The problem of self-correlation in fluvial flux data – the case of nitrate flux from UK rivers

Fred Worrall[1], Tim P.Burt[2], and Nicholas J.K. Howden[3].

1. Dept of Earth Sciences, Science Laboratories, South Road, Durham, DH1 3LE, UK.
2. Dept. of Geography, Science Laboratories, South Road, Durham, DH1 3LE, UK.
3. Dept. of Civil Engineering, University of Bristol, Queens Building, Bristol, UK.

Abstract

This study proposes a general method for testing for self-correlation (also known as spurious or induced correlation) in comparisons where there is a common variable, e.g. the comparison of the fluvial flux of a component with water yield. We considered the case of the fluvial flux of nitrate from 153 catchments from across the UK for which there were at least 10 years of data. The results show that 66% of records (102 catchments) could be rejected as significantly self-correlated (P< 95%). Among the 51 catchments, which proved to be significantly different from the spurious, or self-correlated result, the response was variable with linear, convex, s-curve and mixed results proving the best description. There was no spatial pattern across the UK for the results that were and were not rejected as spurious; the most important predictor of not being self-correlated was the length of record rather than any catchment characteristic. The study shows that biogeochemical stationarity cannot be assumed and that caution should be applied when examining fluvial flux data.

Keywords: spurious correlation; induced correlation; biogeochemical stationarity

[1] Corresponding author: Fred.Worrall@durham.ac.uk; tel. no. +44 (0)191 334 2295; fax no. +44 (0)191 334 2301.

Introduction

Significant correlations can occur in data that are entirely spurious and not related to any causal or physical relationships between the variables being compared (Kenney, 1982). Such spurious correlations (also referred to as induced or self-correlation) will occur where there is common variable in the comparison, e.g. A vs. A/B, A*B vs. B or A/B vs. C/B). The strength of the self-correlation will increase where the variance in the common variable is large in comparison to that of the other or unique variables. Kenney (1982) pointed out that the reverse can also be the case: self-correlation can weaken a strong relationship when the variance of the common variable is equal to or less than that of the other or unique variable. Furthermore, spurious correlation is enhanced whenever log-transformation has been used. Interpretations based upon plots or relationships with common variables are frequently applied and their occurrence has been discussed relative to atmospheric sciences (e.g. Baas et al., 2006); is raised as an issue in geomorphological data (Gani et al., 2007); and the correlations of parts and wholes in ecology (eg. stand biomass and tree measurements – Dean and Cao, 2003).

Several studies have proposed methods for identifying spurious correlation. The strength of relationship between two variables can be tested by standard statistical tests of the correlation coefficient. Pearson (1897), in the original study of spurious correlation, gives an approximation of the correlation coefficient that would be expected for the correlation of two variables both ratioed to a third. However, Pearson's formula is only an approximation and so tests based upon randomisation have become more common. McCuen and Surbeck (2008) observed that many environmental models (e.g. Michaelis-Menten kinetics) are calibrated by linearization based upon plots that include a common variable; they recommend avoiding such a linearization step and used non-linear fitting methods but provided no test of spurious correlation. Lenahan et al. (2011) discussed "induced correlations" with respect to hydrochemical data and especially the comparison of ratio data and summed data, and used randomisation to view self-correlation but, they provided no formal account of how randomisation was performed nor how to compare between the observed and the randomised data. Jackson and Somers (1991) placed randomisation tests in the same context as the null hypothesis and therefore stated that, when comparing ratioed variables, the null hypothesis used not that the regression coefficient was zero but rather the hull hypothesis was that the regression coefficient was the expected value (usually the arithmetic mean) of the distribution of regression coefficients resulting from randomisation of the data, however, Jackson and Somers (1991) provided no suggestion for making this comparison. Kenney (1982) formulated the self-correlated regression coefficient for normally distributed data in the comparison A+B vs. B and Vickers et al. (2009a) used this approach to consider surface exchange of CO_2; they considered the test to be that the regression coefficient was greater than the value predicted by the formulation of Kenney (1982).

