Identification of a periodic time series from an environmental proxy record

Anouk de Brauwere [a,*], Fjo De Ridder [a,b,1], Rik Pintelon [b], Jeroen Meersmans [c], Johan Schoukens [b], Frank Dehairs [a]

[a] Department of Environmental and Analytical Chemistry, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium
[b] Department of Fundamental Electricity and Instrumentation, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium
[c] Department of Geography, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium

ARTICLE INFO

ABSTRACT

The past environment is often reconstructed by measuring a certain proxy (e.g. $\delta^{18}O$) in an environmental archive, i.e. a species that gradually accumulates mass and records the current environment during this mass formation (e.g. corals, shells, trees, etc.). When such an environmental proxy is measured, its values are known on a distance grid. However, to relate the data to environmental variations, the date associated with each measurement has to be known too. This transformation from distance to time is not straightforward to solve, since species usually do not grow at constant or known rates. In this paper, we investigate this problem for environmental archives exhibiting a certain periodicity. In practice, the method will be applicable to most annually resolved archives because these contain a seasonal component, e.g. clams, corals, sediment cores or trees. Due to variations in accretion rate the data along the distance axis have a disturbed periodic profile. In this paper, a method is developed to extract information about the accretion rate, such that the original (periodic, but further unknown) signal as a function of time can be recovered. The final methodology is quasi-independent of choices made by the investigator and is designed to deliver the most precise and accurate result. Every step in the procedure is described in detail, the results are tested on a Monte-Carlo simulation, and finally the method is exemplified on a real world example.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

A fundamental problem often encountered in proxy records is the reconstruction of the time series, starting from a measured distance series. Variations in accretion rate squeeze and stretch the distance series (Fig. 1). This distortion results in the lengthening and shortening of individual features present in the signal, as can be seen on the horizontal graph. Mostly, investigators are interested in the time series (vertical graph). Indeed, the associated time series is much more interpretable than the original measured distance series, especially when

(i) two different proxy series are to be compared, because each series has its unique accretion rate, or
(ii) a proxy is to be compared with an instrumental record measured as a function of time, because both must have the same grid (time).

In other cases the accretion itself is of interest (diagonal graph on Fig. 1). This can hold important information because the sedimentation or growth rate is often influenced by the environment itself and can thus be
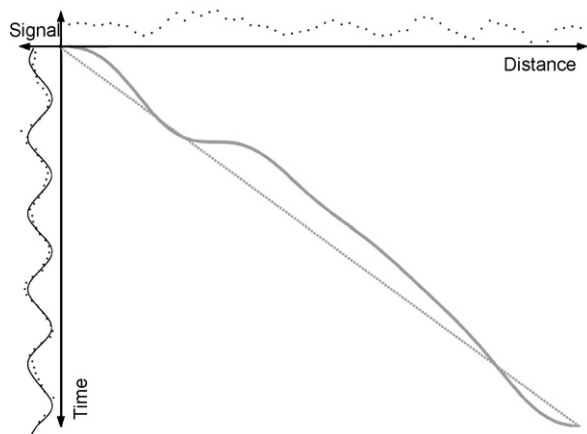

Fig. 1. Concept of transformation of a record along a distance grid (dots on horizontal axis) to a time series (dots on vertical axis). Dashed diagonal line represents constant growth, while curved gray line along diagonal represents accretion optimized by method described in this paper, which assumes that true time series is periodic (full black line in vertical graph). Mirroring distance series on diagonal curve results in time series.

used as an additional source of information. The problem handled in this paper is the reconstruction of the time series and accretion (rate) for a given distance series. As can be seen in Fig. 1, it is impossible to estimate the vertical and diagonal graphs if only the horizontal one is given. How to overcome this problem in a systematic way is the core of this research.

The new method presented in this study was developed in an attempt to simultaneously reduce stochastic and systematic errors, and still keep a flexible and user-independent (i.e. objective) product. The methods proposed or used in the literature so far appear to be unsatisfactory with respect to at least one of these criteria.

The most frequently observed approach is to remove distortions of the time base by correlating by eye. A number of reference points are chosen in the observed distance series and in a priorly assumed time series, and the observation dates are optimized by stretching and shortening the observation records between the anchor points until this new record is most similar to the predefined time series. This, however, works only in the simplest cases, and is not readily quantified and is generally limited to low resolution. What makes things worse is that two investigators, who are dating the same record with this same method, will come to different conclusions, because they may have selected different tuning points. The selection of these "anchor points" is not only subjective but also absolute, i.e. they are assumed to be correct and ignore the fact that the data are subject to stochastic uncertainty. Therefore, systematic errors are easily introduced in this procedure (wrong anchor points) and the measurement noise is fully propagated into the resulting time series. Finally, due to the use of fixed anchor points, the resulting accretion rate is a step function, which can hardly be called realistic for a natural process like the growth of a shell or the accumulation of tree biomass.

More refined methods have been proposed in literature, including those by Martinson et al. (1982), Lisiecki and Lisiecki (2002), Wilkinson and Ivany (2002) and De Ridder et al. (2004). Most of these methods display the serious weakness that they need the time series to be known in advance. Consequently, the problem of dating the observations becomes a task of finding what was postulated. Furthermore, some input of the user is always crucial in order to find reasonable results. For instance, the method proposed by Martinson et al. (1982) is based on an optimization routine which is highly sensitive to initial values. For the method described in Lisiecki and Lisiecki (2002) the user has to decide how many anchor points must be used; in Wilkinson and Ivany (2002) the user has to define the width of the time window in advance; the non-parametric method described in De Ridder et al. (2004) needs the width of a spectral window to be specified by the user.

