

Sucharita Ghosh

Regression-based age estimation of a stratigraphic isotope sequence in Switzerland

Received: 2 September 2005 / Accepted: 8 March 2006 / Published online: 27 July 2006
© Springer-Verlag 2006

Abstract Multi-proxy data such as pollen percentages, aquatic biota, and stable isotope ratios in lake sediments in conjunction with climate transfer functions can be used to reconstruct past climate. This has been the subject of some recent projects. Often, stable-isotope ratios of oxygen are used as an independent proxy for climate. Past climate so reconstructed can in turn be used to assess the response of past vegetation to climatic oscillations, for instance near the epoch boundaries. One important intermediate step is to establish the age of the stratigraphic sequence. Strong similarities between the $\delta^{18}\text{O}$ records from European lake sediments and the Greenland ice cores are of interest. The Greenland ice-core project (GRIP) provided $\delta^{18}\text{O}$ data that were dated using an ice-flow model. Although the physical laws behind the isotope series from ice and lake sediment are different, statistical methods can be used to match the two series. In this paper, a regression-based approach is suggested for series matching. The method is illustrated by analyzing a series of $\delta^{18}\text{O}$ records covering the Late-glacial interstadial (ca. 15,000–13,000 years B.P. [1950]) from Gerzensee, Switzerland. Regression methods for age-depth modelling have also been recommended by other authors. Such an approach leads to reproducible and statistically founded age estimates and can easily be updated to include new data and information as needed. In this paper, the modelling step is preceded by identifying comparable sub-sections in the two isotope series by empirically matching the local minima and maxima in the smoothed isotope values; regression models are then used locally for each sub-section. This accommodates for local differences in the parameters. Variations in the final age estimates caused by different choices of the smoothing (bandwidth) parameters used in the intermediate nonparametric smoothing step are also taken into account in this algorithm.

Keywords Age-depth modelling · Rapid warming · Greenland ice core project (GRIP) · Regression · Smoothing

Introduction

Stable oxygen isotope ratios are widely used for identifying climate signals from stratigraphies (Siegenthaler et al. 1984). The oxygen isotope values are then used as proxies for temperature (Lotter et al. 2000) and correlated with other biological responses found in stratigraphic samples such as pollen. This has made major contributions towards understanding how vegetation may have changed in the past in response to climate (also see Birks 1995). For such a study however, the sediment cores need to be dated. However, if no high-resolution age estimates are available (Gulliksen et al. 1998), one may synchronize separate sequences of measurements, one of which must have been dated using a reliable method. In the literature, this approach is known as wiggle-matching and a wide variety of methods can be found for matching stratigraphic sequences (Schwander et al. 2000). One option is to use regression. Regression methods are simple to use and can be programmed by virtually any statistical software. Birks and Heegard (2003) give an overview, and other references can be found in the literature (Heegard et al. 2005). In this paper, a method is proposed that combines local polynomial smoothing with piecewise parametric regression. Various open questions remain, such as optimal calculation of confidence intervals and other issues concerning choice of the smoothing (bandwidth) parameters. A detailed comparison of the new method with other statistical approaches (Bennett 1994; Telford et al. 2004) will be pursued elsewhere.

The isotope series to be dated is from lake sediments in Gerzensee, Switzerland. Gerzensee is located in the Swiss plateau at an elevation of about 600 m a.s.l., in a glacially-formed landscape (Eicher 1979). This isotope series is synchronous with a part of the Greenland Ice Core Project (GRIP) $\delta^{18}\text{O}$ series, covering the Late-glacial interstadial

Communicated by Pim van der Knaap

S. Ghosh (✉)
Swiss Federal Research Institute WSL,
CH 8903 Birmensdorf, Switzerland
e-mail: rita.ghosh@wsl.ch

(ca. 15,000-13,000 yrs B.P. [1950]; Johnsen et al. 1997). The entire GRIP isotope series was dated by correlating with the DYE-3 core using volcanic markers, layer counting and an ice-flow model (*ibid.*). DYE-3 was the primary core of the GISP initiative (Greenland Ice Sheet Project).

