RESEARCH

Modeling Complex Concentration-Discharge Relationships with Generalized Additive Models

Claudia von Brömssen[1] · Jens Fölster[2] · Katarina Kyllmar[3] · Magdalena Bieroza[3]

Abstract

Concentration-discharge relationships in water chemical time series can provide important insights into sources, mobilization, and delivery of solutes and particulates into stream networks. The observed relationships are often complex, including nonlinear and hysteretic patterns reflecting seasonal, climatic, and land management changes in biogeochemical release and hydrological transport of solutes and particulates to streams. Using standard single concentration-discharge (c-q) slopes can obscure this wealth of information. In this study, we suggest a new approach using generalized additive models for evaluation of complex c-q patterns in low-frequency water quality data (only monthly or biweekly observations). We used these models to estimate c-q slopes together with their uncertainty, to provide evidence of changes in c-q behaviors and their controls. We estimated c-q slopes for a selection of Swedish streams and evaluated their nonlinear, seasonal, and temporal structure as indicators of changing hydrological or biochemical drivers.

Keywords Concentration-discharge relationship · Generalized additive models · Time-varying relationships · Nonlinearity · Water quality

Highlights

- Complex concentration-discharge relationships (c-q) are often observed.
- Generalized additive models can be a robust tool for analyzing complex c-q data.
- Seasonality in c-q slopes can help identify chemical sources and delivery patterns.
- Identifying trends and variation in c-q slopes can inform management decisions.
- Uncertainty in c-q slopes should be accounted for in decision-making.

✉ Claudia von Brömssen
Claudia.von.Bromssen@slu.se

1 Division of Applied Statistics and Mathematics, Department of Energy and Technology, Swedish University of Agricultural Sciences, PO Box 7032, 750 07 Uppsala, Sweden

2 Section for Geochemistry and Hydrology, Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, PO Box 7050, 750 07 Uppsala, Sweden

3 Agricultural Water Management, Department of Soil and Environment, Swedish University of Agricultural Sciences, PO Box 7014, 750 07 Uppsala, Sweden

1 Introduction

Concentration-discharge (c-q) relationships evaluated on a storm-event, seasonal, and multi-annual basis provide valuable information on the functioning of individual streams, but also entire stream networks and their catchments [1–3]. They can be used to identify dominant sources, mobilization, and delivery pathways for a wide range of solutes and particulates [4], which in turn can help to adjust water quality monitoring frequencies [5] and plan and evaluate the effectiveness of catchment management interventions [6]. The slope (b) of a c-q relationship on the logarithmic scale is typically used to identify dominant sources and delivery pathways and is often categorized into three dominant patterns: chemostatic (b approx. 0), chemodynamic leading to dilution ($b < 0$), or to concentration ($b > 0$) [7, 8]. This information is used to link estimated slopes with stream and catchment properties in order to identify the sources and behavior of solutes and particulates [3, 6–8].

Early work on c-q relationships [7, 9, 10] focused mainly on understanding spatial differences between catchments and solutes, based on single and constant c-q slopes calculated from readily available low-frequency (seasonally to weekly) water quality data. More recently, the widespread availability of longer time series, often obtained at higher monitoring

frequencies (sub-daily), has allowed for better and more detailed evaluation of c-q relationships and their slopes, and a greater focus on their temporal dynamics. Accounting for temporal dynamics in c-q relationships provides a better understanding of underlying drivers, including changes in the relative proportions of different sources of solutes/particulates (e.g., point vs. diffuse sources) and processes (e.g., hydrological vs. biogeochemical) [11], and seasonally variable weather conditions and management practices.

Due to greater c-q data availability, several approaches for improved estimation of non-constant c-q slopes have been suggested [2, 3]. For example, separate slope estimates can be determined for individual sub-datasets where the data are split according to discharge, using (1) the baseflow recursive filter method [3], (2) the inflection point in the c-q relationship method [6], or (3) the median of discharge threshold levels [12]. More advanced approaches include, e.g., piecewise linear regression to determine the c-q slope below and above a flow discharge threshold that is determined by iterative maximum likelihood fitting [4, 13], or a combined model for baseflow and quick flow using a two-component mixing equation [14]. Recently, nonlinear c-q relationships have been investigated using smoother splines [15] to determine if the underlying c-q relationship is linear, convex, or concave, in order to link this behavior with in-stream uptake processes. To evaluate changes in c-q relationships over time, data series are often divided into two or several non-overlapping sub-series, for which the c-q slope is determined separately. For example, Bieroza et al. [6] split series in half when studying the effect of a trend and Fork et al. [16] calculated c-q slopes for each calendar year to illustrate changes over time and linked them to catchment properties. For most of the approaches mentioned above, uncertainties, e.g., standard error or confidence band for c-q slopes, are not provided or even discussed.

Existing techniques fail to address several important issues that could be improved when estimating c-q relationships. One approach that accounts for a number of features, i.e., temporal trends, nonlinear relationships, and time-varying relationships, is weighted regressions on time, season, and discharge (WRTDS) [17–19]. Hirsch et al. [20] developed a bootstrap approach for determining the uncertainty of trend estimation for the WRTDS model and that approach was adapted by Zhang et al. [21] to present uncertainties for c-q slopes. A disadvantage with the WRTDS method is that it relies on daily data and cannot be used directly for low-frequency data. Time-varying and nonlinear relationships were addressed by Zimmer et al. [22] and Fazekas et al. [23] using a moving window approach, which also applied to high-frequency data. Generalized additive models (GAMs) [24, 25] are a commonly used tool to estimate complex relationships between a response and one or several explanatory variables. The form of the relationship is completely determined by the available data and thus

does not require any prespecified functional form, e.g., linear or exponential, as parametric regression models do. The most common use of GAMs is to model temporal trends in concentration time series [26–28]. Beyond flexible nonlinear relationships, GAMs can also estimate relationships that vary in time, using so-called time-varying coefficient models [29] and are sometimes used to study relationships between water quality concentrations and influencing variables [30, 31]. However, the method has so far not been explicitly applied to study or quantify c-q relationships.

In this study, we examined how GAMs can improve estimation and evaluation of complex c-q relationships compared with the standard single c-q slope approach. As an example, we used low-frequency (monthly or biweekly) data from several Swedish agricultural streams exhibiting nonlinear and/or time-varying c-q slopes. We examined whether explicit inclusion of temporal trends in concentration was necessary to obtain reliable estimates of c-q slopes and whether changes in flow needed to be addressed in interpretation. Estimates of c-q slopes for nitrate-nitrogen and total phosphorus were plotted in informative diagrams, together with the corresponding uncertainties. These diagrams can be a useful tool to support interpretation of solute and particulate behavior in response to climatic and management drivers.

