

# Modelling palaeoecological time series using generalized additive models

Gavin L. Simpson

[gavin.simpson@uregina.ca](mailto:gavin.simpson@uregina.ca)<sup>1</sup>

<sup>1</sup>Institute of Environmental Change and Society, University of Regina,  
Regina, Saskatchewan, Canada, S4S 0A2

May 11, 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 The vast majority of data produced by palaeoecologists and palaeolimnologists is in the form  
37 of time-ordered observations on one or more proxies or biological taxa (Birks, 2012b; Smol,  
38 2008; Smol et al., 2012). Typically these data are arranged irregularly in time; in the absence of  
39 annual laminae or varves, the sediment core is sectioned at regular depth intervals (Glew et al.,  
40 2001), which, owing to variation in accumulation rates over time and compaction by overlying  
41 sediments, results in an uneven sampling in time. An under-appreciated secondary feature  
42 of such sampling is that younger sediments often have larger variance than older sediments;  
43 each section of core represents fewer lake years in newer samples, relative to older samples.  
44 This variable time averaging acts as a variable low-pass (high-cut) filter of the raw, annual  
45 depositional signal.

46 The irregular intervals between samples means that the customary time-series analysis meth-  
47 ods of autoregressive or moving average processes, in the form of autoregressive integrated  
48 moving average (ARIMA) models, are practically impossible to apply because software imple-  
49 menting these methods typically requires even spacing of observations in time. Dutilleul et al.  
50 (2012) and Birks (2012a), eschewing the term *time series*, prefer to call such data *temporal series*  
51 on account of the irregular spacing of samples, a distinction that I find unnecessary.

52 Where statistical approaches have been applied to trend estimation in palaeoenvironmental  
53 time series, the most commonly used has been LOESS (Birks, 1998, 2012a; Cleveland, 1979; Jug-  
54 gins and Telford, 2012). LOESS, or locally weighted scatterplot smoother, as it's name suggests,  
55 was developed to smooth x-y scatterplot data, ostensibly as part of exploratory data analysis  
56 (Cleveland, 1979). The method fits a smooth line through data by fitting weighted least squares  
57 (WLS) models to observations within a particular, user-specified window of the target point,  
58 whose width is typically expressed as a proportion  $\alpha$  of the  $n$  data points. Weights are deter-  
59 mined by how close (in the x-axis only) an observation in the window is to the focal point,  
60 with a tricubic function, by default, giving greatest weight to points closest to the focal point.  
61 The interim LOESS-smoothed value for the focal point is the predicted value from the weighted  
62 regression at the focal point. The interim values are updated using weights based on how far  
63 in the y-axis direction, the interim smoothed value lies from the observed value plus the x-axis  
64 distance weights; this has the effect of down-weighting the effects of outlier observations. The  
65 final LOESS is obtained by joining up the smoothed (fitted) values. The user has to choose how  
66 large a window to use, whether to fit degree 1 (linear) or degree 2 (quadratic) polynomials in  
67 the WLS model, and how to weight points in the x-axis. When used in an exploratory mode,  
68 the user has considerable freedom to choose the detail of the LOESS fit; the window width, for  
69 example, can be infinitely tweaked to give as close a fit to the data, as assessed by eye, as is de-  
70 sired. Using cross-validation (CV) to choose  $\alpha$  or the degree of polynomial in the WLS model  
71 is complicated for a number of reasons, not least that the CV scheme used must involve the  
72 time ordering of the data (e.g. Bergmeir et al., 2018). This level of subjectivity is problematic

73 however once we wish to move beyond exploratory data analysis to statistically identify trends  
74 and to test hypotheses involving those trend estimates.

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

85 These are by no means the only methods that have been used to estimated trends in strati-  
86 graphic series. Another common approach involves fitting a simple linear trend using ordi-  
87 nary least squares regression and use the resulting  $t$  statistic as evidence against the null  
88 hypothesis of no trend despite the statistical assumptions of the method being almost surely  
89 violated due to dependence among observations arising from temporal autocorrelation in the  
90 series. The Pearson correlation coefficient,  $r$ , is also often used to detect trends in palaeo time  
91 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; Gautheir,  
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  $\rho$  and Kendall's  $\tau$  assumes independent observations.

100 Palaeoecology and palaeolimnology have moved away from being descriptive disciplines,  
101 rapidly adopting new statistical developments in the 1990s and beyond (@ Smol et al., 2012).  
102 Less development has been observed in the area of trend estimation in palaeoenvironmental  
103 time series. Here, I describe the use of generalized additive models (GAMs; Hastie and  
104 Tibshirani, 1986, 1990; Ruppert et al., 2003; Wood, 2017; Yee and Mitchell, 1991) for trend  
105 estimation. GAMs, like simple linear regression, are a formal regression-based method for  
106 estimating trends, yet they are also, superficially at least, similar to LOESS. GAMs and LOESS  
107 estimate smooth, non-linear trends in time series and both can handle the irregular spacing of  
108 samples in time, yet GAMs do not suffer from the subjectivity that plagues LOESS as a method  
109 of formal statistical inference.

110 In the subsequent sections, I present an introduction to GAMs and how they are able to es-  
111 timate smooth, non-linear trends from time series. Next the issue of uncertainty in model-  
112 estimated trends is discussed, as is the topic of posterior simulation from a regression model.  
113 The issue of identifying periods of significant environmental change is considered and I show  
114 how this can be achieved using the first derivative of the estimated trend. Two non-standard  
115 types of spline — adaptive smoothers and Gaussian process splines — that are especially appli-

<sup>116</sup> cable to GAMs in the palaeoenvironmental setting are subsequently described, followed by an  
<sup>117</sup> assessment of the the impact of age-model uncertainty on trend estimation via GAMs. Finally,  
<sup>118</sup> I briefly discuss the application of GAM trend analysis to multivariate species abundance and  
<sup>119</sup> compositional data.

## <sup>120</sup> 1.1 Example time series

<sup>121</sup> To illustrate the estimation of trends in palaeoenvironmental data using GAMs, I use two proxy  
<sup>122</sup> time series; a 150-year bulk organic matter  $\delta^{15}\text{N}$  time series from Small Water, and a 3000-  
<sup>123</sup> year alkenone record from Braya-Sø. Between them, the two examples, combine many of the  
<sup>124</sup> features of interest to palaeoecologists that motivate the use of GAMs; non-linear trends and  
<sup>125</sup> the question of when changes in the measured proxy occurred. The example analyses were  
<sup>126</sup> all performed using the *mgcv* package and R, and the supplementary material contains a fully  
<sup>127</sup> annotated document showing the R code used to replicate all the analyses described in the  
<sup>128</sup> remainder of the paper.

### <sup>129</sup> 1.1.1 $\delta^{15}\text{N}$ time series from Small Water

<sup>130</sup> Figure 1a shows 48 nitrogen stable isotope measurements on the bulk organic matter of a sed-  
<sup>131</sup> iment core collected from Small Water, a small corrie lake located in the English Lake District,  
<sup>132</sup> UK. The data were collected to investigate disturbance of nitrogen (N) cycling in remote, olig-  
<sup>133</sup> otrophic lakes by N deposited from the atmosphere (Simpson, unpublished data). The data  
<sup>134</sup> are shown on a  $^{210}\text{Pb}$  time scale. Questions that might be asked about this series are; what is  
<sup>135</sup> 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  
<sup>136</sup>  $\delta^{15}\text{N}$  values in the uppermost section of the core a real change?

### <sup>137</sup> 1.1.2 Braya-Sø alkenone time series

<sup>138</sup> The second example time series is a 3,000 year record of alkenone unsaturation,  $U_{37}^K$ , from  
<sup>139</sup> Braya-Sø, a meromictic lake in West Greenland (D'Andrea et al., 2011). Alkenones are long-  
<sup>140</sup> chained unsaturated organic compounds that are produced by a small number of planktonic  
<sup>141</sup> organisms known as haptophytes. The  $U_{37}^K$  unsaturation index is (Brassell, 1993)

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

<sup>142</sup> where  $[C_{37:x}]$  is the concentration of the alkenone with 37 carbon atoms and  $x$  double carbon  
<sup>143</sup> bonds. The relative abundance of these alkenones is known to vary with changes in water  
<sup>144</sup> temperature (Brassell, 1993; Chu et al., 2005; Toney et al., 2010; Zink et al., 2001), and as a  
<sup>145</sup> result  $U_{37}^K$  is used as a proxy for lake- and sea-surface temperatures. For further details on the  
<sup>146</sup> Braya-Sø  $U_{37}^K$  record and age model see D'Andrea et al. (2011). Here I use the 3,000 year  $U_{37}^K$   
<sup>147</sup> record from the PAGES 2K database (PAGES 2K Consortium, 2013). The data are presented in  
<sup>148</sup> Figure 1b.

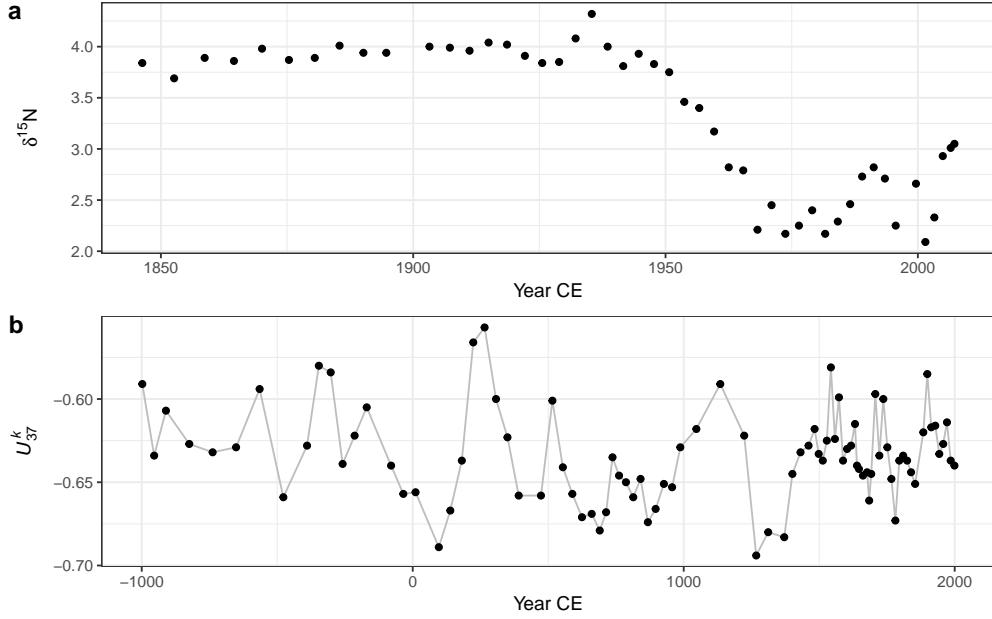


Figure 1: Example time series; a) Small Water bulk organic matter  $\delta^{15}\text{N}$  time series on a  $^{210}\text{Pb}$  time scale, and b) Braya-Sø  $U_{37}^K$  time series on a calibrated  $^{14}\text{C}$  time scale. The observations  $U_{37}^K$  time series have been joined by lines purely as a visual aid to highlight potential trends.

## <sup>149</sup> 2 Regression models for palaeoenvironmental time series

<sup>150</sup> A linear model for a trend in a series of  $T$  observations  $y_t$  at observation times  $x_t$  with  $t =$   
<sup>151</sup>  $1, 2, \dots, T$  is

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

<sup>152</sup> where  $\beta_0$  is a constant term, the model *intercept*, representing the expected value of  $y_t$  where  
<sup>153</sup>  $x_t$  is 0.  $\beta_1$  is the *slope* of the best fit line through the data; it measures the rate of change in  
<sup>154</sup>  $y$  for a unit increase in  $x$ . The unknowns, the  $\beta_j$  are commonly estimated using least squares  
<sup>155</sup> by minimising the sum of squared errors,  $\sum_t \varepsilon_t^2$ . If we want to ask if the estimated trend  $\beta_1$   
<sup>156</sup> is statistically significant, a process called *inference*, we must make further assumptions about  
<sup>157</sup> the data (conditional upon the fitted model) or the model errors (residuals);  $\varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$ .  
<sup>158</sup> This notation indicates that the residuals  $\varepsilon_t$  are *independent* and *identically distributed* Gaussian  
<sup>159</sup> random variables with mean equal to 0 and constant variance  $\sigma^2$ . In the time series setting,  
<sup>160</sup> the assumption of independence of model residuals is often violated.

