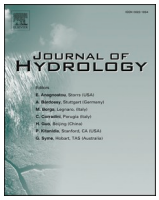
Journal of Hydrology 651 (2025) 132550

Contents lists available at ScienceDirect

Journal of Hydrology

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



A Bayesian hierarchical approach to combine interpolation and extrapolation methods for fluvial flux calculation

Fred Worrall [a,*], Nicholas J.K. Howden [b,c], Tim P. Burt [c,d,e]

[a] Dept of Earth Sciences, Science Laboratories, South Road, Durham DH1 3LE, UK
[b] School of Civil, Aerospace and Design Engineering (CADE), University of Bristol, Queens Building, Bristol, UK
[c] Cabot Institute, Royal Fort House, Bristol BS8 1UH, UK
[d] Dept of Geography, Science Laboratories, South Road, Durham DH1 3LE, UK
[e] School of Geographical Sciences, University Road, Bristol BS8 1SS, UK

ARTICLE INFO

This manuscript was handled by N. Basu, Editor-in-Chief, with the assistance of Magdalena Bieroza, Associate Editor

Keywords:
Nitrate
Suspended solids
Rivers
Flux

ABSTRACT

Calculation of the flux of material through catchments of all scales is a key measure used to understand how our environment functions and how it is changing. Methods of fluvial flux calculation have tended to be either based upon extrapolation against a well-known driving variable, typically river discharge; or interpolation where expected values (e.g., arithmetic mean) are used. The choice of method has been a matter of personal choice and perception of strength of controls on the determinand of interest. In this study we propose that this dichotomy of methods is unnecessary and show how a Bayesian hierarchical method can combine approaches. Our method uses factors of month, year and monitoring site to modify the concentration-discharge (C-Q) relationship for the determinand of interest, i.e. the method uses all the available data to define the C-Q relationship at each site for each level of the included factors. If the C-Q relationship is not significant, then the method returns the expected value for the particular month, year and monitoring site combination. The method was applied to 161 catchments across England for both suspended solids (SS) and nitrate ($N\text{-}NO_3$) concentrations, and results compared to existing extrapolation methods. The method developed here demonstrates: i) Greater sensitivity in that it was better able to identify significant C-Q relationships. ii) The ability to detect smaller significant slopes for C-Q relationships and with improved precision by at least 76%. iii) Scaling results to the national level implies that existing extrapolation methods would overestimate national fluxes.

The proposed method uses no more data than would be used by either an extrapolation method or an interpolation method but draws upon the information from all data to strengthen the analysis.

1. Introduction

The rate at which chemicals and sediment are exported from river catchments has been estimated using a variety of techniques, and there has been much discussion about which are the most appropriate techniques to use and where to apply them to obtain the maximum understanding (Walling & Webb, 1985, 1988; Littlewood, 1995; Johnes, 2007; Worrall et al., 2012, 2013a). Because concentration data are usually only sparsely sampled due to practical or economical constraints and continuous records of river discharge data are readily available, there is a challenge for determining the optimal method of calculating fluvial flux as sparse samples do not capture the full variation of the concentration time series.

There are two aspects of flux calculation techniques that should be considered when calculating fluvial flux. Firstly, the precision of the approach indicates the dispersion of the load estimations around a central value; secondly, the accuracy of the method depicts its ability to estimate the given value – the systematic bias. Rather than bias or accuracy, precision is described in many studies. Littlewood et al. (1998) provide an example of this; they were only able to track precision with varying sample frequencies using "indicative" curves, and they were unable to address the accuracy of their methods due to the lack of a "true" number. Johnes (2007) examined 17 catchments with daily phosphorus measurements, but no sub-daily data, and so assumed that the true value was determined by "method 5″ (Littlewood, 1995). Where "method 5" is based on sum of the product of the measured

* Corresponding author.
E-mail address: Fred.Worrall@durham.ac.uk (F. Worrall).

https://doi.org/10.1016/j.jhydrol.2024.132550
Received 22 June 2024; Received in revised form 28 October 2024; Accepted 23 November 2024
Available online 16 December 2024

concentration of the determinand and the river discharge at which it was collected, corrected for ratio of the total discharge to the total discharge at the time of sampling:

$$F = K\left[\frac{\sum_{i=1}^{n} C_i Q_i \widehat{Q}}{\sum_{i=1}^{n} Q_i}\right] \quad (1)$$

Where: F = flux (tonnes/yr); C_i = concentration of the determinanded of interest the sampling site at time i (mg/l); Q_i = the river discharge at the sampling pojnt at time i (m^3/s); $\overline{Q}$ = total discharge (m^3/s); and K = unit conversion constant.

In several studies (Johnes, 2007; Cassidy & Jordan, 2011; Moatar et al., 2012; Worrall et al., 2013a), a "true" flux series is calculated using the highest resolution data available. Successive series are then calculated using sub-samples of the concentration series, and the resulting (sub-sampled) flux series is compared to the true flux series to identify the uncertainty introduced due to sub-sampling (Johnes, 2007; Cassidy & Jordan, 2011; Moatar et al., 2012; Worrall et al., 2013a). The magnitude of such uncertainties has been investigated frequently, usually in locations with near-continuous discharge and concentration time series. These methods have yielded enhanced flux estimation techniques (Verma et al., 2012) as well as suggested methods to lower uncertainty for different flux computations involving sub-sampled concentration series (e.g., Quilbé et al., 2006). The inability to compare to a "true" value restricts the ability to evaluate the accuracy of a method and change in accuracy with changing sampling frequency. Using high-frequency data, Cassidy and Jordan (2011) examined bias and precision in their methodology and demonstrated bias increasing with decreasing sampling frequency, reaching up to 60 % on monthly sampling. The "method 5" has been frequently cited as the best interpolation technique (e.g. Johnes, 2007); however, the method has a clear bias because it adjusts apparently simpler methods by presuming that river discharge follows a normal distribution. Because river flow is not normally distributed, fluxes at high sampling frequencies (f > 1 sample per 7 days) were overestimated by "method 5". The technique overestimated the expected value of the determinand, which led to this overestimation of the fluvial flux. Alternatively, Worrall et al. (2013a) considered a three-year long record of high-frequency (f = 1 per hour) DOC concentration measurements in a river. By taking into consideration a range of extrapolation and interpolation methods and the sources of variation that occur when two variables are multiplied (Goodman, 1960; Howden et al., 2018), they showed that the best method was "method 2" (Littlewood, 1995). "Method 2" had both high accuracy (−2% at f = 1 per month) and very high precision (±8% for f = 1 per month). "Method 2" is based upon knowing the total annual water yield and the expected value of the determinand:

$$F = KE(C)Q_{total} \quad (2)$$

where: E(C) = the expected value of the determinand of interest over the period of study; Q_{total} = the total river discharge over the period of study; and K = a unit conversion constant.

The LOADEST approach (Runkel et al., 2004) uses a linear regression approach to predict the concentration, while they recommend that a single explanatory variable (log streamflow) was sufficient for prediction of suspended-sediment (Crawford, 1991); for nutrients it was recommended to include variables based upon timing −such as month of the year (Cohn et al., 1992). Kim et al. (2018) used upto nine variables including terms in sine, cosine, powers and the river discharge expressed as the percentile. So LOADEST is a more advanced extrapolation method building upon C-Q relationships. In its most common approach, a power-law relationship is used to explain concentration variations as function of flow. Decreasing concentrations with increasing flow are typically referred to as "dilution" behaviour and are often considered indicative of solutes derived from deeper groundwater storage and/or solutes that are source limited (e.g., Zhi and Li, 2020, Stewart et al., 2022). Conversely, patterns of increasing concentration with increasing flow are termed "mobilisation" or "flushing" behaviour, and the respective solutes are often assumed to be located in the near-surface parts of the landscape. Finally, a "chemostatic" pattern is identified if concentrations vary very little with (or at least not as function of) flow (Godsey et al., 2009).

Proposed flux computation techniques generally fall into two categories: interpolation and extrapolation. When it came to suspended sediment flux estimation, Webb et al. (1997) looked at two extrapolation and five interpolation strategies. They found that extrapolation techniques generated the least biased results, and that bias increased with decreasing sample frequency. As an alternative to routine sampling, Kronvang and Bruhn (1996) suggested gathering samples "hydrologically," and a number of studies (e.g. Cooper and Watts, 2002; Skarbøvik et al., 2012) have suggested combining routine sampling with flood samples. While these sampling methods will augment derived C-Q relationships and so aid any extrapolation method, such sampling approaches cannot be included when assessing long-term data because they are not practical when considering region-wide and multi-catchment systems due to their high cost and low acceptance rate.

