

Modelling palaeoecological time series using generalized additive models

Gavin L. Simpson

¹gavin.simpson@uregina.ca

¹Institute of Environmental Change and Society, University of Regina,
Regina, Saskatchewan, Canada, S4S 0A2

August 15, 2018

8 *Keywords*— time series; generalized additive model; simultaneous interval; spline; environmental change
9

10 Abstract

In the absence of annual laminations, time series generated from lake sediments or other similar stratigraphic sequences are irregularly spaced in time, which complicates formal analysis using classical statistical time series models. In lieu, statistical analyses of trends in palaeoenvironmental time series, if done at all, have typically used simpler linear regressions or (non-) parametric correlations with little regard for the violation of assumptions that almost surely occurs due to temporal dependencies in the data or that correlations do not provide estimates of the magnitude of change, just whether or not there is a linear or monotonic trend. Alternative approaches have used LoESS-estimated trends to justify data interpretations or test hypotheses as to the causal factors without considering the inherent subjectivity of the choice of parameters used to achieve the LoESS fit (e.g. span width, degree of polynomial). Generalized additive models (GAMs) are statistical models that can be used to estimate trends as smooth functions of time. Unlike LoESS, GAMs use automatic smoothness selection methods to objectively determine the complexity of the fitted trend, and as formal statistical models, GAMs, allow for potentially complex, non-linear trends, a proper accounting of model uncertainty, and the identification of periods of significant temporal change. Here, I present a consistent and modern approach to the estimation of trends in palaeoenvironmental time series using GAMs, illustrating features of the methodology with two example time series of contrasting complexity; a 150-year bulk organic matter $\delta^{15}\text{N}$ time series from Small Water, UK, and a 3000-year alkenone record from Braya-Sø, Greenland. I discuss the underlying mechanics of GAMs that allow them to learn the shape of the trend from the data themselves and how simultaneous confidence intervals and the first derivatives of the trend are used to properly account for

32 model uncertainty and identify periods of change. It is hoped that by using GAMs greater at-
33 tention is paid to the statistical estimation of trends in palaeoenvironmental time series leading
34 to more a robust and reproducible palaeoscience.

35 1 Introduction

36 Palaeoecology and palaeolimnology have moved away from being descriptive disciplines,
37 rapidly adopting new statistical developments in the 1990s and beyond (@ Smol et al., 2012).
38 Less development has been observed in the area of trend estimation in palaeoenvironmental
39 time series. The vast majority of data produced by palaeoecologists and palaeolimnologists
40 is in the form of time-ordered observations on one or more proxies or biological taxa (Birks,
41 2012b; Smol, 2008; Smol et al., 2012). Typically these data are arranged irregularly in time;
42 in the absence of annual laminae or varves, the sediment core is sectioned at regular depth
43 intervals (Glew et al., 2001), which, owing to variation in accumulation rates over time
44 and compaction by overlying sediments, results in an uneven sampling in time. An under-
45 appreciated feature of such sampling is that younger sediments often have larger variance
46 than older sediments; each section of core represents fewer lake years in newer samples,
47 relative to older samples. This variable averaging acts as a time-varying low-pass (high-cut)
48 filter of the annual depositional signal.

49 Irregular intervals between samples means that the time-series analysis methods of autore-
50 gressive or moving average processes, in the form of autoregressive integrated moving aver-
51 age (ARIMA) models, are practically impossible to apply because software typically requires
52 even spacing of observations in time. Dutilleul et al. (2012) and Birks (2012a), eschewing the
53 term *time series*, prefer to call such data *temporal series* on account of the irregular spacing of
54 samples, a distinction that I find unnecessary.

55 Where statistical approaches have been applied to trend estimation in palaeoenvironmental
56 time series, a commonly-used method is LoESS (Birks, 1998, 2012a; Cleveland, 1979; Juggins
57 and Telford, 2012). LoESS, locally weighted scatterplot smoother, as it's name suggests, was
58 developed to smooth x-y scatterplot data (Cleveland, 1979). The method fits a smooth line
59 through data by fitting weighted least squares (WLS) models to observations within a partic-
60 ular, user-specified window of the focal point, whose width is typically expressed as a pro-
61 portion α of the n data points. Weights are determined by how close (in the x-axis only) an
62 observation in the window is to the focal point giving greatest weight given to points closest
63 to the focal point. The interim LoESS-smoothed value for the focal point is the predicted value
64 from the weighted regression at the focal point. The interim values are updated using weights
65 based on how far in the y-axis direction the interim smoothed value lies from the observed
66 value plus the x-axis distance weights; this has the effect of down-weighting outlier observa-
67 tions. The final LoESS is obtained by joining the smoothed values. The user has to choose how
68 large a window to use, whether to fit degree 1 (linear) or degree 2 (quadratic) polynomials in
69 the WLS model, and how to weight points in the x-axis. When used in an exploratory mode,
70 the user has considerable freedom to choose the detail of the LoESS fit; the window width, for
71 example, can be infinitely tweaked to give as close a fit to the data, as assessed by eye, as is de-
72 sired. Using cross-validation (CV) to choose α or the degree of polynomial in the WLS model

73 is complicated for a number of reasons, not least because the CV scheme used must involve
74 the time ordering of the data (e.g. Bergmeir et al., 2018). This subjectivity is problematic how-
75 ever once we wish to move beyond exploratory analysis and statistically identify trends to test
76 hypotheses involving those trend estimates.

77 Running means or other types of filter (Juggins and Telford, 2012) have also been used exten-
78 sively to smooth palaeoenvironmental time series, but, as with LOESS, their behaviour depends
79 on a number of factors, including the filter width. Furthermore, the width of the filter causes
80 boundary issues; with a centred filter, of width five, the filtered time series would be two data
81 points shorter at both ends of the series because the filter values are not defined for the first
82 and last two observations of the original series as these extra time points were not observed.
83 Considerable research effort has been expended to identify ways to pad the original time series
84 at one or both ends to maintain the original length in the filtered series, without introducing
85 bias due to the padding (e.g. Mann, 2004, 2008; Mills, 2006, 2007, 2010).

86 These are not the only methods that have been used to estimate trends in stratigraphic series.
87 Another common approach involves fitting a simple linear trend using ordinary least squares
88 regression and use the resulting t statistic as evidence against the null hypothesis of no trend
89 despite the statistical assumptions being almost surely violated due to dependence among
90 observations. The Pearson correlation coefficient, r , is also often used to detect trends in palaeo
91 time series (Birks, 2012a), despite the fact that r provides no information as to the magnitude of
92 the estimated trend, and the same temporal autocorrelation problem that dogs ordinary least
93 squares similarly plagues significance testing for r (Tian et al., 2011). Additionally, both the
94 simple least squares trend line and r are tests for *linear* trends only, and yet we typically face
95 data sets with potentially far more complex trends than can be identified by these methods.
96 Instead, non-parametric rank correlation coefficients have been used (Birks, 2012a; Gautier,
97 2001), and whilst these do allow for the detection of non-linear trends, trends are restricted to
98 be monotonic, no magnitude of the trend is provided, and the theory underlying significance
99 testing of Spearman's ρ and Kendall's τ assumes independent observations.

100 Here, I describe generalized additive models (GAMs; Hastie and Tibshirani, 1986, 1990; Rup-
101 pert et al., 2003; Wood, 2017; Yee and Mitchell, 1991) for trend estimation. GAMs, like simple
102 linear regression, are a regression-based method for estimating trends, yet they are also, su-
103 perficially at least, similar to LOESS. GAMs and LOESS estimate smooth, non-linear trends in
104 time series and both can handle the irregular spacing of samples in time, yet GAMs do not
105 suffer from the subjectivity that plagues LOESS as a method of formal statistical inference.

106 In the subsequent sections, I present an introduction to GAMs and discuss the issue of uncer-
107 tainty in model-estimated trends, the topic of posterior simulation from a regression model
108 and how to identify periods of significant environmental change using the first derivative of
109 the estimated trend. The main steps in the analysis of palaeoenvironmental time series using
110 GAMs are illustrated in Figure 1. Two non-standard types of spline — adaptive smoothers and
111 Gaussian process splines — that are especially applicable to GAMs in the palaeoenvironmental
112 setting are subsequently described, followed by an assessment of the impact of age-model
113 uncertainty on trend estimation via GAMs. Finally, I briefly discuss the application of GAM
114 trend analysis to multivariate species abundance and compositional data.

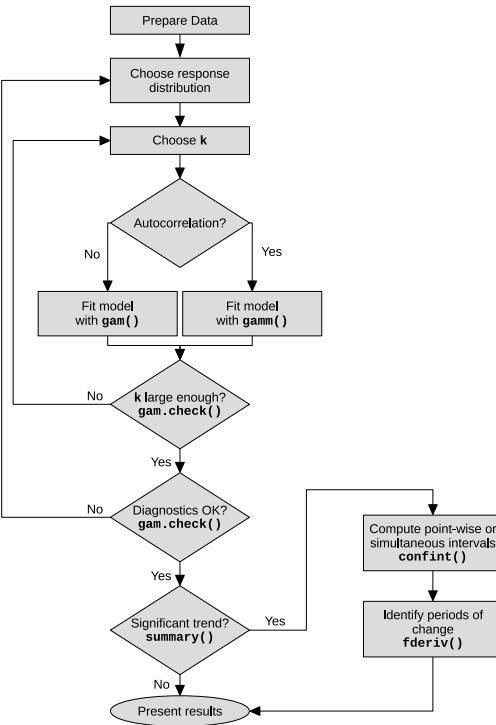


Figure 1: Flowchart showing the main steps in the analysis of time series using generalized additive models.

115 1.1 Example time series

116 To illustrate trend estimation in palaeoenvironmental data using GAMs, I use two proxy time
 117 series; a 150-year bulk organic matter $\delta^{15}\text{N}$ record from Small Water, and a 3000-year alkenone
 118 record from Braya-Sø. Between them, the two examples, combine many of the features of
 119 interest to palaeoecologists that motivate the use of GAMs; non-linear trends and the question
 120 of when changes in the measured proxy occurred. The example analyses were all performed
 121 using the *mgcv* package (version 1.8.23; Wood, 2017) and R (version 3.4.4; R Core Team, 2018),
 122 and the supplementary material contains a fully annotated document showing the R code used
 123 to replicate all the analyses described in the remainder of the paper.

124 1.1.1 $\delta^{15}\text{N}$ time series from Small Water

125 Figure 2a shows 48 nitrogen stable isotope measurements on the bulk organic matter of a sed-
 126 iment core collected from Small Water, a small corrie lake located in the English Lake District,
 127 UK. The data were collected to investigate disturbance of nitrogen (N) cycling in remote, olig-
 128 otrophic lakes by N deposited from the atmosphere (Simpson, unpublished data). The data
 129 are shown on a ^{210}Pb time scale. Questions that might be asked about this series are; what is
 130 the trend in $\delta^{15}\text{N}$?, when do we first see evidence for a change in $\delta^{15}\text{N}$?, and is the reversal in
 131 $\delta^{15}\text{N}$ values in the uppermost section of the core a real change?

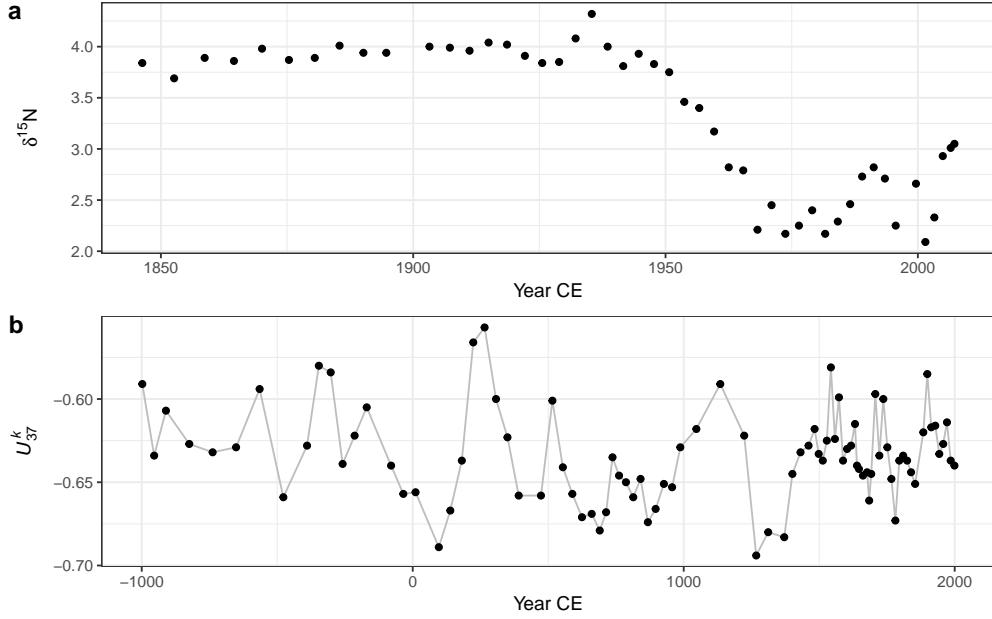


Figure 2: Example time series; a) Small Water bulk organic matter $\delta^{15}\text{N}$ time series on a ^{210}Pb time scale, and b) Braya-Sø U_{37}^K time series on a calibrated ^{14}C time scale. The observations U_{37}^K time series have been joined by lines purely as a visual aid to highlight potential trends.

¹³² 1.1.2 Braya-Sø alkenone time series

¹³³ The second example time series is a 3,000 year record of alkenone unsaturation, U_{37}^K , from
¹³⁴ Braya-Sø, a meromictic lake in West Greenland (D'Andrea et al., 2011). Alkenones are long-
¹³⁵ chained unsaturated organic compounds that are produced by a small number of planktonic
¹³⁶ organisms known as haptophytes. The U_{37}^K unsaturation index is (Brassell, 1993)

$$U_{37}^K = \frac{[C_{37:2}] - [C_{37:4}]}{[C_{37:2}] + [C_{37:3}] + [C_{37:4}]}$$

¹³⁷ where $[C_{37:x}]$ is the concentration of the alkenone with 37 carbon atoms and x double carbon
¹³⁸ bonds. The relative abundance of these alkenones is known to vary with changes in water
¹³⁹ temperature (Brassell, 1993; Chu et al., 2005; Toney et al., 2010; Zink et al., 2001), and as a
¹⁴⁰ result U_{37}^K is used as a proxy for lake- and sea-surface temperatures. For further details on the
¹⁴¹ Braya-Sø U_{37}^K record and age model see D'Andrea et al. (2011). Here I use the 3,000 year U_{37}^K
¹⁴² record from the PAGES 2K database (PAGES 2K Consortium, 2013). The data are presented in
¹⁴³ Figure 2b.