<sup>161</sup> The linear model described above is quite restrictive in terms of the types of trend it can fit;  
<sup>162</sup> essentially linear increasing or decreasing trends, or, trivially, a null trend of no change. This  
<sup>163</sup> model can be extended to allow for non-linear trends, most notably by making  $y_t$  depend on  
<sup>164</sup> polynomials of  $x_t$ , for example

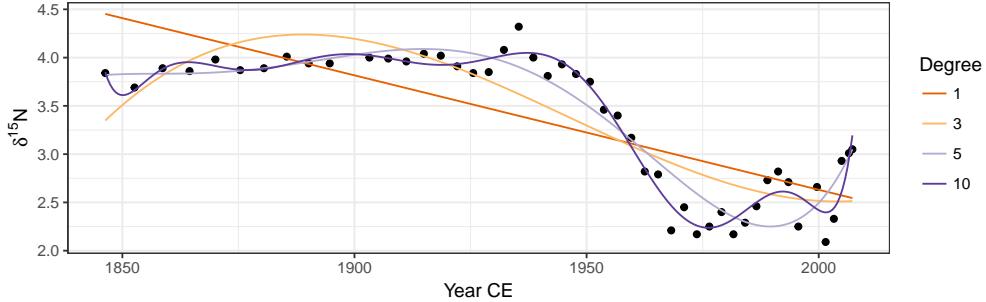


Figure 2: Linear models with various orders of polynomial of the covariate Year fitted using ordinary least squares to the  $\delta^{15}\text{N}$  time series from Small Water. The degree of polynomial is indicated, with the degree 1 line equal to a simple linear regression model.

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

where polynomials of  $x_t$  up to order  $P$  are used. This model allows for more complex trends but it remains a fully parametric model and suffers from several problems, especially the behaviour of the fitted polynomial at the start and end of the observed series.

Linear models using a range of polynomials fitted to the Small Water data set are shown in Figure 2. The low-order models ( $P \in \{1, 3\}$ ) result in very poor fit to the data. The model with  $P = 5$  does a reasonable job of capturing the gross pattern in the time series, but fails to adapt quickly enough to the decrease in  $\delta^{15}\text{N}$  that begins  $\sim 1940$  CE, and the estimated trend is quite biased as a result. The  $P = 10$ -th-order polynomial model is well able to capture this period of rapid change, but it does so at the expense of increased complexity in the estimated trend prior to  $\sim 1940$ . Additionally, this model ( $P = 10$ ) has undesirable behaviour at the ends of the series, significantly overfitting the data, a commonly observed problem in polynomial models such as these (Epperson, 1987; Runge, 1901). Finally, the choice of what order of polynomial to fit is an additional choice left for the analyst to specify; choosing the optimal  $P$  is not a trivial task when the data are a time series and residual autocorrelation is likely present.

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

### 3 Generalized additive models

The GAM version of the linear model (1) is

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

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

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

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

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

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

197 where  $\mu_t$  is the expected value (e.g. the mean count or the probability of occurrence) of the  
 198 random variable  $Y_t$  ( $\mu_t \equiv \mathbb{E}(Y_t)$ ) of which we have observations  $y_t$ .  $g$  is the link function, an  
 199 invertible, monotonic function, such as the natural logarithm, and  $g^{-1}$  is its inverse. The link  
 200 function maps values from the response scale on to the scale of the linear predictor, whilst  
 201 the inverse of the link function provides the reverse mapping. For example, count data are  
 202 strictly non-negative integer values and are commonly modelled as a Poisson GLM/GAM us-  
 203 ing the natural log link function. On the log scale, the response can take any real value between  
 204  $-\infty$  and  $+\infty$ , and it is on this scale that model fitting actually occurs (i.e. using equation (4b)).  
 205 However we need to map these unbounded values back on to the non-negative response scale.  
 206 The inverse of the log link function, the exponential function, achieves this and maps values  
 207 to the interval  $0-\infty$  (equation (4c)).

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

### 217 3.1 Basis functions

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

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

229 where  $\beta_j$  is the weight applied to the  $j$ th basis function. Note that here we concern ourselves  
230 only with univariate smooths involving a single covariate, the time variable  $x_t$ .

231 The polynomial model we encountered earlier is an example of a statistical model that uses a  
232 basis expansion. For the cubic polynomial ( $P = 3$ ) fit shown in Figure 2 there are in fact 4 basis  
233 functions:  $b_1(x_t) = x_t^0 = 1$ ,  $b_2(x_t) = x_t$ ,  $b_3(x_t) = x_t^2$ , and  $b_4(x_t) = x_t^3$ . Note that  $b_1(x_t)$  is constant  
234 and is linked to the model intercept,  $\beta_0$ , in the linear model (2), and further, that the weights  
235 are the estimated coefficients in the model, the  $\beta_j$ .

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

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

245 A cubic spline is a smooth curve comprised of sections of cubic polynomials ( $P = 3$ ), where  
246 the sections are joined together at some specified locations — known as *knots* — in such a way  
247 that at the joins, the two sections of cubic polynomial that meet have the same value as well  
248 as the same first and second derivative. These properties mean that the sections join smoothly  
249 and differentiably at the knots (Wood, 2017, 5.3.1).

250 The CRS can be parameterized in a number of different ways. One requires a knot at each  
251 unique data value in  $x_t$ , which is computationally inefficient. Another way of specifying a  
252 CRS basis is to parameterize in terms of the value of the spline at the knots. Typically in  
253 this parametrization there are many fewer knots than unique data, with the knots distributed

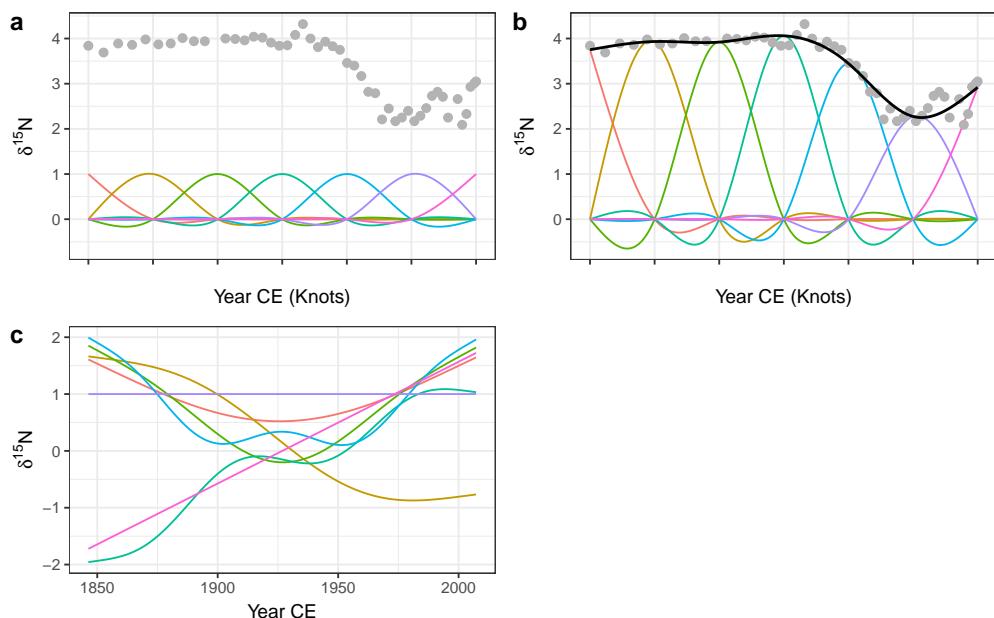


Figure 3: Basis functions for the time covariate and the Small Water  $\delta^{15}\text{N}$  time series. A rank (size) 7 cubic regression spline (CRS) basis expansion is show in a), with knots, indicated by tick marks on the x-axis, spread evenly through the rang of the data. b) shows the same CRS basis functions weighted by the estimated coefficients  $\beta_j$ , 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.

evenly over the range of  $x_t$  or at the quantiles of  $x_t$ . Placing knots at the quantiles of  $x_t$  has the effect of placing a greater number of knots where the data is densest.

A CRS basis expansion comprised of 7 basis functions for the time covariate in the Small Water series, is shown in Figure 3a. The tick marks on the x-axis show the locations of the knots, which are located at the ends of the series and evenly in between. Notice that in this particular parametrization, the  $j$ th basis function takes a value of 1 at the  $j$ th knot and at all other knots a value of 0.

To estimate a model using this basis expansion each basis function forms a column in the model matrix  $X$  and the weights  $\beta_j$  can be found using least squares regression (assuming a Gaussian response). Note that in order to estimate a coefficient for each basis function the model has to be fitted without an intercept term. In practice we would include an intercept term in the model and therefore the basis functions are modified via an identifiability constraint. This has the effect of making the basis orthogonal to the intercept but results in more complicated basis functions than those shown in in Figure 3a.

Having estimated the weight for each basis function, the  $j$ th basis function  $b_j$  is scaled (weighted) by its coefficient  $\beta_j$ . The scaled CRS basis functions for the Small Water time series are shown in Figure 3b. The solid line passing through the data points is formed by summing up the values of the scaled basis functions ( $b_j(x_t)\beta_j$ ) at any value of  $x_t$  (time).

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

To address this issue, thin plate regression splines (TPRS) have been suggested which truncate the space of the thin plate spline basis to some lower number of basis functions whilst preserving much of the advantage of the original basis as an optimally-fitting spline (Wood, 2003). A rank 7 TPRS basis (i.e. one containing 7 basis functions) is shown in Figure 3c for the Small Water time series. The truncation is achieved by performing an eigen-decomposition of the basis functions and retaining the eigenvectors associated with the  $k$  largest eigenvalues. This is similar to the way principal components analysis decomposes a data set into axes of variation (eigenvectors) in decreasing order of variance explained. The truncated basis can preserve much of the space of functions spanned by the original basis but at the cost of using far fewer basis functions (Wood, 2003, 2017, 5.5.1). Note the horizontal TPRS basis function (at

<sup>297</sup>  $\delta^{15}\text{N} = 1$ ) in Figure 3c; this basis function is confounded with the intercept term and, after the  
<sup>298</sup> application of identifiability constraints, ends up being removed from the set of basis functions  
<sup>299</sup> used to fit the model.

<sup>300</sup> The truncation suggested by Wood (2003) is not without cost; the eigen-decomposition and  
<sup>301</sup> related steps can be relatively costly for large data sets. For data sets of similar size to the two  
<sup>302</sup> examples used here, the additional computational effort required to set up the TPRS basis over  
<sup>303</sup> the CRS basis will not be noticeable. For highly resolved series containing more than  $\sim 1000$   
<sup>304</sup> observations the truncation may be too costly computationally. In such instances, little is lost  
<sup>305</sup> by moving to the CRS basis, with the same number of knots as the rank of the desired TPRS,  
<sup>306</sup> with the benefit of considerably reduced set up time for the basis.

<sup>307</sup> To fit a GAM using either of the two regression spline bases described above, the analyst is  
<sup>308</sup> generally only required to specify the size (rank) of the basis expansion required to rep-  
<sup>309</sup> resent or closely approximate the true function  $f$ . With practice and some knowledge of the  
<sup>310</sup> system from which the observations arise, it can be relatively easy to put an upper limit on  
<sup>311</sup> the expected complexity of the true trend in the data. Additionally, the number available data  
<sup>312</sup> points places a constraint on the upper limit of the size of basis expansion that can be used.

<sup>313</sup> In practice, the size of the basis is an upper limit on the expected complexity of the trend,  
<sup>314</sup> and a simple test is available to check if the basis used was sufficiently large (Pya and Wood,  
<sup>315</sup> 2016). This test is available via the `gam.check()` function in *mgcv* for example, which essen-  
<sup>316</sup> tially looks at whether there is any additional nonlinearity or structure in the residuals that  
<sup>317</sup> can be explained by a further smooth of  $x_t$ . Should a smooth term in the fitted model fail this  
<sup>318</sup> test the model can be refitted using a larger basis expansion, say by doubling the value of  $k$  (the  
<sup>319</sup> rank) used to fit the original. Note also that a smooth might fail this test whilst using fewer  
<sup>320</sup> effective degrees of freedom than the maximum possible for the dimension of basis used. This  
<sup>321</sup> may happen when the true function lies at the upper limit of the set of functions encompassed  
<sup>322</sup> by the size of basis used. Additionally, a basis of size  $2k$  encompasses a richer space of func-  
<sup>323</sup> tions of a given complexity than a basis of size  $k$  (Wood, 2017); increasing the basis dimension  
<sup>324</sup> used to fit the model may unlock this additional function space resulting in a better fitting  
<sup>325</sup> model whilst using a similar number of effective degrees of freedom.

