term, low-frequency water quality sampling: avoiding throwing the baby out with the bathwater. Hydrological Processes 25 (5), 828–830.

Cassidy, R., Jordan, P., 2011. Limitations of instantaneous water quality sampling in surface-water catchments: Comparison with near-continuous phosphorus time-series data. Journal of Hydrology 405 (1), 182–193.

Cohn, T., 2005. Estimating contaminant loads in rivers: An application of adjusted maximum likelihood to type 1 censored data. Water Resources Research 41 (7), W07003.

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

Cooper, D., Watts, C., 2002. A comparison of river load estimation techniques: application to dissolved organic carbon. Environmetrics 13 (7), 733–750.

Crawford, C., 1991. Estimation of suspended-sediment rating curves and mean suspended-sediment loads. Journal of Hydrology 129 (1-4), 331–348.

Davis, J., Koop, K., 2006. Eutrophication in australian rivers, reservoirs and estuaries–a southern hemisphere perspective on the science and its implications. Hydrobiologia 559 (1), 23–76.

de Gruijter, J., Brus, D., Bierkens, M., Knotters, M., et al., 2006. Sampling for natural resource monitoring. Springer Verlag.

Drewry, J., Newham, L., Croke, B., 2009. Suspended sediment, nitrogen and phosphorus concentrations and exports during storm-events to the Tuross Estuary, Australia. Journal of Environmental Management 90 (2), 879–887.

Francois, R., Eddelbuettel, D., Bates, D., 2011. Rcpparmadillo: Rcpp integration for Armadillo templated linear algebra library. R package version 0.2 34.

Gao, P., 2008. Understanding watershed suspended sediment transport. Progress in Physical Geography 32 (3), 243–263.

Gebbers, R., Bruin, S., 2010. Application of geostatistical simulation in precision agriculture. Geostatistical Applications for Precision Agriculture, 269–303.

Hopmans, P., Bren, L., 2007. Long-term changes in water quality and solute exports in headwater streams of intensively managed radiata pine and natural eucalypt forest catchments in south-eastern Australia. Forest Ecology and Management 253 (1-3), 244–261.

Johnes, P., 2007. Uncertainties in annual riverine phosphorus load estimation: impact of load estimation methodology, sampling frequency, baseflow index and catchment population density. Journal of Hydrology 332 (1-2), 241–258.

Jones, A., Stevens, D., Horsburgh, J., Mesner, N., 2011. Surrogate measures for providing high frequency estimates of total suspended solids and total phosphorus concentrations. JAWRA Journal of the American Water Resources Association 47 (2), 239–253.

Kim, J., Furumai, H., 2012. Improved calibration of a rainfall-pollutant-runoff model using turbidity and electrical conductivity as surrogate parameters for total nitrogen. Water and Environment Journal.

Kristiana, R., Vilhena, L., Begg, G., Antenucci, J., Imberger, J., 2011. The management of lake burragorang in a changing climate: The application of the index of sustainable functionality. Lake and Reservoir Management 27 (1), 70–86.

Kronvang, B., Bruhn, A., 1996. Choice of sampling strategy and estimation method for calculating nitrogen and phosphorus transport in small lowland streams. Hydrological Processes 10 (11), 1483–1501.

Kuhnert, P., Henderson, B., Lewis, S., Bainbridge, Z., Wilkinson, S., Brodie, J., 2012. Quantifying total suspended sediment export from the Burdekin river catchment using the loads regression estimator tool. Water Resources Research 48 (4), W04533.

Lark, R., Cullis, B., 2004. Model-based analysis using REML for inference from systematically sampled data on soil. European Journal of Soil Science 55 (4), 799–813.

Lawrence, I., Lin, K., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics, 255–268.

Lewis, J., 1996. Turbidity-controlled suspended sediment sampling for runoff-event load estimation. Water Resources Research 32 (7), 2299–2310.

Marchant, B., Saby, N., Lark, R., Bellamy, P., Jolivet, C., Arrouays, D., 2010. Robust analysis of soil properties at the national scale: cadmium content of French soils. European Journal of Soil Science 61 (1), 144–152.

Marsh, N., Waters, D., 2009. Comparison of load estimation methods and their associated error. In: 18th IMACS World Congress MODSIM. Vol. 4. Cairns.

Minasny, B., McBratney, A., 2007. Spatial prediction of soil properties using eblup with the matérn covariance function. Geoderma 140 (4), 324–336.

Papritz, A., Künsch, H., Schwierz, C., Stahel, W., 2012. Robust geostatistical analysis of spatial data. In: EGU General Assembly Conference Abstracts. Vol. 14. p. 3025.

Pebesma, E., 2004. Multivariable geostatistics in s: the gstat package. Computers & Geosciences 30 (7), 683–691.

R Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. http://www.R-project.org/

Ribeiro Jr, P., Diggle, P., 2001. geoR: A package for geostatistical analysis. R news 1 (2), 14–18.

Salles, C., Tournoud, M., Chu, Y., 2008. Estimating nutrient and sediment flood loads in a small mediterranean river. Hydrological Processes 22 (2), 242–253.

Sanderson, C., 2010. Armadillo: An open source c++ linear algebra library for fast prototyping and computationally intensive experiments. Tech. rep., Technical report, NICTA.

Sliva, L., Williams, D., 2001. Buffer zone versus whole catchment approaches to studying land use impact on river water quality. Water Research 35 (14), 3462–3472.

Standard Methods, 1995. Standard Methods for the Examination of Water and Wastewater.

Thomas, R., 1985. Estimating total suspended sediment yield with probability sampling. Water Resources Research 21 (9), 1381–1388.

Thomas, R., 1988. Monitoring baseline suspended sediment in forested basins: the effects of sampling. Hydrological sciences journal 33 (5), 499–514.

Thomas, R., Lewis, J., 1993. A comparison of selection at list time and time-stratified sampling for estimating suspended sediment loads. Water Resources Research 29 (4), 1247–1256.

Thomas, R., Lewis, J., 1995. An evaluation of flow-stratified sampling for estimating suspended sediment loads. Journal of Hydrology 170 (1-4), 27–45.

Wang, Y., Kuhnert, P., Henderson, B., 2011. Load estimation with uncertainties from opportunistic sampling data–a semiparametric approach. Journal of Hydrology 396 (1), 148–157.

Webb, B., Phillips, J., Walling, D., 2000. A new approach to deriving best-estimate chemical fluxes for rivers draining the lois study area. Science of the total environment 251, 45–54.

Chapter 4

Estimating the effect of sample size on the precision of total phosphorus mean estimates with limited historical data.

4.1 Summary

The effect of the sample size on the precision of the mean is well understood. However many water quality studies have relied on continuous sampling to assess this relationship. Therefore, catchment managers require a method to estimate this relationship using historical data. By characterising the effect of sample size on the precision of the mean it is possible for catchment managers to adjust the sample size in relation to both the cost and the precision of the mean estimate. Historical data is often sparse and generally collected using several different sampling schemes requiring a model-based approach to provide meaningful analyses. Using total phosphorus data from 17 sub-catchments in south east Australia the ability of a model-based approach to estimate the effect of sample size on the precision of mean concentrations is examined. The effect of sample size on the precision of the mean estimate is examined for base *and event-flow conditions. The effect of catchment characteristics on the precision of the mean estimates was also examined. The results showed annual base flow conditions require similar sample sizes to event-flow estimates, with little improvement in precision above 12 samples for both flow conditions. In addition the precision of both base and event flow was related to the percentage of urban cover in each subcatchment. This chapter outlines a method in which a model-based approach can be used to examine the relationship between the sample size and mean total phosphorus estimates.*

4.2 Introduction

Water quality monitoring is undertaken to understand the state of a catchment, to help with management strategies and to identify water quality trends. Most monitoring schemes attempt to determine *how much* and *when* nutrients are exported from a river. In many Australian catchments the majority of nutrients are exported during rainfall events (Drewry et al., 2009). Many monitoring schemes reflect this with an increased sample size during these periods. These samples are often used to provide estimates of the exported nutrient load. The accuracy of these load estimates is directly related to the monitoring scheme, in particular the sample size. Therefore a critical question that must be answered for a monitoring scheme is; *how many* samples should be taken during events and base flow?

The effect of the sampling schemes on water quality studies are often carried out based on near continuous data (Strobl and Robillard, 2008). These studies are limited, as the collection of near continuous data is extremely expensive. The general approach of these studies is;

1. collect near-continuous water quality data,
2. sample the water quality data using different sampling schemes,
3. use various load estimation techniques to estimate the load based on the sampled data for a given period, and

4. use the accuracy and bias of the predictions to compare the different sampling schemes.

Kronvang and Bruhn (1996) investigated different estimation methods and sampling schemes on two catchments in Denmark. In this study several different sampling schemes were compared including a combination of event and base-flow sampling; where several samples are collected during an event while fewer samples are collected during base-flow. The study found that the most accurate sampling scheme was fortnightly sampling. A study conducted by Johnes (2007) compared several different sampling schemes for 17 catchments in the UK and found daily sampling schemes were the most effective, but highly unlikely to be implemented due to costs. With this Johnes (2007) suggests the use of a combination of weekly sampling and a form of annual high flow sampling as a compromise. More recently Cassidy and Jordan (2011) examined different sampling schemes and load estimation methods using near-continuous total phosphorus (TP) data from 3 small catchments in Ireland. Their results showed that sampling schemes needed to include storm events of different sizes to minimise uncertainty in load estimates. These studies have highlighted the importance of sampling events and sampling frequencies on the accuracy of load estimates, however this approach requires near continuous observations of water quality properties. A disadvantage of these studies are that the results are site and scale specific and do not necessarily give good estimates of sample size at other sites. They are also expensive so cannot be performed at all water quality monitoring sites.

The site-specific nature of variation in WQ is shown by several studies that have investigated the relationship between catchment characteristics (CCs) and water quality variables. These variables include catchment size, slope, rainfall, stream discharge, land use and land cover. The aim of these studies is to determine if a relationship exists between water quality concentrations across multiple catchments. Relationships between phosphorus concentration and agriculture (Ahearn et al., 2005) and catchment size (Banner et al., 2009) have been shown to exist in a variety of catchments. In a study of 25 sites within Kansas, USA, Banner et al. (2009) found no relationship between annual stream discharge and phosphorus, but did find a relationship between phosphorus and land use and catchment topography. Grassland cover has also been found to have a relationship with nitrogen concentrations (Ahearn et al., 2005). Sliva and Williams (2001) found a relationship between both forest cover and slope with several different water quality variables. In addition they found percentage of urbanisation had the greatest effect on water quality. These relationships offer valuable information for the design and implementation of sampling schemes in catchments without prior information.

Sampling schemes can be divided into two categories; design-based and model-based (de Gruijter et al., 2006). Design-based sampling relies on a form of randomised sampling with known inclusion probabilities to estimate the sample statistics and their associated uncertainty, while model-based sampling relies on a model to provide unbiased estimates of the uncertainty (de Gruijter et al., 2006). An example of design-based sampling is given by Thomas and Lewis (1993); in which stratified random sampling is used during events. Thomas and Lewis (1993) explain that the benefit of this method is that the mean and its variance are unbiased and simple to derive because there is no requirement of a model to estimate the uncertainty. Most monitoring schemes do not have access to continuously sampled water quality properties, however many do have access to historical data. It is therefore possible to examine the relationship between samples sizes and the standard error of the mean (SEM). By using this information it is possible to determine the required sample sizes to meet a desired accuracy for estimation of the mean concentration. In addition to this, it is possible to provide estimates of the associated costs of the required sample sizes (de Gruijter et al., 2006). This information is required to improve monitoring schemes in relation to monitoring costs and to improve the accuracy of mean estimates. The method outlined in this study can be extended to relate the uncertainty of the mean estimate to the financial costs related to the sample size. These methods are outlined in Särndal et al. (2003, p 106). As it is difficult to characterise the costs associated with water quality sampling such as travel times, laboratory analysis.

It is common for water quality monitoring schemes to have used several different sampling schemes or not have a record of the sampling design (Bartley et al., 2012).

Therefore a method that uses historical data regardless of the sampling scheme is required to estimate the effect of sample size on the precision of mean estimates is required. A model-based approach is provided by Domburg et al. (1994), where a variogram is used to estimate the SEM of the sampled property. They provide an example of using a variogram to model the variance of the mean phosphate in soil, and the effect of sample size. By combining land use maps with a small sample Visschers et al. (2007) used the procedure of Domburg et al. (1994) to design a more efficient sampling design for measuring the phosphate sorption capacity of soil. By using this method the sampling cost was reduced by 13% Visschers et al. (2007).

Using TP data collected from 17 sub-catchments west of Sydney, Australia, this chapter will apply the method outlined by Domburg et al. (1994) to estimate the SEM for different sample sizes. The analysis will use this method to model the variance between the samples and estimate SEM of TP focusing on the following questions;

1. Is there a difference in sample size requirements between event-flow and base-flow?
2. Can the relationship between sample size and the precision of mean estimates be linked to CCs?

4.3 Methods

4.3.1 Catchment description

The 17 sub-catchments are within the main supply area for Sydney's drinking water. The region is located west of Sydney with the location of each stream gauging station shown in figure 4.1. Table 4.1 provides a summary of the topographical and hydrological characteristics of each sub-catchment. In table 4.1 each sub-catchment has been grouped into a larger catchment based on the main river of the catchment. The location of each sub-catchment in table 4.1 is provided in figure 4.1. There is a wide variety of sub-catchment sizes, ranging from 7.27 km^2 to 4852.12 km^2. The rainfall was determined for each sub-catchment using Thessen polygons with data obtained from nearby Bureau of Meteorology weather stations. Across the region annual rainfall tends to be higher to the east, near the coastline and in the north west area. Forest is the dominant land cover for all sub-catchments except for the sub-catchments within the Wollondilly catchment which are dominated by grass/pasture. The distribution of event lengths for each sub-catchment is positively skewed. The mean and the median event lengths for each sub-catchment is provided in table 4.1.

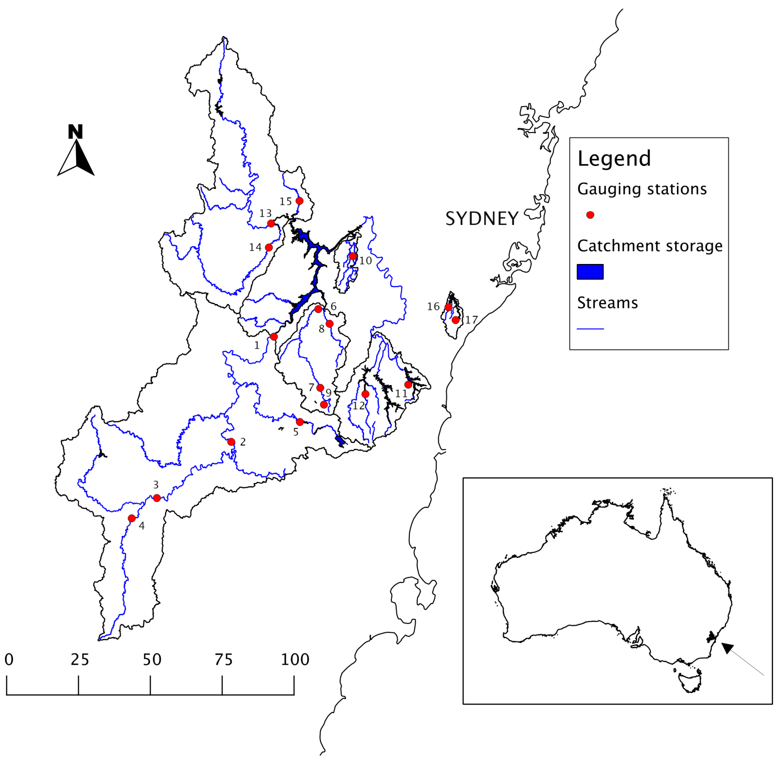

Figure 4.1 – Location of sub-catchment stream gauging stations.

4.3.2 Data description

For each sub-catchment water quality samples were collected using a combination of monthly grab samples and event-based samples. The event-based samples were

Table 4.1 – Key features for each sub-catchment

Site name	sub-catchment number	Area (km^2)	Minimum elevation (m)	Maximum elevation (m)	Annual discharge ($GL\ yr^{-1}$)	Annual rainfall ($mm\ yr^{-1}$)	Mean event length (days)	Median event length (days)	Dominant land cover	Urban (%)
Wollondilly catchment										
Jooriland	1	4825.1	108	1179	4086.9	708.2	4.61	2.71	*grassland*	1
Golden Valley	2	1568.6	599	1020	881.1	614.6	4.54	2.92	*grassland*	1
Golden Valley	3	2192.4	544	1020	1995.3	631.0	3.23	2.19	*grassland*	1
Towers Weir	4	604.7	599	1020	313.8	611.4	4.53	3.42	*grassland*	0
Berrima Weir	5	201.4	620	868	1249.2	959.3	3.38	1.5	*grassland*	10
Nattai catchment										
Smallwoods Crossing	6	436.9	105	867	576.7	857.2	4.45	2.54	*forest*	2
The Crags	7	90.8	429	867	193.1	878.5	2.59	1.52	*forest*	8
Little River	8	104	175	632	138	824.7	4.23	2.94	*forest*	0
Mittagong	9	8	581	867	35.5	879.2	1.05	0.63	*forest*	31
Werriberri catchment										
Werombi	10	56.4	203	519	240.9	695	4.72	3	*forest*	1
Upper Nepean catchment										
Sandy Creek	11	7.3	326	518	37.1	947	3.17	1.79	*forest*	0
Burke River	12	88.3	328	777	420.9	1182.5	4	2.25	*forest*	1
Coxs catchment										
Kelpie Point	13	1447	116	1342	2968.8	886.7	5.17	3.69	*forest*	0
Cedar Ford	14	719.1	169	1382	3445.5	932.4	4.83	3.21	*forest*	0
Kedumba Crossing	15	72.5	152	1061	341.5	1263.6	2.07	1.13	*forest*	13
Woronora catchment										
Woronora River	16	13.1	159	380	50.4	1383.5	3.3	2.04	*forest*	0
Flatrock Crossing	17	20.2	187	403	131.1	1387.9	2.8	2.04	*forest*	0

mostly collected using an automatic sampler designed to collect samples during rainfall events. For this study we focus on TP and we divided the observations based on base and event-flow, where event-flow was defined as greater than the annual 90th percentile of stream discharge. Table 4.2 provides an summary of base and event-flow observations. In most sub-catchments the event-flow mean is greater than the base-flow mean. In three sub-catchments noticeable step changes were observed in concentration at three sampling locations directly downstream of sewage treatment plants (STP)s. These changes coincided with upgrades or the decommissioning of these STPs. Due to these changes data from sampling sites downstream of these STPs were restricted to the period after the change to ensure land use was consistent for the period of sampling.

4.3.3 Statistical analysis

In water quality studies with known inclusion probabilities the mean of a sampled property $\bar{z}$ is estimated using;

$$\hat{\bar{z}} = \sum_{i=1}^{n} \pi_i z(t_i) \tag{4.1}$$

where n is the sample size, π_i is the weight of sample $z(t_i)$ of the i^{th} sample through time t. The weights of each sample are based on the inclusion probabilities of each sample using design p where $1 = \sum_{i=1}^{n} \pi_i$ (Domburg et al., 1994). Sampling design p is unbiased over repeated sampling and therefore the expected estimation of the mean under this design is defined as;

$$E_p\left[\hat{\bar{z}}\right] = \bar{z}. \tag{4.2}$$

Using this, a measure of accuracy of the estimated mean can be assessed using the standard error of the mean (SEM);

$$r = \sqrt{E_p\left[\left(\hat{\bar{z}} - \bar{z}\right)^2\right]}. \tag{4.3}$$

Table 4.2 – Sub-catchment total phosphorus summary statistics.

Site number	Sampling period		base-flow			event-flow		
	start	end	mean (mg L^{-1})	std. dev. (mg L^{-1})	n	mean (mg L^{-1})	std. dev. (mg L^{-1})	n
Wollondilly catchment								
Jooriland	Dec 1993	Jan 2012	0.02	0.03	285	0.09	0.17	330
Golden Valley	Jan 1991	Apr 2012	0.02	0.04	352	0.09	0.15	360
Murrays Flat	Jan 2005	Mar 2012	0.05	0.08	124	0.09	0.08	334
Towers Weir	Jan 1991	Mar 2012	0.06	0.07	236	0.08	0.05	223
Berrima Weir	Dec 1993	Dec 2011	0.05	0.05	244	0.10	0.10	255
Nattai catchment								
Smallwoods Crossing	Jan 1991	Apr 2010	0.04	0.18	299	0.08	0.09	153
The Crags	Jan 2002	Jan 2012	0.04	0.04	164	0.08	0.10	320
Little River	Aug 1991	May 2007	0.01	0.01	208	0.07	0.31	109
Mittagong	Sep 2001	Dec 2011	0.20	0.53	127	0.24	0.23	40
Werriberri catchment								
Werombi	Jan 1991	Feb 2012	0.01	0.02	306	0.04	0.06	431
Upper Nepean catchment								
Sandy Creek	Jul 2007	Jan 2012	0.01	0.01	58	0.01	0.01	32
Burke River	Jan 2003	Jan 2012	0.01	0.01	181	0.01	0.01	240
Coxs catchment								
Kelpie Point	Jan 1991	Feb 2012	0.02	0.10	315	0.05	0.14	297
Cedar Ford	Jan 1991	Apr 2012	0.04	0.21	405	0.04	0.09	448
Kedumba Crossing	Jan 2002	Mar 2012	0.02	0.04	132	0.04	0.03	208
Woronora catchment								
Woronora River	Mar 2007	Feb 2012	0.01	0.01	90	0.01	0.01	152
Flatrock Crossing	Mar 2007	Mar 2012	0.01	0.01	62	0.02	0.01	158

If the period of time and the mean $\hat{\hat{z}}$ are assumed to be fixed, r equals the sampling variance (Domburg et al., 1994). Many water quality sampling schemes do not use a form of probabilistic sampling (i.e the inclusion probabilities are unknown) (Thomas and Lewis, 1995). Therefore the samples cannot be assumed to be independent and a form of temporal correlation must be considered. In this situation, where non-probabilistic sampling has been performed it is necessary to use a model to estimate the variance of the samples (Lark and Cullis, 2004; Domburg et al., 1994). The procedure outlined by Domburg et al. (1994) provides a method which uses a model-based approach to estimate the design-based SEM. Figure 4.2 is a flowchart which outlines the methodology used in this chapter. We will now outline the steps in detail.