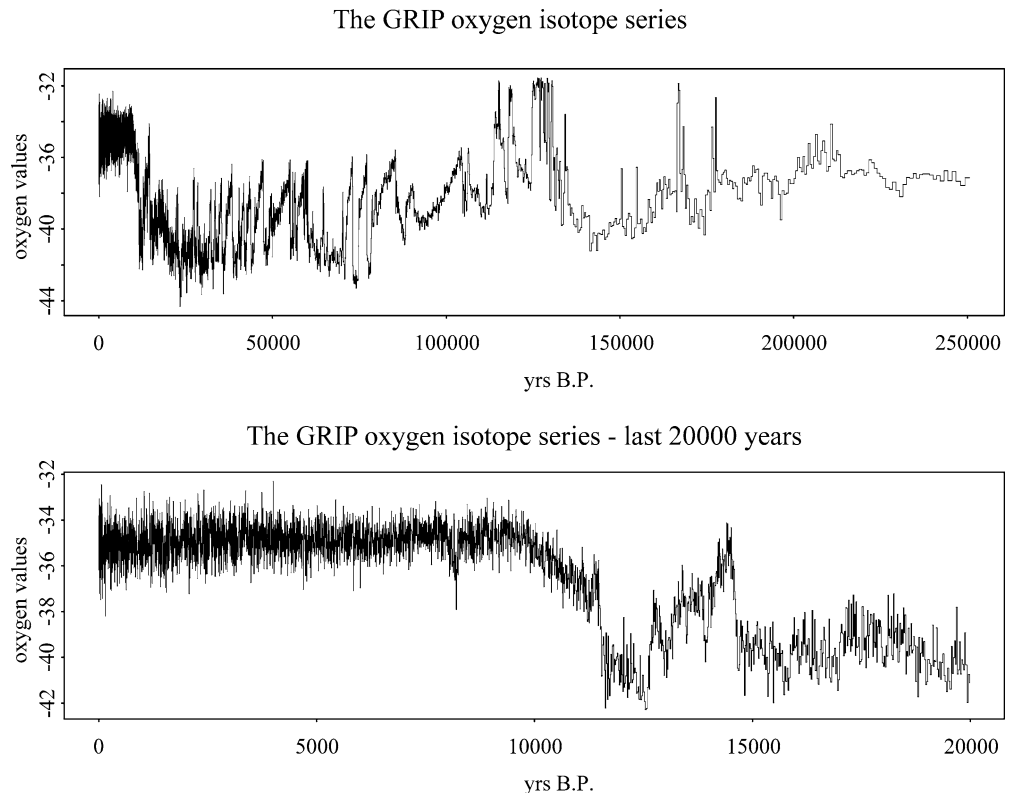
The basic principle behind the proposed regression model is that the functional relationship between age and depth of the Greenland series can be carried over to the Swiss series. The task of modelling is to identify that functional relation. Since polynomials are known to be good approximates for many mathematical functions, a polynomial regression model is used.

Data

The algorithm for estimating age by synchronizing two series is defined in the next section. Its application will be illustrated by the following data.

One of the two series is a portion taken from the 3 km deep oxygen isotope series from Greenland. These oxygen isotope values are plotted in Fig. 1. The series consists of 55 cm averages of $\delta^{18}\text{O}$ values in ice from the GRIP core (bag samples, 0.55 m = one bag). The complete GRIP series has been dated using an ice-flow model (Johnsen et al. 1997). The results of that analysis and the raw GRIP oxygen isotope ratio data can be downloaded from the internet site of the Glaciology Group of the University of Copenhagen; see <http://www.glaciology.gfy.ku.dk/data/grip18o.txt>. The entire GRIP core covers an estimated time span of more than 250,600 years (up to 2983 m below surface). It should

Fig. 1 GRIP $\delta^{18}\text{O}$ values (per mille) plotted against their estimated age (yrs B.P.; Johnsen et al. 1997): the whole series and the last 20000 years



be noted that there have also been other publications on depth-age models for the GRIP core. For stratigraphic age estimates see Hammer et al. (1997).