## <sup>326</sup> 3.2 Smoothness selection

<sup>327</sup> Having identified regression splines as a useful way to represent  $f$ , we next need a way to  
<sup>328</sup> decide how wiggly the fitted trend should be. A backwards elimination approach to sequen-  
<sup>329</sup> tially remove knots or basis functions might seem appropriate, however such an approach  
<sup>330</sup> would likely fail as the resulting sequence of models would not be strictly nested, precluding  
<sup>331</sup> many forms of statistical comparison (Wood, 2017). Alternatively, we could keep the basis  
<sup>332</sup> dimension at a fixed size but guard against fitting very complex models through the use of a  
<sup>333</sup> wigginess penalty.

<sup>334</sup> The default wigginess penalty used in GAMs is on the second derivative of the spline, which  
<sup>335</sup> measures the rate of change of the slope, or the curvature, of the spline at any infinitesimal  
<sup>336</sup> point in the interval spanned by  $x_t$ . The actual penalty used is the integrated squared second  
<sup>337</sup> derivative of the spline

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

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

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

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

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

354 Figure 4 illustrates how the smoothness parameter  $\lambda$  controls the degree of wigginess in the  
 355 fitted spline. Four models are shown, each fitted with a fixed value of  $\lambda$ ; 10000, 1, 0.01, and  
 356 0.00001. At  $\lambda = 10000$  the model effectively fits a linear model through the data. As the value  
 357 of  $\lambda$  decreases, the fitted spline becomes increasingly wiggly. As  $\lambda$  becomes very small, the  
 358 resulting spline passes through most of the  $\delta^{15}\text{N}$  observations resulting in a model that is  
 359 clearly over fitted to the data.

360 To fully automate smoothness selection for  $f(x_t)$  we need to estimate  $\lambda$ . There are two main  
 361 ways that  $\lambda$  can be automatically chosen during model fitting. The first way is to choose  $\lambda$   
 362 such that it minimises the prediction error of the model. This can be achieved by choosing  $\lambda$   
 363 to minimise Akaike's information criterion (AIC) or via cross-validation (CV) or generalized  
 364 cross-validation (GCV; Craven and Wahba, 1978). GCV avoids the computational overhead  
 365 inherent to CV of having to repeatedly refit the model with one or more observations left out  
 366 as a test set. Minimising the GCV score will, with a sufficiently large data set, find a model  
 367 with the minimal prediction error (Wood, 2017). The second approach is to treat the smooth  
 368 as a random effect, in which  $\lambda$  is now a variance parameter to be estimated using maximum  
 369 likelihood (ML) or restricted maximum likelihood (REML; Wood, 2011; Wood et al., 2016).

370 Several recent results have shown that GCV, under certain circumstances, has a tendency to  
 371 under smooth, resulting in fitted splines that are overly wiggly (Reiss and Ogden, 2009). Much  
 372 better behaviour has been observed for REML and ML smoothness selection, in that order  
 373 (Wood, 2011). REML is therefore the recommended means of fitting GAMs, though, where  
 374 models have different fixed effects (covariates) they cannot be compared using REML, and

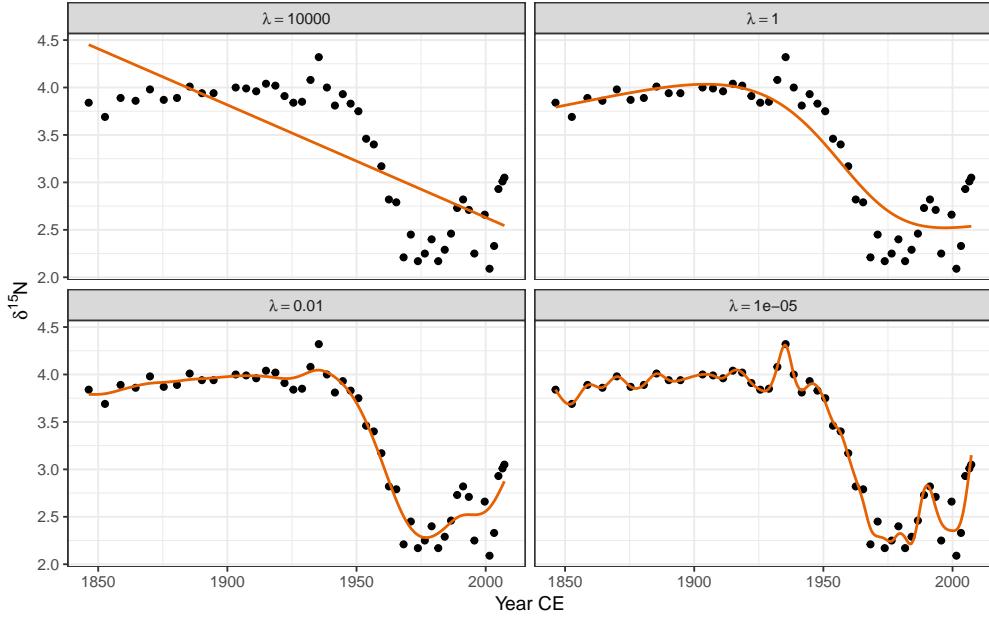


Figure 4: The effect of the smoothness parameter  $\lambda$  on the resulting wigginess of the estimated spline. Large values of  $\lambda$  penalize wigginess strongly, resulting in smooth trends (upper row), while smaller values allow increasingly wiggly trends. The aim of automatic smoothness selection is to find an optimal value of  $\lambda$  that balances the fit of the model with model complexity to avoid overfitting.

375 ML selection should be used instead. In the sorts of data examples considered here there is  
 376 only a single covariate  $x_t$  as our models contain a single estimated trend so REML smoothness  
 377 selection is used throughout unless otherwise stated.

## 378 4 Fitting GAMs

### 379 4.1 Small Water

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

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

386 Note that now I have relaxed the i.i.d. assumption and have introduced  $\Lambda$ , a correlation matrix  
 387 that is used to model autocorrelation in the residuals. The  $\delta^{15}\text{N}$  values are irregularly spaced  
 388 in time and a correlation structure that can handle the uneven spacing of the samples is needed

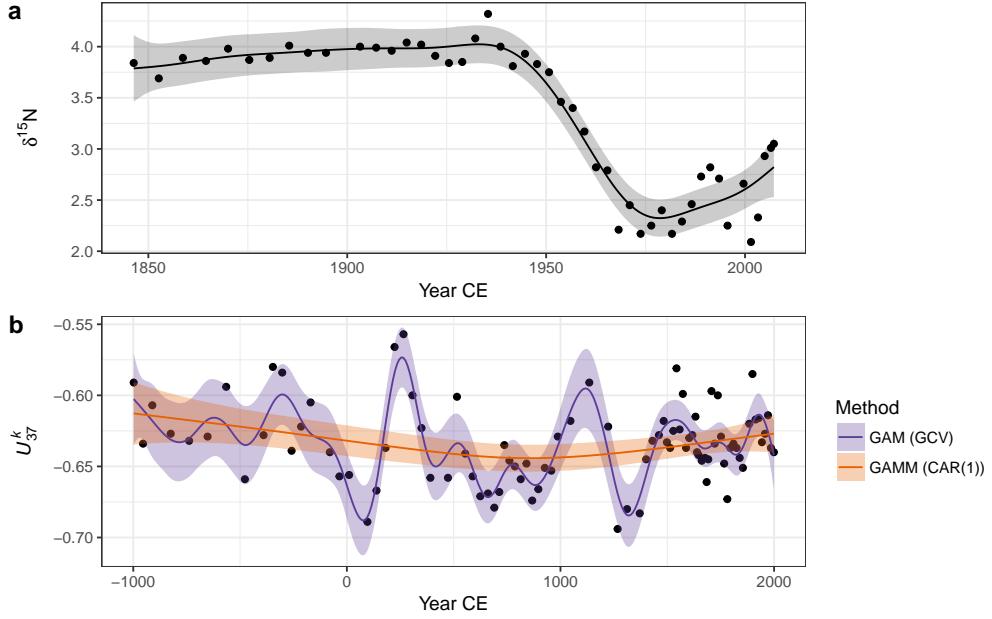


Figure 5: GAM-based trends fitted to the Small Water  $\delta^{15}\text{N}$  (a) and Braya-Sø  $U_{37}^K$  (b) time series. The shaded bands surrounding the estimated trends are approximate 95% across-the-function confidence intervals. For the  $U_{37}^K$  series, two models are shown; the orange fit 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.

(Pinheiro and Bates, 2000). A continuous time first-order autoregressive process (CAR(1)) is a reasonable choice; it is the continuous-time equivalent of the first-order autoregressive process (AR(1)) and, simply stated, models the correlation between any two residuals as an exponentially decreasing function of  $h(\phi^h)$ , where  $h$  is the amount of separation in time between the residuals (Pinheiro and Bates, 2000).  $h$  may be a real valued number in the CAR(1), which is how it can accommodate the irregular separation of samples in time.  $\phi$  controls how quickly the correlation between any two residuals declines as a function of their separation in time and is an additional parameter that will be estimated during model fitting. The model in (6) was fitted using the `gamm()` function (Wood, 2004) in the `mgcv` package (Wood, 2017) for R (R Core Team, 2017).

The fitted trend is shown in Figure 5a, and well-captures the strong pattern in the data. The trend is statistically significant (estimated degrees of freedom = 7.95;  $F = 47.44$ , approximate  $p$  value =  $\ll 0.0001$ ). However further analysis of the fitted model is required to answer the other questions posed earlier about the timing of change and whether features in the trend can be distinguished from random noise. I discuss these issues shortly.

404 4.2 Braya-Sø

405 The  $U_{37}^K$  data present a more difficult data analysis challenge than the  $\delta^{15}\text{N}$  time series because  
406 of the much more complex variation present. Fitting the same model as the Small Water ex-  
407 ample, (6), to the  $U_{37}^K$  data resulted in the unsatisfactory fit shown as the very smooth line in  
408 Figure 5b (labelled GAMM (CAR(1))). Further problems were evident with this model fit —  
409 the covariance matrix of the model was non-positive definite, a sure sign of problems with the  
410 fitted model. Refitting with a smaller basis dimension ( $k = 20$ ) for the trend term resulted in  
411 a model with a positive-definite covariance matrix for the model variance-covariance terms,  
412 but the estimated value of the CAR(1) parameter  $\phi = 0.2$  was exceedingly uncertain (95%  
413 confidence interval 0 – 1!).

414 Fitting this model as a standard GAM with REML smoothness selection resulted in the same  
415 fitted trend as the GAM with CAR(1) errors (not shown), whilst using GCV smoothness selec-  
416 tion resulted in a much more satisfactory fitted trend. There are two potential problems with  
417 the GCV-selected trend: i) GCV is sensitive to the profile of the GCV score and has been shown  
418 to badly under smooth data in situations where the profile is flat around the minimum GCV  
419 score, and ii) the model fitted assumes that the observations are independent, an assumption  
420 that is certainly violated in the  $U_{37}^K$  time series.

421 To investigate the first issue, the GCV and REML scores for an increasing sequence of values  
422 of the smoothness parameter ( $\lambda$ ) were evaluated for the standard GAM (equation (4)) fit to the  
423  $U_{37}^K$  time series. The resulting profiles are shown in Figure 6, with the optimal value of the  
424 parameter shown by the vertical line. The GCV score profile suggests that the potential for  
425 under smoothing identified by Reiss and Ogden (2009) is unlikely to apply here as there is a  
426 well-defined minimum in profile. Addressing the violation of the independence assumption  
427 is more difficult; the standard errors of the model coefficients are most likely anti-conservative,  
428 which presents a significant problem if we wish to determine which of the wiggles in the fitted  
429  $U_{37}^K$  trend can be identified as real.

430 To understand the reason why the GAM plus CAR(1) and the simple GAM with REML smooth-  
431 ness selection performed poorly with the  $U_{37}^K$  time series we need to delve a little deeper into  
432 what is happening when we are fitting these two models.

433 The primary issue leading to poor fit is that neither model accounts for the different variance  
434 (known as (heteroscedasticity) of each observation in the  $U_{37}^K$  record. The sediments in Braya-  
435 Sø are not annually laminated and therefore the core was sliced at regular depth intervals.  
436 Owing to compaction of older sediments and variation in accumulation rates over time, each  
437 sediment slice represents a different number of “lake years”. We can think of older samples as  
438 representing some average of many years of sediment deposition, whilst younger samples are  
439 representative of fewer of these lake years. The average of a larger set of numbers is estimated  
440 more precisely than the average of a smaller set, all things equal. A direct result of this variable  
441 averaging of lake years is that some samples are more precise and therefore have lower variance  
442 than other samples and yet the model assumed that the variance was constant across samples.

