Journal of Hydrology: Regional Studies 4 (2015) 439–451

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

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

J.S. Lessels[a,b,*], T.F.A. Bishop[b]

[a] Northern Rivers Institute, University of Aberdeen, Aberdeen AB24 3UF, UK
[b] Faculty of Agriculture and Environment, University of Sydney, Sydney, NSW, Australia

ARTICLE INFO

Article history:
Received 18 March 2015
Received in revised form 25 June 2015
Accepted 27 June 2015

Keywords:
Australia
Water quality
Sampling
Linear mixed models
Simulation
Total phosphorous

ABSTRACT

Study region: South eastern Australia.
Study focus: This region is characterised with rainfall events that are associated with large exports of nutrients and sediments. Many water quality monitoring schemes use a form of event-based sampling to quantify these exports. Previous water quality studies that 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 need to develop a method that assesses the importance of sampling events using information from limited historical data. This work presents a simulation based 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 phosphorus.
New hydrological insights for the region: The inclusion of event-based sampling improved annual load estimates of 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 in this region. Using this approach, different sampling schemes can be compared based on limited historical data.

1. 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

* Corresponding author.
E-mail addresses: jason.lessels@abdn.ac.uk (J.S. Lessels), thomas.bishop@sydney.edu.au (T.F.A. Bishop).

http://dx.doi.org/10.1016/j.ejrh.2015.06.020

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 catchment with large events separated by long periods 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 six 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 five 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; Mehdi et al., 2015; Saaltink et al., 2014; Sliva and Williams, 2001). Relationships between phosphorus and the proportion of land under agriculture (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 in combination with event sampling (for examples see Salles et al. (2008), Birkel et al. (2011)). This is one of the most common 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). 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 under these situations 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). Similarly to simple linear models, LMM also provide the ability to incorporate additional covariates (e.g. stream discharge and turbidity) to improve predictions (Lessels and Bishop, 2013)

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 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 in load estimates using event-based sampling.

2. Materials and methods

2.1. 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 Fig. 1. A summary of the topographical and hydrological features of each catchment is provided in Table 1. The number of each monitoring site in Table 1 is indicated in Fig. 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.

