

1           **A test for non-linearity in temperature proxy records**

                  BÅRD STØVE,

2                   *Department of Mathematics, University of Bergen, Bergen, Norway*

                  FREDRIK CHARPENTIER LJUNGQVIST

3                   *Department of History, Stockholm University, Stockholm, Sweden*

                  AND PETER THEJLL\*

4                   *Danish Climate Centre, Copenhagen, Denmark*

---

\* *Corresponding author address:* Peter Thejll, Danish Meteorological Institute, Danish Climate Centre, Lyngbyvej 100, DK-2100 Copenhagen Ø, Denmark.

E-mail: pth@dmi.dk

## ABSTRACT

6 Are temperature proxy records linear recorders of past temperature conditions? We apply a  
7 statistical test for linearity to 15 millennial-long proxy records with annual resolution that  
8 were shown to significantly respond to Northern Hemisphere annual mean temperature  
9 selected from a collection of 30 proxies. The test, based on generalized additive modeling,  
10 shows that most of the proxies can indeed be shown to be linear functions of annual mean  
11 temperature, but two proxy records do not appear to have a linear relationship with tem-  
12 perature – this supports the assumption of linearity in most climate reconstruction work.  
13 The method tests for non-linearity in a proxy relative to the group of proxies it is being used  
14 together with. We test robustness of the results and find that the results are stable to choice  
15 of proxies. The linearity-testing method is quite general and could in the future be used for  
16 larger and more extensive sets of proxies.

# 1. Introduction

Considering the fact that a systematic network of instrumental temperature measurements around the globe started as late as in the second half of the 19th century we are dependent on other types of information to understand temperature variability prior to that. Such information is essential in order to evaluate whether the recent global warming falls outside the range of the natural variability of the last one or two millennia in either magnitude or rate (NRC (2006); Jansen et al. (2007)), as well as providing material that may be used to evaluate paleo-climate model output and the paleo-model forcings. Our knowledge of past temperature variability must be drawn from temperature-sensitive proxy data. Such data can be extracted from both historical records and from various natural recorders of climate variability such as corals, fossil pollen, ice-cores, lake and marine sediments, speleothems, and tree-ring width and density. A review of different types of temperature proxy data is given in, for example, Bradley (1999) and Jones et al. (2009), and the availability of millennia-long temperature proxy records in the Northern Hemisphere is synthesized in Ljungqvist et al. (2012).

Multi-proxy temperature reconstructions are based on training a model on data-sets that include both observations and proxy data. Then, using the trained model on proxies alone a model-based value for the observational quantity is extracted. Most current reconstruction methods used are linear – a linear proportionality between a signal in the proxy and in the real world is assumed. But do we know *a priori* whether the relationship actually is linear? We expect that linear methods will produce better reconstructions if the proxies are linear temperature recorders. While knowledge of the biological and geological systems that lay down the proxy records helps to understand which systems are linear (Frank et al. 2010) it is of interest to have quantitative tests for linearity.

There has been an increasing awareness that climate proxy records do not always show a linear relationship to temperature (Tingley et al. 2012). This has been most evident in the field of dendroclimatology where the so-called divergence problem, the fact that some high-

44 latitude tree-ring records show a lessened or even negative response to higher temperatures  
45 in the late 20th century, has been studied by numerous researchers (Andreu-Hayles et al.  
46 (2011); Briffa et al. (1998); Briffa et al. (1998); D’Arrigo et al. (2008); Loehle (2009); Wilson  
47 et al. (2007) and the references there-in). The non-linear properties of some tree-ring width  
48 records have in pioneering studies been assessed by, for example, Carrer and Urbinati (2001),  
49 Fritts (1976), Graumlich and Brubaker (1986), and D’Arrigo et al. (2004). The non-linear  
50 response of some temperature sensitive proxy records to temperature has also been brought  
51 to attention regarding low-resolution proxy archives such as chironomid, diatom and pollen.  
52 Non-parametric methods have been applied for reconstructing temperature from such proxies  
53 by, among others, Birks (1995), and Birks et al. (2010) and Bayesian reconstruction tech-  
54 niques that feature non-linear modeling have been used by Haslett et al. (2006), Korhola  
55 (2002), Toivonen et al. (2001), and Vasko et al. (2000).

56 We present here methods and tests that can reveal non-linear relationship between prox-  
57 ies and temperature, so-called non-parametric methods. These methods are well-known in  
58 statistics, see e.g. chapter 10 in Teräsvirta et al. (2010), but only occasionally utilized in  
59 paleoclimatic research. By using a non-parametric method the aim is to avoid assumptions  
60 on the parametric form of the relationship in question. We let the data ’speak for themselves’  
61 and it enables us to find a function that describes the available data well. This is in contrast  
62 to parametric modelling, where a specific model with parameters is assumed to generate the  
63 data in question. One such method in climate reconstruction is the simple ”direct regres-  
64 sion” method (pioneered by Groveman and Landsberg (1979)). In such a parametric model,  
65 it can be easy to do inference and great gains in efficiency are possible, however, only if the  
66 model is (almost) true. If the assumed model is incorrect, inferences can be useless, leading  
67 to misleading interpretations of the data. Non-parametric models provide a simple way to  
68 find structures in data sets without imposing a parametric model, and it is also possible to  
69 test whether relationships are linear.

70 In this paper we will examine two such models; a non-parametric additive model and a

71 semi-parametric additive model which allows for some linear and some non-linear regressors.  
72 For a purely illustrative purpose we will compare these two methods with reconstructions  
73 using the simple "direct regression" model. The rest of the paper is organized as follows.  
74 In section 2 we give a brief overview of the non-parametric models, and the corresponding  
75 estimation methodology. In section 3 we present the proxy data and temperature series used,  
76 while section 4 reports the results of the fitted models and tests of linearity. Finally, section  
77 5 concludes. In the Appendix we demonstrate the additive non-parametric method in a test  
78 example, to show that it works on realistic data with known properties.

## 79 2. Methods

### 80 a. *Linear regression model*

81 In the multivariate regression – used for instance already in Groveman and Landsberg  
82 (1979) – a NH mean temperature  $Y$  is expressed as a linear sum over  $d$  selected proxies  $X$ ,

$$Y = \alpha + \sum_{j=1}^d \beta_j \times X_j + \epsilon. \quad (1)$$

83 "Training", to determine the coefficients  $(\alpha, \beta_j)$ , is performed during an interval where  
84 both  $Y$  and the  $X$ 's are available, and the determined coefficients are then used to build a  
85 model for use at all other times. Assuming data stationarity is central to this step. This  
86 method is still in some use although several alternatives exist – see Christiansen et al. (2009)  
87 for a current review. We stress that it is used here for illustrative purposes only.

### 88 b. *Non-parametric regression models*

89 The aim of non-parametric models is to relax assumptions on the form of a regression  
90 function, and to let the data search for a suitable function that describes the available data  
91 well. These approaches are powerful in exploring fine structural relationships and provide

92 very useful diagnostic tools for parametric models.

In a multivariate regression problem we want to study the relationship between the response variable  $Y$  and the vector of co-variates  $\mathbf{X} = (X_1, \dots, X_d)^T$  via

$$m(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x}).$$

It is useful to model the unknown regression function  $m(\mathbf{x})$  additively, that is,

$$m(\mathbf{x}) = \sum_{j=1}^d m_j(x_j). \quad (2)$$

Usually an intercept term is added, i.e.  $E(Y) = \alpha$ . This gives us the following additive non-parametric model,

$$Y = \alpha + \sum_{j=1}^d m_j(X_j) + e, \quad (3)$$

where  $m_1, \dots, m_d$  are unknown uni-variate functions,  $E(e) = 0$ ,  $\text{Var}(e) = \sigma^2$  and  $e$  is independent of the vector of co-variates  $\mathbf{X}$ . To ensure identifiability,  $m_1, \dots, m_d$  are required to satisfy

$$E[m_j(X_j)] = 0, \quad j = 1, \dots, d, \quad (4)$$

93 which implies that  $E(Y) = \alpha$ .

94 Estimation of the unknown functions  $m_1, \dots, m_d$  is done by the back-fitting algorithm,  
95 introduced by Breiman and Friedman (1985) and Buja et al. (1989). Note first, that if the  
96 additive model, (3), is correct then

$$E[Y - \alpha - \sum_{j \neq k} m_j(X_j) | X_k] = m_k(X_k), \quad k = 1, \dots, d. \quad (5)$$

97 This relationship suggests an iterative procedure for the estimation of the unknown func-  
98 tions. Thus for a known constant  $\alpha$  and given functions  $m_j$ ,  $j \neq k$ , the function  $m_k$  can  
99 be estimated by a uni-variate regression fit based on the observations  $(X_k^i, Y_i)$ ,  $i = 1, \dots, n$ ,  
100 where  $X_k^i$  is the  $i$ th observation of the  $k$ th additive variable. Denote the uni-variate smoother  
101 of  $m_k$  by  $S_k$ . The algorithm works as follows:

102

103 **Step 1.** Initialization:  $\hat{\alpha} = n^{-1} \sum_{i=1}^n Y_i$ ,  $\hat{m}_k = m_k^0$  for  $k = 1, \dots, d$ .

**Step 2.** Find new transformations: For  $k = 1, \dots, d$ :

$$\hat{m}_k = S_k[Y - \hat{\alpha} - \sum_{j \neq k} \hat{m}_j(X_j) | X_k];$$

centre the estimator to obtain  $\hat{m}_k^* = \hat{m}_k - n^{-1} \sum_{i=1}^n \hat{m}_k(X_k^i)$ ,

$$\text{and } \hat{\alpha}^* = \hat{\alpha} + n^{-1} \sum_{i=1}^n \hat{m}_k(X_k^i).$$

104 **Step 3.** Repeat step 2 until convergence.

