Clim. Past Discuss., 8, C1717–C1753, 2012
www.clim-past-discuss.net/8/C1717/2012/

Climate of the Past Discussions

Interactive comment on "COnstructing Proxy-Record Age models (COPRA)" *by* S. F. M. Breitenbach et al.

S. F. M. Breitenbach et al.

breitenbach@erdw.ethz.ch

Received and published: 26 September 2012

[a4paper,12pt]scrartcl color,graphicx, epstopdf,todonotes hyperref longtable, lscape amsmath,amsfonts,amssymb [normalem]ulem

Summary of changes and response to referees' reports for manuscript **CP-2012-70** "COnstructing Proxy-Record from Age models (COPRA)"

S.F.M. Breitenbach et al.

26 September 2012

1 General response

Dear Editor,
thank you very much for giving us the opportunity to revise our manuscript. The two reviewers both have some issues with the novelty of the statistical methods employed, but both like the introduction of a strategy to include layer counting to improve the overall age model in question. Since we did not provide the software for the reviewers (reviewer 1 asked via the editor for a test version of COPRA), we understand why reviewer 2 has some issues following our line of argument. We have considered all questions of the reviewers and changed our manuscript accordingly, further improved it

and clarified our points. Below, we give detailed responses to the reviewers comments. We improved the manuscript by thoroughly revising sections that needed clarification. We also adjusted the abstract and the title (please note the "from") to transmit our message more clearly.
Thank you very much for your consideration. We also thank both reviewers for their detailed comments.
Yours sincerely,
Sebastian Breitenbach

2 Response to the first referee

(Original report cited in italics)

General Comments:
The authors present an age-depth modelling approach (COPRA) designed for climate archives, which age uncertainties are symmetrical and might provide the possibility to form annual layers. An outlier analysis is included and the program can deal with age inversions. During the age modelling the program translates age uncertainties of the proxy record into proxy uncertainties. The authors argue that this step is justified in order to obtain a time certain, 'true' time scale, which allows comparison with other records. In its present form COPRA is suited to determine age-depth relationships for stalagmites. The approach is tested on an artificial time series and on two unpublished stalagmites from the Georgian Caucasus and Southern Belize. To my opinion the outlier detection and age-depth model routine (MC approach) is not new and nothing special, as honestly admitted by the authors and should be improved in follow-up versions of COPRA. What makes their method suitable for publication in CP is the integration of laminar counted sections of the stalagmites into the age-depth model. In very recent literature laminar counted sections are included in stalagmite age models (Scholz et al., 2012, CPD, http://www.clim-past-discuss.net/8/909/2012/cpd-8-909-2012.pdf ; Dominguez-Villar et al., 2012, Quaternary Geochronology, http://dx.doi.org/10.1016/j.quageo.2012.04.019), each with their own stalagmite-specialised method. Therefore, establishing a general method, which is able to fit one or more floating laminar counted sections to dated depths, is valuable work and deserves publication. However, before I can advise to consider this manuscript for publication, the authors are kindly asked to address my points, listed below.

Specific Comments:
P2373; l 20-25: I really like the idea to translate the age errors into proxy errors. However, I doubt that it will be possible to compare records better with this approach. For me, the error bands produced in the according figures look similar to what I would expect if the proxy signal would be plotted with age errors. Unfortunately, there are relatively few studies available, which present their data in a time uncertain domain (one exception is e.g. Blaauw et al. 2007, Holocene, 17, pp 283-288). Both error bands, addressing either only proxy errors or only age errors, look pretty similar to my opinion. Related to this: If I understand your uncertainty-transfer method (P2380; l 2-9) correctly, I can not agree with the error bands shown in Figure 6 and 9. How is it possible that the error in the proxy is larger than the highest measured value (e.g. Fig. 9 at 36ka the upper limit exceeds the measured values by several tenth of a permil, the same with the low values of the lower error band right before the hiatus)? Please, explain this behaviour in more detail.

We thank the referee for the critical remark, pointing to the similarities between Blaauw's approach of using age uncertainties and our approach of using proxy uncertainties. Although, at a first glance, there might be similarities, our approach differs significantly from his.

Comparing different proxy records requires a common time scale. For example, most of the available time series analysis methods require that the time series share the same time scale. Only recent developments try to overcome such limitations (e.g., Rehfeld et al. NPG 2011). Therefore, translation of the age error into a proxy error is an important step toward reliable time series analysis. We understand that this point of view is new and will cause further concerns by our colleagues. But we are convinced that for certain applications this procedure will allow for more sophisticated analysis and interpretation, considering the uncertainties of the archives.

More technically, even though our representations of proxy records after applying the COPRA methodology (e.g., Figs. 9 and 11) closely resemble the representation of age errors in Fig. 2 of Blaauw et al. 2007, Holocene 17, pp 283-288, the two are quite different. In our case, the errors have been transferred to the proxy axis using conditional probability ($Prob(A|B)$ where A and B are probabilistic events) whereas this is not so in the study by Blaauw.

We believe it is possible that the error in the proxy exceeds the highest and lowest measured proxy values because it takes into account 'all' the age uncertainties (even those that might be far away from the most probable age for that proxy). The final estimated uncertainty in the proxy using our approach incorporates all the age uncertainties in a non-trivial, nonlinear manner and hence it might exceed the highest proxy measured values depending on the actual age uncertainties, the variability of the proxy, the depth-age relation, etc. We would also like to note that the 95% confidence bounds used throughout the manuscript were constructed using $\pm$2-sigma deviations from the median curves. This is now mentioned explicitly in the relevant figure captions.

A further (small) note on this topic (P2392, l 28): I agree, that with large age (or proxy) errors it is not possible (and shouldn't be tried) to make some statements about the

high frequency variability of the proxy or to compare the record with other records on short time scales. However, one can say with confidence that there are high frequency variations. This is more than 'cannot say anything with confidence'. If you have a laminar counted section it is even possible to make statements about the frequency of the proxy variations, although the âĂŽtrueâĂŽ age of the section is not well constrained and comparison with other records is still difficult.