2 Methods

2.1 Estimation of Constant c-q Slopes Using Regression Models

Power-law relationships, of the form $C = A \cdot Q^b \cdot E$, are typically assumed between concentration and discharge (see, e.g., [8]), where C is concentration, Q is discharge, and E is the model's random error term. The concentration-discharge slope is denoted b. For estimation of c-q slopes, this relationship is log-transformed to:

$$\log(C) = \log(A) + b \cdot \log(Q) + \log(E) \tag{1}$$

Substituting c, a, q, and ε for $\log(C)$, $\log(A)$, $\log(Q)$, and $\log(E)$, respectively, gives the following model ($\mathrm{LM_{constant}}$):

$$c = a + b \cdot q + \varepsilon \tag{2}$$

The c-q slope, b, can be estimated by a linear regression model and is then assumed constant. It describes the change in c when q increases by one unit. The intercept, a, indicates the level of c when q is equal to 0. The standard error of b, i.e., the uncertainty of the c-q slope, is computed as:

$$SE_b = \sqrt{\frac{s_\varepsilon^2}{\sum_i (q_i - \bar{q})^2}} \tag{3}$$

where s_ε^2 denotes the residual variance (i.e., the variance of the error term ε in the regression model), q_i are the individual observations of q, and $\overline{q}$ is the overall mean of q. These standard errors can be used to determine confidence intervals for b or to test if the true value of c-q slope is significantly different from zero.

The dataset used to estimate the c-q slope is usually a time series, which means that using a conventional regression model to determine the slope b and its uncertainty is not optimal, as the available observations are not independent. Instead, the model must be extended to include a time series structure in the error term, e.g.:

$$c = a + b \cdot q + \varepsilon \ \ \varepsilon \sim N(0, \Lambda) \tag{4}$$

where $\varepsilon \sim N(0, \Lambda)$ denotes the distribution of the error term, which is assumed to be normal with mean 0 and a variance–covariance matrix Λ. The typical choice for this variance–covariance matrix is to assume that correlations between two different observations in time are a function of the time difference. This is also called an autoregressive process of lag 1. The correlations can be estimated in the modeling process using a mixed model or generalized least squares estimates. Inclusion of correlations in time (autocorrelation or serial correlation) in the model improves the estimate for the uncertainty in c-q slope coefficients (see, e.g., [32]).

2.2 Generalized Additive Models

Generalized additive mixed models facilitate modeling of environmental time series without prior definition of the shape between the response variable, c, and any available explanatory variable, e.g., q or time [24, 25]. Relationships of a concentration variable to time or to explanatory variables are assumed to be smooth, and cross-validation is used to determine the best fit. All GAMs are fitted using the package *mgcv* [33, 34] in R software [35]. Derivatives of smooth functions are computed using the package *gratia* [36].

2.3 Estimation of Constant *c-q* Slopes While Accounting for a Trend in Concentration

A trend in concentration data can often be observed in addition to a relationship between concentration and discharge. Ignoring such trends could influence the accuracy of estimation of c-q slope and therefore it is advisable to include it in all models when estimating the c-q relationship. In the basic model, a trend is added, while assuming that the c-q slope is constant, using the following generalized additive model (GAM_constant):

$$c = a + f_1(time) + b \cdot q + \varepsilon \ \ \varepsilon \sim N(0, \Lambda) \tag{5}$$

where the trend component $f_1(time)$ is a smooth function in time. The smooth function can be modeled as a thin plate spline or cubic regression spline. As a trend in concentration will not always be present, we chose a thin plate spline with a smoothing penalty (bs = "ts"), a so-called shrinkage factor, which allows the model to remove the trend component when not necessary. The slope estimate b was fitted and interpreted as in conventional regression models. An autocorrelation estimate was incorporated in the same way as before, which is necessary to estimate the uncertainty of the slope appropriately and to avoid too much wiggliness, i.e., too much variation, in estimated trends, which would make them more difficult to interpret [37]. Uncertainty estimates of the c-q slope b can be extracted from the model, as usual in regression models, and a test for the hypothesis that the slope coefficient differs significantly from zero is given as the standard model output in R. Confidence intervals for the estimate b can be computed from model results.

2.4 Estimation of Nonlinear *c-q* Relationships

If the c-q slope cannot be assumed to be constant over the range of all observed discharge measurements, GAMs allow estimation of a smooth relationship between concentration and discharge. This can be achieved by the model (GAM_nonlinear):

$$c = \mu + f_1(time) + f_2(q) + \varepsilon \ \ \varepsilon \sim N(0, \Lambda) \tag{6}$$

where the concentration-discharge relationship is now estimated by the smooth function $f_2(q)$. An attractive feature of GAMs is that if $f_2(q)$ is in fact linear, rather than curved, the model will identify it as such and suggest a constant c-q slope as the best alternative. As this model only contains smooth terms, the intercept a is replaced by μ, which represents the overall mean of the concentration variable. Other considerations, such as the autoregressive structure of the error term, remain as for the models presented earlier (Eqs. 4 and 5).

The c-q slope b can be seen as the derivative of the estimated c-q relationship $f_2(q)$, both when the c-q relationship is linear and when it is smoothly changing with discharge. First derivatives can be computed by finite differencing, and their uncertainties can be deduced from the variance–covariance matrix of the original smooth function, e.g., see documentation for the *predict.gam* function in the *mgcv* package in R [33, 38]. Based on such uncertainty computations, confidence bands can be obtained for the derived c-q slope and can be used to determine whether the slope indicates a chemostatic or chemodynamic relationship at different levels of discharge. For individual values of observed discharge, it is possible to check, e.g., if the entire confidence band of the derivative lies below zero (chemodynamic, dilution pattern),

lies above zero (chemodynamic, concentration pattern), or is not significantly different from zero (chemostatic).

Nonlinear *c-q* relationships may show different behavior during different seasons, due to, e.g., differences in temperature, leading to varying nutrient cycling efficiency or agricultural practices. For low-frequency data, a way to investigate this is to define a few distinct sub-datasets, e.g., for the four seasons, and then estimate *c-q* slopes separately for each season. For this, model $GAM_{nonlinear}$ is extended by a categorical variable indicating season, here denoted $seas_{cat}$, and its interaction to discharge, giving the model $GAM_{nonlinear\ and\ seasonal}$:

$$c = \mu + f_1(time) + seas_{cat} + f_2(q) \cdot seas_{cat} + \varepsilon \ \ \varepsilon \sim N(0, \Lambda) \tag{7}$$

This leads to separate nonlinear *c-q* slopes for each of the seasons. Derivatives and their uncertainties are again computed to determine the slope estimate and its confidence bands.

Another possibility is to let the nonlinear *c-q* relationships vary smoothly with month and with discharge, leading to a two-dimensional smooth function. This approach is not discussed further here, because extraction of the slope coefficients from such a model is more cumbersome as it relies on partial derivatives. In addition, the approach requires more data and is probably not a feasible choice with low-frequency datasets.