In this paper, the aim is to achieve a flexible, user-independent, accurate and precise reconstruction of the time series. This means the use of anchor points is avoided, since these are essentially subjective and result in unrealistic stepwise accretion rates. Instead, some other hypothesis about the shape of the unknown accretion rate or time series must be made. Depending on the application, different hypotheses may be suitable. Yet, in order to be flexible enough, the hypothesis should not be too strict. In the current paper, a periodic time series is assumed, which is often the case for annually resolved archives, because of the presence of a seasonal cycle. Such an assumption is much less restrictive than assuming it is a sine, and certainly than assuming it is fully known. This (or any other) hypothesis can be quantified in a parametric model, containing time explicitly. The problem can then be solved by a simultaneous estimation of the signal model parameters and the time base for the measurements.

The new method is most similar to the strategy proposed by Brüggeman (1992), which simultaneously optimizes the time–depth (or time–distance) relation and a linear dynamic model for the time series. This is the only method, so far, which does not assume that the time series itself is known in advance, but only assumes it to belong to a certain class, described by a parametric model. The time–depth relation is optimized by tuning anchor points, which is the main difference with the new method, where the time base is also described by a parametric model. So, with the new method the estimated accretion rate will not be stepwise constant, but a continuous representation of the accretion rate is ensured. The use of a parametric model also reduces the influence of stochastic errors and thus precision of the results is improved. Besides this, the risk of model errors is limited because the shape of the time series is not fixed. Indeed, a parametric model is used, which has a number of free parameters to be optimized to find the best correspondence with the observations. In addition, not one but a number of models will be considered simultaneously in order to reduce the risk of model errors even more. The most appropriate model is selected using a statistical criterion.

The remainder of this article is structured as follows. In Section 2 parametric models for the time base and time series are proposed. Estimating the model parameters maximizes precision by filtering off the measurement noise. The problem of finding appropriate starting values and other issues of the algorithm are discussed in Section 3. Time reversal and optimizing model complexity are discussed, respectively, in Sections 4 and 5. Section 6 consists of a Monte-Carlo simulation to test the algorithm, and in Section 7 real data are processed. In order to illustrate the subsequent choices and strategies, real data have been used throughout the paper.

2. The signal model and time base

If we made no assumptions at all about the signal as a function of time it would be impossible to reconstruct a time base for the measurements. This is the reason why many methods are based on an a priori known target function (e.g. Martinson et al., 1982; Paillard et al., 1996; Yu and Ding, 1998; Lisiecki and Lisiecki, 2002; Ivany et al., 2003). However, the time series is usually not exactly known and may even be the unknown of interest. Therefore, it is important to use a method that does not fully fix the time series in advance, but some assumption will still be necessary. Here it is assumed that the signal is periodic, which is an obvious assumption for annually resolved archives, because they often exhibit a seasonal cycle. Other assumptions can be made, but for clarity we limit this discussion to periodic models. In this section a formal description of the method is given.

The measurements are assumed to be sampled along an equidistant distance grid. These observations are modeled by a discrete time signal $s_{model}(t(n))$, in which the assumption of periodicity is contained:

$$s_{model}(t(n)) = A_0 + \sum_{k=1}^{h} \left[A_k \sin(k\omega t(n)) + A_{k+h} \cos(k\omega t(n)) \right]$$
(1)

where $t(n)$ is the unknown time variable at sample position $n \in \{1,\ldots,N\}$, A_0 is the offset, A_k and A_{k+h} are the unknown amplitudes of the kth harmonic, ω is the unknown fundamental angular frequency and h is the number of harmonics, yet to be identified. Changing the number of harmonics will change the complexity of the model. How to choose the optimal value for h is discussed in Section 5. Note that Eq. (1) does not generally represent a Fourier analysis, since this would require that $\omega = 2\pi/T$ (with T the period spanning the whole record) and that the observations are sampled equidistantly in *time*.

Although the samples are equidistantly spaced along the distance axis, the time instance between two subsequent observations is not constant, due to variations in accretion rate. Note that it is practically impossible to estimate the time base distortion (TBD) directly, by comparing the measurements with the signal model (Eq. (1)), because

Problem (A): the measured record is disturbed by stochastic noise, which would be propagated into the

TBD. This would result in a very low precision on the time base and would make the results practically useless; and

Problem (B): the signal model itself is unknown (even if we assume it is periodic, the abovementioned parameters are still unknown).

To circumvent these two problems, the time instances $t(n)$ are modeled as follows:

$$t(n) = (n + \delta(n))T_s$$
(2)

where T_s is the average sample period (time between two subsequent observations) and where the twists of the time base are modeled by a function $\delta(n)$, called the TBD. In Fig. 1, the TBD is the difference between the dotted line and the full line along the diagonal. To characterize this TBD, it is expanded in a set of basis functions:

$$\delta(n) = \sum_{m=1}^{b} B_m \beta_m(n)$$
(3)

where β is the set of b basis functions (see Fig. 2 for an example), B is a vector of length b with the TBD parameters. The parameter b defines the complexity of the time base function, which still has to be determined (see Section 5).