The second series is composed of $\delta^{18}\text{O}$ measurements in 5–20 mm sediment samples from a single lake sediment core, at a depth of 270 cm to 414 cm below surface, from Gerzensee, Switzerland. The raw data are plotted in Fig. 2. For references to the data, see Schwander et al. (2000). We wish to establish the age of this series.

The Gerzensee series is known to cover a much shorter time period than the GRIP series. Therefore, to start our algorithm, we let ourselves be guided by the findings of Jakob Schwander (personal communication) and select a

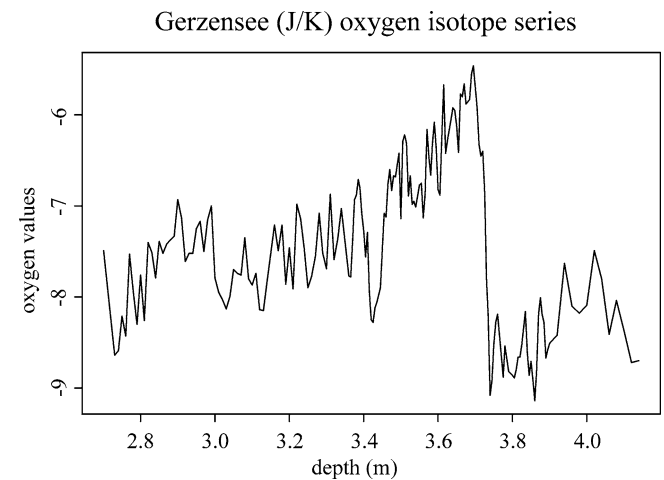


Fig. 2 Gerzensee (J/K) $\delta^{18}\text{O}$ values plotted against depth (cm)

part of the original GRIP isotope series, covering about 15,300 to 12,800 yrs B.P. If no such guideline for choosing the start and end points in the longer series (longer in age) was available, exploratory analysis combined with the maximum matching algorithm proposed later in this paper on trial subdivisions of the complete GRIP series would be required.

Methods and results

The age-estimation algorithm presented here has a smoothing step where raw oxygen values are smoothed as functions of depth (Gerzensee) and age (Greenland). It turns out that the resulting age estimates depend on the choice of the bandwidth in the smoothing steps, so that a combined optimal bandwidth selection method would be required. An alternative idea is to adopt a Bayesian approach and combine the age estimates obtained using different bandwidths; this will be pursued elsewhere. Further variations of this idea are possible. We adopt a robust version by taking the median instead of the mean. This is discussed below.

The main principle adopted here is that the functional relationship between age, rescaled depth, and rescaled oxygen-isotope values for the GRIP data carry over to the Swiss data. It is assumed that the age of the Greenland series has already been established by a reliable method so that this series can be used as a standard for further comparison. As in the case of other similar methods for age-depth modelling, the success of this method relies on the fact that the end points of the two series to be synchronized are chosen well so that reliable and meaningful estimation is possible. For this purpose, the two series are further subdivided by matching major swings in the two data sets. In order to concentrate on major swings at this stage, small local variations in the isotope values are eliminated by nonparametric smoothing using local polynomial regression (Fan and Gijbels 1996).

This smoothing is done with isotope values in the vertical axis and the rescaled age (for GRIP) and depth (Gerzensee) values in the horizontal axis. As described in the algorithm below, at this stage, for the purpose of smoothing, the age (GRIP) and the depth (Gerzensee) values need to be rescaled to a unit interval between zero and one. Formally, the implication of this smoothing operation is that the isotope series is decomposed into a smooth trend function $g(t)$ and local variations $e(t)$. In the case of GRIP, t refers to rescaled age, whereas for Gerzensee, t refers to rescaled depth. In a smoothing method, rescaling of the time index to the unit interval as discussed above is a standard procedure (Wand and Jones 1994). A bandwidth or window-width denoted by b is used for smoothing each data set. It turns out that the chosen bandwidth affects the variance of the final age estimate. This will be addressed again later.

After this smoothing step, the most prominent local minima and maxima in the estimated trend functions are detected. For example, when one maximum and one minimum are detected in a series, the series is divided into

three subseries. Subdivision of each isotope series is done separately. The regression modelling is then done for each pair of isotope subseries. Thus the overall approach may be viewed as a piecewise regression method.