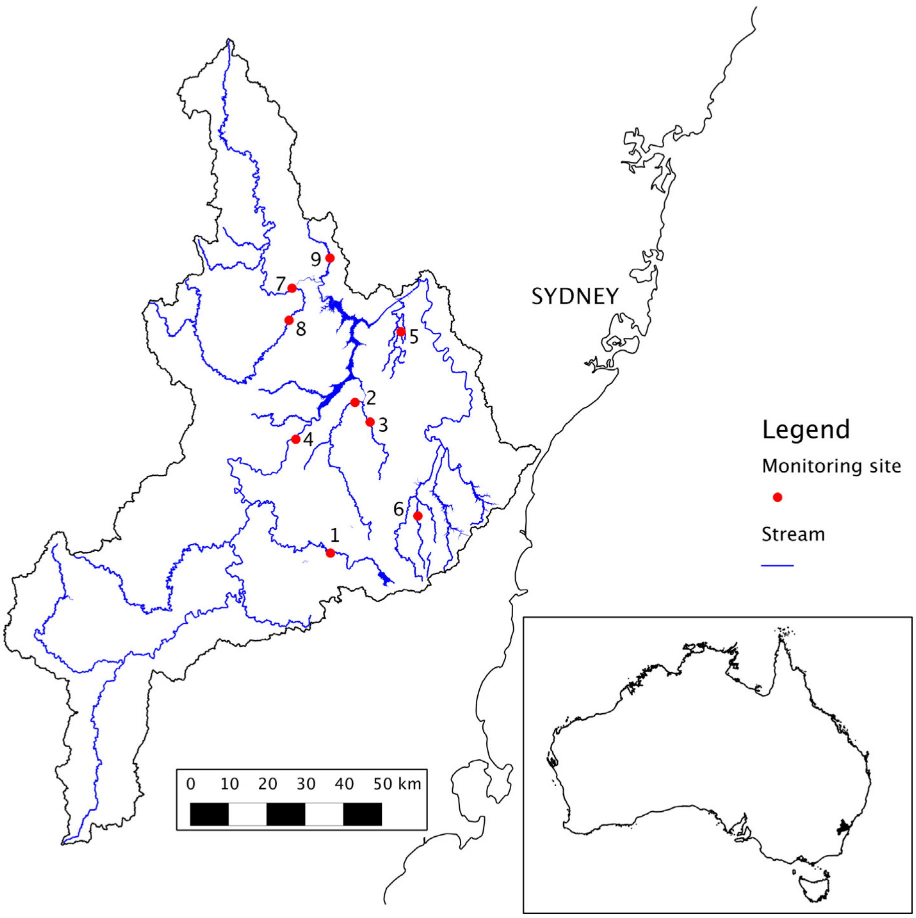

Fig. 1. Map of greater catchment area and location of monitoring sites and the location of the catchment within Australia.

Table 1
Key features of each sub-catchment.

Site	Area (km^2)	Minimum elevation (m)	Maximum elevation (m)	Annual rainfall (mm)	Annual discharge (mm)	Grassland cover (%)	Forest cover (%)	Urban cover (%)	Population density (people km^{-2})
1. Berrima Weir	201.4	620	868	959	272	69	15	10	80
2. Smallwoods Crossing	436.9	105	867	857.2	56	11	86	2	17.6
3. Little River	104	175	632	825	85	2	97	0	8.9
4. Jooriland	4825.1	108	1179	708	42	57	39	1	11.1
5. Werombi	56.4	203	519	695	105	28	68	1	25
6. Burke River	88.3	328	777	1183	195	4	95	1	3.7
7. Kelpie Point	1447	116	1342	887	79	25	70	0	15.2
8. Cedar Ford	719.1	169	1382	932	105	9	80	0	0.8
9. Kedumba Crossing	72.5	152	1061	1264	201	0	85	13	52.9

Table 2
Water quality data summary.

Site	Start date	End date	Mean (mg L^{-1})	Min (mg L^{-1})	Max (mg L^{-1})	SD (mg L^{-1})	Skewness (mg L^{-1})	n
1. Berrima Weir	Dec 1993	Dec 2010	0.08	0.006	0.87	0.09	4.75	392
2. Smallwoods Crossing	Jan 1991	Mar 2010	0.06	0.006	0.70	0.09	4.21	323
3. Little River	Aug 1991	Apr 2007	0.02	0.006	0.46	0.05	6.35	153
4. Jooriland	Feb 1994	Dec 2010	0.06	0.006	0.45	0.07	2.38	527
5. Werombi	Jan 1991	Dec 2010	0.04	0.006	0.73	0.05	6.99	535
6. Burke River	Oct 1991	Dec 2010	0.02	0.006	0.17	0.01	4.94	454
7. Kelpie Point	Jan 1991	Dec 2010	0.03	0.006	0.92	0.07	7.50	392
8. Cedar Ford	Jan 1991	Dec 2010	0.04	0.006	0.87	0.10	5.33	553
9. Kedumba Crossing	Jan 2002	Dec 2010	0.04	0.006	0.67	0.06	5.52	241

2.2. Data description

Table 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 Eaton et al. (1995). The minimum total phosphorus included in this analysis for each site was restricted 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.

2.3. Statistical analysis

The simulation procedure used to compare the two sampling schemes is provided as a flowchart in Fig. 2. The procedure comprises 5 steps which are described below.