Problem (A) is circumvented by the introduction of basis functions (Eq. (3)). We tested trigonometric functions, Legendre polynomials and splines as basis functions (Abramowitz and Segun, 1968; Dierckx, 1995). The last seem to work best. This choice has two reasons:

(i) If the TBD is large for a certain observation, we may intuitively assume that it will be large as well for its neighboring observations and that this correlation decreases proportionally to the distance between the observations. This type of behavior can exactly be described by splines (see Fig. 2). On the other hand,
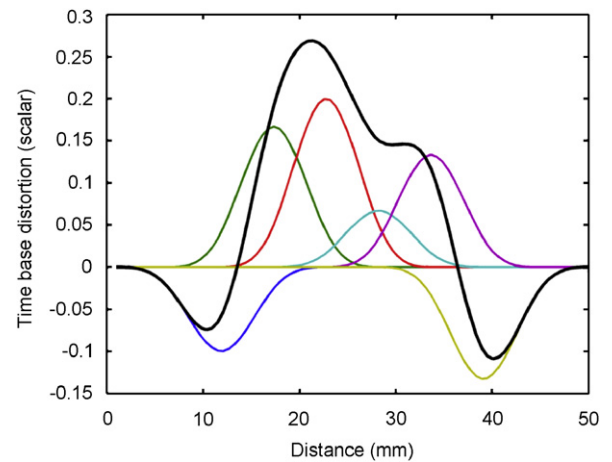

Fig. 2. Six spline basis functions are shown in colors. These Gaussian-like functions have been multiplied with their parameters B, $B = [-0.15, 0.25, 0.3, 0.1, 0.2, -0.2]$. If TBD for observation around 20 mm is large and positive, then red spline function would obtain a large positive parameter value B_3. Consequently, neighboring observations would have a large TBD as well, because of the width of the spline function. TBD is found by summing all scaled basis functions and is shown in black.

keep in mind that hiatuses (growth stops) are not very well described by this type of basis functions. The use of polynomials would be beneficial if slow trends are present in the time base and trigonometric functions would outperform if periodic variations are expected.

(ii) The optimization routine is much more effective for splines basis functions than for the other tested functions. More particularly, local minima seem to be a recurring problem for these other basis functions.

In order to overcome Problem (B), the signal model and time base model are optimized simultaneously, i.e. all unknowns ω, A, B (in practice, these are stored in a vector $\theta = [\omega, A^{\mathrm{T}}, B^{\mathrm{T}}]^{\mathrm{T}}$) are estimated simultaneously. This avoids circular reasoning. Indeed, if the time base is known, the signal model parameters can easily be estimated, but to know the time base it is necessary to have identified the signal model. The optimal values of all these parameters can be calculated by a numerical minimization algorithm, which minimizes a least squares (LS) cost function (objective function). More details about the optimization are given in the next section.

While the central concepts of the new approach are given above, a number of additional issues are important to consider for the results to be both realistic and reliable. These issues are discussed in the following sections.

3. Initialization and optimization strategy

Minimizing the LS cost function will be successful only if one can start from a reasonable set of initial values. Finding good initial values is not trivial and this was one of the main shortcomings of the method proposed by Martinson et al. (1982). However, since their publication some non-parametric methods have been proposed (Wilkinson and Ivany, 2002; De Ridder et al., 2004). Such methods are more sensitive to noise, but do not need initial guesses for the parameters. We used the phase demodulation method (De Ridder et al., 2004) to find a non-parametric estimate of the TBD, which can serve as the starting value for the parametric model optimization. In order to minimize the risk of converging towards a bad local minimum, the optimization strategy is performed in five steps (for a structure of the algorithm, see Fig. 3):

1. Initialization of the angular frequency ω: a non-parametric estimate of the TBD and the corresponding frequency can be gathered for periodic signal records, e.g. following the guidelines of De Ridder et al. (2004).
2. Initialization of the TBD parameters B: initial values for the TBD parameters can be gathered by matching Eq. (3) on the non-parametric TBD. This can easily be done, because Eq. (4) is linear in B. Next, Eq. (2) is used to obtain more precise dates of the observations.
3. Initialization of the signal parameters A: these are gathered by matching Eq. (1) on the observations employing the previously estimated time base. An efficient algorithm is described in Pintelon and Schoukens (1996).
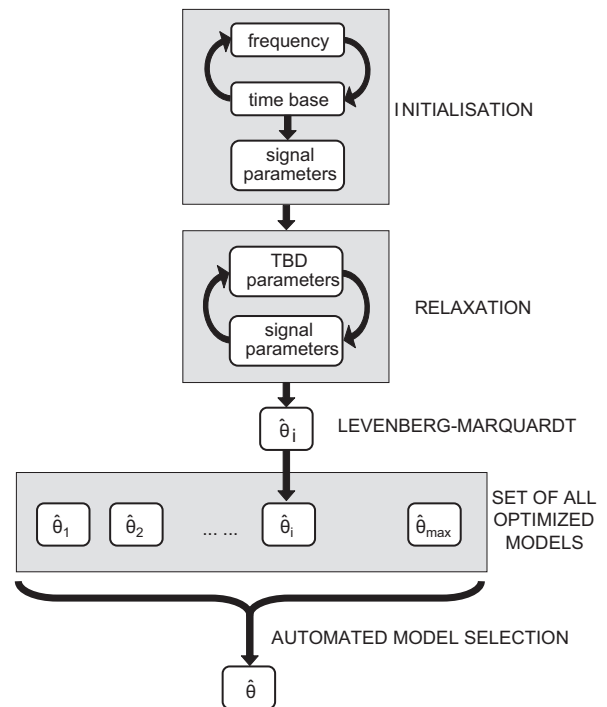

Fig. 3. Structure of algorithm. Initialization, relaxation and Levenberg–Marquardt optimization must be performed for all models $i \in \{1, \ldots, \max\}$, having complexity (h_i, b_i), delivering optimal values $\hat{\theta}_i$ for their parameters. Using an automated model selection procedure (here based on the $\mathrm{MDL_c}$ criterion) optimal model complexity can be selected, giving one final optimal parameter set $\hat{\theta}$.