443 Accounting for heteroscedasticity within the model is relatively simple via the use of observa-  
444 tional weights. The number of lake years represented by each slice is estimated by assigning a  
445 date to the top and bottom of each sediment slice. The variance of each observation should be

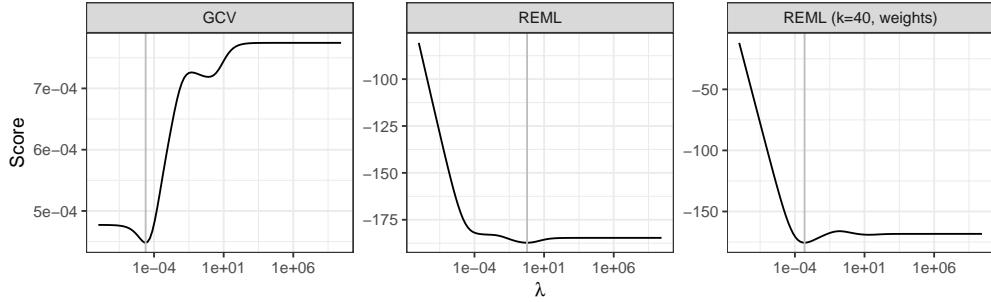


Figure 6: GCV and REML scores as a function of the smoothness parameter  $\lambda$ . From left to right, GAMs were estimated using GCV and REML smoothness selection, and REML using a basis dimension of 40 and observational weights to account for heterogeneity in the  $U_{37}^K$  times series. The selected value of  $\lambda$  for each model is indicated by the vertical grey line.

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 wiggliness because the larger basis contains a richer family of functions and the excess complexity is reduced because of the smoothness penalty in play.)

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 5b). The trace of the REML score for this model (Figure 6c) shows a pronounced minimum at a much smaller value of  $\lambda$  than the original REML fit (Figure 6b), 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).

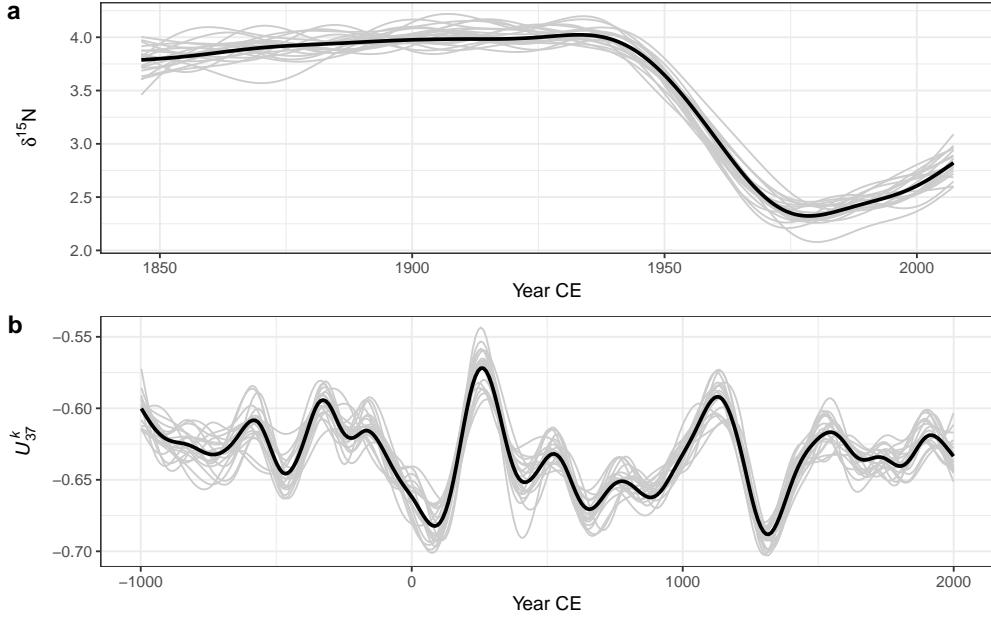


Figure 7: Estimated trends (thick black lines) and 20 random draws (grey lines) from the posterior distribution of the GAM fitted to the Small Water  $\delta^{15}\text{N}$  (a) and Braya-Sø  $U_{37}^K$  (b) time series.

### 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  $\alpha$  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 are 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 approach to fitting GAMs described here has a Bayesian interpretation (Kimeldorf and Wahba, 1970; Silverman, 1985; Wahba, 1983, 1990) and therefore the typical frequentist interpretation does not apply. Nychka (1988) investigated the properties of a confidence interval created as described above using standard errors derived from the Bayesian posterior covariance matrix for the estimated mode parameters. Such intervals have the interesting property that they have good *across-the-function* coverage when considered from a frequentist perspective. This means that, when averaged over the range of the function, the Bayesian credible intervals shown in Figure 5 have close to the expected 95% coverage. However, to achieve this some parts of the function may have more or less than 95%-coverage. Marra and Wood (2012) recently explained Nychka's (1988) surprising results and extended them to the case of generalized models (non-Gaussian responses).

- 496 Whilst the *across-the-function* or average coverage frequentist interpretation of the Bayesian  
 497 credible intervals is useful, it may be important to have an interval that contains the entirety  
 498 of the true function with some state probability ( $1 - \alpha$ ). Such an interval is known as a *simulta-*  
 499 *taneous* interval. A  $(1 - \alpha)100\%$  *simultaneous* confidence interval contains *in their entirety*  $1 - \alpha$  of  
 500 all random draws from the posterior distribution of the fitted model.
- 501 Fitting a GAM involves finding estimates for coefficients of the basis functions. Together, these  
 502 coefficients are distributed multivariate normal with mean vector and covariance matrix spec-  
 503 ified by the model estimates of the coefficients and their covariances respectively. Random  
 504 draws from this distribution can be taken, where each random draw represents a new trend  
 505 that is consistent with the fitted trend but also includes the uncertainty in the estimated trend.  
 506 This process is known as *posterior simulation*.
- 507 Figure 7 shows 20 random draws from the posterior distribution of the GAMs fitted to the  
 508 Small Water and Braya-Sø data sets. In the early period of the  $\delta^{15}\text{N}$  time series many of the  
 509 posterior simulations exhibit short periods of increasing and decreasing trend, balancing out  
 510 to the relatively flat trend estimated by the GAM (Fig. 7a). Reflecting this uncertainty, we might  
 511 expect relatively wide simultaneous intervals during this period in order to contain the vast  
 512 majority of the simulated trends. Conversely, the decreasing  $\delta^{15}\text{N}$  trend starting at ~1945 is  
 513 consistently reproduced in the posterior simulations, suggesting that this feature of the time  
 514 series is both real and statistically significant, and that the rate of change in  $\delta^{15}\text{N}$  is relatively  
 515 precisely estimated. We see a similar pattern in Figure 7b for the Braya-Sø record; the large  
 516 peak in  $U_{37}^K$  at ~250CE and the strong decline at ~1200CE are well defined in the posterior  
 517 simulations, whereas most of the localised trends that are smaller magnitude changes in  $y_t$   
 518 are associated with posterior simulations that are less well constrained with the ends of the  
 519 record in particular showing considerable variation in the strength, timing, and even sign of  
 520 simulated trends, reflecting the greater uncertainty in estimated trend during these periods.  
 521 For the random draws illustrated in Figure 7, a simultaneous interval should contain the entire  
 522 function for on average 19 of the 20 draws.
- 523 There are a number of ways in which a simultaneous interval can be computed. Here I follow  
 524 the simulation approach described by Ruppert et al. (2003) and present only the basic detail; a  
 525 fuller description is contained in Appendix 1. The general idea is that if we want to create an  
 526 interval that contains the whole of the true function with  $1 - \alpha$  probability, we need to increase  
 527 the standard Bayesian credible interval by some amount. We could simulate a large number  
 528 of functions from the posterior distribution of the model and then search for the value of  $m_{1-\alpha}$   
 529 that when multiplied by  $\text{SE}(\hat{f}(x_t))$  yielded an interval that contained the whole function for  
 530  $(1 - \alpha) 100\%$  of the functions simulated. In practice, the simulation method of Ruppert et al.  
 531 (2003) does not involve a direct search, but yields the critical value  $m_{1-\alpha}$  required.
- 532 Simultaneous intervals computed using the method described are show in Figure 8 alongside  
 533 the *across-the-function* confidence intervals for the trends fitted to both example data sets. As  
 534 expected, the simultaneous interval is somewhat wider than the *across-the-function* interval.  
 535 The critical value  $m_{1-\alpha}$  for the simultaneous interval of the estimated trend in  $\delta^{15}\text{N}$  is 3.08,  
 536 whilst the same value for the  $U_{37}^K$  series is 3.42, leading to intervals that are approximately  
 537  $\pm 50\%$  and  $\pm 75\%$  wider than the equivalent across-the-function intervals.

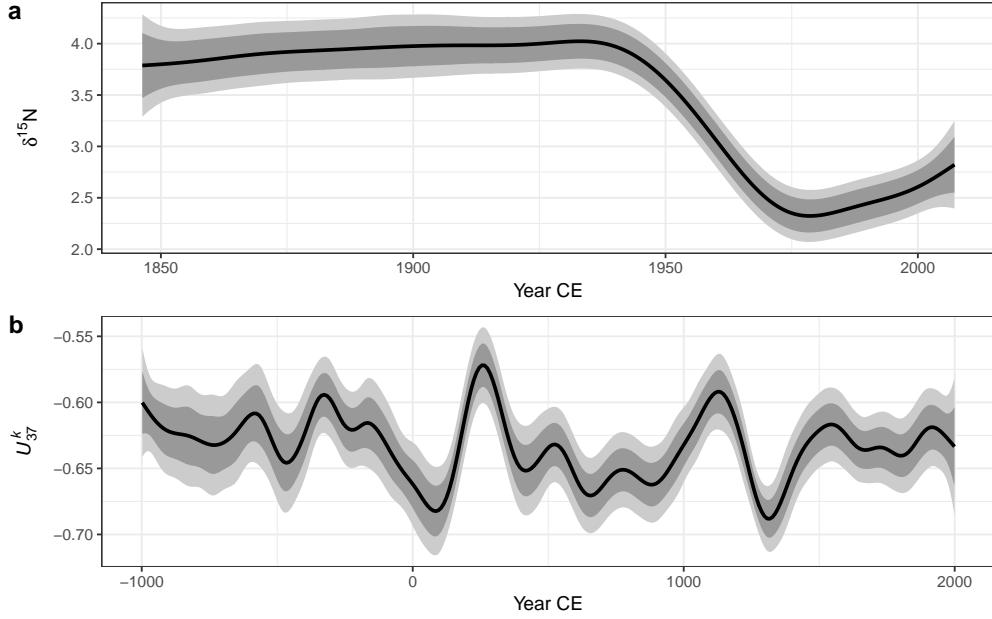


Figure 8: 95% simultaneous confidence intervals (light grey bands) and across-the-function confidence intervals (dark grey bands) on the estimated trends (black lines) for the Small Water  $\delta^{15}\text{N}$  (a) and Braya-Sø  $U_{37}^K$  (b) time series.

#### 538 4.4 Identifying periods change

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

547 In the GAM, the fitted trend need not be linear; the slope of the trend is potentially different  
 548 at every point in the time series. As such we might reasonably ask *where* in the series the  
 549 response  $y_t$  is changing, if at all? Mirroring the linear model we can answer this question by  
 550 determining whether or not the first derivative at any time point  $x_t$  of the fitted trend at any  
 551 time point is consistent with a null hypothesis of no change. We want to know whether or not  
 552 the first derivative is indistinguishable from a value of 0 — no trend — given the uncertainty  
 553 in the estimate of the derivative.

554 Derivatives of the fitted spline are not easily available analytically, but they can be estimated  
 555 using the method of finite differences. Two values of the estimated trend, separated by a very  
 556 small time-shift ( $\Delta_t$ ), are predicted from the model; the difference between the estimated values  
 557 for the two time points is an approximation of the true first derivative of the trend. As  $\Delta_t \rightarrow 0$   
 558 the approximation becomes increasingly accurate. In practice, the first derivative of the fitted

559 trend is evaluated using finite differences at a large number of points in the time series. An  
560 approximate  $(1 - \alpha)100\%$  pointwise confidence interval can be calculated for the derivative  
561 estimates using standard theory (i.e.  $\pm 1.96 \times \text{SE}(\hat{y}_t)$  for a 85% interval) and the covariance matrix  
562 of the spline coefficients. A  $(1 - \alpha)100\%$  simultaneous interval for the derivatives can also be  
563 computed using the method described earlier. Periods of significant change are identified as  
564 those time points where the (simultaneous) confidence interval on the first derivative does not  
565 include zero.

