Review of "DQPB: software for calculating disequilibrium U-Pb ages" by T. J. Pollard *et al.*'

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

October 28, 2022

This paper presents methods and software to correct U–Pb data for U-series disequilibrium. The methods are based on the classic Bateman (1908) equations, whereas the software combines `Python` with `Excel`. As the author of the open source `IsoplotR` package for geochronological data processing, I applaud the authors' decision to share their source code on `GitHub`. I was also excited to to see that DQPB is part of a software suite called `pysoplot`. It would be great if the geochronology community had multiple, independently developed `Isoplot` alternatives, as this would stimulate further innovation and allow users to cross-check results. `DQPB` is easy to install on `Windows` and `Mac OS`, but not on `Linux`. This is, presumably, because `DQPB`'s graphical user interface (GUI) interacts with `Excel`, which is not available on `Linux`.

`DQPB`'s online documentation is detailed and extensive, but only covers the GUI. It would be useful if it also covered the command-line API. Not only would this allow `Linux` users to access `DQPB`, but it would also benefit power users on other operating systems. For example, in the second half of this review, I will carry out some numerical experiments using `R`. Had there been a `DQPB` API, then I would have been able to use `Python` instead, which would have been a more efficient way to make my points.

I tested `DQPB` on a `Windows` machine and found the software easy to install and straightforward to use, even without reading the instructions. The integration with `Excel` reminded me why `Isoplot` was so successful: it makes a lot of sense to group the data processing tool together with the data itself, in a spreadsheet. I haven't really been able to replicate this experience in `IsoplotR`. One slightly annoying issue is that `DQPB` overprints the data with the results. This problem probably only occurs on `Windows` (`DQPB`'s lead developer appears to use `Mac OS`) and should be easy to fix.

I tested `DQPB` on a number of samples and got similar results to `IsoplotR`. This is not surprising given that the two programs use, essentially, the same equations, although `IsoplotR` casts them in a matrix form. The manuscript is a bit dismissive of `IsoplotR`'s disequilibrium corrections, even though these are more extensive than `DQPB`'s current capabilities and include 3-dimensional 'Total U-Pb' isochron regression and Ludwig (1998)-style error propagation, neither of which are implemented in `DQPB`. However, there are two things that `DQPB` does, and which `IsoplotR` doesn't.

The first of these is robust isochron regression as an alternative to `IsoplotR`'s model-3 regression, using the spine algorithm of Powell et al. (2020). I have no objections to or comments on this. The second difference is that `IsoplotR` currently does not propagate the uncertainty of the initial disequilibrium correction. The reason for this limitation is that there is no easy analytical solution to the error propagation problem, and numerical solutions are too slow for a software package that can be run online. `DQPB` estimates the uncertainty of the disequilibrium correction by Monte Carlo simulation. I believe that this is a similar approach to that taken by `Isoplot`. The Monte Carlo approach works as follows:

1. Given a linear array of isotopic data in Tera-Wasserburg (i.e. $^{238}U/^{206}Pb$ vs. $^{207}Pb/^{206}Pb$) space, and the measured $^{234}U/^{238}U$ activity ratio ($\mu_{4/8}$) and its uncertainty ($\sigma_{4/8}$);
2. Draw a random value ($[4/8]_m$, say) from a normal distribution with mean $\mu_{4/8}$ and standard deviation $\sigma_{4/8}$;
3. Carry out a linear regression through the data and find the initial $^{234}U/^{238}U$ activity ratio ($[4/8]_i$) and isochron age (t) that is consistent with both the U–Pb data and $[4/8]_m$;
4. Repeat steps 2 and 3 until the entire distribution of measured $^{234}U/^{238}U$ activity ratios has been sampled;
5. If step 3 fails, or produces physically impossible results (e.g., $t < 0$), then ignore the corresponding $[4/8]_m$. Otherwise add the $[4/8]_i$ and t-values to a list of a acceptable results;
6. Use the spread of the acceptable $[4/8]_i$ and t-values to quantify their respective uncertainties.