4.3.3.1 Variogram estimation

To estimate the temporal correlation between the samples, the observed concentration through time is modelled as;

$$z(\mathbf{t}) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}, \tag{4.4}$$

where the $(n \times 1)$ design matrix $(\mathbf{X})$ is a vector of 1's and $\boldsymbol{\beta}$ is the estimate of the mean conditional on the correlation structure of $\boldsymbol{\eta}$ using the REML likelihood function (see Lark and Cullis (2004) for a detailed description). The covariance structure $\boldsymbol{\eta} \sim \mathcal{N}(0, \mathbf{V})$ is assumed to be in the form of an exponential function where $\mathbf{V}$ is defined as;

$$\begin{aligned} \mathbf{V}_{i,j} = \quad & \sigma^2 s \exp\left(-\tfrac{|\mathbf{x_i}-\mathbf{x_j}|}{a}\right), \quad i \neq j \\ & \sigma^2, \qquad\qquad\qquad\qquad i = j, \end{aligned} \tag{4.5}$$

where the diagonal is equal to the variance σ^2 and the covariance between two samples is defined by the exponential function where $|\mathbf{x_i} - \mathbf{x_j}|$ is the temporal distance between samples and a is the distance parameter of the exponential function and

$$s = \frac{c}{c_0 + c}, \tag{4.6}$$

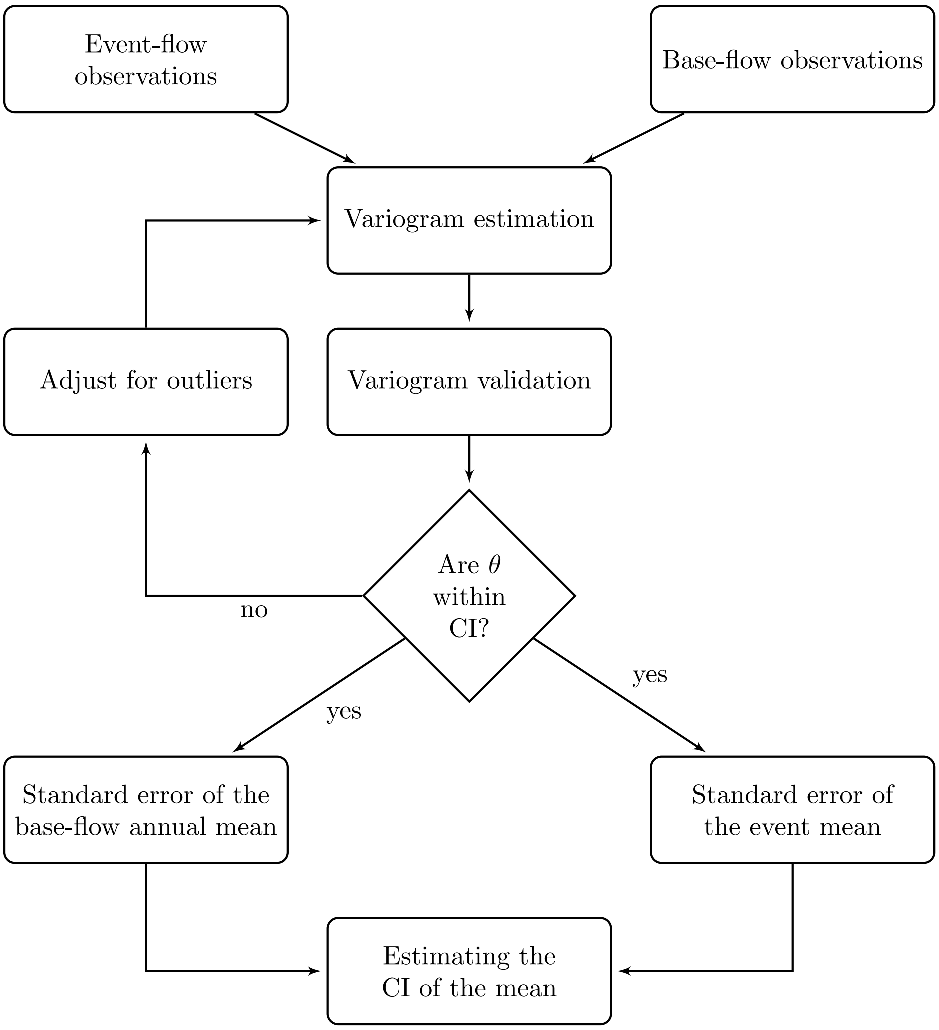

Figure 4.2 – Procedure for estimating the standard error of the mean.

where c_0 is the unexplained temporal variance (nugget) and $c_0 + c$ is the maximum variance between two samples (Lark and Cullis, 2004). It would also be possible to used a different model form to account for the auto-correlation between samples in the LMM. Equation 6.4 assumes the observed $z(\mathbf{t})$ has a Gaussian distribution, however this is highly unlikely for water quality concentration values. A flexible transformation to overcome this problem is a box-cox transformation (Box and Cox, 1964). The box-cox transformed concentration is defined as;

$$z(\mathbf{t})^* = \begin{cases} \log z(\mathbf{t}) & \text{if } \lambda = 0 \\ \frac{z(\mathbf{t})^\lambda - 1}{\lambda} & \text{otherwise,} \end{cases} \tag{4.7}$$

where the value of λ is estimated within the fitting of the model. The model was fitted using the geoR package (Ribeiro Jr and Diggle, 2001) within the R statistical software (R Core Team, 2012)

4.3.3.2 Variogram validation

There are two steps in the validation of each model:

1. The temporal correlation structure is tested for statistical significance, and
2. leave-one-out-cross-validation (LOOCV) is undertaken to validate the prediction variance estimates of each model.

To determine if the temporal structure of the model is significant, the model is fitted with and without the temporal correlation structure. It is then possible to use the Akaike Information Criteria (AIC) (Akaike, 1974) of these two models to determine if the temporal structure of the model is significant. Using the predictions of the LOOCV it is possible to calculate,

$$\theta_i = \frac{\{z(\mathbf{t}_i)^* - \tilde{Z}^*_{(-i)}\}^2}{\sigma^2_{(-i)}}, \tag{4.8}$$

where $z(\mathbf{t}_i)^*$ is the i^{th} transformed concentration observation, $\tilde{Z}^*_{(-i)}$ is the corresponding estimated concentration based on all other samples and $\sigma^2_{(-i)}$ is the estimated kriging variance. Since the approach here is dependent on the semivariogram representation of the variation in the data, the model needs to be valid. Here a perfect model has a mean $\bar{\theta} = 1$ and median $\breve{\theta} = 0.455$. We use the working of Marchant et al. (2010) to estimate the confidence intervals for the mean and median θ. Essentially if our prediction variance (the denominator in eqn 4.8) represents the actual error (numerator in eqn 4.8) then the model is representing the variation well. If our theta statistics fall with the CI then we assume the model is valid.

4.3.3.3 Adjusting for statistical outliers

A potential problem in water quality modelling is the occurrence of observations which appear to be temporally localised anomalies. These observations may be due to incorrect data entry, measurement error or due to insufficient observations of the underlying processes (Díaz Muñiz et al., 2012). In the last situation this could occur because there are too few observations of extreme events, so the ones that do exist look like outliers relative to other observations. Based on the work presented by Marchant et al. (2010) we have applied a winsorising method for models with suspected statistical outliers. The winsoring method is applied to models where the θ statistics do not fall within the 95% CI.

1. Using a similar approach to LOOCV the kriging weights ($kw_{j(-i)} = 1, ..., i-1, i+1, ..., n$) are estimated, as is the kriging variance $\sigma_{(-i)}$of each observation.

2. For each observation the weighted median is calculated. The weighted median solves;

$$\sum_{j=1, j\neq i}^{n} kw_{j(-i)} \mathrm{sign}\{\breve{z}_{(-i)}(t_i) - z(t_j)\} = 0 \tag{4.9}$$

3. Using the weighted median the data $z(\mathbf{t}_i)$ is winsorized using;

$$z_{1.5}(t_i) = \begin{cases} \breve{z}_{(-i)} + z\sigma_{-i} & \text{if } z_i(t_i) - \breve{z}_{(-i)} > 1.5\sigma_{(-i)} \\ z(t_i) & \text{if } |z_i(t_i) - \breve{z}_{(-i)}| \leq 1.5\sigma_{(-i)} \\ \breve{z}_{(-i)} - z\sigma_{-i} & \text{if } z_i(t_i) - \breve{z}_{(-i)} < -1.5\sigma_{(-i)}, \end{cases} \tag{4.10}$$

where $\breve{z}_{(-i)}$ is the weighted median for the observation at t_i, estimated without the inclusion of this observation. The model is then fitted using the winsorized data. With the updated model LOOCV is performed and if both $\bar{\theta}$ and $\breve{\theta}$ are within the 95% CI the model is accepted.

4.3.3.4 Estimating the standard error of the mean

Domburg et al. (1994) describes the process to use a fitted variogram model to estimate the design based SEM. The estimated SEM ($\bar{r}$) is determined using the model based variance from the fitted variogram by modifying eqn 4.3:

$$\bar{r} = E\left[\frac{\bar{\gamma}_T}{n}\right], \tag{4.11}$$

where $\bar{\gamma}_T$ is the mean semivariance of the model, T is the temporal length of interest (e.g. a year) and n is the number of samples. Based on Domburg et al. (1994) the procedure to estimate the sample variance using a variogram model is;

1. Estimate $\bar{\gamma}_T$ by drawing two random times within period T and determining the semivariance between the two samples. Repeat this 10,000 times to estimate the mean semivariance between all pairs in T.

2. Estimate the SEM ($\bar{r}$) for different sample sizes (n).

The values of n are different for both event and base-flow observations, and based on commonly applied sample sizes. The sample design assumes simple random sampling over the period of T, which is defined as 365 days for base-flow model, and the mean event flow duration based on the annual 90th percentile for event-flow models. For each site the mean event flow duration was estimated based on the duration of events which were greater than the annual 90% stream discharge. The mean event duration for each site is presented in Table 4.1. Sample sizes are defined as 365, 52, 24, 12, 4 and 48, 24, 12, 6 for base-flow and event-flow respectively. The sizes were chosen based on commonly used sample frequencies daily, weekly, fortnightly, monthly and quarterly for routine based sampling. Sample sizes for event-based models are based on the use of an automatic sampler capable of collecting differing numbers of samples per event.

4.3.3.5 Estimating the CI of the mean.

In order to compare the effect of sample size on the precision of the estimate of the mean (as represented by the 95% confidence interval) both the mean and the confidence intervals (CI) are back transformed to the original scale. The mean is taken as the β_0 estimate of the model and is back-transformed using;

$$\mu = (1 + \lambda\beta_0)^{\frac{1}{\lambda-2}} (1 + \lambda\beta_0)^2 + \frac{1-\lambda}{2}\sigma^2. \tag{4.12}$$

The upper and lower bounds of confidence intervals are estimated using;

$$CI = \beta_0 \pm t_{0.975}^{n-1}\sqrt{\bar{r}n}, \tag{4.13}$$

where $t_{0.975}^{n-1}$ is the t-statistic for the 0.975 quantile with $n-1$ degrees of freedom. The CI is then back-transformed by substituting β_0 with the lower and upper bound of the CI in equation 4.12.

4.3.3.6 Relating the precision of the estimate of the mean to catchment characteristics

To explore the relationship between the precision of the mean estimate and the catchment characteristics (CC) a standardised CI is used. Standardisation of the CI for each sub-catchment is required as box-cox transformations with different lambda values have been applied. In addition to different transformations, each catchment has a different mean TP. The standardisation is performed using;

$$CI_{std} = \frac{CI_n}{\beta_0}, \tag{4.14}$$

where n is 12 observations for base-flow and 24 observations for event-flow. Eqn 4.14 standardizes each CI relative to its mean enabling a comparison between catchments. It is based on the coefficient of variation which enables comparison of variation between datasets with different means. Two sample sizes are chosen as they reflect typical values; monthly sampling for base-flow and 24 samples for events as this reflects the amount of samples within an automatic sampler.

4.4 Results

Table 4.3 provides a summary of the coefficients of each model (if valid) and indicates if winsoring was applied. Based on the results of the cross validation analysis 12 base-flow models and 7 event-flow models were accepted. Of the accepted models 5 base-flow and 2 event-flow models required the data to be winsorised. Only 5 of the 17 sub-catchments had both a base-flow and event-flow model pass the cross validation requirements. Outliers at some locations and observations at the minimum detection limit may have contributed to the rejection of some of the models. In addition to these potential causes, some gauging stations had limited samples during events, which may have been identified as outliers.

The mean temporal correlation of the base-flow models was 38.9 days with 9 of the 12 models having a temporal correlation greater than 30 days. Event-flow models had a shorter temporal correlation structure with a mean of 7.1 days and 4 of the 7 models with a shorter correlation than 9 days.

Table 4.3 – Summary of model coefficients for event and base-flow models (models fitted using winsorized data are indicated by *).

	base-flow						event-flow					
Site number	$\beta 0$	c_0	$c_0 + c$	a	λ	w	$\beta 0$	c_0	$c_0 + c$	a	λ	w
Wollondilly catchment												
Jooriland	-8.05	1.38	4.31	44.05	-0.24	*						
Golden Valley	-6.83	0.51	2.26	46.97	-0.19	*						
Murrays Flat	-3.43	0.27	1.32	62.61	0.01		-2.15	0.01	0.53	10.24	0.18	*
Towers Weir	-5.81	1.00	4.18	47.78	-0.40	*						
Berrima Weir	-3.12	0.02	0.22	30.25	0.03	*						
Nattai catchment												
Smallwoods Crossing							-5.44	0.91	5.26	4.20	-0.30	
The Crags	-2.38	0.04	0.15	60.85	0.24							
Little River	-10.40	3.81	7.83	61.50	-0.21		-9.50	5.11	26.02	11.84	-0.27	
Mittagong	-3.02	0.05	2.38	5.11	-0.18		-1.87	0.87	0.98	2.93	0.02	
Werriberri catchment												
Werombi												
Upper Nepean catchment												
Sandy Creek	-44.97	1.99	190.14	8.01	-0.65	*	-132.12	173.00	9496.04	10.69	-0.95	
Burke River												
Coxs catchment												
Kelpie Point							-5.81	0.43	4.35	8.23	-0.19	*
Cedar Ford												
Kedumba Crossing	-20.47	49.71	71.58	63.37	-0.49		-3.27	0.11	0.65	1.74	0.05	
Woronora catchment												
Woronora River	-93.40	734.54	1738.04	1.34	-0.78							
Flatrock Crossing	-137.48	1139.90	4184.68	34.87	-0.88							

4.4.1 Differences between the precision in the estimates of the base and event flow means

Using Kedumba Crossing (sub-catchment 15) a detailed explanation is given of how the estimates of estimates of the precision of the mean is estimated using the procedure outlined in figure 4.2. The first stage of the process is to estimate a variogram model which reflects the variance of the observed data. Figure 4.3 shows the exponential model of the base-flow and event-flow used to model the temporal variance structure.The figure shows how the temporal range a is different for both models. For observations further apart than $\sim 3a$ there is no temporal correlation. The temporal range for the base-flow and event-flow models is 63.37 and 1.74 days respectively.

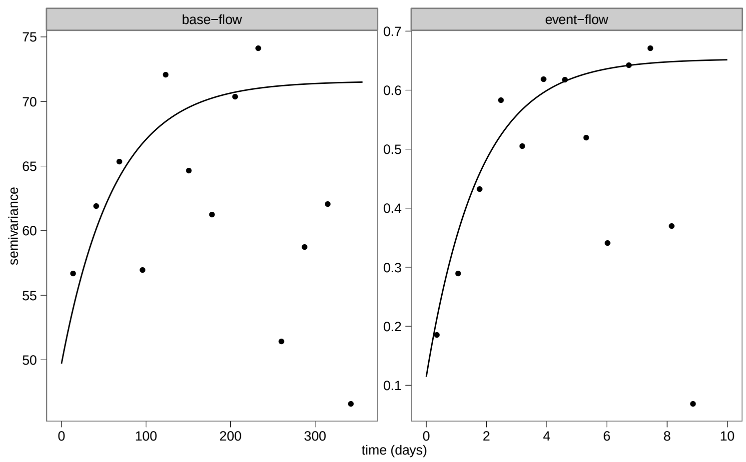

Figure 4.3 – Variogram models of base and event-flow for Kedumba Crossing.

Using the modelled variance the next stage is the estimation the SEM. Figure 4.4 shows the estimated SEM for both the base and event-flow models. The SEM is on the box-cox transformed scale, therefore direct comparisons of the standard error cannot be made between event and base-flow. However the trend between sample size and the standard error can be observed. The reduction of the base-flow standard error is greatest when $n < 100$, increasing n when $n > 100$ has little effect on the SEM.

To simplify the effect of sample size on the precision of the mean estimates, and allow

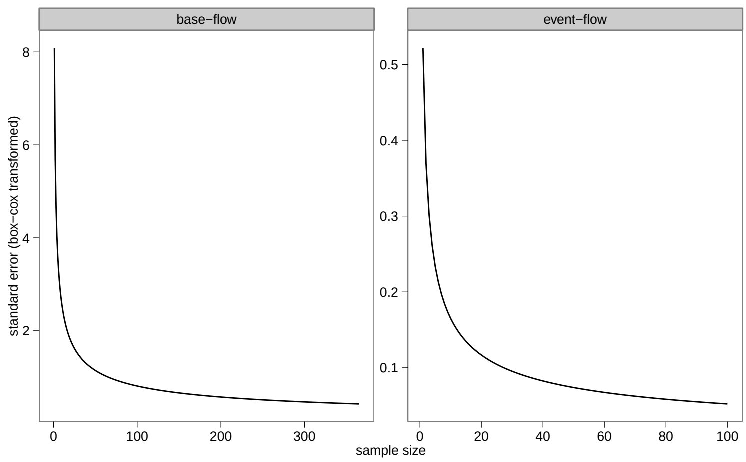

Figure 4.4 – Estimated standard error of base and event-flow for Kedumba Crossing.

comparisons to be made, the model estimates of the mean and the CI were back-transformed. Figure 4.5 shows the back-transformed means, and CI of base-flow and event-flow for commonly used sample sizes. The mean TP is larger during event-flow than during base-flow conditions. The figure shows that the CI of the mean based on 48 event-flow samples is larger than the CI of 12 base-flow samples. The CI of base-flow mean with 4 samples for a year is relatively large compared with the CI of 12 samples per year. This difference is partly contributed by the different *df* associated with each of CI, where the *df* $= 3$ when $n = 4$ and *df* $= 11$ when $n = 12$. Based on the selected sample sizes of event-flow, there is not a linear reduction of the CI when the sample size is increased. For example, doubling the sample size from 12 to 24 samples does not half the CI of the mean. The CI of the mean for different sample sizes decreases as the sample size increases with a noticeable reduction in uncertainty between 6 and 12 samples. There is a reduction of the CI between 24 and 48 samples, however for most water quality sampling schemes a sample size of 48 would require 2 automatic samplers which would double the installation costs.

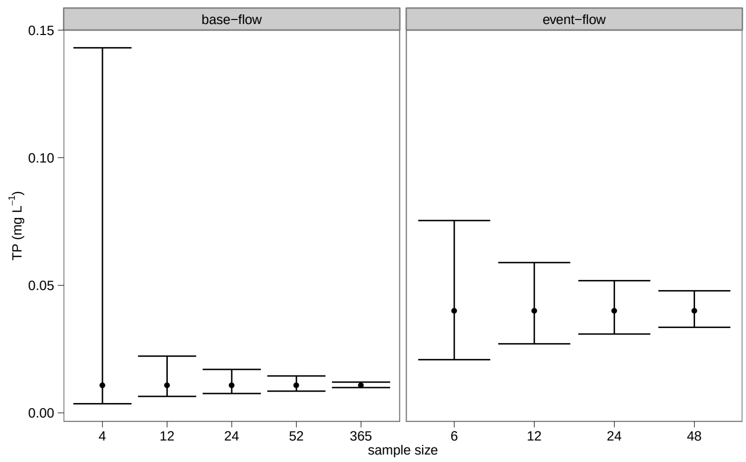

Figure 4.5 – Estimated CI and model estimated means of base and event-flow for Kedumba Crossing.

4.4.2 Sub-catchment comparisons

Figure 4.6 shows the mean and the CI of the base-flow models for each sub-catchment. As expected the the CI decreases as the sample size increases for all sub-catchments. The effect of sample sizes on the CI is different for each sub-catchment. For example Berrima Weir and The Crags sub-catchments have similar means, 0.043 and 0.038 respectively and similar CI for each sample size. The sub-catchments generally have similar CI if the mean concentrations are similar. The Mittagong sub-catchment is quite different to the other sub-catchments. This sub-catchment has the highest mean TP (0.14) of all sub-catchments and the largest CI width of 1.19 with $n = 4$. The sampling site for this sub-catchment is directly downstream of a STP, which may explain the high TP concentration.

The mean and CI of the event-flow models of each sub-catchment are shown in figure 4.7. Kelpie Point and Kedumbda Crossing sub-catchments have similar means, 0.03 and 0.04 respectively and similar CI widths for $n = 6$ of 0.04 and 0.06 respectively. As with base-flow, the Mittagong sub-catchment has the highest mean (0.23) of the sub-catchments, and the largest CI width (0.53) with $n = 6$. This result indicates waste water treatment plant effects both base and event-flow conditions. In addition to the Mittagong sub-catchment, Murrays Flat gauging station is also downstream of a waste water treatment plant, this site had the second highest mean TP concentration during events (0.1).