2.3.1. Fitting a linear mixed model

In this work a LMM is used to model the relationship between TP and stream discharge. In addition, the LMM accounts for the temporal auto-correlation between samples. The LMM used here has the form:

$$z(\mathbf{t}) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta}, \tag{1}$$

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),

$$z(\mathbf{t})^* = \begin{cases} \log z(\mathbf{t}) & \text{if } \lambda = 0 \\ \dfrac{z(\mathbf{t})^{\lambda} - 1}{\lambda} & \text{otherwise.} \end{cases} \tag{2}$$

The LMM fitted in this work uses restricted maximum likelihood (REML) to fit the model using the geoR package (Ribeiro 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

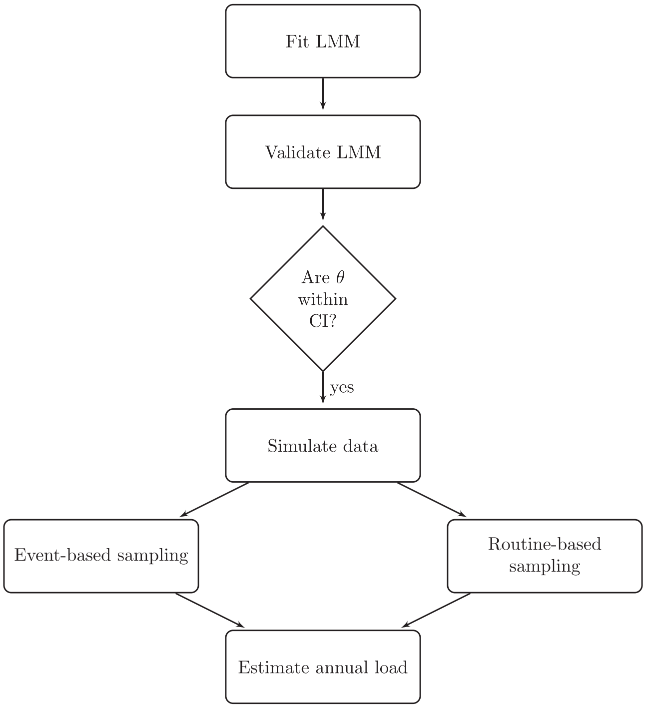

Fig. 2. Flow chart of the simulation process for each site.

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;

$$\mathbf{V}_{i,j} = \begin{cases} \sigma^2 s \exp\left(-\dfrac{\left|\mathrm{x_i} - \mathrm{x_j}\right|}{a}\right), & i \neq j \\ \sigma^2, & i = j, \end{cases} \tag{3}$$

where the variance σ^2 is the diagonal and $\left|\mathrm{x_i} - \mathrm{x_j}\right|$ 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;

$$s = \frac{c}{c_0 + c}, \tag{4}$$

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.

2.3.2. 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) adequately represents 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;

$$\theta_i = \frac{\{c(\mathbf{t}_i)^* - \tilde{C}^*_{(-i)}\}^2}{\sigma^2_{(-i)}}, \tag{5}$$

where $c(\mathbf{t}_i)^*$ is the ith 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 are an accurate representation of the process, $\boldsymbol{\theta}$ has a mean $\bar{\boldsymbol{\theta}} = 1$ and median $\breve{\boldsymbol{\theta}} = 0.455$ (Marchant et al., 2010). Based on the methods of Marchant et al. (2010) it is possible to estimate 95% confidence intervals for the median and mean $\boldsymbol{\theta}$. If the $\boldsymbol{\theta}$ statistics are within the confidence interval, it is assumed the models are valid.

2.3.3. 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. (1)). The simulations for each site are determined using;

$$r(\mathbf{t})_j = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{s}_j, \tag{6}$$

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.

2.3.4. 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.

2.3.5. Load estimation

As the two sampling schemes are non-probabilistic, LMM are used to estimate TP using a log linear relationship with stream discharge, which is equivalent to a Box-Cox transformation where $\lambda = 0$. 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 (Eddelbuettel and Sanderson, 2014) 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 simulated and predicted annual load of each sampling method.

2.4. *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 1. The correlation between the mean difference of the RMSE and the catchment characteristics is examined. In addition, backwards elimination of each linear model based on the AIC is used to find the most parsimonious model to explain the variation in the difference in RMSE, between catchments.

3. Results

3.1. *LMM used for simulation*

Table 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 seven days for seven of the nine 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.