Since `DQPB` does not work on my computer, I implemented my own version of this algorithm, using `R` and `IsoplotR`. The only major difference between my code and `DQPB` is that it does not sample the $[4/8]_m$-distribution randomly, but uses a targeted approach to sample $[4/8]_m$ as a sequence of regularly spaced normal quantiles. This has two advantages. First, it requires orders of magnitude fewer iterations (50 vs. 30,000). Second, it produces a deterministic result, unlike the Monte Carlo approach, whose results depend on the seed of a random number generator. To verify that the `R` code produces equivalent results to the `Python` code, Figure 1 analyses the 'Corchia' dataset from the manuscript. The results are, essentially, the same as those that I obtained using `DQPB` on `Windows`:

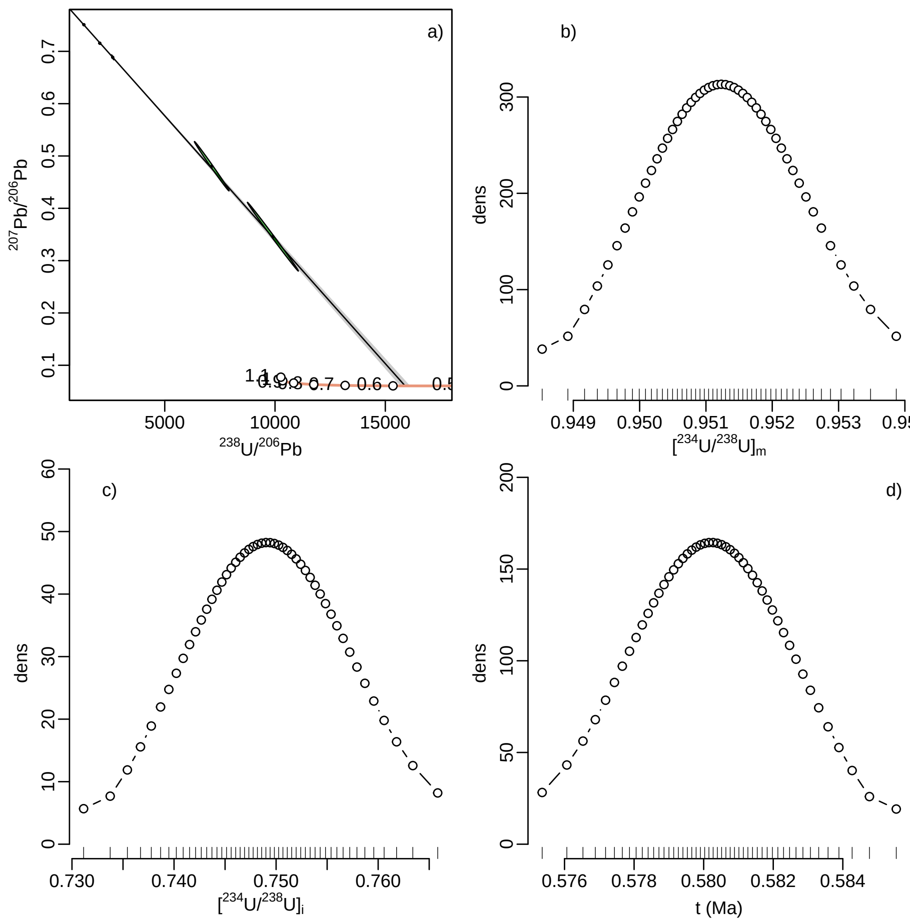

Figure 1: Output of a `DQPB`-like algorithm for the Corchia dataset. a) Tera-Wasserburg concordia diagram with disequilibrium-corrected isochron (age=0.5803 Ma); b) 50 representative samples from the $^{234}U/^{238}U$ measurement distribution; c) The corresponding initial $^{234}U/^{238}U$ activity ratios; d) The isochron ages corresponding to the initial $^{234}U/^{238}U$ activity ratios presented in panel c.

