Geochronology Discuss.,
https://doi.org/10.5194/gchron-2020-25-SC2, 2020

Interactive comment on "GeoChronR – an R package to model, analyze and visualize age-uncertain paleoscientific data" *by* Nicholas P. McKay et al.

Manfred Mudelsee

mudelsee@climate-risk-analysis.com

Received and published: 17 October 2020

Having myself worked for over thirty years in the scientific area targeted by this interesting paper – age-uncertain paleoscientific data analysis – it is my pleasure and privilege to add some remarks. Geochronology Discussions is one of the EGU's family of discussion journals, which are important to have since they allow our students to get a more intimate understanding of how the science of the paleoclimate and other disciplines evolves, how we paleoclimate researchers struggle for the truth. Inevitably, my contribution is biased towards own work, in particular my book on climate time series analysis (Mudelsee, 2014), to which for the sake of brevity here I sometimes refer to

as "Book", and of which – this is for those people interested in the history and also in who wrote first on a subject – the first edition appeared in 2010.

I find the paper well written, easy to read and understand, despite (because of?) the fact that nearly no equations are given. It shows that still one can speak in a rational manner about quantitative methods of data analysis. What I like most is the emphasis the paper puts on the usefulness of having available, in addition to a timescale, $\{t(i)\}_{i=1}^{n}$, where t is time or age, i is a counter, and n the sample size, also a set of simulated timescales $\{t^{*}(i)\}_{i=1}^{n}$. This set of simulation results is useful to have for the determination of the *full* errors of statistical estimations on paleoclimatic time series. That means, we have to take into account not only (1) measurement error of a climate proxy variable, $\{x(i)\}_{i=1}^{n}$, and (2) proxy error (which usually exceeds the measurement error by far), but also (3) timescale errors. The software GeoChronR presented by McKay et al. (2020) serves well those meticulous researchers embarking on a full error determination, since GeoChronR (1) comprises several timescale construction algorithms and (2) is written in the R language, which is gaining much popularity now in geosciences.

However, the meticulous researcher should be informed that the supply of $\{t^{*}(i)\}_{i=1}^{n}$ by GeoChronR is just the first of two milestones she or he has to reach. The second milestone – the development and utilization of statistical methods for processing $\{t^{*}(i)\}_{i=1}^{n}$ – this is in my view the major one. It has to be mentioned that, unfortunately, method development and theoretical derivations for uncertain timescales has never been high on the agenda of statistical science – at least this is my impression from the study of the literature in this science and from the news for members of the Royal Statistical Society (to which I belonged for about 20 years). In what follows I will mention some statistical estimation methods, which are described by McKay et al. (2020: Section 3 therein). It is the privilege of an external commenter in this discussion journal to not having to perform an exhaustive review but be allowed to pick out what is deemed interesting. I may add that I fail to find exhaustive the other comments in the interactive discussion (date of writing, 17 October 2020).

1 Correlation

A major problem imposed by uncertain ages occurs only if the timescales for X and Y are not completely dependent. A minor problem for completely dependent timescales may be that autocorrelation estimates, such as the persistence time for an AR(1) process on an unevenly spaced time grid (Mudelsee, 2002), may be influenced – but this can be safely ignored since this affects only block length selection in the bootstrap resampling for uncertainty determination, and block length selection is not very influential here (Mudelsee, 2014: Table 7.2 therein). A practical example for completely dependent (although uncertain) timescales is a marine sediment core, where on sample material from identical depths proxy measurements are done, such as oxygen (X) and carbon (Y) isotopic compositions.

For the case of not completely dependent timescales, McKay et al. (2020) present the binning approach. It should be mentioned that this works only in the presence of autocorrelation in the data (X, Y) generating process because then also time-distant points may "know" to some degree about the current point. In my Book (Mudelsee, 2014: Section 7.5 therein), I give bin width selection rules based on the persistence time estimates (for X and Y) and I show Monte Carlo simulation results obtained on artificial data in order to test the method. One results is that binning outperforms interpolation. A recent implementation of "my" binning approach for correlation estimation is the R software BINCOR (Polanco-Martinez et al., 2019).