Worrall et al. (2020) proposed a different approach to maximise information from available often sparse data: Bayesian generalised linear modelling. Bayesian modelling approaches, including hierarchical approaches, have been successfully developed and applied elsewhere for environmental management, for example: for detecting dissolved CH_4 (Wilson et al., 2020); ecological risk assessment (McDonald et al., 2015); stream water quality (Wan et al., 2014); and impacts of climate change (Hobbs, 1997). Because the best low-biased method ("method 2" – Worrall et al., 2013a) only requires the estimation of the annual expected value of the determinand of interest, then this can be predicted by statistical modelling of the available data. This Bayesian approach can use all data without the need for censoring based upon sampling frequency, as an expected value could be calculated for each monitoring site for each year. Bayesian methods are probabilistic and so uncertainty is quantified throughout, including uncertainty resulting from analytical methodologies. Further, generalised linear modelling allows for the inclusion of factors, meaning changes between sampling sites or time of sampling can be included to improve estimation. However, the approach of Worrall et al. (2020) was an interpolation method and so what would happen for determinands whose flux is more dependent on the river discharge, i.e. those determinands that would more naturally suit an extrapolation approach? Further, the difference between interpolation and extrapolation is a false dichotomy as some catchments, years or seasons may be more or less controlled by river discharge. Therefore, this study proposes an approach that is flexible to the nature of the relationship with river discharge, and a method that draws upon both interpolation and extrapolation techniques.

2. Approach and methodology

The approach considered was to use a Bayesian hierarchical approach to estimate the rating curve for each site, year and month. The method was applied to two determinands that were considered *a priori* to have contrasting relationships with river discharge: suspended solids (SS) and nitrate N-NO_3: SS has been assumed to have a closer relationship with river discharge than N-NO_3.

The Bayesian approach allows the use of factorial information, so the rating curve can be adapted for changes over a season or longer-term trends over years. Factorial information can also be included in a linear mixed-effects model; however, the Bayesian approach means that information and predictive power is drawn from the whole dataset (from multiple sites) rather than restricted to the data available for an individual site. The rating curve can be simply expressed as:

$$log_e(C_{xt}) = \alpha + \beta log_e(Q_{xt}) \quad (3)$$

where: C_{xt} = the concentration at site x at time t; Q_{xt} = the river discharge at site x and at time t; and a and b are constants.

For interpolation methods, Worrall et al. (2013) showed the lowest bias method for calculation of an annual flux (Equation ii).

Alternatively, the Bayesian formulation is:

$$log_e(C_{xt}) = N\left(\alpha(Site, Year, Month) + \beta(Site, Year, Month)\left(\log_e \Delta[Q]_{xt}\right), \frac{1}{\sigma^2}\right) \quad (4)$$

$$\Delta[Q]_{xt} = [Q]_{xt} - \overline{[Q]} \quad (5)$$

where: Site, Year and Month are factors representing the different monitoring sites and the month and year in which data were available. In this way α and β are calculated for each monitoring site, for each year and month across the period of available data. Further, the flow covariate is expressed as the difference between the river discharge at site x and time t from the mean flow for that site. By centring the flow data, values of α_{xt} are not the y-intercept, i.e. the value of C at Q = 0, rather α_{xt} is the value of C at the average river discharge. An approach using $\Delta[Q]_{xt}$ has two advantages. Firstly, it makes estimation of α more precise as α now sits in the middle of the observations rather than at one extreme as would be the case if α_{xt} was the y-intercept. Secondly, it means that if there is no significant relationship between C and Q then the approach returns the expected value of C for that site, year and month combination, i.e. an extrapolation method becomes an interpolation method.

2.1. Bayesian modelling

Markov Chain Monte Carlo (MCMC) simulation was used to estimate the posterior distribution of the determinands. All Bayesian analysis was implemented using Jags code called from R using the R2Jags library (example R and JAGS code are included in supplementary material). Models were considered with and without the three readily available factors – Site, Month and Year. The length of the MCMC chain was 10,000 iterations after 2,000 burn-in cycles with samples saved every 10 cycles and with 3 chains. Model fit was tested using a number of approaches. Firstly, the adequacy of the MCMC process was assessed using the $\widehat{R}$, the convergence statistic: values of $\widehat{R} < 1.1$ were considered acceptable. If $\widehat{R} > 1.1$, then the burn-in process and number of iterations were increased. Secondly, the 95 % credible interval for any factor does not include zero and this is henceforward referred to as that factor being significantly different from zero at a probability of 95 %. For ease of understanding and for comparison with other methods, the credible interval is henceforward referred to as the confidence interval. Thirdly, inclusion of the factor or interaction caused the total model deviance to decrease. Fourthly, the inclusion of an additional factor or interaction decreased the deviance information criterion (DIC). It is generally true that inclusion of factors will decrease the total deviance of a model as the inclusion means greater degrees of freedom for fitting and so the DIC accounts for the inclusion of more fitting parameters against the additional fit of the model. Fifthly, the effective number of parameters (pD) was monitored. As a factor was added to the model, then the number of effective parameters would be expected to increase. If, as a factor is added, the pD does not increase then that parameter is having no effect and can be removed from the model. Furthermore, if all parameters are contributing, then pD should be close to the ideal case; thus, the calculated pD can be expressed as a percentage of the ideal case – this value can never be greater than 100 %. Finally, the fit of any model was judged using a posterior prediction check, i.e. the output of the model was plotted against the observed values and the fitted line between these two examined; it would be expected that a good fit model would give a 1:1 line between observed and posterior predicted values; and that line would have a gradient = 1.

For all models, the prior distribution for values of β were set as normal distributions with a mean of zero so that both positive and negative trends over time were equally favoured at the outset. The prior distributions for the values of α were also set as a normal distribution but with the mean set to be the mean of all the data (from all sites) and a standard deviation chosen to make negative values unlikely. The choice of such a distribution is justified because, if values of β are small, then α is the prediction of a particular monitoring site's expected value and thus should approximate the expected value of the distribution of the data as a whole. A half-t distribution was used for the prior distributions of the standard deviations for all terms as half-t distributions mean that a negative value of the standard deviation cannot occur.

2.2. Extrapolation method

To demonstrate the benefit of the Bayesian approach, this study also applied the extrapolation method to the dataset. The best-fit log–log CQ relationship (Equation iii) was calculated for each site where there was SS or N-NO_3 data and this relationship was then applied to the river discharge record for the site to calculate the annual flux for each site for each year for which there was a flow record. The uncertainty in the estimated best-fit C-Q relationship for each site was used to estimate the uncertainty in the annual flux estimate for the site.

2.3. Concentration data

The study used data for the Harmonised Monitoring Scheme sites within England (HMS – Bellamy and Wilkinson, 2001). There are 161 HMS sites in England that have ever been monitored for SS and N-NO_3 (Fig. 1). Monitoring sites were included in the original HMS monitoring programme if they were at the tidal limit of the river (the farthest point upstream where a river is affected by tidal fluctuations (Davis and Dalrymple, 2011) with an average annual discharge greater than 2 m^3s^{-1}, or any tributaries with a mean annual discharge above 2 m^3s^{-1} (Bellamy and Wilkinson, 2001). This discharge criterion means that there may be several HMS monitoring sites in a given catchment (particularly in large catchments): there were 98 and 88 HMS monitoring sites which are at the tidal limit and where SS and N-NO_3, respectively, had been monitored. The HMS ended in 2014 but the sites have been continued to be monitored by the Environment Agency – the UK government's environmental monitoring agency in England (Environment Agency, 2021). Data from all these sites were considered from 1974 to 2022 inclusive, with sample collection typically being monthly, but this was not always the case and sampling was not necessarily the same for SS and N-NO_3. Sample collection through 2020 was limited and largely absent due to COVID restrictions and then by release of river flow data. The analytical methodology; including quality assurance procedures; the methods for harmonising results across the country, for the HMS sites was prescribed in DoE (1972) and outlined in Simpson (1980) and current procedures are outlined and controlled by the UK government's Standing Committee of Analysts (http://standingcommitteeofanalysts.co.uk/). Given the decades over which the data was collected it is likely that there was land-use and water management change within the study catchment. The use of Year as factor can account for those differences over time in Equation (iv) between sites.