There are several problems with these approaches used to assess self-correlation; firstly, some of the above provide no formal test method at all; secondly, the formulations so far provided are either approximations or for specific comparisons; thirdly, many methods provide no formal test of the difference between the randomised results and the observed. Alternatively, we propose a different method and test for the detection of induced correlation. For the randomisation we propose that we do not assume a normal distribution and that other distributions or no distribution at all would be more appropriate for some datasets. The difference between the randomly generated and the observed data will be statistically tested and, furthermore the nature of the best-fit line will be considered relative to the randomised data.

Within hydrology, the self-correlation is evident in a number of common approaches to analysing and interpreting data. The authors of this paper have themselves previously used comparisons that we might consider upon reflection to be vulnerable to self-correlation. Worrall et al. (2008) and Worrall and Burt (2008) compared changes in annual DOC flux in a range of catchments over periods of severe droughts to the changes in annual discharge as a means of testing whether there was a biogeochemical response to drought. Worrall et al. (2012) modelled 125 years of Ca flux data by comparing it to annual discharge and, perhaps not surprisingly, annual discharge was the most important factor in explaining the Ca flux. The USEPA (2005) actually recommend using the correlation between pollutant flux and discharge in order to improve flux estimation and Shivers and Moglen (2008) provide an alternative approach on this basis. However, a strong linear relationship between component flux (e.g. nitrate) and annual discharge has been used to suggest biogeochemical stationarity (sometimes also referred to as chemiostasis, eg. Stackpoole et al., 2014) as an emergent property of catchments (Basu et al. 2010). However, Godsey et al. (2009) argued for chemiostasis on the basis of concentration discharge relationships. Basu et al. (2010) use the term biogeochemical stationarity to a low variation in concentration of a component relative to hydrological variation. This biogeochemical stationarity is taken to arise from a legacy of available material present in the catchment that means no matter which pathways water takes through a catchment the results is very similar, i.e. stationary. A strong linear relationship between component flux and discharge (annual water yield) suggests a single concentration of the component exists across a wide range of flows. This test of stationarity has been used for dissolved carbon (Giesler et al., 2014, Jantze et al., 2013).

The sediment delivery ratio (SDR) is a common approach used to explain changes in sediment flux through a catchment (Roehl, 1962; Burt and Allison, 20~~09~~10). The SDR approach is vulnerable to self-correlation yet correlations based upon the variation of sediment yield with catchment area are commonly used, interpreted and discussed (e.g. Tetzlaff et al., 2013). Worrall et al. (2014) have shown that such approaches do suffer from self-correlation. In the SDR context, a negative relationship between SDR and catchment area would be predicted and so positive relationships might be thought to be free of self-correlation. Such relationships between SDR and catchment area have been observed in a number of studies (e.g. Church and Slaymaker, 1989).

Prairie and Bird (1989) in their defence of part versus whole analysis in biology warn of “not throwing the baby out with the bath water”, an attitude also taken by Francey et al. (2010, 2011) in their analysis of pollution loads. Therefore, in the hope of making the most of the available information this study considers how self-correlated, or spurious correlation, can tested for and how to consider relationships in comparisons vulnerable to spurious correlation.

Methods

Data set

The Harmonised Monitoring Scheme (HMS) was established in 1974 to measure important hydrochemical fluxes to the North Atlantic and to allow their trends to be monitored (Simpson, 1980). These measurements met the UK’s commitment to a series of international agreements and treaties (Bellamy and Wilkinson, 2001); and standards and consistency of measurement over time and space are defined within the HMS programme. There are 56 HMS sites in Scotland and 214 sites in England and Wales. Monitoring sites were placed at the tidal limits of all rivers with an average annual discharge of over 2 m^3s^{-1}, with additional sites placed on major tributaries. These criteria means that there is good spatial coverage of the coast of England and Wales but in Scotland many of the west coast rivers are too small to warrant inclusion in the HMS. A range of water quality parameters are measured at these sites: pertinent to this study, the HMS measures nitrogen as nitrate and river discharge (instantaneous discharge and daily average discharge).