105

106 The idea behind this algorithm is to carry out a fit, calculate partial residuals from that  
107 fit, and refit again. That is why the iteration scheme is called back-fitting. The starting  
108 functions  $m_1^0, \dots, m_d^0$  can be obtained in various ways, for example, from a linear regression  
109 fit of  $Y$  on the co-variates  $X_k$ . The smoothing operator  $S_k$  can be other non-parametric  
110 regression estimators such as kernel methods, see e.g. Fan and Gijbels (1996).

This way of modelling non-linear relationships between a response and several predictors has been extensively used within the statistical community for some years. In fact, the additive model above, is a version of a wider model, called generalized additive model (GAM), see Hastie and Tibshirani (1990). Here the conditional mean ( $m(\mathbf{X})$ ) of a response  $Y$  is related to an additive function of the predictors via a link function  $g$ :

$$g[m(\mathbf{X})] = \alpha + m_1(X_1) + \dots + m_p(X_d). \tag{6}$$

111 See also Hastie and Tibshirani (1990) and Hastie et al. (2009) for further technical details  
112 regarding the back-fitting algorithm and non-parametric regression models. An alternative  
113 to the back-fitting algorithm is the marginal integration method, see Tjøstheim and Auestad  
114 (1994).

115 In the linear regression model each regressor represents one degree of freedom – in the  
116 additive model more degrees of freedom are used up, see Hastie and Tibshirani (1990) – it is

117 therefore generally advantageous to consider models that contain both linear and non-linear  
118 terms; these are called semi-parametric models. The methods are implemented in R – see  
119 [www.r-project.org](http://www.r-project.org) – and the software used in this paper is available upon request.

### 120 **3. Data**

121 A wide range of different kinds of proxies with annual resolution have been used in this  
122 study: one speleothem micro-layer thickness record, three ice-core  $\delta^{18}O$  records (one of which  
123 a composite of several others, see below), seven varved thickness sediment records, thirteen  
124 tree-ring width records, four tree-ring maximum latewood density records, one tree-ring  
125 height-increment record, and one tree-ring  $\delta^{13}C$  record. Information about each of these  
126 30 proxy records is given in Table 1, such as name of the proxy, latitude and longitude,  
127 type of proxy, season bias, dating uncertainty, and reference to the original publication. We  
128 only use proxies with annual resolution – this places the proxies on a comparable basis to  
129 the instrumental data they are calibrated against – but excludes most presently available  
130 millennia-long proxies.

131 A general problem for any reconstruction is the risk of 'over-fitting' – i.e. that we have  
132 so many proxies for which we have to find a coefficient or, in the case of GAM, an additive  
133 smoothed model-term, that we are using up all or almost all the degrees of freedom and  
134 therefore can, in the extreme case, explain all the variance in the calibration data set and  
135 at the same time have a limited or even reduced explanatory power in independent data.  
136 As we have restricted our proxies to those with annual resolution we have eliminated many  
137 other, long, series and thereby minimized the risk of over-fitting.

138 We test the set of 30 proxies for relevance by screening them in terms of how well they  
139 correlate to both the Northern Hemispheric mean temperature and the local temperatures –  
140 See Table 1. As noted by Juckes et al. (2007), choosing proxies by screening their correlation  
141 to temperature is not an entirely unproblematic approach. Even when screening proxies to



142 temperature, there still exists a risk that poor proxies are included because of insufficient in-  
143 strumental temperature data for the screening process. The uncertainty estimation becomes  
144 more problematic if temperature measurements have been used already in the data selection.  
145 We found that 15 of the proxies are both significantly correlated (standard two-sided T-test,  
146  $p=0.01$ ) to global and local temperatures, and these form the set we use in the subsequent  
147 analysis. We use a two-sided test since, *a priori*, we do not know if the proxy responds  
148 with an increase or a decrease when the temperature rises. Most temperature proxy records  
149 have temporal and seasonal limitations (Bradley (1999); Bradley and Jones (1992); Jones  
150 et al. (2009)). The proxies have different optimal season response and few proxies are likely  
151 to fully reflect annual mean temperature (Ljungqvist 2010). The response to the proxys  
152 optimal season can be higher than the correlation to annual mean temperature used here.

153 Most of the series we have here are continuous, but the series for Iceberg Lake and  
154 Jämtland have gaps, and the Greenland composite and Southern Sierra Nevada end before  
155 1990 – the missing values, in the standardized series, are replaced with 0’s. This is an  
156 unacceptable practice were the reconstructions to be used as such – but they are calculated  
157 and shown in this paper only to help illustrate the non-linearity testing method. To train  
158 the models we use the Northern Hemisphere annual mean temperature data from the  $5^\circ \times 5^\circ$   
159 gridded HadCRUT3v data set from Brohan et al. (2006).

## 160 4. Results

161 In this section we use the non-parametric additive method to reveal whether there exist  
162 non-linear relationships between 15 selected proxies and the mean Northern Hemisphere  
163 (NH) temperature. We also perform a multi-proxy temperature reconstruction based on a  
164 multiple linear regression, a non-parametric model and a semi-parametric model based on  
165 observed values of the proxies from AD 800. It is well known that low-frequency variability  
166 is underestimated by the multiple linear regression method (Christiansen et al. 2009) so we

167 stress that the reconstruction is performed to have a reference against which the non-linearity  
168 test can be viewed – the direct reconstruction is not presented here for other uses.

169 To calibrate our models, we use annual observations of the NH mean temperature data  
170 from the  $5^\circ \times 5^\circ$  gridded HadCRUT3v data set (Brohan et al. 2006) and the proxies from 1850  
171 through 1969, i.e. 120 observations, which is the calibration period. The NH temperature is  
172 centred over the calibration period. The proxies are centred and normalized with mean and  
173 standard deviation from the calibration period. This choice of calibration period ensures  
174 that there is no missing data in any of the proxies used, i.e. there is no use of zero-padding  
175 in the calibration of the models.

176 We first fit the classical multiple linear regression model to our data, also known as the  
177 direct global method, i.e. a regression with the NH temperature as the response and selected  
178 proxies as predictors. The results are given in Table 2. Here the estimated coefficients and  
179 corresponding p-value for the significance test is reported.

180 Next, we fit a non-parametric model, i.e. the model from equation (3), to the proxies  
181 selected above. We use the software program R, and the function `gam()` from the package  
182 `gam`, with default settings. That is, we use the identity link function and smoothing splines  
183 with 4 degrees of freedom, see e.g. Green and Silverman (1994), as the smoothing operator.  
184 An approximate F test is used to evaluate the significance of the non-linearity, to determine  
185 whether including the non-linear component of each smooth term in the model resulted  
186 in a significantly better fit than a linear relationship. Although the test statistics do not  
187 have exact or even asymptotic F distributions, Hastie and Tibshirani (1990) report that  
188 simulations show them to be useful approximations. The test results for non-linear effects  
189 are shown in Table 2, and shows two significantly non-linear proxies: the tree-ring width  
190 records Indigirka and Yamal. Figure 1 shows the fitted functions  $\hat{m}_k$  for each proxy, and  
191 some of them seem to have a non-linear relationship with the temperature, however, only  
192 two of them are found significantly non-linear. The dashed curves in each plot are the upper  
193 and lower point-wise twice-standard-error curves. The vertical marks along the bottom of

194 each graph illustrates the distribution of the values of the proxies. Further details regarding  
195 the non-linear effects test are found in Hastie and Tibshirani (1990).

We compare these two models, that is, the linear model and the non-parametric model, by a comparison test— see Table 3. The model comparison test is an approximate F-test, where the test statistic is

$$F = \frac{(RSS_1 - RSS_2)/RSS_2}{(DF_1 - DF_2)/DF_2},$$

196 where  $RSS_i$  is the residual sum of square for model  $i$  and  $DF_i$  is the approximate degrees  
197 of freedom for model  $i$ , see Hastie and Tibshirani (1990) for further details. We see that the  
198 reduction in RSS (Residuals Sum of Squares) for the non-parametric model compared with  
199 the linear model is not significant.

200 Since we have found that two of the 15 proxies relates non-linearly to temperature, we fit  
201 a semi-parametric model, i.e. the other 13 proxies are modelled linearly. We thus also reduce  
202 problems with over-fitting. The result for the semi-parametric model are seen in Table 2, i.e.  
203 the same non-linear effects test as above, are here performed for the two proxies. Again, the  
204 tests indicate that Indigirka and Yamal relates non-linear to temperature. Figure 2 shows  
205 the estimated linear and non-parametric curves.

206 We then compare the explanatory power of the linear and the semi-parametric model in  
207 a comparison test as above (Table 3). We find that the reduction in residual sum of squares  
208 *is* significant, indicating a better fit for the semi-parametric model over the linear model.

209 Finally we calculate the in-sample correlation (i.e. during the training interval) of the  
210 three methods and the NH temperature – see Table 4. These non-parametric methods give  
211 better in-sample fit than the linear method.

#### 212 *a. Robustness test*

213 The results regarding proxy non-linearity are interpretable in a relative sense – i.e. the  
214 non-linearity is present in a proxy with respect to the other proxies, and a natural question

215 is whether the list of non-linear proxies detected will be different if a different set of proxies  
216 is used. We test this in a very simple way, by removing one proxy at the time and re-  
217 calculating, noting which proxies test positive for non-linearity, then replacing the omitted  
218 proxy and going on to the next one. Thus, 15 tests are performed. Table 5 shows the  
219 results. Evidently there are two proxies that overwhelmingly test positive for non-linearity  
220 – Indigirka and Yamal, while a handful of others now and then also pass the test, depending  
221 on which proxy is omitted in the leave-one-out analysis. This gives a robust indication that  
222 the two proxies are special with respect to the other proxies as a group.