2.4. National-scale budget

The estimation of the national-scale flux was calculated from all the available site-year combination of flux calculations for each of the HMS sites at the tidal limit in England (Fig. 2). The previous attempts at national-scale flux calculation have been hampered by patchiness of site-year combinations meaning that for some years there was no estimate for some regions which led to the need for extrapolation. However, in this approach this method solves that problem and calculates a flux for each possible site-year combination, no matter how uncertain those estimates are and therefore no extrapolation between sampling sites or years is required. However, there remains the problem of how to scale up from the area covered by all the study catchments to the total area of the UK? Worrall and Burt (2007) introduced the method of scaling by

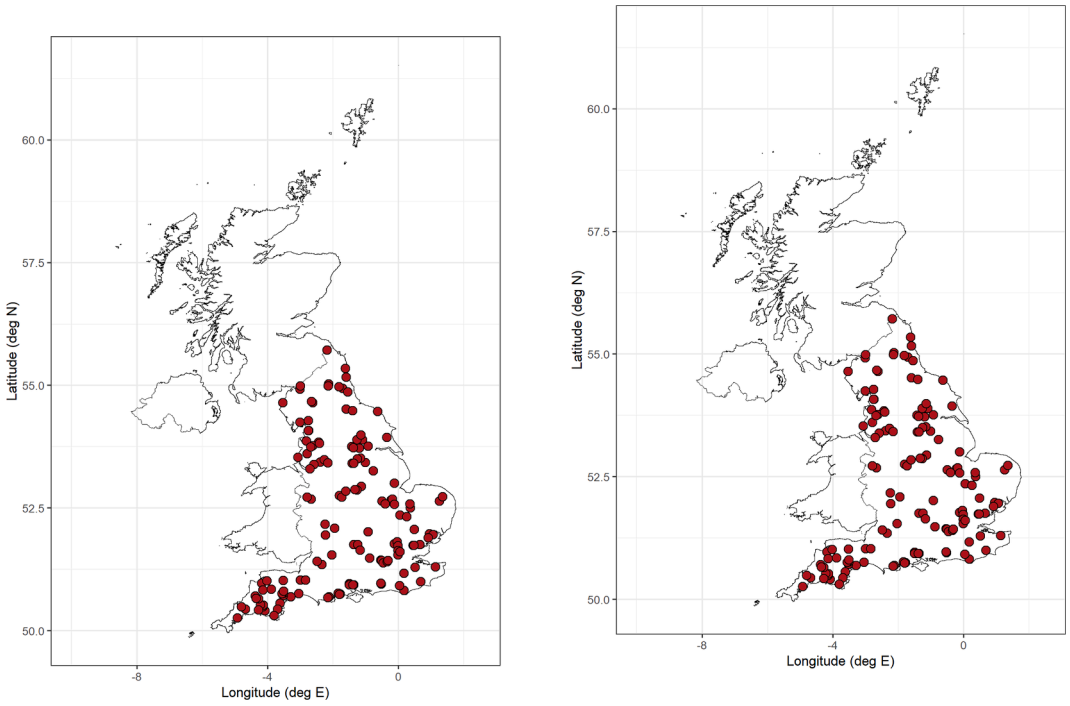


Fig. 1. The location of the sampling locations that could be used in this study for a) SS; and b) N-NO_3.

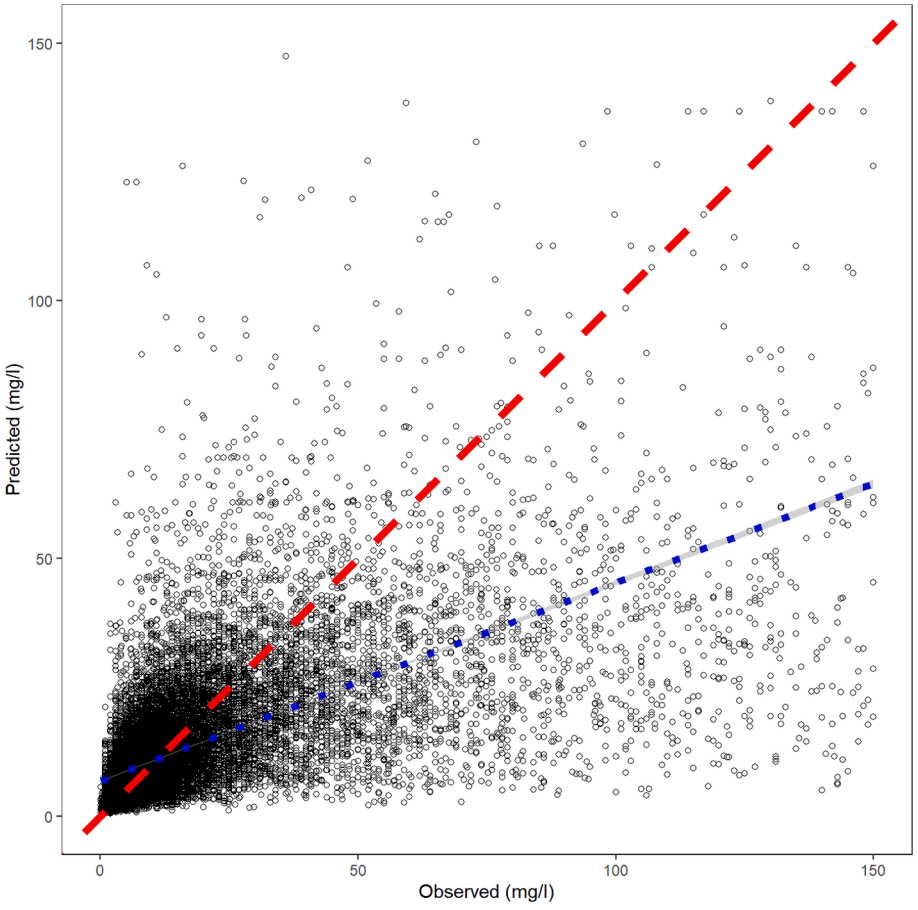


Fig. 2. The comparison between predicted and observed SS based upon logC-logQ regression for individual sites. (—) 1:1 line; and (···) the best-fit line.

the unsampled area by region which meant that catchments are therefore taken as representative of their local area rather than of the country as a whole.

3. Results

In total there were 172,640 measurements of SS concentration in the Harmonised Monitoring Scheme sites in England between 1974 and 2022, and 133,018 measurements of N-NO_3 concentration data. The least sampled year was 2020 (512 samples) and the most sampled year was 1977 (6,218 samples). Outline details of these sites and their respective catchments are provided in Supplementary Material – Table S1.

The median of SS over all sites and the years 1974 to 2022 was 7.3 mg /l (the arithmetic mean = 20.6 mg/l) with the 2.5th to 97.5th percentile range as 1 to 105 mg/l. The time series of the data shows that the highest annual median was in 1982 while the minimum value of the annual median was in 2003. The median number of samples per site was 574 with a range of 15 to 2,385.

For N-NO_3, the least sampled year was 2020 (312 samples) and the most sampled year was 2019 (4,446 samples). The median of N-NO_3 over all sites and the years 1974 to 2022 was 4.9 mg N /l (the arithmetic mean = 5.5 mg N/l) with the 2.5th to 97.5th percentile range as 0.5 to 14 mg N/l. The time series of the data shows that the highest annual median was in 1982 while the minimum value of the annual median was in 2003. The median number of samples per site was 777.5 with a range of 46 to 1637.

3.1. Extrapolation method

3.1.1. Suspended solids (SS)

It was possible to calculate a rating curve at 153 individual sites; the value of β ranged from −0.37 to 1.99 with a median of 0.76 (arithmetic mean of 0.72) and 2.5th to 97.5th percentile range as 0.004 to 1.57. For 145 sites the value of β was significantly different from zero at a 95 % probability – for the values that were significant, the values of β ranged from 0.22 to 1.99; the lower value may considered as a detection limit for this approach. None of the negative values of β were found to be significantly different from zero.

The values of α ranged from −2.21 to 3.21; of the 153 estimates 134 were significantly different from zero at the 95 % probability; those found to be significant varied between −2.21 and 3.21. The lowest significant value of α was 0.04. For all the sites where there was not a significant value of β the value of α was also not significant.

A comparison between the predicted and observed values across all of the C-Q relationships showed that, although the fit was significant (ln(Predicted) = 0.91ln(Observed), $r^2 = 0.91$, n = 25600) the gradient was significantly different from 1 (Fig. 2).