4. Relaxation: alternating, the TBD parameters and the signal parameters are optimized, while the other set is remained fixed. Note that optimizing the TBD parameters, while the signal parameters are constant, is, in fact, the method proposed by Martinson et al. (1982). This relaxation algorithm is stopped when the largest relative variation in the parameter vector $\theta = [\omega, A^{\mathrm{T}}, B^{\mathrm{T}}]^{\mathrm{T}}$ is lower than a numerical stop criterion (typically 10^{-3}). This step in the optimization is implemented to increase the calculation speed but it will not influence the final results.
5. Final simultaneous parameter estimation: all parameter values are estimated together employing a Levenberg–Marquardt algorithm (this algorithm is preferred over the Newton–Gauss algorithm, because the convergence area is larger). Further, each row in the Jacobian was scaled by its standard deviation in order to improve the numerical conditioning.

4. Local time reversal problem

If the observations are ordered along the distance grid, logically this same order should be found along the time grid. So far, this is not necessarily true. The method as presented until now is too powerful and can alter the observations' order. In this paragraph a solution is presented.
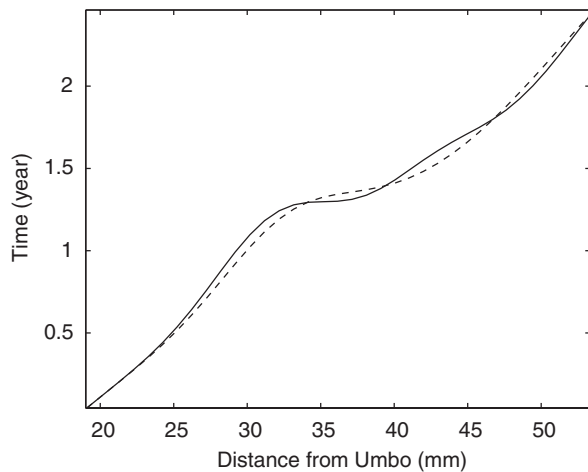
Fig. 4. Around 35 mm from the Umbo, estimated time is constant (full line), which would correspond to an infinite accretion rate. In order to avoid this type of un-physical solutions, an inequality constraint optimization is implemented: if negative accretion rates occur, time base is forced to remain slightly positive (dotted line). Accretion curves shown here are calculated while processing clam 1, with a high complexity $(h,b) = (4,20)$.

The method is especially vulnerable for this weakness when the number of TBD parameters is relatively high (an example is given in Fig. 4). This is not surprising, because the noise sensitivity is larger in this case (more parameters have to be estimated from the same number of observations). Since we known this reversal behavior is physically impossible, an inequality constraint optimization was implemented (e.g. Fletcher, 1991): in each step of the Levenberg–Marquardt algorithm a check is performed to verify if any time reversals have occurred. If the minimal sample period is lower than 20% of the average sample period, the constraints become active and the TBD at these samples is fixed at this minimum for this step of the optimization routine. The dotted line in Fig. 4 shows the result after the implementation of the inequality constraint optimization. This 20% is an arbitrary boundary. If the dates of two observations coincide, the accretion rate at that time is infinite, which is not realistic. Even if two observations do not coincide but are dated too closely to each other, the corresponding accretion rate is no longer realistic. So, at a certain accretion rate we have to control the further evolution of the time base during the optimization process. The constraints become active when the time between two neighboring observations becomes smaller than 20% of the average one. Obviously, the value of this parameter can be changed if necessary.

5. Selecting the optimal model complexities *h* and *b*

If we stopped developing the algorithm at this point, the complexity of the signal model and time base model, quantified by *h* and *b*, respectively, are still chosen by the user. Fig. 5 shows some examples of the accretion rate and Fig. 6 the associated signal models, estimated from the same measurement record, with identically the same algorithm, but with different levels of complexity.
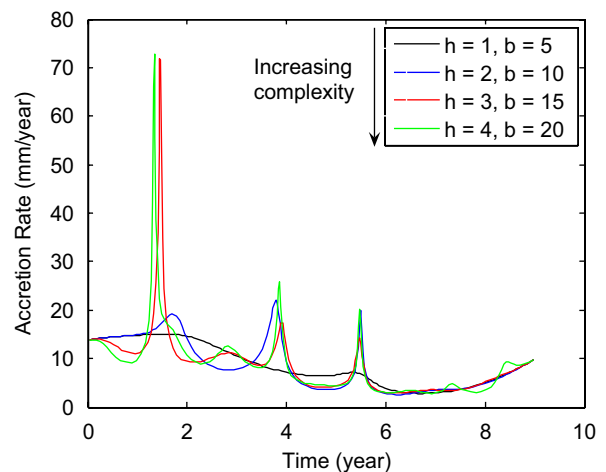

Fig. 5. Accretion rates for different model complexities (all matched on clam 1). Variations increase with increase in complexity, but which of these accretion rates describes reality best? Accretion rates of about 70 mm/year are fixed at these levels due to activated constraints.
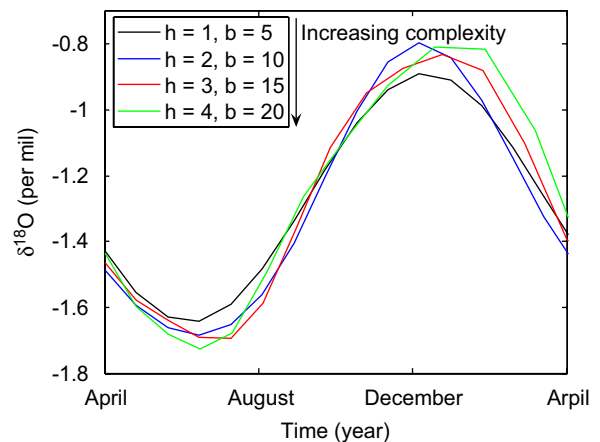

Fig. 6. One period of signals with different complexities (all matched on clam 1). Black signal amplitude is different from green signal amplitude. Signal minima correspond quite well (June), but maxima change with about 1 month. Which of these models is close to reality?