We agree with the reviewer in that layer counted intervals allow confident statements on high-frequency variability, which is in line with our argumentation. We feel that one can say that there are high-frequency variations by looking at the proxy signal (in the depth domain) but the validity of this statement is restricted to the depth domain. To be able to make a confident comment about the high frequency variations in the time domain one has to be able to resolve it first (with either extremely precise dating or with additional laminar counted sections).

P 2378, l 17: 100 MC simulations seems to be too less for me. Is there an explanation why this number is used? I recommend to use at least 2000 as default value. This is a number Efron and Tibshirani (1993, An Introduction to the Bootstrap. Chapman & Hall, New York) suggested for MC simulation in order to suppress simulation noise. Is this small number the reason, why both age-depth realisations in Figure 6a provide such large differences at the oldest part of the artificially constructed archive? The inclusion of the layer count section at the top of this stalagmite should not have a strong influence on the errors at the other end of the stalagmite. The same enlargement of the error envelope is present at about 100 mm distance from top in YOK-G (compare Fig 9a and b).

(a) Thank you for pointing this out. To illustrate the importance of the number of MC simulations we include the Fig.1 of this response letter, which clearly indicates that the minimum number of MC runs needed for a reliable estimate in our case is around 2000.
(b) The low number of MC simulations was probably the most likely reason for the difference in the oldest part of the constructed archive in Fig. 6a of the manuscript as well as for the oldest part of the age model for YOK-G. Following the suggestion of the reviewer and the evidence from Fig.1, we use 2000 MC simulations throughout the revised manuscript. The new figures in the manuscript do not show this anomalous behaviour.

P 2384; l 15-21: I agree that this is one possibility to fit the floating laminar counted part. However, my concern with your approach is that the result is biased by the kind of interpolation of age-depth model A. e.g. if the laminar counted section reveals a relatively linear growth history, but the user decided to apply a spline between dated depths. An alternative would be to calculate the least square only between the laminated section and the dated depths. I admit that this causes some trouble when no age is measured within the laminar counted section. Do you have any arguments, why you choose an integration procedure as described?

(a) We follow the reviewers concern that the results might be biased by the choice of interpolation of depth-age model A. However, if the laminar counted sections and the radiometrically dated points are measured accurately – and if the laminar counted shows a relatively linear growth history, the dated points should also be relatively linear, irrespective of the kind of interpolation used (to a certain degree of mismatch). In other words, if the accuracy of the radiometric dates is low the large errors allow for many different interpolations to be realized fulfilling the stratigraphic requirements. This

results in larger error margins (without layer counting). If a segment with counted layers is included, this segment will markedly improve the final error estimate, again regardless the chosen interpolation method.
(b) Initially, we had considered the idea of calculating the least squares between the laminar counted sections and the dated points – however this has the obvious problem of no age being measured at all laminar counted depths. We finally chose to use the approach described in the manuscript as it provided several advantages:

1. It allows for a certain level of objectivity to positioning the laminar counted section within the primary age model (barring dependence on certain parameter choices and interpolation methods).
2. It also allows to represent the laminar counted sections – which are measured with an error in depth – as a set of measurements with errors only in age (and hence could be then trivially included alongside the dated points).
3. It inherently contains the fundamental idea that the dated points and the laminar counted sections were independent measurements on the same archive. This means that the dating errors of the layer counting are uncorrelated to the dating errors of the radiometric dates. This independence is a huge advantage that allow us to use this procedure.

Related to this point: Do you take the age errors of A and B into account for the least squares calculation? Do you assign an error to the vector A0 and do you add this error to the laminar counted error? At this point, I don't request the authors to modify their code according to the aforementioned question, but I think that it is necessary to briefly mention these points and maybe they could be considered for further stages of development of the software.

We thank the reviewer for raising this critical point.
To answer the questions, (a) no, at present, we do not take into account the errors of Age modes A and B for the least squares calculation. We only consider the mean/median age model (either one).
(b) Thus, we also do not assign an error to the vector A_0.
We acknowledge that the incorporation of the age model errors for the least squares estimate, and thus assigning an error to vector A_0 (and thereby incorporating this error as well into the final positioning of the laminar counted section within the dating table) is crucial for a more reliable chronology. We now mention these issues in the last portion of Sec.2.5 of the revised manuscript.

Section 3.1.1: In your description of the artificial record construction (and Fig. 6a) it seems that the midpoint of the dating uncertainty always reflects the 'true', prescribed age. It is highly unlikely that all dating points of a real archive are perfect like this. Other age-depth modelling studies construct their artificially constructed age models in a more reliable way (first there is the growth history and than the âĂŽmeasuredâĂŽ ages at certain depths are randomly determined âĂŽÄì often with large deviations from the âĂŽtrueâĂŽ growth history, see e.g., Scholz and Hoffmann (2011) âĂŽÄì although the randomisation of ages do not have to be as sophisticated as described therein). To my opinion it would be fair enough to test COPRA on an artificial age-depth model, constructed with dating points that are not as perfect as in the present manuscript version. I think in this way we all could learn more about the performance of COPRA.

We agree with the reviewer that the chosen example in Fig. 6a presents a simple and unproblematic depth-age modeling case. However, this chronology was intended to

illustrate the effect of layer counting on the confidence intervals, and not necessarily to test the performance of COPRA quantitatively. Considering the YOK-G chronology (Fig. 5 in the manuscript) we observe that, while it might be unlikely, it still is possible to find linear and straightforward accumulation histories like this test case.
We do not consider to introduce a further competitor for depth-age modeling. Our aim is to introduce a framework which can include different age-modeling schemes. Therefore, the selection of the examples is based on the demonstration of the complementary features of such a framework, as the age reversals, or hiatus detection.

Section 3.1.2, 3.1.3: This is the first publication of the dates of both stalagmites. Therefore, it is essential that the Th/U ages are explicitly given in a table (not necessarily in the paper, but at least as supplementary material). I would appreciate if you can state if the given errors are 1 sigma or 2 sigma. Usually Th/U dates are given with 2 sigma uncertainty, but with respect to figure 8 a lot of the 100(?) MC realisations (more than five) plot outside the given age errors. Therefore, I assume you show 1 sigma errors in your figures.

