Review of "Combined linear regression and Monte Carlo approach to modelling exposure age depth profiles" by Yiran Wang and Michael E. Oskin

Pieter Vermeesch
University College London
p.vermeesch@ucl.ac.uk

January 26, 2022

First of all, I would like to apologise for the lengthy review process. Your paper is quite technical and it was not so easy to find qualified reviewers for it. Unfortunately, most of those qualified reviewers are very busy. I am still awaiting one more review, but in anticipation of this, I will already share my own thoughts on your manuscript.

In your paper, you present an improved method to fit cosmogenic nuclide depth profiles. As explained in the introduction, there currently are two ways to do so:

1. Linear regression, ignoring muons.

2. Iterative forward-inverse models, including muons.

The manuscript under consideration presents a hybrid approach, in which the linear regression approach is modified to account for muons, and Monte Carlo simulations are used to estimate uncertainties.

1 Linear fitting with muons

The manuscript shows that, if the thickness of the eroded layer (D) is known, it is possible to compute a depth-dependent 'effective production rate' that accounts for both the neutron and the (slow+fast) muons. This is achieved using a second order Taylor series approximation. That clever trick is the main novelty in this paper.

One problem with the new method is that it requires prior knowledge of the eroded thickness, which violates one of the principles set out in the introductory sections of the paper. According to lines 35–36 of the manuscript, the main advantage of the linear regression approach to depth-profile modelling is that it can "determine an exposure age without any prior knowledge". The second order Taylor approximation trick has removed this advantage. Then why not switch to iterative methods entirely?

In other words, I must agree with the first reviewer that it is not clear what value the new method adds to the cosmogenic nuclide toolbox. Perhaps your method does have advantages, but these are not clear to me at the moment. Probably the best way to demonstrate the alleged superiority of the proposed method is with synthetic examples. Can you come up with a realistic scenario in which the modified linear regression will provide better (i.e. more accurate and/or precise) results than the iterative method?

2 Monte Carlo error estimation

The paper presents an ad-hoc algorithm to estimate the analytical uncertainties of their exposure age estimates by 'Monte Carlo simulation'. My understanding is that this algorithm works as follows:

1. Given a number of measurements and parameters, fit the data by least squares regression of Equation 10. Then calculate the exposure age from that fit using Equation 11.
2. Create new parameter values by drawing random numbers from user-specified prior distributions. Similarly, create new data by drawing random numbers from normal distributions with means corresponding to the measurements, and standard deviations given by the analytical uncertainties.
3. Refit the data.
4. Repeat steps 2 and 3 a few thousand times and investigate the distribution of fit parameters.

This method seems problematic to me because it mixes two kinds of uncertainties:

1. Parameter uncertainties such as the thickness of the eroded layer.
2. Analytical uncertainty of the ^{10}Be measurements.

I'm not sure if you can treat these two sources of uncertainty equally. Parameter uncertainties represent some kind of prior belief about the plausible true values. In contrast, analytical uncertainties represent the standard error of the mean given a finite number of measurements, where the mean is offset from the true value by an unknown amount.

Instead of splitting the calculation into two steps (regression & estimation), parameter estimation could also be done in one step by replacing Equation 11 with the following log-likelihood function:

$$LL(t, C_{inh}, D|C(z), s[C(z)]) = \text{Constant} - \frac{1}{2}\sum_{j=1}^{n}\left(\frac{C(z_j) - P_e(z_j)T_{en}(t, D) - C_{inh}}{s[C(z_j)]}\right)^2$$

where $z = \{z_1, z_2, ..., z_n\}$, $P_e(z_j)$ stands for P_{ze} in the manuscript (at depth z_j), and $T_{en}(t, D)$ is shorthand for Equation 8 (with $i = n$). The 'best' values for t, C_{inh} and D are the ones that maximise LL. The (co)variance (matrix) of t and C_{inh} can then be estimated by inverting the matrix of second derivatives of $-LL$ with respect to t, C_{inh} and D at the maximum likelihood values.

An alternative approach is to use the log-likelihood function in a Bayesian framework, e.g. using the Metropolis-Hastings algorithm. You can then assign your preferred probability distribution for the parameters as 'prior information'. MCMC modelling will yield posterior distributions for t, C_{inh} and D that will look somewhat similar to (but not exactly the same as) Figures 2–5 of the manuscript.

Of course, both of these approaches effectively remove the key characteristic of your linear regression. So, in effect, it turns your hybrid approach to depth profile modelling into a standard forward-inverse modelling approach.

3 Negative inheritance

Lines 177–178 of the manuscript state that "some inversion results yield non-physical predictions with negative inheritance. These negative inheritance predictions are necessary to estimate the full distribution of the exposure age, but we exclude these from the final inheritance results."

This does not make sense. An inversion result that yields non-physical predictions is wrong. A pragmatic way to fix this issue is to parameterise your problem in terms of the logarithm of inheritence, rather than inheritence itself. After the optimisation, you can then exponentiate the maximum likelihood value to ensure a strictly positive result.

4 Scaling models

Your method is based on an exponential approximation for the depth-dependence of muon production, but as far as I can tell, the parameters of the exponential approximation are not varied with elevation or location. This is incorrect. At shallow depths the absolute magnitude of the muon production rate is quite variable with elevation, so even if you use an exponential approximation the value for the surface production rate for each reaction would have to be different at each site. This could be easily fixed by using a more complicated muon production model at the beginning to compute site-specific exponentials for each site. But without fixing this, the age estimates are expected to be fairly inaccurate in lots of cases.