223 *b. Reconstruction uncertainties*

224 The reconstructions presented here are for comparative purposes only, but it is of interest  
225 to see how specific they are given the proxies used and whether they are individually different  
226 once a picture of their uncertainties are at hand. To obtain this information we perform  
227 bootstrapping with replacement on the observations. This is a well-known procedure which  
228 is occasionally coming into use in climate reconstruction circles (Till and Guiot 1990), and  
229 (Guiot 2005). We performed 1000 re-samples, performed the training and reconstruction  
230 each time, and compiled the reconstructions. In the end we had 1000 values for each year  
231 of the reconstruction period and found the lower 2.5% and the upper 97.5% percentile, and  
232 plot these, along with the reconstructions, in Figure 3.

233 The reconstructions are from 800–1990, in order to minimize problems with missing proxy  
234 data. However, for these reconstructions, zero-padding in the standardized proxy series are  
235 applied for missing observations in the Greenland composite (1974–1990), the Southern  
236 Sierra Nevada (1989–1990), Jämtland (888–908) and the Columbia Ice Field (800–949). As  
237 already mentioned, such a replacment of missing data may affect the reconstructions. In  
238 order to estimate such an effect, we have excluded the Columbia Ice Field tree-ring record  
239 in a separate analysis, as this is the proxy with the most missing data. That is, first, we  
240 have used the other 14 proxies and the mean NH temperature to calibrate three models; a

241 complete linear, a non-parametric and a semi-parametric model with Indigirka and Yamal as  
242 non-linear. Second, reconstructions are performed with these three models. A comparison  
243 between these reconstructions (with 14 proxies) and the reconstructions in Figure 3 (with  
244 15 proxies), indicates that the zero-padding has almost no effect on the linear and semi-  
245 parametric reconstructions, and just a small effect on the non-parametric reconstruction.  
246 That leads us to believe that even though zero-padding has been applied, the reconstructions  
247 in this case seem relatively robust.

248 We note that the intervals of uncertainty in Figure 3 are quite wide – so wide that here is  
249 hardly any difference between the reconstructions – but particularly wide in the case of the  
250 fully non-linear method. This is consistent with the possibility of over-fitting in the training  
251 interval. Note the much larger uncertainty, relatively, for the period near 1000 AD in the  
252 non-linear method.

## 253 5. Summary and Discussion

254 We have introduced testing for non-linearity in proxy-based temperature reconstructions,  
255 and a non-linear reconstruction method. An alternative to the approximate test for non-  
256 linear effects we have used here, is the generalized likelihood ratio (GLR) test developed by  
257 Fan and Jiang (2005). By using such a test it is possible to test a parametric null hypothesis,  
258 e.g. that the mean NH temperature is a linear model of the available proxies, against an  
259 alternative, not a linear model. We note that even though the asymptotic null distribution  
260 of the GLR statistic is available, in finite samples one would resort to conditional bootstrap  
261 methods to obtain the null distributions. We therefore defer application of this test for future  
262 research.

263 We show that these non-parametric methods do give better in-sample fit and further, we  
264 should expect that temperature reconstructions utilizing such methods also could be more  
265 correct. Of course, we should have in mind problems with over-fitting when using non-linear

266 models.

267 Furthermore, auto-correlation in the residuals can cause problems during reconstructions,  
268 such as biased level or biased uncertainty estimates. This problem may still be present, and  
269 in future research we aim to control for this fact by introducing auto-regressive error terms.  
270 Also, using a better selection procedure for relevant proxies for the non-parametric additive  
271 regression model would be of interest, but we note that inference in additive models is not  
272 well developed, see e.g. Fan and Jiang (2005).

273 Our analysis shows that relative to the present group of proxies two are non-linear relative  
274 to the NH mean temperature. The series are the tree-ring width records Indigirka and Yamal.  
275 One possible physical explanation for the non-linear behaviour of the Yamal tree-ring width  
276 record (Briffa (2000); Briffa et al. (2008)) may be the observed change in growth form of the  
277 trees in this region during the latter part of the 20th century caused by the warmer and thus  
278 more favourable growth conditions (Devi et al. 2008). Concerning the Indigirka tree-ring  
279 width record (Moberg et al. 2006) only the last 600 years of the record have actually been  
280 published in a specific article that critically evaluates the record (under the name Yakutia)  
281 (Hughes et al. 1999). It is interesting to note, however, that the correlation between the two  
282 records only amounts to 0.77 over the period AD 1400–1993 that they both have in common.

283 That only two out of 15 proxies tested positive for non-linearity is support for the general  
284 assumption that temperature proxies can be used in linear reconstruction attempts. Loehle  
285 (2009) discusses how non-linearity can come about in the context of trees' response functions;  
286 it is evident how the response can be not only non-linear but monotonic, but also bi-valued  
287 although then with oppositely-signed response rates. Our own result for Yamal seems to  
288 be of the former kind (Devi et al. 2008), while the result for Indigirka hints at the latter  
289 (or an even more complex case). We note that data non-linearities do not just arise due to  
290 direct causes affecting trees and their environment but can also be due to mundane things  
291 like data-collection issues (Esper et al. 2012). None of the non-linearities we have detected  
292 were found in other forms of temperature proxies.

293 To validate the models we would like to introduce cross-validation on data not used to  
294 train the model – a common and important procedure in the reconstruction field, but difficult  
295 to carry out due to the scarcity of independent data. ‘Cross-validation’ a.k.a. ‘leave one out’  
296 validation is a possibility but is complicated by the presence of auto-correlation in the data.

297 Finally, testing reconstruction methods on ensembles (Christiansen et al. 2009) of pseudo-  
298 data allows one to find not only the scatter of the reconstructions but also the bias. The use  
299 of the Bootstrap routine is, however, a very economic way of extracting information on the  
300 scatter of a reconstruction, from limited data. Further research should also include comparing  
301 non-parametric reconstruction methods with the more commonly used reconstruction meth-  
302 ods, such as the iterative regularized expectation maximization (RegEM) method (Dempster  
303 et al. (1977); Schneider (2001); Mann and Rutherford (2002); Smerdon and Kaplan (2007)),  
304 and also to examine their performance in an ensemble study as in (Christiansen et al. 2009).

305 *Acknowledgments.*

306 We wish to express our gratitude to the scholars that provided us with proxy data that  
307 are not available from public databases: Christophe Corona, CNRS/Aix-Marseille Universit,  
308 and Ivan A. Kalugin, Siberian Branch of the Russian Academy of Sciences. This work was  
309 supported by the Danish Climate Centre at the Danish Meteorological Institute.

## REFERENCES

- 312 Andreu-Hayles, L., R. D'Arrigo, K. J. Anchukaitis, P. S. A. Beck, D. Frank, and S. Goetz,  
313 2011: Varying boreal forest response to Arctic environmental change at the Firth River,  
314 Alaska. *Environmental Research Letters*, **6** (4), 049502.
- 315 Bird, B. W., M. B. Abbott, B. P. Finney, and B. Kutchko, 2009: A 2000 year varve-based  
316 climate record from the central Brooks Range, Alaska. *Journal of Paleolimnology*, **41**,  
317 25–41.
- 318 Birks, H., O. Heiri, H. Seppä, and A. Bjune, 2010: Strengths and Weaknesses of Quantitative  
319 Climate Reconstructions Based on Late-Quaternary Biological Proxies. *The Open Ecology*  
320 *Journal*, **3**, 68–110.
- 321 Birks, H. J. B., 1995: Quantitative palaeoenvironmental reconstructions. *Statistical Mod-*  
322 *elling of Quaternary Science Data (D. Maddy and J. S. Brew, eds.) Technical Guide 5*  
323 *161-254. Quaternary Research Association, Cambridge*, 161–254.
- 324 Bradley, R. S., 1999: *Paleoclimatology: Reconstructing climates of the Quaternary*. Academic  
325 Press.
- 326 Bradley, R. S. and P. D. Jones, (Eds.) , 1992: *Climatic variations over the last 500 years*,  
327 *in: Climate Since A.D. 1500*. Routledge: London, 649–665 pp.
- 328 Breiman, L. and J. Friedman, 1985: Estimating optimal transformations for multiple regres-  
329 sion and correlation (with discussion). *Journal of the American Statistical Association*,  
330 **80**, 580–619.
- 331 Briffa, K. R., 2000: Annual climate variability in the Holocene: interpreting the message of  
332 ancient trees. *Quat. Sci. Rev.*, **19**, 87.



333 Briffa, K. R., P. D. Jones, F. H. Schweingruber, S. G. Shiyatov, and E. R. Cook, 1995: Un-  
334 usual twentieth-century summer warmth in a 1,000-year temperature record from Siberia.  
335 *Nature*, **376**, 156–159.

336 Briffa, K. R., F. H. Schweingruber, P. D. Jones, T. J. Osborn, I. C. Harris, S. G. Shiyatov,  
337 E. A. Vaganov, and H. Grudd, 1998: Trees tell of past climates: but are they speaking  
338 less clearly today? *Philosophical Transactions of the Royal Society of London. Series B:*  
339 *Biological Sciences*, **353 (1365)**, 65–73.

340 Briffa, K. R., F. H. Schweingruber, P. D. Jones, T. J. Osborn, S. G. Shiyatov, and E. A.  
341 Vaganov, 1998: Reduced sensitivity of recent tree-growth to temperature at high northern  
342 latitudes. *Nature*, **391**, 678–682.