Next, let us apply the same approach to a much older sample, such as the 'Hoogland' data set of Pickering et al. (2019):

[38/06]	2σ%[38/06]	[07/06]	2σ%[07/06]	ρ
843.8	7.59420	0.0697	0.00742305	-0.9991
843.9	4.64145	0.0673	0.00450910	-0.9972
737.4	6.26790	0.1671	0.00618270	-0.9982
834.6	10.43250	0.0749	0.01048600	-0.9994
845.5	9.72325	0.0696	0.00967440	-0.9995
781.6	5.47120	0.1189	0.00552885	-0.9984
787.1	6.69035	0.1212	0.00672660	-0.9578
834.5	6.25875	0.0719	0.00611150	-0.9976

The uncorrected U–Pb isochron age for this sample is 7.407 Ma, which is 30 half-lives of ^{234}U. Consequently, the measured present-day ^{234}U/^{238}U activity ratio ($[4/8]_m$) is statistically indistinguishable from secular equilibrium, at 1.0016 ± 0.001 (1se). However, when we apply the `DQPB` approach to this data set:

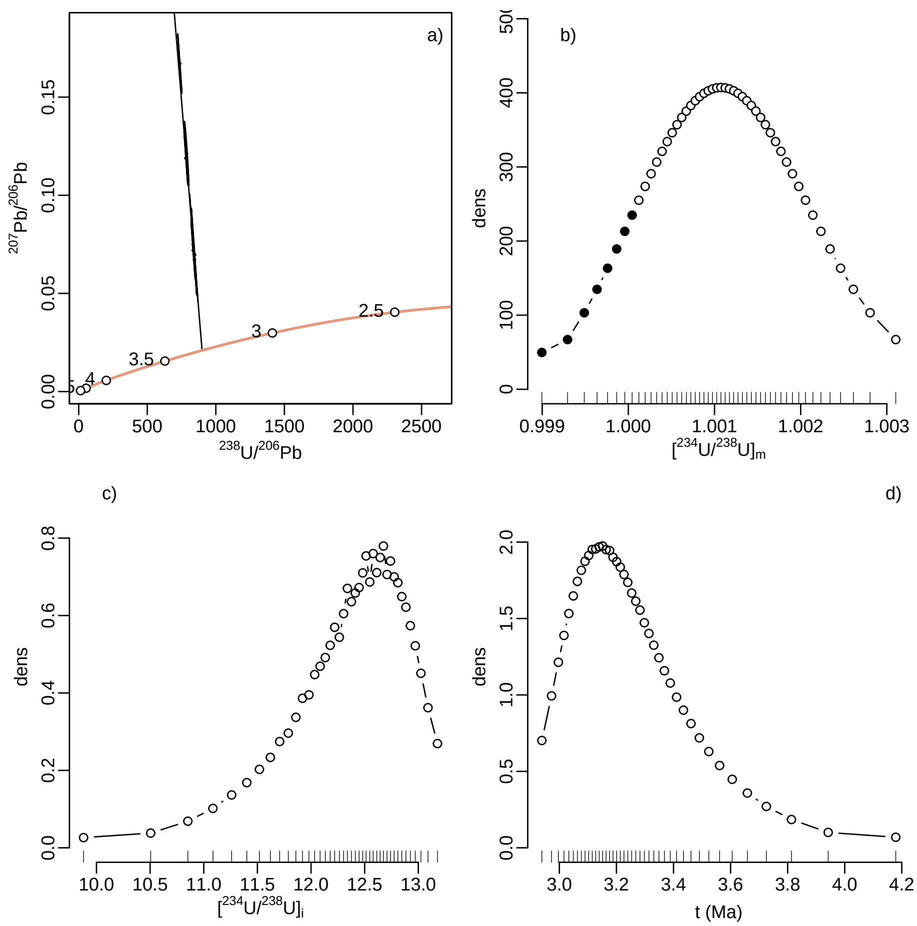

Figure 2: Output of the DQPB-like algorithm for the Hoogland dataset of Pickering et al. (2019). Panels a)-d) are as in Figure 1. The black dots in panel b) mark synthetic replicates that are rejected because they yield physically impossible $[4/8]_i$ and/or t-values.

Despite the lack of measurable disequilibrium, the DQPB approach appears to have successfully applied a disequilibrium correction, resulting in a corrected age that is less than half the uncorrected age. Although the magnitude of the correction is huge, the precision of the corrected age appears to be surprisingly good, at less than 15%. How is this possible? The answer lies in the rejected solutions (step 5 of the algorithm), which are marked in black in Figure 2.b.