2.5 Estimation of Time-Varying *c-q* Relationships

Time-varying coefficient models allow relationships between two variables to change with time [29]. This can be relevant if the *c-q* relationship changes over time due to, for example, a change in relation between point and non-point sources. To estimate a time-varying *c-q* slope, we can use the following model ($GAM_{time-varying}$):

$$c = a + f_1(time) + b(time) \cdot q + \varepsilon \ \ \varepsilon \sim N(0, \Lambda) \tag{8}$$

where $f_1(time)$ represents the trend in concentration, $b(time)$ indicates that the *c-q* slope (*b*) varies smoothly with time, while at each specific point in time it is assumed that the relationship between concentration and discharge is linear. Autocorrelation for the error term is included in the same way as before. Uncertainties for the *c-q* slope and for the trend estimate can be extracted directly from the fitted model for any given time point.

Seasonality can be included in the model by defining a two-dimensional smooth function ($GAM_{time-varying\ and\ seasonal}$) using both year and month as indicator of time:

$$c = a + f_1(year, month) + b(year, month) \cdot q + \varepsilon \ \ \varepsilon \sim N(0, \Lambda) \tag{9}$$

i.e., the *c-q* slope varies smoothly over time (years) and over seasons (month). For the monthly component, it is common to use a circular smooth function, e.g., a circular cubic regression spline, to guarantee that December and January values are connected without a shift. It is recommended that the two-dimensional smooth functions are tensor products [25, 39], in order to allow different amounts of smoothing in the two directions (year and month). However, this type of model is rather complex and relies on sufficient data to produce stable estimates, so it is advisable to use it on series with more than one observation per month and year. Autocorrelation for the error term is included in the same way as in other models.

2.6 Mean-Centering of Discharge

An appealing feature of the nonlinear models $GAM_{nonlinear}$ and $GAM_{nonlinear\ and\ seasonal}$ is that they present the overall (log-transformed) concentration mean as model output. The other models present the mean at levels when the log-transformed discharge is equal to 0, i.e., when discharge is equal to 1. Mean-centering of *q* can be used to adjust these models to estimate the overall mean μ, e.g., for the $GAM_{time-varying}$ model:

$$c = \mu + f_1(time) + b(time) \cdot \left(q - \overline{q}\right) + \varepsilon \ \ \varepsilon \sim N(0, \Lambda) \tag{10}$$

Model estimates, i.e., time-varying slope and the trend component, are not affected by this change, other than that the optimization routine can produce slightly altered values as *q* is rescaled. We used mean-centering of *q* for estimation of time-varying *c-q* slopes.

2.7 Hydrochemical Data

We applied the proposed models to examine *c-q* relationships in water quality (total phosphorus (TP), nitrate-nitrogen (NO_3-N)), using flow discharge data from the Swedish Environmental Monitoring Programme on larger rivers (https://miljodata.slu.se/MVM/) and small agricultural catchments (http://jordbruksvatten.slu.se) (Table 1).

The three catchments included in this study vary in size, agricultural land use, and catchment properties (Table 2). The Fyrisån catchment is the recipient of treated sewage water from the city of Uppsala (~200,000 inhabitants). At the city's sewage treatment plant, improvements in phosphorus removal were made in 1972 and in nitrogen removal in 1999 [40]. Water quality measurements in Fyrisån are made on a monthly basis and daily modeled water flow adjusted for local measurements is downloaded from SMHI (https://vattenwebb.smhi.se/nadia/). For the other two catchments, O14 and H29, with a high proportion of agricultural land,

Table 1 Time series used in the study

Catchment	Parameter	Non-linear	Time-varying	Seasonal	Model/s used	Reason
O14	Nitrate-nitrogen (NO_3-N)	Yes	No	Yes	$GAM_{nonlinear}$ $GAM_{nonlinear\ and\ seasonal}$	Multiple N sources resulting in nonlinear relationship with discharge
Fyrisån	Total phosphorus (TP)	No	Yes	No	$GAM_{time\text{-}varying}$	Removal of point sources could change the *c-q* relationship over time
H29	Nitrate-nitrogen (NO_3-N)	No	Yes	Yes	$GAM_{time\text{-}varying\ and\ seasonal}$	Changes in agricultural management over time

samples are taken at least biweekly as grab samples and water flow is measured every quarter hour and averaged to daily values [41]. For these catchments, both concentrations and discharge were collected within the environmental monitoring program on small agricultural catchments.

The discharge and concentration data were both transformed using a base 10 logarithm before analysis. The datasets and all R code used for models and figures are available online [42].

2.8 Simulation Study

To illustrate the effect of a prevailing trend in concentration on estimation and variation of modeled *c-q* relationships, we performed a simulation study. For this, we generated a discharge variable (*q*) assuming a normal distribution with mean 0 and variance 1, and obtained the value of the response variable from the model $c = \beta \cdot q + r$, where *r* is a random error term that is also normally distributed, with mean 0 and variance 1. We set the value of *β* to 0.3 and simulated 1000 time series for three distinct scenarios: (i) no trend in concentration or discharge; (ii) a trend in concentration; and (iii) a trend in discharge. Any trend simulated was linear, with an increase of 0.1 units per time point. We then fitted each of the simulated series by an $LM_{constant}$ and a $GAM_{constant}$ model, to evaluate how well the true *c-q* slope could be estimated in the three scenarios.

In the absence of a trend in either concentration or discharge (scenario i), the model that did not include any trend estimate ($LM_{constant}$) and the model that allowed explicit estimation of a trend in concentration ($GAM_{constant}$) both gave very similar results (Fig. 1, left). In fact, in 86% of the simulations the two models provided exactly the same estimate, while in the remaining cases the difference was at most 0.05 units. When a trend in concentration was simulated (scenario ii) but ignored in estimation of the *c-q* slope ($LM_{constant}$), the variation in estimates increased substantially (Fig. 1, center). In this case, $GAM_{constant}$ can be considered the correct model and $LM_{constant}$ provided identical estimates in 50% of all simulated series, while the remaining estimates differed by up to one unit from the target value. When data exhibited a trend in discharge (scenario iii), both models again gave comparable results and 82% of the simulations led to identical values of the *c-q* slope. In 65% of these cases, the $GAM_{constant}$ model correctly identified that there was no concentration trend present and removed this component in the estimation. However, in some simulated datasets, the estimated *c-q* slope deviated from the correct slope for this scenario (determined by $LM_{constant}$) by up to 0.3 units. The amount of variation observed in the *c-q* slope generally depended on the magnitude of the trend and on the variation in concentration and discharge (not shown).