566 Figure 9 shows the estimated first derivative of the fitted trend in the Small Water (9a) and  
567 Braya-Sø (9b) time series. Although the estimated trend suggests a slight increase in  $\delta^{15}\text{N}$   
568 from the start of the record to  $\sim 1940$ , the estimated trend is sufficiently uncertain that the  
569 simultaneous interval on the first derivative includes 0 throughout this period. We can under-  
570 stand why this is so by looking at the posterior simulations in Figure 7a; there is considerable  
571 variation in the shape of the simulated trends throughout this period. From  $\sim 1925$  the deriva-  
572 tive of the trend becomes negative, however it is not until  $\sim 1940$  that the simultaneous interval  
573 doesn't include 0. At this point we have evidence to reject the null hypothesis of no change.  
574 This time point may be taken as the first evidence for change in  $\delta^{15}\text{N}$  in the Small Water core.  
575 The simultaneous interval on the first derivative of the trend in  $\delta^{15}\text{N}$  is bounded away from  
576 0 between  $\sim 1940$  and  $\sim 1975$ , covering the major decline in values evident in the observations.  
577 The simultaneous interval includes 0 from  $\sim 1975$  onward, suggesting that, whilst quite pro-  
578 nounced, the recent increase in  $\delta^{15}\text{N}$  is not statistically significant. To determine whether or  
579 not the recent increase is real, we would require considerably more samples with which to  
580 (hopefully) more-precisely estimate the trend during this period. Alternatively, we might just  
581 have to wait until sufficient additional sedimentation has occurred to warrant recoresing Small  
582 Water and reestimating the trend in  $\delta^{15}\text{N}$ .

583 The estimated trend at Braya-Sø exhibited a number of oscillations in  $U_{37}^K$ . As we saw previ-  
584 ously in Figures 7b and 8b, many of these are subject to significant uncertainty and it is impor-  
585 tant therefore to discern which, if any, of the oscillations in the response can be identified given  
586 the model uncertainty. In Figure 9b only two features of the estimated trend are considered  
587 significant based on the derivatives of the smooth; one centred on  $\sim 250\text{CE}$  and a second at  
588  $\sim 1150\text{CE}$ . In both these periods, the simultaneous interval for the first derivative of the trend  
589 does not include zero. In the first case we detect the large peak and subsequent decline in  
590  $U_{37}^K$  at  $\sim 250\text{CE}$ , whilst at  $\sim 1150\text{CE}$  the large trough is identified, but not the increasing trend  
591 immediately prior to this excursion to lower  $U_{37}^K$ . Recall that these intervals are simultaneous  
592 in nature, strongly guarding against false positives, and as such we can be confident in the  
593 estimation of these two features, whilst care must be taken to not over-interpret the remaining  
594 variations in the estimated trend.

## 595 4.5 Residual autocorrelation and model identification

596 The GAM fitted to the  $\delta^{15}\text{N}$  time series contained a CAR(1) process to model residual temporal  
597 autocorrelation in the residuals. The estimated magnitude of the autocorrelation is given by  
598 the parameter  $\phi$ . The estimated value of  $\phi$  for the  $\delta^{15}\text{N}$  series is 0.6 with 95% confidence in-  
599 terval 0.28–0.85, indicating moderate to strong residual autocorrelation about the fitted trend.

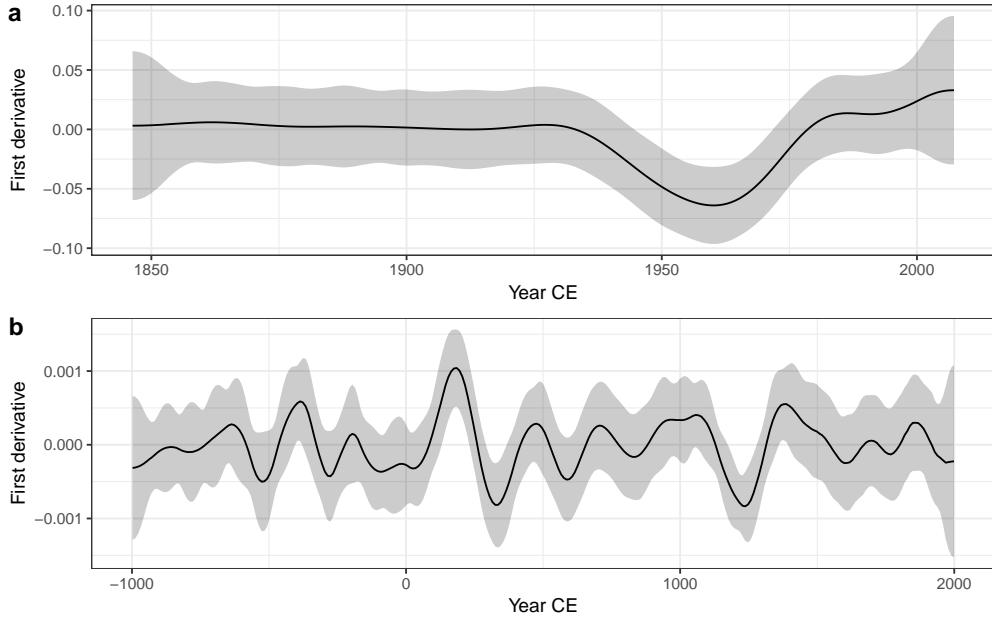


Figure 9: Estimated first derivatives (black lines) and 95% simultaneous confidence intervals of the GAM trends fitted to the Small Water  $\delta^{15}\text{N}$  (a) and Braya-Sø  $U_{37}^K$  (b) time series. Where the simultaneous interval does not include 0, the models detect significant temporal change in the response.

- 600 The correlation function is an exponentially decreasing function of temporal separation ( $\Delta_t$ ),  
 601 and whilst observations that are a few year apart are quite strongly dependent on one another,  
 602 this dependence drops off rapidly as  $\Delta_t$  increases and is effectively zero when samples are  
 603 separated by a decade or more (Figure 10).
- 604 Failure to account for the dependencies in the  $\delta^{15}\text{N}$  time series could lead to the estimation  
 605 of a more wiggly trend than the one shown in Figure 5a which would negatively impact the  
 606 confidence place on the inferences we might draw from the fitted model. Importantly, failing  
 607 to account for the strong dependency in the residuals would lead to smaller uncertainties in  
 608 the estimated spline coefficients, which would propagate through to narrower confidence in-  
 609 tervals on the fitted trend and on the first derivatives, and ultimately to the identification of  
 610 significant periods of change. The end result would be a tendency toward anti-conservative  
 611 identification of periods of change; the coverage probability would be lower than the antici-  
 612 pated  $1 - \alpha$ , leading to a greater chance of false positive results.
- 613 Problems estimating the GAM plus CAR(1) model were encountered when this was fitted to  
 614 the  $U_{37}^K$  time series; both a smooth trend in the mean  $U_{37}^K$  and a CAR(1) process in the resi-  
 615 duals lead to an unidentifiable model. What makes a model with a spline-based trend and an  
 616 autocorrelation process like the CAR(1) potentially unidentifiable?
- 617 Consider again the basic GAM for a smooth trend, (3). In that equation the correlation ma-  
 618 trix  $\Lambda$  was omitted for the sake of simplicity. As I did in (6), I reintroduce it and restate the  
 619 distributional assumptions of this model

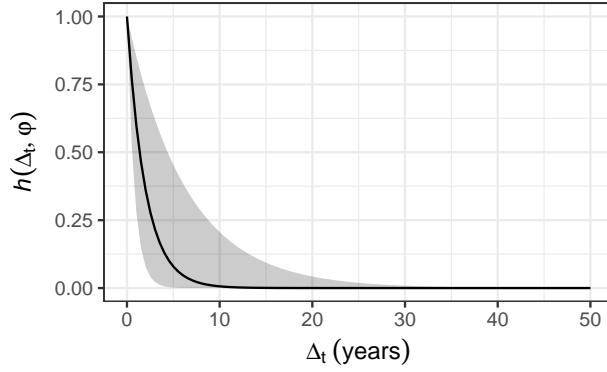


Figure 10: Estimated CAR(1) process from the GAM fitted to the Small Water  $\delta^{15}\text{N}$  time series.  $h(\Delta_t, \phi)$  is the correlation between residuals separated by  $\Delta_t$  years, where  $\hat{\phi} = 0.6$ . The shaded band is a 95% pointwise confidence interval on the estimated correlation  $h$ .

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

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

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

621 which is where the independence assumption of the model comes from; a model residual is  
622 perfectly correlated with itself (the 1s on the diagonal), but uncorrelated with any other resi-  
623 dential (the off-diagonal 0s). In the GAM plus CAR(1) model, an alternative correlation function  
624 for  $\Lambda$  was used — the CAR(1) with correlation parameter  $\phi$ . Fahrmeir and Kneib (2008) show  
625 that where the stochastic structure of  $f$  and  $\Lambda$  approach one another, i.e. where we have a  
626 potentially wiggly trend or strong autocorrelation as  $\phi \rightarrow 1$ , the two processes can quickly  
627 become unidentifiable (see also Fahrmeir et al., 2013). By unidentifiable, we mean that it be-  
628 comes increasingly difficult to distinguish between a wiggly trend or strong autocorrelation  
629 because these two processes are very similar to one another in appearance. This leads to model  
630 estimation problems of the sort encountered with fitting the GAM plus CAR(1) model to the  
631 Braya-sø  $U_{37}^K$  series.

632 Why might this be so? Autocorrelation is the tendency for a large (small) value of  $y_t$  at time  
633  $x_t$  to be followed by a likewise large (small) value at time  $x_{t+1}$ . This leads to runs of values  
634 that are consistently greater (less) than the overall mean. Short runs would indicate weaker  
635 autocorrelation whilst longer runs are associated with stronger autocorrelation, and long runs  
636 of values greater (less) than the mean would evident as non-linear trends in the time series. As  
637 a result, a wiggly trend and an autocorrelation function with large  $\phi$  are two ways to describe  
638 the same pattern of values in a time series, and without any further information to constrain  
639 either the model is unable to distinguish both components uniquely.

640 Situations where it may be possible to uniquely identify separate wiggly trends and autocor-  
 641 relation are exemplified by the Small Water  $\delta^{15}\text{N}$  time series. The non-linear trend and the  
 642 autocorrelation operate at very different scales; the trend represents decadal-scale variation in  
 643 mean  $\delta^{15}\text{N}$ , whilst the CAR(1) process represents the much smaller-scale tendency for values  
 644 of the response to be followed in time by similar values. That such a pattern is observed in  
 645 the Small Water core is the result of the high resolution of the sampling in time relative to the  
 646 long-term trend. In contrast, the Braya-Sø record is sampled at far lower resolution relative  
 647 to the fluctuations in the mean response, and consequently the data do not contain sufficient  
 648 information to separate trend and autocorrelation.

## 649 4.6 Gaussian process smooths

650 In the world of machine learning, Gaussian processes (Golding and Purse, 2016; Rasmussen  
 651 and Williams, 2006) are a widely-used method for fitting smooth non-parametric regression  
 652 models. A Gaussian process is a distribution over all possible smooth functions  $f(x)$ . In the  
 653 field of spatial statistics, Gaussian processes are known by name *kriging*.  
 654 With a Gaussian process we are interested in fitting a smooth temporal trend by modelling the  
 655 way the correlation between pairs of observations varies as a function of the distance,  $h$ , in time  
 656 that separates ( $h$ ) the observations. The correlation between pairs of observations decreases  
 657 with increasing separation, which is modelled using a correlation function,  $c(h)$ . Figure 11  
 658 shows examples of two different correlation functions; the *power exponential* (Figure 11a), and  
 659 the Matérn (Figure 11b) correlation functions. These functions are smooth and monotonic-  
 660 decreasing, meaning that the value of the correlation function decreases with increasing sepa-  
 661 ration ( $h$ ). When  $h = 0$ , the correlation is equal to 1 ( $c(0) = 1$ ); two samples taken at exactly the  
 662 same time point are perfectly correlated. As  $h \rightarrow \infty$ , the correlation tends to zero ( $c(h) \rightarrow 0$ );  
 663 two samples separated by a large amount of time tend to be uncorrelated. Often we are in-  
 664 terested in learning how large the separation in time needs to be before, on average, a pair of  
 665 observations is effectively uncorrelated (i.e. where  $c(h)$  is sufficiently close to zero).  
 666 Several functions can be used to represent  $c(h)$ . Two common ones are the power exponential  
 667 function and the Matérn family of correlation functions. The power exponential function at  
 668 separation distance  $h$  is

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

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

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

672 whilst for  $\kappa = 2.5$  it is

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

673 and for  $\kappa = 3.5$

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

674 In all cases,  $\phi$  is the effective range parameter, which sets the distance beyond which the cor-  
 675 relation function is effectively zero.