Monitoring is the responsibility of regional offices of the Environment Agency in England and Wales and the Scottish Environment Protection Agency in Scotland. As a result, sampling frequencies vary ranging from sub-weekly to monthly (or even less frequently in some cases). Data from any year at any site where fewer than 12 samples were collected in that year were excluded from the analysis. Consequently, although there are 270 HMS sites across Great Britain, the number of sites which could be included in any one year was variable: the distribution of sites from which data were used is shown in Figure 1. Furthermore, because self-correlation relies on examining a correlation and so only sites with at least 10 years of annual flux data were considered.

In addition to the use of data from the HMS sites, this study considered the World's longest water quality record, the Thames at Teddington (Howden et al., 2011). Howden et al. (2011) have demonstrated the consistency and coherence of this record over the 126 years. The record at Teddington consists of monthly average nitrate concentrations since 1867 but river discharge records were only available for complete calendar years from 1883 to 2008 – 126 years of data. Therefore, the correlation between annual nitrate flux from the Thames at Teddington and the annual water yield was analysed for self-correlation in the same way as data from the HMS.

Flux calculation

For cases where data are relatively sparse, such as in much of the HMS, Littlewood et al. (1998) suggested that the product of flow weighted concentration and the annual discharge was most appropriate. However, HMS sampling is generally aperiodic and the following method (Rodda and Jones, 1983) is more appropriate:

$$F_{jy} = KA_y \sum_{1}^{N} n_y C_i Q_i \quad \text{(i)}$$

$$n_y = \frac{A_y}{N_y} \qquad \text{(ii)}$$

where F_{jy} = the annual flux at the site j for a given year *y(tonnes N/yr*); C_i = the measured concentration at the site at time *i (mg N/l)*; Q_i= the river discharge at time *i* (m^3/s); K = a conversion factor which takes into account the units used (0.0864); n_y = average number of days between samples (days); N_y = the number of samples at the site in year *y*; and A_y = the number of days in year *y* (can vary with a leap year). This approach assumes that each sample taken at a site is equally likely to be representative of an equal proportion of the year as any other sample.

For the purpose of this study no attempt is made to sample bias correct the estimates of nitrate flux or of annual water yield for two reasons. Previous studies that have compared flux vs. yield plots have not sample bias corrected their estimates. Secondly, given the pairing of the data for concentration and flow there would be a sample bias in both that would be of similar order of magnitude which would mean the overall effect may be small.

Estimation of self-correlation

Vickers et al. (2009a) suggested a method for testing the occurrence and magnitude of self-correlation. To apply this method to the problem of comparing flux and annual water yield, a single value of concentration is drawn at random from the normal distribution fitted to the observed concentration data and paired with a value of stream discharge drawn at random from the normal distribution of stream discharge data. The process is repeated to derive the required number of random pairs for calculation of an annual flux and annual water yield (in the case here a number of pairs equal to or greater than 12). The process can be repeated as many times as desired so that sufficient pairs of estimates of annual flux and annual water yield exist and these can be plotted against each other and compared to the observed pairs of annual flux and annual water yield. If there is no significant difference between the best-fit line for the observed data and the best-fit line from the randomised data then any line fitted to the observed data, can be dismissed as spurious due to self-correlation. However, there are several problems with this approach. Firstly, the randomisation process assumes that the data from which the calculation was derived are normally distributed. This situation was true for the gas flux data which Vickers et al. (2009a) studied but this is unlikely to be true for the concentration and stream discharge data that need to be considered here. Therefore, this study considered normal, log normal, and gamma distributions in order to select random pairs of data from the observed concentration and stream discharge. However, the assumption of normality made by Vickers et al. (2009a), or indeed the assumption of any distribution assumes that sufficient data would be available such that a distribution of whatever sort could be accurately fitted. Given that in this study annual flux was calculated based on as few as 12 samples per year, fitting complex distributions with repeatable accuracy would be difficult. As an alternative approach, this study selected at random concentration and stream discharge data not from a distribution fitted to the observed but taken at random from each of the actually observed data series. This second approach requires no assumption about the distribution of the data. Thirdly, it was assumed that a straight line relationship exists but that this is not necessarily the case and indeed more complex descriptions of the relationship between flux and annual water yield may be found and so this model considers linear, power law, exponential and sigmoidal relationships (Weibull function). The Aikike Information Criterion (AIC) was used to decide between relationships given the additional degrees of freedom from 2 to 4.