3 Results

3.1 Estimation of *c-q* Relationships in the Presence of a Trend in Concentration Data

Constant *c-q* slopes for three time series were analyzed with the two models $LM_{constant}$ and $GAM_{constant}$ to illustrate the effect of a trend in concentration in actual hydrochemical
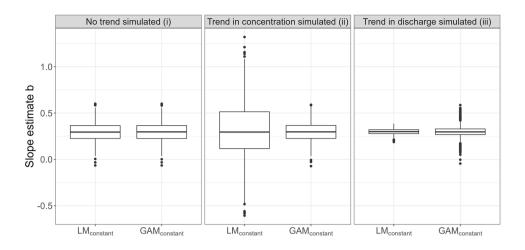
Table 2 Geographical, land use, physical, and chemical characteristics of the three selected catchments

Catchment	Area (km^2)	Location	Soil texture	Agricultural land (%)	Water discharge (mm)[a]	Nitrogen (mg L^{-1})[a]	Phosphorus (mg L^{-1})[a]	Water quality monitoring period	Sampling frequency
O14	10.1	West Sweden	Silt loam	72	299	4.7	0.17	1993–2020	Biweekly
Fyrisån	2006	East Sweden	Clay	25	232 (230[b])	2.7 (3[b])	0.07 (0.12[b])	1965–2020	Monthly
H29	7.0	East Sweden	Sandy loam	73	124	9.1	0.15	1995–2020	Biweekly

[a]Mean values computed for 1995–2020

[b]Mean value computed for 1965–2020

data. One of the series, TP concentration in Fyrisån, showed distinct trends in concentration (Fig. S1 in SI) and when this trend was not accounted for the estimation of the *c-q* slope was slightly lower (0.012 vs. 0.016) (Table 3). While the difference was not large, the R^2 values indicated a substantially better fit when the trend was included in the model. Catchment O14 did not show any clear temporal trend in NO$_3$-N, and nor did catchment H29, but in the latter catchment discharge decreased over time (Table S1, Fig. S2). The *c-q* slope estimated with the two models was very similar in these catchments.

Based on the single slope *c-q* relationships, catchments O14 and H29 showed chemodynamic concentration patterns, while the *c-q* relationship in Fyrisån was chemostatic. However, as we will show in the following sections, these relationships were more complex than implied by constant *c-q* slope.

3.2 Estimation of Nonlinear *c-q* Relationships

On visualizing the relationship between log-transformed discharge and NO$_3$-N in catchment O14 using the GAM$_{nonlinear}$ model, we found that the estimated *c-q* relationship was not completely linear, with flattening at high-flow conditions (Fig. 2, left). This was better illustrated by plotting

the estimate of the *c-q* slope (*b*) as function of discharge (Fig. 2, right). The estimated relationship was chemodynamic (concentration) for low-flow conditions, with an estimated *c-q* slope of almost 0.8 for the lowest discharge, but chemostatic at high flow (above log-transformed discharge of 2). The uncertainty of the estimated slope coefficient was quite large, especially for very low values of discharge as a consequence of lower data availability. The adjusted R^2 value for the model (0.46) was slightly improved compared with the linear model ($R^2 = 0.38$). There was some indication of residuals having higher values during autumn and lower during spring and summer, meaning that there was some seasonality in the data that could not be captured in this model. In catchment O14, discharge did not show any temporal trend (Table S1, Fig. S2 in SI).

To study variation in *c-q* slopes due to seasonal effect, we calculated separate nonlinear *c-q* relationships for different seasons (spring: March–May, summer: June–August, autumn: September–November, and winter: December–February) using the GAM$_{nonlinear and seasonal}$ model. There was some difference in the *c-q* slopes identified for spring compared with other seasons during low- and high-flow conditions (Fig. 3, left). During spring, the *c-q* slope at low discharge was quite high (~ 2.5), but with little data support and broad uncertainty bands at discharge levels below

Table 3 Constant *c-q* slope (*b*) and its standard error estimated from models without and with inclusion of a concentration trend (LM$_{constant}$ and GAM$_{constant}$), respectively

Site	Variable	Trend estimate not included in the model (LM$_{constant}$)			Trend estimate included in the model (GAM$_{constant}$)		
		Slope *b*	Standard error of *b*	Model R^2 (adj)	Slope *b*	Standard error of *b*	Model R^2 (adj)
O14	Nitrate-nitrogen (NO$_3$-N)	0.34	0.019	0.33	0.35	0.018	0.38
Fyrisån	Total phosphorus (TP)	0.012	0.025	0	0.016	0.016	0.58
H29	Nitrate-nitrogen (NO$_3$-N)	0.78	0.028	0.56	0.82	0.029	0.60

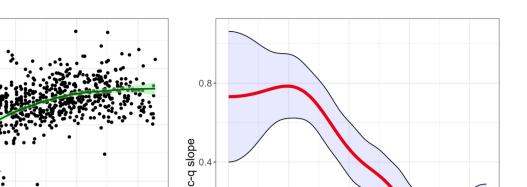**Fig. 2** (Left) Log-transformed nitrate-nitrogen (NO_3-N) concentration as a function of log-transformed discharge from catchment O14 based on a smooth, nonlinear c-q relationship and (right) estimate of the c-q slope as a function of log-transformed discharge

log-transformed discharge of 0. The c-q relationship for spring remained chemodynamic for almost all flow conditions. In contrast, the estimated pattern for summer and autumn indicated smaller c-q slopes for low-flow conditions and chemostatic patterns for high flow (Fig. 3, center and right). The estimated c-q slope for winter was constant, with a value of 0.08 (standard error: 0.03). As the seasonal curves were based on fewer observations compared with whole dataset, the uncertainty around the estimated c-q slope was broader than in the non-seasonal model (Fig. 3, lower row).

Overall, the $GAM_{\text{nonlinear and seasonal}}$ model gave more consistent predictions than the constant slope model (GAM_{constant}) (Fig. S4), as the latter did not reproduce very low or high concentrations accurately. The $GAM_{\text{nonlinear and seasonal}}$ model also had a higher adjusted R^2 value (0.55), but showed increased variation for low concentration values, which usually occur in low-flow conditions that are not observed often. As the seasonal model was considerably more complex, the risk of overfitting must be taken into account. For example, there were only three observations above a log-transformed discharge of 3 during summer months and the modeled curve adapted to these by bending downwards (Fig. 3, upper row, center). It cannot be assumed that the model fit would be the same if there were more observations available within this discharge range.

3.3 Estimation of Time-Varying c-q Relationships

Point sources of TP to the Fyrisån catchment were significantly reduced in the 1970s leading to a clear downward trend in concentrations (see Fig. S1). c-q relationships were plotted for five separate sub-periods (Fig. 4, left), confirming that also the c-q relationship changed over time. It was negative during the 1960s and 1970s, while it turned positive during later years. To model this time-varying c-q relationship explicitly, we used $GAM_{\text{time-varying}}$. The c-q slope estimates obtained exhibited large changes, from approximately -0.5 in early 1970 to almost 0.2 in 2009 (Fig. 4, right). A minor decrease in the c-q slope was observed for the most recent decade. We used the uncertainty of the estimates to determine whether the slope parameter b was significantly different from zero at different time points, i.e., whether the entire confidence band lay below or above zero. Using that approach, we found that Fyrisån showed a clear dilution pattern between the start of the series and the late 1970s, while a concentration pattern prevailed after about 1990.