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. The cross validation indicated that smaller TP observations are overestimated by the LMM and there is some underestimation of TP at higher concentrations. Using the estimated LMMs of each site, LOOCV was performed and is summarised in Table 4. The table reports the estimated

Table 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)
1. Berrima Weir	−6.45	0.26*	30.00	0.92	2.45	−0.30
2. Smallwoods Crossing	−14.59	1.07*	5.56	5.33	14.47	−0.50
3. Little River	−71.82	5.65*	19.44	391.54	629.82	−0.86
4. Jooriland	−7.44	0.34*	2.96	0.60	2.16	−0.17
5. Werombi	−12.03	0.89*	2.33	1.49	6.27	−0.35
6. Burke River	−44.76	1.69*	2.25	31.71	166.71	−0.73
7. Kelpie Point	−28.6	1.96*	2.25	4.92	31.95	−0.54
8. Cedar Ford	−25.06	1.3*	4.20	3.52	45.54	−0.54
9. Kedumba Crossing	−15.08	1.01*	2.33	3.62	19.14	−0.44

Table 4
Cross validation statistics of LMMs.

Site	$\bar{\theta}$	$\hat{\theta}$	Confidence interval $\bar{\theta}_{95\%}$	Confidence interval $\hat{\theta}_{95\%}$	Correlation (Lin's)
1. Berrima Weir	1.01	0.25	(0.84–1.15)	(0.36–0.57)	0.60
2. Smallwoods Crossing	0.91	0.36	(0.86–1.17)	(0.35–0.58)	0.70
3. Little River	0.97	0.55	(0.78–1.23)	(0.3–0.64)	0.49
4. Jooriland	0.95	0.25	(0.8–1.21)	(0.31–0.63)	0.76
5. Werombi	0.99	0.30	(0.88–1.14)	(0.37–0.56)	0.77
6. Burke River	1.03	0.40	(0.87–1.14)	(0.36–0.56)	0.60
7. Kelpie Point	1.02	0.44	(0.86–1.16)	(0.36–0.58)	0.68
8.Cedar Ford	0.99	0.33	(0.88–1.14)	(0.37–0.56)	0.76
9. Kedumba Crossing	1.02	0.46	(0.86–1.14)	(0.36–0.56)	0.63

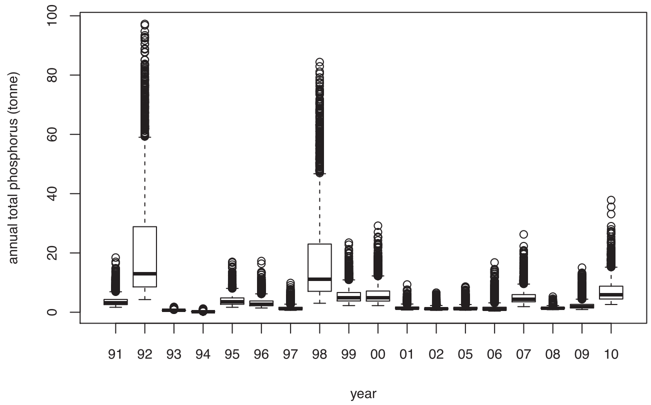

Fig. 3. Simulated annual load for the Kelpie Point catchment.

mean ($\bar{\theta}$)and median ($\hat{\theta}$) (Eq. (5)), and the estimated 95% confidence interval for both $\bar{\theta}$ and $\hat{\theta}$. In addition, Lin's correlation coefficient (Lawrence and Lin, 1989) between the simulated 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 confidence interval. Based on the $\hat{\theta}$ of the LOOCV only five sites fell within the estimated confidence interval. 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 simulated error. Therefore, simulations of TP were only conducted at the following 5 sites:

2. Smallwoods Crossing,
3. Little River,
6. Burke River,
7. Kelpie Point,
9. Kedumba Crossing.

3.2. 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. Fig. 3 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.

3.3. Estimating annual load

Table 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 three and four days for four of the five sites and less than four 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 ten times as large for the Kedumba Crossing catchment.