The regression steps are as follows, where notations have been named with reference to the data set at hand.

$$\text{age}_{\text{grip}} = f(\text{isotope}_{\text{grip}}, \text{depth}_{\text{grip}}) + \text{error}.$$

Here f is the regression function and \hat{f} is its regression fit. The parameters of this model are used to estimate (predict) the age of the Swiss series as

$$\text{estimated.age}_{\text{swiss}} = \hat{f}(\text{isotope}_{\text{swiss}}, \text{depth}_{\text{swiss}}).$$

This is justified by the assumption that the corresponding global climatic shifts (covering ca. 15,000–13,000 yrs B.P. [1950]) are recorded synchronously in the Greenland ice and in central European calcareous lake sediments. Prior to this regression modelling, all depth values and the isotope values are standardized. The standardization takes care of the fact that the absolute values of the isotope records or the depths in the two series are not comparable.

Algorithm

In what follows, the algorithm is presented explicitly. For ease of presentation, the notations refer to the current data sets. However, the algorithm is generally applicable. In our example, let $x_{\text{iso,green}}$ and $x_{\text{iso,gerzen}}$ be the isotope series for Greenland and Gerzensee, respectively. Also let $x_{\text{age,green}}$ and $x_{\text{depth,green}}$ be the corresponding age and depth series for Greenland and $x_{\text{depth,gerzen}}$ the Gerzensee depth series. For other data sets, the notations would have to be renamed. In what follows, we assume that the data do not have any pathological characteristics so that standard smoothing and regression type methods can be applied. Essentially this implies the existence of finite second moments (variances and covariances).

Step 1: smoothing

To obtain a first rough match the most important features of the two isotope series $x_{\text{iso,green}}$ and $x_{\text{iso,gerzen}}$ are synchronized. This is done as follows:

- Step 1a: Let $t_{\text{green}}(i) \in [0, 1]$, $i = 1, 2, \dots, n_1$ denote age for the Greenland series rescaled to the range $[0, 1]$. Similarly let $t_{\text{gerzen}}(i) \in [0, 1]$, $i = 1, 2, \dots, n_2$ be depth in the Gerzensee series, also rescaled to the range $[0, 1]$. Suppose that $g_{\text{green}}(t)$ and $g_{\text{gerzen}}(t)$ are the trend functions in the isotope curves for the two series.
- Step 1b: Fix a bandwidth b and using this bandwidth, fit a smooth curve by local polynomial fitting to each of the paired data sets $(t_{\text{green}}(i), x_{\text{iso,green}}(i))$ and $(t_{\text{gerzen}}(i), x_{\text{iso,gerzen}}(i))$. Thus the estimated smooth

curves $\hat{g}_{\text{green}}(t_{\text{green}}(i))$ and $\hat{g}_{\text{gerzen}}(t_{\text{gerzen}}(i))$ are obtained. In this analysis, *S-Plus* function *loess* is used.

- Step 1c: Interpolate both curves \hat{g}_{green} and \hat{g}_{gerzen} to the same grid on $[0,1]$. For the specific data here a regular grid of 1000 points is used.

Step 2: max-min matching

- Step 2a: In each series, find the k_1 most extreme local maxima and k_2 most extreme local minima. In this analysis, $k_1 = k_2 = 1$ was used, that is one minimum and one maximum (three subseries) were identified. How extreme the local maxima or the local minima are, is defined as follows. First, local peaks on moving windows were defined on the smooth curves. Then the relative depth of each peak was assessed based on a discrete measure of left and right derivatives at the peak. In other words, the local maxima or the minima need not be the global maxima or minima, respectively.
- Step 2b: Identify the t -values corresponding to the maxima and the minima in the two series. The subseries defined by these maxima and minima are used for the parametric regression modelling. Note that the regression modelling described in the next step uses the raw data and not the smoothed fitted trend functions. The smoothed trend functions are only used for identifying major swings in the two isotope curves.