For additional comparison the value of the self-correlated regression for normally distributed data was calculated (Kenney, 1982):

$$r^2 = \frac{1}{\left(1 + \frac{\sigma_{com}}{\sigma_{uni}}\right)} \qquad \text{(iii)}$$

Where: σ_{com} = the standard deviation of the common variable (for this study riverflow); σ_{uni} = the standard deviation of the unique variable (for this study nitrate concentration).

Catchment characteristics

To help understand the occurrence of self-correlation, the results were compared to a range of catchment properties including soil, land use and hydrological characteristics. The dominant soil of each 1 km^2 grid square in Great Britain was classified into mineral, organo-mineral and organic soil based upon the classification system of Hodgson (1997); note that by this definition peat soils are a subset of organic soils. The land use for each 1 km^2 square of Great Britain was classified into: arable, grass and urban based upon the June Agricultural Census for 2004. Note that values for forested land are not available from this census. In addition, the number of cattle and sheep in each cell were counted from the June Agricultural Census for 2004. Catchment areas were calculated from the CEH Wallingford digital terrain model which has a 50 m grid interval and a 0.1 m altitude interval. The soil and land-use characteristics for each 1 km^2 grid square were summed across catchment areas and expressed as percentages of each catchment area. Within the catchments for which N flux information was available both soil and land-use properties were expressed as percentages of catchment area. For livestock, equivalent sheep per hectare values were calculated based on a ratio of 3.1 sheep per cow (Johnes and Heathwaite, 1997). In addition, it was possible to give a range of hydrological characteristics for each catchment. Based upon data from the National River Flow Archive (www.nrfa.ac.uk), hydroclimatic measures used were: base flow index (BFI; Gustard et al., 1992); average actual evaporation, the average annual rainfall; and by difference the average runoff for each catchment for which flux data were available. Also included in the analysis were: the sample size for each catchment, the ratio of the concentration variance to flow variance for each catchment; and the self-correlation regression coefficient as predicted by Equation (iii).

The presence, or absence, of self-correlation was then compared to the catchment characteristics using logistic regression analysis, Logistic regression analysis was fitted to the binary response variable (presence/absence of self-correlation) using maximum likelihood techniques; the fit of the analysis was assessed using correct classification; and importance of variables within any logistic regression was measured using the odds ratio.

Results

Across all HMS data from 1974 to 2010, it was possible to assess 105019 pairs of concentration and flow data in 153 catchments (Figure 1) – these catchments ranged in scale from 4 to 9885 km^2. For the 153 catchments, the median number of years that could be considered was 31, with an inter-quartile range of 24 to 34 years; the minimum number of years at any one site was 12 years. At the 95% probability that the null hypothesis can be rejected, i.e. a 95% probability that the observed data relationship is different from that due to spurious or self-correlation, then 51 out of 153 catchments show a relationship significantly different from the random relationship. 136 catchments had a better than 50% chance of being different from random correlation, leaving 17 catchments with no better than 50% chance of being spuriously correlated. The self-correlated regression coefficient, as predicted by Equation (iii), varied from 0.2 to 0.999; 69 catchments had an $r^2_{sc} > 0.99$ and the median $r^2_{sc} = 0.95$, i.e. it would be very difficult using the approach of Kenney (1982) to prove anything other than self-correlated data. The ratio of the concentration variance to the flow variance has a median value 2.6% with an inter-quartile range between 0.2 and 13.6%, i.e. the

Formatted: Superscript

calculation is dominated by the flow variance which is the common variable. The spatial distribution of the self-correlation shows no obvious spatial distinction (Figure 2) although two regions showed only self-correlated catchments and they were north Scotland and the Scottish Borders. The region with the lowest proportion of self-correlated catchments was Wales.