¹⁴⁴ 2 Regression models for palaeoenvironmental time series

¹⁴⁵ A linear model for a trend in a series of T observations y_t at observation times x_t with $t =$
¹⁴⁶ $1, 2, \dots, T$ is

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, \quad (1)$$

147 where β_0 is a constant term, the model *intercept*, representing the expected value of y_t where
 148 x_t is 0. β_1 is the *slope* of the best fit line through the data; it measures the rate of change in y
 149 for a unit increase in x . The unknowns, the β_j , are commonly estimated using least squares
 150 by minimising the sum of squared errors, $\sum_t \varepsilon_t^2$. If we want to ask if the estimated trend β_1 is
 151 statistically significant we must make further assumptions about the data (conditional upon
 152 the fitted model) or the model errors (residuals); $\varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$. This notation indicates that
 153 the residuals ε_t are *independent* and *identically distributed* Gaussian random variables with mean
 154 equal to 0 and constant variance σ^2 . In the time series setting, the assumption of independence
 155 of model residuals is often violated.

156 The linear model described above is quite restrictive in terms of the types of trend it can fit;
 157 essentially linear increasing or decreasing trends, or, trivially, a null trend of no change. This
 158 model can be extended to allow for non-linear trends by making y_t depend on polynomials of
 159 x_t , for example

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \cdots + \beta_P x_t^P + \varepsilon_t \\ &= \beta_0 + \sum_{p=1}^P \beta_p x_t^p + \varepsilon_t, \end{aligned} \quad (2)$$

160 where polynomials of x_t up to order P are used. This model allows for more complex trends
 161 but it remains a fully parametric model and suffers from several problems, especially the be-
 162 haviour of the fitted trend at the start and end of the observed series.

163 Linear models using a range of polynomials fitted to the Small Water data set are shown in
 164 Figure 3. The low-order models ($P \in \{1, 3\}$) result in very poor fit to the data. The model with
 165 $P = 5$ does a reasonable job of capturing the gross pattern in the time series, but fails to adapt
 166 quickly enough to the decrease in $\delta^{15}\text{N}$ that begins ~ 1940 CE, and the estimated trend is quite
 167 biased as a result. The $P = 10$ -th-order polynomial model is well able to capture this period
 168 of rapid change, but it does so at the expense of increased complexity in the estimated trend
 169 prior to ~ 1940 . Additionally, this model ($P = 10$) has undesirable behaviour at the ends of the
 170 series, significantly overfitting the data, a commonly observed problem in polynomial models
 171 such as these (Epperson, 1987; Runge, 1901). Finally, the choice of what order of polynomial to
 172 fit is an additional choice left for the analyst to specify; choosing the optimal P is not a trivial
 173 task when the data are a time series and residual autocorrelation is likely present.

174 Can we do better than these polynomial fits? In the remainder, I hope to demonstrate that
 175 the answer to that question is emphatically “yes”! Below I describe a coherent and consistent
 176 approach to modelling palaeoenvironmental time series using generalized additive models
 177 that builds upon the linear regression framework.

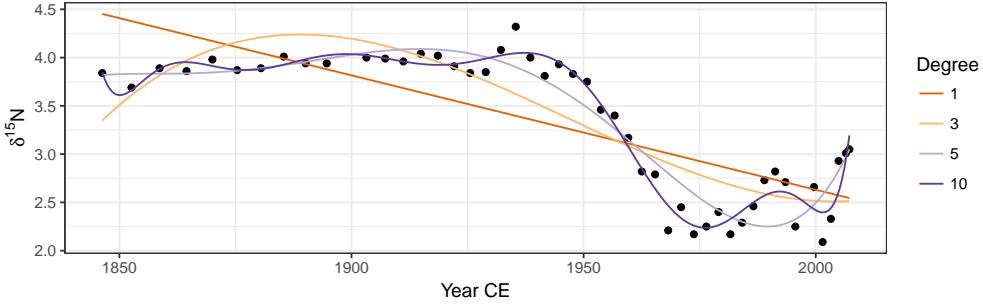


Figure 3: Linear models with various orders of polynomial of the covariate Year fitted using ordinary least squares to the $\delta^{15}\text{N}$ time series from Small Water. The degree of polynomial is indicated, with the degree 1 line equal to a simple linear regression model.

178 3 Generalized additive models

179 The GAM version of the linear model (1) is

$$y_t = \beta_0 + f(x_t) + \varepsilon_t, \quad (3)$$

180 where the linear effect of time (the $\beta_1 x_t$ part) has been replaced by a smooth function of time,
 181 $f(x_t)$. The immediate advantage of the GAM is that we are no longer restricted to the shapes
 182 of trends that can be fitted via global polynomial functions such as (2). Instead, the shape of
 183 the fitted trend will be estimated from the data itself.

184 The linear model is a special case of a broader class, known as the generalized linear model
 185 (GLM; McCullagh and Nelder, 1989). The GLM provides a common framework for modelling
 186 a wide range of types of data, such as count, proportions, or binary (presence/absence) data,
 187 that are not conditionally distributed Gaussian. GLMs are, like the linear model, parametric
 188 in nature; the types of trends that we can fit using a GLM are the linear or polynomial mod-
 189 els. GAMs extend the GLM by relaxing this parametric assumption; in a GAM some, or all,
 190 of the parametric terms, the β_p , are replace by smooth functions f_j of the covariates x_j . For
 191 completeness then, we can write (3) as a GLM/GAM

$$y_t \sim \text{EF}(\mu_t, \Theta) \quad (4a)$$

$$g(\mu_t) = \beta_0 + f(x_t) \quad (4b)$$

$$\mu_t = g^{-1}(\beta_0 + f(x_t)), \quad (4c)$$

192 where μ_t is the expected value (e.g. the mean count or the probability of occurrence) of the
 193 random variable Y_t ($\mu_t \equiv \mathbb{E}(Y_t)$) of which we have observations y_t . g is the link function, an
 194 invertible, monotonic function, such as the natural logarithm, and g^{-1} is its inverse. The link
 195 function maps values from the response scale on to the scale of the linear predictor, whilst the
 196 inverse of the link function provides the reverse mapping. For example, count data are strictly
 197 non-negative integer values and are commonly modelled as a Poisson GLM/GAM using the

198 natural log link function. On the log scale, the response can take any real value between $-\infty$
 199 and $+\infty$, and it is on this scale that model fitting actually occurs (i.e. using equation (4b)).
 200 However we need to map these unbounded values back on to the non-negative response scale.
 201 The inverse of the log link function, the exponential function, achieves this and maps values
 202 to the interval $0-\infty$ (equation (4c)).

203 In (4a), we further assume that the observations are drawn from a member of the exponential
 204 family of distributions — such as the Poisson for count data, the binomial for presence/absence
 205 or counts from a total — with expected value μ_t and possibly some additional parameters Θ
 206 ($y_t \sim \text{EF}(\mu_t, \Theta)$). Additionally, many software implementations of the above model also allow
 207 for distributions that are not within the exponential family but which can be fitted using an
 208 algorithm superficially similar to the one used to fit GAMs to members of the exponential
 209 family (e.g. Wood et al., 2016). Common examples of such extended families include the
 210 negative binomial distribution (for overdispersed counts) and the beta distribution (for true
 211 proportions or other interval-bounded data).

212 3.1 Basis functions

213 It is clear from the plots of the data (Figure 2) that we require the fitted trends for the Small
 214 Water $\delta^{15}\text{N}$ and Braya-Sø U_{37}^K time series to be non-linear functions, but it is less clear how to
 215 specify the actual shape require. Ideally, we'd like to learn the shape of the trend from the
 216 data themselves. We will refer to these non-linear functions as *smooth functions*, or just *smooths*
 217 for short, and we will denote a smooth using $f(x_t)$. Further, we would like to represent the
 218 smooths in a way that (4) is represented parametrically so that it can be estimate within the
 219 well-studied GLM framework. This is achieved by representing the smooth using a *basis*. A
 220 basis is a set of functions that collectively span a space of smooths that, we hope, contains the
 221 true $f(x_t)$ or a close approximation to it. The functions in the basis are known as *basis functions*,
 222 and arise from a *basis expansion* of a covariate. Writing $b_j(x_t)$ as the j th basis function of x_t , the
 223 smooth $f(x_t)$ can be represented as a weighted sum of basis functions

$$f(x_t) = \sum_{j=1}^k b_j(x_t)\beta_j,$$

224 where β_j is the weight applied to the j th basis function.

225 The polynomial model is an example of a statistical model that uses a basis expansion. For the
 226 cubic polynomial ($P = 3$) fit shown in Figure 3 there are in fact 4 basis functions: $b_1(x_t) = x_t^0 = 1$,
 227 $b_2(x_t) = x_t$, $b_3(x_t) = x_t^2$, and $b_4(x_t) = x_t^3$. Note that $b_1(x_t)$ is constant and is linked to the
 228 model intercept, β_0 , in the linear model (2), and further, that the basis function weights are the
 229 estimated coefficients in the model, the β_j .

230 As we have already seen, polynomial basis expansions do not necessarily lead to well-fitting
 231 models unless the true function f is itself a polynomial. One of the primary criticisms is that
 232 polynomial basis functions are global (Magee, 1998); the value of f at time point x_t affects the
 233 value of f at time point x_{t+s} even if the two time points are at opposite ends of the series. There
 234 are many other bases we could use; here I discuss one such set of bases, that of splines.

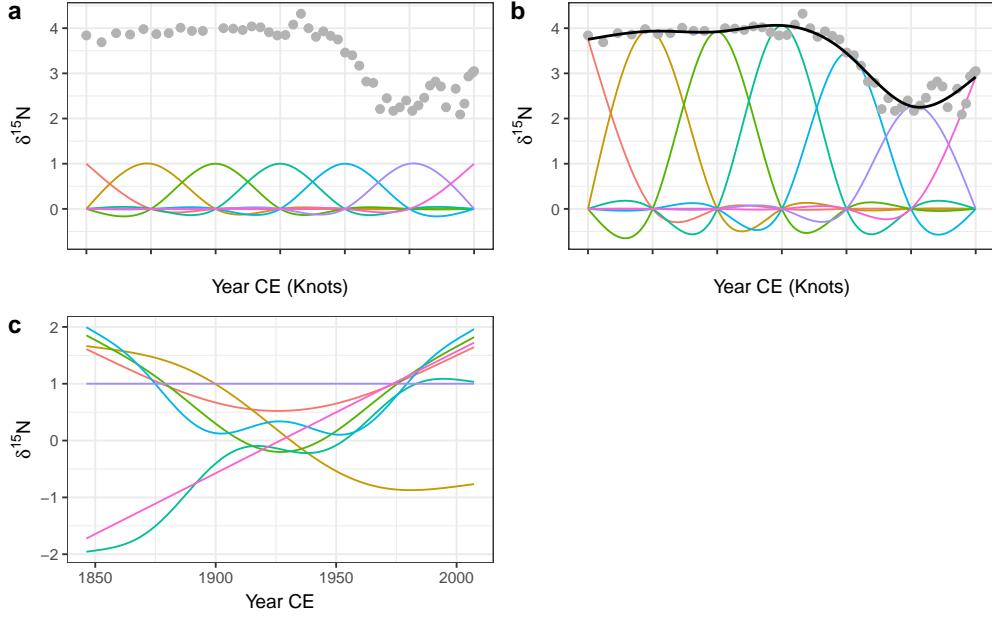


Figure 4: Basis functions for the time covariate and the Small Water $\delta^{15}\text{N}$ time series. A rank (size) 7 cubic regression spline (CRS) basis expansion is show in a), with knots, indicated by tick marks on the x-axis, spread evenly through the rang of the data. b) shows the same CRS basis functions weighted by the estimated coefficients β_j , plus the resulting GAM trend line (black line drawn through the data). The grey points in both panels are the observed $\delta^{15}\text{N}$ values. c) A rank 7 thin plate regression spline basis for the same data.

235 There are a bewildering array of different types of spline. In the models discussed below we
 236 will largely restrict ourselves to cubic regression splines (CRS) and thin plate regression splines
 237 (TPRS). In addition, I also discuss two special types of spline basis, an adaptive spline basis
 238 and a Gaussian process spline basis.

239 A cubic spline is a smooth curve comprised of sections of cubic polynomials, where the sections
 240 are joined together at some specified locations — known as *knots* — in such a way that at
 241 the joins, the two sections of cubic polynomial that meet have the same value as well as the
 242 same first and second derivative. These properties mean that the sections join smoothly and
 243 differentiably at the knots (Wood, 2017, 5.3.1).

244 The CRS can be parameterized in a number of different ways. One requires a knot at each
 245 unique data value in x_t , which is computationally inefficient. Another way of specifying a
 246 CRS basis is to parameterize in terms of the value of the spline at the knots. Typically in
 247 this parametrization there are many fewer knots than unique data, with the knots distributed
 248 evenly over the range of x_t or at the quantiles of x_t . Placing knots at the quantiles of x_t has the
 249 effect of placing a greater number of knots where the data is densest.

250 A CRS basis expansion comprised of 7 basis functions for the time covariate in the Small Water
 251 series, is shown in Figure 4a. The tick marks on the x-axis show the locations of the knots,
 252 which are located at the ends of the series and evenly in between. Notice that in this particular
 253 parametrization, the j th basis function takes a value of 1 at the j th knot and at all other knots

254 a value of 0.

255 To estimate a model using this basis expansion each basis function forms a column in the model
256 matrix \mathbf{X} and the weights β_j can be found using least squares regression (assuming a Gaussian
257 response). Note that in order to estimate a coefficient for each basis function the model has
258 to be fitted without an intercept term. In practice we would include an intercept term in the
259 model and therefore the basis functions are modified via an identifiability constraint (@ Wood,
260 2017). This has the effect of making the basis orthogonal to the intercept but results in more
261 complicated basis functions than those shown in in Figure 4a.

262 Having estimated the weight for each basis function, the j th basis function b_j is scaled
263 (weighted) by its coefficient β_j . The scaled CRS basis functions for the Small Water time series
264 are shown in Figure 4b. The solid line passing through the data points is formed by summing
265 up the values of the scaled basis functions ($b_j(x_t)\beta_j$) at any value of x_t (time).

266 Cubic regression splines, as well as many other types of spline, require the analyst to choose
267 the number and location of the knots that parametrise the basis. Thin plate regression splines
268 (TPRS) remove this element of subjectivity when fitting GAMs. Thin plate splines were in-
269 troduced by Duchon (1977) and, as well as solving the knot selection problem, have several
270 additional attractive properties in terms of optimality and their ability to estimate a smooth
271 function of two or more variables, leading to smooth interactions between covariates. How-
272 ever, thin plate splines have one key disadvantage over CRS; thin plate splines have as many
273 unknown parameters as there are unique combinations of covariate values in a data set (Wood,
274 2017, 5.5.1). It is unlikely that any real data problem would involve functions of such complex-
275 ity that they require as many basis functions as data. It is much more likely that the true func-
276 tions that we attempt to estimate are far simpler than the set of functions representable by 1
277 basis function per unique data value. From a practical point of view, it is also highly inefficient
278 to carry around all these basis functions whilst model fitting, and the available computational
279 resources would become quickly exhausted for large time series with many observations.