We fully agree with the reviewer concerning the publication of the dates of the stalagmites. Details on the radiometric dates are now included in supplementing material.The dating errors of the two stalagmites used in this study are provided as 2-sigma and thus, the visualizations of the age models and the final proxy records also use 2-sigma confidence bounds. Please note however, that the input file with the dating table must be given with 1 σ errors in order to let COPRA run correct calculations. From the modeling point of view, however, the issue of whether the input and output errors are 1- or 2-sigma is redundant because the Monte Carlo simulations require only the value of the standard deviation, *viz. sigma*. This is true also for the construction of the proxy records from the age model.
The number of MC realizations that plot outside the individual error margins of a dated age is dependent not only on the error margins of the date itself, but also on the align-

ment of previous ages, and their errors, due to the monotonicity requirement.

Section 3.2.1; 3.2.2; 3.2.3: Please provide information on the parameters used for the construction of the age-depth models (e.g., Number of MC simulation, kind of interpolation, . . .)

All examples in the revised manuscript have been run using 2000 MC simulations and cubic interpolation with the exception of the layer counting where we used linear interpolation. The reason for this is the large number of iterations (and thus model run time) needed to reach a monotonous chronology using cubic interpolation.

Technical Comments

P2371; l 4: Unfortunately it is not entirely true that âĂŽavailable modelling algorithms do not allow incorporation of layer counted intervalsâĂŽ. Very recently Dominguez-Villar et al. (2012, Quaternary Geochronology,

http://dx.doi.org/10.1016/j.quageo.2012.04.019)

published an approach.

We thank the reviewer for pointing to this new paper by Dominguez-Villar et al.; this paper was unknown to us during submission and we refer to it now in the new submission. Nevertheless, their approach tackles the problem of combining the U-series chronology and layer-counted intervals in a different manner.
In their study, Dominguez-Villar et al. anchor a floating layer counted chronology using a least squares fit of a linear relation between the two. Even while we adopt a similar perspective of least squares fit, we use it to estimate the minimum 'distance' between the radiometric age model and the layer counted age model. This allows us to create a final dating table that combines both the radiometric dates (and uncertainties) and the

anchored layer counted dates (and their uncertainties) for a last round of age modeling to get the final age model. This is a critical difference between their approach and the one described in our study.

P2373; l 8-9: It is not true that StalAge is able to construct age models for lake sediments and ice cores, since StalAge accounts only for Gaussian age uncertainties and calibrated 14C ages are not Gaussian distributed.

It is correct that StalAge is not suited yet to build non-Gaussian 14C-based chronologies. Other algorithms, like Oxcal, are especially adopted for this task. We adjusted the text accordingly. As it stands, COPRA is best suited for U-series (or better, Gaussian-distributed) chronologies, but we point out that it is planned to include non-Gaussian uncertainty calculations in a subsequent COPRA version, which will broaden the spectrum of applicability for COPRA.

P2374; l 6: Please, state already in this line, which kind of interpolation COPRA provides.

We state the different interpolation possibilities in this item (linear, spline, cubic).

P2375; l 19-24 and P2391; l 23): I would prefer, if you could remove all text passages where you mention that COPRA can build age models for 14C dated archives. In general, calibrated 14C ages are highly asymmetric and COPRA can, at its present state, not deal adequately with such uncertainty distributions.

All text passages are adjusted in this regard.

P2376; l 15-17: This sentence sounds somewhat weird.

This sentence has been clarified.

P2377; l 1: Please, be more specific and explain what âĂŽproper treatmentâĂŽ means.

'Proper treatment' here is meant in the context of establishing depth-age relationships. The scientist needs to prepare the ages in a way so that a) a monotonic depth-age relationship can possibly be established and b) additional knowledge s/he might possess about the ages are reflected. This would mean that dates that s/he does not trust would be either removed (if seriously false) or if the analytical error does not reflect the uncertainty felt by the scientist the error margin should be increased.

P2378; l 4: Please, delete âĂŽsmallâĂŽ since the error has not to be small.

We deleted 'small'.

P2378; l 11-12: It seems to me that you speak in both lines of the depths of the proxy profile. Therefore, I suggest to use already in this line another index (not âĂŽi, as this is attributed to the depths of the age determinations). Maybe it is appropriate to use âĂŽjâĂŽ instead as you did on the next page.

The use of the index j indeed seems to be more appropriate here. We have changed the index variable to j, as suggested.

Section 2.3.2: Formally, I do not see a difference in the terms âĂŽoutlierâĂŽ and âĂŽnon-tractable reversalâĂŽ. Both describe the same problem in the same way. Please, rearrange this section accordingly or provide more information how to differen-

tiate between both anomalies.

A reversal is the observation that the monotonicity is violated. A tractable reversal still allows for construction of a MC age model, while a non-tractable reversal does not. The non-tractable reversal describes a situation where the datapoint in question does not overlap with the the neighboring data points. This could be either caused by an outlier, or by one of the neighboring data points, which must be evaluated by the user. It is important to differentiate between reversals and outliers. An outlier is something far away from a likely situation or which deviates significantly from the rest of the data (here: from the general depth-age relationship), mostly caused by measurement issues. It is possible that an outlier causes an age reversal. However, the opposite is also not necessarily true: A non-tractable reversal does not necessarily have to be an outlier.
Because an outlier does not need to be a reversal, the simple reversal check is not enough. Therefore, in future versions of COPRA, a robust automatic statistical outlier test will be incorporated into the modeling procedure. At the moment, in COPRA outliers can only be detected by human guidance.
The treatment remains subject to individual analysis: If the data point is identified as an outlier, it must be removed from the dating sequence. Alternatively, if the data point is not clearly identified as outlier, it remains in the sequence and its error margins are to be increased. This is the "treatment" we refer to on page 2381, l14. If the data point is ok, but a neighboring point caused the violation of monotonicity, that point instead must be handled (again, either removing it (given additional information is at hand) or by increasing the error margin).
We have clarified the explanations on reversals and outliers in Sect. 2.3.