An important aspect of the LMMs is the modelling of the temporal auto-correlation. Table 6 summarises the mean temporal range of the models fitted to the two sampling schemes. The table also indicates the percentage of realisations

Table 5
Sample scheme characteristics of each site.

Site	Trigger level (mm day^{-1})	Event length (days)	Event samples	Routine samples
2. Smallwoods Crossing	0.005	3.39	521	213
3. Little River	0.23	3.31	736	181
6. Burke River	0.91	2.64	994	233
7. Kelpie Point	0.54	3.89	784	237
9. Kedumba Crossing	1.06	1.48	1050	102

Table 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 (%)
2. Smallwoods Crossing	2.50	90.30	112.62	5.35
3. Little River	2.08	52.05	72.67	18.75
6. Burke River	1.43	99.65	126.93	5.75
7. Kelpie Point	1.52	98.75	116.36	5.00
9. Kedumba Crossing	1.76	100.00	116.68	4.50

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 three 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.

Fig. 4 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. 4a) shows how the variance of the predicted annual load increases as the simulated annual load increases. Fig. 4b shows the routine based annual load estimates. The routine based estimates underestimate annual loads by 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. For each realisation if the difference between the RMSE is positive, the event-based estimate is deemed more accurate than the routine estimate. Table 7 provides a summary of the differences between the RMSE of the catchments. The table also includes the improvement of event-based sampling which is the percentage of realisations where the inclusion of event-based sampling improved the accuracy of load estimates. The improvement of event-based sampling of all sites, except the Little River catchment are greater less than 90% and 79.1% for the Little River catchment. This indicates that the inclusion of event-based sampling increases the accuracy of the 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.

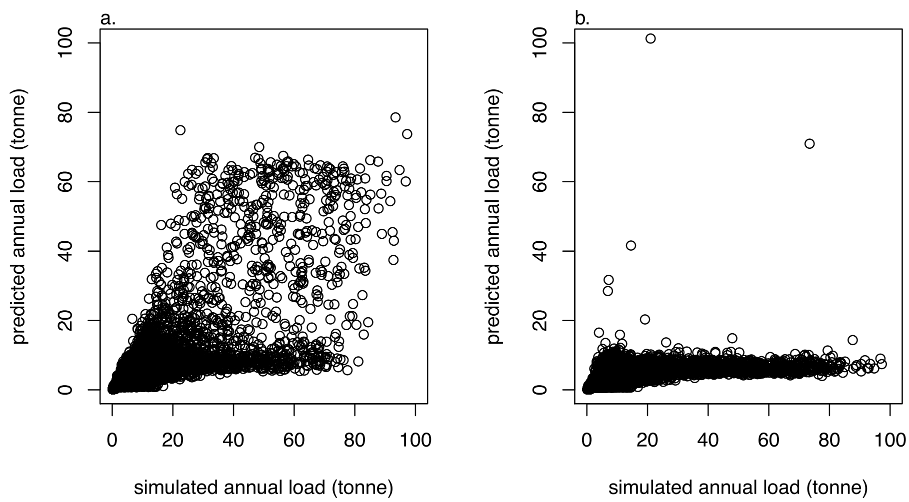

Fig. 4. Event-based (a) and routine (b) annual load estimates for the Kelpie Point catchment.

Table 7
Summary of differences between routine and event.

Site	min RMSE difference (tonne)	max RMSE difference (tonne)	mean RMSE difference (tonne)	Improvement of event-based sampling (%)
2. Smallwoods Crossing	−1.77	5.94	0.55	91.6
3. Little River	−0.74	1.54	0.11	79.1
6. Burke River	−0.12	0.39	0.05	93.1
7. Kelpie Point	−10.42	16.11	1.93	94.4
9. Kedumba Crossing	−0.14	0.74	0.17	96.65