To demonstrate that the result of Figure 2 is wrong, let us replace the measured $^{234}U/^{238}U$ activity ratio with the equilibrium ratio. Thus

$$\mu_{4/8} = \frac{\lambda_{234}}{\lambda_{234} - \lambda_{238}} = 1.000055(\pm 0.001)$$

Plugging this value into the DQPB algorithm yields the following result:

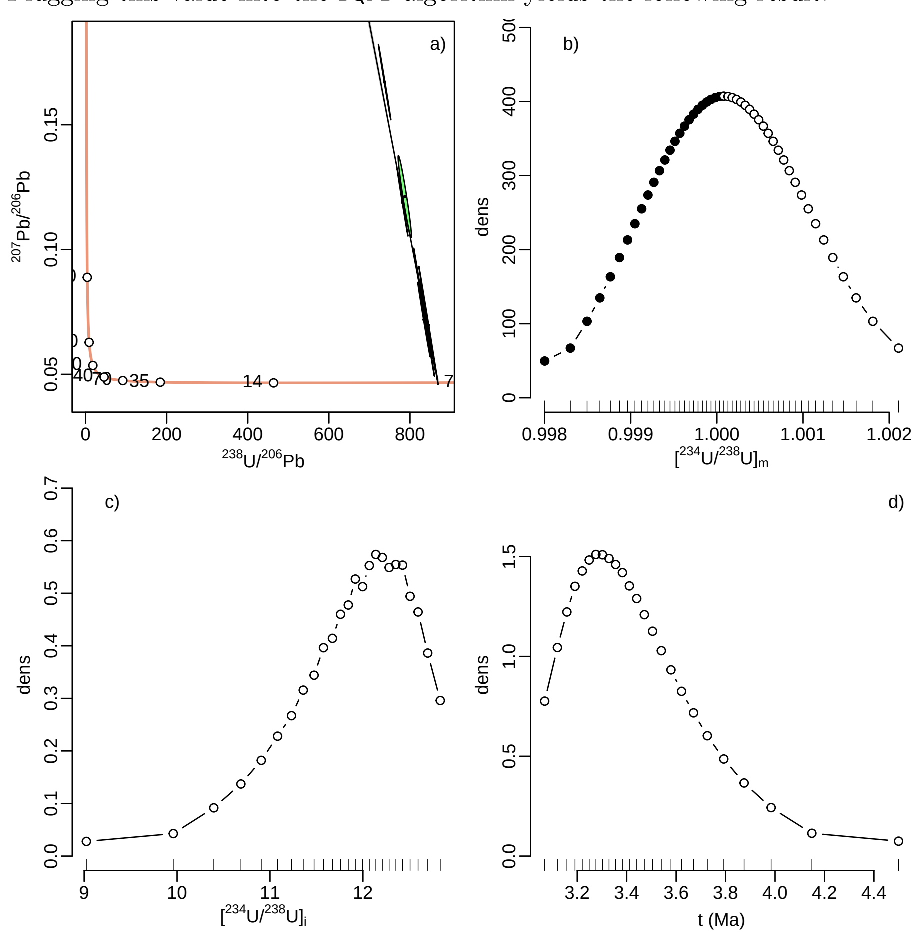

Figure 3: The same data as Figure 2, but assuming perfect equilibrium of the measured $^{234}U/^{238}U$ activity ratio. Note that half of the synthetic replicates have been rejected (black circles). Even though there is absolutely no evidence for disequilibrium, the `DQPB` approach nevertheless seems very confident that the isochron age should be halved compared to the uncorrected data.

Once again, it appears that the algorithm has achieved the impossible, and has corrected the disequilibrium effect in the absence of any detectable deviation from the equilibrium ratio. It has done so by ignoring 50% of the $[4/8]_m$ distribution. This is because, for this old sample, essentially any $[4/8]_m$ value that is less than the equilibrium ratio would require a negative $[4/8]_i$ ratio, or a negative isochron age t.

In fairness to `DQPB`, the program does report the number of failed inversions. I would argue that solutions that include rejected tries should not be trusted. It would be better to replace the ad-hoc Monte Carlo approach with a more rigorous alternative. I believe that there are two approaches for doing this. The first option is to use the traditional ('frequentist') approach, which uses the method of maximum likelihood. I have attempted this but got stuck: the problem quickly becomes too complex for a non-statistician. The second option is to use a Bayesian approach. This is much easier to implement, as I will demonstrate in a basic form:

1. Define a prior distribution for $[4/8]_i$. In the following example, I will use a uniform distribution from m to M (e.g., $m = 0$ and $M = 20$ for $[4/8]_i$-values running from 0 to 20), but this could easily be replaced by a more informative prior.