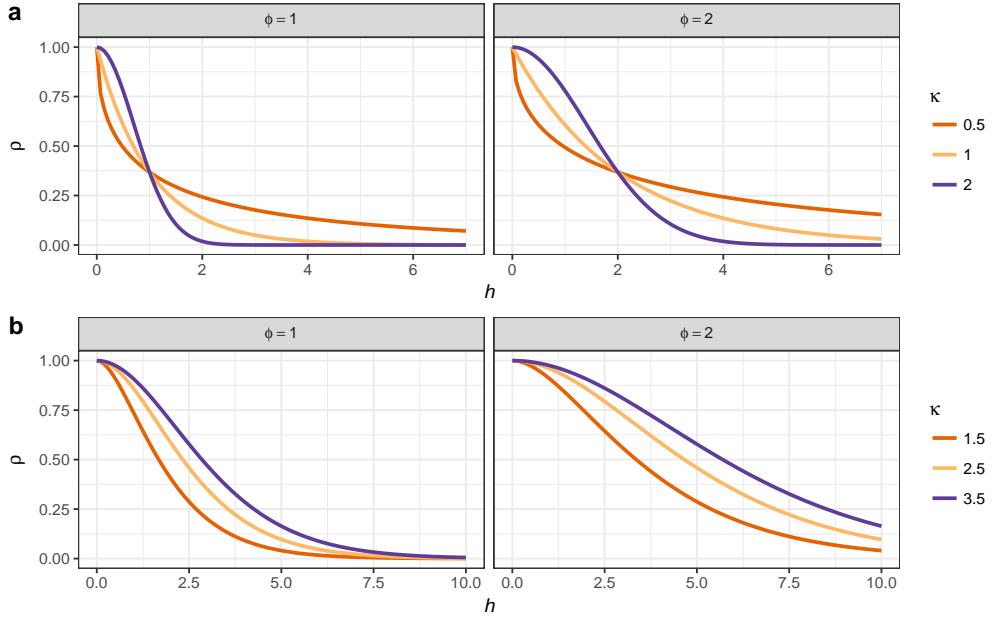


Figure 11: Power exponential (a) and Matérn (b) correlation functions for observation separation distance  $h$ . Two values of the effective range parameter ( $\phi$ ) are shown for each function. For the power exponential function,  $\kappa$  is the power in the power exponential function. For the Matérn correlation function,  $\kappa$  distinguishes the member of the Matérn family.

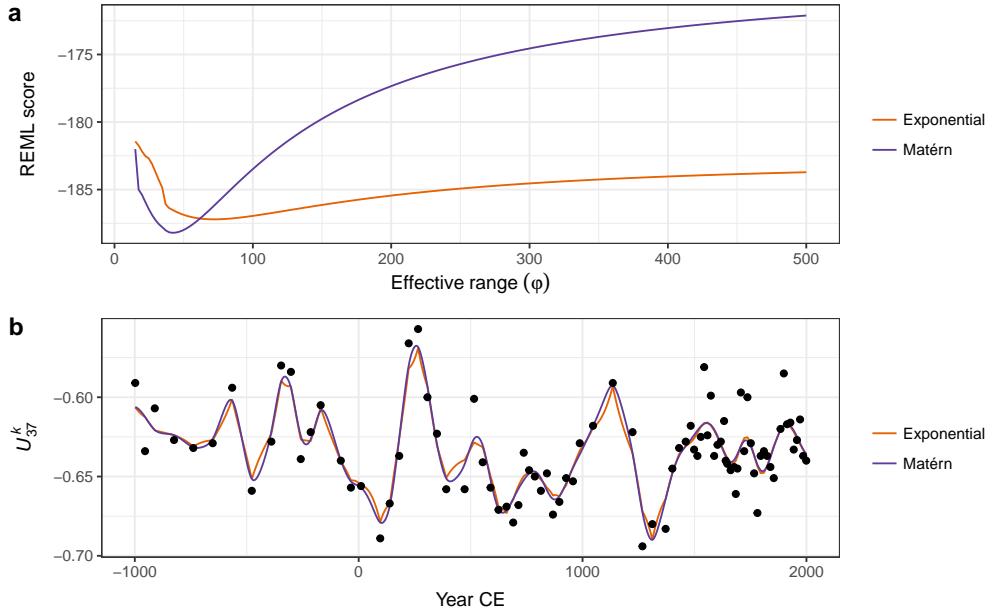


Figure 12: Gaussian process smooths fitted to the  $U_{37}^K$  time series. REML score traces for GAMs fitted using power exponential ( $\kappa = 1$ ) or Matérn ( $\kappa = 1.5$ ) correlation functions as a function of the effective range parameter ( $\phi$ ) are shown (a). The optimal model for each function is that with the lowest REML score. b) shows the resulting trends estimated using the respective correlation function with the value of  $\phi$  set to the optimal value.

676 Gaussian processes and GAMs share many similarities and we can fit a Gaussian process us-  
677 ing the techniques already described above for splines (Handcock et al., 1994; Kammann and  
678 Wand, 2003). It can be shown (e.g. Fahrmeir et al., 2013 section/pages) that the Gaussian  
679 process model has the same penalised likelihood form as the GAM that we discussed earlier;  
680 the observations are the knots of the smoother and each has a basis function in the form of  
681 a correlation function. The equivalence is only true if the basis functions do not depend on  
682 any other parameters of the model, which is only achievable if the value of  $\phi$  is fixed and  
683 known (Fahrmeir et al., 2013). In general, however, we would like to estimate  $\phi$  as part of  
684 model fitting. To achieve this we can maximise the profile likelihood or score statistic of the  
685 model over a range of values of  $\phi$  (Wood, 2017, 362–363). This involves proposing a value of  
686  $\phi$  for the effective range of the correlation function and then estimating the resulting GAM  
687 by minimising the penalised log-likelihood conditional upon this value of  $\phi$ . The model, and  
688 its corresponding value of  $\phi$ , with lowest penalised log-likelihood or score statistic is then re-  
689 tained as the estimated GAM. Figure 12a shows the REML score for models estimated using a  
690 Gaussian process smooth with a Matérn correlation function ( $\kappa = 1.5$ ) for a sequence of values  
691 of  $\phi$  between 15 and 1000 years. There is a clear minimum around 40 years separation, with  
692 the minimum REML score being observed at  $\phi = 41.81$ ). Also shown are the REML scores for  
693 models using the power exponential function ( $\kappa = 1$ ) with the minimum score observed at a  
694 somewhat higher effective range of  $\phi = 71.06$ .

695 Figure 12b shows the estimated trends for the  $U_{37}^K$  time series using Gaussian process smooths  
696 with exponential and Matérn correlations functions, both using  $\phi$  values at their respective  
697 optimal value as assessed using the REML score. The estimated trends are very similar to  
698 one another, although there is a noticeable difference in behaviour, with the power exponen-  
699 tial ( $\kappa = 1$ ) version being noticeably less-smooth than the Matérn version. This difference is  
700 attributable to the shapes of the respective correlation functions; the Matérn approaches a cor-  
701 relation of 1 smoothly as  $h$  approaches 0, whilst the power exponential with  $\kappa = 1$  approaches  
702 a correlation of 1 increasingly quickly with decreasing  $h$ . The power exponential with  $\kappa =$   
703 2, like the Matérn, approaches  $\phi = 1$  smoothly, and consequently the trend estimated using  
704 this correlation function is qualitatively similar to that estimated using the Matérn correlation  
705 function.

## 706 4.7 Adaptive smoothing

707 Each of the spline types that I have discussed so far share a common feature; the degree of  
708 wiggliness over the time series is fixed due to the use of a single smoothness parameter,  $\lambda$ . The  
709 definition of wigginess, as the integrated squared second derivative of the spline, ensures that  
710 fitted smoother does not jump about wildly. This assumes that the data themselves are well  
711 described by a smoothly varying trend. If we anticipate abrupt change or step-like responses to  
712 environmental forcing this underlying assumption of the GAM would suggest that the method  
713 is ill-suited to modelling palaeo time series in which such features are evident or expected.

714 While there is not much we can do within the GAM framework to model a series that contains  
715 both smooth trends and step-like responses, an adaptive smoother can help address problems  
716 where the time series consists of periods of rapid change in the mean combined with periods

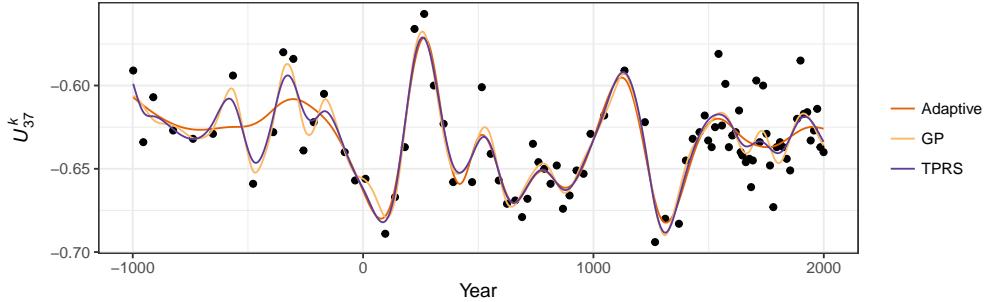


Figure 13: Comparison of trends estimated using i) adaptive smooth, ii) Gaussian process, and iii) thin plate regression spline bases for the  $U_{37}^K$  time series.

of complacency or relatively little change. As suggested by their name, adaptive smoothers can adjust to changes in the wiggliness of the time series. This adaptive behaviour is achieved by making the smoothness parameter  $\lambda$  itself depend smoothly on  $x_t$  (Ruppert et al., 2003, 17; Wood, 2017, 5.3.5); in other words, the adaptive smoother allows the wiggliness of the estimated trend to vary smoothly over time. Whilst this does allow the estimated trend to adapt to periods of rapid change in the response, adaptive smoothers make significant demands on the data (Wood, 2017, 5.3.5); if we used  $m$  smoothness penalties to allow the wiggliness to vary over a time series, it would be like estimating  $m$  separate smooths from chunks of the original series each of length  $n/m$ . In a practical sense, this limits the use of adaptive splines in palaeoecology to proxies that are readily enumerated, such as the biogeochemical proxies used in the two example data sets.

Figure 13 compares trends for the Braya-Sø time series estimated using GAMs with the three main types of spline discussed; i) TPRS, ii) Gaussian process smooths, and iii) an adaptive smoother using 45 basis functions and 5 smoothing parameters. There is a clear difference in the behaviour of the adaptive and non-adaptive smoothers for the first 1000 years of the record, with the adaptive smooth exhibiting much less variation compared with either the TPRS or Gaussian process splines. Over the remaining two thirds of the series, there is much closer agreement in the three smooths.

The behaviour of the TPRS and Gaussian process splines for these data is the result of requiring a large amount of wigginess (a small  $\lambda$ ) to adapt to the large oscillations in  $U_{37}^K$  present around year 250CE and again at ~900–1500CE. This large degree of wigginess allows the splines to potentially over-fit individual data points much earlier in the record. Because the adaptive smoother, in contrast, can adapt to these periods of rapid change in the response it is much less susceptible to this “chasing” behaviour — we don’t need to waste effective degrees of freedom in periods with little or no change just to be able to fit the data well when there is a lot of change.

This potential for over-fitting in such situations is undesirable, yet if we recall Figure 9 and the discussion around the use of the first derivative to identify periods of significant change, we would not interpret the oscillations in the early part of the  $U_{37}^K$  record as being statistically significant. Owing to the paucity of data in this part of the series the trends fitted using the non-adaptive smoothers are subject to such a large degree of uncertainty that the alternative

748 of no trend through the first 1000 years of the record is also a plausible explanation of the data.  
749 The trend estimated using the adaptive smooth reflects this. Therefore, should we conclude  
750 that there is no trend in  $U_{37}^K$  and thence climate in this period? I believe that to be too-strong a  
751 statement; those oscillations in  $U_{37}^K$  may be real responses to climate forcing but we may simply  
752 lack the statistical power to distinguish them from the null hypothesis of no trend through this  
753 period. The adaptive smoother is only adjusting to the data available to it; just because it does  
754 not detect a trend during this period does not lend itself to an interpretation of stable climate  
755 forcing or complacency in the lake's response to forcing. If there were particular interest in the  
756 climate of this particular period we might take from the Braya-Sø record that there is potential  
757 early variation in climate forcing, but that additional data from this or other sites is required  
758 before any definitive conclusion can be drawn.