Table 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

3.4. Relating event-based sampling improvement with catchment characteristics

Finding relationships between the mean RMSE difference and catchment characteristics is important for the improvement of sampling schemes at sites without historical data. The correlation between catchment characteristics and the mean RMSE difference at each site is provided in Table 8. Annual rainfall, elevation range, urban cover and population density are the most correlated with the mean RMSE difference. All catchment characteristics except elevation range are negatively correlated with this. In addition the percentage of urban land cover and population density had similar correlation values of −0.51 and −0.5 respectively. The Backwards elimination of the catchment characteristics resulted in a final linear model including annual rainfall and elevation range. Fig. 5 shows the predicted values against the mean RMSE difference. 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 mean RMSE difference and catchment characteristics. The results of the backwards elimination indicate that annual rainfall and elevation range are linearly related to the improvement of load estimates with the inclusion of event based samples. Therefore, the sampling of high flow events has a greater effect of reducing load estimate uncertainties in catchments with higher annual rainfall and elevation ranges. 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.

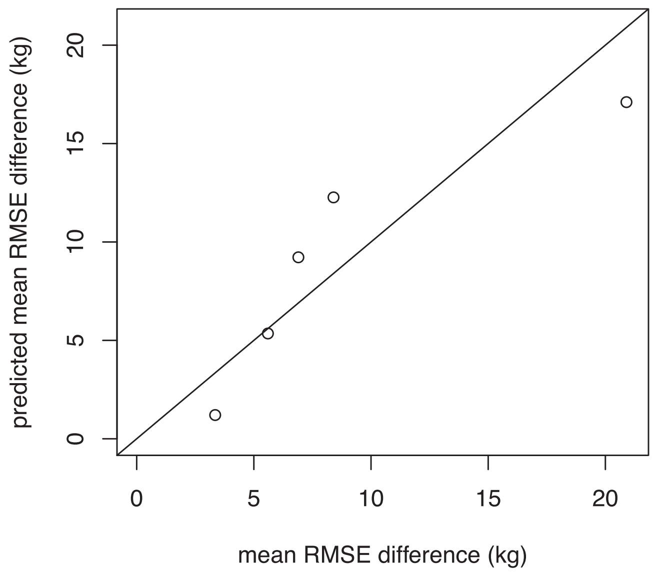

Fig. 5. Relationship between mean RMSE difference and predicted mean RMSE difference.

4. Discussion

It is important to consider the temporal auto-correlation of water quality observations (Kuhnert et al., 2012; 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 nine catchments it was only possible to fit a valid statistical model for five 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 water quality. 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 water quality data. For example, it should be possible for this procedure to be applied to total nitrogen which has a similar overall trend with TP. 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 low-cost covariates such as electrical conductivity and turbidity.

Linking the mean RMSE difference to the catchment characteristics is important to identify the controlling hydrological processes (Todd et al., 2007). 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 mean RMSE difference 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.

5. 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. The study revealed that there was a relationship between the increased currency with event-based sampling and the catchment characteristics. However, this was limited to static land-uses, and future studies should consider the potential affect of temporal land-use changes. Methods such as this are beneficial in situations with limited data, but should not be substituted for high frequency sampling where it is feasible.

Conflicts of interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgements

The authors would like to thank Dr. Rob Mann and the Sydney Catchment Authority for providing the data and information used in this work.