Nitrate (N-NO_3).

For the N-NO_3 concentrations, it was possible to calculate a rating curve at 138 individual sites of which 126 sites showed a significant C-Q relationship. Of the 126 sites where there was a significant relationship, 56 had a significant negative slope (β) while 71 sites had a significant positive slope. In contrast, there were no significant negative slopes values for the SS data. For those sites where the β was significantly negative, the values of β ranged from −0.56 to −0.002 with a median of −0.107 (arithmetic mean of −0.15) and 2.5th to 97.5th percentile range as −0.004 to −0.418. For 71 sites, where β was significant, the value of β ranged from 0.005 to 0.869 with a median of 0.169 (arithmetic mean of 0.205) and 2.5th to 97.5th percentile range as 0.009 to 0.642.

The values of α had a median of 1.54 and ranged from −1.57 to 3.04. Of the 138 estimates, 108 were significantly different from zero at the 95 % probability; of those found to be significant the value of α varied between −1.57 and 3.04, with a median = 1.58. The lowest significant value of α was 0.27. For all the sites where there was not a significant value of β the value of α was also significant.

For N-NO_3 concentrations, the comparison between the predicted and observed values across all of the C-Q relationships showed that, although the fit was significant (ln(Predicted) = 1.39ln(Observed), $r^2 = 0.69$, n = 18) although the gradient is significantly different from 1 and the approach would significantly underestimate NO_3 (Fig. 3).

3.2. Bayesian hierarchical method

3.2.1. Suspended solids (SS)

For SS, all models had $\widehat{R} < 1.01$ and so the fitting process for all models was deemed adequate. The results of model fits relative to each other is detailed in Table 1. The pD will increase with the number of parameters included in the model, but the percentage of the expected pD may decrease as more parameters are included. The model with the most possible parameters (Year + Month + Site) was the one with the lowest percentage of the expected pD. The model with the least parameters (Site) was the one with highest expected pD. The difference between the pD values for the different models should be a multiple of the number of levels in each additional factor. Between the Site and Site + Year models the difference should be a factor of 49, but it was 28, implying that while the difference between years was significant in the majority of cases, it was not in all cases. Similarly, between the Site + Year and Site + Year + Month models, the difference should be a multiple of 12 but it was three, showing that only the minority of months were making a significant difference to the model. The DIC value is at a minimum for the Site model and increases with the inclusion of both the Year and Month factors, i.e. the most efficient model is the Site model. With respect to deviance the best-fit model was the Year + Month + Site model; however, the reduction in deviance upon adding the additional factors was less than 10 %. The final test to be considered was the posterior prediction comparison (Fig. 4) and this suggests that the Site model was the best predictor; even though it had the lowest r^2 the gradient of the posterior prediction was not significantly different from 1. While the r^2 improved for the Site + Year and Site + Year + Month models, the gradient of the posterior prediction plot was significantly different from 1 (Fig. 4).

The values β_{Bayes} varied from 1.38 to −0.005; of the 154 sites where an estimate could be made, 153 showed a significant value of β_{Bayes} (where significance was judged at a 95 % probability of being different from zero). For β_{Bayes} the range of values that were significantly different from zero was 0.07 to 1.389, i.e. significant values of b_{Bayes} were all positive, and the median value was 0.75 with a 95th percentile range of 0.19 to 1.16. The smallest significant value was values of β_{Bayes} was 0.07.

The values of α had a median of 1.03 and ranged from −1.04 to 2.83; of the 154 estimates 153 were significantly different from zero at the 95 % probability of those found to be significant the value of α varied between −1.71 and 3.20. The site where β_{Bayes} was not significantly different from zero was not the same site where α was not significantly different from zero.

Comparing β_{ext} and β_{Bayes} for SS shows there is a significant relationship between the two sets of estimates:

$$\beta_{Bayes} = 0.92\beta_{ext} \mathrm{n} = 153, \mathrm{r2} = 0.97 \quad (6)$$

(0.01)

where the terms are defined earlier in the text. The relationship is significant but also significantly different from 1 and shows that on average β_{Bayes} is 8 % lower than β_{ext} (Fig. 5). Conversely, there was no significant relationship between the confidence intervals for β_{Bayes} and β_{ext} for: the median percent confidence interval β_{Bayes} was 4 % while for log–log, extrapolation method, it was 43 % (Fig. 5), i.e. the Bayesian method was 90 % more precise than the extrapolation method.

3.2.2. Nitrate (N-NO_3)

For the N-NO_3 concentrations, all models had $\widehat{R} < 1.01$ and so the fitting process for all models was deemed adequate. The results of model

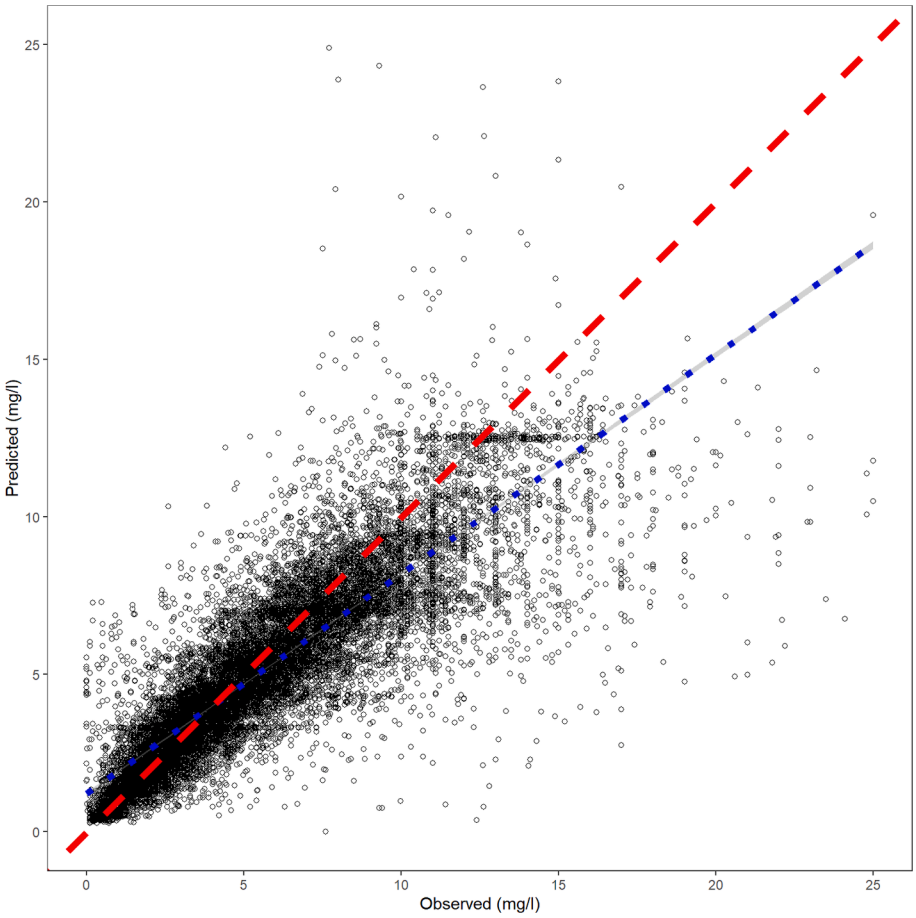


Fig. 3. The comparison between predicted and observed N-NO_3 concentration based upon logC-logQ regression for individual sites. (—) 1:1 line; and (···) the best-fit line.

Table 1
Fitting properties of the model combinations applied. The pD is expressed as both its absolute value and the % of that which could expected if all new parameters included in the model were effective. The R^2 is the plot of posterior predictions, i.e. the Predicted vs. Observed and the gradient of the the posterior predictions plot.

Factors	pD (% expected)	DIC	Deviance	R^2 (gradient)
Suspended solids (SS)				
Site	420 (80)	42,616	42,198	0.91 (1.00)
Site + Year	11,910 (50)	51,760	39,619	0.94 (1.04)
Site + Year + Month	35,207 (12)	74,706	39,601	0.94 (1.02)
Nitrate (N-NO_3)				
Site	274 (52)	111,108	110,834	0.77 (1.00)
Site + Year	11,059 (46)	86,999	75,939	0.79 (1.02)
Site + Year + Month	141,047 (48)	159,300	18,253	0.95 (1.05)

fits relative to each other are detailed in Table 1. There was very little difference in expected pD but comparison between models suggests that, while about 50 % of sites were significantly different, the comparison between Site and Site + Year models was a factor of 40 different, so implying that most years were making a significant difference to prediction in the model. Further, the comparison between Site + Year and the Site + Year + Month model is a factor of 12, i.e. all Months influenced the model. The DIC is lowest for the Site + Year model, but the lowest deviance was for the Site + Year + Month model. The posterior prediction plots show that the best-fit was for the Site + Year + Month model, although the gradient of the best-fit line was significantly different from 1 and suggests that the model would tend to overpredict the N-NO_3 concentration (Fig. 6).