Within the 51 catchments identified as having a greater than 95% chance of not being self-correlated a range of behaviours were then identified using the AIC to identify the best-fit response. Of the 51, 18 showed a straight line response, 5 showed a curve response, and 3 catchments showed a sigmoidal (s-curve) response (Figure 3). For 23 catchments the response would better be described as triangular. The straight line responses (Figure 4) all show a line of significant lower gradient than predicted on the basis of randomisation alone. The curved response (Figure 5) was always convex up in the significant results, i.e. as flow increased the flux decreased and this can be interpreted as dilution at higher flows either due to exhaustion of the nitrate supply or the bypassing of the nitrate reservoir. The s-curve response (Figure 6) could be interpreted as a mixture of sources. The triangular responses appear to be bounded by two trends one of which was very close to the line predicted from randomisation and the other at a lower gradient than that predicted for self-correlation (Figure 7). The spatial distribution of the type of response (Figure 3) does suggest a differentiation either side of a north-east to south-west axis with the triangular or mixed response dominating in north west England and Wales while linear and convex up, curved responses dominate in south east England. Such a north west to south east division follows the geology of the UK with younger, more permeable geology dominating in south east England and less permeable, Palaeozoic geology dominating to the north west.

The results from the 126 years of record for the Thames show a mixed response in comparison to the results from the HMS records (Figure 7). Firstly, there was no single response for the Thames and by the standards of this study the results must be dismissed as self-correlated; however, this may belie a more complex response. As was observed for many of the catchments in the HMS dataset for many of the years in the Thames record there is a straight line response between flux and yield at values below that predicted from randomisation but the response is also bound by a sigmoidal response above the line predicted by randomisation.

Applying logistic regression analysis to the binary response defined by the 95% probability of not being self-correlated showed that the probability of not being self-correlated is best-predicted by:

$$log_e\left(\frac{\theta}{1-\theta}\right) = 2.24log_e n + 0.51log_e(Area) - 0.39log_e(Arable) - 10.4 \qquad \text{(iv)}$$

(0.9) (0.25) (0.12) (3.3)

Where θ = the probability that the catchment is not self-correlated; n = the number of years in the record; Area = catchment area (km^2); Arable = area of arable land within the catchment (km^2); Only characteristics found to be significantly different from zero at the 95% probability were included and the values in the brackets below each term are the standard errors in the coefficient or the constant. Equation (iv) correctly classified 75% of the 153 catchments but it should be remembered that, if the equation classified all catchments as being self-correlated, then it would get 67% correct classification. Indeed, Equation (iv) correctly classified 13 out 51 non-self-correlated catchments. The odds ratio suggests that the most important variable is the length of the record (n) followed by the catchment area (Area). However, the odds ratio for the arable variable is less than 1 suggesting that it is only in combination with the other variables that it is significant

Discussion

A number of objections to the idea of self-correlation have been raised. Prairie and Bird (1989) claim that self-correlation for ratio data is not problematic, and will only occur when large measurement errors are present in the variables; that log transformation will reduce spurious correlation; that the variables are meaningful and represent concepts of interest. How the latter is itself spurious as the concept of interest may only have arisen as the result of a spurious correlation and indeed biogeochemical stationarity is a case in point. Lasslop et al. (2009) argued that the case raised by Vickers et al. (2009a) was not a case of spurious correlation given that the component is not part of the derived variable (gross primary production compared to ecosystem respiration), unlike the case where the whole is compared to a part (body weight to liver weight). Vickers et al. (2009b) have refuted the arguments of Lasslop et al. (2009) as irrelevant because self-correlation was demonstrated and that GPP was not measured independently of the ecosystem respiration. Indeed, self-correlated results appear and persist often because they are mechanistically plausible, e.g. one would expect the flux of nitrate to increase with increased annual discharge.