4.4.3 Is there a difference between the precision of the base and event-flow mean estimates?

Figure 4.8 shows the distribution of CI for each sample size of both base and event-flow based models. There is a clear difference in the CI between base and event flow estimates. By examining the distribution of CI with $n = 12$ it is clear that there is more uncertainty in the estimate of the mean for event-flow periods, even though the samples are taken within a shorter period of time. The reduction of the uncertainty

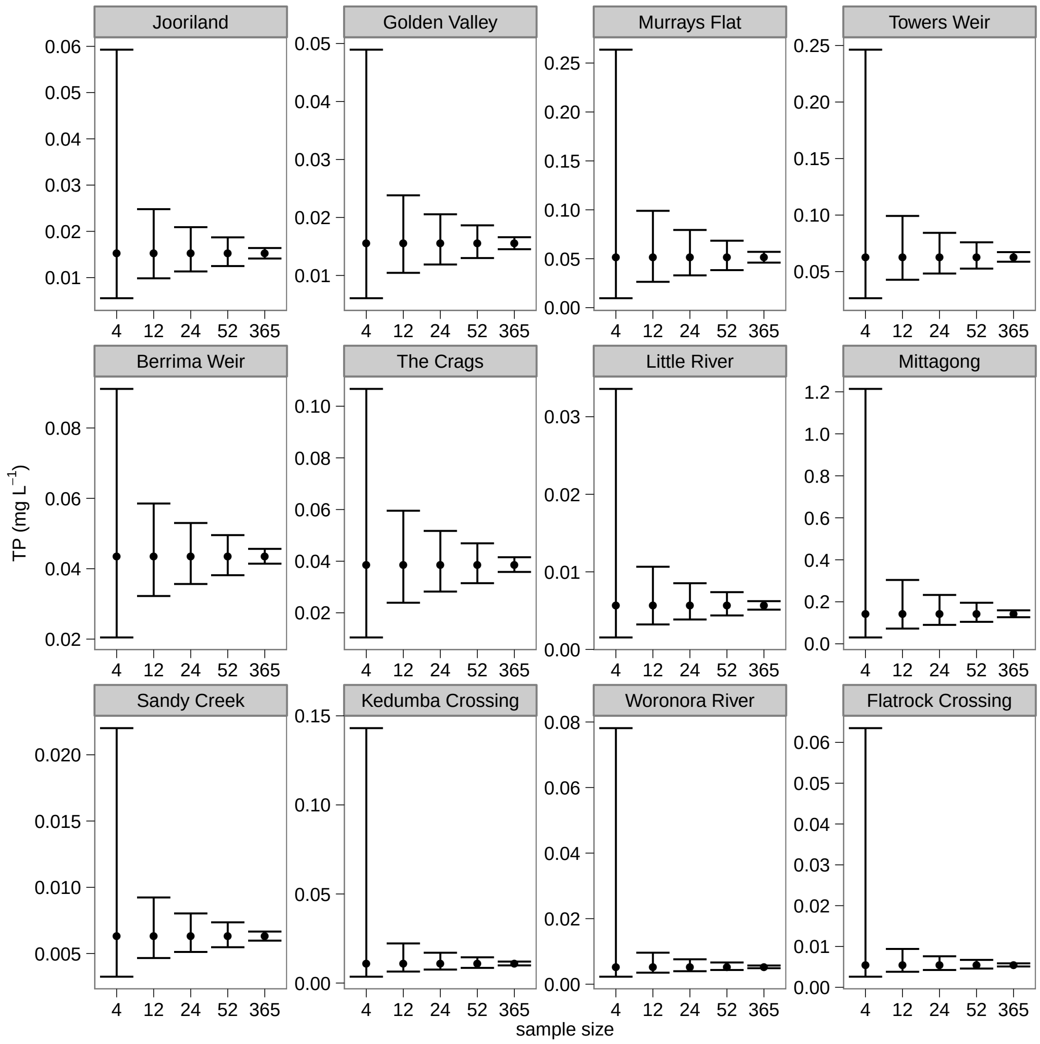

Figure 4.6 – Back-transformed base-flow mean and CI of each sub-catchment.

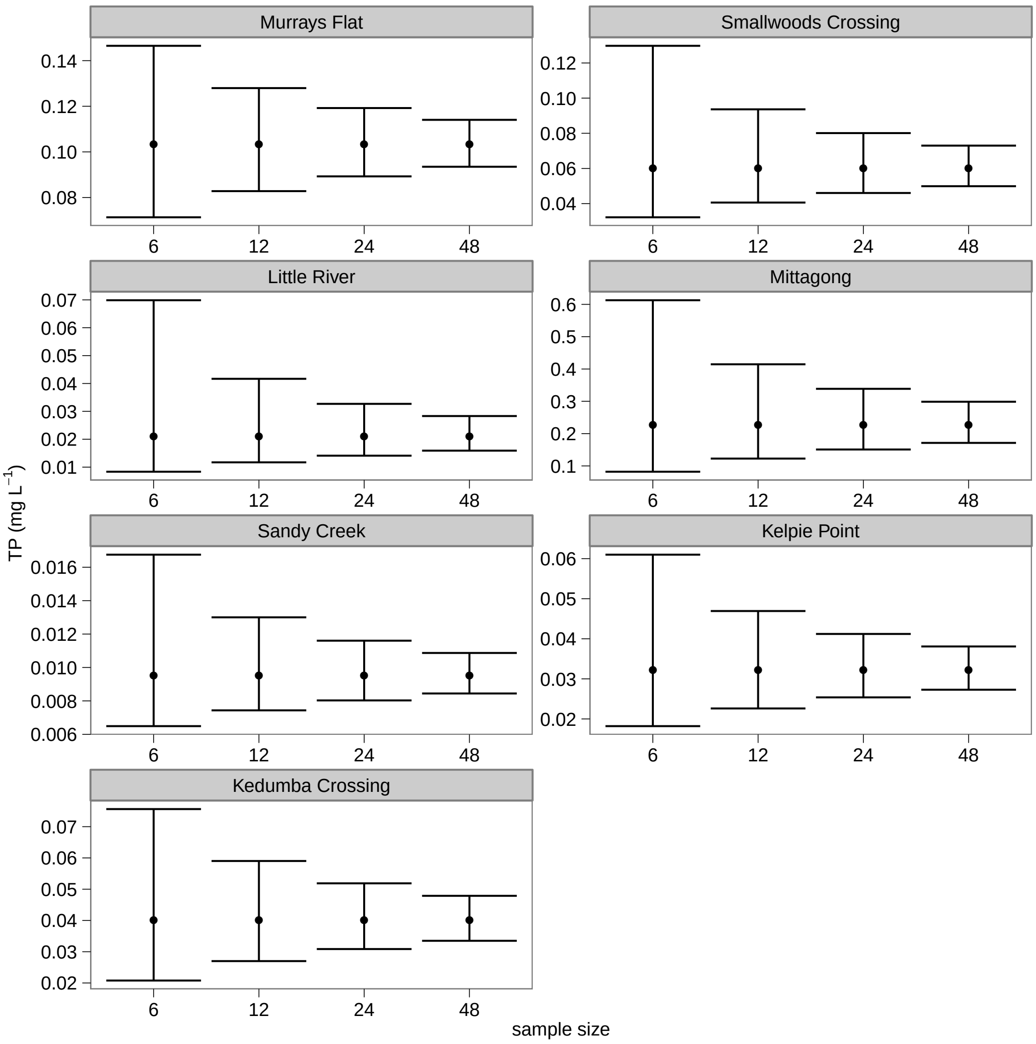

Figure 4.7 – Back-transformed event-flow mean and CI of each sub-catchment.

of the mean estimate for base-flow shows that there is little improvement for all catchments between $n = 24$ and $n = 365$ sample sizes. The outlier for both event and base-flow comparisons is the Mittagong sub-catchment which has an STP located directly upstream of the sampling site. In addition to this sub-catchment, Murrays Flat sub-catchment has a STP upstream of the gauge station. After sub-catchment Mittagong, Murrays Flat had the highest mean and CI for both base and event-flow conditions.

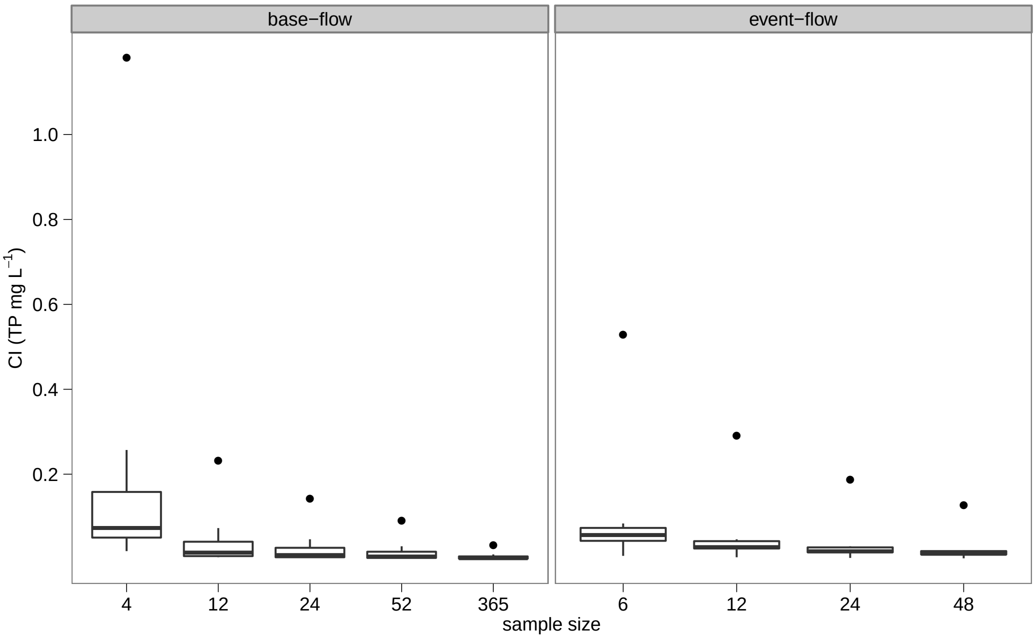

Figure 4.8 – Box-plot of the back-transformed standardised CI across all sub-catchments.

4.4.4 Can the relationship between sample size and the precision of mean estimates be linked to CCs?

If it is possible to relate the precision of the mean estimates to CCs it may help with future sample designs and provide help determining factors which control TP mean uncertainty. Table 4.4 provides Pearson correlation values between the standardised CI and common catchment characteristics. The percentage of urban cover is the most highly correlated CC with base-flow CIs with a correlation of 0.49. Grassland percentage cover and catchment size are highly correlated with event-flow CIs with correlations of -0.52 and -0.57 respectively. The correlation between annual flow and event and base-flow CIs are similar with correlations of -0.32 and -0.27 respectively. Annual flow, rainfall and urban cover are have similar correlations for both base and event-flow CIs. For both base and event-flow, the percentage of urban cover is the strongest positively correlated CC. In addition to the correlation between CCs and CIs the correlation to the sample size used for each model was determined as -0.32 and -0.54 for base and event-flow respectively. The sample size being the number of observations used to fit the linear mixed models.

Table 4.4 – Correlation between standardised CI and catchment characteristics.

covariate	base-flow	event-flow
annual rainfall (mm)	0.11	0.11
elevation range (m)	0.20	-0.12
urban cover (%)	0.49	0.47
forest cover (%)	0.20	0.34
grassland cover (%)	-0.32	-0.52
catchment size (km^2)	-0.13	-0.57
annual flow (ML)	-0.27	-0.32
n	-0.32	-0.54

In addition to the correlation between individual CCs, a linear model using backwards elimination was fitted to estimate the standardised CI for base and event flow. Figure 4.9 shows the predicted and observed standardised CI based on the linear models with the number indicating each sub-catchment. The Lin's correlation concordance (Lawrence and Lin, 1989) for the base and event-flow models were 0.83 and 0.43 respectively. It is not the intention of these linear models to be used to estimate the standardised CI of catchments, but to indicate that differences between the CIs can be related to common CCs. The final model for the base-flow standardised CI included elevation range, percentage of urban cover, catchment size and annual discharge. All CC except stream discharge were positively related to the standardised CI. This indicates that the uncertainty of the mean decreases in relation to the annual amount of stream discharge. Percentage of urban cover was the only CC not to be removed from the backward elimination process. Table 4.4 shows that urban cover is not the most correlated CC, however the backwards elimination process highlights the importance of the examination of the combined effect of CCs. Urban cover was positively related to the standardised CI, indicating the larger the percentage of urban area the higher the variation of the water quality during events.

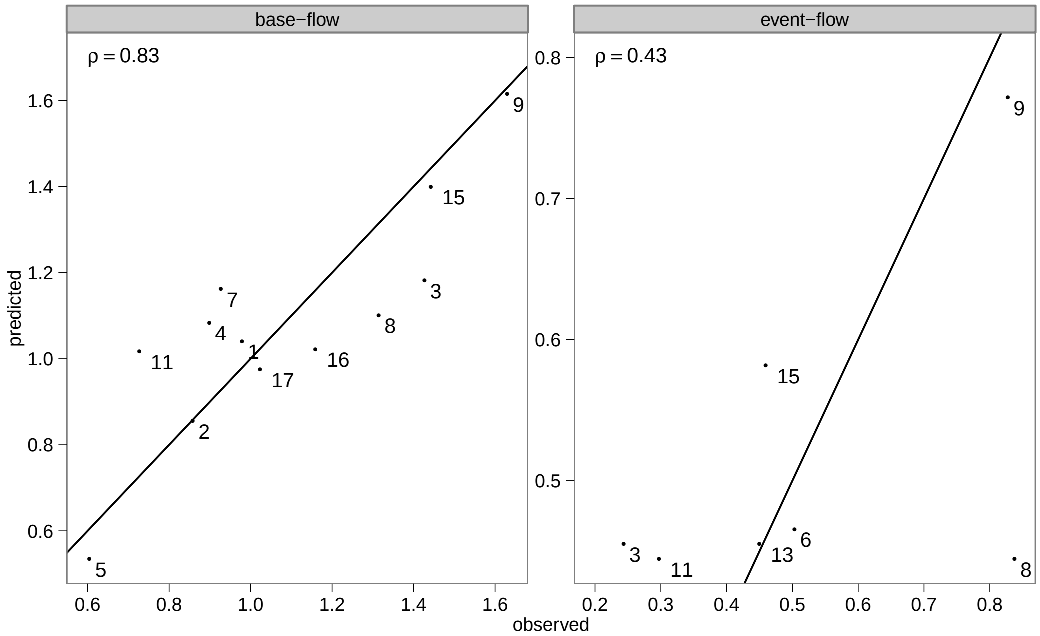

Figure 4.9 – Observed and predicted standardised CI of each sub-catchment.

4.5 Discussion

The results indicate that approach shown here can be used to model the temporal variance of TP data. This is very beneficial as it provides a method to estimate the variance of legacy water quality data where the inclusion probability is unknown. By applying the simulation based method of Domburg et al. (1994) it was possible to estimate the precision of the estimate of the mean with different sample sizes. With this approach it is possible to describe the usefulness of different sample sizes without the need of near-continuous datasets or the use of load estimation methods. When evaluating water quality sampling schemes it is important to consider the effect of sample size on the precision of the mean estimates, especially when this is often the main goal of water quality monitoring schemes.

The precision is different between base and event-flow conditions, with a worse precision for estimating event-flow mean concentration. The worse precision during event-flow is to be expected as streams in south east Australia are known to have larger exports during these periods and are also affected by other factors including hysteresis (Kuhnert et al., 2012). This finding also supports previous studies which have stressed the importance of event-flow water quality sampling (Cassidy and Jordan, 2011; Drewry et al., 2009). The base-flow mean precision indicated that 365 samples a year had the best precision, which is similar to the findings of Cassidy and Jordan (2011).

There was a good relationship between the event mean precision and catchment characteristics which indicates different catchments have different sample size requirements. The percentage of urban cover is positively correlated with the standardised CI for both base and event-flow. This indicates that catchments with larger urban coverage require larger sample sizes in order to account for the increased variation. The correlation of catchment size with the standardised CI suggests that as the size of the catchment increases, the estimate of the mean becomes more precise, which may be a reflection of how smaller catchments tend to experience shorter events. It is difficult to relate these findings to previous studies as previous studies have focused on mean concentration, and loads rather than the preciseness of the estimates of the mean concentrations. In terms of allocating sampling effort between a group of monitoring sites the approach here can be used to identify which sites require more or fewer samples. In many applications each of these sub-catchments would have similar sample sizes. A common alternative to this would be to use a catchment specific sampling design based on expert knowledge. The approach here can base it on the the existing data and repeatable statistics. In this study we assumed simple random sampling throughout the sample periods. Simple random sampling would most likely not be appropriate in practice especially over a year. A more likely probabilistic sampling approach would use a form of stratified sampling to reduce the variance as described by Thomas and Lewis (1993). For the purposes of this study the assumption of simple random sampling provides valuable information which can be used to assess current monitoring schemes. Given the financial constraints of most water quality monitoring schemes it is unlikely that it would be affordable for most monitoring schemes to use a sampling scheme where $n > 52$ a year for base-flow sampling. The method outlined in this paper can be used to estimate a suitable sample size of each sub-catchment to provide a mean within a pre-determined precision.

The results of the standardised CI suggest that there is little benefit from sample sizes greater than 12 for both base and event flow conditions, however the Mittagong and Murrays Flat sub-catchments require a larger sample size to provide estimates of the mean with similar precision as these two sub-catchments have waste water treatment plants.

Water quality data is generally highly skewed and legacy data often contains outliers, however it is often difficult to determine which observations should be classified as outliers. We have used a winsorising approach to remove the effect of outliers. This step was necessary for several sub-catchments to ensure the prediction variance reflected the actual error. The model for several sub-catchments failed the requirements of the cross-validation and it was not possible to estimate the precision of the mean for these sub-catchments. Research is being undertaken to develop robust methods which can be used to fit models to data with outliers. (Papritz et al., 2011; Marchant et al., 2010).

4.6 Conclusions

As it is not financially possible to apply continuous sampling for all water quality monitoring schemes a method is required to characterise the effect of the sample size on the precision associated with mean estimates. With many sampling schemes using a form of non-probabilistic sampling, a model is required to obtain an unbiased estimate of the variance of the samples. By using the methods of Domburg et al. (1994) we have shown that;

- variograms can be used to model the temporal variance of water quality and provide estimates of the SEM of the mean.
- the estimates of the event-flow means are less precise than the estimates of the base-flow means.
- the precision of the mean estimates for base and event-flow can be related to catchment characteristics.

As found in previous studies, variograms are sensitive to outliers. It was not possible to use this method for all sub-catchments, however the results indicated that it is possible to use this method to estimate the effect of sample size on the SEM using historical data.

4.7 References

Ahearn, D., Sheibley, R., Dahlgren, R., Anderson, M., Johnson, J., Tate, K., 2005. Land use and land cover influence on water quality in the last free-flowing river draining the Western Sierra Nevada, California. Journal of Hydrology 313 (3), 234–247.

Akaike, H. 1974 A new look at the statistical model identification. Automatic Control, IEEE Transactions on, 19(6), 716-723.

Banner, E., Stahl, A., Dodds, W., 2009. Stream discharge and riparian land use influence in-stream concentrations and loads of phosphorus from central plains watersheds. Environmental Management 44 (3), 552–565.

Bartley, R., Speirs, W., Ellis, T., Waters, D., 2012. A review of sediment and nutrient concentration data from Australia for use in catchment water quality models. Marine Pollution Bulletin 65 (4), 101–116.

Box, G., Cox, D., 1964. An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological), 211–252.

Cochran, W., 2007. Sampling techniques. John Wiley & Sons.

Cassidy, R., Jordan, P., 2011. Limitations of instantaneous water quality sampling in surface-water catchments: Comparison with near-continuous phosphorus time-series data. Journal of Hydrology 405 (1), 182–193.

de Gruijter, J., Brus, D., Bierkens, M., Knotters, M., et al., 2006. Sampling for natural resource monitoring. Springer Verlag.

Díaz Muñiz, C., García Nieto, P., Alonso Fernández, J., Martínez Torres, J., Taboada, J., 2012. Detection of outliers in water quality monitoring samples using functional data analysis in San Esteban Estuary (Northern Spain). Science of The Total Environment 439, 54–61.

Domburg, P., De Gruijter, J., Brus, D., 1994. A structured approach to designing soil survey schemes with prediction of sampling error from variograms. Geoderma 62 (1), 151–164.

Drewry, J., Newham, L., Croke, B., 2009. Suspended sediment, nitrogen and phosphorus concentrations and exports during storm-events to the Tuross Estuary, Australia. Journal of environmental management 90 (2), 879–887.

Johnes, P., 2007. Uncertainties in annual riverine phosphorus load estimation: impact of load estimation methodology, sampling frequency, baseflow index and catchment population density. Journal of Hydrology 332 (1-2), 241–258.

Kronvang, B., Bruhn, A., 1996. Choice of sampling strategy and estimation method for calculating nitrogen and phosphorus transport in small lowland streams. Hydrological Processes 10 (11), 1483–1501.

Kuhnert, P., Henderson, B., Lewis, S., Bainbridge, Z., Wilkinson, S., Brodie, J., 2012. Quantifying total suspended sediment export from the Burdekin River catchment using the loads regression estimator tool. Water Resources Research 48 (4), W04533.

Lark, R., Cullis, B., 2004. Model-based analysis using REML for inference from systematically sampled data on soil. European Journal of Soil Science 55 (4), 799–813.

Lawrence, I., Lin, K., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics, 255–268.

Marchant, B., Saby, N., Lark, R., Bellamy, P., Jolivet, C., Arrouays, D., 2010. Robust analysis of soil properties at the national scale: cadmium content of french soils. European Journal of Soil Science 61 (1), 144–152.

Papritz, A., Künsch, H., Schwierz, C., Stahel, W., 8 2011. Robust estimation of the external drift and the variogram of spatial data. In: 58th World Statistics Congress of the International Statistical Institute. Dublin.

R Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. http://www.R-project.org/

Ribeiro Jr, P., Diggle, P., 2001. geoR: A package for geostatistical analysis. R news 1 (2), 14–18.

Särndal, C. E., Swensson, B., Wretman, J. (2003). Model assisted survey sampling. Springer.

Sliva, L., Williams, D., 2001. Buffer zone versus whole catchment approaches to studying land use impact on river water quality. Water Research 35 (14), 3462–3472.

Strobl, R., Robillard, P., 2008. Network design for water quality monitoring of surface freshwaters: A review. Journal of environmental management 87 (4), 639–648.