343 Briffa, K. R., V. V. Shishov, T. M. Melvin, E. A. Vaganov, H. Grudd, R. M. Hantemirov,  
344 M. Eronen, and M. M. Naurzbaev, 2008: Trends in recent temperature and radial tree  
345 growth spanning 2000 years across northwest Eurasia. *Phil. Trans. R. Soc. London B*,  
346 **363**, 2269–2282.

347 Brohan, P., J. Kennedy, I. Harris, S. Tett, and P. Jones, 2006: Uncertainty estimates in  
348 regional and global observed temperature changes: a new dataset from 1850. *J. Geophys.*  
349 *Res. D*, **111**.

350 Buja, A., T. Hastie, and R. Tibshirani, 1989: Linear smoothers and additive models (with  
351 discussion). *Annals of Statistics*, **17**, 453–555.

352 Büntgen, U., D. C. Frank, D. Nievergelt, and J. Esper, 2006: Summer temperature variations  
353 in the European Alps, A.D. 755–2004. *J. Climate*, **19**, 5606.

354 Büntgen, U., et al., 2011: 2500 Years of European Climate Variability and Human Suscep-  
355 tibility. *Science*, **331**, 578–582.

- 356 Carrer, M. and C. Urbinati, 2001: Spatial analysis of structural and tree-ring related pa-  
357 rameters in a timberline forest in the italian alps. *Journal of Vegetation Science*, **12** (5),  
358 643–652.
- 359 Christiansen, B., T. Schmith, and P. Thejll, 2009: A surrogate Ensemble Study of Climate  
360 Reconstruction Methods: Stochasticity and Robustness. *J. Climate*, **22**, 951–976.
- 361 Cook, T. L., R. S. Bradley, J. S. Stoner, and P. Francus, 2009: Five thousand years of  
362 sediment transfer in a high arctic watershed recorded in annually laminated sediments  
363 from Lower Murray Lake, Ellesmere Island, Nunavut, Canada. *J. Paleolimnology*, **41**,  
364 77–94.
- 365 Corona, C., J.-L. Edouard, G. Guibal, J. Guiot, S. Bernard, A. Thomas, and N. Denelle,  
366 2011: Long-term summer (AD 751–2008) temperature fluctuation in the French Alps based  
367 on tree-ring data. *Boreas*, **40**, 351–366.
- 368 D’Arrigo, R., R. Wilson, and G. Jacoby, 2006: On the long-term context for late twentieth  
369 century warming. *J. Geophys. Res.*, **111**, 10.1029/2005JD006352.
- 370 D’Arrigo, R., R. Wilson, B. Liepert, and P. Cherubini, 2008: On the ‘Divergence Problem’  
371 in Northern Forests: A review of the tree-ring evidence and possible causes. *Global and*  
372 *Planetary Change*, **60**, 289–305.
- 373 D’Arrigo, R., et al., 2001: 1738 years of Mongolian temperature variability inferred from a  
374 tree-ring width chronology of Siberian pine. *Geophys. Res. Lett.*, **28**, 543–546.
- 375 D’Arrigo, R. D., R. K. Kaufmann, N. Davi, G. C. Jacoby, C. Laskowski, R. B. Myneni, and  
376 P. Cherubini, 2004: Thresholds for warming-induced growth decline at elevational tree  
377 line in the Yukon Territory, Canada. *Global Biogeochemical Cycles*, **18**, GB3021.
- 378 Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977: Maximum likelihood from incomplete  
379 data via the EM algorithm. *J. R. Statist. Soc. B*, **39**, 1–38.

- 380 Devi, N., F. Hagedorn, P. Moiseev, H. Bugmann, S. Shiyatov, V. Mazepa, and A. Rigling,  
381 2008: Expanding forests and changing growth forms of Siberian larch at the Polar Urals  
382 treeline during the 20th century. *Change Biology*, **14**, 1581–1591.
- 383 Esper, J., U. Büntgen, M. Timonen, and D. C. Frank, 2012: Variability and extremes of north-  
384 ern Scandinavian summer temperatures over the past two millennia. *Global and Planetary*  
385 *Change*, in press.
- 386 Esper, J., E. R. Cook, and F. H. Schweingruber, 2002: Low-frequency signals in long tree-ring  
387 chronologies for reconstructing past temperature variability. *Science*, **295**, 2250.
- 388 Fan, J. and I. Gijbels, 1996: *Local Polynomial Modelling and Its Applications*. Chapman and  
389 Hall.
- 390 Fan, J. and J. Jiang, 2005: Nonparametric inferences for additive models. *Journal of the*  
391 *American Statistical Association*, **100**, 890–907.
- 392 Frank, D., J. Esper, E. Zorita, and R. Wilson, 2010: A noodle, hockey stick, and spaghetti  
393 plate: a perspective on high-resolution paleoclimatology. *Wiley Interdisciplinary Reviews:*  
394 *Climate Change*, **1**, 507–516.
- 395 Fritts, H. C., 1976: *Tree rings and climate*. Academic Press.
- 396 Graumlich, L. and L. Brubaker, 1986: Reconstruction of annual temperature (1590–1979)  
397 for Longmire, Washington, derived from tree rings. *Quat. Res.*, **25**, 223–234.
- 398 Graumlich, L. J., 1993: A 1000-year record of temperature and precipitation in the Sierra  
399 Nevada. *Quat. Res.*, **39**, 249–255.
- 400 Green, P. J. and B. W. Silverman, 1994: *Nonparametric regression and generalized linear*  
401 *models*. Chapman & Hall, London.
- 402 Groveman, B. and H. E. Landsberg, 1979: Simulated Northern Hemisphere Temperature  
403 Departures 1579–1880. *Geophys. Res. Lett.*, **6**, 767–769.

- 404 Grudd, H., 2008: Torneträsk tree-ring width and density AD 500–2004: a test of climatic  
405 sensitivity and a new 1500-year reconstruction of north Fennoscandian summers. *Climate*  
406 *Dyn.*, **31**, 843–857.
- 407 Guiot, J., 2005: Last-millennium summer-temperature variations in western Europe based  
408 on proxy data. *The Holocene*, **15**, 489–500.
- 409 Haslett, J., M. Whiley, S. Bhattacharya, M. Salter-Townshend, S. P. Wilson, J. R. M. Allen,  
410 B. Huntley, and F. J. G. Mitchell, 2006: Bayesian palaeoclimate reconstruction. *Journal*  
411 *of the Royal Statistical Society: Series A (Statistics in Society)*, **169 (3)**, 395–438.
- 412 Hastie, T., R. Tibshirani, and J. Friedman, 2009: *The elements of statistical learning: Data*  
413 *mining, inference and prediction. 2nd ed.* Springer, New York.
- 414 Hastie, T. J. and R. Tibshirani, 1990: *Generalized Additive Models*. Chapman and Hall,  
415 London.
- 416 Helama, S., M. M. Fauria, K. Mielikinen, M. Timonen, and M. Eronen, 2010: Sub-  
417 Milankovitch solar forcing of past climates: mid and late Holocene perspectives. *Geological*  
418 *Society of America Bulletin*, **122**, 1981–1988.
- 419 Hughes, M., E. Vaganov, S. Shiyatov, R. Touchan, and G. Funkhouser, 1999: Twentieth-  
420 century summer warmth in northern Yakutia in a 600-year context. *The Holocene*, **9 (5)**,  
421 629–634.
- 422 Jansen, E., et al., 2007: Paleoclimate. *Climate Change 2007: The Physical Science Basis*,  
423 Cambridge University Press, Cambridge, UK, chap. 6, URL <http://www.ipcc.ch/>.
- 424 Jones, P., et al., 2009: High-resolution palaeoclimatology of the last millennium: A review  
425 of current status and future prospects. *The Holocene*, **19**, 3–49.
- 426 Juckes, M. N., M. R. Allen, K. R. Briffa, J. Esper, G. C. Hegerl, A. Moberg, T. J. Osborn, and

- 427 S. L. Weber, 2007: Millennial temperature reconstruction intercomparison and evaluation.  
428 *Climate of the Past*, **3**, 591–609.
- 429 Kalugin, I. A., A. V. Daryin, and V. V. Babich, 2009: Reconstruction of annual air temper-  
430 atures for three thousand years in Altai region by lithological and geochemical Indicators  
431 in Teletskoe Lake sediments. *Dokl. Earth Sci.*, **426**, 681–684.
- 432 Korhola, A., 2002: Holocene temperature changes in northern Fennoscandia reconstructed  
433 from chironomids using Bayesian modelling. *Quat. Sci. Rev.*, **21**, 1841–1860.
- 434 Lamoureux, S. F. and R. S. Bradley, 1996: A late Holocene varved sediment record of  
435 environmental change from northern Ellesmere Island, Canada. *Journal of Paleolimnology*,  
436 **16**, 239–255.
- 437 Linderholm, H. W. and B. E. Gunnarson, 2005: Summer temperature variability in central  
438 Scandinavia during the last 3600 years. *Geogr. Ann.*, **87A**, 231–241.
- 439 Lindholm, M., R. Jalkanen, H. Salminen, T. Aalto, and M. Ogurtsov, 2011: The height-  
440 increment record of summer temperature extended over the last millennium in Fennoscan-  
441 dia. *The Holocene*, **21**, 319–326.
- 442 Ljungqvist, F., 2010: A new reconstruction of temperature variability in the extra-tropical  
443 Northern Hemisphere during the last two millennia. *Geografiska Annaler: Series A, Phys-  
444 ical Geography*, **92 (3)**, 339–351.
- 445 Ljungqvist, F. C., P. J. Krusic, G. Brattström, and H. S. Sundqvist, 2012: Northern Hemi-  
446 sphere temperature patterns in the last 12 centuries. *Climate of the Past*, **8 (5)**, in press.
- 447 Lloyd, A. H. and L. J. Graumlich, 1997: Holocene dynamics of treeline forests in the Sierra  
448 Nevada. *Ecology*, **78**, 1199–1210.
- 449 Loehle, C., 2009: A mathematical analysis of the divergence problem in dendroclimatology.  
450 *Climatic Change*, **94**, 233–245, 10.1007/s10584-008-9488-8.

