# Review of “Robust Isochron Calculation” Version 4 Revision by Powell et al. (gchron-2020-4-manuscript version4)

September 3, 2020

# 1 Summary

This is a greatly improved version of the manuscript. The abstract and introduction now provide a much better lead-in for the main parts of the paper. Furthermore, the authors have paid careful attention to this reviewer's specific recommendations for improving the manuscript, and for the most part have responded quite positively. The Section 1.1 On ISOPLOT and Section 1.2 Replacing ISOPLOT are very effective. The over-arching value of the paper is that: (1) introduces the use of the well-known Huber robust regression method in Geochronology research, and (2) it provides a novel algorithm, including Python code, for computing a Huber robust regression estimator where the measurements are subject to error. As such, I recommend publication of the paper.

That said, there are a few small and moderate issues and one major issue, described below, that the authors need to take care of.

# 2 Small Issues

1. The term "excess scatter" used in the Abstract and several places in the Section 1 Introduction is far too vague particularly when the focus is dropping the normality assumption for the tails of the distribution. It would be better to use language such as "when the data distribution has non-normal tails (or fat tails) associated with outliers", or "when the data distribution has higher variability in the tails than is indicated by the variability of the central part of the distribution".

2. The Abstract contains the statement: "A statistical test is provided to ensure that a central spine of the uncertainty distribution data is Gaussian." Then in the discussion connected with Table 1, it should be pointed out exactly what is the statistical test.

3. In several places the authors state in effect that a robust method produces "identical" results as least squares, this seldom literally true, and better term would be "nearly identical".

4. The author's use the term "resistant" to indicated a highly robust (e.g., with regard to breakdown point and bias control) but rather inefficient estimate. In fact "resistant" is just an alternative data-oriented term for "robust", the latter typically meaning with respect to analytic measures such as variance/efficiency, bias and mean-squared error. So the term resistant should be dropped, and in its place use "highly robust (toward outliers, or with respect to bias) but inefficient (in terms of variances) at a normal distribution".

5. The two-component normal mixture distribution was used in the seminal paper Tukey, J.W. (1960), *A survey of sampling from contaminated distributions, in: Contributions to Probability and Statistics, I. Olkin, Ed., Stanford University Press*. to show that simple trimmed means perform better than the sample mean in terms of efficiency for such distributions, including the normal distribution as a special case. This paper inspired a lot of subsequent research, e.g., by Huber and Hampel and beyond. So it wouldn't hurt to reference it.

6. What is the "Tera-Wasserburg isochron space"? Maybe all readers know what this is? Can't this be stated in a way that is for sure clear to all possible readers?

7. The sentence just below equation (2) states: "with the constant normalizing the result to be like the standard deviation for Gaussian-distributed $r$. The use of the word "like" is evidently to make this paper easier to read by a broad audience of readers. However, for basic concepts such as "an unbiased estimator", I think it is best to be precise. In this particular case, the choice of the constant 1.4826 makes `nmad(r)` an unbiased estimator of the standard deviation as the sample size goes to infinity when the data is normally distributed, and approximately unbiased in finite samples. Something like this should be put in a footnote.

8. It would be good to consistently use circumflex's to indicate estimators, e.g., $\hat{s}$ instead of $s$ in equation (2).

# 3 Moderate Issues

1. In Section 2.4 the authors state: "Efficiency at the Gaussian distribution is the ratio of the variance obtained by the optimal estimator (YORK), divided by the variance using the chosen robust estimator (in this case, SPINE).", which is quite correct. However, the subsequent sentence: "Obviously, SPINE has optimal efficiency when all r in a data set have $|r_k| < h$, when it is identical to YORK, but there is an efficiency loss associated with using SPINE for an isochron-yielding data set with any $|r_k| < h$." is quite nonsensical. Efficiency of an estimator is defined at a data distribution, not at a data set!

2. The practice of trying to define an errorchron versus an isochron based on the character of the residuals from a straight line fit seems to be a rabbit hole that one should not be caught in. After all, the robust regression could be of quite high quality (see next item) with regression residuals being quite fat-tailed and with outliers.

3. The use of mswd in the broad range of regression practice is seldom used (seems to have arisen mostly in connection with categorical data and two-way tables). But R-squared is always used as a measure of goodness of a regression fit. So why is that not used? And if it is not used on purpose, because mswd is deemed to be better, what is the justification. Something needs to be said about this.

4. The idea to somehow use `nmad(r)` as defined by equation (2) is a good one. However, using it for testing whether or not the central part of the distribution is normal, as the authors do, is not very useful. After all, the central part is usually well-modeled by a normal distribution, and it is fat non-normal tails that are the issue. One might consider tests based on the ratio of the standard deviation estimate to the `nmad(r)`, or a test based on the difference between 95% and 75% ordered data. But at the small sample sizes encountered, such a statistic may not be very useful.

## 4 Main Issue

At the end of Section 2.1 the authors state: "Note that with the sample sizes provided by most modern geochronological techniques, it is not possible to test for Gaussian behaviour, or such small departures from Gaussian behavior." For really small sample sizes, say 10 - 20 or so, formal tests such as the Kolmogorov-Smirnov test or the Anderson-Darling test are not likely to be very useful. However, normal qqplots with point-wise 95% confidence bands are very useful for detecting fat-tails and outliers.

To this I would add: I have reviewed many papers on applications of robust statistics over the years and this is the first time I have seen a paper promoting the use of robust regression without a data example showing that there is at least one potentially influential outlier (and often more than one).

NOTE: I ran least squares and a robust regression estimator on the pairs of mean values of the ellipses in Figure 7, with the result that there is essentially no difference between the LS and robust regressions, and no indication what-so-ever of outliers in the regression residuals.

So the paper needs one convincing example of a data set where there are one or more outliers that result in a clear difference between the LS and robust fits (as measured most simply by the robust standard error of the robust regression coefficients), and outliers in the regression residuals as indicated by a normal qqplot with 95% point-wise confidence intervals. Otherwise, the paper presents a solution in search of a problem.