P 2385; l 4: Please delete '(not shown in Fig. 3)'.

We deleted '(not shown in Fig. 3)'.

P 2386; l 15: Why do you ascribe an error of 1mm at the top? This reads strange, since it suggests that you do not know were the top of the stalagmite is.

The reviewer correctly points to an unrealistically large uncertainty for the top/ beginning of the synthetic layer counts. Normally, the distance can be measured highly accurate and an uncertainty is much less than 1 mm (probably in the order of 0.05-0.1mm, depending on the stalagmite surface). Therefore, we changed the synthetic dataset input and attribute only 0.1mm as the uncertainty for "how good do we know the starting point for the counting", plus an additional 1% of the distance to the first layer counting depth, so that the error increases with depth. This is the natural uncertainty one would encounter if the layer counting was performed repeatedly/ along different (albeit parallel) depth axes.

Fig. 5: Why seems the red layer count started at about 28 mm and stopped at about 53mm? The section where counting is possible should not change - only the number of counted layers/years could change.

The counted interval has been plotted correctly. To explain why the depth is different for each counting track, we give here some details on the counting process:
Two people counted from the top downward, assigning counting years and recording corresponding depths. Consequently, the depth is different in each individual count because layers are assigned subjectively and hence each "counter" will find a different depth for a certain layer. The final counted interval will present itself as a time axis (years) (which remains fixed), assigned with each "counter's" depth (accumulated depth). One example: If the top of the stalagmite is 2006AD, a counted layer "number 40" would correspond to 1967AD, but would be found at two different depths by two

individual counters.
The standard solution is to count the laminated segments multiple (at least three) times, so as to get an estimate of the expected depth a year is assigned and its standard deviation. As only two layer counting runs are currently available for YOK-G, we used the *maximum absolute deviation* of the two layer counting samples in order to have an estimate of the depth-assignments in the layer counted interval. We have added a paragraph in the manuscript in the introduction of YOK-G as a dataset to illustrate, how layer counting data should and can be prepared.

3 Response to the second referee

(Original report cited in italics)

The authors present a manuscript describing procedures to derive an age model with associated uncertainties for U-Th dated speleothems. A âĂŽsoftwareâĂŽ (i.e. a MatLab code) for these procedures is announced to be provided. An interesting feature of the presented procedures is the incorporation of additional chronological information, such as layer counting. This can greatly improve an age model. The authors recognise the importance of error margins for an age model and provide a code that does a Monte Carlo (MC) variation on linear, spline or cubic interpolation.

General comments:

The COPRA code itself is new, but the basic idea behind (i.e. design a code to derive *an age model for speleothems based on U-series dates with MC-based confidence levels) is not. This manuscript follows other recent publications on the same topic (Scholz and Hoffmann, 2011, QG; Scholz et al., in press, QG; Dominguez et al., in press, QG; Hercman and Pawlak, in press, QG). The radiocarbon dating community has also developed various tools for distance-age modeling (e.g. Heegaard et al., 2005; Bronk Ramsey, 2008; Blaauw, 2010). Novel to some extend is the realisation how to incorporate layer counting as additional constraint (although see Dominguez-Villar et al., in press, QG). The title and abstract should be focused on the aspect of including layer counting rather than selling the code as the solution for âĂŽANY and ALLâĂŽ proxy records.*

We thank the reviewer for pointing to the very recent and important works by Scholz et al., Dominguez et al. and Hercman & Pawlak (all in press). All these studies have not yet been available during writing our manuscript and we have, therefore, not been aware of them. This parallel research on the same problem (depth-age modeling) emphasizes the importance and the needs of considering uncertainty in age models and proxy records.
The reviewer points to the fact that we make use of established statistics, rather then implementing a new way to estimate uncertainties. Although it is true that we build upon existing ideas of Monte Carlo-based age modeling techniques, there are also several novel features in our approach. Our approach differs from other '*age modeling*' procedures in that it translates dating uncertainties to proxy uncertainties (and we refer to this more as a '*proxy uncertainty modeling*'). This is one step towards strict comparison between different time series with a common fixed (error-free) time axis.
As acknowledged by the referee, one of the very new features is the inclusion of floating (layer counted) chronologies into the age modeling procedure. This goes in the same direction as very recently presented by Dominguez-Villar et al., in press, but whereas these authors estimate an uncertainty value for the entire (layer counted, inserted into an U-series chronology) record, we calculate an error estimate for each proxy value. Thus, both methods are heading in the same direction, but the solutions are rather

different.
We adjusted the abstract and introduction in order to more carefully stress the critical advantages of COPRA.

Generally more information about the interpolation itself is needed. What spline or cubic interpolation is used, how are parameters constrained? And on what basis is the interpolation method chosen? Any constraints, advice, guidance? Please show for your examples potential differences for the different possible interpolation options. Also, how do different choices for parameters for spline or cubic interpolation affect the results and how is a decision made? In its current version the manuscripts leaves too many questions open and many aspects of the age modeling seem to be somewhat random. This needs to be clarified, especially in the light of a claim for a reproducible approach.

The interpolation method used should avoid abrupt changes in the depth-age relationship, which happens if e.g. linear interpolation is used. Our goal in this study is to present a framework, wherein the specific interpolation scheme is trivial and hence we allow the user to choose different interpolation methods. The different interpolation schemes each have advantages and disadvantages, depending on the data. The reproducibility is ensured by logging all relevant choices and inputs in the metadata.

The proof of concept is not yet convincing. A synthetic data set should be used to address critical issues and show strengths and weaknesses of such an approach. However, the authors fail to do so. The example using synthetic data is not demanding at all. No âĂŽoutlier treatmentâĂŽ is really addressed, although this is one of the critical aspects. This can be seen by one of their stalagmite examples (TSAL-1) where deleted âĂŽoutliersâĂŽ are not necessarily outliers and could / should not be deleted.