As regards the assessment of the significance of the correlation coefficient (which, after all, is an *estimate* of the true but unknown population correlation coefficient), McKay et al. (2020) are correct in their assessment that the standard way is via a statistical test (of the null "population correlation coefficient equals zero") that assumes

normal shapes (for X and Y) and serial independences (of X and Y) – unfortunately! What the Monte Carlo experiments on correlation estimation, summarized in the Book (Mudelsee, 2014: Section 7.3 therein), teach us is sobering.

1. A confidence interval as uncertainty measure is superior to a P-value of a significance test because it bears more quantitative information. It allows you to compare two correlations, whether or not one association is stronger than another. A practical example is when you wish to construct a rank list of surface-air temperature measurement stations in Europe in terms of the strength of the correlation with the North Atlantic Oscillation.

2. Serial dependence can be taken into account by pairwise block bootstrap resampling (Mudelsee, 2014: p. 279 therein). The resampling preserves the distributional shapes, and the pairwise manner preserves (over the block length) the serial dependence or autocorrelation structures. Serial dependence can also be taken into account in a "classical" manner (i.e., using formulas instead of employing resampling or simulation approaches) via the effective number of degrees of freedom, as mentioned by McKay et al. (2020: Section 3.1 therein) as the first approach of GeoChronR.

3. The presence of non-Gaussian shapes *completely* invalidates determined classical uncertainty measures. Then you cannot trust at all a P-value or a confidence interval obtained in a classical manner. (This is what you get if you press the button in standard "office software".) The only remedy that yields acceptably accurate uncertainty measures is pairwise block bootstrap resampling enhanced by computing-intensive calibration (Mudelsee, 2014: Chapter 7 therein). There is a Fortran 90 software on this that employs parallel computing (Ólafsdóttir and Mudelsee, 2014).

4. Spearman's (1904, 1906) rank correlation coefficient is more robust (i.e., it delivers more accurate uncertainty measures in the presence of violations of made

distributional assumptions) than Pearson's (1896) coefficient. (By the way, it is interesting from a philosophy-of-science viewpoint to see how Pearson reacted when he became aware of Spearman's papers.)

2 Regression

First, the observation by McKay et al. (2020: Section 3.2 therein) that age-uncertainties plague also the calibration of proxy variables is important. GeoChronR's error propagation into a set of calibration curves is useful. However, a problem with proxy calibration is that here both variables (the proxy and the indicated climate variable) do show measurement uncertainties – and this leads in ordinary least squares regression (as done by GeoChronR) to a down-biased slope estimate. What is needed is a bias correction, and the Book (Mudelsee, 2014: Chapter 8 therein) gives two ways to perform this.

As regards regression, one may also mention trend estimation, for example, changepoint detection on time series (Mudelsee, 2000; Mudelsee, 2009). Also this estimation target becomes noisier in the presence of timescale errors. However, using a hybrid resampling approach (block bootstrap for obtaining $\{x^*(i)\}_{i=1}^n$, parametric for $\{t^*(i)\}_{i=1}^n$), for which GeoChronR can be utilized, reliable uncertainty determination can be achieved, as Monte Carlo simulations demonstrate (Mudelsee, 2014: Chapter 4 therein). One may further mention nonparametric regression, where this hybrid approach also works; as an example, the paper by Mudelsee et al. (2012) is entitled "Effects of dating errors on nonparametric trend analyses of speleothem time series."

3 Spectral Analysis

The estimation of the true (but unknown) spectrum of the random component in a climate-data generating process is important for reasons also McKay et al. (2020) give: you can study peaks versus background noise; external drivers; leads and lags; and so forth – a spectrum allows you to learn about the physics of the system. Typical records from paleoclimate archives are unevenly spaced since the accumulation process of an archive usually is not constant over time and constant depths are sampled because of material requirements for the measurements. The various timescale construction algorithms, implemented in GeoChronR or elsewhere, can be used to study accumulation in detail. Even in the "untypical" situation of even spacing, it is then the age-uncertainty that introduces uneven spacing into the $\{t^*(i)\}_{i=1}^n$. In any way, spectral estimation for paleoclimate time series has to deal with uneven temporal spacing.