Of the 138 sites for which β_{Bayes} values could be estimated, 129 had a significant value of β_{Bayes} (where significance was judged at a 95 % probability of being different from zero). Of the 129, 46 had significant positive slopes and 83 had significant negative slopes. The range of the significant values was −0.49 to 0.31 with a median of −0.05 with a 95th percentile range of −0.46 to 0.27. The largest value of β_{Bayes} that was not significantly different from zero was 0.30 and the smallest value that was significant was −0.004.

Of the 138 sites for which α_{Bayes} values could be estimated, all 138 had a significant value of α_{Bayes} (where significance was judged at a 95 % probability of being different from zero). The range of the significant values was −11.6 to 34.9 with a median of 19.9 with a 95th percentile range of −6.41 to 32.3.

Comparing β_{ext} and β_{Bayes} for N-NO_3 shows there is a significant relationship between the two sets of estimates:

$$\underset{(0.03)}{\beta_{Bayes} = 0.56\beta_{ext}} \mathrm{n} = 138,\ \mathrm{r2} = 0.66 \quad (10)$$

where terms are defined as in the text. The relationship is significant but also significantly different from 1 and shows that on average β_{Bayes} is 44 % lower than β_{ext} (Fig. 7). Conversely there was no significant relationship between the confidence intervals for β_{Bayes} and β_{ext} for N-NO_3: the median percent confidence interval β_{Bayes} was 10 % while for log–log, extrapolation method, it was 45 %, i.e. 76 % more precise than the extrapolation method (Fig. 7).

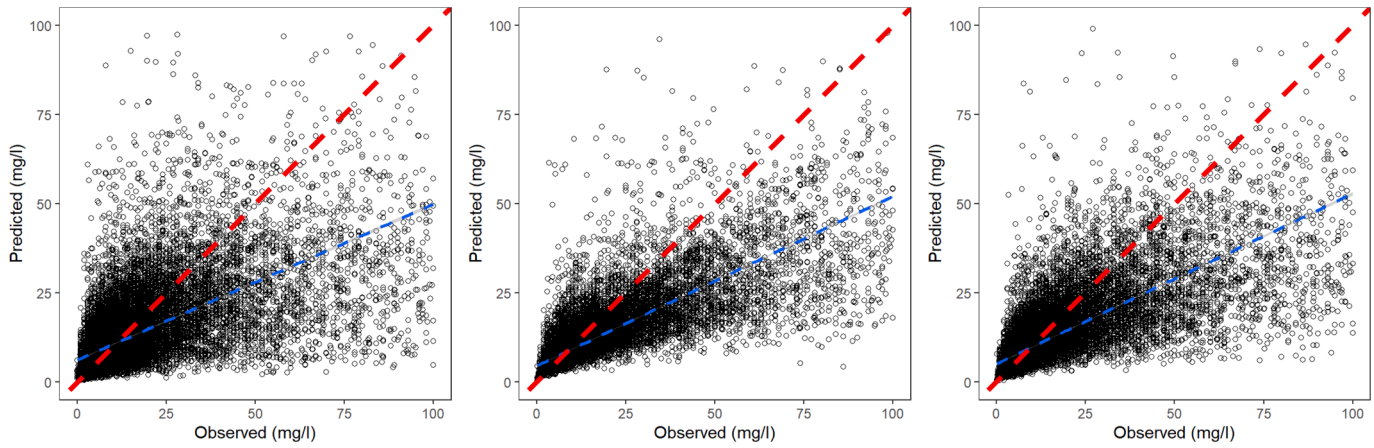


Fig. 4. The posterior prediction plots for SS for the Site, Site + Year, and Site + Year + Month models. The data hexagonally binned for clarity. (---) is the 1:1 line, while ---) is the best-fit line.

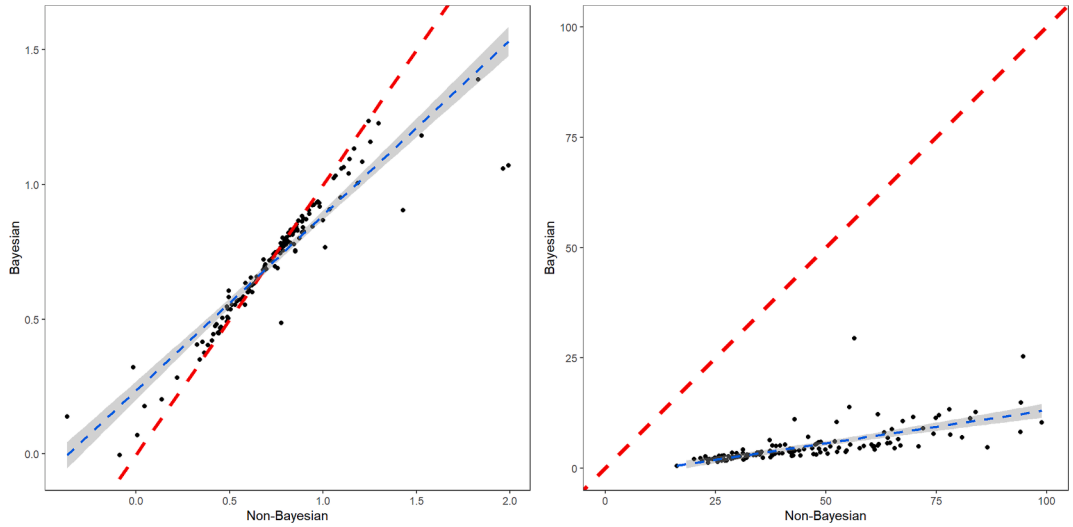


Fig. 5. Comparison of the slope and percentage confidence intervals for Bayesian (β_{Bayes}) vs. extrapolation (β_{fext}) for SS.

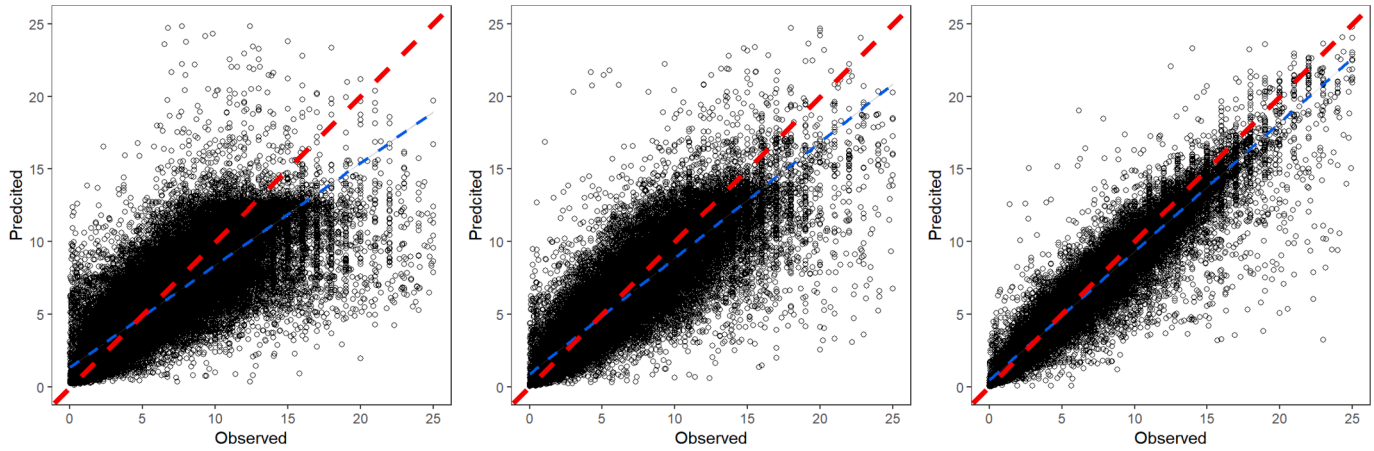


Fig. 6. The posterior prediction plots for N-NO_3 concentration for the Site, Site + Year, and Site + Year + Month models. The data hexagonally binned for clarity. (---) is the 1:1 line, while (---) is the best-fit line.