Comparison of observed and predicted concentrations (Fig. S4) indicated that $GAM_{\text{time-varying}}$ provided more consistent results, and thus more reliable estimates of the c-q slope than the model with constant slope (GAM_{constant}),

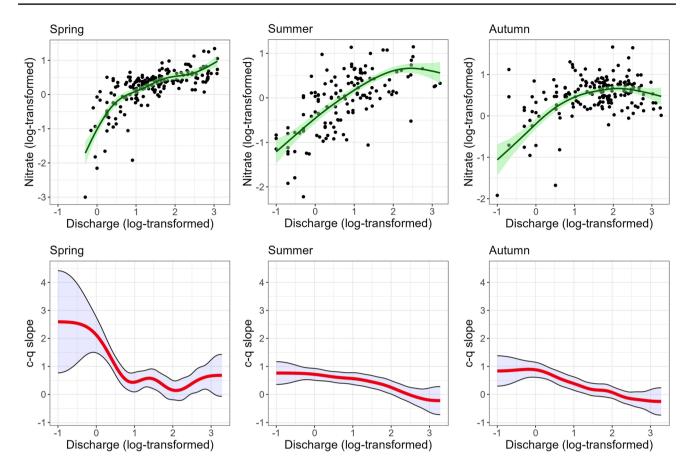**Fig. 3** (Upper row) Log transformed nitrate-nitrogen (NO$_3$-N) concentration as a function of log-transformed discharge from catchment O14 in (left) spring, 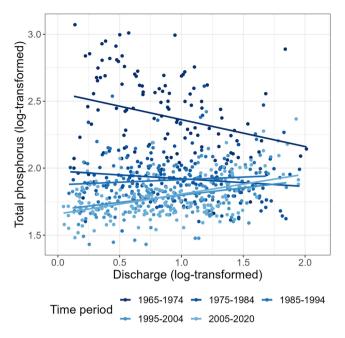(center) summer, and (right) autumn, based on a nonlinear *c-q* relationship and (lower row) nonlinear *c-q* slope estimates as a function of log-transformed discharge for the same three seasons

even though discharge exhibited a slight downward trend over time (Fig. S2, Table S2). The adjusted R^2 value of the GAM$_{time-varying}$ model was higher (0.64) and Akaike information criterion (AIC) was lower (-326) than for the constant slope model ($R^2 = 0.59$, AIC $= -295$).

For catchment H29, which was monitored with at least two observations per month, we estimated the *c-q* relationships using a two-dimensional smooth function (GAM$_{time-varying\ and\ seasonal}$), i.e., allowing the *c-q* slope to vary smoothly over years, but also over months. We found that the highest estimates for the *c-q* slope lay around 1 during late summer months in the 1990s (Fig. 5, left, in yellow) and then decreased steadily over the years for all seasons, with the most pronounced changes in early autumn and early spring.

To analyze this further, *c-q* slopes and their uncertainty estimates were extracted for two specific times in the year and presented as changes over years together with confidence bands (Fig. 5, right). The estimates and confidence bands for mid-February and mid-August suggested that the relationship was chemodynamic, with a concentration pattern for both months and all years. The *c-q* slopes during August were generally of much higher magnitude, but decreased strongly after 2005 and reached similar *c-q* slope values as in February towards the end of the series. The *c-q* slope for February was estimated to be rather stable over the years, but with a dip around 2010. The uncertainty estimates for the slopes revealed a clear significant difference between estimated *c-q* slope magnitude for the two selected months except during the last years. Residuals were normally distributed, but showed two outliers during November 2015, when unusually low NO$_3$-N concentrations were recorded. The modeled trend component also showed strong seasonality, as well as a decrease in NO$_3$-N levels (Fig. S3 in SI).

Discharge in catchment H29 exhibited a clear trend over the years (Fig. S2, Table S1), with low values during summer and autumn in the end of the series. This could have affected the efficiency of the estimated slope and uncertainty parameters. Within the applied model, we allowed the trend component to decline to zero if no trend in concentration was present and the *c-q* slope to reduce to linear if that represented the best fit. A clear trend component was
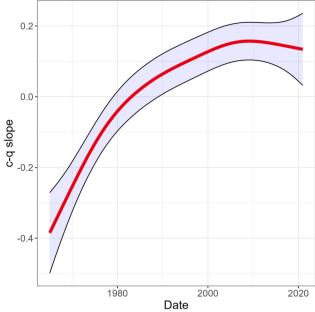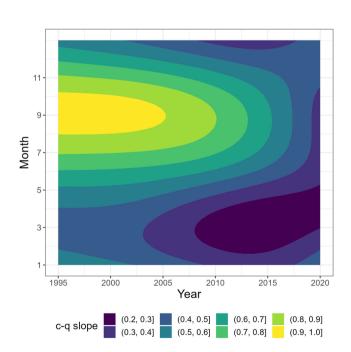
Fig. 4 Concentration-discharge relationships for the Fyrisån catchment, 1965–2021. (Left) Log-transformed total phosphorus concentration as a function of log-transformed discharge with darker colors indicating observations in the beginning of the observed series and lighter colors indicating observations towards the end. (Right) Estimated c-q slope changing smoothly over the years

still identified (effective degrees of freedom 9.5) and the c-q slope was also determined to not be linear (effective degrees of freedom 10.5). To further determine the difference between a constant c-q slope model and the current

model, $GAM_{time\text{-}varying\ and\ seasonal}$ was compared with a modified $GAM_{constant}$ model that allowed a trend component to vary smoothly with both month and year, while the c-q slope was held constant. The $GAM_{time\text{-}varying\ and\ seasonal}$
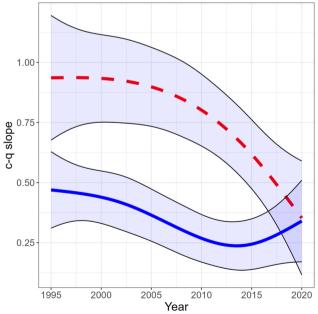

Fig. 5 (Left) Variation in concentration-discharge relationships for nitrate-nitrogen (NO_3-N) in catchment H29 over year and month, where colors indicate the value of the c-q slope. (Right) Examples of estimated c-q slopes for mid-August (red, dashed) and mid-February (blue, solid), and associated confidence bands (shaded)

had an adjusted R^2 value of 0.76 and AIC = 239, while the modified GAM$_{constant}$ model had an adjusted R^2 value of 0.73 and AIC = 280. All parametric parameters and smooth functions in the two models were highly significant. On comparing the predictions for the log-transformed concentrations of NO$_3$-N, we observed that the constant slope model slightly overestimated concentrations at high flow, while it underestimated concentrations at medium flow (Fig. S4). GAM$_{time-varying and seasonal}$ generally performed better, but both models had difficulties in reproducing concentrations at low flow.