## 759 4.8 Accounting for age model uncertainty

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

772 Figure 14a shows the estimated dates (in Years CE) for 12 levels in the Small Water core dated  
773 using  $^{210}\text{Pb}$ . The vertical bars show the estimated age uncertainty of each level. The solid line  
774 through the data points is an additive model fitted to the observations, with prior weights  
775 given by the estimated age uncertainties. The fitted age-depth model is constrained to be  
776 monotonically decreasing with increasing depth, following the method of (Pya and Wood,  
777 2015) using the *scam* package (Pya, 2017). Also shown are 25 simulations from the model pos-  
778 terior of the monotonically-constrained GAM. Each simulation from the posterior distribution  
779 of the age-model is itself a potential age-depth model, which can be used to assign dates to the  
780 Small Water core. The trend model in (4) can be fitted to the  $\delta^{15}\text{N}$  data using these new dates  
781 as  $x_t$ , and the whole process repeated for a large number of simulations from the age model.

782 Figure 14b shows the trend in  $\delta^{15}\text{N}$  for the observed age-depth model, plus trends estimated  
783 via the same model using 100 draws from the posterior distribution of the age model. In this  
784 case, the age-depth model is relatively simple with little variation in the posterior draws, re-  
785 sulting in trends that match closely that obtained from the estimated age-depth relationship.  
786 Even so, this additional uncertainty suggests that the timing of the decline in  $\delta^{15}\text{N}$  covers the  
787 interval ~1935–1945.

788 The uncertainty in the trend estimates illustrated in Figure 14b only reflects the variation in

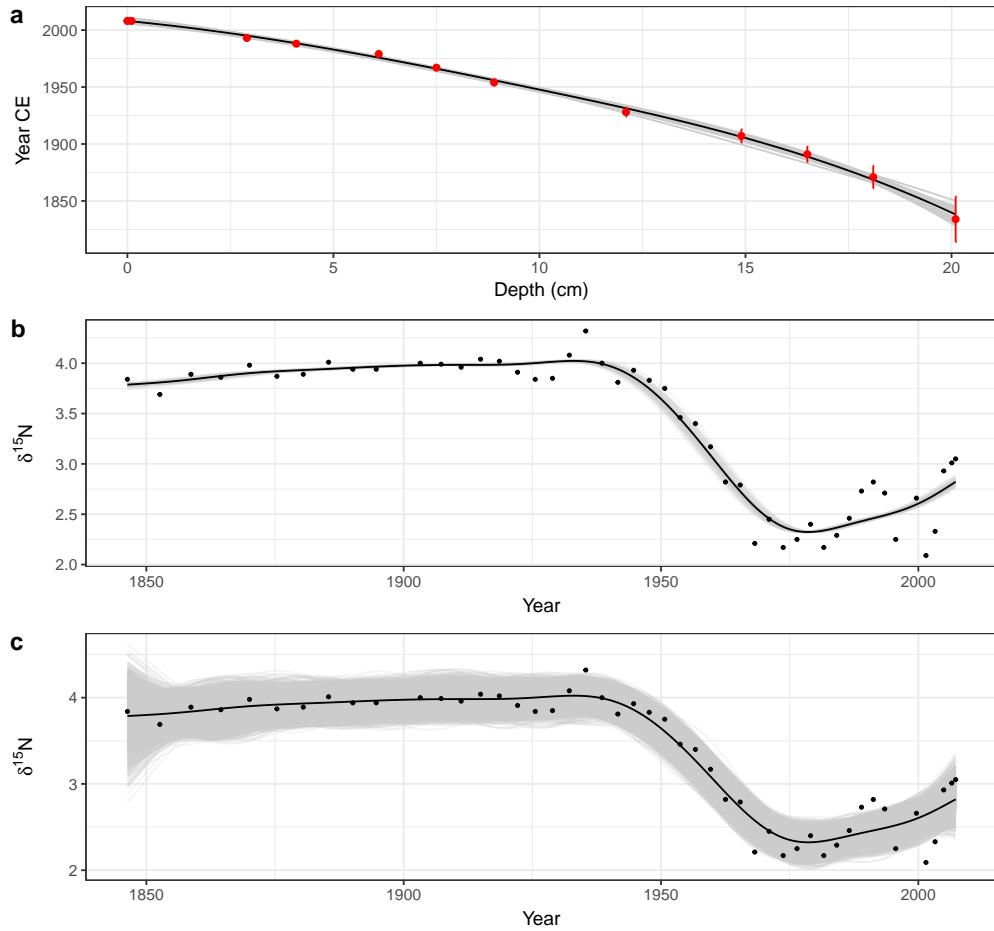


Figure 14: Accounting for uncertainty in age estimates whilst fitting a smooth trend to the Small Water  $\delta^{15}\text{N}$  time series. (a) Estimated age model using a monotonically-constrained spline fitted to  $^{210}\text{Pb}$  inferred ages for selected depths in the sediment core (red points). The uncertainty in the  $^{210}\text{Pb}$  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.

trends fitted to the uncertain dates of the sediment samples. To fully visualise the uncertainty in the trend estimates, incorporating both age model uncertainty *and* uncertainty in the estimated model coefficients themselves, 50 simulations from the posterior distribution of each of the 100 estimated trends shown in Figure 14b were performed, resulting in 5,000 trend estimates for the  $\delta^{15}\text{N}$  series. These are shown in Figure 14c, where the two obvious changes over the same simulations without accounting for uncertainty in  $x_t$  (Figure 7a) are that the uncertainty band traced out by the simulations is approximately 50% wider and, not surprisingly, the uncertainty in the estimated trend is most pronounced in the least accurately-dated section of the core. Despite this additional uncertainty however, the main result holds; a marked decline of  $\sim 1.5\%$  that occurred between approximately 1930 and 1945, with mild evidence of a small increase in  $\delta^{15}\text{N}$  post 2000 CE.

## 4.9 Multivariate data

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

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

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

A pragmatic although inelegant approach that has been used to estimate trends in multivariate palaeoecological data is to first summarise the response data using an unconstrained ordination via a PCA, CA, or principal curve and then fit separate GAM models to the site (sample) scores of the first few ordination axes or principal curve (Beck et al., 2018; Bennion et al., 2015). Whilst this two-step approach is relatively easy to implement and builds on approaches that palaeoecologists already use to summarise multivariate stratigraphic data, it is best thought of

830 as modelling changes in abundance or relative composition at the community level. It is less  
831 well suited to unpicking taxon-specific trends however, because the ordination step combines  
832 individual species information into latent variables (axes) that are linear combinations of *all*  
833 species and it is these latent variables that are then modelled using GAM.

## 834 5 Conclusions

835 Formal statistical estimation of trends in palaeoenvironmental data has been hampered by the  
836 nature of the data that comprise the time series; the uneven spacing of samples in time makes  
837 it, if not impossible, difficult to fit classical statistical time series models like ARIMA. This has  
838 lead palaeoecologists and palaeolimnologists to fall back on basic statistical methods such as  
839 linear parametric and non-parametric correlations or simple linear regression models, where  
840 the assumptions of the method are often grossly violated by the dependencies inherent to  
841 time series data. GAMs, whilst similar to the popular Loess smoother, provide a superior al-  
842 ternative approach to trend estimation in palaeoenvironmental time series. They can estimate  
843 non-linear trends, provide estimates of the magnitude of change as well as allow the identifica-  
844 tion of periods of change, can account for the lack of independence (either via autocorrelation  
845 processes or via the fitting of a wiggly trend), and provide a formal framework for statistical  
846 inference on each of these features.

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

## 851 Conflict of interest statement

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

## 854 Acknowledgements

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

863 **References**

- 864 Anchukaitis, K. J., and Tierney, J. E. (2013). Identifying coherent spatiotemporal modes in time-  
865 uncertain proxy paleoclimate records. *Climate Dynamics* 41, 1291–1306. doi:[10.1007/s00382-012-1483-0](https://doi.org/10.1007/s00382-012-1483-0).
- 866
- 867 Beck, K. K., Fletcher, M.-S., Gadd, P. S., Heijnis, H., Saunders, K. M., Simpson, G. L., et al.  
868 (2018). Variance and Rate-of-Change as early warning signals for a critical transition in an  
869 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).
- 870
- 871 Bennion, H., Simpson, G. L., and Goldsmith, B. J. (2015). Assessing degradation and recovery  
872 pathways in lakes impacted by eutrophication using the sediment record. *Frontiers in Ecology  
873 and Evolution* 3. doi:[10.3389/fevo.2015.00094](https://doi.org/10.3389/fevo.2015.00094).
- 874 Bergmeir, C., Hyndman, R. J., and Koo, B. (2018). A note on the validity of cross-validation for  
875 evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis* 120,  
876 70–83. doi:[10.1016/j.csda.2017.11.003](https://doi.org/10.1016/j.csda.2017.11.003).
- 877 Birks, H. J. B. (1998). Numerical tools in palaeolimnology — progress, potentialities, and prob-  
878 lems. *Journal of Paleolimnology* 20, 307–332. doi:[10.1023/A:1008038808690](https://doi.org/10.1023/A:1008038808690).
- 879 Birks, H. J. B. (2012a). “Introduction and overview of part III,” in *Tracking environmental change  
880 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).
- 881 Birks, H. J. B. (2012b). “Overview of numerical methods in palaeolimnology,” in *Tracking envi-  
882 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).
- 883
- 884 Blaauw, M., and Heegaard, E. (2012). “Estimation of Age-Depth relationships,” in *Tracking  
885 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).
- 886
- 887 Brassell, S. C. (1993). “Applications of biomarkers for delineating marine paleoclimatic fluctu-  
888 ations during the pleistocene,” in *Organic geochemistry: Principles and applications*, eds. M. H. En-  
889 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).
- 889
- 890 Chu, G., Sun, Q., Li, S., Zheng, M., Jia, X., Lu, C., et al. (2005). Long-chain alkenone distribu-  
891 tions and temperature dependence in lacustrine surface sediments from china. *Geochimica et  
892 Cosmochimica Acta* 69, 4985–5003. doi:[10.1016/j.gca.2005.04.008](https://doi.org/10.1016/j.gca.2005.04.008).
- 893
- 894 Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Jour-  
895 nal of the American Statistical Association* 74, 829–836. doi:[10.1080/01621459.1979.10481038](https://doi.org/10.1080/01621459.1979.10481038).
- 896
- Craven, P., and Wahba, G. (1978). Smoothing noisy data with spline functions. *Numerische  
Mathematik* 31, 377–403. doi:[10.1007/BF01404567](https://doi.org/10.1007/BF01404567).
- 897
- Duchon, J. (1977). “Splines minimizing rotation-invariant semi-norms in sobolev spaces,”  
898 in *Constructive theory of functions of several variables* (Springer, Berlin, Heidelberg), 85–100.  
899 doi:[10.1007/BFb0086566](https://doi.org/10.1007/BFb0086566).
- 900
- Dutilleul, P., Cumming, B. F., and Lontoc-Roy, M. (2012). “Autocorrelogram and periodogram  
901 analyses of palaeolimnological Temporal-Series from lakes in central and western north amer-