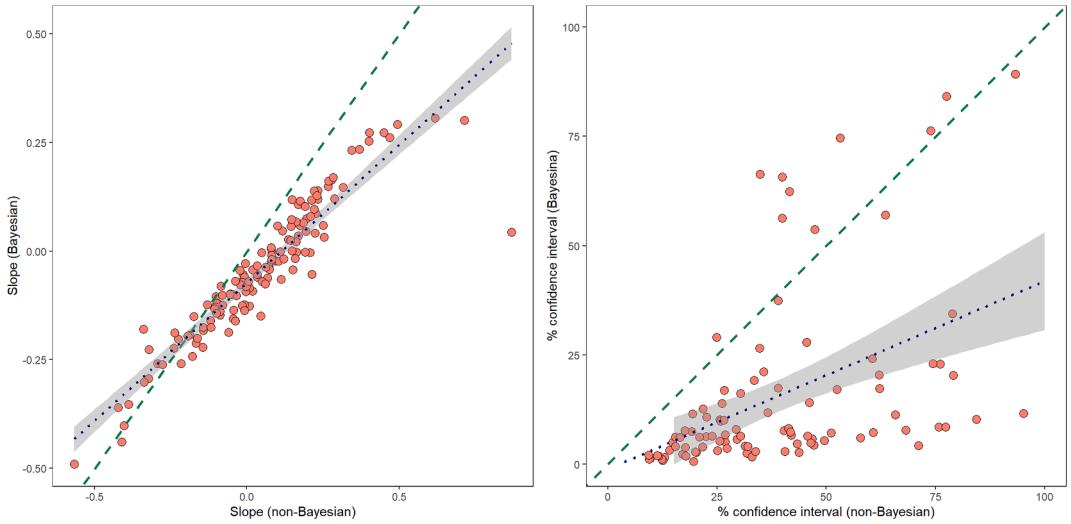


Fig. 7. Comparison of the slope and percentage confidence intervals for Bayesian (β_{Bayes}) vs. extrapolation (β_{ext}) for N-NO_3 concentrations.

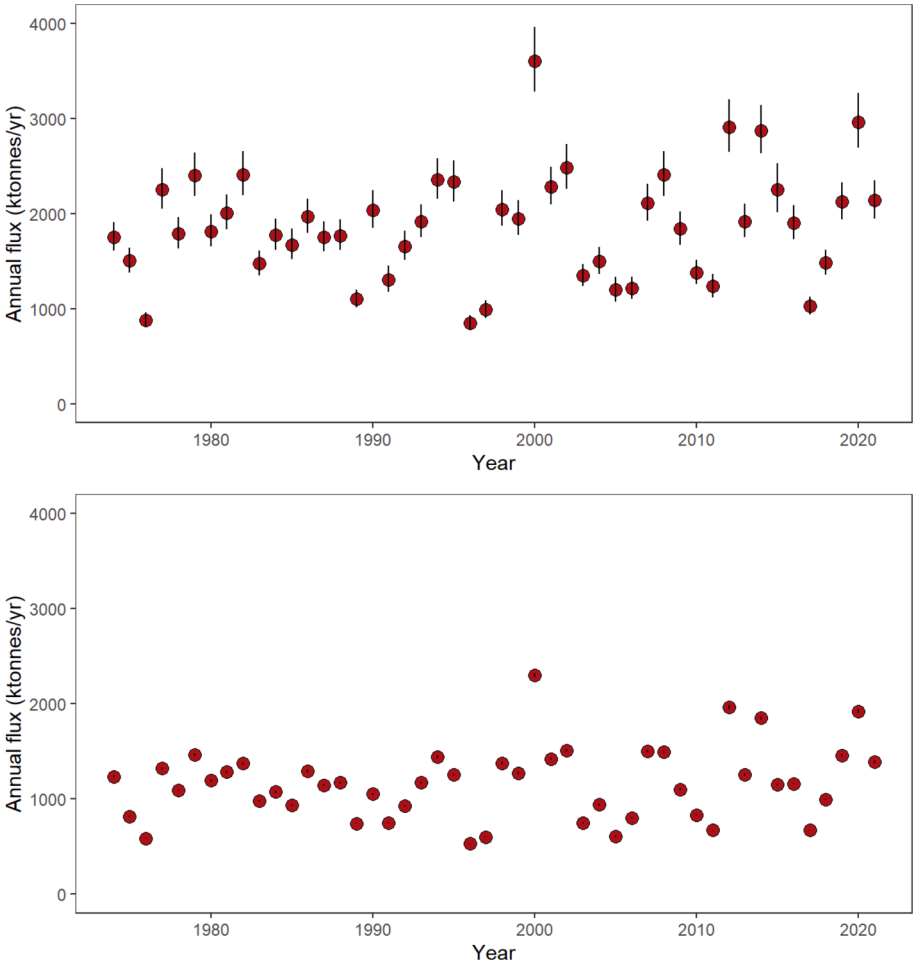


Fig. 8. The national scale flux of SS calculated from the extrapolation and the Bayesian approach.

3.3. National-scale flux

3.3.1. Suspended solids (SS)

The national-scale flux of SS, when calculated from the individual C-Q relationships, gave a median of 1865 ktonnes/yr with a range of 859 to 3599 ktonnes/yr and peaking in the year 2000 (Fig. 8). There was no significant increase over the period of the study and so its arithmetic mean of 1872 ktonnes/yr and 95th percentile range of 898 to 2954 ktonnes/yr may give a better estimate of uncertainty than the 95 % confidence interval on the expected value which itself had an arithmetic mean of 9.3 %.

The median for the Bayesian flux estimate was 1,162 with a range of

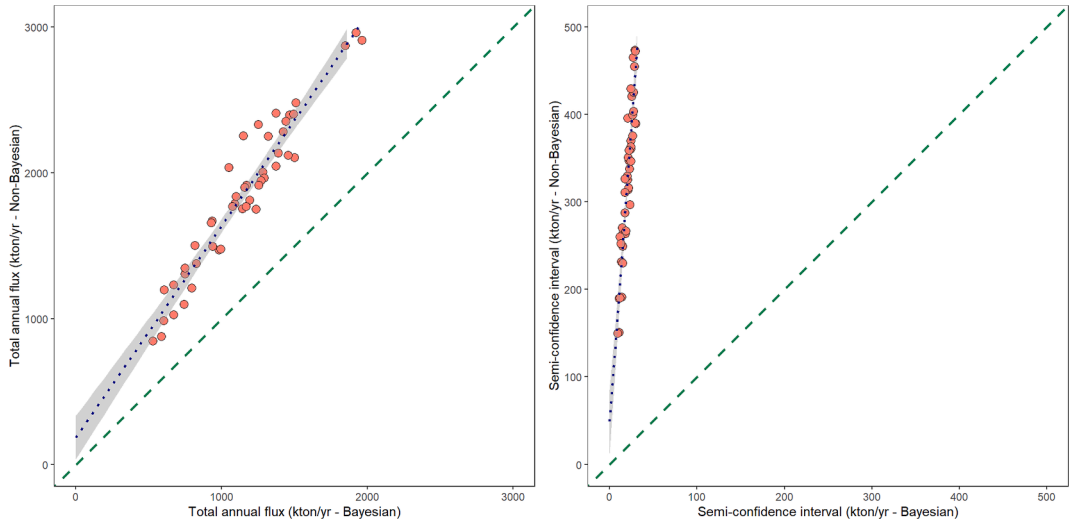


Fig. 9. The national scale flux of SS calculated from the extrapolation and the Bayesian approach.

527 to 2300 ktonnes/yr (Fig. 8). There was no significant increase over the period of the study and the arithmetic mean was 1161 ktonnes/yr with a range of 588 to 1955 ktonnes/yr. The 95 % confidence interval had a median of 1.0 %, i.e. the scaling up to the national flux using the Bayesian approach was 80 % more precise than the extrapolation method. There was a significant relationship between the estimates (r^2 = 0.99, n = 48) with the Bayesian estimate being 62 % of the non-Bayesian flux estimation (Fig. 9).