- 451 Loso, M. G., 2009: Summer temperatures during the Medieval Warm Period and Little Ice  
452 Age inferred from varved proglacial lake sediments in southern Alaska. *J. Paleolimnology*,  
453 **41**, 117–128.
- 454 Luckman, B. H. and R. J. S. Wilson, 2005: Summer temperatures in the Canadian Rockies  
455 during the last millennium: a revised record. *Climate Dyn.*, **24**, 131.
- 456 Mann, M. E. and S. Rutherford, 2002: Climate reconstruction using pseudoproxies. *Geophys.*  
457 *Res. Lett.*, **29**, 1501.
- 458 Moberg, A., D. M. Sonechkin, K. Holmgren, N. M. Datsenko, W. Karlén, and S.-E. Lauritzen,  
459 2006: Corrigendum: Highly variable Northern Hemisphere temperatures reconstructed  
460 from low- and high-resolution proxy data. *Nature*, **439**, 1014.
- 461 Moore, J. J., K. A. Huguen, G. H. Miller, and J. T. Overpeck, 2001: Little Ice Age recorded  
462 in summer temperature reconstruction from varved sediments of Donard Lake, Baffin  
463 Island, Canada. *J. Paleolimnology*, **25**, 503–517.
- 464 NRC, 2006: *Surface temperature reconstructions for the last 2,000 years*. National Academies  
465 Press.
- 466 Popa, I. and Z. Kern, 2009: Long-term summer temperature reconstruction inferred from  
467 tree-ring records from the Eastern Carpathians. *Climate Dyn.*, **32**, 1107–1117.
- 468 Salzer, M. W. and K. F. Kipfmüller, 2005: Reconstructed temperature and precipitation  
469 on a millennial timescale from tree-rings in the Southern Colorado Plateau, U.S.A. *Clim.*  
470 *Change*, **70**, 465–487.
- 471 Schneider, T., 2001: Analysis of incomplete climate data: Estimation of mean values and  
472 covariance matrices and imputation of missing values. *J. Climate*, **14**, 853–871.
- 473 Smerdon, J. E. and A. Kaplan, 2007: Comments on Testing the fidelity of methods used in

- 474 proxy-based reconstructions of past climate: The role of the standardization interval. *J.*  
475 *Climate*, **20**, 5666–5670.
- 476 Tan, M., T. Liu, J. Hou, X. Qin, H. Zhang, and T. Li, 2003: Cyclic rapid warming on  
477 centennial-scale revealed by a 2650-year stalagmite record of warm season temperature.  
478 *Geophys. Res. Lett.*, **30** (12), 1617.
- 479 Teräsvirta, T., D. Tjøstheim, and C. Granger, 2010: *Modelling nonlinear economic time*  
480 *series*. Oxford University Press.
- 481 Thomas, E. K. and J. P. Briner, 2009: Climate of the past millennium inferred from varved  
482 proglacial lake sediments on northeast Baffin Island, Arctic Canada. *J. Paleolimnology*,  
483 **41**, 209–224.
- 484 Till, C. and J. Guiot, 1990: Reconstruction of precipitation in Morocco since 1100 A.D.  
485 Based on *Cedrus atlantica* tree-ring widths. *Quat. Res.*, **33**, 337–351.
- 486 Tingley, M., P. Craigmile, M. Haran, B. Li, E. Mannshardt-Shamseldin, and B. Rajaratnam,  
487 2012: Piecing together the past: Statistical insights into paleoclimatic reconstructions.  
488 *Quat. Sci. Rev.*, *in press*.
- 489 Tjøstheim, D. and B. Auestad, 1994: Nonparametric identification of nonlinear time series:  
490 Projection. *Journal of the American Statistical Association*, **89**, 1398–1409.
- 491 Toivonen, H., H. Mannila, A. Korhola, and H. Olander, 2001: Applying Bayesian statistics  
492 to organism-based environmental reconstruction. *Ecological Applications*, **11**, 618–630.
- 493 Treydte, K. S., D. C. Frank, M. Saurer, G. Helle, G. H. Schleser, and J. Esper, 2009: Impact  
494 of climate and CO<sub>2</sub> on a millennium-long tree-ring carbon isotope record. *Geochimica et*  
495 *Cosmochimica Acta*, **73**, 4635–4647.
- 496 Vasko, K., H. T. Toivonen, and A. Korhola, 2000: A Bayesian multinomial Gaussian response

497 model for organism-based environmental reconstruction. *Journal of Paleolimnology*, **24**,  
498 243–250.

499 Vinther, B. M., P. D. Jones, K. R. Briffa, H. B. Clausen, K. K. Andersen, D. Dahl-Jensen,  
500 and S. J. Johnsen, 2010: Climatic signals in multiple highly resolved stable isotope records  
501 from Greenland. *Quat. Sci. Rev.*, **29**, 522–538.

502 Vinther, B. M., et al., 2008: Synchronizing ice cores from the Renland and Agassiz ice caps  
503 to the Greenland Ice Core Chronology. *J. Geophys. Res. (Atmospheres)*, **113**, D08115.

504 Wilson, R., et al., 2007: A matter of divergence: Tracking recent warming at hemispheric  
505 scales using tree ring data. *Journal of Geophysical Research D (Atmospheres)*, **112** (11),  
506 D17103.

## 507 APPENDIX

In this appendix we document the application of the additive non-parametric model to an artificial case, i.e. we specify a non-linear model for some of the proxies and then check whether the method detects this model. For simplicity, we assume an additive model consisting of just three standardized proxies,

$$Y = \sum_{j=1}^3 m_j(X_j) + e, \quad (\text{A1})$$

508 where the three proxies  $X_1$ ,  $X_2$  and  $X_3$  are Southern Colorado Plateau (as its coefficient  
509 is found significant in the linear model), Indigirka and Yamal (as these are found signifi-  
510 cantly non-linear), respectively. We further assume that the regression function for South-  
511 ern Colorado Plateau is linear (i.e.  $m_1 = 0.06 \cdot X_1$ ), the regression function for Indigirka  
512 is non-linear ( $m_2 = 0.1 \cdot \cos^2(X_2)$ ) and the regression function for Yamal also non-linear  
513 ( $m_3 = 0.02 \cdot X_3^2 - 0.02 \cdot X_3^3$ ). The functions are chosen to resemble the calibrated functions  
514 from the semiparametric model.



515 By using the calibration data from 1850–1969 for these proxies, we generate the artificial  
516 'temperature'  $Y$  from the model above. In addition, we add noise ( $e$ ), which is iid from a  
517 normal distribution with zero mean and standard deviation equal to 0.15, to the artificial  
518 'temperature'. We then fit a non-parametric model to this data. The result from the  
519 estimation is shown in Table 6 and Figure 4. The test correctly detects  $m_2$  and  $m_3$  to be  
520 non-linear, and rejects  $m_1$  as non-linear. In the figure the estimated functions are plotted as  
521 thin solid curves, and the true underlying functions are plotted as thick solid curves. Clearly,  
522 the method is capable of detecting the true underlying functions, but not perfectly, mainly  
523 due to the added noise. In the bottom right plot, the points are the artificial 'temperature'  
524  $Y$  with noise. Clearly, they deviate quite a bit from the the solid curve, which is the artificial  
525 'temperature'  $Y$  without noise. The dashed curve is the reconstruction obtained from the  
526 calibrated non-parametric model. The reconstruction is quite close to the true curve, and  
527 we conclude that the method works.

## 528 List of Tables

- 529 1 Proxies.  $R1$ – $R5$  are Pearson correlation coefficients.  $R1$  and  $R2$  are the correlations  
530 to NH mean T – with and without linear trends, for AD 1850–1969.  $R3$  and  $R4$  are  
531 the same, but correlated against the local grid-point T and for the years 1870–1969.  
532  $R5$  is by the original author, when available. A ('-') indicates that no information  
533 is available.  $\Delta$  is dating uncertainty in years. Note for Polar Urals: The  $R5$  is  
534 from the Briffa et al. (1995) version of Polar Urals for the May–Sep season AD  
535 1882–1980. Note for Southern Colorado Plateau: The authors stated that  $R5$  was  
536 calculated from the "mean max annual T". This is difficult to find – instead we  
537 used the mean of the monthly-mean Jun–Aug T, although these are based on daily-  
538 mean values, not daily max values. 4 of the high-latitude series are evaluated,  
539 ( $R3$  and  $R4$ ) against averages of grid-point T covering an area: Renland  $-30^{\circ}\text{W}$ –  
540  $20^{\circ}\text{W}/65^{\circ}\text{N}$ – $75^{\circ}\text{N}$ ; Avam-Taimyr  $-92^{\circ}\text{E}$ – $102^{\circ}\text{E}/69^{\circ}\text{N}$ – $73^{\circ}\text{N}$ ; Greenland composite  
541  $-55^{\circ}\text{W}$ – $33\text{W}/60^{\circ}\text{N}$ – $75^{\circ}\text{N}$ ; Gulf of Alaska  $-153^{\circ}\text{W}$ – $131^{\circ}\text{W}/55^{\circ}\text{N}$ – $62^{\circ}\text{N}$ . 27
- 542 2 Calibration results. The first two columns give the estimated coefficients and  
543 corresponding p-value of a standard significance test for the linear model.  
544 The third column gives the p-value for the non-linearity test of the proxies  
545 in the non-parametric model. The fourth column gives the p-value for the  
546 non-linearity test of the proxies in the semi-parametric model. \*\*\*, \*\* and  
547 \* indicate significance on the 1%, 5% and 10% level, respectively. 28
- 548 3 Model comparison tests. The first column reports the degrees of freedom,  
549 second column reports the Residual Sum of Squares and the last column  
550 reports the results from the two tests. \*\*\* denote significance on the 1% level. 29
- 551 4 In-sample correlation between the reconstructions and measured NH temper-  
552 ature (1850–1969) 30