With respect to biogeochemical stationarity, Gall et al. (2013) have shown that biogeochemical stationarity to be mechanistically plausible as sources mix and in-stream processes dominate with increasing scale, leading to decreasing influence of the diversity of behaviours in the headwaters. However, given the possible self-correlated nature of the primary evidence of biogeochemical stationarity, then Occam's razor must apply and so the more complex explanation must not be used until self-correlation has been tested for and rejected. However, the evidence froorm this study of 153 catchments across a range of scales would suggest that even when self-correlation could be rejected, then stationarity is not necessarily the best explanation but rather there may in fact be a range of different explanations.- Gall et al. (2013) predict more biogeochemical stationarity with increase scale and catchment size, however, the reverse was observed in this case. Equation (iv) actually shows that self-correlation was less likely with greater catchment area and self-correlation was less likely if there was not a linear response. This result would suggest that mix of sources becomes more and not less important up to the scales considered by this study, i.e. upto 9885 km^2.

Although straight line responses were commonly found in this analysis the common response was best described as a mixture of sources with different sources operating at different times of the year or at different flow conditions. Jackson and Somers (1991) reminds us that the real hypothesis test when dealing with comparisons with a common variable is the comparison between the line generated from randomisation and the observed line and not just the existence of a significant relationship. Therefore, given the results above, the observation of straight lines is not only that they are best described as straight lines but also they are all at values lower than the randomised line, i.e. nitrate fluxes might appear lower than would be expected and this could simply be that lower concentrations are more likely to be at higher flows, i.e. in general runoff events dilute the main nitrate source. Indeed, whenever a curve response was found then it was always convex up, i.e. as annual water yield increased there was increased dilution of the nitrate responses. However, the most common result was a triangular one where the data were bounded by two trends. Such a response means that it is possible that for any given annual water yield a range of nitrate fluxes could be possible. An easy explanation of this degeneracy is that over time the catchment has changed in terms of the sources of nitrate to the stream, for example, land use has changed from low input pasture to high-input arable; or perhaps there have been improvements at wastewater treatment works in highly urbanised catchments. ~~the convex up results~~

With respect to the prediction of fluxes, use of the self-correlated relationship between flux and yield could ~~be considered to~~ appear to work because, in effect it samples a flux estimate from the distribution of known flux estimates and so therefore it represents a reasonable estimate of the flux, but this estimate could be achieved without reference to annual water yield. Self-correlation might well arise, in these circumstances ~~self-correlation may exist~~ because of the sampling frequency. The scatter in the graph of component flux versus annual water yield~~self-correlation~~ may arise because of limited sampling relative to the nature of the hydrological variation within a monitored catchment and at the~~such~~ low frequency of sampling, common for the catchments in this study, ~~where~~ a sample could be taken at very similar discharges but in very distinct hydrological contexts, e.g. a sample taken on rising limb versus a sample taken on a~~nd~~ recession limb of the storm hydrograph.

Conclusions

This study has proposed a general approach to assess self-correlation (spurious or induced correlation) in situations where there is a common variable and conditions of normality do not hold. Application of this method to nitrate fluvial flux date shows that, even for datasets of more than 10 years, self-correlation was found in 66% of the 153 study catchments. Self-correlation was mainly related to length of record with longer records being less likely to be self-correlated. Amongst those records that were not self-correlated, there were a range of behaviours with the most common being a “triangular” behaviour implying a mixing of sources rather than biogeochemical stationarity.

Acknowledgements

The authors are grateful to Abby Lane and Sarah Wheater of the Environment Agency of England and Wales for supplying the HMS data.

References

Baas, P., Steeneveld, G.J., De Wiel, B.J.H.V., Holtslag, A.A.M., 2006. Exploring self-correlation in flux-gradient relationships for stably-stratified conditions. Journal of Atmospheric Sciences 63, 11, 3045-354.

Basu, N.B., Destouni, G., Jawitz, J.W., Thompson, S.E., Loukinova, N.V., Darracq, A., Zanardo, S., Yaeger, M., Sivapalan, M., Rinaldo, A., Rao, P.S.C., 2010. Nutrient loads exported from managed catchments reveal emergent biogeochemical stationarity. Geophysics Research Letters 37, L23404.

Bellamy, D., Wilkinson, P., 2001. OSPAR 98/3: an environmental turning point or a flawed decision? Marine Pollution Bulletin 49, 87-90.

Burt, T.P., Allison, R.J. (editors) 2010. Sediment cascades: an integrated approach. Wiley-Blackwell, Chichester.