4 Discussion

To quantify c-q relationships from hydrochemical data, standard linear regression models are often used [7, 12]. They assume a constant slope in time and over the range of different flow conditions, which can be relevant if data are limited, e.g., cover a short monitoring period. However, in many cases low-frequency data contain a greater wealth of information about complex c-q relationships, which can be extracted using adequate statistical tools.

4.1 Complex *c-q* Relationships Estimated Based on Low-Frequency Data

In the examples in this study, we showed that non-stationary and nonlinear c-q relationships can be investigated using datasets with low monitoring frequencies. Generalized additive models allow smooth nonlinear and time-varying c-q relationships to be computed, which can improve the ability to identify and interpret the c-q patterns. The models used here are generally well-documented and easily available in the free software R [33, 35]. The c-q slopes calculated with GAM are simple to extract and are by default accompanied by uncertainty estimates, which is important when c-q slopes are used in decision-making. The uncertainty values might underestimate the total uncertainty, since they do not include uncertainty connected to selecting the model smoothness [43], but this is still an improvement compared with conventional approaches where c-q slope uncertainty is usually simply ignored.

Using GAM models, we can model c-q slopes that vary smoothly over time or nonlinearly with the level of flow discharge. Thus, these methods do not suffer from any shift by division into subsets, e.g., based on flow discharge thresholds, which is often performed [6, 12, 16]. This makes it easier to compare results from different studies and reveal some general hydrochemical patterns valid for catchments differing in size, climate, soils, and land use. Another advantage of the smoothly varying c-q slopes is that they simplify analysis when the effects of mitigation measure on water quality are delayed due to, e.g., increased hydrological residence time, pollutant retardation, or ecosystem linkages

[44]. For conventional constant c-q slopes, it is often necessary to define a time period when these delayed effects are expected to be observable, while with the suggested time-varying c-q slopes any change will be visible when it arises.

The proposed methods can be used on hydrochemical time series that have low monitoring frequency, but special care needs to be taken to avoid overfitting. Overfitting occurs when too complex a model is fitted on too small a dataset, e.g., if no or only few observations within a range of flows are available, which is often the case with manual water quality sampling [5]. As a result, the model might fit the available observations well, but would not be able to generalize to new data. For example, in this study we observed signs of overfitting in models that included strong seasonality patterns in addition to nonlinear c-q relationships. In catchment O14, where only three observations were available for high flows during summer, these observations determined the behavior of the c-q slope under high-flow conditions. Additional measurements at high flows, e.g., using high-frequency water quality analyzers and sensors [45], can fill this gap. The uncertainty bands for c-q slopes give some indication about the reliability of the estimates produced, but ultimately it is up to the user to identify potential overfitting issues. Studying the visualizations of c-q relationships side-by-side with the available data is important as a model validation tool.

Untangling time-dependent structures is always a complex task for single time series, and complex models fitted on low-frequency data can lead to a risk of creating artifacts, e.g., by overfitting. With GAMs, it is possible to include a component describing trends in concentrations that can be shrunk to zero if no concentration trend is present in the data. Additionally, thin plate splines used in GAMs reduce to linear relationships if the modeled c-q slopes are not more complex than that. This means that, while the models are flexible and can model complex c-q relationships, they can be expected to provide simple relationships if these are the best fit. For additional model validation, we compared the proposed models with simpler models with constant c-q slopes using adjusted R^2, AIC, and visual inspection of residuals and predictions. The results confirmed that nonlinear and time-varying c-q slopes led to overall better models for our data.

4.2 Trends, Nonlinearities, and Seasonality in *c-q* Slopes

For many hydrochemical time series, using a single linear c-q slope can lead to faulty interpretation of hydrochemical behavior. This was evident in the case of TP data for the Fyrisån catchment, where a single constant c-q slope missed most of the hydrochemical changes that occurred due to substantial improvement in wastewater treatment over more than 50 years of monitoring. In other situations, the method could be used to evaluate or identify effects of changes that cannot

be directly quantified, e.g., due to climate change, agri-environmental mitigation measures, or stream channel interventions. Including data covering different seasons is often critical for interpretation of c-q patterns, their variation, and underlying drivers, and ignoring seasonality has been shown to pose a risk of misclassification [23]. For datasets with biweekly or more frequent data, we found seasonally varying c-q slopes both when analyzing changes over time and when evaluating nonlinear concentration-discharge relationships for NO_3-N in small agricultural catchments (O14, H29). Only such detailed analysis allows identification of the connection between estimated c-q slopes and seasonally varying potential drivers, seasonal climatic conditions (droughts, storms), or agricultural practices that change over the course of the year.

Studying nonlinear c-q slopes makes it possible to determine whether increased discharge affects loss of nutrients differently for low-flow and high-flow conditions. For example, in catchment O14 the c-q relationships showed nonlinear seasonal behavior, with chemodynamic relationships in high-flow conditions during spring, but chemostatic relationships at high-flow conditions in summer, autumn, and winter. This may imply that heavy precipitation events in the beginning of the crop-growing season, when nitrogen uptake is low, increase the risk of losses of nitrogen mineralized in the soil and applied as fertilizer. In contrast, even intense precipitation would lead to smaller nitrogen losses in summer, probably because almost all applied nitrogen had already been assimilated by the crop. This type of information could be used in practice, e.g., when selecting optimal periods for stream channel interventions like dredging or vegetation removal and for optimizing the timing of agricultural operations.

Seasonal and time-varying c-q slopes could also be investigated by other statistical methods that allow more variability over the year or over a longer period, e.g., a moving window approach [22, 23]. However, when only low-frequency data are available, it will not be possible to quantify rapid changes caused by, e.g., storm events, and the approach involving GAM is then useful as it smooths out short-term variation.

4.3 Interpreting *c-q* Relationships

We focused our analysis of c-q slopes over time on three study catchments, for which we expected the c-q slope to be non-constant due to changes in point sources or agricultural management over time and due to the presence of multiple nitrogen sources. While the models we tested identified the expected changes in c-q slopes and usually gave better predictions, and thereby more reliable estimates of c-q slopes, interpretation of c-q patterns is a complex task. If the underlying processes are not yet known, time-varying

and nonlinear c-q slopes can be used as support for source apportionment [3, 6, 23, 46]. Using the approaches we present, it is also possible to separate drivers that influence concentration levels from those that influence c-q relationships, as concentration trends are modeled separately. Trends in discharge, on the other hand, can increase the variation in slope estimates and, in the worst case, lead to flow levels that have not been observed earlier, making purely data-driven interpretation difficult. Therefore, analysis based on c-q slopes requires supplementary modeling approaches and a better mechanistic understanding of individual and combined effects of different drivers on c-q slopes. We recommend interpreting c-q slopes together with analysis of trends in concentration and discharge, and auxiliary meteorological, hydrological, chemical and management data, and confirming the findings with high-frequency hydrochemical measurements. Together, this information can yield important insights into transport and transformations of nutrients, sediments, and other water quality constituents for a wide range of flow conditions.