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. J. Hydrol. 313 (3), 234–247.
Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19 (6), 716–723.
Australian Bureau of Statistics, 2011. Estimated Resident Populations for New South Wales, 2011. Catalogue No. 3218.0. NSW State Office, Sydney.
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. Environ. Manag. 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. Mar. Pollut. Bull. 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. Hydrol. Process. 26 (2), 308–316.
Bivand, R., Pebesma, E., Gómez-Rubio, V., 2008. Applied Spatial Data Analysis with R (Use R). Springer Science, New York, NY.
Box, G., Cox, D., 1964. An analysis of transformations. J. R. Stat. Soc. Ser. B: (Methodol.), 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. Hydrol. Process. 25 (5), 828–830.
Cohn, T., 2005. Estimating contaminant loads in rivers: an application of adjusted maximum likelihood to type 1 censored data. Water Resour. Res. 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 Resour. Res. 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. J. Hydrol. 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, Berlin, Heidelberg.
Drewry, J., Newham, L., Croke, B., 2009. Suspended sediment, nitrogen and phosphorus concentrations and exports during storm-events to the Tuross Estuary, Australia. J. Environ. Manag. 90 (2), 879–887.
Eaton, A.D., Clesceri, L.S., Greenberg, A.E., 1995. Standard Methods for the Examination of Water and Wastewater, 19th ed. United Book Press, Inc., Baltimore, Maryland.
Eddelbuettel, D., Sanderson, C., 2014. RcppArmadillo: Accelerating R with high-performance C++ linear algebra. Comput. Stat. Data Anal. 71, 1054–1063.
Gao, P., 2008. Understanding watershed suspended sediment transport. Prog. Phys. Geogr. 32 (3), 243–263.
Gebbers, R., Bruin, S., 2010. Application of geostatistical simulation in precision agriculture. Geostat. Appl. Precis. Agric., 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. For. Ecol. Manag. 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. J. Hydrol. 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 J. Am. Water Resour. Assoc. 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 Environ. J. 27 (1), 79–85.
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 Reserv. Manag. 27 (1), 70–86.
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 Resour. Res. 48 (4), W04533.
Lark, R., Cullis, B., 2004. Model-based analysis using REML for inference from systematically sampled data on soil. Eur. J. Soil Sci. 55 (4), 799–813.
Lawrence, I., Lin, K., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics, 255–268.
Lessels, J.S., Bishop, T.F.A., 2013. Estimating water quality using linear mixed models with stream discharge and turbidity. J. Hydrol. 498, 13–22.
Lewis, J., 1996. Turbidity-controlled suspended sediment sampling for runoff-event load estimation. Water Resour. Res. 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. Eur. J. Soil Sci. 61 (1), 144–152.
Marsh, N., Waters, D., 2009. Comparison of load estimation methods and their associated error. In: Anderssen, R.S., Braddock, R.D., Newham, L.T.H. (Eds.), 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation. , pp. 3322–3328.
Mehdi, B., Ludwig, R., Lehner, B., 2015. Evaluating the impacts of climate change and crop land use change on streamflow, nitrates and phosphorus: a modeling study in Bavaria. J. Hydrol. Reg. Stud. 4, B, 60–90.
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, p. 3025.
Pebesma, E., 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. 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.
Saaltink, R., van der Velde, Y., Dekker, S.C., Lyon, S.W., Dahlke, H.E., 2014. Societal, land cover and climatic controls on river nutrient flows into the Baltic Sea. J. Hydrol. Reg. Stud. 1, 44–56.
Salles, C., Tournoud, M., Chu, Y., 2008. Estimating nutrient and sediment flood loads in a small mediterranean river. Hydrol. Process. 22 (2), 242–253.
Sliva, L., Williams, D., 2001. Buffer zone versus whole catchment approaches to studying land use impact on river water quality. Water Res. 35 (14), 3462–3472.
Thomas, R., 1985. Estimating total suspended sediment yield with probability sampling. Water Resour. Res. 21 (9), 1381–1388.
Thomas, R., 1988. Monitoring baseline suspended sediment in forested basins: the effects of sampling. Hydrol. Sci. J. 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 Resour. Res. 29 (4), 1247–1256.
Thomas, R., Lewis, J., 1995. An evaluation of flow-stratified sampling for estimating suspended sediment loads. J. Hydrol. 170 (1-4), 27–45.
Todd, C.E.D., Goss, A.M., Tripathy, D., Harbor, J.M., 2007. The effects of landscape transformation in a changing climate on local water resources. Phys. Geogr. 28 (1), 21–36.
Wang, Y., Kuhnert, P., Henderson, B., 2011. Load estimation with uncertainties from opportunistic sampling data – a semiparametric approach. J. Hydrol. 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. Sci. Total Environ. 251, 45–54.