Thomas, R., Lewis, J., 1993. A comparison of selection at list time and time-stratified sampling for estimating suspended sediment loads. Water Resources Research 29 (4), 1247–1256.

Thomas, R., Lewis, J., 1995. An evaluation of flow-stratified sampling for estimating suspended sediment loads. Journal of Hydrology 170 (1-4), 27–45.

Visschers, R., Finke, P., De Gruijter, J., 2007. A soil sampling program for the netherlands. Geoderma 139 (1), 60–72.

Chapter 5

Characterisation of events using stratified random sampling

5.1 Summary

Short rainfall events contribute to large portions of annual sediment and nutrient exports. Most water quality sampling schemes rely on regularly spaced temporal sampling and increasingly monitoring schemes are including a form of event-based sampling. A typical approach is to sample each event using equal intervals in time using an automatic sampler. The use of this form of sampling is systematic in nature and requires model-based statistics to be analysed correctly. Probabilistic based sampling methods allow for easier and more defendable statistical inference as the assumptions are not based on a model, rather they are based on the sample design. Several probabilistic methods have been developed, however these methods commonly require additional hardware to implement. In this paper we present a method using a stratified random sampling procedure for automatic samplers which does not require any additional hardware. Our approach is to divide the mean event hydrograph into strata based on key features such as the rising and falling limbs. Random sampling is applied within each strata. A problem of this approach is that the length of the event and strata must be defined before each event. We therefore outline how the samples can be post-

stratified after each event based on the key hydrological components of each event. The sampling scheme is outlined using continuously sampled electrical conductivity and turbidity data of three events from a creek in south eastern Australia. Limited to 24 samples per event, estimated event mean CIs were within the observed event means for all three events. Both the stratified and re-stratified estimates provided unbiased estimates of key event hydrological components.

5.2 Introduction

It is widely accepted that event-based water quality sampling is required to accurately characterise the temporal variation in water quality. This is especially evident in catchments characterised by long durations of base-flow, where short rainfall events are responsible for exporting the majority of nutrients (Kronvang and Bruhn, 1996; Johnes, 2007; Drewry et al., 2009). Recent studies using continuously measured data have highlighted the importance of high frequency sampling and the information this can provide for understanding hydrological processes (Kirchner et al., 2004; Burt et al., 2011; Cassidy and Jordan, 2011; Wade et al., 2012). However, due to financial constraints it is unlikely that these new technologies will be widely implemented in the near future (Wade et al., 2012), with the majority of catchment managers relying on automatic samplers to sample during rainfall events (Bartley et al., 2012). Furthermore, for many water quality properties there are no options for field based continuous sensors.

Without access to continuously monitored data, catchment managers often rely on a combination of grab samples and automatic samplers to collect samples during rainfall events. However during rainfall events access to water quality monitoring sites is often limited. For example, safety constraints require the Sydney Catchment Authority (in Australia) to use helicopters to access sites after 10-15 mm rainfall events (Sydney Catchment Authority, 2011). Under these restrictions it is necessary for catchment managers to rely on field based equipment to take samples during an event. This equipment generally consists of two devices; a stream discharge measuring device and an automatic sampling device (Wang et al., 2011). The use of these devices introduces additional limitations on the sample scheme. One of the largest constraints of automatic samplers is the number of bottles, which is typically 24, and therefore the maximum samples possible per event.

Common objectives of event-based sampling studies are for estimation of load (Drewry et al., 2009) and event mean concentration (Bartley et al., 2012). It is important to assess the objectives of the sampling schemes to determine the most appropriate design (de Gruijter et al., 2006). The research of Marsh and Waters (2009) found that event load estimation requires samples to be collected during both the rising and falling limb to ensure accurate estimates are obtained. If the objective of event-based sampling is to estimate the event mean concentration then continuous sampling would be preferred. As this is rarely possible it is important that the selected sampling scheme is unbiased and therefore provides unbiased estimates of the mean and its uncertainty (Thomas and Lewis, 1993). Several water quality models now require a form of event mean concentration for parameterisation (Bartley et al., 2012).

In general, sampling schemes can be divided into two categories; probabilistic (design-based) and non-probabilistic (model-based) (de Gruijter et al., 2006). Non-probabilistic sampling designs do not require known inclusion probabilities (i.e. the probability of a sample being taken at time t), however these methods require the use of a model to provide unbiased estimates of the uncertainty of the mean (de Gruijter et al., 2006; Lark and Cullis, 2004). Probabilistic sampling schemes rely on known inclusion probabilities and therefore do not require a model to provide unbiased estimates of the uncertainty of the mean (de Gruijter et al., 2006). The most commonly used event-based sampling scheme is systematic sampling based on equally spaced samples through time or flow volume (Harmel et al., 2006). This form of sampling is known to provide unbiased mean estimates when used with a random starting time (de Gruijter et al., 2006; Lohr, 2009) or by using a model which accounts for the correlation between samples (Lark and Cullis, 2004; de Gruijter et al., 2006). Under this form of sampling one must assume the sample interval is short enough to provide adequate coverage of the sampled property throughout the event, which may not be true especially during the rising limb of an event.

Probabilistic sampling schemes provide unbiased estimates of the event mean and its uncertainty. The use of probabilistic sample designs for water quality sampling has been investigated by Thomas (1985, 1988); Thomas and Lewis (1993, 1995); Arabkhedri et al. (2010). Focusing on suspended sediments they evaluated several different probabilistic methods. The methods evaluated include selection at list time, which is a form of probabilities proportional to size sampling (Thomas, 1985, 1988; Thomas and Lewis, 1995) , simple random sampling (Thomas, 1988), flow proportional (Thomas, 1988), flow-stratified (Thomas and Lewis, 1995) and time-stratified (Thomas and Lewis, 1993, 1995) sampling schemes. More recently Arabkhedri et al. (2010) has demonstrated the use of adaptive cluster sampling to estimate suspended sediments. The results of these studies showed that it is possible to provide accurate unbiased event based estimates using probabilistic sample designs, however these methods have not been widely implemented. We believe the main reason for the lack of use is due to the perceived complexity of the design and possibly due to the focus on suspended sediment exports and the additional hardware and software requirements.

We are therefore proposing a method which is a variation of that of Thomas and Lewis (1993). The time-stratified method outlined by Thomas and Lewis (1993) used a scheduled based approach, with the sampling strata based on the size and extent of real time changes of stream discharge during an event, in combination with short strata with small sample sizes to maximise sample coverage. The implementation of this method requires additional hardware and programming knowledge. Therefore we have attempted to simplify their method by using a simplified design, based on historical stream discharge data. In contrast to predicting strata lengths in real-time, we propose the use of one stratification based on a *mean* hydrograph combined with post-stratification which provides the ability to re-stratify the samples into key components of each observed event hydrograph (e.g. rising and falling limbs).

In this paper we evaluate a simplified time-stratified sampling scheme to estimate different water quality properties affected by different transport pathways. Using turbidity and EC will we evaluate the use of event-based time-stratified sampling to provide estimates that relate to two distinctly different flow transport paths. The two flow paths of interest are overland flow and baseflow, using turbidity and electrical conductivity to reflect these paths respectively. Turbidity has been used to improve estimates of TP (Biggs, 1995) and TSS (Biggs, 1995; Lane et al., 2006) whereas EC has been used for base-flow separation (Pellerin et al., 2008) and to estimate dissolved nitrogen and particulate nitrogen (Kim and Furumai, 2012). By sampling continuously turbidity and EC data we examine the use of stratified sampling for event-based water quality monitoring. Therefore the aims of this paper are to:

1. illustrate the use of stratified sampling for estimating event mean concentrations for different water quality properties, and

2. outline the method used to restratify samples after each event based on the observed stream discharge to provide information on key hydrological components of events.

5.3 Methods

5.3.1 Catchment description

Figure 5.1 presents the Muttama Creek catchment (1061km^2) which is located in the south west slopes of New South Wales in south eastern Australia. Agriculture is the main land use in the catchment with winter cropping and pastures the dominant agricultural practices. A main feature of the catchment are several north-south fault lines on the eastern side of the main stream (Conyers et al., 2008). These fault lines are associated with ultrabasic rocks with frequent outcrops of serpentine (Conyers et al., 2008). In contrast the western side is dominated by slates and rhyolite (Conyers et al., 2008). The geological variety in the catchment has lead to a variety of soil types, of particular importance are Sodosols in the north west of the catchment (Warren et al., 1995) which are highly dispersive soils (Isbell, 2002). These areas are also associated with saline areas (Conyers et al., 2008). Several catchment characteristics are provided in table 5.1. Annual rainfall within the catchment is 507 mm with an annual discharge of 578.98 GL. The main township of Cootamundra is in the north of the catchment with a population of 7729 (Australian Bureau of Statistics, 2010).

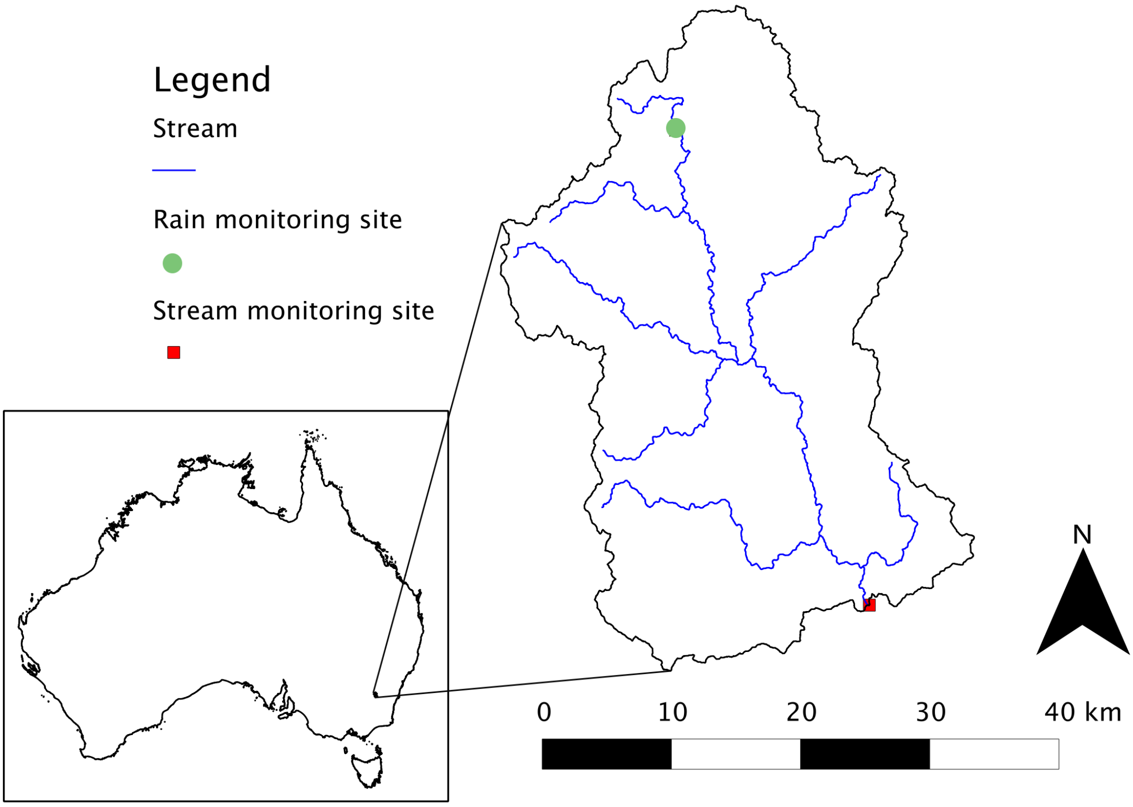

Figure 5.1 – Muttama Creek catchment and location.

5.3.2 Event data description

A YSI sonde (Hydrolab Corporation, Austin, Texas) was used to collect turbidity data for each event. Rainfall, stream discharge and electrical conductivity data was obtained from the government monitoring stations (http://waterinfo.nsw.gov.au). Figure 5.2 shows the observed rainfall, stream discharge, turbidity and EC of the three events. The first event (fig. 5.2 a.) occurred in August 2011, and is representative of events for this catchment. The observed double peak in the hydrograph is a common feature of event hydrographs at this site which is possibly due to the figure 8 like shape of the catchment (Gordon et al., 2004, p. 67). The second event (fig. 5.2 b.) occurred in March 2012 and was the largest event on record. Three peaks occurred during this event with two distinct rainfall events. The third event (fig. 5.2 c.) occurred in July 2012 and is similar but smaller than the first event, however rain during the falling limb caused an additional smaller peak.

Table 5.1 – Summary statistics for the Muttama catchment

Catchment characteristics	Statistic
Topographical features	
Catchment area (km^2)	1061
Maximum overland flow distance to outlet (km)	74.31
Minimum elevation (m)	227
Mean elevation (m)	402.39
Maximum elevation (m)	719
Mean annual rainfall (2004- 2012)	
Berthong Station (mm)[a]	507
Stream discharge statistics (2000 - 2012)	
Mean annual discharge (GL $year^{-1}$)	578.96
Minimum daily discharge (ML d^{-1})	0
Mean daily discharge (ML d^{-1})	67
Maximum daily discharge (ML d^{-1})	15746.43

[a] See Figure 5.1 for rainfall station location.

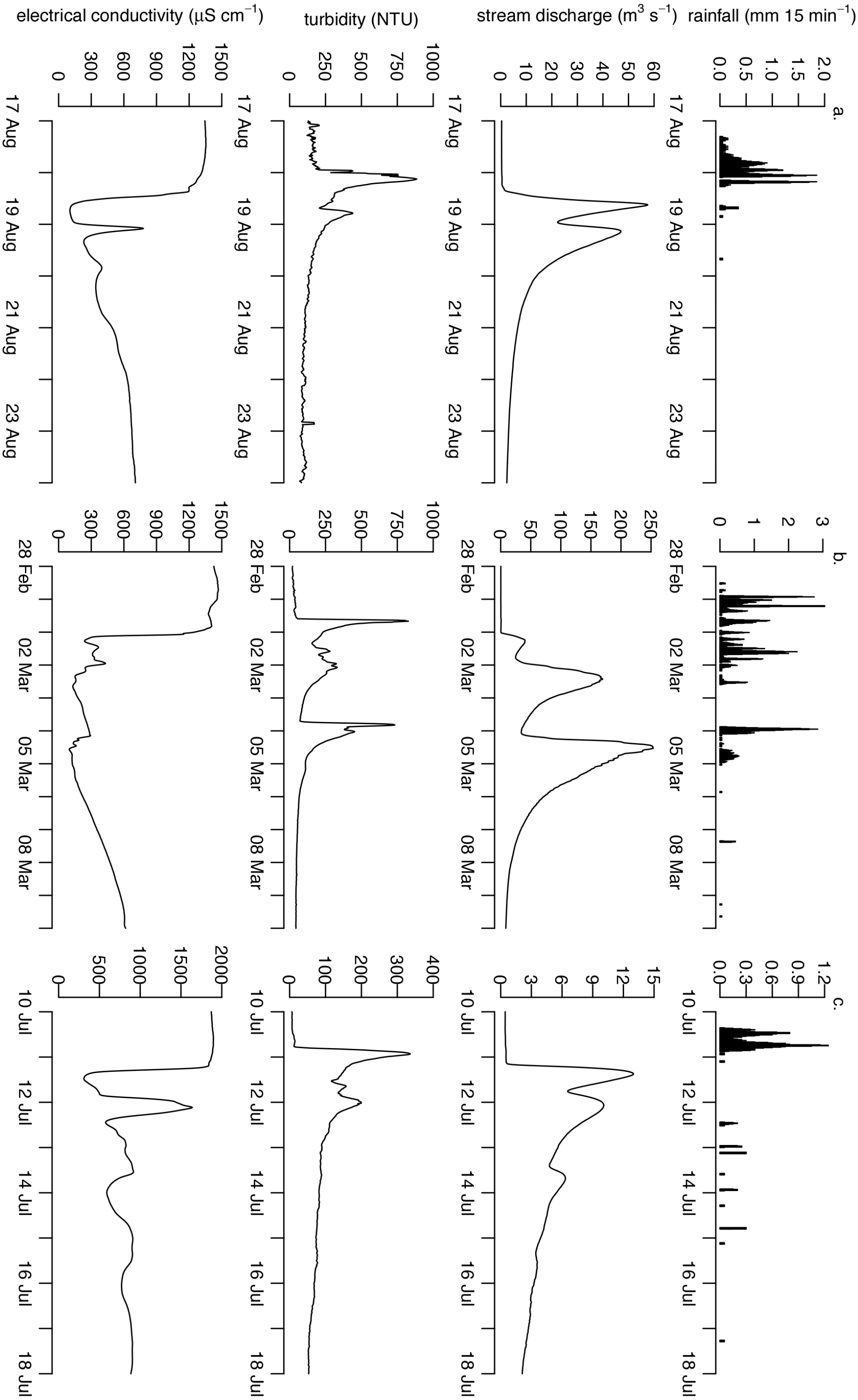

Figure 5.2 – Observed data of the three events (rainfall, stream discharge, turbidity and specific conductivity for; a. events 1,b. event 2 and c. event 3.

5.3.3 Defining strata and sample sizes

To perform event-based sampling it is necessary to define certain aspects of an event. To use a probabilistic based sampling design, the start and end of the event and the sample sizes, length and number of strata are also required. We propose a simple 4 step process to define a general time-stratified sampling scheme;

1. determine the event trigger height based on the upper 5% of the stream height distribution,
2. find all events using the trigger height as the start and end of each event,
3. determine the length of the event using the mean duration of these events, and
4. create a mean event hydrograph using the trigger height and the mean event length. Use this to define strata and sample sizes.

Event based sampling is generally commenced using a trigger consisting of a pre-determined stream height reached during the rising limb. It is therefore important that this height is not too low as to be triggered by small rises of stream height, but not too large which would commence the sampling to late (Harmel et al., 2002). Our approach is to use the stream height which corresponds to the 95th percentile stream height. The flow duration curve for Muttama Creek is based on the last 10 years of observations and is shown in figure 5.3 with the corresponding 95th percentile stream height (1.4 m). To define the duration of the mean event, all events are found by defining the start of the event as the time where the trigger height is exceeded and the end of the event as the time where the stream height recedes below the trigger height. Using these events, the mean event length is defined as the mean of all events which are longer then 12 hours. Figure 5.4 shows the event mean hydrograph based on the trigger height with the mean event strata boundaries. The initial strata are defined based on the mean event hydrograph and are such that they provide good coverage of rapid changes of stream discharge. Strata 1, 2 and 3 are designed to cover the rising limb, initial peak and second peak of the hydrograph. Strata 4 and 5 are longer and designed to capture the tail of the event. Assuming a maximum sample size of 24, the sample size of each strata is defined as; 6, 6, 4, 4 and 4. Larger sample sizes are given to the initial two strata, as these strata are associated with sharp changes in stream discharge.

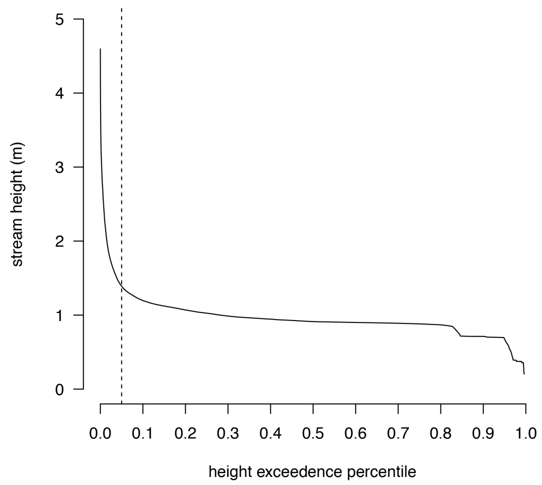

Figure 5.3 – Stream height exceedance percentile plot with event trigger height.

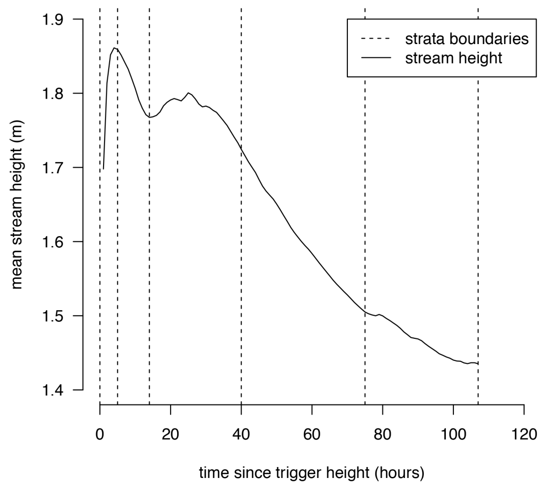

Figure 5.4 – Mean event hydrograph with strata boundaries.

5.3.4 Strata and event analysis

Stratified sampling is a common sampling design, a detailed description of its application is provided by de Gruijter et al. (2006, p. 82). Using this information the mean of an event is estimated using;

$$\hat{\bar{z}}_{st} = \frac{1}{T}\sum_{h=1}^{H} T_h \hat{\bar{z}}_h, \tag{5.1}$$

where H is the number of strata in the event, T is the total time of the event, T_h duration of time in strata h and $\hat{\bar{z}}_h$ is the mean of stratum h. In this study, $T = 107$ hours (the mean event duration), $T_h = \{5, 9, 26, 35, 32\}$ hours. The variance of each stratum mean is estimated by

$$\hat{V}(\hat{\bar{z}}_h) = \frac{1}{n_h(n_h - 1)}\sum_{i=1}^{n_h}(z_{hi} - \hat{\bar{z}}_h)^2, \tag{5.2}$$

where n_h is the sample size of stratum h, therefore $n_h \geq 2$ is needed to estimate the variance of the mean. Using eq. 5.2 the variance of the event mean can be estimated using

$$\hat{V}(\hat{\bar{z}}_{st}) = \frac{1}{T^2}\sum_{h=1}^{H} T_h^2 \hat{V}(\hat{\bar{z}}_h), \tag{5.3}$$

which can be used to estimate the standard error of the event mean ($\sqrt{\hat{V}(\hat{\bar{z}}_{st})}$). Confidence intervals are easily estimated for each strata by;

$$\hat{\bar{z}} \pm t_{1-\alpha/2} \cdot \sqrt{\hat{V}(\hat{\bar{z}}_h)}, \tag{5.4}$$

where $t_{1-\alpha/2}$ is the t-critical value of the student distribution with degrees of freedom and α is the probability value which is set to 0.025 if we wish to estimate the 95% confidence interval around the mean. Similarly, eqn. 5.4, can be used to estimate the confidence interval around the event mean concentration.