3.3.2. *Nitrate (N-NO_3)*

The national scale flux of N-NO_3 from the individual C-Q extrapolation relationships had a median of 2,521 with a range of 1,433 to 3,730 ktonnes/yr and peaked in the year 2000; the expected value was 2,476 with a 95 % interval of 1,500 to 3,459 ktonnes/yr (Fig. 10). There was a significant trend of 1.2 ktonnes/yr^2. The Bayesian approach had a median of 1,151 with a range of 523 to 2,277 ktonnes/yr and peaked in the year 2000, the mean was 1,161 with a 95 % interval of 588 to 1,955. There was a significant trend of 0.58 ktonnes/yr^2. There was a significant relationship between the estimates (r^2 = 0.99, n = 48) with the Bayesian estimate being 47 % of the non-Bayesian flux estimate (Fig. 11).

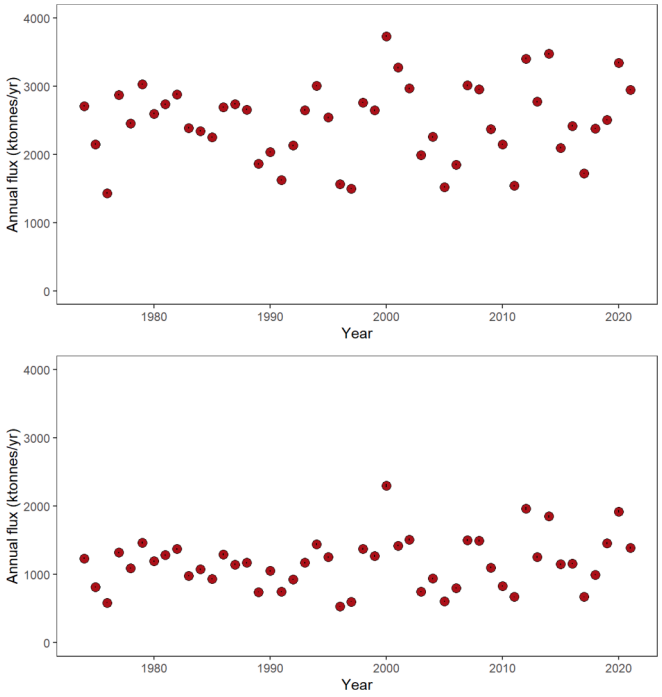


Fig. 10. The national scale flux of N-NO_3 concentration calculated from the extrapolation and the Bayesian approach.

4. Discussion

Using a Bayesian hierarchical approach provides a number of clear advances over previous work. The Bayesian approach was more sensitive than the non-Bayesian, extrapolation approach for both SS and N-$NO_3$where sensitivity is indicated by the number of values of both α and β that the Bayesian approach found significantly different from zero. The Bayesian method was more sensitive. Secondly, the Bayesian method had a lower detection limit for C-Q relationships than the extrapolation method. Finally, the Bayesian approach had lower confidence intervals on slope estimates with the confidence interval being 76 % lower for N-NO_3 and 90 % lower for SS.

The Bayesian approach makes use of all available information and so gains value from the whole monitoring network, i.e. maximum information is obtained from past and ongoing monitoring; there is no censoring of the data for sites below a certain sampling frequency. This advantage has several benefits in itself. First, by using all of the data, this approach gives best value for the money invested in environmental monitoring. Second, the method does not even require that a particular monitoring site be monitored within the time step of interest (e.g. a particular year); all that the method requires is that a monitoring site had been monitored at some stage during the period covered by the analysis and that a site, but not necessarily a particular site, had been sampled during the period of interest, as long as that one site (wherever it was) was sampled in each year of interest. The uncertainty may be larger for those sites not sampled in a given time step compared to those actually sampled in that particular time step. The Bayesian framework not only means that all observations have value, but it means that the results will improve with time as more observations become available. Given the advantages of this approach, then it becomes possible to give more robust interpretation of occurrence, flux and trends from any study region.

Would it be possible to improve on the current approach? It would be possible to include other covariates and other factors within the analysis. This study has already included the year, month, site and flow, but we might make two further improvements. First, for some determinands of interest, there may be other measured determinands that would be

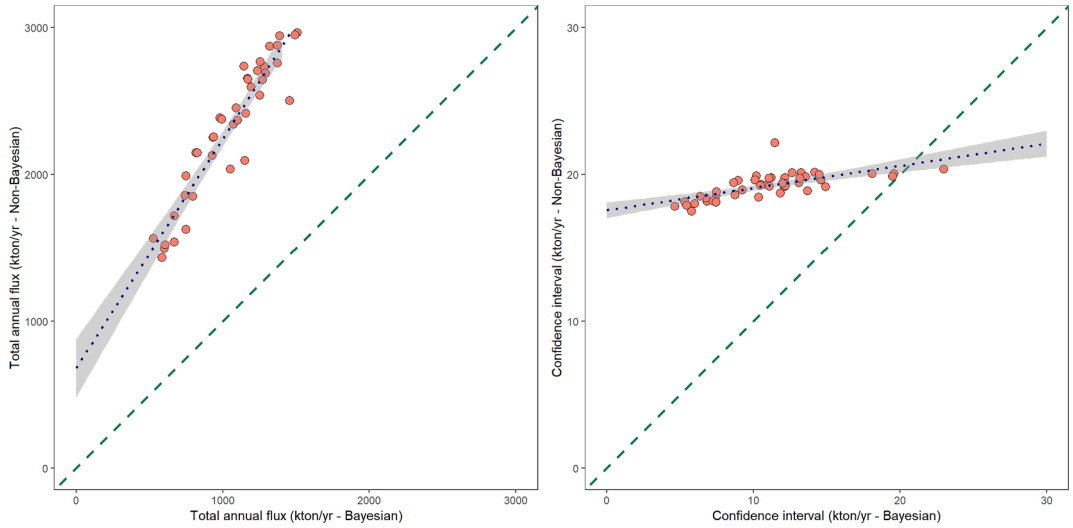


Fig. 11. The national scale flux of N-NO_3 concentrations calculated from the extrapolation and the Bayesian approach.

expected to be related and so inform prediction; for example, other nutrients could aid in the prediction of N-NO_3. Second, there could be a spatial correlation within the monitoring that could relate monitoring sites, e.g. sites on the same river or neighbouring rivers being more closely related to each other than to sites at opposite ends of the study region. Such a spatial correlation in the observations is not included in this modelling approach. Qian et al. (2005) have developed a Bayesian hierarchical model for the calculation of nutrient loads in rivers that incorporated spatial correlation, but the spatial correlation was based on the flow through the river network which provides for one-dimensional and directional correlation not appropriate for the three dimensions of the catchments considered here. This study has compared to the extrapolation approach but a further test of the Bayesian approach would be compare to high frequency water quality monitoring data. High frequency monitoring is available for a range determinands including nitrate (Rode et al., 2016), although for suspended solids high frequency monitoring is complex. High frequency monitoring for suspended solids as it is turbidity that can be monitored and not suspended solids thus requiring site specific calibration (eg. Villa et al., 2019). England's Environment Agency does operate high frequency monitoring sties of which six coincide with the long term monitoring sites used within this study, however, at none of these six was either N-NO_3 or SS measured.

The Bayesian approach was able to be scaled up to give national-scale fluxes. Previous estimates have been made of the SS flux for the UK. Worrall et al. (2013b) used an interpolation method to estimate the flux of suspended sediment from the UK, which varied between 2199 k tonnes and 27550 k tonnes/yr, or approximately 1272 to 15,931 ktonnes/yr for England as distinct from the larger UK area. Again, using an interpolation method, Worrall et al. (2009) estimated that the N-NO_3 flux from Great Britain for 1990 to 2005 varied from 275 to 758 ktonnes N/yr; if corrected for the area of England, then this would be 159 to 458 ktonnes/yr.

5. Conclusions

The study has demonstrated an improved method for the calculation of fluvial fluxes. The approach uses Bayesian hierarchical modelling to account for differences in concentration-discharge (C-Q) relationship between monitoring, seasons and years. Further, the approach centres the data so that when the C-Q relationship is not significant, the approach will return the expected value of the determinand. In comparison to an approach based upon the individual C-Q relationships for sites, the Bayesian method was able to find more statistically significant relationships and do so with greater precision. The approach combines the interpolation and extrapolation methods and so is not only more precise than previous approaches, it is more flexible and can handle diverse hydrological behaviours not only between determinands but also between catchments and years. The Bayesian approach uses exactly the same data as previous approaches, and so therefore, makes more efficient use of publically-funded monitoring data.