553	5	Results of testing the robustness of the non-linearity test, based on 'leave-	
554		one-out' sampling. As there are 15 proxies we can choose 15 different sets	
555		of 14 proxies each and test for non-linearity and see whether a particular	
556		proxy tests positive for non-linearity. We count the number of times a proxy	
557		is found significantly non-linear (at 10 % level) in the 15 possible calibrated	
558		non-parametric models. For instance Yamal was found to be non-linear 10	
559		out of 15 times while Teletskoe Lake was only found non-linear twice out of	
560		the 15 tests.	31
561	6	Results from the approximate non-linear test for the fitted non-parametric	
562		model for the artificial case.	32

TABLE 1. Proxies.  $R1$ – $R5$  are Pearson correlation coefficients.  $R1$  and  $R2$  are the correlations to NH mean T – with and without linear trends, for AD 1850–1969.  $R3$  and  $R4$  are the same, but correlated against the local grid-point T and for the years 1870–1969.  $R5$  is by the original author, when available. A (‘-’) indicates that no information is available.  $\Delta$  is dating uncertainty in years. Note for Polar Urals: The  $R5$  is from the Briffa et al. (1995) version of Polar Urals for the May–Sep season AD 1882–1980. Note for Southern Colorado Plateau: The authors stated that  $R5$  was calculated from the ”mean max annual T”. This is difficult to find – instead we used the mean of the monthly-mean Jun–Aug T, although these are based on daily-mean values, not daily max values. 4 of the high-latitude series are evaluated, ( $R3$  and  $R4$ ) against averages of grid-point T covering an area: Renland  $-30^\circ\text{W}$ – $20^\circ\text{W}/65^\circ\text{N}$ – $75^\circ\text{N}$ ; Avam-Taimyr  $-92^\circ\text{E}$ – $102^\circ\text{E}/69^\circ\text{N}$ – $73^\circ\text{N}$ ; Greenland composite  $-55^\circ\text{W}$ – $33^\circ\text{W}/60^\circ\text{N}$ – $75^\circ\text{N}$ ; Gulf of Alaska  $-153^\circ\text{W}$ – $131^\circ\text{W}/55^\circ\text{N}$ – $62^\circ\text{N}$ .

Proxy record	Season	Long	Lat	Extent	R1	R2	R3	R4	R5	$\Delta$	Proxy type	Reference
Agassiz Ice Cap	Annual	-73.10	80.70	800–1972	<b>0.37</b>	<b>0.25</b>	0.13	-0.01	-	$\pm 0$	Ice-core $\delta^{18}\text{O}$	Vinther et al. (2008)
Avam-Taimyr	July	93.00	70.00	800–2003	<b>0.49</b>	<b>0.30</b>	<b>0.29</b>	<b>0.25</b>	0.39	$\pm 0$	Tree-ring width	Briffa et al. (2008)
Big Round Lake	July to September	-68.50	69.83	971–2003	<b>0.41</b>	<b>0.21</b>	0.03	-0.06	0.46	$\pm 1 - 20$	Varved lake sediment	Thomas and Briner (2009)
Blue Lake	June to August	-150.46	68.08	800–2005	-0.02	-0.15	-0.02	0.01	0.56	$\pm 12$	Varved lake sediment	Bird et al. (2009)
Boreal/Upper Wright	June to August	-118.46	36.54	800–1992	<b>0.36</b>	0.08	0.06	0.06	-	$\pm 0$	Tree-ring width	Lloyd and Graumlich (1997)
Central Europe	June to August	8.00	46.30	800–2003	0.12	0.08	<b>0.53</b>	<b>0.54</b>	0.72	$\pm 0$	Tree-ring width	Büntgen et al. (2011)
Columbia Ice Field	May to August	-117.15	52.15	950–1998	<b>0.18</b>	0.03	<b>0.40</b>	<b>0.38</b>	0.73	$\pm 0$	Tree-ring density	Luckman and Wilson (2005)
Donard Lake	June to August	-61.35	66.66	800–1992	-0.09	-0.01	-0.11	-0.08	0.57	$\pm 1 - 20$	Varved lake sediment	Moore et al. (2001)
Eastern Carpathians	July to August	25.10	47.10	994–2005	0.05	0.11	<b>0.34</b>	<b>0.33</b>	0.42	$\pm 0$	Tree-ring width	Popa and Kern (2009)
Finnish Lapland	June to August	25.00	69.00	800–2005	<b>0.35</b>	<b>0.25</b>	<b>0.54</b>	<b>0.50</b>	0.64	$\pm 0$	Tree-ring width	Helama et al. (2010)
French Alps	June to August	7.00	45.50	800–2008	<b>0.31</b>	<b>0.18</b>	<b>0.59</b>	<b>0.59</b>	0.39	$\pm 0$	Tree-ring width	Corona et al. (2011)
Greenland composite	Annual	-40.00	70.00	800–1973	<b>0.23</b>	0.13	<b>0.31</b>	<b>0.21</b>	0.56	$\pm 0$	Stacked ice-core $\delta^{18}\text{O}$	Vinther et al. (2010)
Gulf of Alaska	January to September	-145.00	60.00	800–1999	<b>0.22</b>	0.00	<b>0.36</b>	<b>0.32</b>	0.48	$\pm 0$	Tree-ring width	D’Arrigo et al. (2006)
Iceberg Lake	May and June	-142.95	60.78	800–1998	0.17	0.11	-0.13	-0.16	0.23	$\pm 32$	Varved lake sediment	Loso (2009)
Indigirka	June to August	148.15	70.53	800–1993	<b>0.32</b>	<b>0.23</b>	<b>0.31</b>	<b>0.28</b>	-	$\pm 0$	Tree-ring width	Moberg et al. (2006)
Jämtland	June to August	13.30	63.10	800–2002	<b>0.37</b>	0.16	<b>0.48</b>	<b>0.50</b>	0.63	$\pm 0$	Tree-ring width	Linderholm and Gunnarson (2005)
Karakorum Mountains	June and July	74.99	36.37	828–1998	<b>0.29</b>	0.08	0.12	<b>0.29</b>	0.48	$\pm 0$	$\delta^{13}\text{C}$ tree-ring	Treydte et al. (2009)
Laanila	June to August	27.30	68.50	800–2007	0.09	0.17	<b>0.53</b>	<b>0.56</b>	0.56	$\pm 0$	Tree-ring height-increment	Lindholm et al. (2011)
Lake C2	June to August	-77.54	82.47	800–1987	0.14	0.04	0.09	0.06	-	$\pm 57$	Varved lake sediment	Lamoureaux and Bradley (1996)
Lower Murray Lake	July	-69.32	81.21	800–1969	<b>0.27</b>	<b>0.23</b>	0.11	0.10	0.78	$\pm 16$	Varved lake sediment	Cook et al. (2009)
Polar Urals	May to September	65.75	66.83	800–1990	<b>0.52</b>	<b>0.38</b>	<b>0.40</b>	<b>0.30</b>	-	$\pm 0$	Tree-ring density	Esper et al. (2002)
Renland	Annual	-26.70	71.30	800–1986	0.09	-0.04	<b>0.39</b>	<b>0.27</b>	-	$\pm 2 - 20$	Ice-core $\delta^{18}\text{O}$	Vinther et al. (2008)
ShiHua Cave	May to August	116.23	39.54	800–1985	<b>0.43</b>	0.12	0.06	0.05	0.65	$\pm 5$	Speleothem layer thickness	Tan et al. (2003)
Sol Dav	April to October	98.93	48.30	800–1999	<b>0.39</b>	-0.28	-0.06	0.02	0.58	$\pm 0$	Tree-ring width	D’Arrigo et al. (2001)
Southern Sierra Nevada	June to August	-118.90	36.90	800–1988	<b>0.20</b>	<b>0.20</b>	<b>0.21</b>	<b>0.21</b>	-	$\pm 0$	Tree-ring width	Graumlich (1993)
Southern Colorado Plateau	June to August	-111.40	35.20	800–1996	<b>0.43</b>	<b>0.27</b>	<b>0.23</b>	0.19	0.68	$\pm 0$	Tree-ring width	Salzer and Kipfmüller (2005)
Teletskoe Lake	Annual	87.61	51.76	800–2002	<b>0.44</b>	0.06	<b>0.31</b>	0.17	-	$\pm 1$	Varved lake sediment	Kalugin et al. (2009)
The Alps	June to September	8.00	46.30	800–2004	<b>0.24</b>	0.14	<b>0.81</b>	<b>0.80</b>	0.69	$\pm 0$	Tree-ring density	Büntgen et al. (2006)
Torneträsk	April to August	19.80	68.31	800–2004	<b>0.38</b>	<b>0.33</b>	<b>0.81</b>	<b>0.79</b>	0.79	$\pm 0$	Tree-ring density	Grudd (2008)
Yamal	June to July	69.17	66.92	800–1996	<b>0.29</b>	0.03	<b>0.61</b>	<b>0.59</b>	0.56	$\pm 0$	Tree-ring width	Briffa (2000)

TABLE 2. Calibration results. The first two columns give the estimated coefficients and corresponding p-value of a standard significance test for the linear model. The third column gives the p-value for the non-linearity test of the proxies in the non-parametric model. The fourth column gives the p-value for the non-linearity test of the proxies in the semi-parametric model. \* \* \*, \*\* and \* indicate significance on the 1%, 5% and 10% level, respectively.

