GeoChronR response to SC2

Monday, November 16, 2020

We thank Dr Mudelsee for offering his valuable perspective on this topic, and spending the time to share these comments. We take up his points in turn:

On uncertainty

We wholeheartedly agree that chronological uncertainties are only one of many types of uncertainty affecting paleoenvironmental timeseries. Indeed, we practice this every day in our own research. We agree that it is important to place chronological uncertainties in this broader context, and will do so in the revised paper.

On correlations

We thank Dr Mudelsee for bringing BINCOR [Polanco-Martinez et al., 2019] to our attention, as we were not aware of it. It seems like a useful and complementary way of going about the problem, and we will investigate incorporating it and its concepts in geoChronR in upcoming releases. The reason why this cannot be done quickly is that it makes several deep assumptions that may or may not be consistent with others that we make, and we need to think this over carefully.

We agree wholeheartedly with points 1, 2, 3, and 4. A confidence interval may be derived from the histogram of correlations, but we will work to export its summary as a a 95% CI, say. We agree that non-gaussianity can be an issue, which is why geoChronR's default behavior is to transform input data to a standard normal (via quantile mapping [van Albada and Robinson, 2007; Emile-Geay and Tingley, 2016]). Lastly, it is true that we could make it easier for users to use other formulations of the correlation coefficient, like Spearman's. We will incorporate rank-based correlation methods into the package on revision of the manuscript. The impact on significance estimates will be the same for the non-parametric tests (isopersistent, isospectral) implemented in geoChronR, so we expect this to be an easy transition.

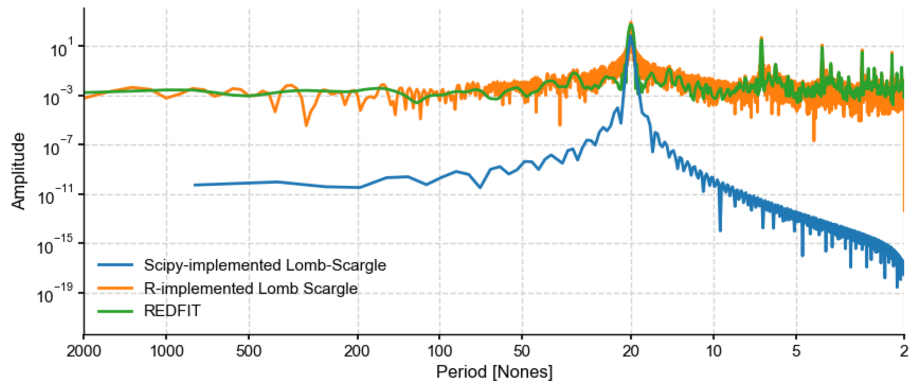

Fig. 1. Benchmarling the Lomb-Scargle implementations in `lomb::lsp()` (orange curve), REDFIT (via the `dplR` package, [Bunn, 2008, green curve]) and `SciPy` [Virtanen et al., 2020, blue curve], as implemented in Khider et al. [2018]. The sample signal is a 20-year sinusoid sampled yearly over 1000 years.

On regression

We are aware that OLS is a biased estimator in the presence of timescale errors. One can imagine extensions like truncated total least squares [Van Huffel and Vandewalle, 1991; Van Huffel, 2004; Markovsky et al., 2010], or the ones suggested in "the book" [Mudelsee, 2013]. We will flag in the revised version of the paper that timescale uncertainties strongly interact with other uncertainties here, in a way that needs to be more rigorously assessed.

On spectral analysis

We disagree with the statement that "the Lomb-Scargle (LSP) method is superior to other spectrum estimation methods since it is regression-based and therefore can be directly applied to unevenly spaced series."

LSP is appropriate in many circumstances VanderPlas [2018], and we've used it in our own work [e.g., Khider et al., 2014], but there are many applications where it can be problematic. First, let us note that the weighted wavelet Z transform [WWZ Foster, 1996; Kirchner and Neal, 2013; Zhu et al., 2019] shares many of these characteristics, and performs very similarly on analytical benchmarks [Khider et al.]. However, there is an important difference between the implementation of the Lomb-Scargle algorithm in R (as used in GeoChronR) and Python (specifically, the SciPy package). This comparison was carried out on a simple, 20-year harmonic in Khider and Emile-Geay [2020], and is summarized in Fig. 1.