2. Draw a random sample from this prior distribution, carry out the constrained isochron regression and report the resulting age (t) and remaining $[4/8]_m$.

3. Calculate the likelihood of the inferred $[4/8]_m$ values under a normal distribution with mean $\mu_{4/8}$ and standard deviation $\sigma_{4/8}$.

4. Repeat steps 2 and 3 and combine the likelihood values with the prior distribution to produce a 'posterior' distribution of $[4/8]_i$ values and the correponding age distributions. This can either be done using a Markov chain, or with a targeted approach of regularly spaced $[4/8]_i$ values.

Applying this approach to the Corchia example yields essentially identical results to the `DQDB` results of Figure 1:

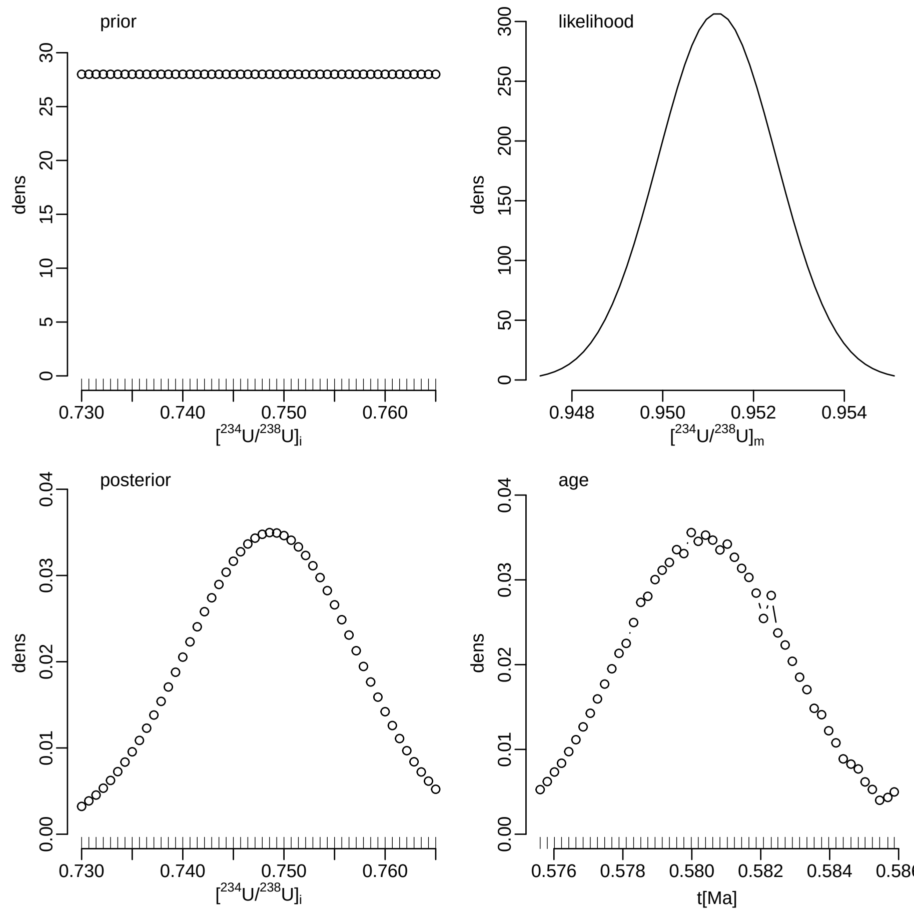

Figure 4: The same data as Figure 1, but using the Bayesian approach. For this sample, the `DQPB` method yields similar results to the Bayesian solution.