5.3.5 Post event stratification

Post event stratification is the term given to the reclassification of samples into new strata. To simplify and reduce confusion new strata are referred to as domains. In this paper, this process is used to combine samples from multiple strata into domains. The major benefit of this process is the ability to perform this task based on the observed stream hydrograph of each individual event. The only limitation of this process is the requirement that at least 2 samples from each stratum fall within the domain to estimate the variance of the domain mean. Using event 2 as an example, the samples from multiple strata are reclassified into domains based on the three peaks of the event. Figure 5.7 shows the hydrograph of the event with the three domains and locations of the samples. In addition to mean estimates, the rising and falling limbs are also examined. In this chapter the rising limb is from the start of the event to the maximum value of the first peak and the falling limb is the remaining part of the event. A detailed description of reclassification of stratified sampling into domains can be found in (Särndal et al., 2003, p. 390). Based on this description the mean of a stratum within each domain can be estimated using

$$\hat{\hat{z}}_{dh} = \frac{\sum_{i=1}^{n_h} I_{dhi} z_{hi}}{n_{dh}}, \tag{5.5}$$

where I_{dhi} is an indicator vector where 1 indicates the sample is within the domain and 0 the sample is not within the domain, z_{hi} is the i^{th} observation of domain h and n_{dh} is the amount of samples in stratum h and domain d. Therefore the mean of the domain is estimated using

$$\hat{\hat{z}}_{d} = \frac{\sum_{h=1}^{H} \frac{T_h n_{dh}}{n_h} \hat{\hat{z}}_{dh}}{\sum_{h=1}^{H} \frac{T_h n_{dh}}{n_h}}. \tag{5.6}$$

The associated standard error of the domain mean can be estimated by

$$se(\hat{\hat{z}}_{d}) = \sqrt{\frac{1}{T_d^2} \sum_{h=1}^{H} T_h^2 \frac{(n_{dh}-1)s_{dh}^2 + n_{dh}(1-\frac{n_{dh}}{n_h})\left(\hat{\hat{z}}_{dh} - \hat{\hat{z}}_{d}\right)^2}{n_d(n_d-1)}} \tag{5.7}$$

where T_d is the duration of time of the domain, n_d is the amount of samples with domain d and the variance is estimated using;

$$s_{dh}^2 = \frac{\sum_{i=1}^{n_h} I_{dhi} \left(z_{di} - \hat{\hat{z}}_d\right)^2}{n_{dh} - 1}. \tag{5.8}$$

It is important to note that each domain requires $n_{dh} \geq 2$ to allow for an estimate of s_{dh}^2 (Särndal et al., 2003).

5.4 Results

5.4.1 Time-stratified sampling: sampling locations

Figure 5.5 shows the location of each sample within each strata of the three observed events. The shape of the hydrograph of the first event (fig. 5.5 a.) closely resembles the mean event hydrograph (fig. 5.4) with the strata boundaries closely matching the desired components of the hydrograph. The second event (fig. 5.5 b.) is not similar to the mean event hydrograph due to additional rainfall during the event and only the first and second strata match the changing conditions of the first peak of the event. In addition to this the pre-determined time of the event did not meet the extended duration of the event. The third event (fig. 5.5 c.) is similar to both the first event and the mean event hydrograph with the strata matching the desired hydrograph components. However a small peak occurred during the falling stage of the hydrograph. The use of the mean event strata closely matched 2 of the 3 observed events. Such variation would be expected and shows the importance of post-stratification if we wish to extract mean values from different components of the event hydrograph.

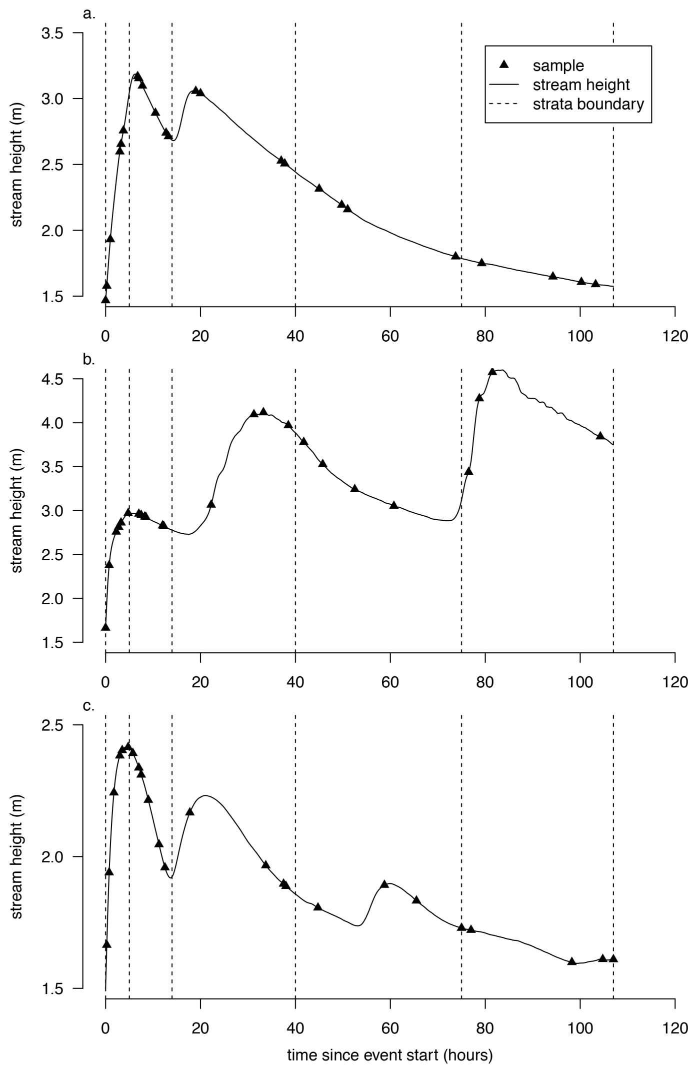

Figure 5.5 – Location of strata and samples for each event; a. event 1,b. event 2,b. event 3.

5.4.2 Stratified estimates

The stratified mean estimates, with confidence intervals, of turbidity and EC of each event are shown in figure 5.6. In addition to the strata estimates, the observed continuous data is also provided with the mean estimates of each strata also provided. Observed means for all strata estimates for both turbidity and EC of all three events are within the associated confidence intervals, except for the third strata for turbidity for event 3 (fig. 5.6 c.). Confidence interval width related to the number of samples and variation within the strata, therefore strata with the greatest variation require additional sampling effort to increase the precision of the mean estimates. The confidence intervals of event 1 (fig. 5.6 a.) indicate that the underlying trend of turbidity varies most during the first and third strata. This variation is shown by the continuously sampled turbidity data. Similarly the first strata of EC in this event was also the largest CI of this event. Event 2 (fig. 5.6 b.) was the largest event on record for the site, in terms of maximum stream height, and it had three distinct peaks within the event hydrograph. The first and last strata have large CI widths indicating a large amount of variation during these times which is confirmed by the continuously monitored data. A similar trend can also be seen in event 3 (fig. 5.6 c.) where the CI width of the first strata is quite wide compared with other strata CI widths of this event. The third strata CI estimate does not cover the observed mean turbidity, with the observed mean falling outside the CI by 9.96 NTU. By examining figure 5.5c., it can be seen that 3 of the 4 samples were collected towards the end of this strata, after the peak of observed turbidity (fig. 5.6 c.). Observed mean EC of each strata do fall within the CI of the estimates for this event, with relatively large CI widths for the first and third strata. These results indicate that this simple application of stratified sampling provides accurate estimates of key hydrological components of event hydrographs and can provide information about transport pathways during events.

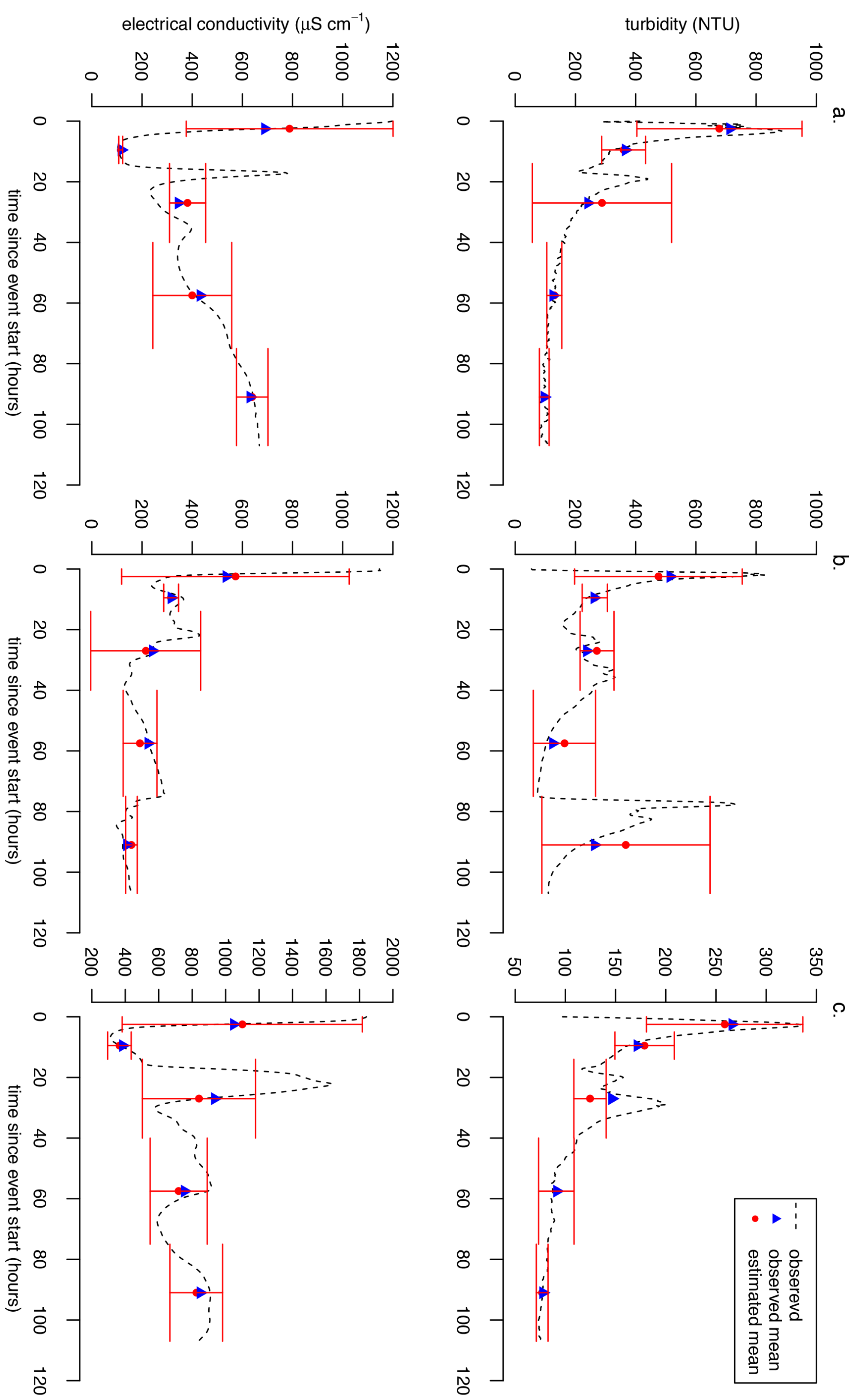

Figure 5.6 – Stratified estimates of turbidity and EC for the three events; a. event 1,b. event 2,b. event 3.

5.4.3 Event estimates

By combining the strata mean estimates it is possible to obtain event mean estimates. Table 5.2 provides event mean estimates and observed means for each event. Using the time-stratified event estimates it is possible to observe the trends between the three events. This trend indicates that the event means of turbidity and EC are directly related to the size of each event, with the largest event (event 2) having the largest turbidity event mean and the smallest event (event 3) having the smallest estimated mean turbidity. This trend is inversely true for EC of the three events. These event estimates are confirmed by the observed turbidity and EC event means. Observed turbidity (224.73) was overestimated by 62.42 for event 2 but within the CI of the estimate. In this situation the large estimated uncertainty provides additional information, indicating the large variation during this event. In contrast to this, the estimated mean turbidity of event 3 was very accurate and had CI of ±18.22 NTU. This accuracy was achieved with a sample size of 24 for the 107 hour period comprising 428 possible sample times, reinforcing the ability of design based methods to provide accurate estimates of event means.

Table 5.2 – Stratified sampling event mean estimates.

Event	turbidity (NTU)			EC (μS cm^{-1})		
	estimated	CI(±)	observed	estimated	CI(±)	observed
1	202.12	81.62	194.38	467.09	96.50	457.11
2	287.15	128.10	224.73	219.1	102.91	227.51
3	111.59	18.22	116.02	748.7	179.31	810.27

5.4.4 Re-stratified domain estimates

By re-stratifying the strata into domains it is possible to provide unbiased estimates of key components of the event hydrograph. Table 5.3 provides the estimates of the key components of each event hydrograph. For each event, estimates were made of the rising limb, falling limb and peak of the event. The boundaries of each peak in an event is based on the start, minima between each peak and the end of the event. For example, the boundaries of each peak for the second event are shown in figure 5.7. Three peaks were observed for the second event, but there was no easily definable rising and falling limb. Therefore the rising and falling limbs were not estimated for this event.

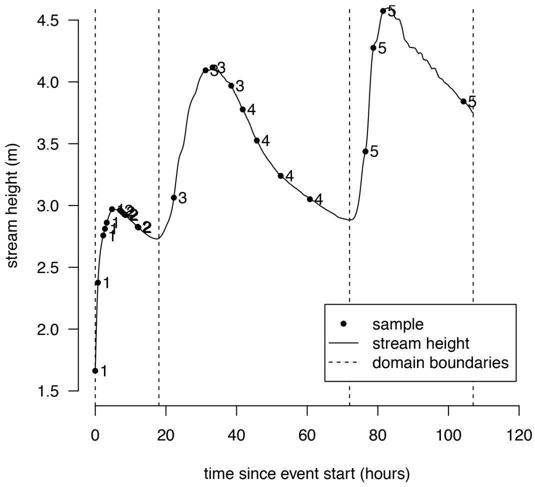

Figure 5.7 – Event 2 with domain boundaries and sample locations and strata number.

The observed EC mean for each domain is within the CI of the estimates for the rising limb, falling limb, and all peak estimates for the three events. The observed means of turbidity were within the estimated CI of the rising limb, falling limb and peak 1 domains for all events. The estimation of the peak 2 of event 1 was over-estimated and the observed mean fell outside the CI. This is most likely due to only 2 samples being available to estimate the domain mean. Peak 2, event 3 was also poorly estimated with the observed mean falling outside the estimated CI.

5.5 Discussion

The majority of event based sampling schemes are non-probabilistic Wang et al. (2011). The results of these sampling schemes are inherently biased and may be affected by large intervals between samples (Thomas and Lewis, 1995). The importance of high frequency samples to accurately characterise variation in water quality is well known (Kirchner et al., 2004; Burt et al., 2011; Cassidy and Jordan, 2011).

Table 5.3 – Estimates of event domains.

Domain	samples	strata	turbidity (NTU)			EC (μ cm^2)		
			estimated	CI($\pm$)	observed	estimated	CI($\pm$)	observed
Event 1:								
rising limb	6	1	678.23	200.86	680.83	788.43	301.76	585.39
falling limb	18	4	180.04	46.37	163.16	445.39	71.41	448.29
peak 1	12	2	473.64	124.5	490.65	355.62	231.7	325.4
peak 2	2	1	412.5	182.45	277.48	382.44	368.64	337.64
Event 2:								
peak 1	12	2	339.75	82.27	318.71	407.59	121.48	381.07
peak 2	8	2	209.64	56.93	178.84	201.85	62.82	223.75
peak 3	4	1	367.76	222.77	246.37	158	18.55	153.86
Event 3:								
rising limb	6	1	258.67	78.06	267.82	1189.94	32.87	1216.164
falling limb	18	4	102.84	18.55	109.07	432.21	67.82	444.33
peak 1	12	2	207.4	34.79	207.67	627.38	310.68	632.15
peak 2	4	1	126.78	22.65	150.13	865.08	515.64	948.05
peak 3	4	1	87.31	8.12	86.55	704	758.58	705.75

However the majority of event-based sampling schemes rely on automatic samplers with a maximum sample size of 24 to sample events. The presented time-stratified sampling scheme offers the ability to provide accurate unbiased estimates under this restriction, with the ability to reclassify strata after each event to estimate mean values for key hydrograph components.

The simplified time-stratified design based on Thomas and Lewis (1993) used a mean event hydrograph to design the strata for each event, however did not require any additional hardware or software modifications to the standard automatic samplers. For the three observed events these strata were quite close to events 1 and 3. The defined strata did not match the shape of the second event. By using domain based estimates it was possible to provide accurate mean estimates of both turbidity and EC for the three peaks of this event, with the observed mean within the CI of each domain. The flexibility of the proposed sampling scheme shows that it is possible to derive accurate estimates using probabilistic sampling schemes.

Often water quality monitoring investigates multiple water quality properties affected by multiple transport pathways, therefore proposed sampling schemes must be flexible enough to reflect this. With access to stratified and domain estimates the controlling processes of various water quality properties can be identified. The results of the stratified sampling shows the mean turbidity of the first strata (rising limb) is greater then the second strata of each event. In addition to this, the domain estimates also help identify these controlling processes. For example during, the second event (fig 5.2 b.) the second peak of turbidity coincided with an additional rainfall event. The controlling processes of EC are also evident. For example, the results of the domain estimates of each of these individual peaks (table 5.3) improves the understanding of groundwater sources of each individual peak. For events 1 and 3 the second peak had a higher EC mean than the first peak of these events, however the second event had a smaller EC mean for the second peak due to the additional rainfall.

Event-based sampling is often used to estimate event mean concentration or event loads for water quality modelling (Wang et al., 2011; Bartley et al., 2012). Ratio based methods are commonly used to estimate *true loads*, however these require probabilistic sample designs (Wang et al., 2011). These results show how a simple time-stratified sample design can provide unbiased accurate mean estimates which can be used in conjunction with ratio based methods, or used to directly estimate event loads (Thomas and Lewis, 1995). Many water quality models now require estimates of event mean concentration (Feikema et al., 2011; Bartley et al., 2012). Time-stratified sampling offers the ability to provide unbiased accurate event mean concentrations for these models.

5.6 Conclusion

We have successfully presented the use of a simple time-stratified sampling scheme which can be implemented in the bounds of current hardware limitations. The method provided accurate mean event estimates of two water quality properties affected by different flow paths. Through this we have shown;

1. Probabilistic sampling schemes do not require great complexity to provide accurate event mean estimates.
2. It is possible to use probabilistic sampling schemes to provide accurate event means of two separate water quality variables affected by two different flow paths without any additional hardware or software.
3. Estimates of key hydrograph components can be made using re-stratification after each event.

Improvements to the sample design are possible by optimising the number of samples per strata in order to get more precise estimates of the event mean. For example, more variable strata would receive more samples. Unbiased meaningful sampling schemes are required as it is highly unlikely that continuous monitoring devices will be widely deployed. Event-based water quality sampling schemes require additional attention to provide unbiased accurate estimates of event exports and should not be limited to systematic sampling schemes.

5.7 References

Arabkhedri, M., Lai, F., Noor-Akma, I., Mohamad-Roslan, M., 2010. An application of adaptive cluster sampling for estimating total suspended sediment load. Hydrology Research 41 (1), 63–73.

Australian Bureau of Statistics, 2010. National regional profile: Cootamundra (a) (statistical local area).
http://www.abs.gov.au/AUSSTATS/abs@nrp.nsf/Latestproducts/150102200Population/People1200
2010?opendocument&tabname=Summary&prodno=150102200&issue=2006-2010

Bartley, R., Speirs, W., Ellis, T., Waters, D., 2012. A review of sediment and nutrient concentration data from Australia for use in catchment water quality models. Marine pollution bulletin 65 (4), 101–116.

Biggs, B., 1995. The contribution of flood disturbance, catchment geology and land use to the habitat template of periphyton in stream ecosystems. Freshwater biology 33 (3), 419–438.

Burt, T., Howden, N., Worrall, F., McDonnell, J., 2011. On the value of long-term, low-frequency water quality sampling: avoiding throwing the baby out with the bathwater. Hydrological Processes 25 (5), 828–830.

Cassidy, R., Jordan, P., 2011. Limitations of instantaneous water quality sampling in surface-water catchments: Comparison with near-continuous phosphorus time-series data. Journal of Hydrology 405 (1), 182–193.

Conyers, M., Hume, I., Summerell, G., Slinger, D., Mitchell, M., Cawley, R., 2008. The ionic composition of the streams of the Mid-Murrumbidgee River: Implications for the management of downstream salinity. Agricultural Water Management 95 (5), 598–606.

de Gruijter, J., Brus, D., Bierkens, M., Knotters, M., et al., 2006. Sampling for natural resource monitoring. Springer Verlag.

Drewry, J., Newham, L., Croke, B., 2009. Suspended sediment, nitrogen and phosphorus concentrations and exports during storm-events to the Tuross Estuary, Australia. Journal of Environmental Management 90 (2), 879–887.

Feikema, P., Sheridan, G., Argent, R., Lane, P., Grayson, R., 2011. Estimating catchment-scale impacts of wildfire on sediment and nutrient loads using the e2 catchment modelling framework. Environmental Modelling & Software 26 (7), 913–928.

Gordon, N., McMahon, T., Finlayson, B., Gippel, C., Nathan, R., 2004. Stream hydrology: an introduction for ecologists. Wiley.

Harmel, R., Cooper, R., Slade, R., Haney, R., Arnold, J., 2006. Cumulative uncertainty in measured streamflow and water quality data for small watersheds. Transactions of the ASAE (American Society of Agricultural Engineers) 49 (3), 689.