Proxy	Linear model		Non-parametric model	Semi-parametric model
	Coefficients	p-value	Non-linear p-value	Non-linear p-value
Avam-Taimyr	0.035	0.026**	0.543	-
Columbia Ice Field	-0.002	0.873	0.466	-
Finnish Lapland	0.037	0.014**	0.473	-
French Alps	0.021	0.131	0.121	-
Greenland composite	0.002	0.878	0.237	-
Gulf of Alaska	0.021	0.133	0.560	-
Indigirka	0.019	0.127	0.084*	0.027**
Jämtland	0.013	0.935	0.647	-
Polar Urals	0.060	0.005***	0.243	-
Southern Sierra Nevada	0.042	0.002***	0.591	-
Southern Colorado Plateau	0.057	0***	0.489	-
Teletskoe Lake	0.002	0.914	0.193	-
The Alps	0.012	0.375	0.610	-
Torneträsk	0.006	0.701	0.358	-
Yamal	-0.033	0.081*	0.079*	0.043**

TABLE 3. Model comparison tests. The first column reports the degrees of freedom, second column reports the Residual Sum of Squares and the last column reports the results from the two tests. \* \* \* denote significance on the 1% level.

Model	Df.	RSS	F-statistic	p-value
Linear	104	1.634	-	-
Non-parametric	59	0.881	1.120	0.339
Linear	104	1.634	-	-
Semi-parametric	98	1.380	3.012	0.009***

TABLE 4. In-sample correlation between the reconstructions and measured NH temperature (1850–1969)

Model	Correlation with NH temperature
Linear model	0.80
Semi-parametric model	0.86
Non-parametric model	0.92

Proxy	Significantly non-linear
French Alps	2
Greenland composite	1
Indigirka	13
Polar Urals	2
Telteskoe Lake	2
Yamal	10
Others	0

TABLE 5. Results of testing the robustness of the non-linearity test, based on 'leave-one-out' sampling. As there are 15 proxies we can choose 15 different sets of 14 proxies each and test for non-linearity and see whether a particular proxy tests positive for non-linearity. We count the number of times a proxy is found significantly non-linear (at 10 % level) in the 15 possible calibrated non-parametric models. For instance Yamal was found to be non-linear 10 out of 15 times while Teletskoe Lake was only found non-linear twice out of the 15 tests.



TABLE 6. Results from the approximate non-linear test for the fitted non-parametric model for the artificial case.

Proxy	Non-linear p-value
Southern Colorado Plateau	0.599
Indigirka	0.001***
Yamal	0.051**

## 563 List of Figures

- 564 1 The estimated functions (i.e.  $\hat{m}$ ) for each proxy for the non-parametric model.  
565 The dashed lines in the plots are twice the pointwise standard error bounds.  
566 The vertical marks along the bottom illustrate the distribution of the proxies. 34
- 567 2 The estimated functions (i.e.  $\hat{m}$ ) for each proxy for the semi-parametric model.  
568 The dashed lines in the plots are twice the pointwise standard error bounds.  
569 The vertical marks along the bottom illustrate the distribution of the proxies. 35
- 570 3 The reconstructed rolling 10-year mean NH temperatures from the linear  
571 method (top plot), the non-linear method (middle plot) and the semi-parametric  
572 method (bottom plot). The gray curves are the the 2.5% lower and upper  
573 97.5% percentiles, see text for details on the bootstrapping performed to gen-  
574 erate these. 36
- 575 4 Top left plot to bottom left plot: The estimated functions (dashed curves) for  
576 the fitted non-parametric model and the true functions (solid curves) for the  
577 artificial case. The bottom right plot: points are the artificial 'temperature'  
578 Y with noise, the solid curve is the artificial 'temperature' Y without noise,  
579 and the dashed curve is the reconstruction obtained from the non-parametric  
580 model. 37

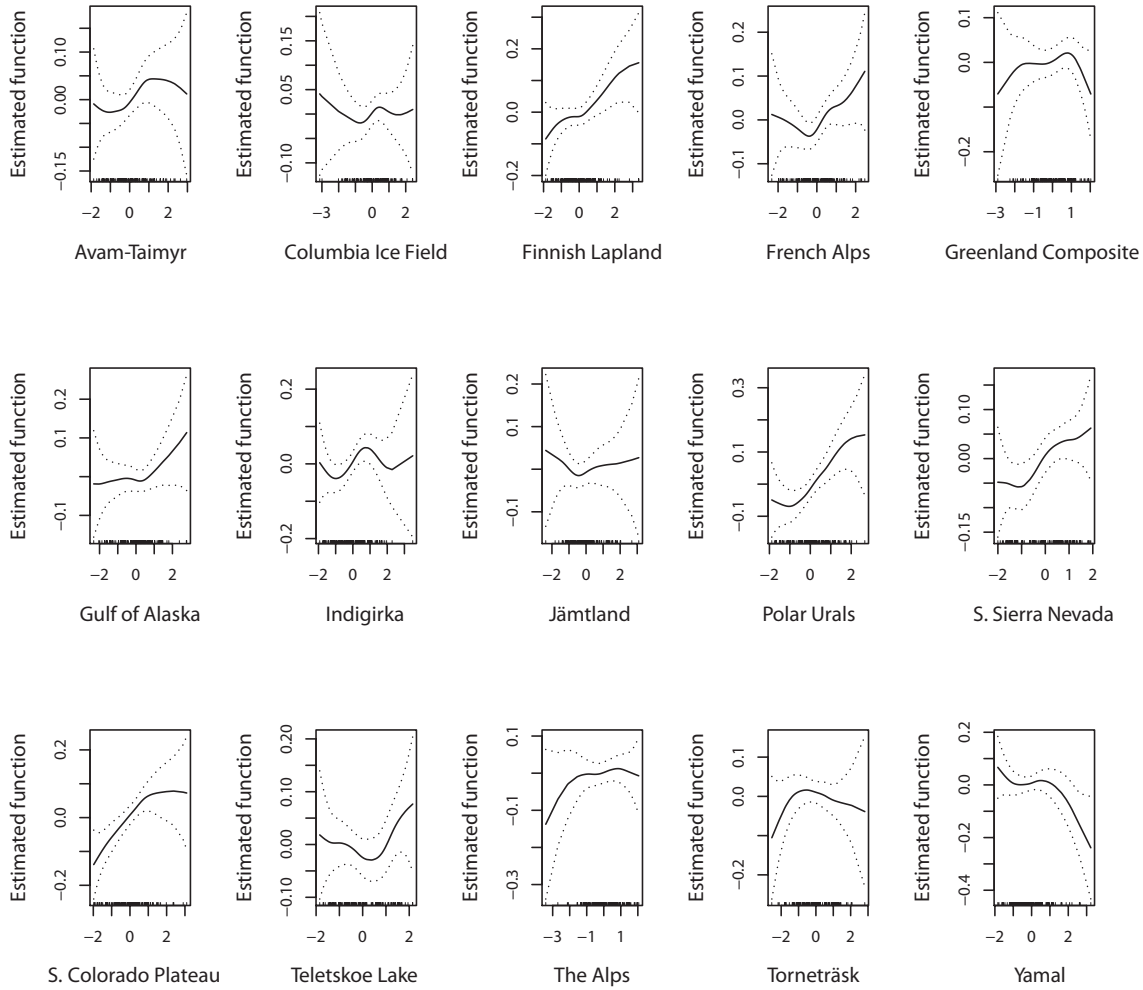


FIG. 1. The estimated functions (i.e.  $\hat{m}$ ) for each proxy for the non-parametric model. The dashed lines in the plots are twice the pointwise standard error bounds. The vertical marks along the bottom illustrate the distribution of the proxies.