Step 3: regression based age estimation

For each subseries, the following regression is carried out. The purpose of the regression modelling is that the parameters can be interpreted in a standard manner and the fitted model can be used for estimating the age of the Swiss series. To obtain scale invariance, the isotope and the depth values of each subseries are standardized by subtracting the mean and dividing by the sample standard deviation. For the Greenland series we use the following notations:

$y = \text{age}$, $z_{11} = \text{vector of standardized depth}$ and $z_{21} = \text{vector of standardized isotope measurements}$. Similarly for Gerzensee, let $z_{12} = \text{vector of standardized depth}$ and $z_{22} = \text{vector of standardized isotope measurements}$. Now the following regression model is fitted to the Greenland series: $y = Z\beta + \varepsilon$ where, $y = (y(1), y(2), \dots)^t$, β the parameter vector of an appropriate length and Z is the design matrix involving powers of z_{11} and z_{21} these powers correspond to a degree m regression based on z_{11} and z_{21} . In this analysis, $m = 3$ was used. Thus an estimated functional relationship

$$\widehat{\text{age}}(j) = \hat{y}(j) = f(z_{11}(j), z_{21}(j))$$

is obtained. Assuming that this basic relationship carries over to Gerzensee, the age estimates for Gerzensee can now be given as

$$\widehat{\text{age}}(j) = \hat{y}(j) = f(z_{12}(j), z_{22}(j)).$$

Table 1 Estimated age (yrs B.P. [1950]) of the Gerzensee oxygen isotope values corresponding to depth (cm)

Depth	Yrs B.P.	Depth	Yrs B.P.	Depth	Yrs B.P.
287	13088.34	340.5	13821.15	372	14551.96
288	13104.28	341	13828.32	372.5	14564.16
289	13120.01	341.5	13834.01	373	14578.18
290	13134.26	342	13839.18	373.5	14592.12
291	13150.57	342.5	13911.5	374	14610.38
292	13166	343	13923.2	374.5	14621.68
293	13181.22	343.5	13934.52	375	14631.63
294	13196.14	344.5	13957.19	375.5	14643.26
295	13210.88	345.5	13974.1	376	14702.17
296	13225.31	346	13984.92	377.5	14758.45
297	13240.04	346.5	13995.02	378	14766.89
298	13254.09	347	14005.82	379	14784.41
299	13267.71	347.5	14012.27	380	14801.78
300	13281.97	348	14020.52	380.5	14810.47
301	13295.23	348.5	14026.38	381	14819.09
302	13308.69	349.5	14113.83	381.5	14827.68
303	13321.86	350	14122.2	382	14836.36
304	13336.37	350.5	14130.76	382.5	14844.97
305	13351.23	351	14139.22	383.5	14862.49
306	13364.57	351.5	14147.1	384	14871.08
307	13377.94	352	14154.66	384.5	14879.96
308	13391.97	352.5	14162.75	385	14888.55
309	13404.4	353	14170.98	385.5	14897.38
310	13417.33	353.5	14179.01	386	14906.11
311	13431.01	354	14187.13	386.5	14914.71
312	13442.04	355	14202.98	387	14923.33
313	13455.05	355.5	14210.98	387.5	14932.38
314	13469.86	356	14219.38	388	14940.72
316	13496.55	356.5	14227	388.5	14949.34
317	13509.64	357	14236.14	389	14958.11
318	13522.41	357.5	14243.19	389.5	14966.75
319	13534.43	358	14250.91	390	14975.41
320	13548.4	358.5	14259.49	392	15010.16
321	13559.98	359	14268.47	394	15047.3
322	13573.13	359.5	14275.37	396	15081.37
323	13586.74	360	14282.87	398	15124.25
324	13600.07	360.5	14290.93	400	15156.54
325	13611.74	361	14302	402	15192.63
326	13625.3	361.5	14313.67	404	15221.91
327	13638.95	362	14324.54	406	15252.08
328	13651.51	362.5	14335.8	408	15286.28

The confidence intervals for the true age at each depth can be given using basic least-squares theory (Rao 1973).