4.4 The Future of Complex *c-q* Slope Estimation

The time-varying c-q slope models we present assume that there is no additional nonlinearity in concentration response to increased discharge. To assess intricate effects of a variety of mitigation measures, there is a need to model not only linear but also nonlinear c-q slopes changing over time. Mathematically, this is not difficult, e.g., smooth estimates of slopes can be produced in three dimensions (year, season, and flow) and specific c-q slopes can be extracted from models by partial derivatives. In practice, this will not be possible with low-frequency data, partly because there is not enough data to fit such complex models and partly because monthly or biweekly monitoring might miss extreme low or high flows [5]. Huntington and Wieczorek [18] studied c-q relationships using the WRTDS model on data interpolated from monthly to daily resolution and linked results to changes in the relative importance of pollution sources. While this approach allowed fitting of complex models, it did not solve the problem of missing observations in extreme flow conditions, and thus cannot replace appropriate monitoring frequency. Increased possibilities to collect high-frequency hydrochemical data over longer periods will enable estimation of complex c-q slopes. Adequate statistical methods in combination with high-frequency data will greatly enhance understanding of concentration-discharge relationships for different flow regimes and how they change over season, years, and even decades. This will be necessary to reliably connect changes in c-q slopes to their causes and provide a basis for evaluating the effects of environmental change on solute/sediment hydrochemical behavior, which are often difficult to quantify directly.

5 Conclusions

Estimation of c-q slopes can be important in revealing underlying stream and catchment hydrological and biogeochemical processes and can be used to support decisions on land and water management. The hydrochemical data underpinning c-q relationships are often collected at low frequency over a long period and with both concentration and discharge data exhibiting seasonal variation. In such circumstances, assuming that the c-q slope is always constant is naïve and more comprehensive models are needed for detailed analysis of how c-q slopes vary with time, discharge, and other potential drivers. The GAM-based approaches suggested in this study allow modeling of the most common deviations from constant slopes, and improve and simplify estimation of c-q slopes for low-frequency hydrochemical data. However, to improve interpretation of the complex c-q patterns obtained, we recommend combining data analysis of c-q slopes with GAMs and catchment modeling approaches to yield information on different solute and particulate sources and their mobilization patterns under different flow conditions.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10666-023-09915-z.

Author Contribution Claudia von Brömssen: conceptualization, methodology, formal analysis, writing—original draft preparation. Jens Fölster: data curation, writing—reviewing and editing. Katarina Kyllmar: data curation, writing—reviewing and editing. Magdalena Bieroza: writing—reviewing and editing.

Funding Open access funding provided by Swedish University of Agricultural Sciences.

Availability of Data and Materials The datasets and all R code used for models and figures are available online [42].

Declarations

Ethics Approval and Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests The authors declare no competing interests.

References

1. Heathwaite, A. L., & Bieroza, M. (2021). Fingerprinting hydrological and biogeochemical drivers of freshwater quality. *Hydrological Processes*, *35*(1), e13973. https://doi.org/10.1002/hyp.13973

2. Knapp, J. L. A., von Freyberg, J., Studer, B., Kiewiet, L., & Kirchner, J. W. (2020). Concentration–discharge relationships vary among hydrological events, reflecting differences in event characteristics. *Hydrology and Earth System Sciences, 24*(5), 2561–2576. https://doi.org/10.5194/hess-24-2561-2020

3. Minaudo, C., Dupas, R., Gascuel-Odoux, C., Roubeix, V., Danis, P.-A., & Moatar, F. (2019). Seasonal and event-based concentration-discharge relationships to identify catchment controls on nutrient export regimes. *Advances in Water Resources*, *131*, 103379. https://doi.org/10.1016/j.advwatres.2019.103379

4. Rose, L. A., Karwan, D. L., & Godsey, S. E. (2018). Concentration–discharge relationships describe solute and sediment mobilization, reaction, and transport at event and longer timescales. *Hydrological Processes, 32*(18), 2829–2844. https://doi.org/10.1002/hyp.13235

5. Bieroza, M. Z., Heathwaite, A. L., Mullinger, N. J., & Keenan, P. O. (2014). Understanding nutrient biogeochemistry in agricultural catchments: The challenge of appropriate monitoring frequencies. *Environmental Science: Processes & Impacts, 16*(7), 1676–1691. https://doi.org/10.1039/C4EM00100A

6. Bieroza, M. Z., Heathwaite, A. L., Bechmann, M., Kyllmar, K., & Jordan, P. (2018). The concentration-discharge slope as a tool for water quality management. *Science of The Total Environment, 630*, 738–749. https://doi.org/10.1016/j.scitotenv.2018.02.256

7. 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. https://doi.org/10.1002/hyp.7315

8. Musolff, A., Schmidt, C., Selle, B., & Fleckenstein, J. H. (2015). Catchment controls on solute export. *Advances in Water Resources, 86*, 133–146. https://doi.org/10.1016/j.advwatres.2015.09.026

9. Evans, C., & Davies, T. D. (1998). Causes of concentration/discharge hysteresis and its potential as a tool for analysis of episode hydrochemistry. *Water Resources Research, 34*(1), 129–137. https://doi.org/10.1029/97WR01881

10. Johnson, N. M., Likens, G. E., Bormann, F. H., Fisher, D. W., & Pierce, R. S. (1969). A working model for the variation in stream water chemistry at the Hubbard Brook Experimental Forest. *New Hampshire. Water Resources Research, 5*(6), 1353–1363. https://doi.org/10.1029/WR005i006p01353

11. Fazekas, H. M., McDowell, W. H., Shanley, J. B., & Wymore, A. S. (2021). Climate variability drives watersheds along a transporter-transformer continuum. *Geophysical Research Letters*, *48*(21), e2021GL094050. https://doi.org/10.1029/2021GL094050

12. Moatar, F., Abbott, B. W., Minaudo, C., Curie, F., & Pinay, G. (2017). Elemental properties, hydrology, and biology interact to shape concentration-discharge curves for carbon, nutrients, sediment, and major ions. *Water Resources Research, 53*(2), 1270–1287. https://doi.org/10.1002/2016WR019635

13. Muggeo, V. M. R. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine, 22*(19), 3055–3071. https://doi.org/10.1002/sim.1545

14. Neira, J. M. T., Tallec, G., Andreassian, V., & Mouchel, J.-M. (2020). A combined mixing model for high-frequency concentration-discharge relationships. *Journal of Hydrology*, *591*, 125559. https://doi.org/10.1016/j.jhydrol.2020.125559

15. Dehaspe, J., Sarrazin, F., Kumar, R., Fleckenstein, J. H., & Musolff, A. (2021). Bending of the concentration discharge relationship can inform about in-stream nitrate removal. *Hydrology and Earth System Sciences Discussions*, 1–43. https://doi.org/10.5194/hess-2021-16