We agree that a general comparison and benchmark testing are important in order to characterize the performance of algorithms. It is, however, beyond the scope of this work to do so and this point deserves more attention in a separate study. The synthetic example we use here is used to explicitly illustrate only the effects of layer counting on the confidence bands.
We regard the two samples U1 and U2 as outliers because both dates are far off ("too old") the general growth trend, compared to the rest of the data. We included them in this example to show the reversals they cause and how a monotonic depth-age relationship is achieved when deleting these data points.
The older date (U2) does not fit into the general growth trend of the other dates bracketing it. If we assume the previous date to be erroneous instead, we can establish a monotonic depth-age relation including the next (above) date. Thus, a decision has to be made how to deal with one of these two dates. If no sample can be clearly identified as outlier, increasing the error margins of these samples is feasible. The sample U1 was excluded based on a note from the laboratory that this date was likely contaminated, and by crystallographic inspection of TSAL-1: except for the hiatus, no evidence is found that would support the changes in growth rate, that would be inferred if the age would remain in the chronology. We argue that identified outlier must be eliminated and that exclusion or error-enlargement of other suspicious samples (that cause reversals) must be based on multiple lines of evidence and that the final decision is that of the scientist.

I would generally criticise that - although the authors do mention that âĂŽoutliers cannot simply be deleted without further evidence - the examples and the design of COPRA implies that deleting outliers is the appropriate standard procedure. Alternative handling / treatment of âĂŽoutlierâĂŽ is not really considered in the manuscript. For example, all cases where data are deleted should be compared to an alternative by increasing error margins instead of deleting.

Outlier removal is the only appropriate choice, if identified as per definition of 'outlier'. We do not agree with the referee in this respect. Increasing errors for outliers would be a very questionable procedure without any physical motivation.
There are however situations where 'age reversals' might allow to increase errors systematically or subjectively. We revised the reversal/outlier section and added a statement focussing on the treatment of outliers and reversals to make this more clear.

Furthermore the authors do not follow their own statements and do not check / treat neighbouring data points of âĂŽÄôoutliersâĂŽÄô, which would be important for the example TSAL-1. The authors vaguely leave many decisions to âĂŽÄôthe experienced researcherâĂŽÄô and fail to define what is meant by âĂŽÄôappropriate data handling / modificationâĂŽÄô.

We apologize if we have not been clear enough. We have added a statement concerning the treatment of outliers that should make this more clear (Section 2.3.3). See above for the answer regarding the TSAL-1 example.

The functionality (and user-friendliness) of the code could not be tested / verified for this review because a final âĂŽuser-friendlyâĂŽ version was not available.

We apologize that the software have not been available for the reviewers. We have now uploaded the latest version of the software to our toolbox website http://tocsy.pik-potsdam.de where it should now be available for the reviewers and the public.

Detailed comments:
line 5/6: The modeling and âĂŽinteractive handling of outliers and hiatusesâĂŽ are

separate issues. It is not the modeling that allows the âĂŽinteractive handlingâĂŽ. Interactive handling is data modification prior to modeling.

We thank the reviewer for pointing this out. We have changed the sentence to: "We present an approach that allows automatic detection and interactive handling of outliers and hiatuses."

line 10: This statement could not be tested/ verified. There was no final, easy to use version available for this review. Also, the easy to use version with gui will require a MatLab licence which should be mentioned.

The software is now available on the TOCSY website: http://tocsy.pik-potsdam.de.

Introduction

p2371 line 17: Generally a sampling position is given with distance to a reference point. Can be depth (for stalagmites distance from top) but also distance from bottom.

We agree that the sampling position is always with respect to a reference point which could be either the topmost or the bottom-most point of the core. However, for simplicity and because in general in geology the stratigraphic order is related to depth, we use the term "depth". Nevertheless, the proposed method is working independently of using the term distance or depth. We have now added a corresponding sentence in the introduction.

p2372 line 11: What is precisely meant by âĂŽdirect network analysesâĂŽ? I could not find a description of this term in the cited reference.

The present spatio-temporal coverage of proxy records would allow for a "meaningful and conceptually simple" network analyses using the theory of spatially embedded

complex networks (see for example Donges et al. (2009): The backbone of the climate network. Europhys. Lett., 87; Newman (2010) Networks, An introduction. Oxford University Press; Watts&Strogatz (1998) Collective dynamics of 'small-world' networks. Nature 393, 409-410). To improve understanding, we deleted the term "direct" in the corresponding sentence. Details on network analyses can be found in the cited paper of Rehfeld et al. (2011) and the very recently published study by Rehfeld et al. (Clim. Dyn.), which is now also referred to.

p2372 line 15-23: This sounds as if the authors introduce the concept of an age model! In line 23 they even claim that the presented method can construct an "absolute" time scale for ANY and ALL proxy records. That sounds quite bold and I pretty much doubt that. The whole paragraph should be deleted.

We deleted the phrase "any and all' from the sentence. However, we argue that: (i) the construction of an "absolute" (i.e., error-free) time scale as discussed in the manuscript is possible, (ii) it is generally applicable to most proxy records, and (iii) it will facilitate proxy record inter-comparison as well as palaeoclimatic complex network studies. Further, to avoid future ambiguity, we now choose to use the term "precise" time scale rather than our earlier choice of "absolute time scale".

p2372 line 25: Why do you change here from depth (or better distance) - age relationship (as correctly described in p2371 line 20) to age-depth relationship? I suggest to use the term distance-age model, relationship etc throughout the whole manuscript. Generally, looking at a stalagmite: every distance has an age, but not every age has a distance on a stalagmite. So the distance is the primary information and we measure (or assign/model) an age to it.

We thank for the hint that we have been inconsistent. We use the common term 'depth-age relationship' throughout the manuscript.

p2373 line 1: This statement is misleading. There are other approaches which also assign uncertainty to the age model.