280 To address this issue low rank thin plate regression splines (TPRS) have been suggested which
281 truncate the space of the thin plate spline basis to some lower number of basis functions whilst
282 preserving much of the advantage of the original basis as an optimally-fitting spline (Wood,
283 2003). A rank 7 TPRS basis (i.e. one containing 7 basis functions) is shown in Figure 4c for the
284 Small Water time series. The truncation is achieved by performing an eigen-decomposition
285 of the basis functions and retaining the eigenvectors associated with the k largest eigenvalues.
286 This is similar to the way principal components analysis decomposes a data set into axes of
287 variation (eigenvectors) in decreasing order of variance explained. The truncated basis can
288 preserve much of the space of functions spanned by the original basis but at the cost of using
289 far fewer basis functions (Wood, 2003, 2017, 5.5.1). Note the horizontal TPRS basis function (at
290 $\delta^{15}\text{N} = 1$) in Figure 4c; this basis function is confounded with the intercept term and, after the
291 application of identifiability constraints, ends up being removed from the set of basis functions
292 used to fit the model.

293 The truncation suggested by Wood (2003) is not without cost; the eigen-decomposition and
294 related steps can be relatively costly for large data sets. For data sets of similar size to the two
295 examples used here, the additional computational effort required to set up the TPRS basis over
296 the CRS basis will not be noticeable. For highly resolved series containing more than ~1000

297 observations the truncation may be costly computationally. In such instances, little is lost by
298 moving to the CRS basis with the same number of knots as the rank of the desired TPRS, with
299 the benefit of considerably reduced set up time for the basis.

300 To fit a GAM using either of the two regression spline bases described above, the analyst is
301 generally only required to specify the size (rank) of the basis expansion required to rep-
302 resent or closely approximate the true function f . With practice and some knowledge of the
303 system from which the observations arise, it can be relatively easy to put an upper limit on
304 the expected complexity of the true trend in the data. Additionally, the number available data
305 points places a constraint on the upper limit of the size of basis expansion that can be used.

306 In practice, the size of the basis is an upper limit on the expected complexity of the trend,
307 and a simple test to check if the basis used was sufficiently large (Pya and Wood, 2016). This
308 test is available via the `gam.check()` function in *mgcv* for example, which essentially looks at
309 whether there is any additional nonlinearity or structure in the residuals that can be explained
310 by a further smooth of x_t . Should a smooth term in the fitted model fail this test the model can
311 be refitted using a larger basis expansion, say by doubling the value of k (the rank) used to fit
312 the original. Note also that a smooth might fail this test whilst using fewer effective degrees
313 of freedom than the maximum possible for the dimension of basis used. This may happen
314 when the true function lies at the upper limit of the set of functions encompassed by the size
315 of basis used. Additionally, a basis of size $2k$ encompasses a richer space of functions of a
316 given complexity than a basis of size k (Wood, 2017); increasing the basis dimension used to fit
317 the model may unlock this additional function space resulting in a better fitting model whilst
318 using a similar number of effective degrees of freedom.

319 3.2 Smoothness selection

320 Having identified low rank regression splines as a useful way to represent f , we next need
321 a way to decide how wiggly the fitted trend should be. A backwards elimination approach
322 to sequentially remove knots or basis functions might seem appropriate, however such an
323 approach would likely fail as the resulting sequence of models would not be strictly nested,
324 precluding many forms of statistical comparison (Wood, 2017). Alternatively, we could keep
325 the basis dimension at a fixed size but guard against fitting very complex models through the
326 use of a wigginess penalty.

327 The default wigginess penalty used in GAMs is on the second derivative of the spline, which
328 measures the rate of change of the slope, or the curvature, of the spline at any infinitesimal
329 point in the interval spanned by x_t . The actual penalty used is the integrated squared second
330 derivative of the spline

$$\int_{\mathbb{R}} [f'']^2 dx = \beta^T \mathbf{S} \beta. \quad (5)$$

331 The right hand side of (5) is the penalty in quadratic form. The convenience of the quadratic
332 form is that it is a function of the estimated coefficients of $f(x_t)$ where \mathbf{S} is known as the penalty
333 matrix. Notice that now both the weights for the basis functions and the wigginess penalty
334 are expressed as functions of the model coefficients.

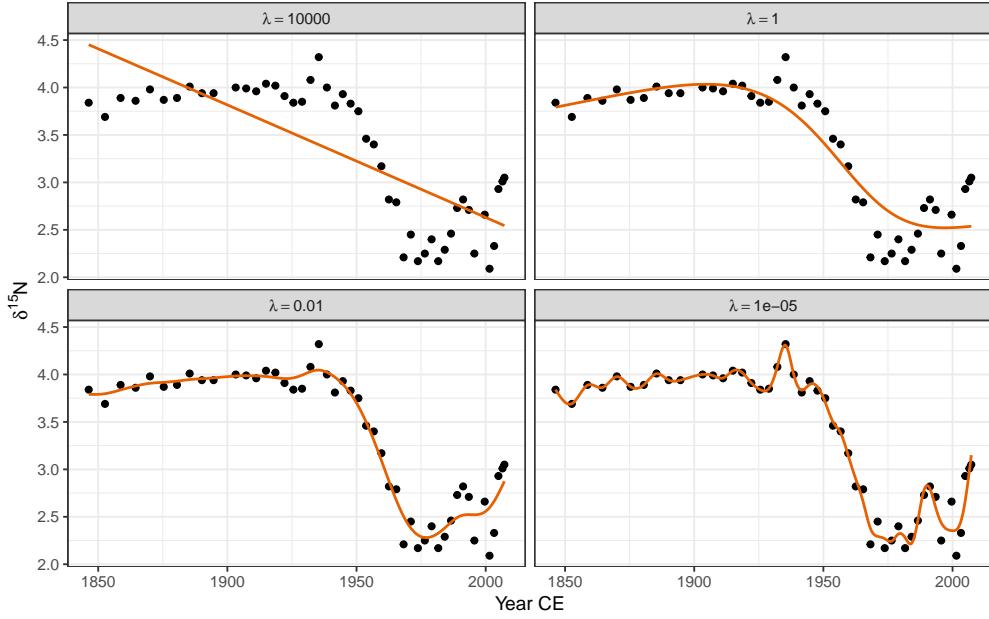


Figure 5: The effect of the smoothness parameter λ on the resulting wigginess of the estimated spline. Large values of λ penalize wigginess strongly, resulting in smooth trends (upper row), while smaller values allow increasingly wiggly trends. The aim of automatic smoothness selection is to find an optimal value of λ that balances the fit of the model with model complexity to avoid overfitting.

Now that we have a convenient way to measure wigginess, it needs to be incorporated into the objective function that will be minimised to fit the GAM. The likelihood of the model given the parameter estimates $\mathcal{L}(\beta)$ is combined with the penalty to create the penalized likelihood $\mathcal{L}_p(\beta)$:

$$\mathcal{L}_p(\beta) = \mathcal{L}(\beta) - \frac{1}{2}\lambda\beta^T S\beta.$$

The fraction of a half is there simply to make the penalised likelihood equal the penalised sum of squares in the case of a Gaussian model. λ is known as the smoothness parameter and controls the extent to which the penalty contributes to the likelihood of the model. In the extreme case of $\lambda = 0$ the penalty has no effect and the penalized likelihood equals the likelihood of the model given the parameters. At the other extreme, as $\lambda \rightarrow \infty$ the penalty comes to dominate $\mathcal{L}_p(\beta)$ and the wigginess of $f(x_t)$ tends to 0 resulting in an infinitely smooth function. In the case of a second derivative penalty, this is a straight line, and we recover the simple linear trend from (1) when assuming a Gaussian response.

Figure 5 illustrates how the smoothness parameter λ controls the degree of wigginess in the fitted spline. Four models are shown, each fitted with a fixed value of λ ; 10000, 1, 0.01, and 0.00001. At $\lambda = 10000$ the model effectively fits a linear model through the data. As the value of λ decreases, the fitted spline becomes increasingly wiggly. As λ becomes very small, the resulting spline passes through most of the $\delta^{15}\text{N}$ observations resulting in a model that is clearly overfitted to the data.

353 To fully automate smoothness selection for $f(x_t)$ we need to estimate λ . There are two main
 354 ways that λ can be automatically chosen during model fitting. The first way is to choose λ
 355 such that it minimises the prediction error of the model. This can be achieved by choosing λ
 356 to minimise Akaike's information criterion (AIC) or via cross-validation (CV) or generalized
 357 cross-validation (GCV; Craven and Wahba, 1978). GCV avoids the computational overhead
 358 inherent to CV of having to repeatedly refit the model with one or more observations left out
 359 as a test set. Minimising the GCV score will, with a sufficiently large data set, find a model
 360 with the minimal prediction error (Wood, 2017). The second approach is to treat the smooth
 361 as a random effect, in which λ is now a variance parameter to be estimated using maximum
 362 likelihood (ML) or restricted maximum likelihood (REML; Wood, 2011; Wood et al., 2016).
 363 Several recent results have shown that GCV, under certain circumstances, has a tendency to
 364 under smooth, resulting in fitted splines that are overly wiggly (Reiss and Ogden, 2009). Much
 365 better behaviour has been observed for REML and ML smoothness selection, in that order
 366 (Wood, 2011). REML is therefore the recommended means of fitting GAMs, though, where
 367 models have different fixed effects (covariates) they cannot be compared using REML, and
 368 ML selection should be used instead. In the sorts of data examples considered here there is
 369 only a single covariate x_t as our models contain a single estimated trend so REML smoothness
 370 selection is used throughout unless otherwise stated.

371 4 Fitting GAMs

372 4.1 Small Water

373 The trend in $\delta^{15}\text{N}$ values is clearly non-linear but it would be difficult to suggest a suitable
 374 polynomial model that would allow for periods of relatively no change in $\delta^{15}\text{N}$ as well as rapid
 375 change. Instead, a GAM is ideally suited to modelling such trends; the data suggest a smoothly
 376 varying change in $\delta^{15}\text{N}$ between 1925 and 1975. It is reasonable to expect some autocorrelation
 377 in the model errors about the fitted trend. Therefore I fitted the following GAM to the $\delta^{15}\text{N}$
 378 time series.

$$y_t = \beta_0 + f(x_t) + \varepsilon, \quad \varepsilon_t \sim (0, \Lambda\sigma^2) \quad (6)$$

379 Now the i.i.d. assumption has been relaxed and a correlation matrix, Λ , has been introduced
 380 that is used to model autocorrelation in the residuals. The $\delta^{15}\text{N}$ values are irregularly spaced in
 381 time and a correlation structure that can handle the uneven spacing is needed (Pinheiro and
 382 Bates, 2000). A continuous time first-order autoregressive process (CAR(1)) is a reasonable
 383 choice; it is the continuous-time equivalent of the first-order autoregressive process (AR(1))
 384 and, simply stated, models the correlation between any two residuals as an exponentially de-
 385 creasing function of $h(\phi^h)$, where h is the amount of separation in time between the residuals
 386 (Pinheiro and Bates, 2000). h may be a real valued number in the CAR(1), which is how it can
 387 accommodate the irregular separation of samples in time. ϕ controls how quickly the corre-
 388 lation between any two residuals declines as a function of their separation in time and is an
 389 additional parameter that will be estimated during model fitting. The model in (6) was fitted

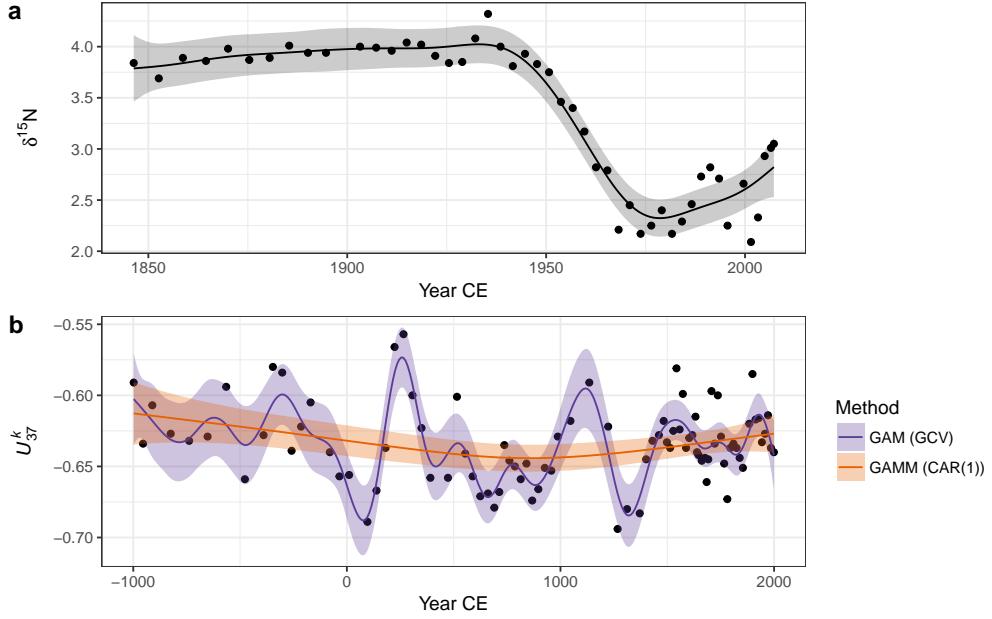


Figure 6: GAM-based trends fitted to the Small Water $\delta^{15}\text{N}$ (a) and Braya-Sø U_{37}^K (b) time series. The shaded bands surrounding the estimated trends are approximate 95% across-the-function confidence intervals. For the U_{37}^K series, two models are shown; the orange fit is the result of a GAM with a continuous-time AR(1) process estimated using REML smoothness selection, while the blue fit is that of a simple GAM with GCV-based smoothness selection. The REML-based fit significantly oversmooths the U_{37}^K time series.

390 using the `gamm()` function (Wood, 2004) in the `mgcv` package (Wood, 2017) for R (R Core Team,
391 2017).

392 The fitted trend is shown in Figure 6a, and well-captures the strong pattern in the data. The
393 trend is statistically significant (estimated degrees of freedom = 7.95; $F = 47.44$, approximate p
394 value = $\ll 0.0001$). However further analysis of the fitted model is required to answer the other
395 questions posed earlier about the timing of change and whether features in the trend can be
396 distinguished from random noise. I discuss these issues shortly.