Church, M., Slaymaker, O., 1989. Disequilibrium of Holocene sediment yieldin glaciated British Columbia. Nature 337, 452-454.

Dean, T.J., Cao, Q.V., 2003. Inherent correlations between stand biomass variables calculated from tree variables. Forest Science 49, 2, 279-284.

Francey, M., Duncan, H., Deletic, A., Fletcher, T.D., 2011. Testing and sensitivity of a simple method for predicting urban pollutant loads. Journal of Environmental Engineering 137, 9, 782-789.

Francey, M., Fletcher, T.D., Deletic, A., Duncan, H., 2010. New insight into the quality of urban storm water in south eastern Australia. Journal of Environmental Engineering 136, 4, 381-390.

Gall, H.E., Park, J., Harman, C.J., Jawitz, J.W., Rao, P.S.C., 2013. Landscape filtering of hydrologic and biogeochemical responses in managed catchments. Landscape Ecology 28, 4, 651-664.

Gani, N.H.S., Gani, M.R., Abdelsalam, M.G., 2007. Blue Nile incision on the Ethiopan plateau: pulsed plateau growth, Pliocene uplift, and Hominin evolution. GSA Today 17, 9, 4-11.

Giesler, R., Lyon, S. W., Morth, C-M., Karlsson, J., Karlsson, E. M., Jantze, E. J., Destouni, G., Humborg, C., 2014. Catchment-scale dissolved carbon concentrations and export estimates across six subarctic streams in northern Sweden. Biogeosciences 11, 2, 525-537.

Godsey, S.E., Kirchner, J.W., Clow, D.W., 2009. Concentration-discharge relationships reflect chemostatic characteristics of US catchments. Hydrological Processes 23, 13, 1844-1864.

Gustard, A., Bullock, A., Dixon, J., 1992. Low flow estimation in the United Kingdom. IH Report 108, Wallingford, UK.

Hodgson, J.M., 1997. Soil Survey Field Handbook: Describing and Sampling Soil Profiles. Soil survey Technical Monograph No. 5. Soil Survey and Land Research Centre, Silsoe. England.

Howden, NJK., Burt, TP., Worrall, F., Mathias, SA., Whelan, M.J., 2011. Nitrate pollution in intensively farmed regions: what are the prospects for sustaining high-quality groundwater? Water Resource Research 47, W00l02.

Jackson, D.A., Somers, K.M., 1991. The spectre of “spurious” correlation. Oecologia 86, 147-151.

Formatted: Default Paragraph Font

Jantze, E.J., Lyon, S.w., Destouni, G., 2013. Subsurface release and transport of dissolved carbon of a discontinuous permafrost region. Hydrology & Earth System Science 17, 3827 – 3839.

Johnes, P.J., Heathwaite, A.L., 1997. Modelling the impact of land use change on water quality in agricultural catchments. Hydrological Processes 11, 3, 269-286.

Kenney, B. C., 1982. Beware of spurious self-correlations. Water Resources Research 18, 1041-1048.

Lasslop, G., Reichstein, M., Detto, M., Richardson, A.D., Baldocchi, D.D., 2009. Comment on Vickers at al.: self-correlation between assimilation and respiration resulting from flux partitioning of eddy covaraince CO_2 fluxes. Agricultural and Forest Meteorology 150, 2, 312-314.

Lenahan, M.J., Bristow, K.L., de Caritat, P., 2011. Detecting induced correlations in hydrochemistry. Chemical Geology 284, 182 – 192.

Littlewood, I.G., Watts, C.D., Custance, J.M., 1998. Systematic application of United Kingdom river flow and quality databases for estimating annual river mass loads (1975 – 1994). Science of the Total Environment 210 (1-6), 21-40.

McCuen, R.H., Surbeck, C.Q., 2008. An alternative to spurious linearization of environmental models. Water Research 42, 4033-4040.

Pearson, K., 1897. On a form of spurious correlation which may arise when indices are used in the measurement or organs. Proceedings of the Royal Society 60, 489-498.