- 902 ica to assess shifts in drought conditions," in *Tracking environmental change using lake sediments*  
903 (Springer, Dordrecht), 523–548. doi:[10.1007/978-94-007-2745-8\16](https://doi.org/10.1007/978-94-007-2745-8_16).
- 904 D'Andrea, W. J., Huang, Y., Fritz, S. C., and Anderson, N. J. (2011). Abrupt holocene climate  
905 change as an important factor for human migration in west greenland. *Proceedings of the National*  
906 *Academy of Sciences* 108, 9765–9769. doi:[10.1073/pnas.1101708108](https://doi.org/10.1073/pnas.1101708108).
- 907 Epperson, J. F. (1987). On the Runge example. *The American Mathematical Monthly* 94, 329–341.  
908 doi:[10.2307/2323093](https://doi.org/10.2307/2323093).
- 909 Fahrmeir, L., and Kneib, T. (2008). "On the identification of trend and correlation in temporal  
910 and spatial regression," in *Recent advances in linear models and related areas* (Physica-Verlag HD),  
911 1–27. doi:[10.1007/978-3-7908-2064-5\1](https://doi.org/10.1007/978-3-7908-2064-5_1).
- 912 Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. (2013). *Regression: Models, methods and applications*.  
913 Springer Berlin Heidelberg doi:[10.1007/978-3-642-34333-9](https://doi.org/10.1007/978-3-642-34333-9).
- 914 Gautheir, T. D. (2001). Detecting trends using spearman's rank correlation coefficient. *Envi-  
915 ronmental Forensics* 2, 359–362. doi:[10.1080/713848278](https://doi.org/10.1080/713848278).
- 916 Glew, J. R., Smol, J. P., and Last, W. M. (2001). "Sediment core collection and extrusion," in  
917 *Tracking environmental change using lake sediments: Basin analysis, coring, and chronological techniques*,  
918 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).
- 920 Golding, N., and Purse, B. V. (2016). Fast and flexible bayesian species distribution modelling  
921 using gaussian processes. *Methods in Ecology and Evolution*. doi:[10.1111/2041-210X.12523](https://doi.org/10.1111/2041-210X.12523).
- 922 Handcock, M. S., Meier, K., and Nychka, D. (1994). Kriging and splines: An empirical compar-  
923 ision of their predictive performance in some applications: Comment. *Journal of the American  
924 Statistical Association* 89, 401–403. doi:[10.2307/2290838](https://doi.org/10.2307/2290838).
- 925 Hastie, T. J., and Tibshirani, R. J. (1990). *Generalized additive models*. Boca Raton, Fl.: Chapman  
926 & Hall / CRC.
- 927 Hastie, T., and Tibshirani, R. (1986). Generalized additive models. *Statistical Science* 1, 297–310.
- 928 Juggins, S., and Telford, R. J. (2012). "Exploratory data analysis and data display," in *Tracking  
929 environmental change using lake sediments* (Springer, Dordrecht), 123–141. doi:[10.1007/978-94-007-2745-8\5](https://doi.org/10.1007/978-94-007-2745-8_5).
- 931 Kammann, E. E., and Wand, M. P. (2003). Geoadditive models. *Journal of the Royal Statistical  
932 Society. Series C, Applied statistics* 52, 1–18. doi:[10.1111/1467-9876.00385](https://doi.org/10.1111/1467-9876.00385).
- 933 Kimeldorf, G. S., and Wahba, G. (1970). A correspondence between bayesian estimation on  
934 stochastic processes and smoothing by splines. *Annals of Mathematical Statistics* 41, 495–502.
- 935 Magee, L. (1998). Nonlocal behavior in polynomial regressions. *The American Statistician* 52,  
936 20–22. doi:[10.2307/2685560](https://doi.org/10.2307/2685560).
- 937 Mann, M. E. (2004). On smoothing potentially non-stationary climate time series. *Geophysical  
938 Research Letters* 31, L07214. doi:[10.1029/2004GL019569](https://doi.org/10.1029/2004GL019569).
- 939 Mann, M. E. (2008). Smoothing of climate time series revisited. *Geophysical Research Letters* 35,

- 940 L16708. doi:[10.1029/2008GL034716](https://doi.org/10.1029/2008GL034716).
- 941 Marra, G., and Wood, S. N. (2012). Coverage properties of confidence intervals for generalized  
942 additive model components. *Scandinavian Journal of Statistics, Theory and Applications* 39, 53–74.  
943 doi:[10.1111/j.1467-9469.2011.00760.x](https://doi.org/10.1111/j.1467-9469.2011.00760.x).
- 944 McCullagh, P., and Nelder, J. A. (1989). *Generalized linear models, second edition*. CRC Press.
- 945 Mills, T. C. (2006). Modelling current trends in Northern Hemisphere temperatures. *International  
946 Journal of Climatology* 26, 867–884. doi:[10.1002/joc.1286](https://doi.org/10.1002/joc.1286).
- 947 Mills, T. C. (2007). A note on trend decomposition: The “classical” approach revisited  
948 with an application to surface temperature trends. *Journal of applied statistics* 34, 963–972.  
949 doi:[10.1080/02664760701590418](https://doi.org/10.1080/02664760701590418).
- 950 Mills, T. C. (2010). “Skinning a cat”: Alternative models of representing temperature trends.  
951 *Climatic Change* 101, 415–426.
- 952 Nychka, D. (1988). Bayesian confidence intervals for smoothing splines. *Journal of the American  
953 Statistical Association* 83, 1134–1143. doi:[10.1080/01621459.1988.10478711](https://doi.org/10.1080/01621459.1988.10478711).
- 954 PAGES 2K Consortium (2013). Continental-scale temperature variability during the past two  
955 millennia. *Nature Geoscience* 6, 339–346. doi:[10.1038/ngeo1797](https://doi.org/10.1038/ngeo1797).
- 956 Pinheiro, J. C., and Bates, D. M. (2000). *Mixed-Effects models in S and S-PLUS*. Springer Science  
957 & Business Media.
- 958 Pya, N. (2017). *Scam: Shape constrained additive models*. Available at: <https://CRAN.R-project.org/package=scam>.
- 959
- 960 Pya, N., and Wood, S. N. (2015). Shape constrained additive models. *Statistics and Computing*  
961 25, 543–559. doi:[10.1007/s11222-013-9448-7](https://doi.org/10.1007/s11222-013-9448-7).
- 962 Pya, N., and Wood, S. N. (2016). A note on basis dimension selection in generalized additive  
963 modelling. *ArXiv e-prints*.
- 964 R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R  
965 Foundation for Statistical Computing Available at: <https://www.R-project.org/>.
- 966 Rasmussen, C. E., and Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT  
967 Press.
- 968 Reiss, P. T., and Ogden, R. T. (2009). Smoothing parameter selection for a class of semipara-  
969 metric linear models. *Journal of the Royal Statistical Society. Series B, Statistical methodology* 71,  
970 505–523. doi:[10.1111/j.1467-9868.2008.00695.x](https://doi.org/10.1111/j.1467-9868.2008.00695.x).
- 971 Runge, C. (1901). Über empirische funktionen und die interpolation zwischen äquidistanten  
972 ordinaten. *Zeitschrift fur Angewandte Mathematik und Physik* 46, 224–243.
- 973 Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge Uni-  
974 versity Press.
- 975 Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric  
976 regression curve fitting. *Journal of the Royal Statistical Society. Series B, Statistical methodology* 47,

- 977 1–52.
- 978 Smol, J. P. (2008). *Pollution of lakes and rivers: A paleoenvironmental perspective*. Blackwell Pub.
- 979 Smol, J. P., Birks, H. J. B., Lotter, A. F., and Juggins, S. (2012). “The march towards the quanti-  
980 tative analysis of palaeolimnological data,” in *Tracking environmental change using lake sediments*  
981 (Springer, Dordrecht), 3–17. doi:[10.1007/978-94-007-2745-8\1](https://doi.org/10.1007/978-94-007-2745-8_1).
- 982 Tian, J., Nelson, D. M., and Hu, F. S. (2011). How well do sediment indicators record past  
983 climate? An evaluation using annually laminated sediments. *Journal of Paleolimnology* 45, 73–  
984 84. doi:[10.1007/s10933-010-9481-x](https://doi.org/10.1007/s10933-010-9481-x).
- 985 Toney, J. L., Huang, Y., Fritz, S. C., Baker, P. A., Grimm, E., and Nyren, P. (2010). Cli-  
986 matic and environmental controls on the occurrence and distributions of long chain  
987 alkenones in lakes of the interior united states. *Geochimica et Cosmochimica Acta* 74, 1563–1578.  
988 doi:[10.1016/j.gca.2009.11.021](https://doi.org/10.1016/j.gca.2009.11.021).
- 989 Wahba, G. (1983). Bayesian “confidence intervals” for the Cross-Validated smoothing spline.  
990 *Journal of the Royal Statistical Society. Series B, Statistical methodology* 45, 133–150.
- 991 Wahba, G. (1990). *Spline models for observational data*. SIAM.
- 992 Wang, Y., Naumann, U., Wright, S. T., and Warton, D. I. (2012). Mvabund- an R package for  
993 model-based analysis of multivariate abundance data. *Methods in Ecology and Evolution* 3, 471–  
994 474. doi:[10.1111/j.2041-210X.2012.00190.x](https://doi.org/10.1111/j.2041-210X.2012.00190.x).
- 995 Warton, D. I. (2011). Regularized sandwich estimators for analysis of high-dimensional  
996 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).
- 997 998 Warton, D. I., Blanchet, F. G., O’Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., et al.  
999 (2015). So many variables: Joint modeling in community ecology. *Trends in Ecology & Evolution*.  
1000 doi:[10.1016/j.tree.2015.09.007](https://doi.org/10.1016/j.tree.2015.09.007).
- 1001 1002 Warton, D. I., Wright, S. T., and Wang, Y. (2012). Distance-based multivariate analy-  
1003 ses confound location and dispersion effects. *Methods in Ecology and Evolution* 3, 89–101.  
doi:[10.1111/j.2041-210X.2011.00127.x](https://doi.org/10.1111/j.2041-210X.2011.00127.x).
- 1004 Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society. Series B,*  
1005 *Statistical methodology* 65, 95–114. doi:[10.1111/1467-9868.00374](https://doi.org/10.1111/1467-9868.00374).
- 1006 Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for  
1007 generalized additive models. *Journal of the American Statistical Association* 99, 673–686.  
1008 doi:[10.2307/27590439](https://doi.org/10.2307/27590439).
- 1009 Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estima-  
1010 tion of semiparametric generalized linear models. *Journal of the Royal Statistical Society. Series B,*  
1011 *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).
- 1012 Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R, Second edition*. CRC Press.
- 1013 Wood, S. N., Pya, N., and Säfken, B. (2016). Smoothing parameter and model selection  
1014 for general smooth models. *Journal of the American Statistical Association* 111, 1548–1563.

- 1015 doi:[10.1080/01621459.2016.1180986](https://doi.org/10.1080/01621459.2016.1180986).
- 1016 Yee, T. W., and Mitchell, N. D. (1991). Generalized additive models in plant ecology. *Journal of*  
 1017 *Vegetation Science* 2, 587–602. doi:[10.2307/3236170](https://doi.org/10.2307/3236170).
- 1018 Zink, K.-G., Leythaeuser, D., Melkonian, M., and Schwark, L. (2001). Temperature de-  
 1019 pendency of long-chain alkenone distributions in recent to fossil limnic sediments and  
 1020 in lake waters11Associate editor: J. b. fein. *Geochimica et Cosmochimica Acta* 65, 253–265.  
 1021 doi:[10.1016/S0016-7037\(00\)00509-3](https://doi.org/10.1016/S0016-7037(00)00509-3).

## 1022 Appendix 1 — Simultaneous intervals

- 1023 We proceed by considering a simultaneous confidence interval for a function  $f(x)$  at a set of  $M$   
 1024 locations in  $x$ ; we'll refer to these locations, following the notation of Ruppert et al. (2003) by

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

- 1025 The true function over  $\mathbf{g}$ ,  $\mathbf{f}_g$ , is defined as the vector of evaluations of  $f$  at each of the  $M$  locations

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

- 1026 and the corresponding estimate of the true function given by the fitted GAM denoted by  $\hat{\mathbf{f}}_g$ .  
 1027 The difference between the true function and our unbiased estimator is given by

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

- 1028 where  $\mathbf{C}_g$  is a matrix formed by the evaluation of the basis functions at locations  $\mathbf{g}$ , and the part  
 1029 in square brackets is the bias in the estimated model coefficients, which we assume to be mean 0  
 1030 and distributed, approximately, multivariate normal with mean vector  $\mathbf{0}$  and covariance matrix  
 1031  $\mathbf{V}_b$

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

- 1032 where  $\mathbf{V}_b$  is the Bayesian covariance matrix of the GAM coefficients.

- 1033 Now, the  $(1 - \alpha)100\%$  simultaneous confidence interval is

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

1034 where  $m_{1-\alpha}$  is the  $1 - \alpha$  quantile of the random variable

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

1035 The sup refers to the *supremum* or the *least upper bound*; this is the least value of  $\mathcal{X}$ , the set of all  
1036 values of which we observed subset  $x$ , that is *greater* than all of the values in the subset. Often  
1037 this is the maximum value of the subset. This is what is indicated by the right-hand side of  
1038 the equation; we want the maximum (absolute) value of the ratio over all values in  $g$ .

1039 The fractions in both sides of the equation correspond to the standardized deviation between  
1040 the true function and the model estimate, and we consider the *maximum absolute* standardized  
1041 deviation. We don't usually know the distribution of the maximum absolute standardized  
1042 deviation but we need this to access its quantiles. However, we can closely approximate the  
1043 distribution via simulation. The difference here is that rather than simulating from the poste-  
1044 rior of the model as we did earlier [see section Confidence intervals](#), this time we simulate from  
1045 the multivariate normal distribution with mean vector  $\mathbf{0}$  and covariance matrix  $\mathbf{V}_b$ . For each  
1046 simulation we find the maximum absolute standardized deviation of the fitted function from  
1047 the true function over the grid of  $x$  values we are considering. Then we collect all these max-  
1048 ima, sort them and either take the  $1 - \alpha$  probability quantile of the maxima, or the maximum  
1049 with rank  $\lceil (1 - \alpha)/N \rceil$ .