CRediT authorship contribution statement

Fred Worrall: Writing – original draft, Resources, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Nicholas J.K. Howden:** Writing – review & editing, Conceptualization. **Tim P. Burt:** Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2024.132550.

Data availability

All data is publically available and links to it have been included in the text

References

Bellamy, D., Wilkinson, P., 2001. Ospar 98/3: an environmental turning point or a flawed decision? Mar. Pollut. Bull. 42 (2), 87–90.
Cassidy, R., Jordan, P., 2011. Limitations of instantaneous water quality sampling in surface-water catchments: comparison with near-continuous phosphorus time-series data. J. Hydrol. 405 (1–2), 182–193.
Cohn, T.A., Caulder, D.L., Gilroy, E.J., Zynjuk, L.D., Summers, R.M., 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.M., Watts, C.D., 2002. A comparison of river load estimation techniques: application to dissolved organic carbon. Environmetrics 13 (7), 733–750.
Crawford, C.G., 1991. Estimation of suspended-sediment rating curves and mean suspended-sediment loads. J. Hydrol. 129, 331–348.

Davis, Jr., R.A., Dalrymple, R.W., 2011. Principles of Tidal Sedimentology. Springer Science & Business Media. p. 82. ISBN 978-94-007-0123-6.
Department of the Environment, 1972. Analysis of Raw, Potable and Wastewaters. HMSO.
Environment Agency, 2021. Open WIMS data [Dataset]. https://environment.data.gov.uk/water-quality/view/download/new.
Godsey, S.E., Kirchner, J.W., Clow, D.W., 2009. Concentration–discharge relationships reflect chemostatic characteristics of US catchments. Hydrol. Process. 23, 1844–1864.
Hobbs, B.F., 1997. Bayesian methods for analysing climate change and water resource uncertainties. J. Environ. Manage. 49 (1), 53–72.
Howden, N.J.K., Birgand, F., Burt, T., Worrall, F., 2018. The seven sources of variance in fluvial flux time series. Hydrol. Process. 32 (26), 3996–3997.
Johnes, P.J., 2007. Uncertainties in annual riverine phosphorus load estimation, impact of load estimation methodology, sampling frequency, baseflow index and catchment population density. J. Hydrol. 332, 241–258.
Kim, J., Lim, K.J., Park, Y.S., 2018. Evaluation of regression models of LOADEST and eight-parameter model for nitrogen load estimations. Water Air Soil Pollut. 229, 179–190.
Kronvang, B., Bruhn, A.J., 1996. Choice of sampling strategy and estimation method for calculating nitrogen and phosphorus transport in small lowland streams. Hydrol. Process. 10 (11), 1483–1501.
Littlewood, I.G., 1995. Hydrological regimes, sampling strategies and assessment of errors in mass load estimates for United Kingdom rivers. Environ. Int. 21 (2), 211–220.
Littlewood, I., Watts, C., Custance, J., 1998. Systematic application of United Kingdom river flow and quality databases for estimating annual river mass loads (1975–1994). Sci. Total Environ. 210–211, 21–40.
McDonald, K.S., Ryder, D.S., Tighe, M., 2015. Developing best-practice Bayesian Belief Networks in ecological risk assessments for freshwater and estuarine ecosystems: a quantitative review. J. Environ. Manage. 154, 190–200.
Moatar, F., Meybeck, M., Raymond, S., Birgand, F., Curie, F., 2012. River flux uncertainties predicted by hydrological variability and riverine material behaviour. Hydrol. Process. 27 (25), 3535–3546.
Qian, S.S., Reckhow, K.H., Zhai, J., McMahon, G., 2005. Nonlinear regression modeling of nutrient loads in streams: A Bayesian approach. Water Resour. Res. 41 (7), W07012.
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 (Québec, Canada). J. Hydrol. 326 (1–4), 295–310.
Rode, M., Wade, A.J., Cohen, M.J., Hensley, R.T., Bowes, M.J., Kirchner, J.W., Arhonditsis, G.B., Jordan, P., Kronvang, B., Halliday, S.J., Skeffington, R.A., Rozemeijer, J.C., Aubert, A.H., Karsten Rinke, K., Joma, S., 2016. Sensors in the stream: the high-frequency wave of the present. Environ. Sci. Technol. 50, 10297–10307.
Runkel, R.L., Crawford, C.G., Cohn, T.A., 2004, Load Estimator (LOADEST): A FORTRAN Program for Esti mating Constituent Loads in Streams and Rivers: U.S. Geological Survey Techniques and Methods Book 4, Chap ter A5.
Simpson, E.A., 1980. The Harmonization of the monitoring of the quality of rivers in the United Kingdom. Hydrol. Sci. Bull. 25 (1), 13–23.
Skarbøvik, E., Stånacke, P., Bogen, J., Bønsnes, T.E., 2012. Impact of sampling frequency on mean concentrations and estimated loads of suspended sediment in a Norwegian river: implications for water treatment. Sci. Total Environ. 433, 462–471.
Stewart, B., Shanley, J.B., Kirchner, J.W., Norris, D., Adler, T., Bristol, C., Harpold, A.A., Perdrial, J.N., Rizzo, D.M., Sterle, G., Underwood, K.L., Wen, H., Li, L., 2022. Streams as mirrors: Readingsubsurface water chemistry from streamchemistry. Water Resour. Res. 58, e2021WR029931.
Verma, S., Markus, M., Cooke, R.A., 2012. Development of error correction techniques for nitrate-N load estimation methods. J. Hydrol. 432, 12–25.
Villa, A., Fölster, J., Kyllmar, K. (2019). Determining suspended solids and total phosphorus from turbidity: comparison of high-frequency sampling with conventional monitoring methods. Environ Monit Assess 191, art. no. 605.
Walling, D., Webb, B., 1988. The reliability of rating curve estimates of suspended sediment yield: Some further comments, Sediment Budgets, Proceedings of the Porto Alegre Symposium, December 1988, Publication No. 174. Wallingford, England: International Association of Hydrological Sciences, pp. 337–350.
Walling, D., Webb, B., 1985. Estimating the discharge of contaminants to coastal waters by rivers: Some cautionary comments. Mar. Pollut. Bull. 16 (12), 488–492.
Wan, R., Cai, S., Li, H., Yang, G., Li, Z., Nie, X., 2014. Inferring land use and land cover impact on stream water quality using a Bayesian hierarchical modeling approach in the Xitiaoxi River Watershed. China. J. Environ. Manage. 133, 1–11.
Webb, B.W., Phillips, J.M., Walling, D.E., Littlewood, I.G., Watts, C.D., Leeks, G.J.L., 1997. Load estimation methodologies for British rivers and their relevance to the LOIS RACS(R) programme. Sci. Total Environ. 194–195, 379–389.
Wilson, M.P., Worrall, F., Davies, R.J., Hart, A., 2020. A dynamic baseline for dissolved methane in English groundwater. Sci. Total Environ. 711, Art. No. 134854.
Worrall, F., Burt, T.P., 2007. Flux of dissolved organic carbon from U.K. rivers. Global Biogeochem. Cycles 21, 1, GB1013.
Worrall, F., Burt, T.P., Howden, N.J.K., Whelan, M.J., 2009. The flux of nitrate from Great Britain 1974–2005 in the context of the terrestrial nitrogen budget of Great Britain. Global Biogeochem. Cycles 23, GB3107.
Worrall, F., Davies, H., Bhogal, A., Lilly, A., Evans, M.G., Turner, E.K., Burt, T.P., Barraclough, D., Smith, P., Merrington, G., 2012. The flux of DOC from the UK – predicting the role of soils, land use and in-stream losses. J. Hydrol. 448–449, 149–160.
Worrall, F., Burt, T.P., Howden, N.J.K., 2013a. Assessment of sample frequency bias and precision in fluvial flux calculations – an improved low bias estimation method. J. Hydrol. 503, 101–110.
Worrall, F., Burt, T.P., Howden, N.J.K., 2013b. The flux of suspended sediment from the UK 1974 to 2010. J. Hydrol. 504, 29–39.
Worrall, F., Howden, N.J.K., Burt, T.P., 2020. The dissolved organic carbon flux from the UK – A new Bayesian approach to flux calculation. J. Hydrol. 590, 125511.
Zhi, W., Li L., 2020. The shallow and deep hypothesis: subsurface vertical chemical contrasts shape nitrate export patterns from different land uses. Environ. Sci. Technol. 2020 54 (19), 11915-11928. DOI: 10.1021/acs.est.0c01340.