In this section, the strategy for selecting optimal values for *h* and *b* is presented. A good choice will result in a maximum extraction of information, but a too simple or too complex model will produce less accurate or precise results. Increasing the model complexity will decrease the systematic errors; however, at the same time the model variability increases. In other words: at a certain complexity the additional parameters no longer reduce the systematic errors but are used to fit the actual noise realization of the data. Hence, it is not a good idea to select the model with the smallest cost function (best fit) within the set because it will generally continue to decrease when more parameters are added. For this reason the cost function is extended with a penalty term that compensates for the increasing model variability. In this way, the resulting model selection criterion, called MDL$_c$ (De Ridder et al., 2005), is able to detect undermodeling

(= too simple model) as well as overmodeling (= too complex model). In the examples of this paper the criterion to be minimized has the following expression:

$$\text{MDL}_c(\hat{\theta}, n_\theta, n_c, N) = \frac{K(\hat{\theta})}{N} \exp p_c(n_\theta, n_c, N) \qquad (4)$$

with penalty

$$p_c(n_\theta, n_c, N) = \frac{\ln(N)(n_\theta - n_c + 1)}{N - (n_\theta - n_c) - 2} \qquad (5)$$

and $\hat{\theta}$ the optimized values of the model parameters, $K(\hat{\theta})$ the residual cost function (residual sum of squares), N the total number of observations, n_θ the number of parameters and n_c the number of active constraints. This criterion expresses that if the cost function decreases faster than the exponent of the penalty term, the more complex model is preferred over the simpler one. A detailed description of model selection criteria can be found in Akaike (1974), Rissanen (1978), Schwarz (1978), De Ridder et al. (2005), de Brauwere et al. (2005). Notice that introducing a model selection criterion enables the automatic selection of the optimal model complexity (i.e. the optimal values for h and b), again eliminating an interference from the user, which makes the proposed method more objective and user independent. Especially for the time base complexity b such a criterion is of much help, since this complexity is difficult to interpret in terms of physical constraints or processes.

Practically, the user chooses the maximum values for h and b, i.e. $(h_{\max}, b_{\max})$. Next, all models with a model complexity between $(h,b) = (0,0)$ and this maximum are optimized and Eqs. (4) and (5) are used to select the best model within this set (lowest MDL_c value).

What values should be chosen for $(h_{\max}, b_{\max})$? The theoretical maxima for h and b are given by the number of observations N: $h + b \leqslant N$. However, Eqs. (4) and (5) will never select this most complex model. In any practical case, the maximum complexity of the signal model is limited by the Nyqvist frequency. So, as a rule of thumb, $h_{\max}$ equals the average number of observations in each year or period. For the complexity parameter of the time base, $b_{\max}$, a rule of thumb cannot be provided so easily. From our own experience, we suggest that at least five to ten observations per basis function should be available. So, start with $b_{\max} = N/10$. If the model selection criterion (Eqs. (4) and (5)) selects this maximum, the optimization procedure should be repeated with a higher complexity.

6. Monte-Carlo simulations

To test the reliability of the algorithm and quantify noise propagation, a Monte-Carlo simulation is used (Fishman, 1996). Firstly, performance of the method is tested, followed by an examination of the noise propagation.

In order to test the *accuracy of the method*, we have tested it with a relative standard deviation of 40% (signal-to-noise ratio of 2.5). This is a relatively weak signal. One thousand measurement sets were generated and all models were optimized with complexities varying from zero to five harmonics $h \in \{0, \ldots, 5\}$ and from zero to ten

Table 1
One thousand Monte-Carlo simulations have been run

	0	1	2	3	4	5	6	7	8	9	10
Selected h	0	3	*985*	12	0	0	–	–	–	–	–
Selected b	0	0	0	4	*838*	22	7	108	19	2	25

In this table the distribution of the selected model complexities is shown. The 'true' model consisted of two harmonics $h = 2$, and four time base distortion parameters $b = 4$. In the large majority of simulations (83.3%) the criterion was able to identify the 'true' model (italic).

spline functions $b \in \{0, \ldots, 10\}$. The 'true' simulated model consists of two harmonics and four spline basis functions ($h = 2$, $b = 4$). Table 1 shows the distribution of the selected complexities after performing 1000 Monte-Carlo simulations. In 83.3% the MDL_c criterion was able to identify the complexity of the 'true' model. Note that this value increases to 98% if the relative standard deviation of the stochastic noise is decreased to a more realistic 10% (tested on 100 simulations).

To what extent does measurement *noise influence the actual parameter values*? This can be assessed by Monte-Carlo simulations, but also by linearizing the models and calculating the error propagation directly (Pintelon and Schoukens, 2001). However, the latter method is only approximate, especially if the model is highly nonlinear in the parameters and/or if the noise level is large. Both conditions can be checked simultaneously by comparing the linear analysis with the Monte-Carlo simulation. This comparison for the model parameters is summarized in Table 2. Both seem to correspond well and will correspond even better for lower (more realistic) noise levels. If we compare the uncertainties on the signal and time base (not shown), we come to the next conclusions: the average standard deviation on the residual is 0.14 per mil according to the Monte-Carlo simulation, while it is 0.12 per mil according to the linearized uncertainty estimation. Though, for the time base, the Monte-Carlo simulation finds an uncertainty of 0.07 year, while the linearized uncertainty estimation find 0.02 years. This difference is probably due to the non-linearity of the model with respect to the TBD parameters.