For such a simple and abundantly sampled harmonic signal, any good estimator should return something close to a delta function peaked at the f_0 frequency. We see that the SciPy implementation of LS achieves this, but the

the R and REDFIT implementations are extremely noisy, detecting many spurious peaks at high frequencies. Indeed, the signal to noise ratio is 4 to 15 orders of magnitude smaller in those implementations than in SciPy's, though REDFIT's spectrum is (by virtue of averaging) quite a bit smoother than the standard `lomb` implementation. This is quite a substantial difference, for which we cannot find an easy explanation. Indeed, the algorithm is the same, only the numerical implementation differs. In the revised paper, geoChronR users will be alerted to this important limitation.

We do agree that interpolation has serious downsides, but disagree with the blanket statement that it is always "dangerous". Indeed, in our tests, Lomb-Scargle detects many spurious features. The example from section 5.5 is one where the spacing is nearly equal, so the effects of interpolation are minimal. On benchmarks using synthetic data, the multitaper method (MTM) is far superior to LSP with WOSA on evenly-spaced series, and interpolation – used sparingly – provides a way to access these important features of MTM. Thus, we do not agree with a blanket condemnation of interpolation, though we agree that it must be used very carefully to avoid raising the sample size to spuriously high levels; in our practice, typically err on the side of coarse-graining the time-series to avoid this effect, and only use linear interpolation to avoid introducing spurious oscillations. We do agree that sanity checks of robustness are essential.

On age modeling

We agree with Dr Mudelsee that the proliferation of chronology modeling methods is problematic. This is why geoChronR only considers methods based on explicit statistical models, so that the assumptions may be examined on their scientific merits. Unfortunately, this is not the case for many of the other methods you bring up (e.g. StalAge), and we thus refrain from using those in the package or elsewhere.

Age model intercomparison would indeed an important application of `geoChronR`, as it provides a standardized platform upon which methods can be readily tested and compared. Although a thorough intercomparison is beyond the scope of this article, we very much hope to pursue it in future investigations, or to facilitate this task for other investigators. We will revise our discussion to mention recent work on this topic. While Scholz et al. [2012] have indeed performed a comparison of some methods, Parnell et al. [2011] had published on some of the same methods the year before, coming to somewhat different conclusions as Scholz et al. [2012]. More recently, Trachsel and Telford [2016] compared several of the methods included in geoChronR (including OxCal, BChron, and Bacon) on a reference, varved lake chronology. They conclude that "All methods produce mean age–depth models that are close to the true varve age, but the uncertainty estimation differs considerably among models." In particular, BChron is found to overestimate uncertainties in this context. There is thus plenty more to be done to document, benchmark and understand the effects of these various modeling choices, and our revised paper will point to the existing work.

In regards to *"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"* this is called **coverage rate** in the statistical literature, and is indeed a property that paleoscientists should try to constrain for the various methods.

"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". We share your optimism and thank you for these valuable comments, which helped improve the package and the paper.

References

Bunn, A. G.: A dendrochronology program library in R (dplR), Dendrochronologia, 26, 115–124, https://doi.org/10.1016/j.dendro.2008.01.002, 2008.

Emile-Geay, J. and Tingley, M.: Inferring climate variability from nonlinear proxies: application to palaeo-ENSO studies, Climate of the Past, 12, 31–50, https://doi.org/10.5194/cp-12-31-2016, URL `http://www.clim-past.net/12/31/2016/`, 2016.

Foster, G.: Wavelets for period analysis of unevenly sampled time series, Astron. Jour., 112, 1709, https://doi.org/10.1086/118137, 1996.

Khider, D. and Emile-Geay, J.: Benchmarking the Lomb-Scargle implementation in GeochronR, URL `https://github.com/KnowledgeCaptureAndDiscovery/autoTS/blob/master/notebooks/Methods/Lomb-Scargle%20Performance_RvsPython.ipynb`, 2020.

Khider, D., Athreya, P., and Emile-Geay, J.: Estimation of spectral peaks and continuum – a benchmark, URL `\url{https://github.com/KnowledgeCaptureAndDiscovery/autoTS/blob/master/notebooks/Methods/lomb_scargle_vs_wwz_analytical_benchmarks.ipynb}`.