Step 4: calculation on a grid of bandwidths

The choice of the bandwidth in the loess (local polynomial regression) used in Step 1 affects the final result. In fact, it can be analytically proved (Wand and Jones 1994) that the mean squared error of the estimated age will depend on the bandwidth parameter b . This is taken into account as

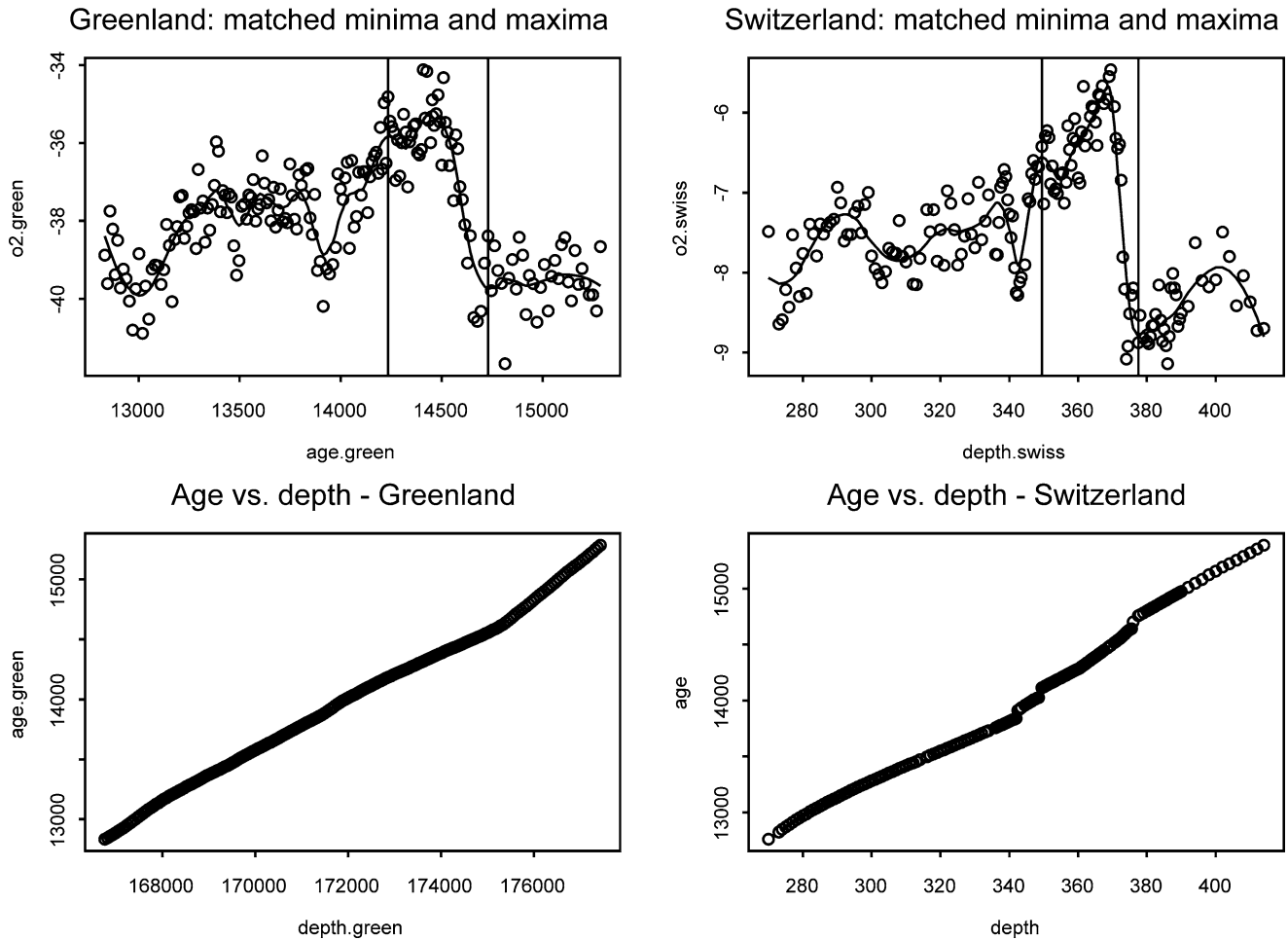


Fig. 3 Matched maxima and minima (*above*) and estimated age (yrs B.P.) vs. depth (cm) for Greenland and Gerzensee (*below*)

follows. Steps 1 to 3 are repeated for a selection of bandwidths. Then, for each depth, the median of the estimated age values obtained by using different bandwidth parameters is chosen. Using the median instead of the mean leads to a robust choice.