## 7. Application: *Saxidomus giganteus*

The method is illustrated on the stable oxygen isotope records ($\delta^{18}\text{O}$) measured in the aragonite shells of *S. giganteus* from Puget Sound (Washington State, USA) sampled in September 2001. For a detailed description of the experimental setup and interpretation of the results, we refer interested readers to Gillikin et al. (2005). This site generally experiences a SST range of 8 °C (8–16 °C) and has a salinity of 27.7 ± 1.06 ‰, with occasional (less than yearly) periods of freshwater input reducing salinity to ca. 20‰. A thick section of the aragonite shell was sampled using a computer-controlled micro-drill from growing tip to half way to the umbo. The carbonate powder ($\pm 100\,\mu\text{g}$ per hole) was processed using an automated carbonate device (Kiel III) coupled to a Finnigan Delta+XL. Data were corrected using an internal laboratory standard and are

Table 2
Summary of the uncertainties on the parameters estimated with a linearized uncertainty propagation method and with a Monte-Carlo simulation

	Period	Offset	Amplitudes				TBD parameters			
	$2\pi/\omega$ (year)	A_0 (‰)	A_1 (‰)	A_2 (‰)	A_3 (‰)	A_4 (‰)	B_1	B_2	B_3	B_4
True parameter values	1	0	1	−0.5	0	0.5	0.4	−0.4	0.2	0.2
Estimated parameters	1.0099	0.05	0.95	0.34	−0.03	0.66	0.31	−0.44	0.15	0.13
Parameter standard deviation from										
Linearization of model	0.0038	0.04	0.06	0.11	0.09	0.08	0.036	0.052	0.054	0.045
Monte-Carlo simulation	0.004	0.04	0.06	0.12	0.09	0.08	0.03	0.05	0.05	0.03

The estimated values are taken from the first run of the simulation.
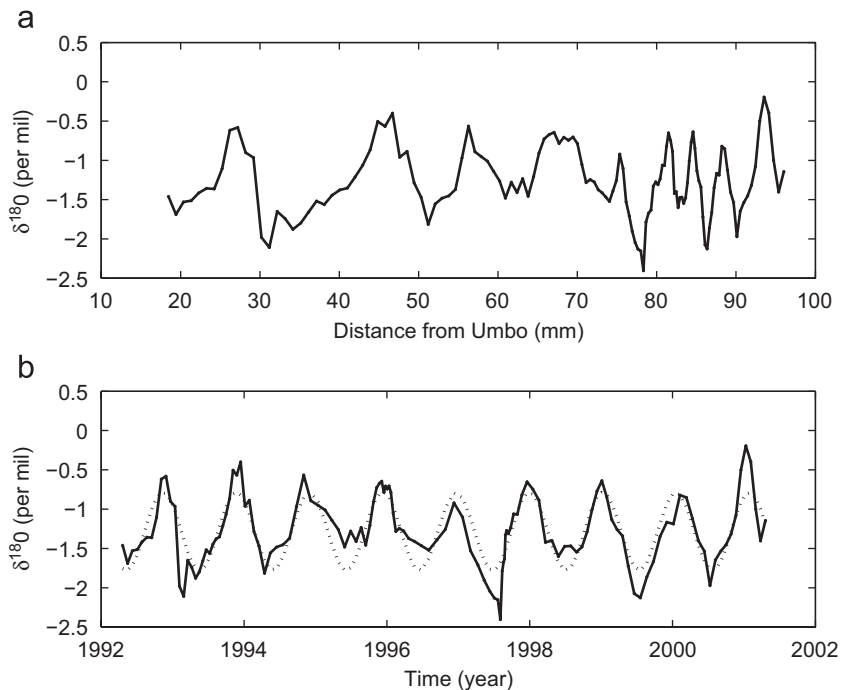
Fig. 7. Clam 1: (a) raw data and (b) signal on constructed time base (full line) and signal model (dotted line).

reported relative to V-PDB in conventional notation. Precision is generally better than 0.08‰. Two different specimens were studied and the raw data are shown in Figs. 7a and 8a (see Gillikin et al. (2005) for more details). Since the specimens come from the same sampling site, it is expected that they have recorded similar environmental conditions. However, due to individually varying growth rates the measured distance series are distinct (Figs. 7a and 8a). The correlation between the reconstructed time records can thus be used to validate the method. The records consist of 190 (clam 1) and 123 (clam 2) observations and cover periods of approximately 9 years. The large winter–summer variations are reflected in these signals and motivate the hypothesis of a periodic signal model.

The maximum model complexity was limited to four harmonics and 20 TBD parameters, i.e. $(h,b)_{\max} = (4,20)$. For certain model complexities, the constraints became active. This was typically the case when the number of

TBD parameters became high. The results are summarized in Table 3: for both clams the lowest model selection criteria were found for a signal model consisting of only one harmonic. Both samples were collected in 2001 and the most recent observation was dated as April 1, so that annual maxima in $\delta^{18}O$ correspond to winter situations. The correlation between both reconstructed records is 84% (see Fig. 9), whereas the original correlation (in the distance domain) was only 27%. The accretion rates of both clams are shown in Fig. 10. Note that

(i) annual variations in the accretion rate may have occurred, but such a TBD model was too complex according to the MDL_c criterion. So, the quality of the data is not sufficient to support annual variations in the TBD;