Harmel, R., King, K., Wolfe, J., Torbert, H., 2002. Minimum flow considerations for automated storm sampling on small watersheds. Texas Journal of Science 54 (2), 177–188.

Isbell, R., 2002. The Australian soil classification. Vol. 4. Csiro Publishing.

Johnes, P., 2007. Uncertainties in annual riverine phosphorus load estimation: impact of load estimation methodology, sampling frequency, baseflow index and catchment population density. Journal of Hydrology 332 (1-2), 241–258.

Kim, J., Furumai, H., 2012. Improved calibration of a rainfall-pollutant-runoff model using turbidity and electrical conductivity as surrogate parameters for total nitrogen. Water and Environment Journal.

Kirchner, J., Feng, X., Neal, C., Robson, A., 2004. The fine structure of water-quality dynamics: The (high-frequency) wave of the future. Hydrological Processes 18 (7), 1353–1359.

Kronvang, B., Bruhn, A., 1996. Choice of sampling strategy and estimation method for calculating nitrogen and phosphorus transport in small lowland streams. Hydrological Processes 10 (11), 1483–1501.

Lane, P., Sheridan, G., Noske, P., 2006. Changes in sediment loads and discharge from small mountain catchments following wildfire in south eastern Australia. Journal of Hydrology 331 (3), 495–510.

Lark, R., Cullis, B., 2004. Model-based analysis using REML for inference from systematically sampled data on soil. European Journal of Soil Science 55 (4), 799–813.

Lohr, S., 2009. Sampling: design and analysis. Thomson.

Marsh, N., Waters, D., 2009. Comparison of load estimation methods and their associated error. In: 18th IMACS World Congress MODSIM. Vol. 4. Cairns.

Pellerin, B., Wollheim, W., Feng, X., Vörösmarty, C., 2008. The application of electrical conductivity as a tracer for hydrograph separation in urban catchments. Hydrological Processes 22 (12), 1810–1818.

Särndal, C., Swensson, B., Wretman, J., 2003. Model assisted survey sampling. Springer Verlag.

Sydney Catchment Authority, 2011. Annual water quality monitoring report annual water quality monitoring report 2010-11. Annual report, Sydney Catchment Authority, Sydney, Australia.

Thomas, R., 1985. Estimating total suspended sediment yield with probability sampling. Water Resources Research 21 (9), 1381–1388.

Thomas, R., 1988. Monitoring baseline suspended sediment in forested basins: the effects of sampling. Hydrological Sciences Journal 33 (5), 499–514.

Thomas, R., Lewis, J., 1993. A comparison of selection at list time and time-stratified sampling for estimating suspended sediment loads. Water Resources Research 29 (4), 1247–1256.

Thomas, R., Lewis, J., 1995. An evaluation of flow-stratified sampling for estimating suspended sediment loads. Journal of Hydrology 170 (1-4), 27–45.

Wade, A., Palmer-Felgate, E., Halliday, S., Skeffington, R., Loewenthal, M., Jarvie, H., Bowes, M., Greenway, G., Haswell, S., Bell, I., et al., 2012. Hydrochemical processes in lowland rivers: insights from in situ, high-resolution monitoring. Hydrology and Earth System Sciences 16 (11), 4323.

Wang, Y., Kuhnert, P., Henderson, B., 2011. Load estimation with uncertainties from opportunistic sampling data–a semiparametric approach. Journal of Hydrology 396 (1), 148–157.

Warren, A., Gilligan, L., Raphael, N., 1995. Explanatory notes, Cootamundra 1:250 000 geological sheet si/55-11, Geological Survey of New South Wales, Sydney.

Chapter 6

Estimating water quality using linear mixed models with stream discharge and turbidity

6.1 Summary

Most water quality monitoring schemes rely on estimation methods as it is often too expensive to monitor water quality properties continuously. Estimations are used to evaluate management strategies and long term trends. It is critical that the methods used provide accurate estimations of water quality and the associated uncertainty. Currently the most common estimation methods assume observations are sampled using a probabilistic sampling scheme, however this assumption is often not met. This chapter evaluated the ability of a linear mixed model to estimate water quality concentration values based on observations collected using non-probabilistic sampling. The linear mixed models were used to predict total phosphorus and total nitrogen observations from two catchments in south east Australia. A comparison between stream discharge and turbidity as predictors is made to investigate the effectiveness of turbidity to estimate water quality. In addition to stream discharge and turbidity, several covariates were derived from stream discharge in an attempt to account for *hydrological processes. To compare models and their covariates leave one out event cross validation was performed. Event cross validation evaluated predictions during periods of high stream discharge. For both catchments the use of turbidity instead of stream discharge increased the accuracy of predictions by at least 15% for total phosphorus and total nitrogen. However, event based cross validation indicated that a combination of both turbidity and stream discharge based variables provided more accurate predictions, increasing the event RMSE by 18% for total phosphorus and 24% for total nitrogen. In catchments characterised by long periods of base-flow and short rainfall events the addition of turbidity measurements provide more accurate predictions during base flow and during events.*

6.2 Introduction

Water quality monitoring provides critical information about the health of a catchment. In many situations catchment managers require accurate information to be able to implement management strategies. In Australia the relationship between stream discharge and other variables is quite complex (Davis and Koop, 2006; Drewry et al., 2009). Total nitrogen and total phosphorus are two key nutrients in Australian catchments, for example in large concentrations these two naturally occurring nutrients can often cause algae blooms (Davis and Koop, 2006; Kristiana et al., 2011). As a result catchment managers require estimates of nutrient fluxes to understand and manage catchment processes.

However environmental sampling is expensive, and water quality is no exception. With large catchments and numerous properties it is unrealistic to expect continuous time series data for all properties at all sampling locations. Combined with large analytical costs larger Australian catchments often have additional expenses due to travel time. The annual expense of water quality monitoring in Australia is estimated to be in excess of $142M (Bartley et al., 2012; Kristiana et al., 2011). Most monitoring schemes can only afford to continuously monitor stream discharge and rely on sparse water quality sampling. Therefore suitable and reliable methods are required to gain an understanding of the processes within a catchment. Many studies rely on load estimation methods to evaluate water quality over a duration of time (e.g. monthly or annually). Australian catchment managers use the Australian and New Zealand guidelines for fresh and marine water quality (ANZECC) guidelines to assess water quality (ANZECC, 2000; Bartley et al., 2012). The ANZECC guidelines provide concentration based thresholds for various variables and catchment types. As the guidelines provide threshold values in the form of concentrations, catchment managers require methods to evaluate the observed data in relation to these thresholds. Catchment managers also rely on load estimations to evaluate management practices and perform trend assessments. Regression based methods use affordable covariates such as stream discharge to estimate temporal water quality concentrations, which can also provide load estimates, all at the frequency of stream discharge.

The most common water quality sampling scheme is based on sampling at equally spaced intervals in time. In south eastern Australian catchments it is common for catchment managers to use a monthly sampling scheme and in some instances a form of storm based sampling as these events correspond to high nutrient exports (Armstrong and Mackenzie, 2002; Drewry et al., 2009; Bartley et al., 2012). South-east Australian rivers are characterised by short rainfall events. These short rainfall events are often separated by long dry periods, increasing the amount of export in following events (Drewry et al., 2009). Hopmans and Bren (2007) discovered that 70% of 6 years suspended sediments was exported during one rainfall event in a part of the Buffalo River catchment in north eastern Victoria. Increasing complexity is introduced as the relationship between water quality and stream discharge differs within and between events (Drewry et al., 2009). One issue within events is the hysteresis between stream discharge and water quality properties, which is caused by different trends during the rising and falling stages of the hydrograph. In addition, the distance between rainfall events can vary and may effect the amount of nutrients exported during the initial rising stage of the event hydrograph.

The importance of load estimation methods for monitoring water quality is evident by the the amount of load estimation methods available. In a single study Marsh and Waters (2009) evaluated 34 different load estimation techniques. More recently artificial neural networks have been shown to provide accurate water quality load estimates (He et al., 2011). However, the majority of load estimation techniques fall into three main categories; average, ratio and regression methods (Marsh and Waters, 2009). In the simplest form averaging methods use the product of the mean concentration and the corresponding mean discharge

$$L = N\left(\bar{q}\bar{c}\right) \tag{6.1}$$

where $\bar{q}$ is the mean stream discharge at the sampling times, $\bar{c}$ is the mean concentration, and N is the total samples of the continuously sampled stream discharge Q (Cooper and Watts, 2002). Ratio based methods extend the averaging methods to include all observed stream discharge values

$$L = \frac{N\left(\bar{q}\bar{c}\right)}{\bar{q}}\bar{Q} \tag{6.2}$$

where $\bar{Q}$ is the mean of all observed stream discharge values (Cooper and Watts, 2002). Both averaging and ratio based methods can only provide load estimations for time intervals which have enough observations to calculate a mean. For example, for monthly sampling this would be a 2 monthly average. On the other hand regression methods provide a continuous concentration estimate and the integral multiplied by Q provides a load estimate. The most commonly applied regression method is

$$c(\mathbf{t}) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{6.3}$$

where c is the log transformed concentration at sampled time $\mathbf{t}$, $\mathbf{X}$ is the design matrix, often only using log transformed stream discharge (q) as a predictor, $\boldsymbol{\beta}$ is a vector of coefficients, and $\boldsymbol{\epsilon}$ is a vector of errors with a normal distribution. When fitted using ordinary least squares it is assumed that the errors are independent and identically distributed (*iid*) and have the following distribution; $\boldsymbol{\epsilon} \sim \mathcal{N}\left(0, \sigma^2\right)$ (Lark and Cullis, 2004).

The three common load estimation methods have been compared in many studies (Cassidy and Jordan, 2011; Marsh and Waters, 2009; Johnes, 2007; Cooper and Watts, 2002; Kronvang and Bruhn, 1996). These studies often examine the effect of sample size on load estimates. Ratio based methods have been shown to provide the most accurate estimates of the three categories by Cassidy and Jordan (2011) and Johnes (2007). Marsh and Waters (2009) found ratio based methods to provide the most accurate load estimates when water quality sample sizes are smaller than 20 and continuous stream discharge data is available. However, both Marsh and Waters (2009) and Quilbé et al. (2006) proposed the use of a regression method in the presence of a strong correlation between stream discharge and the water quality property.

Sampling schemes that use either routine sampling (e.g. monthly) or a combination of routine and event based sampling are not probabilistic (i.e the samples times are not selected randomly). By using this type of sampling scheme there is an unknown inclusion probability of collecting a sample at a particular point in time. The three main types of estimation techniques all assume probabilistic sampling. The bias due to these sampling schemes and the assumptions of the load estimation methods has long been acknowledged (Thomas, 1985, 1988; Crawford, 1991; Cohn et al., 1992; Cooper and Watts, 2002; Cohn, 2005). Average and ratio based methods assume the data is sampled using simple random sampling which is rarely the case (Cooper and Watts, 2002). Regularly used regression methods fitted using ordinary least squares will provide unbiased estimates of coefficients, however the variance estimates will be biased when the sampling scheme is not probability based (Lark and Cullis, 2004). This is a problem when both predictions and the prediction variance are required.

Linear mixed models (LMM) provide the ability to handle non-probability based sampling schemes by using a model-based approach (Lark and Cullis, 2004). Lark and Cullis (2004) compared ordinary least squares and LMM methods for estimating soil attributes. Their results indicated that the variance estimates from OLS were biased, as the OLS methods assumed the samples where independent of each other and had equal inclusion probabilities. They found an increase in variance with the use of LMMs as the model accounts for the non-probabilistic sampling scheme. Water quality sampling shares many similarities to systematic soil sampling, as the samples are non-probabilistic and there is auto-correlation between samples. With these similarities, LMM based estimations should provide less biased estimates of the prediction variance, than conventional methods. Furthermore, since LMMs model the auto-correlation in the model residuals, this can also be used to interpolate the model residuals using kriging. The kriged residuals are added to regression predictions at each prediction location to give an improved prediction (Bivand et al., 2008). Another major benefit from using regression based methods is due to the ability to include covariates other then stream discharge e.g. rising and falling limbs and time since the last rainfall event. In addition to these covariates Wang et al. (2011) also proposed the use of a discounted flow covariate which uses a weighting function to account for stream discharge prior to events. Turbidity has also been used in regression based models to estimate TP as it directly relates to the water quality of the stream. Jones et al. (2011) found in-situ measurements of turbidity were significant covariates for estimating TP in a catchment in Utah. With the existence of relative low cost reliable turbidity sensors it is now feasible for catchment managers to use these sensors for continuous monitoring.

Therefore the aims of this paper are to

- present the use of LMMs for predicting water quality
- compare the use of discharge-related predictor variables proposed by Wang et al. (2011) with turbidity measurements
- focus the comparison on the prediction quality for flow events as these are when most export of nutrients and sediments occurs under Australian conditions.

This will be illustrated with a dataset of TP and TN for 2 sub-catchments draining into Lake Burragorang, the main reservoir for supplying Sydney's drinking water.

6.3 Materials and methods

6.3.1 Catchment description

This study involves the analysis of two sub catchments within the greater Lake Burragorang catchment. Lake Burragorang is located south west of Sydney, Australia. The lake is responsible for delivering 80% of Sydney's drinking water (Armstrong and Mackenzie, 2002; Kristiana et al., 2011). The immediate surrounding area of the lake is closed to the public and mainly consists of native forest. The remaining areas of the catchment mainly consist of agricultural pasture, with small urbanised areas. Two main land forms are found in the catchment; a gorge system nearest the lake, and a plateau in the southern outer reaches (Armstrong and Mackenzie, 2002). Approximately 105000 people reside in the greater catchment (Kristiana et al., 2011).

Figure 6.1 shows the location of the two sub catchments in relation to Lake Burragorang. The two sub catchments are quite different in size, with the southern catchment (Wollondilly) being 4800 km^2 and the narrower northern catchment (Coxs) being 1448 km^2, with lengths of 251 and 106 km respectively. The maximum length of overland flow of each catchment was determined using the methods outlined by Jasiewicz and Metz (2011). Table 6.1 highlights these differences. The Coxs catchment is dominated by gorges, with 70% coverage by forest and only 25% pasture/grassland with a mean elevation of 850 m. In contrast the Wollondilly catchment has 57% pasture/-grassland, mainly in the mid/upper reaches with 39% forest, mainly located nearest to the lake and a mean elevation of 710 m. Table 6.1 shows the northern part of the main Lake Burragorang catchment has a higher annual rainfall as compared to the southern part.

6.3.2 Data description

The outlet of both sub-catchments has a near continuous (15 minute) stream height monitoring device, and a storm based ISCO 3700 automatic sampler (ISCO Inc. Lincoln, Nebraska, America). The current sampling scheme consists of routine monthly

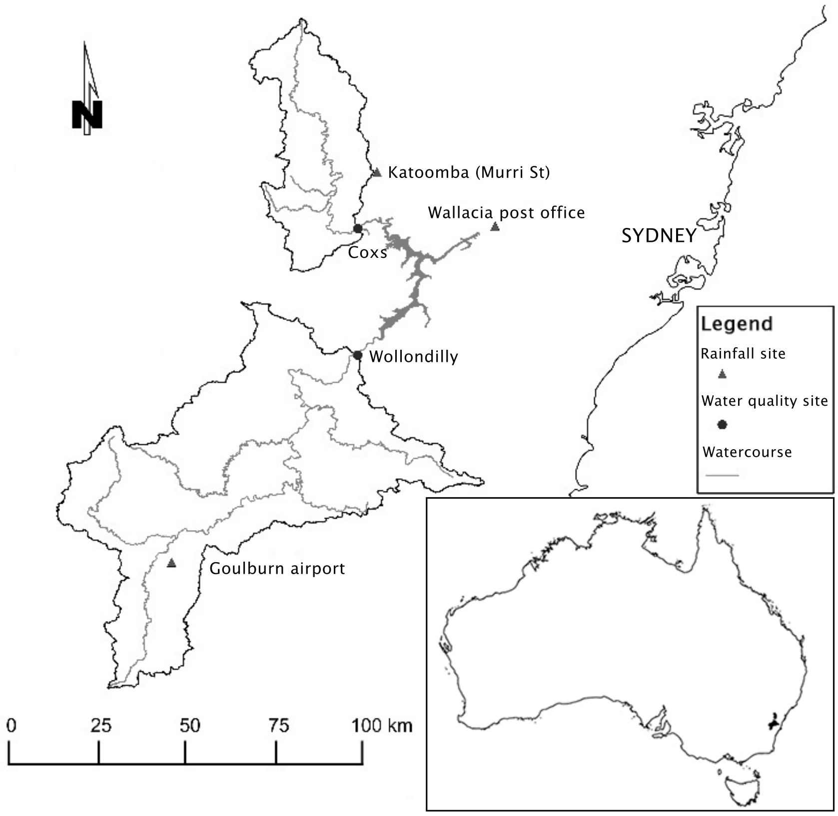

Figure 6.1 – Map indicating location of the two sub-catchments in relation to Lake Burragorang, and location within Australia.

grab samples and storm based sampling which targets the upper 90 percentile of stream discharge. The automatic samplers are designed to trigger at a predetermined stream height which is the 90th percentile. Using two stratifications the automatic samplers are setup to take samples at short regular intervals during the start of an event, and longer regular intervals later in the event. Strata lengths and sampling intervals are based on historical records at both sites, and have changed since the samplers introduction. An almost continuous time series of stream discharge is available for both sub catchments for the study period (1991 - 2008). Total phosphorus and total nitrogen have been sampled on a monthly basis at both sub catchments, using acid digestion methods. Since the introduction of the storm based samplers, both sites have sporadic samples from storm events, however the frequency of these samples are varied, and based on several factors including actual storm frequency.

Table 6.1 – Summary statistics for the two sub-catchments

Sub-catchment characteristics	Wollondilly	Coxs
Landcover (Main types)		
Pasture/grassland (%)	57	25
Forest (%)	39	70
Topographical features		
Maximum overland flow distance to outlet (km)	251	106
Minimum elevation (m)	85	115
Mean elevation (m)	710	850
Maximum elevation (m)	1176	1331
Mean annual rainfall (1991 - 2008)[a]		
Goulburn Airport (mm)	507	
Wallacia Post Office (mm)	740	740
Katoomba (Murri St) (mm)		1241
Mean annual stream discharge statistics (1991 - 2008)		
Total discharge (GL year^{-1})	4609	767
Minimum daily discharge (ML d^{-1})	10	19
Mean daily discharge (ML d^{-1})	529	267
Maximum daily discharge (ML d^{-1})	36004	21501

[a] See Figure 6.1 for rainfall locations

The ANZECC (2000) guidelines for these sub-catchments recommend a threshold of 0.02 mg L^{-1} and 0.25 mg L^{-1} for TP and TN respectively. Table 6.2 shows that at both sites both the observed median and mean (except TP at the Coxs sub-catchment) are above the recommended levels. The total samples collected at each site is higher for the Wollondilly sub-catchment for both TP and TN. The Coxs sub catchment has a maximum observed TN value twice that of the Wollondilly sub-catchment, however both sub catchments have similar observed TP values. For each site the hourly stream discharge was assigned to each water quality observation. Table 6.2 also provides a summary of observed turbidity for each sub-catchment, these values are derived from laboratory analysis for all samples that were collected at the site. At each site the upper detection limit of turbidity was 1000 NTU, therefore the maximum observed value for both catchments is 1000 NTU. As it is not possible to determine the actual turbidity for these samples, they have been excluded from the models. A greater amount of turbidity measurements have been conducted as it is cheaper to analyse. We rely on lab measured turbidity to reflect the possibility of using an in-situ turbidity probe to provide continuous turbidity measurements.

Table 6.2 – Summary statistics of water quality for each sub-catchment .

Sub-catchment	Wollondilly			Coxs		
Variable	TP mg L^{-1}	TN mg L^{-1}	Turbidity NTU	TP mg L^{-1}	TN mg L^{-1}	Turbidity NTU
min	0.001	0.17	0.54	0.001	0.05	0.2
mean	0.06	0.78	69.39	0.03	0.50	17.61
median	0.02	0.58	15.25	0.01	0.23	4.55
max	1.26	6.43	1000*	1.92	13.2	1000*
variance	0.01	0.37	18104.98	0.02	1.07	3342.6
skewness	6.39	3.5	3.97	11.25	9.62	11.12
n	454	464	1242	386	390	1062

* Maximum detection limit

6.3.3 Statistical analysis

The model applied in this paper is often applied in spatial sciences. The LMM is beneficial in spatial sciences as it accounts for the spatial auto-correlation between samples. In this paper the spatial auto-correlation is replaced by the temporal auto-correlation between the water quality observations. It is important to note that the LMM used in this paper relies on a variogram structure to model the temporal auto-correlation and therefore requires samples at temporal distances that help build the model of the temporal auto-correlation structure.

The LMM differs from the linear model (eq. 6.3) by treating all effects not taken into account in the $(n \times p)$ design matrix $\mathbf{X}$ as a random variable $\boldsymbol{\eta}$ (Lark and Cullis, 2004)

$$c(\mathbf{t}) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}. \tag{6.4}$$

There is no requirement of *iid* samples as $\boldsymbol{\beta}$ is now assumed to have a correlation structure $\boldsymbol{\eta} \sim \mathcal{N}(0, \mathbf{V})$, where $\mathbf{V}$ is a positive definite matrix of the variance (σ^2)

and the covariance. There are many different model structures available for the auto-correlation between the samples. In this paper the temporal auto-correlation is assumed to be in the form of an exponential model. Therefore $\mathbf{V}$ has the following structure:

$$\begin{aligned} \mathbf{V}_{i,j} = \quad & \sigma^2 s \exp\left(-\frac{|\mathbf{x_i}-\mathbf{x_j}|}{a}\right), \quad && i \neq j \\ & \sigma^2, && i = j, \end{aligned} \tag{6.5}$$

with the variance σ^2 as the diagonal and all other values characterised by the exponential covariance structure where $|\mathbf{x_i} - \mathbf{x_j}|$ is the temporal distance between two samples, a is the distance parameter of the exponential function, and s is defined as;

$$s = \frac{c}{c_0 + c}, \tag{6.6}$$

where c_0 is the nugget component, which is the unexplained variance, often referred to as sampling error from laboratory analysis, with $c_0 + c$ describing the maximum variance between two variables. As the temporal structure of the model is heavily dependent on the observed samples, it is important that samples are observed at proximities that are within the range of auto-correlation. If the temporal distances between samples are greater then the actual temporal structure of the auto-correlation ($\sim 3a$) then this may result in a nugget model. (Marchant and Lark, 2007).