397 4.2 Braya-Sø

398 The U_{37}^K data present a more difficult data analysis challenge than the $\delta^{15}\text{N}$ time series because
399 of the much more complex variation present. Fitting the same model as the Small Water ex-
400 ample, (6), to the U_{37}^K data resulted in the unsatisfactory fit shown as the very smooth line in
401 Figure 6b (labelled GAMM (CAR(1))). Further problems were evident with this model fit —
402 the covariance matrix of the model was non-positive definite, a sure sign of problems with the
403 fitted model. Refitting with a smaller basis dimension ($k = 20$) for the trend term resulted in
404 a model with a positive-definite covariance matrix for the model variance-covariance terms,
405 but the estimated value of the CAR(1) parameter $\phi = 0.2$ was exceedingly uncertain (95%
406 confidence interval 0 – 1!).

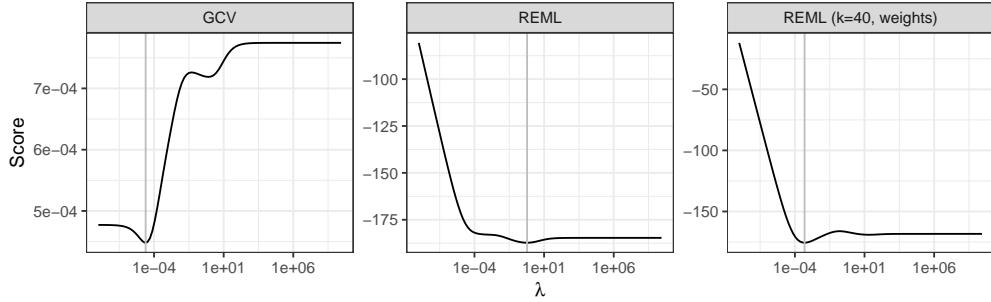


Figure 7: GCV and REML scores as a function of the smoothness parameter λ . From left to right, GAMs were estimated using GCV and REML smoothness selection, and REML using a basis dimension of 40 and observational weights to account for heterogeneity in the U_{37}^K times series. The selected value of λ for each model is indicated by the vertical grey line.

- 407 Fitting this model as a standard GAM with REML smoothness selection resulted in the same
 408 fitted trend as the GAM with CAR(1) errors (not shown), whilst using GCV smoothness se-
 409 lection resulted in a much more satisfactory fitted trend. There are two potential problems
 410 with the GCV-selected trend: i) GCV is sensitive to the profile of the GCV score and has been
 411 shown to under smooth data in situations where the profile is flat around the minimum GCV
 412 score, and ii) the model fitted assumes that the observations are independent, an assumption
 413 that is certainly violated in the U_{37}^K time series.
- 414 To investigate the first issue, the GCV and REML scores for an increasing sequence of values
 415 of the smoothness parameter (λ) were evaluated for the standard GAM (equation (4)) fit to the
 416 U_{37}^K time series. The resulting profiles are shown in Figure 7, with the optimal value of the
 417 parameter shown by the vertical line. The GCV score profile suggests that the potential for
 418 under smoothing identified by Reiss and Ogden (2009) is unlikely to apply here as there is a
 419 well-defined minimum in profile.
- 420 To understand the reason why the GAM plus CAR(1) and the simple GAM with REML smooth-
 421 ness selection performed poorly with the U_{37}^K time series we need to delve a little deeper into
 422 what is happening when we are fitting these two models.
- 423 The primary issue leading to poor fit is that neither model accounts for the different variance
 424 (known as (heteroscedasticity) of each observation in the U_{37}^K record. This seemingly isn't a
 425 problem for GCV which minimizes prediction error. The sediments in Braya-Sø are not an-
 426 nually laminated and therefore the core was sliced at regular depth intervals. Owing to com-
 427 paction of older sediments and variation in accumulation rates over time, each sediment slice
 428 represents a different number of "lake years". We can think of older samples as representing
 429 some average of many years of sediment deposition, whilst younger samples are representa-
 430 tive of fewer of these lake years. The average of a larger set of numbers is estimated more
 431 precisely than the average of a smaller set, all things equal. A direct result of this variable av-
 432 eraging of lake years is that some samples are more precise and therefore have lower variance
 433 than other samples and yet the model assumed that the variance was constant across samples.
- 434 Accounting for heteroscedasticity within the model is relatively simple via the use of observa-
 435 tional weights. The number of lake years represented by each slice is estimated by assigning a

date to the top and bottom of each sediment slice. The variance of each observation should be proportional to the inverse of the number of lake years each sample represents. In the `gam()` function used here, weights should be specified as the number of lake years each sample represents. Other software may require the weights to be specified in a different way.

A secondary problem is the size of the basis dimension used for the time variable. The main user selectable option when fitting a GAM in the penalised likelihood framework of Wood (2004) is how many basis functions to use. As described above, the basis should be large enough to contain the true, but unknown, function or a close approximation to it. For GCV selection the basis used contained 29 basis functions, whilst the CAR(1) model with REML smoothness selection would only converge with a basis containing 20 functions. The size of the basis appears to be sufficient for GCV smoothness selection, but following Wood (2011) REML smoothness selection is preferred. Using the test of Pya and Wood (2016), the basis dimension for the models with REML smoothness selection was too small. To proceed therefore, we must drop the CAR(1) term and increase the basis dimension to 39 functions (by setting `k = 40`; I set it larger than expected because the larger basis contains a richer family of functions and the excess complexity is reduced because of the smoothness penalty.)

With the larger basis dimension and accounting for the non-constant variance of the observations via weights, the model fitted using REML is indistinguishable from that obtained using GCV (Figure 6b). The trace of the REML score for this model shows a pronounced minimum at a much smaller value of λ than the original REML fit (Figure 7), indicating that a more wiggly trend provides a better fit to the Braya-Sø time series. This example illustrates that some care and understanding of the underlying principles of GAMs is required to diagnose potential issues with the estimated model. After standard modelling choices (size of basis to use, correct selection of response distribution and link function, etc.) are checked, it often pays to think carefully about the properties of the data and ensure that the assumptions of the model are met. Here, despite increasing the basis size, it was the failure to appreciate the magnitude of the effect on fitting of the non-constant variance that lead to the initially poor fit and the problems associated with the estimation of the CAR(1) process. I return to the issue of why the GAM plus CAR(1) model encountered problems during fitting later (see [Identifiability](#)).

4.3 Confidence intervals and uncertainty estimation

If we want to ask whether either of the estimated trends is statistically interesting or proceed to identifying periods of significant change, we must address the issue of uncertainty in the estimated model. What uncertainty is associated with the trend estimates? One way to visualise this is through a $1 - \alpha$ confidence interval around the fitted trend, where α is say 0.05 leading to a 95% interval. A 95% interval would be drawn at $\hat{y}_t \pm (m_{1-\alpha} \times \text{SE}(\hat{y}_t))$, with $m_{1-\alpha} = 1.96$, the 0.95 probability quantile of a standard normal distribution, and $\text{SE}(\hat{y}_t)$ is the standard error of the estimated trend at time x_t . This type of confidence interval would normally be described as *pointwise*; the coverage properties of the interval being correct for a single point on the fitted trend, but, if we were to consider additional points on the trend, the coverage would logically be lower than $1 - \alpha$. This is the traditional frequentist interpretation of a confidence interval. However, the GAM described here has a Bayesian interpretation (Kimeldorf and Wahba, 1970;

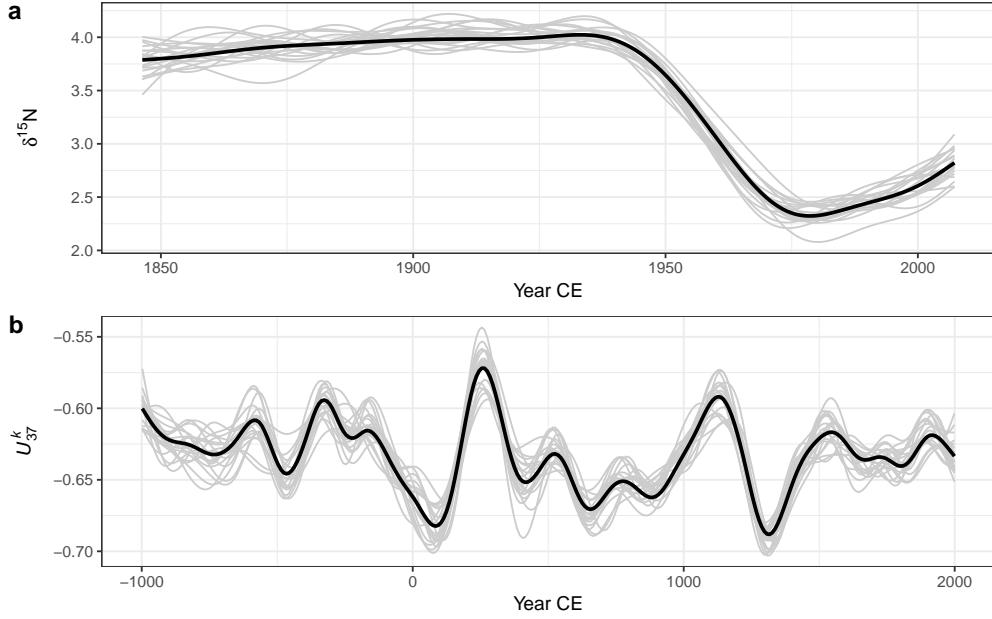


Figure 8: Estimated trends (thick black lines) and 20 random draws (grey lines) from the posterior distribution of the GAM fitted to the Small Water $\delta^{15}\text{N}$ (a) and Braya-Sø U_{37}^K (b) time series.

477 Silverman, 1985; Wahba, 1983, 1990) and therefore the typical frequentist interpretation does
 478 not apply. Nychka (1988) investigated the properties of a confidence interval created as de-
 479 scribed above using standard errors derived from the Bayesian posterior covariance matrix for
 480 the estimated mode parameters. Such intervals have the interesting property that they have
 481 good *across-the-function* coverage when considered from a frequentist perspective. This means
 482 that, when averaged over the range of the function, the Bayesian credible intervals shown in
 483 Figure 6 have close to the expected 95% coverage. However, to achieve this some parts of the
 484 function may have more or less than 95%-coverage. Marra and Wood (2012) recently explained
 485 Nychka's (1988) surprising results and extended them to the case of generalized models (non-
 486 Gaussian responses).

487 Whilst the *across-the-function* frequentist interpretation of the Bayesian credible intervals is use-
 488 ful, it may be important to have an interval that contains the entirety of the true function with
 489 some state probability $(1 - \alpha)$. Such an interval is known as a *simultaneous* interval. A $(1 - \alpha)100\%$
 490 simultaneous confidence interval contains *in their entirety* $1 - \alpha$ of all random draws from the
 491 posterior distribution of the fitted model.

492 Fitting a GAM involves finding estimates for coefficients of the basis functions. Together, these
 493 coefficients are distributed multivariate normal with mean vector and covariance matrix spec-
 494 ified by the model estimates of the coefficients and their covariances respectively. Random
 495 draws from this distribution can be taken, where each random draw represents a new trend
 496 that is consistent with the fitted trend but also reflects the uncertainty in the estimated trend.
 497 This process is known as *posterior simulation*.

498 Figure 8 shows 20 random draws from the posterior distribution of the GAMs fitted to the

499 Small Water and Braya-Sø data sets. In the early period of the $\delta^{15}\text{N}$ time series many of the
500 posterior simulations exhibit short periods of increasing and decreasing trend, balancing out
501 to the relatively flat trend estimated by the GAM (Fig. 8a). Reflecting this uncertainty, we
502 might expect relatively wide simultaneous intervals during this period in order to contain the
503 vast majority of the simulated trends. Conversely, the decreasing $\delta^{15}\text{N}$ trend starting at ~1945
504 is consistently reproduced in the posterior simulations, suggesting that this feature of the time
505 series is both real and statistically significant, and that the rate of change in $\delta^{15}\text{N}$ is relatively
506 precisely estimated. We see a similar pattern in Figure 8b for the Braya-Sø record; the large
507 peak in U_{37}^K at ~250CE and the strong decline at ~1200CE are well defined in the posterior
508 simulations, whereas most of the localised trends that are smaller magnitude changes in y_t
509 are associated with posterior simulations that are less well constrained with the ends of the
510 record in particular showing considerable variation in the strength, timing, and even sign of
511 simulated trends, reflecting the greater uncertainty in estimated trend during these periods.
512 For the random draws illustrated in Figure 8, a $(1 - \alpha)100\%$ simultaneous interval should con-
513 tain the entire function for on average 19 of the 20 draws.

514 There are a number of ways in which a simultaneous interval can be computed. Here I follow
515 the simulation approach described by Ruppert et al. (2003) and present only the basic detail; a
516 fuller description is contained in Appendix 1. The general idea is that if we want to create an
517 interval that contains the whole of the true function with $1 - \alpha$ probability, we need to increase
518 the standard Bayesian credible interval by some amount. We could simulate a large number
519 of functions from the posterior distribution of the model and then search for the value of $m_{1-\alpha}$
520 that when multiplied by $\text{SE}(\hat{f}(x_t))$ yielded an interval that contained the whole function for
521 $(1 - \alpha)$ 100% of the functions simulated. In practice, the simulation method of Ruppert et al.
522 (2003) does not involve a direct search, but yields the critical value $m_{1-\alpha}$ required.

523 Simultaneous intervals computed using the method described are show in Figure 9 alongside
524 the *across-the-function* confidence intervals for the trends fitted to both example data sets. As
525 expected, the simultaneous interval is somewhat wider than the *across-the-function* interval.
526 The critical value $m_{1-\alpha}$ for the simultaneous interval of the estimated trend in $\delta^{15}\text{N}$ is 3.08,
527 whilst the same value for the U_{37}^K series is 3.42, leading to intervals that are approximately
528 $\pm 50\%$ and $\pm 75\%$ wider than the equivalent across-the-function intervals.

529 4.4 Identifying periods change

530 In the simple linear trend model (1) whether the estimated trend constitutes evidence for or
531 against a null hypothesis of no change rests on how large the estimated rate of change in y_t
532 is ($\hat{\beta}_1$) relative to its uncertainty. This is summarised in the t statistic. As the rate of change
533 in y_t is constant over the fitted trend — there is only a singe slope for the fitted trend $\hat{\beta}_1$ — if
534 the t statistic of the test that $\hat{\beta}_1 = 0$ is unusually extreme this would be evidence against the
535 null hypothesis of no change. Importantly, this applies to the whole time series as the linear
536 model implies a constant rate of change throughout. More formally, the estimate $\hat{\beta}_1$ is the first
537 derivative of the fitted trend.