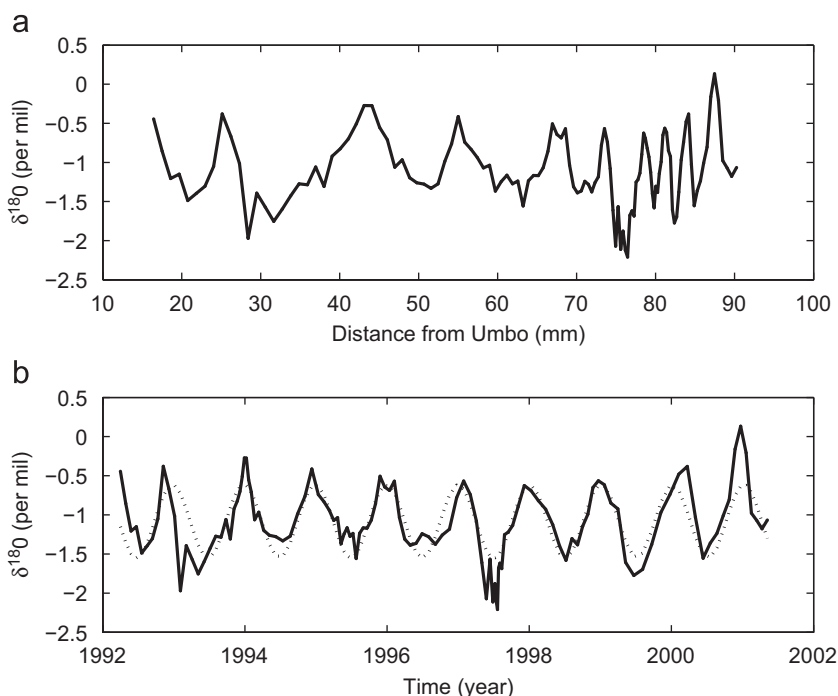(ii) the estimated accretion rates decrease slowly with age, which can be expected;

Fig. 8. Clam 2: (a) raw data and (b) signal on constructed time base (full line) and signal model (dotted line).

Table 3
Summary of the selected models used in the *Saxidomus giganteus* examples

	Clam 1	Clam 2
Residual cost function (per mil)^2, $K(\hat{\theta})$	3.48	4.64
Automated model selection criterion (per mil)^2 MDL$_c$	0.045	0.048
Number of observations, N	133	123
Selected number of harmonics, h	1	1
Selected number of TBD parameters, b	9	10

(iii) most variation occurs at more or less the same moments in both clams.

The latter is reflected in the correlation of 63% between the two accretion rate profiles. The mismatch between the two peaks around 1996 can be due to errors still present in the time base, which is used to construct and date the accretion rate. An alternative explanation can be found in the fact that the accretion rate is a non-linear function of the TBDs. Consequently, small errors on these parameters can have a larger influence on the accretion rate than on the TBD itself.

The relatively high correlation illustrates that these variations are relevant (i.e. it does not reflect the stochastic noise) and that they are determined not only by the age and physiology of the individual specimens but also by external forcing. Otherwise, accretion rate decrease with age would be much smoother or in a random manner with much less similarity between both specimens. Such precise estimations of the accretion rates open the possibility to use this information for climate reconstruction purposes.

8. Summary and conclusion

A new method to reconstruct the time base for periodic archives was presented. It is based on the methods described in Martinson et al. (1982) and De Ridder et al. (2004), and is closely related to the one proposed by Brüggeman (1992). The novelty of this approach is that it estimates the time base together with the signal describing the time series. The method has several advantages:

(i) it is combined with a statistically based model selection criterion (MDL$_c$), to choose the most appropriate model complexity;

(ii) which makes it robust to overmodeling in the signal and time base model; and

(iii) which makes it robust to undermodeling;

(iv) it is robust to stochastic measurement errors, since parametric signal and time base models are used, whose complexities are carefully chosen,

(v) it is robust to non-sinusoidal periodic signals, because overtones are modeled too.

The combination of (i) and (iv) makes it possible to largely separate the stochastic noise from the significant variations. The combination of (i), (ii) and (iii) allows the user to extract the maximum amount of significant information, "hidden" in the record. In addition, all tuning is done by the algorithm, which makes the method user-friendly and more objective. On the other hand, the algorithm does assume that the 'true' record is periodic, which may not be true. A violation of this assumption may bias the final result. However, this assumption is already
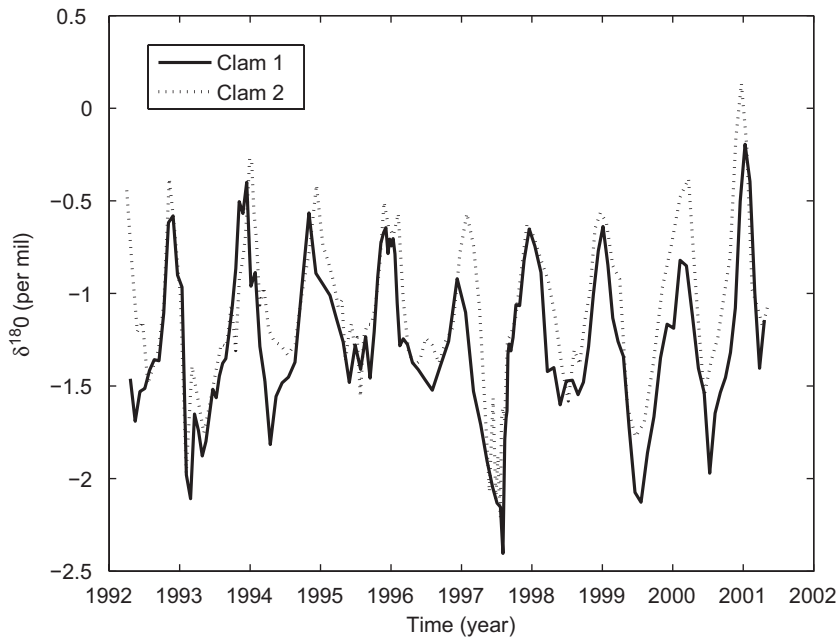
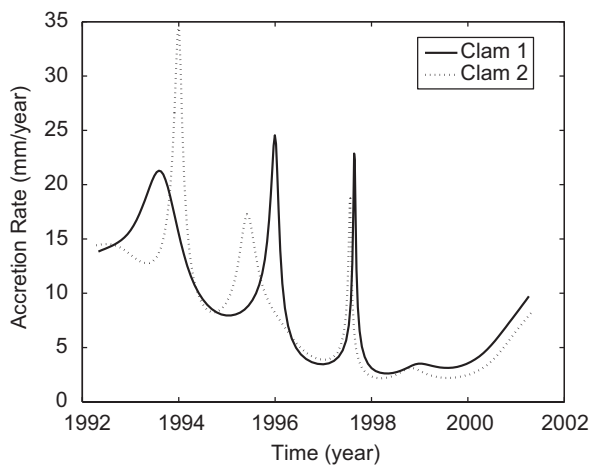**Fig. 9.** δ^{18}O-records from clams 1 and 2 after time base correction.