The final age estimates for the various depths are given in Table 1. The matched maxima and the minima are shown in Fig. 3 (above). The estimated age plotted against depth for the two series are in Fig. 3 (below). Fig. 4 shows the corresponding synchronized isotope series (with the smoothed age-depth relationship).

Discussion

The regression based algorithm estimates the age of the Swiss series by using the empirical dependence between the estimated age, depth and the isotope values in the Greenland series. This being the main principle, further refinements involve selecting regions in each series to be compared. We use a max-min algorithm to identify these regions in the two series. Another technical issue is the role of the bandwidths, as this affects the final age estimates. Some authors in the statistical literature have considered local bandwidths as

opposed to global bandwidths. A thorough approach for doing this for stratigraphic data would involve a possibly parsimonious model for the serial correlations in the data and developing an algorithm based on minimization of the mean squared error of the estimates (for example Ghosh

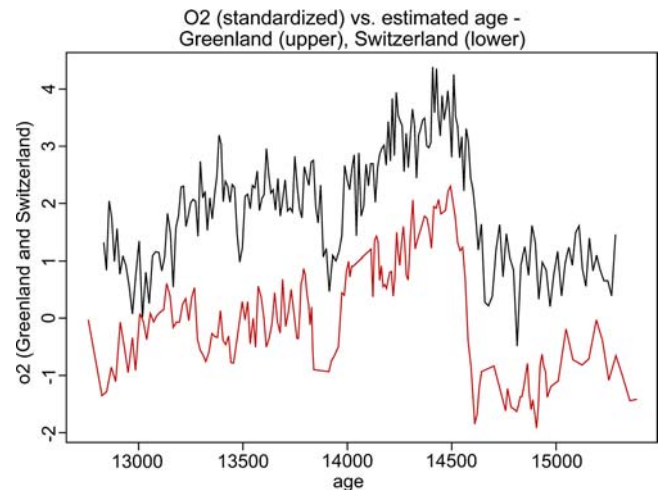


Fig. 4 Synchronized isotope series. For the presentation, the isotope values have been rescaled

and Draghicescu 2002; Beran and Feng 2002). This will be considered elsewhere.

The estimated age can be an ingredient for further studies. For instance, terrestrial vegetation as represented by the fossil pollen record may be expected to show a delayed response to climatic change (Wright 1984), and appropriate models can be developed to study such a relation (Lotter et al. 2000).

However, one problem with such analyses is that when the biotic response (Ammann 2000; Birks 1995; Tinner and Lotter 2001) is analyzed as a function of time (in this case, estimated age), a major source of uncertainty is in time itself, because time is estimated here. An in-depth analysis would involve developing appropriate methods for time series analysis with non-equidistant and estimated time points. Application of these ideas in palaeoecological research will be the subject of future research.

Acknowledgements This research was motivated by the encouragement of Brigitta Ammann (Bern) who suggested looking into the problem of statistical age estimation of a stratigraphic sequence in the context of her Rapid Warming Project.

J. Beran (Konstanz) kindly provided his max-min SPlus code and W. Tinner (Bern) drew attention to some important literature on age-depth modelling cited in this article. J. Schwander (Bern) read through an earlier draft of the manuscript and provided background information concerning the Greenland Ice Core Project. The author would like to thank all these colleagues for valuable discussions and H.J.B. Birks (Bergen) for detailed editorial remarks that significantly improved presentation. Helpful remarks by two referees are also acknowledged. The GRIP $\delta^{18}\text{O}$ values and the corresponding age estimates can be downloaded from <http://www.glaciology.gfy.ku.dk/data/grip18o.txt> maintained by the Glaciology Group of the University of Copenhagen.