538 In the GAM, the fitted trend need not be linear; the slope of the trend is potentially different
539 at every point in the time series. As such we might reasonably ask *where* in the series the

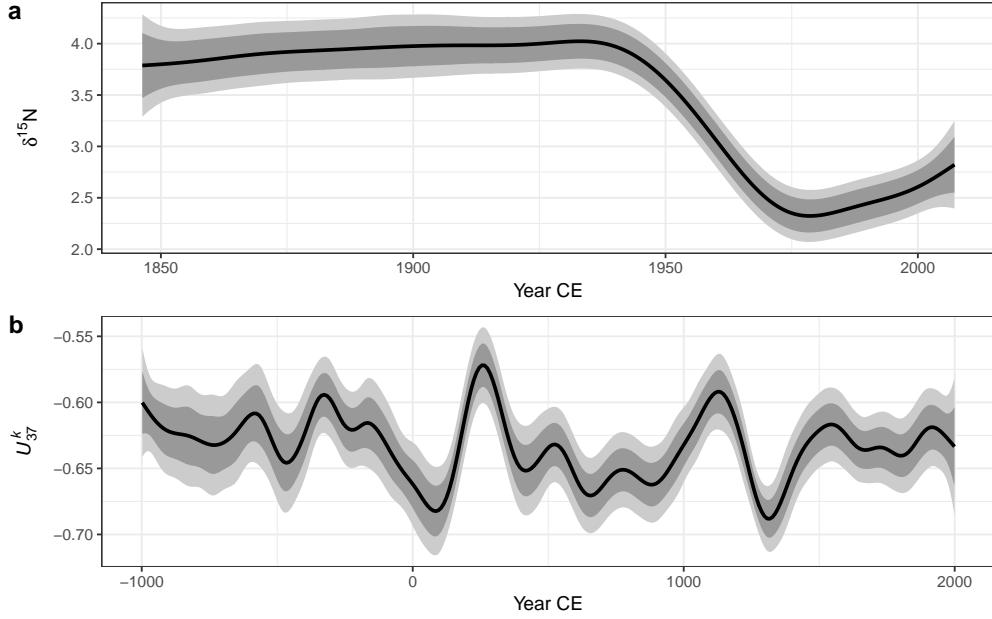


Figure 9: 95% simultaneous confidence intervals (light grey bands) and across-the-function confidence intervals (dark grey bands) on the estimated trends (black lines) for the Small Water $\delta^{15}\text{N}$ (a) and Braya-Sø U_{37}^K (b) time series.

540 response y_t is changing, if at all? Mirroring the linear model we can answer this question by
 541 determining whether or not the first derivative at any time point x_t of the fitted trend at any
 542 time point is consistent with a null hypothesis of no change. We want to know whether or not
 543 the first derivative is indistinguishable from a value of 0 — no trend — given the uncertainty
 544 in the estimate of the derivative.

545 Derivatives of the fitted spline are not easily available analytically, but they can be estimated
 546 using the method of finite differences. Two values of the estimated trend, separated by a very
 547 small time-shift (Δ_t), are predicted from the model; the difference between the estimated values
 548 for the two time points is an approximation of the true first derivative of the trend. As $\Delta_t \rightarrow 0$
 549 the approximation becomes increasingly accurate. In practice, the first derivative of the fitted
 550 trend is evaluated using finite differences at a large number of points in the time series. An
 551 approximate $(1 - \alpha)100\%$ pointwise confidence interval can be calculated for the derivative
 552 estimates using standard theory (i.e. $\pm 1.96 \times \text{SE}(\hat{y}_t)$ for a 85% interval) and the covariance matrix
 553 of the spline coefficients. A $(1 - \alpha)100\%$ simultaneous interval for the derivatives can also be
 554 computed using the method described [earlier](#). Periods of significant change are identified as
 555 those time points where the (simultaneous) confidence interval on the first derivative does not
 556 include zero.

557 Figure 10 shows the estimated first derivative of the fitted trend in the Small Water (10a) and
 558 Braya-Sø (10b) time series. Although the estimated trend suggests a slight increase in $\delta^{15}\text{N}$
 559 from the start of the record to ~ 1940 , the estimated trend is sufficiently uncertain that the si-
 560 multaneous interval on the first derivative includes 0 throughout. We can understand why this
 561 is so by looking at the posterior simulations in Figure 8a; there is considerable variation in the

shape of the simulated trends throughout this period. From ~1925 the derivative of the trend becomes negative, however it is not until ~1940 that the simultaneous interval doesn't include 0. At this point we have evidence to reject the null hypothesis of no change. This time point may be taken as the first evidence for change in $\delta^{15}\text{N}$ in the Small Water core. The simultaneous interval on the first derivative of the trend in $\delta^{15}\text{N}$ is bounded away from 0 between ~1940 and ~1975, covering the major decline in values evident in the observations. The simultaneous interval includes 0 from ~1975 onward, suggesting that, whilst quite pronounced, the recent increase in $\delta^{15}\text{N}$ is not statistically significant. To determine whether or not the recent increase is real, we would require considerably more samples with which to (hopefully) more-precisely estimate the trend during this period. Alternatively, we might just have to wait until sufficient additional sedimentation has occurred to warrant recoresing Small Water and reestimating the trend in $\delta^{15}\text{N}$.

The estimated trend at Braya-Sø exhibited a number of oscillations in U_{37}^K . As we saw previously in Figures 8b and 9b, many of these are subject to significant uncertainty and it is important therefore to discern which, if any, of the oscillations in the response can be identified given the model uncertainty. In Figure 10b only two features of the estimated trend are considered significant based on the derivatives of the smooth; one centred on ~250CE and a second at ~1150CE. In both these periods, the simultaneous interval for the first derivative of the trend does not include zero. In the first case we detect the large peak and subsequent decline in U_{37}^K at ~250CE, whilst at ~1150CE the large trough is identified, but not the increasing trend immediately prior to this excursion to lower U_{37}^K . Recall that these intervals are simultaneous in nature, strongly guarding against false positives, and as such we can be confident in the estimation of these two features, whilst care must be taken to not over-interpret the remaining variations in the estimated trend.

4.5 Residual autocorrelation and model identification

The GAM fitted to the $\delta^{15}\text{N}$ time series contained a CAR(1) process to model residual temporal autocorrelation in the residuals. The estimated magnitude of the autocorrelation is given by the parameter ϕ . The estimated value of ϕ for the $\delta^{15}\text{N}$ series is 0.6 with 95% confidence interval 0.28–0.85, indicating moderate to strong residual autocorrelation about the fitted trend. The correlation function is an exponentially decreasing function of temporal separation (Δ_t), and whilst observations that are a few years apart are quite strongly dependent on one another, this dependence drops off rapidly as Δ_t increases and is effectively zero when samples are separated by a decade or more (Figure 11).

Failure to account for the dependencies in the $\delta^{15}\text{N}$ time series could lead to the estimation of a more wiggly trend than the one shown in Figure 6a which would negatively impact the confidence placed on the inferences we might draw from the fitted model. Importantly, failing to account for the strong dependency in the residuals would lead to smaller uncertainties in the estimated spline coefficients, which would propagate through to narrower confidence intervals on the fitted trend and on the first derivatives, and ultimately to the identification of significant periods of change. The end result would be a tendency toward anti-conservative identification of periods of change; the coverage probability would be lower than the antici-

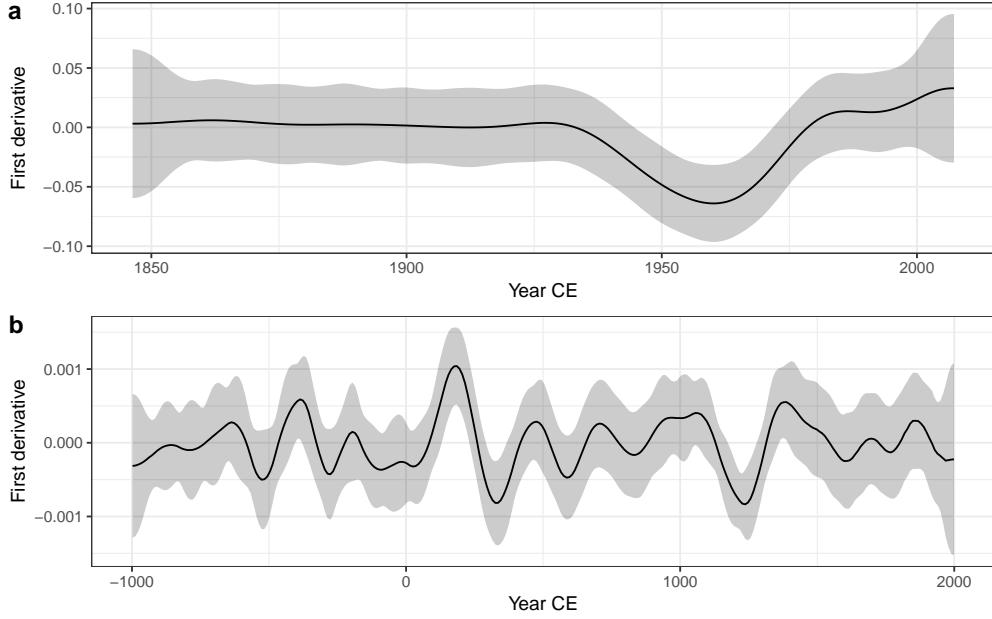


Figure 10: Estimated first derivatives (black lines) and 95% simultaneous confidence intervals of the GAM trends fitted to the Small Water $\delta^{15}\text{N}$ (a) and Braya-Sø U_{37}^K (b) time series. Where the simultaneous interval does not include 0, the models detect significant temporal change in the response.

603 pated $1 - \alpha$, leading to a greater chance of false positive results.

604 Problems estimating the GAM plus CAR(1) model were encountered when this was fitted to
 605 the U_{37}^K time series; including both a smooth trend in the mean U_{37}^K and a CAR(1) process in
 606 the residuals lead to an unidentifiable model. What makes a model with a spline-based trend
 607 and an autocorrelation process like the CAR(1) potentially unidentifiable?

608 Consider again the basic GAM for a smooth trend, (3). In that equation the correlation ma-
 609 trix Λ was omitted for the sake of simplicity. As I did in (6), I reintroduce it and restate the
 610 distributional assumptions of this model

$$y_t = \beta_0 + f(x_t) + \varepsilon_t, \quad \varepsilon \sim (0, \Lambda\sigma^2) \quad (7)$$

611 In the basic GAM, $\Lambda \equiv \mathbf{I}$ is an identity matrix, a matrix with 1s on the diagonal and 0s elsewhere

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix},$$

612 which is where the independence assumption of the model comes from; a model residual is
 613 perfectly correlated with itself (the 1s on the diagonal), but uncorrelated with any other resid-
 614 ual (the off-diagonal 0s). In the GAM plus CAR(1) model, an alternative correlation function

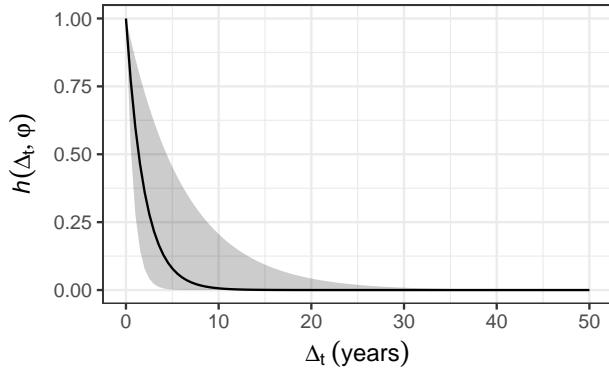


Figure 11: Estimated CAR(1) process from the GAM fitted to the Small Water $\delta^{15}\text{N}$ time series. $h(\Delta_t, \phi)$ is the correlation between residuals separated by Δ_t years, where $\hat{\phi} = 0.6$. The shaded band is a 95% pointwise confidence interval on the estimated correlation h .

- for Λ was used — the CAR(1) with correlation parameter ϕ . Fahrmeir and Kneib (2008) show that where the stochastic structure of f and Λ approach one another, i.e. where we have a potentially wiggly trend or strong autocorrelation as $\phi \rightarrow 1$, the two processes can quickly become unidentifiable (see also Fahrmeir et al., 2013). By unidentifiable, we mean that it becomes increasingly difficult to distinguish between a wiggly trend or strong autocorrelation because these two processes are very similar to one another in appearance. This leads to model estimation problems of the sort encountered with fitting the GAM plus CAR(1) model to the Braya-sø U_{37}^K series.
- Why might this be so? Autocorrelation is the tendency for a large (small) value of y_t at time x_t to be followed by a likewise large (small) value at time x_{t+1} . This leads to runs of values that are consistently greater (less) than the overall mean. Short runs would indicate weaker autocorrelation whilst longer runs are associated with stronger autocorrelation, and long runs of values greater (less) than the mean would be evident as non-linear trends in the time series. As a result, a wiggly trend and an autocorrelation function with large ϕ are two ways to describe the same pattern of values in a time series, and without any further information to constrain either the model is unable to distinguish both components uniquely.
- Situations where it may be possible to uniquely identify separate wiggly trends and autocorrelation are exemplified by the Small Water $\delta^{15}\text{N}$ time series. The non-linear trend and the autocorrelation operate at very different scales; the trend represents decadal-scale variation in mean $\delta^{15}\text{N}$, whilst the CAR(1) process represents the much smaller-scale tendency for values of the response to be followed in time by similar values. That such a pattern is observed in the Small Water core is the result of the high resolution of the sampling in time relative to the long-term trend. In contrast, the Braya-Sø record is sampled at far lower resolution relative to the fluctuations in the mean response, and consequently the data do not contain sufficient information to separate trend and autocorrelation.

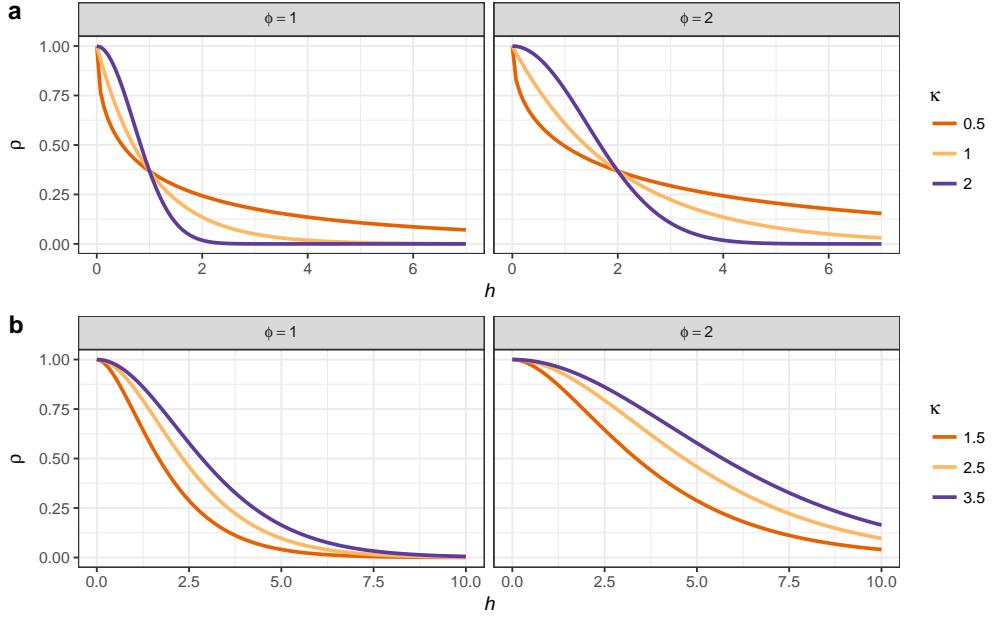


Figure 12: Power exponential (a) and Matérn (b) correlation functions for observation separation distance h . Two values of the effective range parameter (ϕ) are shown for each function. For the power exponential function, κ is the power in the power exponential function. For the Matérn correlation function, κ distinguishes the member of the Matérn family.