16. Fork, M. L., Sponseller, R. A., & Laudon, H. (2020). Changing source-transport dynamics drive differential browning trends in a boreal stream network. *Water Resources Research*, *56*(2), e2019WR026336. https://doi.org/10.1029/2019WR026336

17. Hirsch, R. M., Moyer, D. L., & Archfield, S. A. (2010). Weighted regressions on time, discharge, and season (WRTDS), with an application to Chesapeake Bay river inputs1. *JAWRA Journal of the American Water Resources Association, 46*(5), 857–880. https://doi.org/10.1111/j.1752-1688.2010.00482.x

18. Huntington, T. G., & Wieczorek, M. E. (2021). An increase in the slope of the concentration-discharge relation for total organic carbon in major rivers in New England, 1973 to 2019. *Science of The Total Environment*, *778*, 146149. https://doi.org/10.1016/j.scitotenv.2021.146149

19. Zhang, Q. (2018). Synthesis of nutrient and sediment export patterns in the Chesapeake Bay watershed: Complex and non-stationary concentration-discharge relationships. *Science of The Total Environment, 618*, 1268–1283. https://doi.org/10.1016/j.scitotenv.2017.09.221

20. Hirsch, R. M., Archfield, S. A., & De Cicco, L. A. (2015). A bootstrap method for estimating uncertainty of water quality trends. *Environmental Modelling & Software, 73*, 148–166. https://doi.org/10.1016/j.envsoft.2015.07.017

21. Zhang, Q., Harman, C. J., & Ball, W. P. (2016). An improved method for interpretation of riverine concentration-discharge relationships indicates long-term shifts in reservoir sediment trapping. *Geophysical Research Letters, 43*(19), 10215–10224. https://doi.org/10.1002/2016GL069945

22. Zimmer, M. A., Pellerin, B., Burns, D. A., & Petrochenkov, G. (2019). Temporal variability in nitrate-discharge relationships in large rivers as revealed by high-frequency data. *Water Resources Research, 55*(2), 973–989. https://doi.org/10.1029/2018WR023478

23. Fazekas, H. M., Wymore, A. S., & McDowell, W. H. (2020). Dissolved organic carbon and nitrate concentration-discharge behavior across scales: Land use, excursions, and misclassification. *Water Resources Research*, *56*(8), e2019WR027028. https://doi.org/10.1029/2019WR027028

24. Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical Science, 1*(3), 297–310. https://doi.org/10.1214/ss/1177013604

25. Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). CRC Press/Taylor & Francis Group.

26. Monteith, D. T., Stoddard, J. L., Evans, C. D., de Wit, H. A., Forsius, M., Høgåsen, T., … & Vesely, J. (2007). Dissolved organic carbon trends resulting from changes in atmospheric deposition chemistry. *Nature*, *450*(7169), 537–540. https://doi.org/10.1038/nature06316

27. Orr, H. G., Simpson, G. L., Clers, S., & des, Watts, G., Hughes, M., Hannaford, J., … & Evans, R. (2015). Detecting changing river temperatures in England and Wales. *Hydrological Processes, 29*(5), 752–766. https://doi.org/10.1002/hyp.10181

28. von Brömssen, C., Betnér, S., Fölster, J., & Eklöf, K. (2021). A toolbox for visualizing trends in large-scale environmental data. *Environmental Modelling & Software*, *136*, 104949. https://doi.org/10.1016/j.envsoft.2020.104949

29. Hastie, T., & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society Series B-Methodological, 55*(4), 757–796.

30. Carvalho, L., Miller nee Ferguson, C. A., Scott, E. M., Codd, G. A., Davies, P. S., & Tyler, A. N. (2011). Cyanobacterial blooms: Statistical models describing risk factors for national-scale lake assessment and lake management. *The Science of the Total Environment, 409*(24), 5353–5358. https://doi.org/10.1016/j.scitotenv.2011.09.030

31. Gong, Y., Tang, X., Shao, K., Hu, Y., & Gao, G. (2017). Dynamics of bacterial abundance and the related environmental factors in large shallow eutrophic Lake Taihu. *Journal of Freshwater Ecology, 32*(1), 133–145. https://doi.org/10.1080/02705060.2016.1248506

32. Neter, J., Wasserman, W., & Kutner, M. H. (1990). *Applied linear statistical models: Regression, analysis of variance, and experimental designs*. Irwin.

33. Wood, S. (2022). mgcv: Mixed GAM computation vehicle with automatic smoothness estimation. Retrieved from https://CRAN.R-project.org/package=mgcv

34. Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Series B-Statistical Methodology, 73*, 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x

35. R Core Team. (2022). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

36. Simpson, G. L. (2023). *gratia: Graceful ggplot-based graphics and other functions for GAMs fitted using mgcv*. Retrieved from https://gavinsimpson.github.io/gratia/

37. Simpson, G. L. (2018). *Modelling palaeoecological time series using generalized additive models*. https://doi.org/10.1101/322248

38. Marra, G., & Wood, S. N. (2012). Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics, 39*(1), 53–74. https://doi.org/10.1111/j.1467-9469.2011.00760.x

39. de Boor, C. (1978). *A practical guide to splines*. Springer, New York, NY.

40. Persson, I., Pirard, J., Larsson, A., Holm, C., & Lousa-Alvin, A. (2012). Kväveavskiljningens effekt på Ekoln. Svenskt Vatten AB. Retrieved from vav.griffel.net/filer/SVU-rapport_2012–1

41. Kyllmar, K., Forsberg, L. S., Andersson, S., & Mårtensson, K. (2014). Small agricultural monitoring catchments in Sweden representing environmental impact. *Agriculture, Ecosystems & Environment, 198*, 25–35. https://doi.org/10.1016/j.agee.2014.05.016

42. von Brömssen, C., Fölster, J., Kyllmar, K., & Bieroza, M. (2023). R code and data for modelling complex-c-q-relationships using generalized additive models. *Zenodo*. https://doi.org/10.5281/zenodo.7798939

43. Wood, S. N. (2013). On p-values for smooth components of an extended generalized additive model. *Biometrika, 100*(1), 221–228. https://doi.org/10.1093/biomet/ass048

44. Meals, D. W., Dressing, S. A., & Davenport, T. E. (2010). Lag time in water quality response to best management practices: A review. *Journal of Environmental Quality, 39*(1), 85–96. https://doi.org/10.2134/jeq2009.0108

45. Rode, M., Wade, A. J., Cohen, M. J., Hensley, R. T., Bowes, M. J., Kirchner, J. W., … & Jomaa, S. (2016). Sensors in the stream: The high-frequency wave of the present. *Environmental Science & Technology*, *50*(19), 10297–10307. https://doi.org/10.1021/acs.est.6b02155

46. Stålnacke, P., & Grimvall, A. (2001). Semiparametric approaches to flow normalization and source apportionment of substance transport in rivers. *Environmetrics, 12*(3), 233–250. https://doi.org/10.1002/env.459