Khider, D., Jackson, C. S., and Stott, L. D.: Assessing millennial-scale variability during the Holocene: A perspective from the western tropical Pacific, Paleoceanography, 29, 143–159, https://doi.org/10.1002/2013PA002534, URL `http://dx.doi.org/10.1002/2013PA002534`, 2014.

Khider, D., Zhu, F., Hu, J., and Emile-Geay, J.: LinkedEarth/Pyleoclim_util: Pyleoclim release v0.4.0, https://doi.org/10.5281/zenodo.1205662, URL `https://doi.org/10.5281/zenodo.1205662`, 2018.

Kirchner, J. W. and Neal, C.: Universal fractal scaling in stream chemistry and its implications for solute transport and water quality trend detection, Proceedings of the National Academy of Sciences, 110, 12 213–12 218, https://doi.org/10.1073/pnas.1304328110, 2013.

Markovsky, I., Sima, D. M., and Van Huffel, S.: Total least squares methods, Wiley Interdisciplinary Reviews: Computational Statistics, 2, 212–217, https://doi.org/10.1002/wics.65, 2010.

Mudelsee, M.: Climate time series analysis: Classical Statistical and Bootstrap Methods, Springer, 2nd edn., 2013.

Parnell, A. C., Buck, C. E., and Doan, T. K.: A review of statistical chronology models for high-resolution, proxy-based Holocene palaeoenvironmental reconstruction, Quaternary Science Reviews, 30, 2948 – 2960, https://doi.org/10.1016/j.quascirev.2011.07.024, URL http://www.sciencedirect.com/science/article/pii/S0277379111002356, 2011.

Polanco-Martinez, J. M., Medina-Elizalde, M. A., Goni, M. F. S., and Mudelsee, M.: BINCOR: An R package for Estimating the Correlation between Two Unevenly Spaced Time Series, The R Journal, 11, 170–184, https://doi.org/10.32614/RJ-2019-035, URL https://doi.org/10.32614/RJ-2019-035, 2019.

Scholz, D., Hoffmann, D. L., Hellstrom, J., and Bronk Ramsey, C.: A comparison of different methods for speleothem age modelling, Quaternary Geochronology, 14, 94–104, https://doi.org/10.1016/j.quageo.2012.03.015, URL http://www.sciencedirect.com/science/article/pii/S1871101412000684, 2012.

Trachsel, M. and Telford, R. J.: All age–depth models are wrong, but are getting better, The Holocene, 27, 860–869, https://doi.org/10.1177/0959683616675939, URL https://doi.org/10.1177/0959683616675939, 2016.

van Albada, S. and Robinson, P.: Transformation of arbitrary distributions to the normal distribution with application to {EEG} test–retest reliability, Journal of Neuroscience Methods, 161, 205 – 211, https://doi.org/10.1016/j.jneumeth.2006.11.004, 2007.

Van Huffel, S.: Total least squares and errors-in-variables modeling: bridging the gap between statistics, computational mathematics and engineering, pp. 539–555, Physica-Verlag HD, https://doi.org/10.1007/978-3-7908-2656-2_44, 2004.

Van Huffel, S. and Vandewalle, J.: The Total Least Squares Problem: Computational Aspects and Analysis, vol. 9 of *Frontiers in Applied Mathematics*, SIAM, Philadelphia, PA, 1991.

VanderPlas, J. T.: Understanding the Lomb–Scargle Periodogram, The Astrophysical Journal Supplement Series, 236, 16, https://doi.org/10.3847/1538-4365/aab766, URL http://stacks.iop.org/0067-0049/236/i=1/a=16, 2018.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nature Methods, 17, 261–272, https://doi.org/10.1038/s41592-019-0686-2, 2020.

Zhu, F., Emile-Geay, J., McKay, N. P., Hakim, G. J., Khider, D., Ault, T. R., Steig, E. J., Dee, S., and Kirchner, J. W.: Climate models can correctly simulate the continuum of global-average temperature variability, Proceedings of the National Academy of Sciences, 116, 8728, https://doi.org/10.1073/pnas.1809959116, URL `http://www.pnas.org/content/116/18/8728.abstract`, 2019.