However, when the Bayesian approach is applied to the Hoogland data, it produces a very different, and I would argue more sensible, result than `DQPB`:

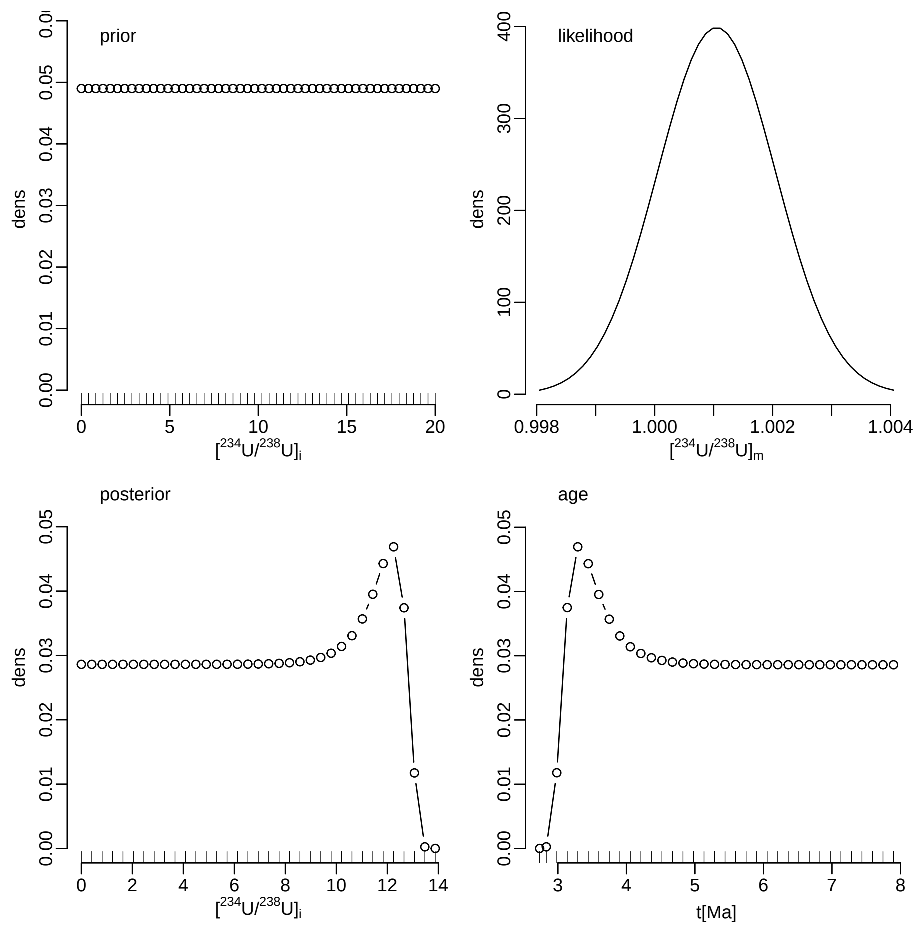

Figure 5: Application of the Bayesian approach to the Hoogland data of Pickering et al. (2019). The maximum posterior likelihood agrees with the mode of the `DQPB` solution of Figure 2. However, whereas `DQPB` suggests a high degree of confidence in the disequilibrium correction, the Bayesian result shows that one cannot rule out a much older age. In fact, the uncorrected U–Pb age of 7.4 Ma is almost as likely as the corrected age of 3.3 Ma.

Note how the posterior distribution still has a maximum at $t = 3.3$ Ma and $[4/8]_i = 12$, just like the `DQPB` solution of Figure 2. But unlike the `DQPB` solution, the Bayesian solution assigns a significant likelihood to older ages, including the uncorrected age of 7.4 Ma (and older!). The similarity of the posterior distribution to the prior distribution reflects the fact that the measured $^{234}U/^{238}U$ activity ratio contains relatively little information. The resulting uncertainties are huge but a correct reflection of our ignorance about the true extent of the disequilibrium in this case.

Finally, changing the $\mu_{4/8}$ ratio to the equilibrium value:

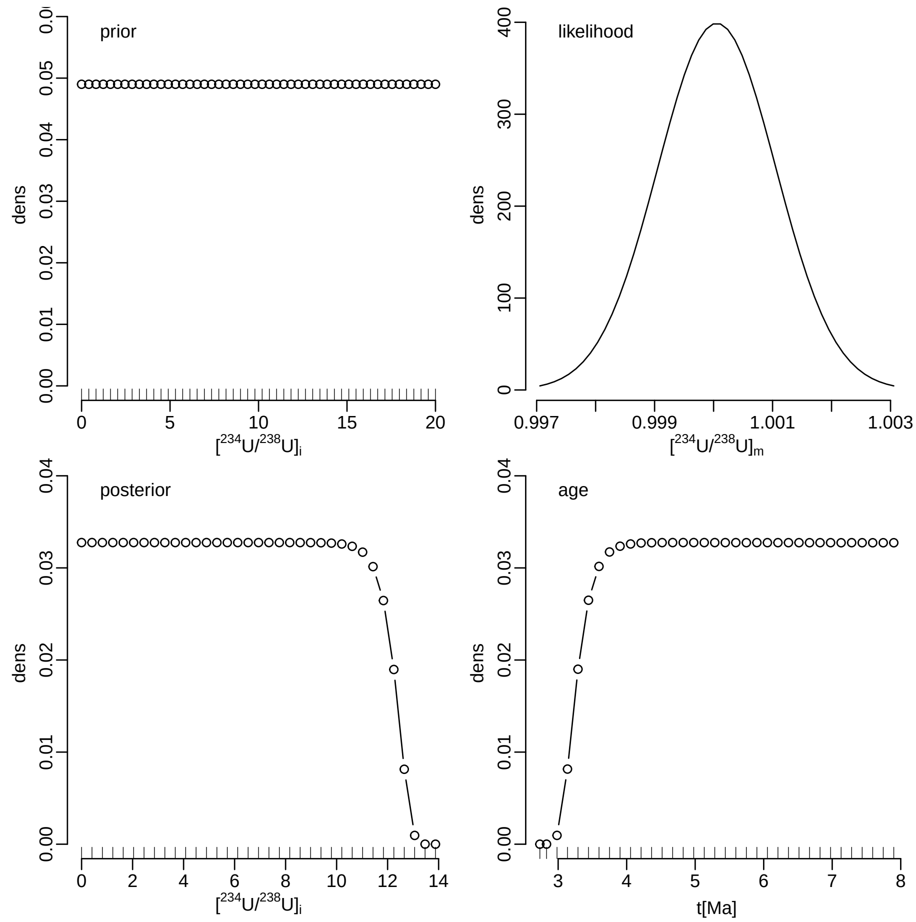

Figure 6: Bayesian alternative to the modified Hoogland example of Figure 3. The posterior distribution is, essentially, identical to the prior distribution.

The similarity of the posterior distribution to the prior means that there is no information in the likelihood function. In other words: the measured $^{234}U/^{238}U$ activity ratio does not tell us anything about the initial disequilibrium. If we truly believe that the sample may have experienced extreme $^{234}U/^{238}U$ disequilibrium, then it is not possible to undo the effects of this disequilibrium on the $^{206}Pb/^{238}U$ clock with the current $^{234}U/^{238}U$ activity ratio.

In conclusion, I think that the authors have two options:

1. If they want to stick with the Monte Carlo approach, then they should modify their code to return an error message when the analytical uncertainty of the $^{234}U/^{238}U$ measurement overlaps with the equilibrium ratio.

2. Alternatively (and preferably), they could replace the ad-hoc Monte Carlo approach with a Bayesian solution along the lines of what I have sketched

in this review. Note, however, that the simple solution presented above is limited in the sense that it only quantifies the uncertainty associated with the disequilibrium correction, and ignores the uncertainty of the actual isochron fit. Doing so would be relatively straightforward using an MCMC algorithm, but I lacked the time to do so for this review.

References

Bateman, H. The solution of a system of differential equations occurring in the theory of radio-active transformations. *Proc. Cambridge Phil. Soc., 1908*, 15: 423–427, 1908.

Ludwig, K. R. On the treatment of concordant uranium-lead ages. *Geochimica et Cosmochimica Acta*, 62:665–676, 1998. doi: 10.1016/S0016-7037(98)00059-3.

Pickering, R., Herries, A. I., Woodhead, J. D., Hellstrom, J. C., Green, H. E., Paul, B., Ritzman, T., Strait, D. S., Schoville, B. J., and Hancox, P. J. U–Pb-dated flowstones restrict South African early hominin record to dry climate phases. *Nature*, 565(7738):226, 2019.

Powell, R., Green, E. C., Marillo Sialer, E., and Woodhead, J. Robust isochron calculation. *Geochronology*, 2(2):325–342, 2020.