Fig. 10. Estimated accretion rate in clams 1 and 2.

much less stringent than assuming the full time series to be known. Moreover, the general strategy proposed here is not limited to periodic signal models, although gathering initial values for arbitrary models can become a hard task.

The method has been exemplified on two records of δ^{18}O, measured in clams. Both clams lived in the same environment, so the time series of δ^{18}O could be used to check the robustness of the method. After independent optimization of the two records the correlation between both records increased from 27% to 84%. Furthermore, the correlation between the independently estimated accretion rates was 63%. This could indicate that also the accretion rate is changing with varying environmental conditions in a deterministic manner.

With these arguments and results, we hope to have convinced the reader of the importance of a correct time base reconstruction. As seen in the clam example, this is necessary for intercomparing records, but it is also critical for any quantitative understanding of processes underlying the proxy data. In paleoclimate studies this issue may have considerable influence on the actual potential of e.g. multi-proxy approaches. In applications where multiple elements or isotopes are measured in the same sample, specific signal models may be used for each proxy (i.e. not necessarily periodic as in Eq. (1), but e.g. exponentially decaying) but all proxies must have the same time base model.

References

Abramowitz, M., Segun, I.A., 1968. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover Publications, New York, 1046pp.
Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control AC 19 (6), 716–723.

Brüggeman, W., 1992. A minimal cost function method for optimizing the age–depth relation of deep-sea sediment cores. Paleoceanography 7 (4), 476–487.

de Brauwere, A., De Ridder, F., Pintelon, R., Elskens, M., Schoukens, J., Baeyens, W., 2005. Model selection through a statistical analysis of the minimum of a Weighted Least Squares cost function. Chemometrics and Intelligent Laboratory Systems 76, 163–173.

De Ridder, F., Pintelon, R., Schoukens, J., Gillikin, D.P., André, L., Baeyens, W., de Brauwere, A., Dehairs, F., 2004. Decoding nonlinear growth rates in biogenic environmental archives. Geochemistry Geophysics Geosystems 5 (12) number Q12015.

De Ridder, F., Pintelon, R., Schoukens, J., Gillikin, D.P., 2005. Modified AIC and MDL model selection criteria for short data records. IEEE Transactions on Instrumentation and Measurement 54 (1), 144–150.

Dierckx, P., 1995. Curve and Surface Fitting With Splines. Clarendon Press, Oxford, 285pp.

Fishman, G.S., 1996. Monte Carlo: Concepts, Algorithms, and Applications. Springer, New York, 728pp.

Fletcher, R., 1991. Practical Methods of Optimization. Wiley, Chichester, 436pp.

Gillikin, D.P., De Ridder, F., Ulens, H., Elskens, M., Keppens, E., Baeyens, W., Dehairs, F., 2005. Assessing the reproducibility and reliability of estuarine bivalve shells (*Saxidomus giganteus*) for sea surface temperature reconstruction: implications for paleoclimate studies. Palaeogeography Palaeoclimatology Palaeoecology 228, 70–85.

Ivany, L.C., Wilkinson, B.H., Jones, D.S., 2003. Using stable isotopic data to resolve rate and duration of growth throughout ontogeny: an example from the Surf Clam, *Spisula solidissima*. Palaios 18, 126–137.

Lisiecki, L.E., Lisiecki, P., 2002. Application of dynamical programming to the correlation of paleoclimate records. Paleoceanography 17 (D4) article number 1049.

Martinson, D.G., Menke, W., Stoffa, P., 1982. An inverse approach to signal correlation. Journal of Geophysical Research 87 (B6), 4807–4818.

Paillard, D., Labeyrie, L., Yiou, P., 1996. Macintosh program performs time-series analysis. Eos Transactions AGU 77 (39), 379.

Pintelon, R., Schoukens, J., 1996. An improved sine-wave fitting procedure for characterizing data acquisition channels. IEEE Transactions on Instrumentation and Measurement 45 (2), 588–593.

Pintelon, R., Schoukens, J., 2001. System Identification: A Frequency Domain Approach. IEEE press, New York, 605pp.

Rissanen, J., 1978. Modeling by shortest data description. Automatica 14, 465–471.

Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6 (2), 461–464.

Wilkinson, B.H., Ivany, L.C., 2002. Paleoclimatic inference from stable isotope profiles of accretionary biogenic hardparts—a quantitative approach to the evaluation of incomplete data. Palaeography, Palaeoclimatology, Palaeoecology 185, 95–114.

Yu, Z.W., Ding, Z.I., 1998. An automatic orbital tuning method for paleoclimate records. Geophysical Research Letters 25, 4525–4528.