We changed this sentence to: "Most available approaches use the mean or median of the age model to construct the final proxy record. This leaves a disjoint between the errors of their constructed age models and the final proxy uncertainties."

p2373 line 6: The authors do not adequately describe potential of other already published age model approaches by stating âĂŽ...helps to interpolate...âĂŽ. As far as I can tell, COPRA is largely following previous publications such as StalAge but uses different interpolation(s) and has a different philosophy about data handling.

We apologize if we have given the impression to not adequately appreciating existing age model approaches. We have now modified the corresponding text and added important features of these approaches.
However, we disagree that we follow the mentioned previous publication. The development of COPRA goes back several years and was already used for several publications (e.g. Adkins et al. 2011, AGU San Francisco: Holocene History of δ^{18}O and Grayscale from a Stalagmite in NE India, with Implications for Monsoon and ENSO Variability; Marwan & Breitenbach 2007, EGU Vienna: Detection of Climate Transitions in Asia Derived from Speleothems; Marwan et al. 2010, EGU Vienna: Synchronous climate transitions during the Holocene in Asia derived from speleothems; Marwan et al. 2012, EGU Vienna: Reconstruction of uncertain age-depth relationships). Hence, the development was either earlier or in parallel to StalAge, but not yet published (an earlier version of the software was submitted to Journal of Speleology but is still in revision stage).

StalAge and COPRA use different interpolation schemes. Whereas StalAge uses a piecewise linear interpolation, leading to abrupt changes at the dating points, COPRA offers the use of continuous interpolation, and, depending on the interpolation algorithm, allows for a smooth depth-age relationship without abrupt changes. Moreover, with COPRA we introduce the concept of translating age errors into proxy value uncertainties, which is a novel feature most other age modeling approaches do not provide.
Moreover, the used Monte Carlo technique was not invented by the authors of StalAge (or another age-modeling approach), but goes back until Enrico Fermi, Stanislaw Ulam, and John von Neumann in the mid of the last century (Andrieu et al. (2003) An introduction to MCMC for machine learning, in: Machine learning, Springer).

p2374 line 25f: Following above statement - this is a bit overstating a novelty. The code itself is new, but the idea behind (design a method to derive an age model with MC-based confidence levels) is not.

We are unsure how to interpret this comment by the reviewer. When just looking at the Monte Carlo simulation, COPRA is as novel as StalAge or ModAge (both published in Quaternary Geochronology). The editors and reviewers of Quat. Geochron. found both methods novel enough to be publishable. We are a bit surprised that COPRA should not be novel enough, in particular as the reviewer correctly stated that we combined layer counting with radiometric age modeling and transferring age errors into proxy errors – two concepts which are novel indeed.
We have to explicitly state that COPRA is not reimplementing or just copying the ideas presented in Scholz & Hoffmann (2011) or Hercman & Pawlak (2012) as the reviewer is insinuating. In contrast, COPRA is not only an approach of creating a depth-age model, but it is a more general framework, allowing also data preprocessing (outlier exclusion, reversal treatment etc). It has such generality that COPRA might be perhaps extended in the future to include also the StalAge and/or ModAge algorithms for the specific task of depth-age modeling! We have added a corresponding sentence in the conclusion

and modified abstract and introduction to make it more clear that COPRA is a more general framework.

Methods

There are a lot of repetitions between âĂŽintroductionâĂŽ and âĂŽgeneral remarksâĂŽ. This can be substantially shortened.

We have now rearranged the methods and introduction sections and tried to shorten repeating explanations. For example, we have moved the description of the different dating possibilities (absolute dating, relative dating) into the introduction.

p2375 line 3: Why are proxy data needed for an age model? The age model (i.e. assigning an age to distances of the stalagmite) does not depend on proxy data. Actually, the age model itself is some sort of a proxy since it provides timing of growth phases, changes of growth rate etc.

We thank the reviewer for pointing this out. We changed the sentence to:
"For age modeling and subsequent assignment of proxy uncertainties in an absolute time frame, two datasets are needed. One includes dated points, and another proxy values, each with their respective depths from the stalagmite top (or base). Additional information on marker layers (e.g., hiatuses), and other specific information might be provided in a third file."
The reviewer correctly noticed that an age model can serve as proxy itself, but we point out again, that in COPRA, after a first depth-age relationship, we go a step further and include the proxy data to translate the uncertainty from age to proxy. This has been clarified in the manuscript.

p2375 footnote: I agree with the footnote, but why do the authors use the term âĂŽÄôabsoluteâĂŽÄô dating in some places (e.g. where the footnote is placed; p2372 line 12; p2375 line 20)? This could be part of the introduction making a case for not using this term.

Following the reviewer we omit the term "absolute" to avoid confusion. So we refer to "precise" when we mean "true", i.e. "without uncertainty".

p2376 line 25f: This is a vague and thus dangerous âĂŽÄôfundamental assumptionâĂŽÄô: What is âĂŽÄôproper treatmentâĂŽÄô? The manuscript only describes deletion of data points as a treatment, although it also mentions changing error margins. However, the alternative of changing error margins is not described further anywhere and all descriptions of data treatment in the text lead to the conclusion that âĂŽÄôoutliersâĂŽÄô have to be deleted when using COPRA (e.g. p2378 line 20).

We repeat here the response to a similar question by reviewer 1:
'Proper treatment' here is meant in the context of establishing depth-age relationships. The scientist needs to prepare the ages in a way so that a) a monotonic depth-age relationship can possibly be established and b) additional knowledge he/she might possess about the ages are reflected. This would mean that dates that s/he does not trust would be either removed (if seriously false) or if the analytical error does not reflect the uncertainty felt by the scientist the error margin should be increased.

p2377 line 7f: Reproducibility strongly depends on the user. Using COPRA requires decisions made by âĂŽÄôthe experienced researcherâĂŽÄô (select interpolation, delete data points...). This is the case for every age model and thus they are also reproducible in case ALL decisions and treatments are documented.