References

- Ammann B (2000) Biotic responses to rapid climate changes: introduction to a multidisciplinary study of the Younger Dryas and minor oscillations on an altitudinal transect in the Swiss Alps. *Palaeogeogr Palaeoclimatol Palaeoecol* 159:191–201
- Bennett KD (1994) Confidence intervals for age estimates and deposition times in late-Quaternary sediment sequences. *The Holocene* 4:337–348
- Beran J, Feng YH (2002) Local polynomial fitting with long memory, short memory and antipersistent errors. *Ann Inst Stat Math* 54:291–311
- Birks HJB (1995) Quantitative palaeoenvironmental reconstructions. In: Maddy D, Brew JS (eds) *Statistical Modelling of Quaternary Science Data*. Quaternary Research Association, Cambridge, pp 161–254
- Birks HJB, Heegard E (2003) Developments in Age-Depth modelling of Holocene stratigraphical sequences. *PAGES News* 11:7–8
- Eicher U (1979) Die $^{18}\text{O}/^{16}\text{O}$ - und $^{13}\text{C}/^{12}\text{C}$ -Isotopenverhältnisse in spätglazialen Süßwasserkarbonaten und ihr Zusammenhang mit den Ergebnissen der Pollenanalyse. Thesis, Bern
- Fan J, Gijbels I (1996) Local polynomial modelling and its applications. Chapman and Hall, London
- Ghosh S, Draghicescu D (2002) An algorithm for optimal bandwidth selection for smooth nonparametric quantiles and distribution functions. In: *Statistics in industry and technology: statistical data analysis based on the L_1 -norm and related methods*. Birkhäuser, Basel, pp 161–168
- Gulliksen S, Birks HH, Possnert G, Mangerud J (1998) A calendar age estimate of the Younger Dryas-Holocene boundary at Kråkenes, western Norway. *The Holocene* 8:249–259
- Hammer CU, Andersen KK, Clausen HB, Dahl-Jensen D, Hvidberg CS, Iversen P (1997) The stratigraphic dating of the GRIP ice core. Special report of the Geophysical Department, Niels Bohr Institute for Astronomy, Physics and Geophysics, University of Copenhagen
- Heegard E, Birks HJB, Telford RJ (2005) Relationship between calibrated ages and depth in stratigraphical sequences: an estimation procedure by mixed-effect regression. *The Holocene* 15:612–618
- Johnsen SJ, Clausen HB, Dansgaard W, Gundestrup NS, Hammer CU, Andersen U, Andersen KK, Hvidberg CS, Dahl-Jensen D, Steffensen JP, Shoji H, Sveinbjörnsdóttir AE, White JWC, Jouzel J, Fisher D (1997) The $\delta^{18}\text{O}$ record along the Greenland Ice Core Project: deep ice core and the problem of possible Eemian climatic instability. *J Geophys Res* 102:26397–26410
- Lotter AF, Birks HJB, Eicher U, Hofmann W, Schwander J, Wick L (2000) Younger Dryas and Allerød summer temperatures at Gerzensee (Switzerland) inferred from fossil pollen and cladoceran assemblages. *Palaeogeogr Palaeoclimatol Palaeoecol* 159:349–361
- Rao CR (1973) *Linear statistical inference and its applications*, 2nd edn. Wiley, New York
- Schwander J, Eicher U, Ammann B (2000) Oxygen isotopes of lake marl at Gerzensee and Leysin (Switzerland), covering the Younger Dryas and two minor oscillations, and their correlation to the GRIP ice core. *Palaeogeogr Palaeoclimatol Palaeoecol* 159:203–214
- Siegenthaler U, Eicher U, Oeschger H, Dansgaard W (1984) Lake sediments as continental $\delta^{18}\text{O}$ records from the glacial/post-glacial transition. *Ann Glaciol* 5:149–152
- Telford RJ, Heegard E, Birks HJB (2004) All age-depth models are wrong: but how badly? *Quat Sci Rev* 23:1–5
- Tinner W, Lotter AF (2001) Central European vegetation response to abrupt climate change at 8.2 ka. *Geology* 29:551–554
- Wand M, Jones MC (1994) *Kernel smoothing*. Chapman and Hall, New York
- Wright HE (1984) Sensitivity and response time of natural systems to climatic change in the late Quaternary. *Quat Sci Rev* 3:91–131