The estimation of the model requires the estimation of both the model coefficients ($\boldsymbol{\beta}$) and the parameters ($\boldsymbol{\theta}$) of the variance-covariance matrix ($\mathbf{V}$). In this paper residual maximum likelihood (REML) is used. As explained by Lark and Cullis (2004) the estimation of $\boldsymbol{\beta}$ is conditional on $\boldsymbol{\theta}$ when using REML. The models were fitted using the optimisation methods of the geoR package (Ribeiro Jr and Diggle, 2001).

It is possible to predict concentration (c) at an un-sampled time t_0 using a $1 \times p$ row vector of $x(t_0)$ of the prediction design matrix (x). Based on the LMM (equation 6.4) the prediction equation is defined as follows;

$$c(t_0) = x(t_0)\hat{\boldsymbol{\beta}} + \mathbf{v}'\mathbf{V}^{-1}\left(c(\mathbf{t}) - \mathbf{X}\hat{\beta}\right), \tag{6.7}$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}c(\mathbf{t})$ is the least squares estimate of the trend coefficients. The equation is a combination of the regression estimate at the location $x(t_0)\hat{\boldsymbol{\beta}}$ and the kriged residuals of the model $(c(\mathbf{t})-\mathbf{X}\hat{\beta})$ estimated from the auto-correlation of the observed and prediction locations $(\mathbf{v}'\mathbf{V}^{-1})$ (Bivand et al., 2008). Using this method the residuals of nearby observations are used to improve the model predictions. The associated prediction variance is;

$$\sigma^2(t_0) = \sigma_0^2 - \mathbf{v}'\mathbf{V}^{-1}\mathbf{v} + \delta\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\delta', \tag{6.8}$$

where σ_0^2 is the variance of $c(t_0)$ and $\delta = x(t_0) - \mathbf{v}'\mathbf{V}^{-1}\mathbf{X}$.

When all observations are uncorrelated with the prediction location $\mathbf{v}'\mathbf{V}^{-1}\mathbf{v}$ equals zero and equals σ_0^2 when the prediction location coincides with an observation. The last term in equation 6.8 is the estimation error, i.e var $\left(\hat{\beta}-\beta\right) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$ (Bivand et al., 2008). In the context of this study the concentration values were log transformed for the model. The back transformation of the predicted concentration (c) is;

$$T(c(t_0)) = \exp\left(c(t_0) + \frac{\sigma^2(t_0)}{2}\right) \tag{6.9}$$

and the back-transformed prediction variance is;

$$T(\sigma^2(t_0)) = \exp\left(2c(t_0) + \sigma^2(t_0)\right)\left(exp(\sigma^2(t_0) - 1\right). \tag{6.10}$$

The prediction quality of each model was assessed using leave-one-out-cross-validation (LOOCV), where each sample is removed from the model and then predicted based on the remaining samples. Prediction quality was assessed using three different measures;

Mean squared error (ME)

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\bar{c}}(t_i) - c(t_i)\right) \tag{6.11}$$

Residual mean square error (RMSE)

$$= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\bar{c}(t_i) - c(t_i)\right)^2} \tag{6.12}$$

Mean square deviation ratio (MSDR)

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{\left(\hat{\hat{c}}(t_i) - c(t_i)\right)^2}{\hat{\sigma}^2(t_i)}, \tag{6.13}$$

where $\hat{\hat{c}}(t_i)$ is the prediction and $c(t_i)$ is the observation at the time of the i_{th} sample based on the LOOCV results. The ME reflects the bias of the estimates and should be close to zero, accuracy of the model is assessed using the RMSE, where smaller values are desired. The MSDR reflects how the prediction variance represents the actual error and should be close to 1. For each model the cross-validation was performed using the predictions on the log scale. In addition to performing LOOCV on the entire data set, cross validation was also performed on event based samples. As large amounts of nutrients are exported during events it is important that the performance of a model during these periods are examined. The cross validation of events was performed by removing all observations within an event and predicting at these sample times. This process was performed for each event and was assessed using the same measures as the LOOCV.

6.3.4 Model covariates

Several covariates were examined for TP and TN at each site to explain additional variation that is not explained by stream discharge. Four models were produced for TP and TN at both sub-catchments. The covariates of the four models are based on;

model 1: stream discharge,

model 2: stream discharge and derivatives of stream discharge,

model 3: turbidity and

model 4: a combination of model 2 and model 3.

Stream discharge was examined first, as this is commonly measured and used in load estimation studies. As explained by Wang et al. (2011) it is possible to derive covariates to reflect controlling factors of nutrient exports and these were included in model 2. To account for hysteresis of the water quality-discharge relationship during events, a covariate representing the direction of change of stream discharge during an event was used to reflect if the stream was rising or falling. The "first flush" phenomena is the relationship with larger exports during the start of the first event after extended dry periods. Wang et al. (2011) describes this in regards to tropical environments where during dry periods there is an accumulation of biomass in the catchment, which is exported during the initial stages of the first rainfall event of the wet-season. As south-east Australian catchments experience long dry periods the first-flush effect will also be investigated. The first flush covariate was formed based on the distance between periods when discharge was greater then the annual 90th percentile. Another approach proposed by Wang et al. (2011) to account for different duration of base-flow between events is to use discounted flow. The formula for discounted flow is

$$DF(d) = \frac{\sum_{i=1}^{j} d^{j+1-i}\hat{q}_j}{\sum_{i=1}^{j} d^{j+1-i}}, \tag{6.14}$$

where several different levels of d can be used. The formula is effectively the average of j historical observations with the weight of each observation i decreasing exponentially back to the jth time. Based on Wang et al. (2011) five levels of d were examined (0.50, 0.75, 0.9, 0.95, 0.99), where 0.95 per day is equivalent to 0.5 per fortnight and less then 0.5 is the current stream discharge (Wang et al., 2011).

Models 1 and 3 only included stream discharge and turbidity respectively. Model 2 was fitted using stream discharge, time since last event and the event direction as a base model. Each DF level was added one-by-one to determine which DF covariate was the most significant. Wald tests were used to assess the significance of each covariate, this procedure is outlined by Lark and Cullis (2004). At this stage the model included stream discharge, time since last event, event stream direction and the most significant DF covariate. For each model backwards elimination was used to determine which covariates were significant variables. Model 4 was a combination of model 3 and model 2. With the combined model backwards elimination was used to determine the effectiveness of turbidity as a covariate, and examine which if any covariates would be dropped from the model.

6.4 Results

6.4.1 Stream discharge models

By using a LMM it is assumed that the relationship between stream discharge and the water quality variables form a linear relationship. Figure 6.2 shows the linear relationship between the transformed stream discharge and water quality variables at each site, with model 1 shown by the line. The R^2 of the fitted values only using the linear components of the model is also shown for each model. There is scatter at low stream discharge levels in the relationship for both TN and TP at each site. This scatter appears to affect the slope of the linear relationship causing underestimation at higher stream discharge levels. In addition to the effect of the scattering at low stream discharge there are also other factors controlling TP and TN during events. Figure 6.4 a illustrates this as it shows the observed TP-discharge relationship for the Wollondilly sub-catchment during an event in 2007 with the corresponding predictions of the stream discharge model (model 1). The observed values during this event are well above the model, and the model does not intersect with the observed samples, indicating that predictions based on this model would underestimate TP during events.

A major benefit of using LMM is the ability to model the temporal auto-correlation between the residuals, and use this to krige the residuals to improve the predictions. Figure 6.3 shows the modelled correlation structure (line) of the stream discharge based models (model 1) overlaying the semivariogram. The sill of both TP models indicate the maximum variance between the residuals with the auto-correlation between samples of a greater distance then ($\sim 3a$) equal to that of the fitted sill.

Table 6.3 summaries the stream discharge based models (model 2) fitted using backward elimination with the most significant discounted flow variable. For all models the inclusion of the additional covariates reduced the RMSE for Wollondilly (TP, TN)

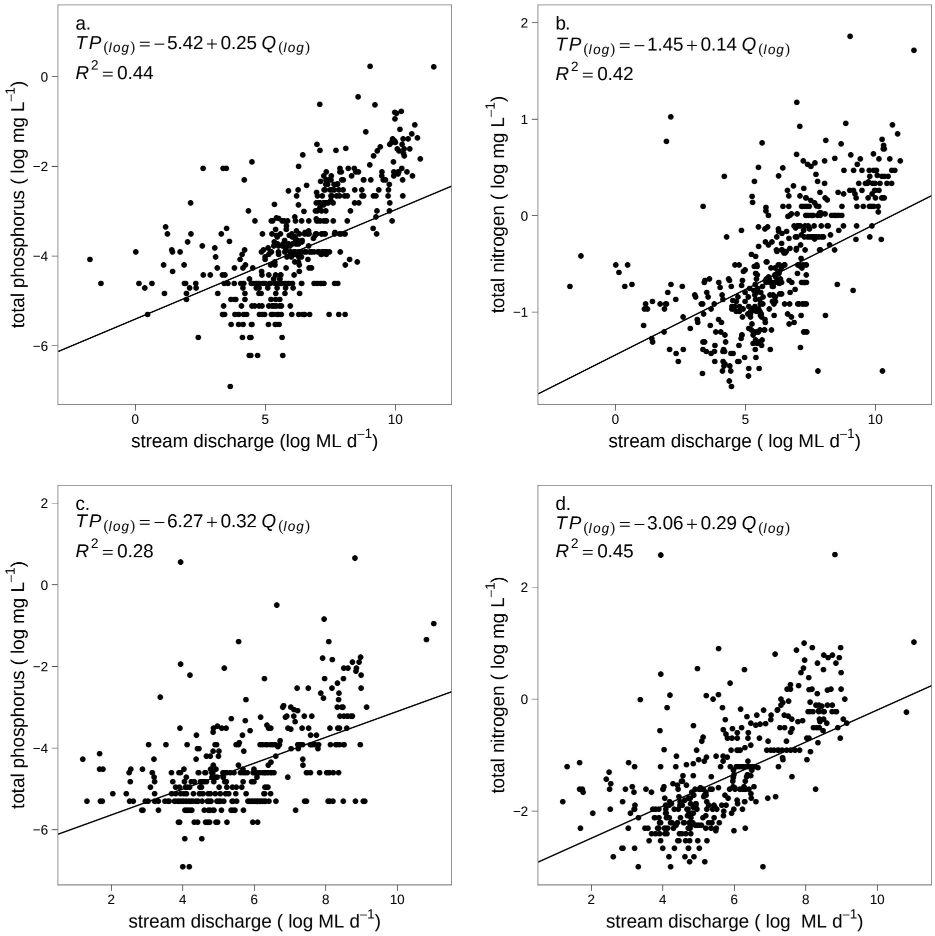

Figure 6.2 – Linear model between stream discharge and water quality variables (model 1). a) Total phosphorus for Wollondilly sub-catchment. b) Total nitrogen for Wollondilly sub-catchment. c) Total phosphorus for Coxs sub-catchment. d) Total nitrogen for Coxs sub-catchment.

and Coxs (TP, TN) from (0.78, 0.41, 0.82, 0.60) to (0.73, 0.39, 0.75, 0.58) respectively. For each model the most significant discounted flow variable stayed in each model with both Wollondilly models including discounted flow with d = 0.5. Event distance remained in the final model for both TN and TP models at Wollondilly. The event direction covariate remained in the TN models at both sites and the TP model in the Coxs sub-catchment. The temporal range of the variogram structure was larger for both models in the Wollondilly sub-catchment, where the range was less then 4 days for both TN and TP in the Coxs sub-catchment. LOOCV results show that there is little bias in each model and based on the MSDR the predicted variance re-

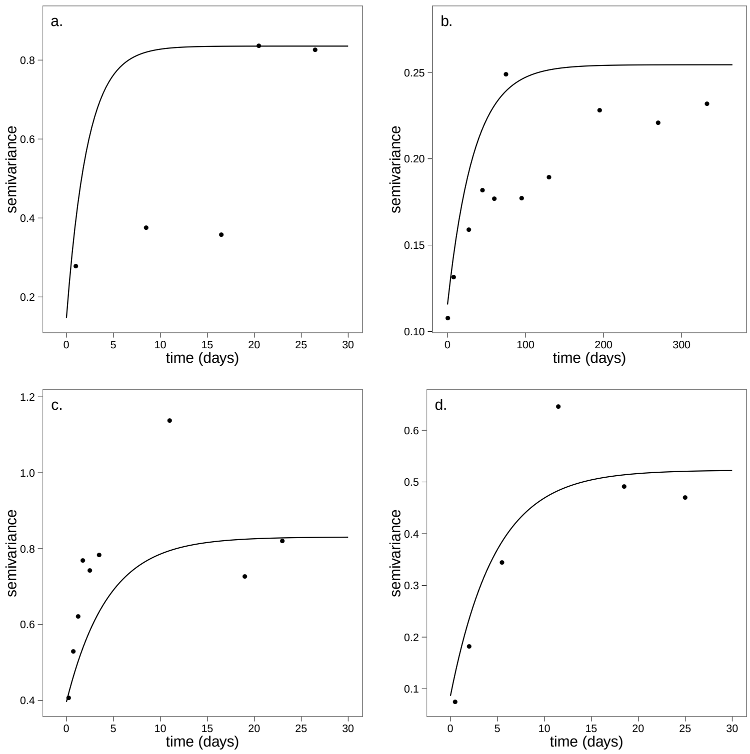

Figure 6.3 – Semivariogram of residuals for discharge only models. a) Total phosphorus for Wollondilly sub-catchment. b) Total nitrogen for Wollondilly sub-catchment. c) Total phosphorus for Coxs sub-catchment. d) Total nitrogen for Coxs sub-catchment.

flects the actual error. Each model must be evaluated in regards to periods of higher flows which relate to short events. The accuracy of predictions during events without the support of event based samples showed that the accuracy of all models decreased, however the MSDR indicates that the predicted variance of TN Coxs underestimated the actual error, whereas the other three models over-estimated the actual error.

6.4.2 Turbidity models

Previous research in the catchment had found strong relationships between nutrients and turbidity measurements. For each catchment, TN and TP were modelled using turbidity instead of stream discharge. At each site model 3 using only turbidity showed that turbidity was a significant covariate for TN and TP. By adding turbidity

Table 6.3 – Fitted models using backward elimination based on stream discharge variables (model 2). Discharge only models provided within parentheses.

Statistic	Wollondilly		Coxs	
Variable	TP	TN	TP	TN
Partial regression coefficients				
intercept	-4.79	-0.68	-5.50	-1.93
discharge	0.48	0.23	0.68	0.36
event distance	-0.06	0.1	-	-
event rising	-	-0.54	0.34	0.37
event falling	-	-0.48	0.32	0.32
DF 0.5	-0.33	-0.14	-	-
DF 0.75	-	-	-0.54	-
DF 0.9	-	-	-	-
DF 0.95	-	-	-	-
DF 0.99	-	-	-	-0.3
AIC	-2244.09 (-2243.93)	67.35 (113.81)	-2501.31 (-2441.56)	-305.91 (-248.19)
Variance structure				
nugget (c_0)	0.34	0.11	0.35	0.06
variance (σ^2)	0.45	0.13	0.36	0.35
range (a)	34.55	51.94	3.97	1.99
LOOCV results				
ME	-0.02 (-0.03)	-0.007 (-0.01)	-0.005 (-0.01)	-0.01 (-0.03)
RMSE	0.74 (0.78)	0.38 (0.41)	0.75 (0.82)	0.58 (0.6)
MSDR	1 (0.98)	0.98 (0.99)	0.94 (0.96)	1.04 (1)
Event CV results				
ME	0.27 (0.26)	0.05 (0.2)	0.001 (0.09)	0.07 (0.32)
RMSE	0.88 (0.88)	0.45 (0.5)	0.97 (1)	0.59 (0.73)
MSDR	1.29 (1.27)	1.18 (1.3)	1.31 (1.2)	0.81 (1.08)

to the stream discharge model (model 2) several covariates were dropped from the models. The discounted flow covariates were dropped from the TP and TN models at Wollondilly. The TN model at Wollondilly also dropped stream discharge, but kept event distance and event direction in the model. With the inclusion of turbidity in both TP models both event direction and event distance were dropped from the models. In regards to LOOCV turbidity (model 3) improved the accuracy by reducing the RMSE of the best discharge based model (model 2) from (0.73, 0.39, 0.75, 0.58) to (0.62, 0.33, 0.62, 0.47) for Wollondilly; TP, TN and Coxs; TP, TN respectively. Table 6.4 summaries the combination of stream discharge and turbidity models (model 4), with values within parentheses representing results of the turbidity models (model 3). For model 4, there was little change in the RMSE for both TN and TP at both catchments. Based on observed data non-soluble sources (nitrate, nitrite and ammonium) of TN only contribute an average of 18% and 22% at the Wollondilly and Coxs catchments respectively therefore turbidity is a good predictor. TP was plotted against turbidity (figure 6.4 b) for the same 2007 event as stream discharge (figure 6.4 a). The main difference between the relationship between the TP-turbidity and the TP-discharge relationship is that the turbidity based model intersects the observed samples. The R^2 of the model with the observations shown in figure 6.4 is 0.29 and 0.41 for stream discharge and turbidity model respectively.

The RMSE of the models with only turbidity and the turbidity-discharge models (table 6.4) shows that there is little difference between the accuracy of the two model types. However there are differences between the turbidity (model 3) and combined turbidity models (model 4) in regards to event cross validation. All models except TP at the Coxs sub-catchment were more accurate when hydrological based covariates were included in the model. Separated by hydrological state the soluble sources of nitrogen are higher during events at both sub-catchments, contributing (17%, 25%,33%) and (13%, 24% ,24%) during base-flow, rising and falling stages at Coxs and Wollondilly sub-catchments respectively. These differences between the hydrograph stages is why the event direction covariate was kept in the final model and why the extra variables give better predictions than turbidity alone. The increase in soluble nitrogen

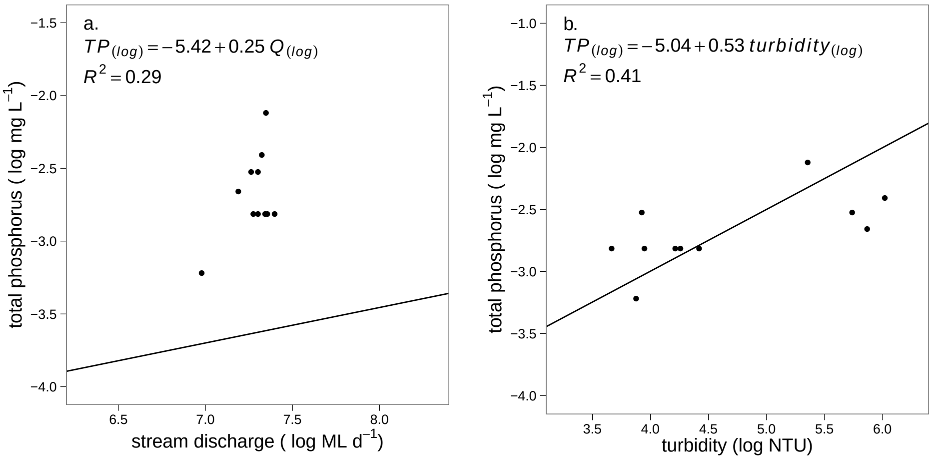

Figure 6.4 – An event in November 2007 at the Wollondilly sub-catchment with global model and R^2 for the observations of the event shown. a) Stream discharge relationship. b) Turbidity relationship.

sources may be due to a delayed response from soluble nitrogen entering the streams from through-flow and groundwater interactions. Total phosphorus models that included turbidity and hydrological based covariates had temporal correlation between samples of less then 10 days. The temporal correlation between TN samples was longer for the Coxs sub-catchment with correlation between samples up to 15.3 days. The temporal correlation between samples for TN in the Wollondilly sub-catchment was much greater then the other models with correlation between samples up to 244 days. This extended temporal correlation for TN may be due a combination of catchment size and soluble sources of nitrogen which may effect the auto-correlation between the samples at greater temporal distances.

6.4.3 Event concentration estimation

Using the best stream discharge based model (model 2) and the best turbidity based model (model 4) predictions for a set of rainfall events were made for TP in the Wollondilly sub-catchment. For the set of events, the predictions where made in the absence of all the observed TP during the three events, to reflect the effectiveness of each model when there are no observations made. In this scenario the predic-

Table 6.4 – Fitted models using backward elimination based on stream discharge variables and turbidity (model 4). Turbidity only model provided within parentheses.

Statistic	Wollondilly		Coxs	
Variable	TP	TN	TP	TN
Partial regression coefficients				
intercept	-5.29	-0.86	-4.89	-1.57
discharge	0.06	-	0.15	0.11
event distance	-	0.07	-	-
event rising	-	-0.35	-	0.2
event falling	-	-0.37	-	0.26
DF 0.5	-	-	-	-
DF 0.75	-	-	-0.17	-
DF 0.9	-	-	-	-
DF 0.95	-	-	-	-
DF 0.99	-	-	-	-0.17
turbidity	0.48	0.24	0.53	0.37
AIC	-2418.70 (-2411.53)	-94.28 (-79.39)	-2666.04 (-2659.89)	-465.9 (-432.71)
Variance structure				
nugget (c_0)	0.22	0.1	0.25	0.1
variance (σ^2)	0.26	0.07	0.19	0.16
range (a)	1.85	81.26	3.25	5.11
LOOCV results				
ME	-0.003 (-0.005)	-0.002 (-0.003)	0.003 (0.002)	-0.007 (-0.02)
RMSE	0.61 (0.62)	0.33 (0.33)	0.61 (0.62)	0.47 (0.47)
MSDR	0.99 (0.98)	0.98 (0.98)	0.95 (0.96)	1.01 (0.99)
Event CV results				
ME	0.15 (0.24)	0.01 (0.13)	-0.08 (-0.11)	0.04 (0.26)
RMSE	0.63 (0.66)	0.38 (0.42)	0.82 (0.8)	0.46 (0.54)
MSDR	0.92 (1.01)	1.24 (1.4)	1.49 (1.42)	0.87 (1.13)

tion method does not benefit from the kriging of residuals based on the temporal auto-correlation as there are no nearby observations. This comparison is designed to compare the two methods in a situation where no observations have been made, and an estimation is required. Figure 6.5 shows the observed stream discharge for the period, the observed TP, stream discharge and turbidity based predictions. The stream discharge based model over estimated the first smaller event, and underestimated the next three events. The turbidity based model estimations were closer to the observed TP values during the first event, and also underestimated TP for the following three events. Based on the observations the mean of this period was 0.15 mg L^{-1}. Using the predictions of the corresponding observations the stream discharge based estimations had estimated a mean TP of 0.11 mg L^{-1} and the turbidity based model estimated a mean of 0.12 mg L^{-1}. Using Lin's correlation coefficient as a measure of accuracy (Lawrence and Lin, 1989), the accuracy of the predictions is 0.63 and 0.17 for the turbidity and stream discharge models respectively. The main difference between the predictions of the two models is that during an event the trend of the turbidity based predictions tends to follow the trend of the observed TP during each event. This is because turbidity is directly measuring water quality so factors such as hysteresis and antecedent conditions that add noise to the discharge model do not matter as much.