Prairie, Y.T., Bird, D.E., 1989. Some misconceptions about the spurious correlation problem in the ecological literature. Oecologia 81, 285-288.

Rodda, J., Jones, N.G., 1983. Preliminary estimates of loads carried by rivers to estuaries and coastal waters around Great Britain derived from the Harmonized Monitoring Scheme. Journal of the Institute Water Engineering and Science 37, 529-539.

Roehl, J.E., 1962. Sediment source areas, delivery ratios and influencing morphological factors. IAHS publication 59, 202-213.

Shivers, D.E., Moglen, G.E., 2008. Spurious correlation in the USEPA rating curve method for estimating pollutant loads. Journal of Environmental Engineering 134, 8, 610 – 618.

Simpson, E.A., 1980. The harmonization of the monitoring of the quality of rivers in the United Kingdom. Hydrological Sciences Bulletin, 25:1, 13-23,

: Stackpoole, S. M., Stets, E. G., Striegl, R. G., 2014. The impact of climate and reservoirs on longitudinal riverine carbon fluxes from two major watersheds in the Central and Intermontane West. Journal of Geophysical Research - Biogeosciences 119, 5, 848-863.

Tetzlaff, B., Friedrich, K., Vorderbrugge, T., Vereecken, H, Wendland, F., 2013. Distributed modelling of mean annual soil erosion and sediment delivery rates to surface waters. Catena 102, 13-20.

US Environmental Protection Agency, 2005. Draft handbook for developing watershed plans to restore and protect our waters. EPA 841-B-05-005, Office of Water, Nonpoint Source Control Branch, Washington, USA.

Vickers, D., Thomas, C. K., Martin, J. G., Law, B., 2009a. Self-correlation between assimilation and respiration resulting from flux partitioning of eddy-covariance CO_2 fluxes. Agricultural and Forest Meteorology 149, 1552-1555.

Vickers, D., Thomas, C. K., Martin, J. G., Law, B., 2009b. Reply to the comment on Vickers et al. (2009): self-correlation between assimilation and respiration resulting from flux partitioning of eddy covariance CO_2 fluxes. Agricultural and Forest Meteorology 150, 315-317.

Worrall, F., Burt, T.P., 2008. The effect of severe drought on the dissolved organic carbon (DOC) concentration and flux from British rivers. Journal of Hydrology 361, 262-274.

Worrall, F., Burt, T.P., Adamson, J.K., 2008. Long-term records of dissolved organic carbon flux from peat-covered catchments: evidence for a drought effect? Hydrological Processes 22, 3181-3193.

Worrall, F., Howden, N.J.K., Burt, T.P., 2012. A 125 year record of fluvial calcium flux from a temperate catchment: Interplay of climate, land-use change and atmospheric deposition. Journal of Hydrology 468–469, 249-256.

Worrall, F., Burt T.P., Howden, N.J.K., Hancock, G.R., 2014, Variation in suspended sediment yield across the UK – a failure of the concept and interpretation of the sediment delivery ratio. Journal of Hydrology 519, 1985-1996.

Figure 1. The study catchments that could be used within this study.

Figure 2. The distribution of sites found not to be significantly self-correlated at the 95% probability in comparison to those found to be significantly self-correlated..

Figure 3. The distribution of the best-fit curve type for those sites which were shown not to be significantly self-correlated.

Figure 4. Example of a linear response between nitrate flux and annual water yield in comparison to the line predicted from randomized data (---). The catchment is the River Lee at Lee Valley Road.

Figure 5. Example of a convex up, curve response between nitrate flux and annual water yield in comparison to the line predicted from randomized data. The catchment is the River Lee at Ware Lock.

Figure 6. Example of an s-curve (ogive) response between nitrate flux and annual water yield in comparison to the line predicted from randomized data. The catchment is the River Severn at Haw Bridge.

Figure 7. Example of an triangular response between nitrate flux and annual water yield. The response predicted for randomized data is shown (---) in comparison to the other proposed bounding trend. The catchment is the River Conwy at Cwm Llanerch.

Figure 8. The nitrate flux vs. annual water yield plot for the River Thames in comparison to the line predicted by randomisation.