640 4.6 Gaussian process smooths

641 In the world of machine learning, Gaussian processes (Golding and Purse, 2016; Rasmussen
 642 and Williams, 2006) are a widely-used method for fitting smooth non-parametric regression
 643 models. A Gaussian process is a distribution over all possible smooth functions $f(x)$. In the
 644 field of spatial statistics, Gaussian processes are known by name *kriging*.

645 With a Gaussian process we are interested in fitting a smooth temporal trend by modelling
 646 the way the correlation between pairs of observations varies as a function of the distance, h , in
 647 time that separates the observations. The correlation between pairs of observations decreases
 648 with increasing separation, which is modelled using a correlation function, $c(h)$.

649 Several functions can be used to represent $c(h)$. Two common ones are the power exponential
 650 function and the Matérn family of correlation functions. The power exponential function at
 651 separation distance h is

$$c(h) = \exp\{(-h/\phi)^\kappa\}$$

652 where $0 < \kappa \leq 2$. The Matérn correlation function is actually a family of functions with closed-
 653 forms only available for a subset of the family, distinguished by κ . When $\kappa = 1.5$, the Matérn
 654 correlation function is

$$c(h) = (1 + h/\phi) \exp(-h/\phi)$$

655 whilst for $\kappa = 2.5$ it is

$$c(h) = \{1 + h/\phi + (h/\phi)^2/3\} \exp(-h/\phi)$$

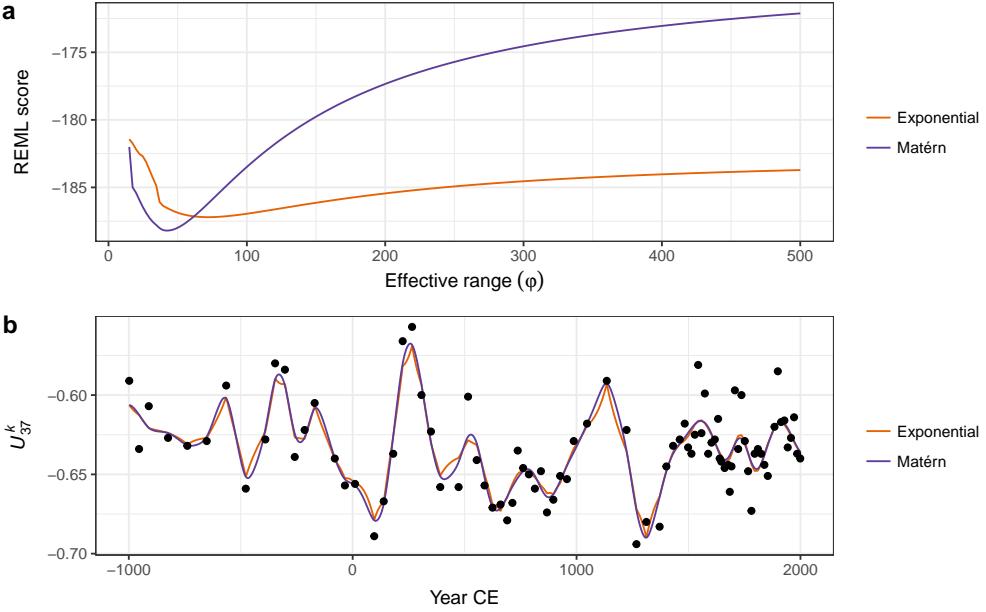


Figure 13: Gaussian process smooths fitted to the U_{37}^K time series. REML score traces for GAMs fitted using power exponential ($\kappa = 1$) or Matérn ($\kappa = 1.5$) correlation functions as a function of the effective range parameter (ϕ) are shown (a). The optimal model for each function is that with the lowest REML score. b) shows the resulting trends estimated using the respective correlation function with the value of ϕ set to the optimal value.

and for $\kappa = 3.5$

$$c(h) = \{1 + h/\phi + 2(h/\phi)^2/5 + (h/\phi)^3/15\} \exp(-h/\phi).$$

In all cases, ϕ is the effective range parameter, which sets the distance beyond which the correlation function is effectively zero.

Figure 12 shows examples of two different correlation functions; the *power exponential* (Figure 12a), and the Matérn (Figure 12b) correlation functions. These functions are smooth and monotonic-decreasing, meaning that the value of the correlation function decreases with increasing separation (h). When $h = 0$, the correlation is equal to 1 ($c(0) = 1$); two samples taken at exactly the same time point are perfectly correlated. As $h \rightarrow \infty$, the correlation tends to zero ($c(h) \rightarrow 0$); two samples separated by a large amount of time tend to be uncorrelated. Often we are interested in learning how large the separation in time needs to be before, on average, a pair of observations is effectively uncorrelated (i.e. where $c(h)$ is sufficiently close to zero).

Gaussian processes and GAMs share many similarities and we can fit a Gaussian process using the techniques already described above for splines (Handcock et al., 1994; Kammann and Wand, 2003). It can be shown (e.g. Fahrmeir et al., 2013) that the Gaussian process model has the same penalised likelihood form as the GAM that we discussed earlier; the observations are the knots of the smoother and each has a basis function in the form of a correlation function. The equivalence is only true if the basis functions do not depend on any other parameters of the model, which is only achievable if the value of ϕ is fixed and known (Fahrmeir et al., 2013). In general, however, we would like to estimate ϕ as part of model fitting. To achieve this we

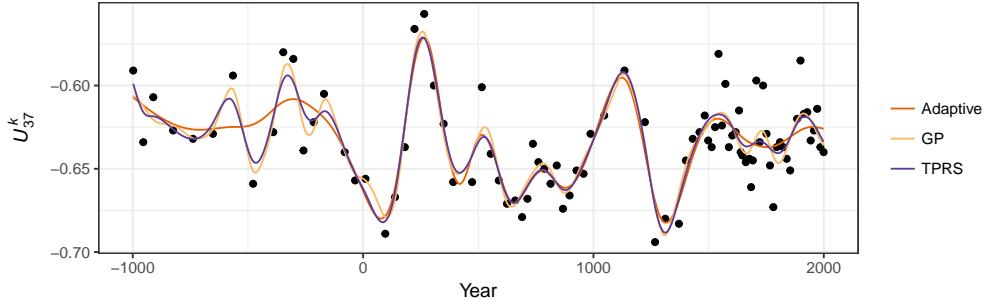


Figure 14: Comparison of trends estimated using i) adaptive smooth, ii) Gaussian process, and iii) thin plate regression spline bases for the U_{37}^k time series.

can maximise the profile likelihood or score statistic of the model over a range of values of ϕ (Wood, 2017, 362–363). This involves proposing a value of ϕ for the effective range of the correlation function and then estimating the resulting GAM by minimising the penalised log-likelihood conditional upon this value of ϕ and repeating for a range of values for ϕ . The model, and its corresponding value of ϕ , with lowest penalised log-likelihood or score statistic is then retained as the estimated GAM. Figure 13a shows the REML score for models estimated using a Gaussian process smooth with a Matérn correlation function ($\kappa = 1.5$) for a sequence of values of ϕ between 15 and 1000 years. There is a clear minimum around 40 years separation, with the minimum REML score being observed at $\phi = 41.81$). Also shown are the REML scores for models using the power exponential function ($\kappa = 1$) with the minimum score observed at a somewhat higher effective range of $\phi = 71.06$.

Figure 13b shows the estimated trends for the U_{37}^k time series using Gaussian process smooths with exponential and Matérn correlations functions, both using ϕ values at their respective optimal value as assessed using the REML score. The estimated trends are very similar to one another, although there is a noticeable difference in behaviour, with the power exponential ($\kappa = 1$) version being noticeably less-smooth than the Matérn version. This difference is attributable to the shapes of the respective correlation functions; the Matérn approaches a correlation of 1 smoothly as h approaches 0, whilst the power exponential with $\kappa = 1$ approaches a correlation of 1 increasingly quickly with decreasing h . The power exponential with $\kappa = 2$, like the Matérn, approaches $\phi = 1$ smoothly, and consequently the trend estimated using this correlation function is qualitatively similar to that estimated using the Matérn correlation function.

4.7 Adaptive smoothing

Each of the spline types that I have discussed so far share a common feature; the degree of wiggliness over the time series is fixed due to the use of a single smoothness parameter, λ . The definition of wiggliness, as the integrated squared second derivative of the spline, ensures that the fitted smoother does not jump about wildly. This assumes that the data themselves are well described by a smoothly varying trend. If we anticipate abrupt change or step-like responses to environmental forcing this underlying assumption of the GAM would suggest that the method

- 704 is ill-suited to modelling palaeo time series in which such features are evident or expected.
- 705 While there is not much we can do within the GAM framework to model a series that contains
706 both smooth trends and step-like responses, an adaptive smoother can help address problems
707 where the time series consists of periods of rapid change in the mean combined with periods
708 of complacency or relatively little change. As suggested by their name, adaptive smoothers
709 can adjust to changes in the wigginess of the time series. This adaptive behaviour is achieved
710 by making the smoothness parameter λ itself depend smoothly on x_t (Ruppert et al., 2003,
711 17; Wood, 2017, 5.3.5); in other words, the adaptive smoother allows the wigginess of the
712 estimated trend to vary smoothly over time. Whilst this allows the estimated trend to adapt
713 to periods of rapid change in the response, adaptive smoothers make significant demands on
714 the data (Wood, 2017, 5.3.5); if we used m smoothness penalties to allow the wigginess to
715 vary over a time series, it would be like estimating m separate smooths from chunks of the
716 original series each of length n/m . In a practical sense, this limits the use of adaptive splines
717 in palaeoecology to proxies that are readily enumerated, such as the biogeochemical proxies
718 used in the two example data sets.
- 719 Figure 14 compares trends for the Braya-Sø time series estimated using GAMs with the three
720 main types of spline discussed; i) TPRS, ii) Gaussian process smooths, and iii) an adaptive
721 smoother using 45 basis functions and 5 smoothing parameters. There is a clear difference
722 in the behaviour of the adaptive and non-adaptive smoothers for the first 1000 years of the
723 record, with the adaptive smooth exhibiting much less variation compared with either the
724 TPRS or Gaussian process splines. Over the remaining two thirds of the series, there is much
725 closer agreement in the three smooths.
- 726 The behaviour of the TPRS and Gaussian process splines for these data is the result of requiring
727 a large amount of wigginess (a small λ) to adapt to the large oscillations in U_{37}^K present around
728 year 250CE and again at ~900–1500CE. This large degree of wigginess allows the splines to
729 potentially over-fit individual data points much earlier in the record. Because the adaptive
730 smoother, in contrast, can adapt to these periods of rapid change in the response it is much
731 less susceptible to this “chasing” behaviour — we don’t need to waste effective degrees of
732 freedom in periods with little or no change just to be able to fit the data well when there is a
733 lot of change.
- 734 This potential for over-fitting in such situations is undesirable, yet if we recall Figure 10 and
735 the discussion around the use of the first derivative to identify periods of significant change,
736 we would not interpret the oscillations in the early part of the U_{37}^K record as being statistically
737 significant. Owing to the paucity of data in this part of the series the trends fitted using the
738 non-adaptive smoothers are subject to such a large degree of uncertainty that the alternative
739 of no trend through the first 1000 years of the record is also a plausible explanation of the data.
740 The trend estimated using the adaptive smooth reflects this. Therefore, should we conclude
741 that there is no trend in U_{37}^K and thence climate in this period? I believe that to be too-strong
742 a statement; those oscillations in U_{37}^K may be real responses to climate forcing but may simply
743 lack the statistical power to distinguish them from the null hypothesis of no trend through this
744 period. The adaptive smoother is only adjusting to the data available to it; just because it does
745 not detect a trend during this period does not lend itself to an interpretation of stable climate
746 forcing or complacency in the lake’s response to forcing. If there were particular interest in the

747 climate of this particular period we might take from the Braya-Sø record that there is potential
748 early variation in climate forcing, but that additional data from this or other sites is required
749 before any definitive conclusion can be drawn.

750 4.8 Accounting for age model uncertainty

751 Thus far, the trend models that I have described and illustrated assumed that the time covariate
752 (x_t) was fixed and known. In both examples, and more generally for most palaeoecological
753 records, this assumption is violated. Unless the record is annually laminated, assigning an
754 age to a sediment interval requires the development of an age model from observations of the
755 relationship between depth down the sediment core and estimates of the age of the sample
756 arrived at using any of a number of techniques, for example ^{210}Pb or ^{14}C radiometric dating.
757 This age-depth relationship is itself uncertain, usually being derived from a mathematical or
758 statistical model applied to point age estimates (e.g. Blaauw and Heegaard, 2012). Incorporating
759 this additional component of uncertainty complicates the estimation of statistical models
760 from palaeoenvironmental data. In this section I illustrate a simulation based approach to
761 quantify and account for age-model uncertainty as part of the trend estimation using a GAM
762 (see Anchukaitis and Tierney (2013) for a similar, non-GAM related idea).

763 Figure 15a shows the estimated dates (in Years CE) for 12 levels in the Small Water core dated
764 using ^{210}Pb . The vertical bars show the estimated age uncertainty of each level. The solid line
765 through the data points is an additive model fitted to the observations, with prior weights
766 given by the estimated age uncertainties. The fitted age-depth model is constrained to be
767 monotonically decreasing with increasing depth, following the method of (Pya and Wood,
768 2015) using the *scam* package (Pya, 2017). Also shown are 25 simulations from the posterior
769 distribution of the monotonically-constrained GAM. Each simulation from the posterior dis-
770 tribution of the age-model is itself a potential age-depth model, which can be used to assign
771 dates to the Small Water core. The trend model in (4) can be fitted to the $\delta^{15}\text{N}$ data using these
772 new dates as x_t , and the whole process repeated for a large number of simulations from the
773 age model.

774 Figure 15b shows the trend in $\delta^{15}\text{N}$ for the observed age-depth model, plus trends estimated
775 via the same model using 100 draws from the posterior distribution of the age model. In this
776 case, the age-depth model is relatively simple with little variation in the posterior draws, re-
777 sulting in trends that match closely that obtained from the estimated age-depth relationship.
778 Even so, this additional uncertainty suggests that the timing of the decline in $\delta^{15}\text{N}$ covers the
779 interval ~1935–1945.