In my view, the Lomb-Scargle method is superior to other spectrum estimation methods since it is regression-based and therefore can be directly applied to unevenly spaced series. However, the raw periodogram (GeoChronR's first spectral approach) is bad to employ since it renders estimates with 100% relative error and also is an inconsistent estimator (i.e., the estimation error does not decrease with n), as we know for decades, see for example the Book (Chapter 5 therein) and the references cited therein. The periodogram therefore has to be combined with a segmenting approach, which trades spectral resolution for reduced estimation error. This advantage, not mentioned by McKay et al. (2020), combined with a test against AR(1) red noise, is the reason of the success of the REDFIT implementation (Schulz and Mudelsee, 2002); which constitutes GeoChronR's second spectral approach. It may be noted that there exists a version called REDFIT-X for cross-spectral analysis (Ólafsdóttir et al., 2016). For the purpose of studying timescale error effects on spectral estimates, I made experiments with adaptations of the REDFIT code (Mudelsee et al., 2009; Mudelsee, 2014: Section 5.2.8 therein) to quantify (1) how much wider spectral peaks get and (2) how much higher the red-noise upper pertenciles get, but there is not yet a publishable code on this.

The other two spectral approaches mentioned by McKay et al. (2020) are wavelet decomposition and Thomson's (1982) multitaper. To the best of my knowledge, for both approaches there is (yet) no reliable implementation or code that is directly applicable to uneven temporal spacing. Likely it is this situation that led McKay et al. (2020: p. 11, line 8–9 therein) to advocate and put into GeoChronR the data pretreatment of "efficient linear interpolation". As many authors before me or Michael Schulz have demonstrated (e.g., Horowitz, 1974), and what I also regularly teach in my courses on climate time series analysis – interpolation means a step away from the original data. It is a dangerous activity. It introduces autocorrelation where there has been none before. If researchers are not trained, then some of them may be even entrapped to employ interpolation in order to boost up the sizes of the samples. This leads to too small uncertainty measures of estimations, to overstatements about the climate system, it damages the credibility of climate research. On top of that, there do exist various interpolation methods (linear, cubic spline, Akima spline, etc.). Which interpolation method to use? Just the "most efficient one"? – Dangerous. In my view the *only* way of having interpolation (to even spacing) in the arsenal of methods is if a meticulous researcher does embark on coding and performing extensive Monte Carlo simulations. In such experiments, the true properties of a data generating process can be prescribed. The properties (e.g., spectrum, n, spacing, noise level) should be close to what data you have, what the prior knowledge about the data indicates or what the geological–physical intuition tells you, the researcher. Then the distorting effects of the various interpolation methods can be quantified and compared. And only if the distorting effects can be shown as negligible, then one may safely proceed with interpolation and present results that are robust in the original statistical sense.

4 Outlook

It is great to see that the engagement of professional statisticians in paleoclimate research is growing. Certainly they can contribute to the issue of the statistical analysis of age-uncertain paleoscientific data.

However, it appears that the various climatological communities, who partly tend to associate themselves with the employed type of paleoclimate archive, are developing a tendency towards using timescale construction algorithms that have been designed within their own community: Oxcal (Ramsey, 2008) or BChron (Haslett and Parnell, 2008) for the radiocarbon community; Wheatley et al.'s (2012) method or BAM (Comboul et al., 2014) for the layered archives such as corals or ice cores; StalAge (Scholz and Hoffmann, 2011), iscam (Fohlmeister, 2012) or MOD-AGE (Hercman and Pawlak, 2012) for the speleothem community. This tendency is unhealthy for scientific progress because it reduces the open exchange among researchers about the optimal way of timescale construction.

Certainly specific archives have their own peculiarities (e.g., about whether hiatuses may occur, whether there exists prior geological-physical knowledge about minimum or maximum accumulation rates) – however, the situation could be improved if the various methods are allowed to be compared against each other (in terms of bias, standard error, etc.) within a Monte Carlo experiment. Scholz et al. (2012) have already done a comparison of methods, but a step beyond that work would be achieved if also the delivered uncertainty bands (around the age–depth curve) could be tested. For example, whether or not a 95% confidence interval for an age at a certain depth does indeed include the true (and known, since it is prescribed) age in 95% of the simulations (within simulation noise). We paleoclimate researchers, as all applied researchers, should apply robust tools that yield reliable estimations and reliable error bars also in the presence of (1) non-Gaussian distributions, (2) autocorrelation and (3) age-uncertainties. I am optimistic that the paper by McKay et al. (2020) and the supplied GeoChronR tool can help to achieve this “Timescale Monte Carlo Comparison Experiment”.