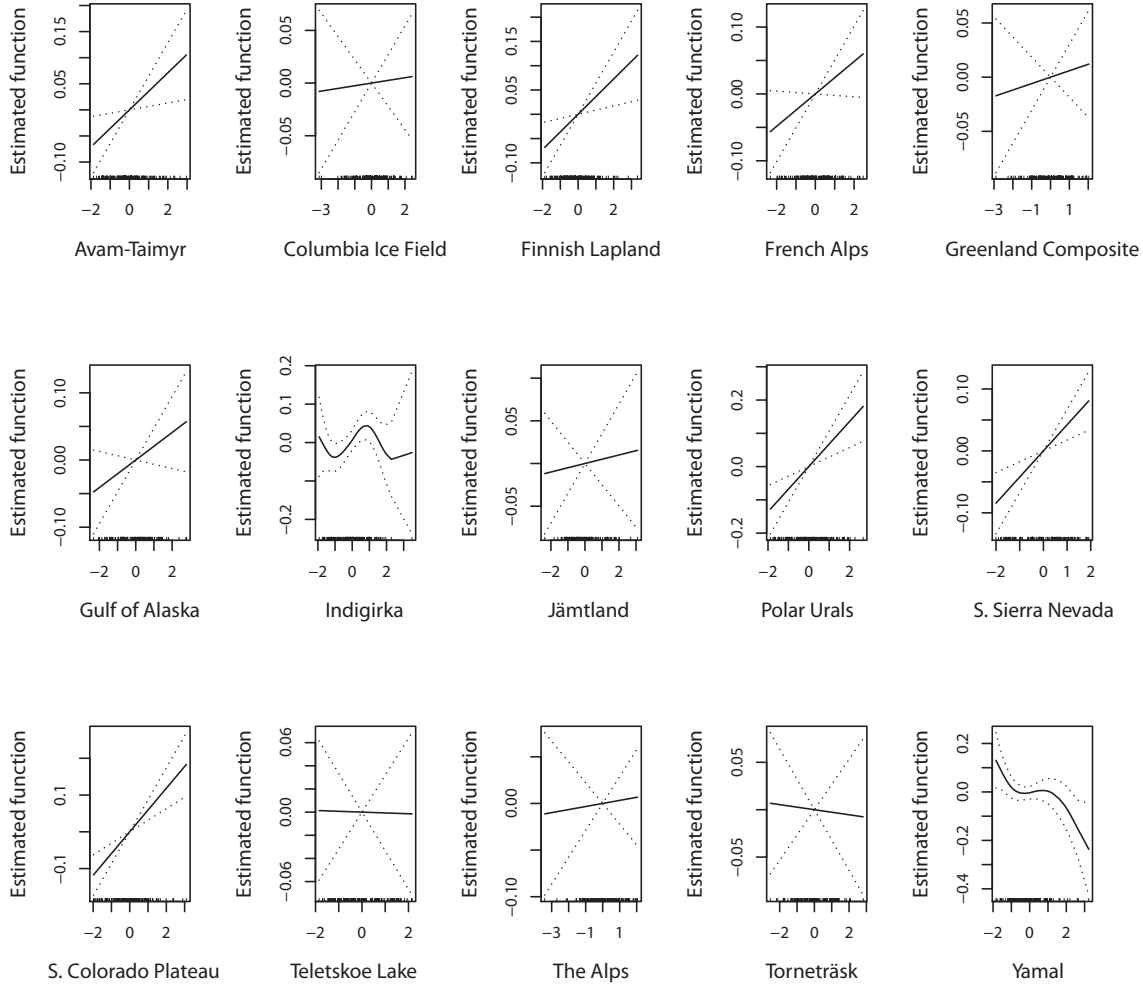


FIG. 2. The estimated functions (i.e.  $\hat{m}$ ) for each proxy for the semi-parametric model. The dashed lines in the plots are twice the pointwise standard error bounds. The vertical marks along the bottom illustrate the distribution of the proxies.

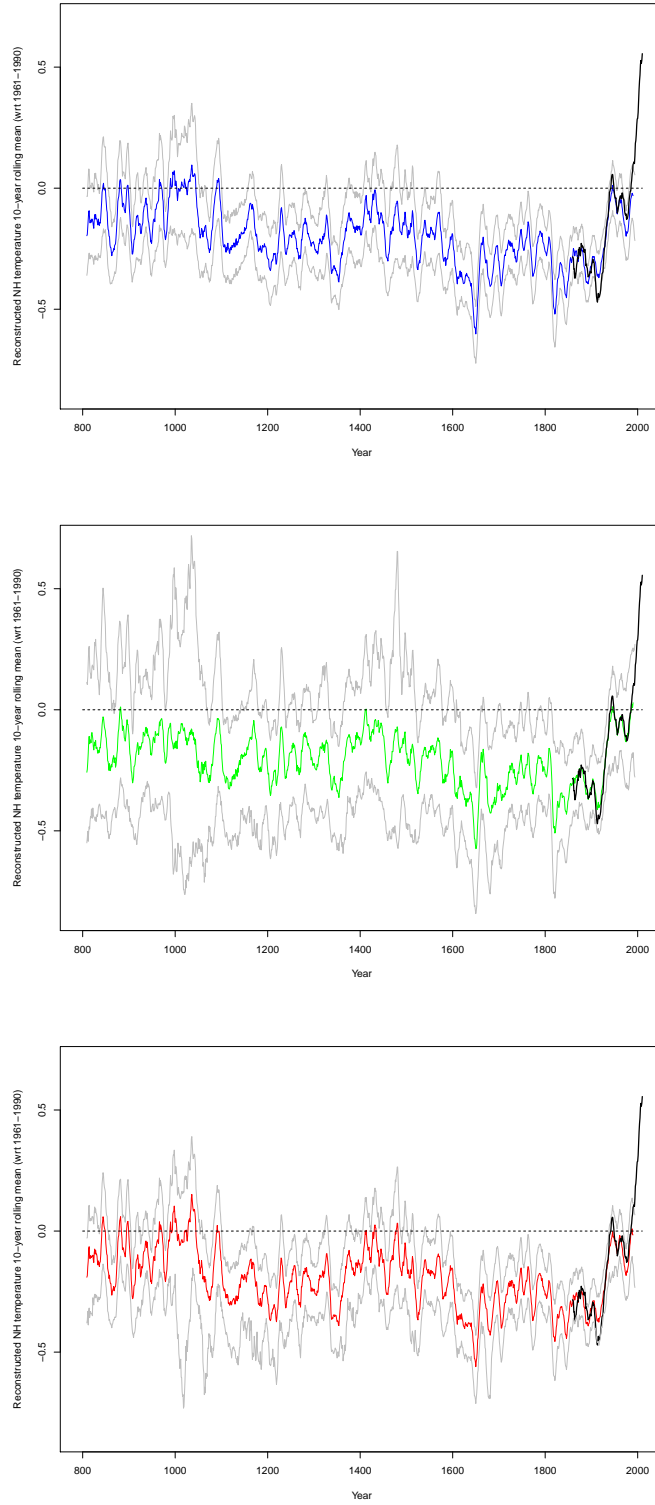


FIG. 3. The reconstructed rolling 10-year mean NH temperatures from the linear method (top plot), the non-linear method (middle plot) and the semi-parametric method (bottom plot). The gray curves are the the 2.5% lower and upper 97.5% percentiles, see text for details on the bootstrapping performed to generate these.

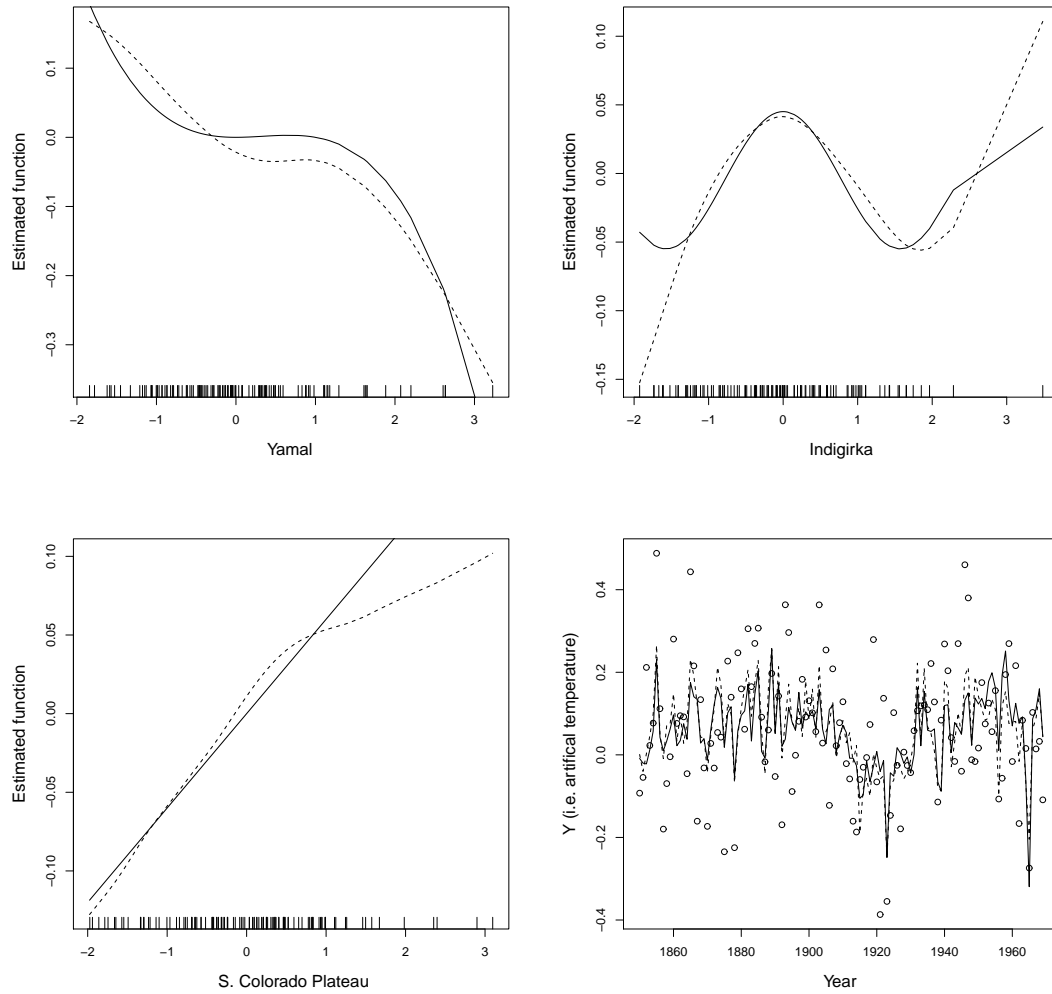


FIG. 4. Top left plot to bottom left plot: The estimated functions (dashed curves) for the fitted non-parametric model and the true functions (solid curves) for the artificial case. The bottom right plot: points are the artificial 'temperature'  $Y$  with noise, the solid curve is the artificial 'temperature'  $Y$  without noise, and the dashed curve is the reconstruction obtained from the non-parametric model.