780 The uncertainty in the trend estimates illustrated in Figure 15b only reflects the variation in
781 trends fitted to the uncertain dates of the sediment samples. To fully visualise the uncertainty
782 in the trend estimates, incorporating both age model uncertainty *and* uncertainty in the esti-
783 mated model coefficients themselves, 50 simulations from the posterior distribution of each
784 of the 100 estimated trends shown in Figure 15b were performed, resulting in 5,000 trend esti-
785 mates for the $\delta^{15}\text{N}$ series. These are shown in Figure 15c, where the two obvious changes over
786 the same simulations without accounting for uncertainty in x_t (Figure 8a) are that the uncer-
787 tainty band traced out by the simulations is approximately 50% wider and, not surprisingly,

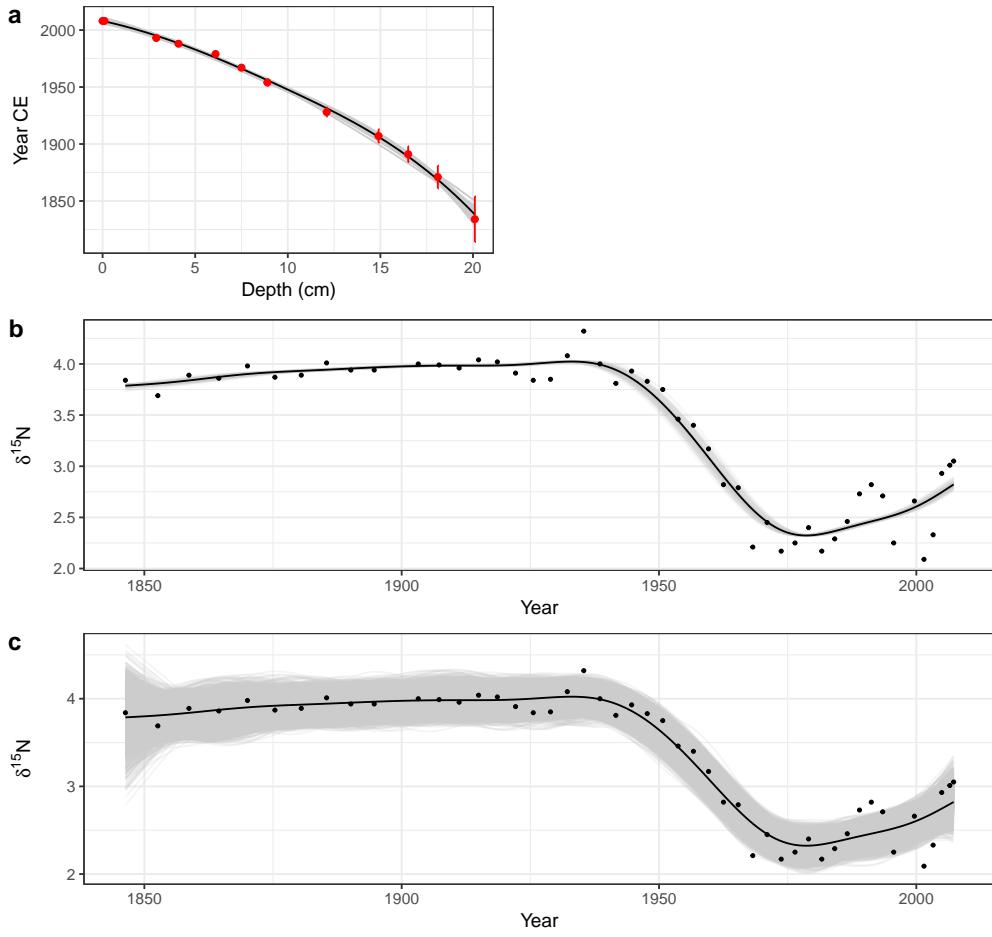


Figure 15: Accounting for uncertainty in age estimates whilst fitting a smooth trend to the Small Water $\delta^{15}\text{N}$ time series. (a) Estimated age model using a monotonically-constrained spline fitted to ^{210}Pb inferred ages for selected depths in the sediment core (red points). The uncertainty in the ^{210}Pb inferred age is show by the red vertical bars. The fitted age model is illustrated by the solid black line. The faint grey lines are 25 random draws from the posterior distribution of the monotonically constrained GAM. The effect of age uncertainty on trend estimation is shown in b); for 100 simulations from the posterior distribution of the age model in a) a trend was estimated using a GAM with a thin plate regression spline basis and a CAR(1) process in the residuals. These trends are shown as grey lines. The combined effect of age model and fitted GAM uncertainty on the trends for the $\delta^{15}\text{N}$ time series is shown in c). The grey lines in c) are based on 50 random draws from the model posterior distribution for each of the 100 trends shown in b). For both b) and c) the black line shows the trend estimated assuming the ages of each sediment sample are known and fixed.

788 the uncertainty in the estimated trend is most pronounced in the least accurately-dated sec-
789 tion of the core. Despite this additional uncertainty however, the main result holds; a marked
790 decline of ~1.5‰ that occurred between approximately 1930 and 1945, with mild evidence of
791 a small increase in $\delta^{15}\text{N}$ post 2000 CE.

792 4.9 Multivariate data

793 A large proportion of the palaeoenvironmental data generated today is multivariate in nature
794 and yet the two examples used to illustrate GAMs were univariate. Can the approach de-
795 scribed here be used for multivariate data? Yes, and no. With one main exception it is not
796 possible to directly apply the GAM methodology described here to multivariate abundance
797 data, where the aim is to model all species at once. The *mgcv* software, for example, is not able
798 to estimate the penalized GAM for multiple non-Gaussian responses. The exception is for a
799 small number of correlated Gaussian responses; these could be modelled as being distributed
800 multivariate normal conditional upon the covariates. Such a model would estimate the ex-
801 pected values of each response and the correlations between them. For example, we could
802 jointly model $\delta^{15}\text{N}$ and $\delta^{13}\text{C}$ series using this approach.

803 Formal multivariate versions of GLM or GAMs are currently an important area of research
804 within ecology (see Warton et al. (2015) for a recent review), where they go by the name joint
805 species distribution models (JSDMs). Whilst undoubtedly powerful, our knowledge regard-
806 ing JSDMs and their availability in software are still in their relative infancy and they require
807 considerable expertise to implement. As such, JSDMs are currently beyond the reach of most
808 palaeoecologists. Despite this, we should be watching JSDM research as developments are
809 ongoing and a degree of method maturation occurring.

810 One currently available avenue for fitting a multivariate GAM is via regularized sandwich
811 estimators and GLMs (Warton, 2011), which involves fitting separate GLMs (or GAMs) to
812 each response variable and subsequently using resampling-based hypothesis tests to deter-
813 mine which covariates are related to variation at the community level and for individual taxa
814 (Wang et al., 2012; Warton, 2011; Warton et al., 2012). The *mvabund* package (Wang et al., 2012)
815 implements this approach within R and can use *mgcv* to fit GAMs to each species.

816 A pragmatic although inelegant approach that has been used to estimate trends in multivariate
817 palaeoecological data is to first summarise the response data using an unconstrained ordina-
818 tion via a PCA, CA, or principal curve and then fit separate GAM models to the site (sample)
819 scores of the first few ordination axes or principal curve (Beck et al., 2018; Bennion et al., 2015).
820 Whilst this two-step approach is relatively easy to implement and builds on approaches that
821 palaeoecologists already use to summarise multivariate stratigraphic data, it is best thought of
822 as modelling changes in abundance or relative composition at the community level. It is less
823 well suited to unpicking taxon-specific trends however, because the ordination step combines
824 individual species information into latent variables (axes) that are linear combinations of *all*
825 species and it is these latent variables that are then modelled using GAM.

826 5 Conclusions

827 Formal statistical estimation of trends in palaeoenvironmental data has been hampered by
828 the nature of the data that comprise the time series; the uneven spacing of samples in time
829 makes it, if not impossible, difficult to fit classical statistical time series models like ARIMA.
830 This has lead palaeoecologists and palaeolimnologists to fall back on basic statistical methods
831 such as linear parametric and non-parametric correlations or simple linear regression models,
832 where the assumptions of the method are often grossly violated by the dependencies inherent
833 to time series data. GAMs, whilst similar to the popular Loess smoother, provide a superior
834 alternative approach to trend estimation in palaeoenvironmental time series. GAMs can es-
835 timate non-linear trends, provide estimates of the magnitude of change as well as allow the
836 identification of periods of change, can account for the lack of independence (either via auto-
837 correlation processes or via the fitting of a wiggly trend), and provide a formal framework for
838 statistical inference on each of these features.

839 In presenting the GAM with specific palaeoenvironmental examples and addressing the issues
840 that arise in palaeoenvironmental time series, it is hoped that palaeoecologists and palaeolim-
841 nologists will be motivated to give greater consideration to the estimation of trends and the
842 identification of change in stratigraphic time series.

843 Conflict of interest statement

844 The author declares that the research was conducted in the absence of any commercial or
845 financial relationships that could be construed as a potential conflict of interest.

846 Acknowledgements

847 The ideas expressed in this paper are the result of many fruitful conversations with colleagues
848 at the Environmental Change Research Centre, UCL, and the University of Regina. In particu-
849 lar I am indebted to Helen Bennion, Rick Battarbee, and Peter Leavitt for their collaborations
850 on projects over many years, and to David Miller, Eric Pedersen, and Noam Ross, my GAM
851 workshop partners in crime. Without Simon Wood's *mgcv* software and his research on GAMs,
852 the application of these models to palaeo time series would not be as straight forward. This
853 work was supported by a Natural Sciences and Engineering Council of Canada (NSERC) Dis-
854 covery Grant to the author (RGPIN-2014-04032).

855 References

856 Anchukaitis, K. J., and Tierney, J. E. (2013). Identifying coherent spatiotemporal modes in time-
857 uncertain proxy paleoclimate records. *Climate Dynamics* 41, 1291–1306. doi:[10.1007/s00382-013-2000-0](https://doi.org/10.1007/s00382-013-2000-0)

- 858 012-1483-0.
- 859 Beck, K. K., Fletcher, M.-S., Gadd, P. S., Heijnis, H., Saunders, K. M., Simpson, G. L., et al.
860 (2018). Variance and Rate-of-Change as early warning signals for a critical transition in an
861 aquatic ecosystem state: A test case from tasmania, australia. *Journal of Geophysical Research: Biogeosciences* 123, 2017JG004135. doi:[10.1002/2017JG004135](https://doi.org/10.1002/2017JG004135).
- 863 Bennion, H., Simpson, G. L., and Goldsmith, B. J. (2015). Assessing degradation and recovery
864 pathways in lakes impacted by eutrophication using the sediment record. *Frontiers in Ecology
865 and Evolution* 3. doi:[10.3389/fevo.2015.00094](https://doi.org/10.3389/fevo.2015.00094).
- 866 Bergmeir, C., Hyndman, R. J., and Koo, B. (2018). A note on the validity of cross-validation for
867 evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis* 120,
868 70–83. doi:[10.1016/j.csda.2017.11.003](https://doi.org/10.1016/j.csda.2017.11.003).
- 869 Birks, H. J. B. (1998). Numerical tools in palaeolimnology — progress, potentialities, and prob-
870 lems. *Journal of Paleolimnology* 20, 307–332. doi:[10.1023/A:1008038808690](https://doi.org/10.1023/A:1008038808690).
- 871 Birks, H. J. B. (2012a). “Introduction and overview of part III,” in *Tracking environmental change
872 using lake sediments* (Springer, Dordrecht), 331–353. doi:[10.1007/978-94-007-2745-8_10](https://doi.org/10.1007/978-94-007-2745-8_10).
- 873 Birks, H. J. B. (2012b). “Overview of numerical methods in palaeolimnology,” in *Tracking envi-
874 ronmental change using lake sediments* (Springer, Dordrecht), 19–92. doi:[10.1007/978-94-007-2745-8_2](https://doi.org/10.1007/978-94-007-2745-8_2).
- 876 Blaauw, M., and Heegaard, E. (2012). “Estimation of Age-Depth relationships,” in *Tracking
877 environmental change using lake sediments* (Springer, Dordrecht), 379–413. doi:[10.1007/978-94-007-2745-8_12](https://doi.org/10.1007/978-94-007-2745-8_12).
- 879 Brassell, S. C. (1993). “Applications of biomarkers for delineating marine paleoclimatic fluctu-
880 ations during the pleistocene,” in *Organic geochemistry: Principles and applications*, eds. M. H. En-
881 gel and S. A. Macko (Boston, MA: Springer US), 699–738. doi:[10.1007/978-1-4615-2890-6_34](https://doi.org/10.1007/978-1-4615-2890-6_34).
- 882 Chu, G., Sun, Q., Li, S., Zheng, M., Jia, X., Lu, C., et al. (2005). Long-chain alkenone distribu-
883 tions and temperature dependence in lacustrine surface sediments from china. *Geochimica et
884 Cosmochimica Acta* 69, 4985–5003. doi:[10.1016/j.gca.2005.04.008](https://doi.org/10.1016/j.gca.2005.04.008).
- 885 Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Jour-
886 nal of the American Statistical Association* 74, 829–836. doi:[10.1080/01621459.1979.10481038](https://doi.org/10.1080/01621459.1979.10481038).
- 887 Craven, P., and Wahba, G. (1978). Smoothing noisy data with spline functions. *Numerische
888 Mathematik* 31, 377–403. doi:[10.1007/BF01404567](https://doi.org/10.1007/BF01404567).
- 889 Duchon, J. (1977). “Splines minimizing rotation-invariant semi-norms in sobolev spaces,”
890 in *Constructive theory of functions of several variables* (Springer, Berlin, Heidelberg), 85–100.
891 doi:[10.1007/BFb0086566](https://doi.org/10.1007/BFb0086566).
- 892 Dutilleul, P., Cumming, B. F., and Lontoc-Roy, M. (2012). “Autocorrelogram and periodogram
893 analyses of palaeolimnological Temporal-Series from lakes in central and western north amer-
894 ica to assess shifts in drought conditions,” in *Tracking environmental change using lake sediments*
895 (Springer, Dordrecht), 523–548. doi:[10.1007/978-94-007-2745-8_16](https://doi.org/10.1007/978-94-007-2745-8_16).
- 896 D’Andrea, W. J., Huang, Y., Fritz, S. C., and Anderson, N. J. (2011). Abrupt holocene climate