References

Comboul, M., Emile-Geay, J., Evans, M. N., Mirnateghi, N., Cobb, K. M., and Thompson, D. M.: A probabilistic model of chronological errors in layer-counted climate proxies: Applications to annually banded coral archives, Clim. Past, 10, 825–841, 2014.

Fohlmeister, J.: A statistical approach to construct composite climate records of dated archives, Quat. Geochronol., 14, 48–56, 2012.

Haslett, J. and Parnell, A.: A simple monotone process with application to radiocarbon-dated depth chronologies, Appl. Stat., 57, 399–418, 2008.

Hercman, H. and Pawlak, J.: MOD-AGE: An age–depth model construction algorithm, Quat. Geochronol., 12, 1–10, 2012.

Horowitz, L. L.: The effects of spline interpolation on power spectral density, IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-22, 22–27, 1974.

McKay, N., Emile-Geay, J., and Khider, D.: GeoChronR – an R package to model, analyze and visualize age-uncertain paleoscientific data, Geochronol. Discuss., https://doi.org/10.5194/gchron-2020-25, in review, 2020.

Mudelsee, M.: Ramp function regression: A tool for quantifying climate transitions, Comput. Geosci., 26, 293–307, 2000.

Mudelsee, M.: TAUEST: A computer program for estimating persistence in unevenly spaced weather/climate time series, Comput. Geosci., 28, 69–72, 2002.

Mudelsee, M.: Break function regression: A tool for quantifying trend changes in climate time series, Eur. Phys. J.-Spec. Top., 174, 49–63, 2009.

Mudelsee, M., 2014. Climate Time Series Analysis: Classical Statistical and Bootstrap Methods, 2nd edn. Springer, Cham, Switzerland, 454pp.

Mudelsee, M., Scholz, D., Röthlisberger, R., Fleitmann, D., Mangini, A., and Wolff, E. W.: Climate spectrum estimation in the presence of timescale errors, Nonlinear Process. Geophys., 16, 43–56, 2009.

Mudelsee, M., Fohlmeister, J., and Scholz, D.: Effects of dating errors on nonparametric trend analyses of speleothem time series, Clim. Past, 8, 1637–1648, 2012.

Ólafsdóttir, K. B. and Mudelsee, M.: More accurate, calibrated bootstrap confidence intervals for estimating the correlation between two time series, Mathematical Geosciences, 46, 411–427, 2014.

Ólafsdóttir, K. B., Schulz, M., and Mudelsee, M.: REDFIT-X: Cross-spectral analysis of unevenly spaced paleoclimate time series, Comput. Geosci., 91, 11–18, 2016.

Pearson, K.: Mathematical contributions to the theory of evolution—III. Regression, heredity, and panmixia, Phil. Trans. R. Soc. Lond. A, 187, 253–318, 1896.

Polanco-Martinez, J. M., Medina-Elizalde, M. A., Sanchez Goni, M. F., and Mudelsee, M.: BINCOR: An R package for estimating the correlation between two unevenly spaced time series, The R Journal, 11, 170–184, 2019.

Ramsey, C. B.: Deposition models for chronological records, Quat. Sci. Rev., 27, 42–60, 2008.

Scholz, D. and Hoffmann, D. L.: StalAge – an algorithm designed for construction of speleothem age models, Quat. Geochronol., 6, 369–382, 2011.

Scholz, D., Hoffmann, D. L., Hellstrom, J., and Ramsey, C. B.: A comparison of different methods for speleothem age modelling, Quat. Geochronol., 14, 94–104, 2012.

Schulz, M. and Mudelsee, M.: REDFIT: Estimating red-noise spectra directly from unevenly spaced paleoclimatic time series, Comput. Geosci., 28, 421–426, 2002.

Spearman, C.: The proof and measurement of association between two things, Am. J. Psychol., 15, 72–101, 1904.

Spearman, C.: 'Footrule' for measuring correlation, British Journal of Psychology, 2, 89–108, 1906.

Thomson, D. J.: Spectrum estimation and harmonic analysis, Proc. IEEE, 70, 1055–1096, 1982.

Wheatley, J. J., Blackwell, P. G., Abram, N. J., McConnell, J. R., Thomas, E. R., and Wolff, E. W.: Automated ice-core layer-counting with strong univariate signals, Clim. Past, 8, 1869–1879, 2012.

Interactive comment on Geochronology Discuss., https://doi.org/10.5194/gchron-2020-25, 2020.