We thank the reviewer for sharing the same point of view with us that reproducibility is an important issue. Reproducibility strongly depends on the choices that the user makes in constructing the proxy record. This is why we mention in p2374 l13 that because COPRA exports all the necessary meta-data, it can provide “reliable and reproducible age modeling for proxy time series”. To our knowledge, other routines do not collect and store all decisions in a metafile, which leaves the degree of reproducibility to the user – the better he/she describes the used procedure, the higher the degree of reproducibility. However, often, this documentation falls short and thus reproducibility is not secured. COPRA however stores all the required information in a meta-data file, which can then be stored/published together with the model results. This is one way to ensure reproducibility.

p2377 line 14f: A few more references would be good in this paragraph. Using MC to derive uncertainties for non-linear systems is neither new nor novel. It is a good method to use and it is certainly ok to give an overview how it works. However, it could and should also be referenced.

The MC has been referenced in the manuscript. Please see also our response to reviewer 1 on page 3.

p2380 line 14: This is a key statement which unfortunately does not seem to be further considered!

We modified the the subsections on age reversals and outliers in order to clarify our line of argument. Outliers are generally to be excluded, while other suspicious samples which cause reversals might either be excluded or their errors enlarged. Complex statistical outlier analysis is considered to be included in future COPRA versions.

p2381 line 14: Please tell what you mean by âĂŽÄô...âĂŽÄôtreatedâĂŽÄô by the user

in an appropriate way...âĂŽÄô. As far as I read this manuscript this will largely be interpreted as : âĂŽÄôdelete data pointâĂŽÄô.

The section on age reversals and outliers and their handling has been revised thoroughly to clarify this point.

p2381 line 22: This is correct, so why do you not also highlight the neighbouring points as potentially problematic? You should also give guidance / examples for this scenario, e.g. using the example data set TSAL-1 or an appropriate synthetic dataset which includes all kinds of scenarios.

We decided against marking the neighboring points, simply to keep the graph less messy. We explain in Section 2.3 that neighboring points need cross-check if a point is highlighted.

p 2382 line 20: For a statistic a minimum of ten data points are needed and even then: only if it was expected that the values are scattering around a mean value a deviation could be detected. Growth rates can be hugely variable and a mean value is actually meaningless. A huge dataset would be needed to investigate whether there is a mean growth rate in the first place and then detect any significant deviation... Also, with few available data points this will be biased by the data density / distances between dated points. Thus, this approach seems not applicable to me.

We would like to emphasize that it will usually not be possible to detect hiatuses. As the reviewer states, indeed a minimum number of data points (here: entries in the dating table) are needed in order to be able to assess the variability of the growth rate, and too few dating points usually give rise to statistical uncertainty. We have added a note of caution concerning this in Sect. 2.4. Still, while the current test (which is based on an individual test whether one slope interval fits in with all the others) might be further improved in future, we believe that improvement of a routine highlighting potential growth rate changes will be possible and of interest. Within COPRA, it is also tied to the possibility to define growth interruptions manually, and manipulate their treatment (i.e. whether ages should be assigned within the hiatus, or whether the age model should be split).

p 2383 line 1: In what cases does COPRA identify a hiatus where there is none?

COPRA cannot identify hiatuses in a strict sense. It does, however, find growth rate changes, i.e. segments in which the growth was slower that in most others. The significance of the detection will highly depend on the size of the dating table (i.e. the number of slope estimates of the age model). Currently, growth rates are evaluated solely from the dating table. Therefore, if a segment of slow, but non-zero, growth is embedded in an otherwise rapidly growing archive, this will be marked as a potential hiatus if the dating density is high enough. False positives are, however, possible, especially when few dating points are available. COPRA supports the user in the identification of likely hiatuses, but the user has to decide if an interval of slow growth is a hiatus or not. We adjusted the manuscript text accordingly.

p 2383 line 17: You probably mean a more precise age model.

The statement "more precise" is justified here, though "realistic" has the same meaning in this context. We changed the text to "more precise".

Application

p 2385 line 20f: The synthetic dataset is disappointing. Such a dataset should be

used to highlight and demonstrate strengths (and potential weaknesses) of the model. All you show in Fig. 6 is a non-demanding dataset without any problematic feature. This is far from convincing. Although you describe such features (line 7), they are not presented. You need to show: dataset with âĂŽÄôoutlierâĂŽÄô (and no evidence for the data point to be suspicious) and how does COPRA work a) deleting b) keeping the âĂŽÄôoutlierâĂŽÄô. Same for a hiatus...

The synthetic dataset is used to illustrate the effects of layer counting and increasing age uncertainty. Here, we do not intend to perform a quantitative assessment of the algorithm within this manuscript. The necessity to perform a quantitative comparison is, however, now stressed in the discussion section. The algorithmic correctness was tested on synthetic datasets with reversals (as mentioned in Sect. 3.2.2), but were redundant due to the TSAL-1 example.

p 2386 line 26: How do you know that one age is suspicious due to contamination in the laboratory and how can you exclude this for the other data points (i.e. how realistic are the errors for the other data points)? Why would you report an age anyway if there was contamination in the lab?

Laboratory contamination cannot easily be detected. That the one reported sample was contaminated is known because it was caused by some mis-handling and reported from the laboratory. Therefore, it is a “human error" that introduced the contamination. This means that we can be faithful with regard to contamination concerning the other samples, which were not affected by the “human mis-handling". However, the best way to monitor the precision/accuracy of a labs results is cross-check within the frame of inter-laboratory calibrations. This is clearly beyond the scope of our paper however.
The contaminated date has been included in the discussion, as it is a good example how errors are increased due to such problems and is shown for illustration reasons only. However, the reviewer is right that in most, if not all, cases, known contaminated

samples are not used and published.

p 2387 line 6: You cannot assign a hiatus by fabrics. Fabrics indicate a discontinuity, but this could also be without a growth stop (e.g. a single short event). A hiatus can only be assigned by independent dating, the fabrics then help to place the hiatus at a certain distance.

