Geochronology Discuss.,
https://doi.org/10.5194/gchron-2020-4-AC6, 2020

Interactive comment on "Robust Isochron Calculation" *by* Roger Powell et al.

Roger Powell et al.

powell@unimelb.edu.au

Received and published: 3 May 2020

As noted in the review's summary, the algorithm and the code is an important part of the manuscript, so the reviewer's comments on that will be discussed first. But we would like the manuscript to be more accessible too.The reviewer's comments on improving accessibility are appreciated. As noted in the response to review 3, there is a difficulty making the manuscript more accessible to both geologists and statisticians, but we have tried to do this.

Regarding the computational algorithm, it is recognised that the algorithm might not converge, and the code flags if that happens. The logic to find a good starting point—now discussed in the Appendix—is that several different estimates are calculated initially, including $L1$. The estimate with the smallest $\sum \rho(r_k)$ (the sum being minimised in the algorithm) is used as the starting point. The HUBER function allows the user to

Printer-friendly version

Discussion paper

call it with their choice of possible additional starting points. Following the reviewer's suggestion, the Siegel repeated median estimator has been added to the calculations that the HUBER function does to find a good starting point.

The reviewer's suggestion for an improved algorithm is certainly interesting and will be investigated in due course. But we can refer here to our response to review 1: in our experience the algorithm works well, with less than 5 iterations for each simulated dataset in the family of contaminated-Gaussian datasets investigated—and showed no failures to converge in application to hundreds of thousands of such datasets. It is worth repeating, too, that the greatest majority of actual isochron datasets are not chronic, for better or worse having been cleaned of more gross outliers by users.

Regarding improving accessibility, the parts of Section 2 in the review will be considered in turn

2.1: As already noted in the Introduction to the manuscript, testing for the distribution of the scatter in the data is not possible, but with insufficient detail why. The fact is that the greatest majority of geochronological datasets are of the order of 10 datapoints.

The dataset in Fig. 6 is unusual in being so large. The `qqplot` for this appears to be in the range of strictly Gaussian datasets of 50 datapoints, from running several simulations. With the small excess scatter as reflected in the relatively small `mswd`, *more* datapoints in a dataset would be needed to ascertain if it is a Gaussian mixture of the sort simulated in the manuscript.

2.2: With current isochron calculations stemming from the work of York, and the first author's lack of knowledge of that part of the stats literature, he was unaware of the Error-in-Variables/Measurement-Error work. These terms have not been used in motivating isochron calculations in the geochronological literature, though the main idea underpins the Introduction in York (1966). Via Fuller (1987), this

aspect is now covered in the manuscript, allowing readers, including statisticians, to see that link. Unlike many measurement error problems, uncertainties in x can be properly accounted for in isochron calculations as they are known from the analytical work generating the isotopic data.

The data for Fig. 6 are now included in Appendix C, following the Python code. Fitting the data just with ordinary least squares and data uncertainties from the data scatter, as in Model 2 calculations, has a slightly smaller slope than YORK possibly indicating a small downward bias, but just with a difference of -1.2 σ on the slope.

2.3–2.5: Section 2.2–2.3 of the manuscript has been rewritten to cover what the reviewer is suggesting. Clearly it was a mistake to sweep as many of the equations as possible into the Appendix! Efficiency is now explained much better in the text.

2.6: As noted above, the greatest majority of geochronological datasets are of the order of 10 datapoints, so that checking for the form of excess scatter is not possible. Investigating the family of Gaussian mixture distributions seemed like a natural thing to do, as the reviewer agrees?

2.7: The possibility of using a redescending ρ function, and also the other methods advocated in Maronna (2019), Section 5.8.1, is discussed already in the response to review 2. At least reviewer 4 ackowledges that adopting such methods for the error-in-variables calculations in the manuscript would be challenging

2.8 minor comment: Using YORK and HUBER as the names for the two main approaches discussed in the manuscript does have the merit of simplicity. The words around choosing these terms can be improved easily. We are at the mercy of the subeditor here.

Interactive comment on Geochronology Discuss., https://doi.org/10.5194/gchron-2020-4, 2020.