- 897 change as an important factor for human migration in west greenland. *Proceedings of the National*
898 *Academy of Sciences* 108, 9765–9769. doi:[10.1073/pnas.1101708108](https://doi.org/10.1073/pnas.1101708108).
- 899 Epperson, J. F. (1987). On the Runge example. *The American Mathematical Monthly* 94, 329–341.
900 doi:[10.2307/2323093](https://doi.org/10.2307/2323093).
- 901 Fahrmeir, L., and Kneib, T. (2008). “On the identification of trend and correlation in temporal
902 and spatial regression,” in *Recent advances in linear models and related areas* (Physica-Verlag HD),
903 1–27. doi:[10.1007/978-3-7908-2064-5\1](https://doi.org/10.1007/978-3-7908-2064-5_1).
- 904 Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. (2013). *Regression: Models, methods and applications*.
905 Springer Berlin Heidelberg doi:[10.1007/978-3-642-34333-9](https://doi.org/10.1007/978-3-642-34333-9).
- 906 Gautheir, T. D. (2001). Detecting trends using spearman’s rank correlation coefficient. *Envi-
907 ronmental Forensics* 2, 359–362. doi:[10.1080/713848278](https://doi.org/10.1080/713848278).
- 908 Glew, J. R., Smol, J. P., and Last, W. M. (2001). “Sediment core collection and extrusion,” in
909 *Tracking environmental change using lake sediments: Basin analysis, coring, and chronological techniques*,
910 eds. W. M. Last and J. P. Smol (Dordrecht: Springer Netherlands), 73–105. doi:[10.1007/0-306-47669-X\5](https://doi.org/10.1007/0-306-47669-X\5).
- 912 Golding, N., and Purse, B. V. (2016). Fast and flexible bayesian species distribution modelling
913 using gaussian processes. *Methods in Ecology and Evolution*. doi:[10.1111/2041-210X.12523](https://doi.org/10.1111/2041-210X.12523).
- 914 Handcock, M. S., Meier, K., and Nychka, D. (1994). Kriging and splines: An empirical compar-
915 ision of their predictive performance in some applications: Comment. *Journal of the American
916 Statistical Association* 89, 401–403. doi:[10.2307/2290838](https://doi.org/10.2307/2290838).
- 917 Hastie, T. J., and Tibshirani, R. J. (1990). *Generalized additive models*. Boca Raton, Fl.: Chapman
918 & Hall / CRC.
- 919 Hastie, T., and Tibshirani, R. (1986). Generalized additive models. *Statistical Science* 1, 297–310.
- 920 Juggins, S., and Telford, R. J. (2012). “Exploratory data analysis and data display,” in *Tracking
921 environmental change using lake sediments* (Springer, Dordrecht), 123–141. doi:[10.1007/978-94-007-2745-8\5](https://doi.org/10.1007/978-94-
922 007-2745-8\5).
- 923 Kammann, E. E., and Wand, M. P. (2003). Geoadditive models. *Journal of the Royal Statistical
924 Society. Series C, Applied statistics* 52, 1–18. doi:[10.1111/1467-9876.00385](https://doi.org/10.1111/1467-9876.00385).
- 925 Kimeldorf, G. S., and Wahba, G. (1970). A correspondence between bayesian estimation on
926 stochastic processes and smoothing by splines. *Annals of Mathematical Statistics* 41, 495–502.
- 927 Magee, L. (1998). Nonlocal behavior in polynomial regressions. *The American Statistician* 52,
928 20–22. doi:[10.2307/2685560](https://doi.org/10.2307/2685560).
- 929 Mann, M. E. (2004). On smoothing potentially non-stationary climate time series. *Geophysical
930 Research Letters* 31, L07214. doi:[10.1029/2004GL019569](https://doi.org/10.1029/2004GL019569).
- 931 Mann, M. E. (2008). Smoothing of climate time series revisited. *Geophysical Research Letters* 35,
932 L16708. doi:[10.1029/2008GL034716](https://doi.org/10.1029/2008GL034716).
- 933 Marra, G., and Wood, S. N. (2012). Coverage properties of confidence intervals for generalized
934 additive model components. *Scandinavian Journal of Statistics, Theory and Applications* 39, 53–74.

- 935 doi:[10.1111/j.1467-9469.2011.00760.x](https://doi.org/10.1111/j.1467-9469.2011.00760.x).
- 936 McCullagh, P., and Nelder, J. A. (1989). *Generalized linear models, second edition*. CRC Press.
- 937 Mills, T. C. (2006). Modelling current trends in Northern Hemisphere temperatures. *International Journal of Climatology* 26, 867–884. doi:[10.1002/joc.1286](https://doi.org/10.1002/joc.1286).
- 939 Mills, T. C. (2007). A note on trend decomposition: The “classical” approach revisited
940 with an application to surface temperature trends. *Journal of applied statistics* 34, 963–972.
941 doi:[10.1080/02664760701590418](https://doi.org/10.1080/02664760701590418).
- 942 Mills, T. C. (2010). “Skinning a cat”: Alternative models of representing temperature trends.
943 *Climatic Change* 101, 415–426.
- 944 Nychka, D. (1988). Bayesian confidence intervals for smoothing splines. *Journal of the American
945 Statistical Association* 83, 1134–1143. doi:[10.1080/01621459.1988.10478711](https://doi.org/10.1080/01621459.1988.10478711).
- 946 PAGES 2K Consortium (2013). Continental-scale temperature variability during the past two
947 millennia. *Nature Geoscience* 6, 339–346. doi:[10.1038/ngeo1797](https://doi.org/10.1038/ngeo1797).
- 948 Pinheiro, J. C., and Bates, D. M. (2000). *Mixed-Effects models in S and S-PLUS*. Springer Science
949 & Business Media.
- 950 Pya, N. (2017). *Scam: Shape constrained additive models*. Available at: <https://CRAN.R-project.org/package=scam>.
- 952 Pya, N., and Wood, S. N. (2015). Shape constrained additive models. *Statistics and Computing*
953 25, 543–559. doi:[10.1007/s11222-013-9448-7](https://doi.org/10.1007/s11222-013-9448-7).
- 954 Pya, N., and Wood, S. N. (2016). A note on basis dimension selection in generalized additive
955 modelling. *ArXiv e-prints*.
- 956 R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R
957 Foundation for Statistical Computing Available at: <https://www.R-project.org/>.
- 958 R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R
959 Foundation for Statistical Computing Available at: <https://www.R-project.org/>.
- 960 Rasmussen, C. E., and Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT
961 Press.
- 962 Reiss, P. T., and Ogden, R. T. (2009). Smoothing parameter selection for a class of semipara-
963 metric linear models. *Journal of the Royal Statistical Society. Series B, Statistical methodology* 71,
964 505–523. doi:[10.1111/j.1467-9868.2008.00695.x](https://doi.org/10.1111/j.1467-9868.2008.00695.x).
- 965 Runge, C. (1901). Über empirische funktionen und die interpolation zwischen äquidistanten
966 ordinaten. *Zeitschrift fur Angewandte Mathematik und Physik* 46, 224–243.
- 967 Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge Uni-
968 versity Press.
- 969 Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric
970 regression curve fitting. *Journal of the Royal Statistical Society. Series B, Statistical methodology* 47,

- 971 1–52.
- 972 Smol, J. P. (2008). *Pollution of lakes and rivers: A paleoenvironmental perspective*. Blackwell Pub.
- 973 Smol, J. P., Birks, H. J. B., Lotter, A. F., and Juggins, S. (2012). “The march towards the quantitative analysis of palaeolimnological data,” in *Tracking environmental change using lake sediments* (Springer, Dordrecht), 3–17. doi:[10.1007/978-94-007-2745-8\1](https://doi.org/10.1007/978-94-007-2745-8_1).
- 976 Tian, J., Nelson, D. M., and Hu, F. S. (2011). How well do sediment indicators record past
977 climate? An evaluation using annually laminated sediments. *Journal of Paleolimnology* 45, 73–
978 84. doi:[10.1007/s10933-010-9481-x](https://doi.org/10.1007/s10933-010-9481-x).
- 979 Toney, J. L., Huang, Y., Fritz, S. C., Baker, P. A., Grimm, E., and Nyren, P. (2010). Climatic and environmental controls on the occurrence and distributions of long chain
980 alkenones in lakes of the interior united states. *Geochimica et Cosmochimica Acta* 74, 1563–1578.
981 doi:[10.1016/j.gca.2009.11.021](https://doi.org/10.1016/j.gca.2009.11.021).
- 983 Wahba, G. (1983). Bayesian “confidence intervals” for the Cross-Validated smoothing spline.
984 *Journal of the Royal Statistical Society. Series B, Statistical methodology* 45, 133–150.
- 985 Wahba, G. (1990). *Spline models for observational data*. SIAM.
- 986 Wang, Y., Naumann, U., Wright, S. T., and Warton, D. I. (2012). Mvabund- an R package for
987 model-based analysis of multivariate abundance data. *Methods in Ecology and Evolution* 3, 471–
988 474. doi:[10.1111/j.2041-210X.2012.00190.x](https://doi.org/10.1111/j.2041-210X.2012.00190.x).
- 989 Warton, D. I. (2011). Regularized sandwich estimators for analysis of high-dimensional
990 data using generalized estimating equations. *Biometrics* 67, 116–123. doi:[10.1111/j.1541-0420.2010.01438.x](https://doi.org/10.1111/j.1541-0420.2010.01438.x).
- 992 Warton, D. I., Blanchet, F. G., O’Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., et al.
993 (2015). So many variables: Joint modeling in community ecology. *Trends in Ecology & Evolution*.
994 doi:[10.1016/j.tree.2015.09.007](https://doi.org/10.1016/j.tree.2015.09.007).
- 995 Warton, D. I., Wright, S. T., and Wang, Y. (2012). Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution* 3, 89–101.
996 doi:[10.1111/j.2041-210X.2011.00127.x](https://doi.org/10.1111/j.2041-210X.2011.00127.x).
- 998 Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society. Series B, Statistical methodology* 65, 95–114. doi:[10.1111/1467-9868.00374](https://doi.org/10.1111/1467-9868.00374).
- 1000 Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for
1001 generalized additive models. *Journal of the American Statistical Association* 99, 673–686.
1002 doi:[10.2307/27590439](https://doi.org/10.2307/27590439).
- 1003 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. doi:[10.1111/j.1467-9868.2010.00749.x](https://doi.org/10.1111/j.1467-9868.2010.00749.x).
- 1006 Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R, Second edition*. CRC Press.
- 1007 Wood, S. N., Pya, N., and Säfken, B. (2016). Smoothing parameter and model selection
1008 for general smooth models. *Journal of the American Statistical Association* 111, 1548–1563.

- 1009 doi:[10.1080/01621459.2016.1180986](https://doi.org/10.1080/01621459.2016.1180986).
- 1010 Yee, T. W., and Mitchell, N. D. (1991). Generalized additive models in plant ecology. *Journal of*
 1011 *Vegetation Science* 2, 587–602. doi:[10.2307/3236170](https://doi.org/10.2307/3236170).
- 1012 Zink, K.-G., Leythaeuser, D., Melkonian, M., and Schwark, L. (2001). Temperature de-
 1013 pendency of long-chain alkenone distributions in recent to fossil limnic sediments and
 1014 in lake waters¹¹Associate editor: J. b. fein. *Geochimica et Cosmochimica Acta* 65, 253–265.
 1015 doi:[10.1016/S0016-7037\(00\)00509-3](https://doi.org/10.1016/S0016-7037(00)00509-3).

1016 Appendix 1 — Simultaneous intervals

- 1017 We proceed by considering a simultaneous confidence interval for a function $f(x)$ at a set of M
 1018 locations in x ; we'll refer to these locations, following the notation of Ruppert et al. (2003) by

$$\mathbf{g} = (g_1, g_2, \dots, g_M)$$

- 1019 The true function over \mathbf{g} , \mathbf{f}_g , is defined as the vector of evaluations of f at each of the M locations

$$\mathbf{f}_g \equiv \begin{bmatrix} f(g_1) \\ f(g_2) \\ \vdots \\ f(g_M) \end{bmatrix}$$

- 1020 and the corresponding estimate of the true function given by the fitted GAM denoted by $\hat{\mathbf{f}}_g$.
 1021 The difference between the true function and our unbiased estimator is given by

$$\hat{\mathbf{f}}_g - \mathbf{f}_g = \mathbf{C}_g \begin{bmatrix} \hat{\beta} - \beta \\ \hat{\mathbf{u}} - \mathbf{u} \end{bmatrix},$$

- 1022 where \mathbf{C}_g is a matrix formed by the evaluation of the basis functions at locations \mathbf{g} , and the part
 1023 in square brackets is the bias in the estimated model coefficients, which we assume to be mean 0
 1024 and distributed, approximately, multivariate normal with mean vector $\mathbf{0}$ and covariance matrix
 1025 \mathbf{V}_b

$$\begin{bmatrix} \hat{\beta} - \beta \\ \hat{\mathbf{u}} - \mathbf{u} \end{bmatrix} \stackrel{\text{approx.}}{\sim} N(\mathbf{0}, \mathbf{V}_b),$$

- 1026 where \mathbf{V}_b is the Bayesian covariance matrix of the GAM coefficients.

- 1027 Now, the $(1 - \alpha)100\%$ simultaneous confidence interval is

$$\hat{\mathbf{f}}_g \pm m_{1-\alpha} \begin{bmatrix} \widehat{\text{st.dev}}(\hat{f}(g_1) - f(g_1)) \\ \widehat{\text{st.dev}}(\hat{f}(g_2) - f(g_2)) \\ \vdots \\ \widehat{\text{st.dev}}(\hat{f}(g_M) - f(g_M)) \end{bmatrix},$$

1028 where $m_{1-\alpha}$ is the $1 - \alpha$ quantile of the random variable

$$\sup_{x \in x} \left| \frac{\hat{f}(x) - f(x)}{\widehat{\text{st.dev}}(\hat{f}(x) - f(x))} \right| \approx \max_{1 \leq \ell \leq M} \left| \frac{\left(\mathbf{C}_g \begin{bmatrix} \hat{\beta} - \beta \\ \hat{\mathbf{u}} - \mathbf{u} \end{bmatrix} \right)_\ell}{\widehat{\text{st.dev}}(\hat{f}(g_\ell) - f(g_\ell))} \right|$$

1029 The sup refers to the *supremum* or the *least upper bound*; this is the least value of \mathcal{X} , the set of all
1030 values of which we observed subset x , that is *greater* than all of the values in the subset. Often
1031 this is the maximum value of the subset. This is what is indicated by the right-hand side of
1032 the equation; we want the maximum (absolute) value of the ratio over all values in g .

1033 The fractions in both sides of the equation correspond to the standardized deviation between
1034 the true function and the model estimate, and we consider the *maximum absolute* standardized
1035 deviation. We don't usually know the distribution of the maximum absolute standardized
1036 deviation but we need this to access its quantiles. However, we can closely approximate the
1037 distribution via simulation. The difference here is that rather than simulating from the poste-
1038 rior of the model as we did earlier [see section Confidence intervals](#), this time we simulate from
1039 the multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix \mathbf{V}_b . For each
1040 simulation we find the maximum absolute standardized deviation of the fitted function from
1041 the true function over the grid of x values we are considering. Then we collect all these max-
1042 ima, sort them and either take the $1 - \alpha$ probability quantile of the maxima, or the maximum
1043 with rank $\lceil (1 - \alpha)/N \rceil$.