As correctly noticed by the reviewer, one cannot assign hiatuses by fabrics alone. However, the crystallographic information can substantiate suspicions that at a certain spot a growth interruption occurred. The hiatus can hardly be dated directly, because it would be the “dating of no deposition". Therefore they are normally bracketed, although U-dates tend to show larger errors the closer they are at such hiatus. Hiatuses are detected as a reduction in growth rate to zero (or near-zero), with the consequence that unrealistic distances between proxy data points are found. This observation normally leads to closer inspection and this is where fabrics become helpful.
We changed the sentence to help the reader to better follow our line of thought.

p 2388 line 21f: As mentioned before, this paragraph needs more work. The synthetic example should cover and demonstrate handling of problematic features. You show just a nice non-problematic example.

Again, the synthetic example is meant to illustrate the tightening effects of layer counting on the error margins and the improvement of the chronology uncertainty. Please refer also to our comment on the synthetic dataset above.

p 2388 line 25: Why are these âĂŽÄôin fact outliersâĂŽÄô? The decision that these are outliers is based on what grounds? Why not also look at neighbouring points? IN FACT: a close look at Fig. 4 and Fig. 7 is a bit confusing. Maybe it is the quality of the

figures, but error margins do not seem to be the same on the two graphs. In Fig. 4 the age around 100 mm seems to just overlap with ages at greater depth with exception of 140 mm.

We would like to stress once more that an outlier in the context of this manuscript is a “dating point that is not consistent with the growth history of the archive” (cf. Sec. 2.3.2 of the revised manuscript). This can be evident either from the dataset itself or from additional sources of information. The second and seventh date of the TSAL-1 stalagmite are highlighted by COPRA because they cause age reversals, i.e. the 2σ error bars of consecutive points in the growth direction do not overlap. As stated above (in the answer to the first reviewer, page 13), these points were subsequently eliminated because of independent evidence. We assume that detrital ^{230}Th is the most likely reason for these samples to show apparent ages that are "too old" compared to the other dates. Since the lowermost sample in the dating sequence was contaminated in the laboratory and because the two dates discussed here are from the same batch, we suspect that laboratory contamination is the most likely reason for their conspicuous behavior. Additional dating is the best way to improve the dating sequence in this case. Figs. 4 and 7 show the same data, but the aspect ratio of Fig. 4 might make it look slightly different. The error margins of the date at ca. 100 mm does indeed overlap with older dates, though only very slightly. We also refer to the subsection on age reversals and outliers for further clarification.

So why is the point at 100 mm the outlier? Other evidence available? Interestingly in Fig. 7 the overlap disappeared. Anyway, it is mentioned that the age at base has a lab contamination issue, so how confident are you about error margins of the other dates? Maybe errors are larger. Also: on what base did you decide that the second point from top is the outlier and not the third? Both options are possible and it is not clear to me why âĂŽÄôin factâĂŽÄô this point was chosen. This highlights that other users would probably derive a different age model from the same data set (at least very *different uncertainties around the distances where you deleted points), questioning the reproducibility of this approach.*

Please refer to the discussion of the outliers in TSAL-1 directly above and in the answer to the first referee.

Discussion

p 2390 line 16/17: What do you mean by âĂŽÄôinformation-content-based approachâĂŽÄô? Is this meant to be something unique? Actually, all research should be information content based... And only accurate data should be used anyway, so this is no special limitation.

We adjusted the sentence where ‘information-content-based approach’ appeared. With COPRA we present not just another age modeling procedure but a framework that takes the age modeling one step further to derive proxy uncertainties and additionally allows integration of independent ‘floating chronologies’ to improve the final proxy reconstruction even further.

p 2391 line 18f: An increase in accuracy still needs to be shown, e.g. by an appropriate synthetic data set which returns an inaccurate age model without extra information such as layer counting.

We illustrate the accuracy of the resulting age-model in Fig. 3 below, where we compare different interpolation schemes with and without layer counting. Although it shows the performance of linear interpolation to be slightly better than the rest this is a bit of an expected result because the growth history of the constructed archive was, in fact, piecewise linear. A thorough qualitative assessment of the reconstruction capabilities should be performed separately.

p 2392 line 3f: This is an interesting approach. However, more discussion would be good here about the concept and implication to assign an age uncertainty into proxy uncertainty. To some extent it represents a running mean for points within an age range. However, in most cases a structure in proxy data is real if outside proxy data uncertainties, just the relative timing / age is uncertain. Thus, it might also be interesting to show different realisations of the proxy data on possible time scales.

We thank the reviewer for acknowledging our novel philosophy. We do agree that the usefulness depends on the application. If just the time of certain events is of interest, remaining in the depth-domain with age distributions is more appropriate, whereas when comparing different records, the translating to proxy errors is better suited.
We have modified the plot of the TSAL-1 proxy-record (Fig. 9) to show the different Monte-Carlo realizations.

Conclusions

p 2393 line 10: Actually, I think I missed the definition for âĂŽÄôoutlier definitionâĂŽÄô in the manuscript. As the example TSAL-1 shows, the outlier definition for this sample is at least questionable. This underpins your statement in line 6 and COPRA does not overcome the issue in its current version.

In the current COPRA version a statistical detection/treatment procedure for outliers is not included. Please see also section 2.3 in this regard. We plan to implement an automatic outlier detection in a further version of COPRA.

figure-1.pdf

Fig. 1. *Variance of the estimated proxy ages dependent on the number of Monte Carlo simulation.* We estimated 10000 depth-age realizations (matrices of possible age-values for each proxy depth) and bootstrapped this matrix 100 times picking N_{mc} realizations each time. The average age error is obtained as the mean over the standard deviations of the assigned ages at each proxy depth. While $N_{mc} = 100$ shows significant variance, the variance has leveled off at $N_{mc} = 2000$.

figure-2.pdf

Fig. 2. Results from running COPRA on a core with inhomogeneous accumulation history. Red circles indicate points which COPRA identified as outliers and which were subsequently removed. 2000 MC realizations and cubic interpolation was applied.

figure-3.pdf

Fig. 3. Comparison of the resultant mean overall age uncertainty (i.e., width of the confidence