6.5 Discussion

Linear mixed models provide an unbiased variance estimate of the linear based prediction in the presence of systematically collected samples, and provide a method to model the temporal auto-correlation between water quality samples. In additional to providing unbiased water quality estimates, it is possible to use or include covariates other than stream discharge to improve water quality estimates. This ability of using other covariates is beneficial due to complex nutrient transport pathways. The application of LMMs in this chapter has outlined several benefits of its use and how the results of the fitted models can be used to characterise catchments and provide information about the factors controlling water quality.

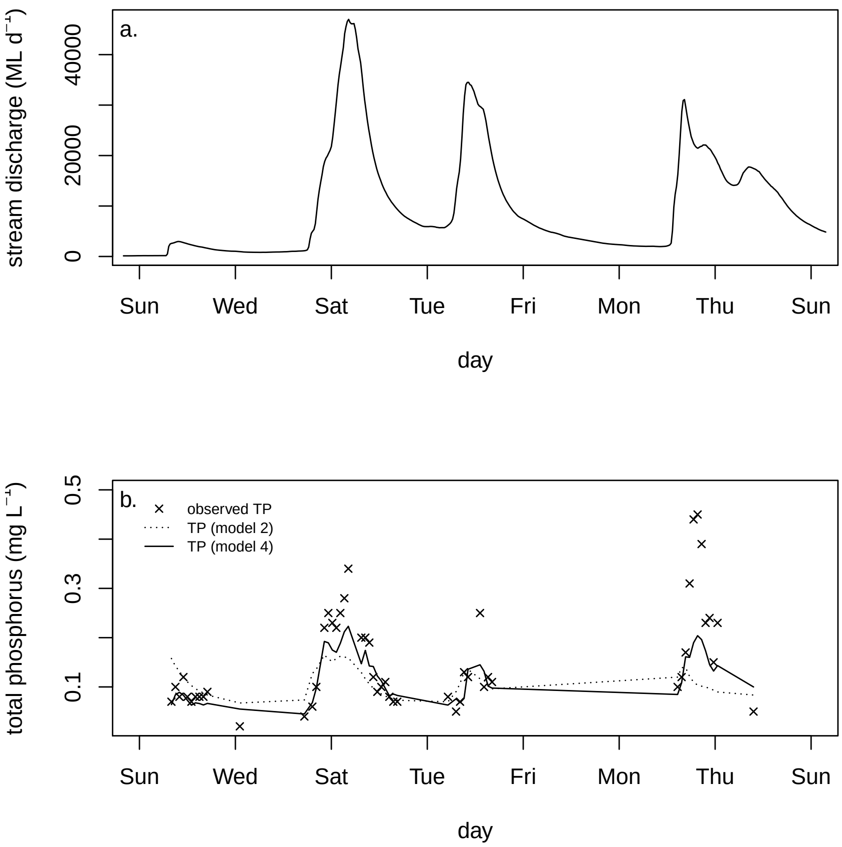

Figure 6.5 – Observed and predicted TP for a series of events in June 2007. a) Observed stream discharge during the period. b) Observed and predicted TP at corresponding sample times. Model 2 predictions are based on the stream discharge model (table 6.3) and model 4 predictions are based on the combined stream discharge and turbidity model (table 6.4).

The results indicate that the inclusion of turbidity was a significant covariate and improved the accuracy of each model for TN and TP. This indicates that in catchments where the main sources of TP and TN are non-soluble forms, turbidity is a significant covariate and provides more accurate predictions then stream discharge alone. In the Wollondilly sub-catchment turbidity accounted for the distance between events, and in the Coxs sub-catchment turbidity replaced the direction of the event for the TP models. This is because turbidity directly measures water quality and stream discharge indirectly relates to water quality.

The LOOCV results indicated that the predicted variance of the final models was equivalent to actual error. This is important for catchment managers as it gives a prediction and a validated uncertainty of prediction. The temporal range of TN in the Wollondilly sub-catchment was quite large and possibly over estimated. Variograms fitted using REML are sensitive to outliers and the fitted temporal structure is heavily dependent on the available samples and the temporal distances between them. Recently Marchant et al. (2010) have evaluated robust variogram estimation methods that are less sensitive to outliers that may prove beneficial to LMM in regards to water quality models.

Turbidity proved to be strong covariate for both TP and TN, however the event cross validation indicated that TN models required a combination of hydrological based covariates and turbidity in respect to predictions of events where no observations are made. This is an important finding as most sampling schemes in Australia do have a form of event sampling, however these sampling schemes do not provide complete coverage of the each event. In catchments with high nutrient exports during short rainfall events it is important to examine load estimation models for accuracy during these periods.

Catchment managers would benefit from the use of low-cost novel turbidity sensors with LMM to provide accurate TP and TN estimates. These high frequency estimates would greatly improve the understanding of the dynamics in the catchment. In regards to regulatory requirements, this approach would provide an estimate of the percentage of time in a certain period that the government threshold is exceeded. In addition, the accuracy of the estimate could also be reported.

The event predictions of the two models indicated that both models underestimate the observed TP, however the inclusion of turbidity in the each model improve the predictions as predictions with turbidity as a covariate tended to follow the trend within each event. With both TN and TP being contributed to by mainly non-soluble sources it should not be surprising that the turbidity was a significant covariate.

6.6 Conclusion

The use of common estimation techniques; average, ratio and regression methods must be applied in situations where systematic sampling has not occurred otherwise these estimation techniques may introduce bias in the estimation of the prediction variance. This investigation has highlighted that;

- there is temporal auto-correlation within TP and TN samples which needs to be accounted for in order to create statistically valid models. Linear mixed models provided the ability to linear based predictions in the presence of this auto-correlation.
- turbidity outperforms discharge and discharge-related predictors for predicting TN and TP. Therefore it's recommended that where possible turbidity sensors should be adopted but in their absence, discharge only methods should be avoided.
- stream discharge derived covariates and turbidity can improve TP and TN estimation models,
- validation of water quality models must examine the effects of the model in relation to storm events.

As many other studies have found estimating TP and TN is difficult during events. The LMMs presented here offer a method to account for non-probabilistic sampling often used in water quality monitoring programs. More work is required to determine optimal sampling schemes for Australian conditions with a particular emphasis on event based samples.

6.7 References

ANZECC, 2000. Australian and New Zealand guidelines for fresh and marine water quality. Australian and New Zealand Environment and Conservation Council and

Agriculture and Resource Management Council of Australia and New Zealand, Canberra, 1–103.

Armstrong, J., Mackenzie, D., 2002. Sediment yields and turbidity records from small upland subcatchments in the Warragamba dam catchment, Southern New South Wales. Soil Research 40 (4), 557–579.

Bartley, R., Speirs, W., Ellis, T., Waters, D., 2012. A review of sediment and nutrient concentration data from Australia for use in catchment water quality models. Marine Pollution Bulletin 65 (4), 101–116.

Bivand, R., Pebesma, E., Gómez-Rubio, V., 2008. Applied spatial data analysis with R. Springer Verlag.

Cassidy, R., Jordan, P., 2011. Limitations of instantaneous water quality sampling in surface-water catchments: Comparison with near-continuous phosphorus time-series data. Journal of Hydrology 405 (1), 182–193.

Cohn, T., 2005. Estimating contaminant loads in rivers: An application of adjusted maximum likelihood to type 1 censored data. Water Resources Research 41 (7), W07003.

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

Cooper, D., Watts, C., 2002. A comparison of river load estimation techniques: application to dissolved organic carbon. Environmetrics 13 (7), 733–750.

Crawford, C., 1991. Estimation of suspended-sediment rating curves and mean suspended-sediment loads. Journal of Hydrology 129 (1-4), 331–348.

Davis, J., Koop, K., 2006. Eutrophication in australian rivers, reservoirs and estuaries–a southern hemisphere perspective on the science and its implications. Hydrobiologia 559 (1), 23–76.

Drewry, J., Newham, L., Croke, B., 2009. Suspended sediment, nitrogen and phosphorus concentrations and exports during storm-events to the Tuross Estuary, Australia. Journal of Environmental Management 90 (2), 879–887.

He, B., Oki, T., Sun, F., Komori, D., Kanae, S., Wang, Y., Kim, H., Yamazaki, D., 2011. Estimating monthly total nitrogen concentration in streams by using artificial neural network. Journal of Environmental Management 92 (1), 172–177.

Hopmans, P., Bren, L., 2007. Long-term changes in water quality and solute exports in headwater streams of intensively managed radiata pine and natural eucalypt forest catchments in south-eastern Australia. Forest Ecology and Management 253 (1-3), 244–261.

Jasiewicz, J., Metz, M., 2011. A new grass gis toolkit for Hortonian analysis of drainage networks. Computers & Geosciences.

Johnes, P., 2007. Uncertainties in annual riverine phosphorus load estimation: impact of load estimation methodology, sampling frequency, baseflow index and catchment population density. Journal of Hydrology 332 (1-2), 241–258.

Jones, A., Stevens, D., Horsburgh, J., Mesner, N., 2011. Surrogate measures for providing high frequency estimates of total suspended solids and total phosphorus concentrations1. JAWRA Journal of the American Water Resources Association 47 (2), 239–253.

Kristiana, R., Vilhena, L., Begg, G., Antenucci, J., Imberger, J., 2011. The management of lake burragorang in a changing climate: The application of the index of sustainable functionality. Lake and Reservoir Management 27 (1), 70–86.

Kronvang, B., Bruhn, A., 1996. Choice of sampling strategy and estimation method for calculating nitrogen and phosphorus transport in small lowland streams. Hydrological Processes 10 (11), 1483–1501.

Lark, R., Cullis, B., 2004. Model-based analysis using REML for inference from systematically sampled data on soil. European Journal of Soil Science 55 (4), 799–813.

Lawrence, I., Lin, K., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics, 255–268.

Marchant, B., Lark, R., 2007. Optimized sample schemes for geostatistical surveys. Mathematical geology 39 (1), 113–134.

Marchant, B., Saby, N., Lark, R., Bellamy, P., Jolivet, C., Arrouays, D., 2010. Robust analysis of soil properties at the national scale: cadmium content of french soils. European Journal of Soil Science 61 (1), 144–152.

Marsh, N., Waters, D., 2009. Comparison of load estimation methods and their associated error. In: 18th IMACS World Congress MODSIM. Vol. 4. Cairns.

Quilbé, R., Rousseau, A., Duchemin, M., Poulin, A., Gangbazo, G., Villeneuve, J., 2006. Selecting a calculation method to estimate sediment and nutrient loads in streams: Application to the Beaurivage River (Quebec, Canada). Journal of Hydrology 326 (1), 295–310.

Ribeiro Jr, P., Diggle, P., 2001. geoR: A package for geostatistical analysis. R news 1 (2), 14–18.

Thomas, R., 1985. Estimating total suspended sediment yield with probability sampling. Water Resources Research 21 (9), 1381–1388.

Thomas, R., 1988. Monitoring baseline suspended sediment in forested basins: the effects of sampling. Hydrological sciences journal 33 (5), 499–514.

Wang, Y., Kuhnert, P., Henderson, B., 2011. Load estimation with uncertainties from opportunistic sampling data–a semiparametric approach. Journal of Hydrology 396 (1), 148–157.

Chapter 7

General discussion and future research

The theme of this thesis has been to improve water quality monitoring with an emphasis on improving the applied sampling scheme and the statistical analysis methods used for reporting the state of the catchment.

7.1 Water quality sampling

7.1.1 Event-based sampling

Monthly routine sampling is the most commonly used sample frequency by catchment managers. However, this form of sampling often misses rainfall events. The sampling of these events is necessary as large proportions of sediments and nutrients are exported during these periods. Therefore catchment managers are increasingly installing automatic sampling equipment which can sample these events. Currently, the majority of studies which have examined the effect of sampling schemes on the accuracy of load estimates require access to continuously sampled data. The collection of this data is extremely expensive, and therefore these studies have only been carried out in a limited amount of catchments. As many catchment managers have access to historical data, statistical methods are required to gain information from this data (Burt et al., 2011).

Chapter 3 presented a simulation based method which used a linear mixed model to describe the relationship between total phosphorus and stream discharge. This simulation method is commonly used in soil science to simulate spatially correlated data (Gebbers and Bruin, 2010). Using this procedure it was possible to examine the effect of including event-based sampling. Using several sites located west of Sydney the results of this analysis found;

- event-based sampling improved the annual load estimates for each site.
- the level of improvement in accuracy was greater for catchments with larger relief and higher annual rainfall.
- the simulation based procedure allows for the comparison of different sampling schemes without the need of continuously monitored water quality data.

7.1.2 Estimating the required sample size

The sample size will directly effect the precision of the mean. As with the evaluation of different sampling schemes, many studies have relied on access to continuous water quality data. In order to improve the efficiency of a monitoring program catchment managers require a method to relate the sample size to the precision with which the mean is known. By characterising this relationship it is then possible to estimate the cost of improving the precision of the mean. Many catchment managers have access to limited historical data, however the sampling schemes are often not probabilistic. Due to this sampling scheme, a model-based approach must be used to estimate the precision of the mean.

Chapter 4 outlines how a variogram can be used to estimate the variance structure of the data. It is necessary to use a variogram because this method does not require equally spaced samples through time. The procedure used in chapter 4 is based on the method presented by Domburg et al. (1994) where the variance structure of the variogram can be used to estimate the precision of the mean using different sample sizes. Using total phosphorus data from 17 sites in south eastern Australia this chapter found;

- variograms provided a suitable method to estimate the temporal auto-correlation of the observations.
- the precision of the mean during events is less than that of base-flow as the variation during events is higher.
- the results indicated there was little improvement from increasing the samples size above 12 samples per year for estimating baseflow means and 12 samples per event or estimating event mean concentrations.

7.1.3 An event-based probabilistic sampling scheme

With the knowledge of the importance of event-based sampling it is vital that the event sampling scheme accurately describes the exports. Many event-based sampling schemes use regular temporal intervals. Systematic sampling schemes for events may introduce bias in the sampling scheme, as this form of sampling assumes the temporal interval is small enough to cover the variation of water quality. In addition, it requires a model-based approach in order to provide unbiased estimates which complicates any statistical analysis. A probabilistic based sampling scheme does not require a model-based approach and provides easily defendable statistics.

Chapter 5 outlines a stratified random sampling scheme for event sampling. A simple but flexible method is presented which can be easily implemented using common automatic samplers. The sampling method stratifies the mean event hydrograph into key hydrological components (e.g. the rising and falling limbs). Random sampling is then applied within each strata. One of the problems with this design is that not every event is identical to the mean event hydrograph. Therefore, post-stratification is used to re-stratify the samples after each event. Using turbidity and EC data the chapter showed;

- probabilistic sampling schemes do not require complex models to provide unbiased estimates.
- probabilistic sampling schemes can be used to provide accurate estimates of the mean value of events for both turbidity and EC.
- re-stratification of the samples allows for the detailed characterisation of individual events.

Optimisation of the samples and strata is also easily achievable without the use of models. For example, if suspended sediments were of interest more sampling and shorter strata could be used during the earlier stages of the event.

7.2 Load estimation

Currently, it is too expensive for catchment managers to sample water quality continuously. Therefore, statistical methods are required to derive information from the limited samples. The use of common estimation techniques; average, ratio and regression methods assume the use of simple random sampling. The use of these methods in combination with systematic sampling schemes may introduce bias in the estimates of the uncertainty of the predictions. In regards to Australian catchments, it is necessary for the estimation method to also provide continuous temporal estimates of concentration in order to relate these estimates to guidelines.

Linear mixed models do not require the use of probabilistic sampling. This method was outlined in chapter 6. In addition, stream discharge and turbidity were examined as explanatory variables for total nitrogen and total phosphorus. Following the research of Kuhnert et al. (2012) and Wang et al. (2011) additional stream discharge derived explanatory variables were examined. To evaluate the accuracy of each model, leave-one-out-cross-validation (LOOCV) was used to evaluate each model. Event cross validation was also performed to assess the accuracy of event predictions for each model. The investigation highlighted that;

- the linear mixed models found temporal auto-correlation for the both total phosphorus and total nitrogen.
- turbidity outperformed stream discharge and stream discharge derived variables for both total nitrogen and total phosphorus.
- using stream discharge derived variables and turbidity in addition to stream discharge can improve model predictions.
- it is important to consider the performance of the models for predicting events.

7.3 Recommendations for catchment managers

This thesis has focused on the use of statistical methods to improve water quality monitoring programs. As many catchment managers do not have access to continuous water quality data, the statistical methods must be able to utilise this limited data. The results of this thesis are of relevance for Australian catchment managers. Emphasis has been given for the design and improvement of sampling schemes and suitable methods for the statistical analysis of these samples.

The design of water quality sampling schemes is important to minimise costs and improve the precision of the estimates. The design of sampling schemes should be an ongoing process which continually incorporates information to improve efficiency. Based on the results of this body of work, it is highly recommended that event-based sampling is used in order to improve water quality predictions. Using the methods of Domburg et al. (1994), this work has found that monthly sampling is adequate in most catchments to monitor base-flow conditions and event-based sampling should collect at least 12 samples per event. It is highly recommended that where possible event-based sampling should be probabilistic. The use of probabilistic sampling for these periods will allow for accurate event mean concentration estimates without the requirement of complicated models. It is recommended that catchment managers should combine monthly sampling with probabilistic event sampling with at least 12 samples per event.

Catchment managers in Australia report on the state of catchments based on the guidelines in the ANZECC. A major limitation of these recommendations is the requirement to report the state of the catchment in regards to concentration values. As catchment managers are required to report on the state of the catchment, and therefore often report based on the percentage of samples which exceed the recommended guidelines. However there are no recommendations on reporting the uncertainty of the estimates. In addition, the highly dynamic nature of water quality it is difficult to determine when these thresholds have been exceeded with limited temporal sampling capabilities. Reporting of the uncertainty of these estimates would help to improve the understanding of the estimates and provide additional information on the dynamic nature of the state of the catchment.

Estimation methods should be used to improve the understanding of the catchments and help with the reporting on the state of the catchment. As routine sampling is sparse and event sampling is dense, it is recommended that estimation methods are used rather than using the percentage of samples above the guidelines. Linear mixed models provide meaningful continuous estimations which account for possible temporal auto-correlation within the data and therefore does not require probabilistic sampling. In addition, this estimation method can take advantage of and improve the precision of estimates using information from low-cost explanatory variables such as turbidity and EC. Another advantage of this method is the unbiased uncertainty estimates which should be used to assess how accurate the estimates are and provide confidence when assessing management strategies.

The final recommendations for Australian catchment managers;

- event-based sampling should be included in monitoring programs. If possible it is highly recommended that probabilistic sampling be used to simplify the analysis.

- combine event-based sampling with routine monthly sampling.
- where possible continuously monitor additional water quality properties with turbidity and EC sensors.
- use linear mixed models to estimate water quality for reporting and evaluating management strategies.

7.4 Future research

This thesis has investigated several statistical methods to improve the information from historical data and improve sampling schemes, however there are several areas which require additional research;

- future research examining water quality sampling must be of relevance to catchment managers and their financial constraints.
- additional methods are required which can gain information from limited historical data. With access to these methods, it will be possible for catchment managers to acquire site specific information to improve the accuracy and efficiency of monitoring programs.
- studies examining sample designs should consider how load estimation methods which can use information from low-cost sensors, can influence the sample requirements.
- as non-normality is a major limitation of estimation methods, there is a requirement for more flexible models (e.g. copulas based methods as presented by Marchant et al. (2011)).

- statistical outliers are commonplace in water quality studies, and therefore more robust methods are required to accommodate these. For example, robust estimation methods presented by Papritz et al. (2011).

7.5 References

Burt, T., Howden, N., Worrall, F., McDonnell, J., 2011. On the value of long-term, low-frequency water quality sampling: avoiding throwing the baby out with the bathwater. Hydrological Processes 25 (5), 828–830.

Domburg, P., De Gruijter, J., Brus, D., 1994. A structured approach to designing soil survey schemes with prediction of sampling error from variograms. Geoderma 62 (1), 151–164.

Gebbers, R., Bruin, S., 2010. Application of geostatistical simulation in precision agriculture. Geostatistical Applications for Precision Agriculture, 269–303.

Kuhnert, P., Henderson, B., Lewis, S., Bainbridge, Z., Wilkinson, S., Brodie, J., 2012. Quantifying total suspended sediment export from the burdekin river catchment using the loads regression estimator tool. Water Resources Research 48 (4), W04533.

Marchant, B., Saby, N., Jolivet, C., Arrouays, D., Lark, R., 2011. Spatial prediction of soil properties with copulas. Geoderma 162 (3), 327–334.

Papritz, A., Künsch, H., Schwierz, C., Stahel, W., 8 2011. Robust estimation of the external drift and the variogram of spatial data. In: 58th World Statistics Congress of the International Statistical Institute. Dublin.

Wang, Y., Kuhnert, P., Henderson, B., 2011. Load estimation with uncertainties from opportunistic sampling data–a semiparametric approach. Journal of Hydrology 396 (1), 148–157.