Response to Dr. Marco Colombo's review of "Too tood to be true: Underdispersion in geochronology"

Pieter Vermeesch

September 1, 2026

I would like to thank the editor for appointing Dr. Colombo as a reviewer. It is important that geochronological studies are evaluated by mathematically trained reviewers. Their approval strengthens the case made by the manuscript under consideration.

1. *the same approach cannot be applied by a single practitioner on their study*

   This is true with the exception of detrital geochronology, as explained in Section 7.4.

2. *It may help to provide a definition of statistical power.*

   The revised manuscript will add a sentence to the first footnote defining power as the probability of rejecting a false null hypothesis.

3. *The MSWD ranges at lines 77–79 are (approximately) valid at $\alpha = 0.05$, not for an arbitrary confidence level, so for clarity that should be stated.*

   Lines 75–77 mention that the standard deviation of the asymptotic MSWD-distribution is $\sqrt{2/\nu}$. Therefore, the cutoff is actually defined at '2-sigma', which is not quite the same as $\alpha = 0.05$.

4. *The arguments presented at the start of Section 5 don't consider correction for multiple testing. Even a plain Bonferroni correction [...] can be enough to correct the distortion in the probability of rejecting a test*

   Correct. In fact the second paragraph of Section 5 contains all the elements needed to derive the Bonferroni correction. It will be reworded to make that connection clearer.

5. *Line 280: this statement is not clear: "More extreme degrees of underdispersion require that the p-value cutoff is raised from the usual value of $\alpha = 0.05$ to higher values (e.g $\alpha = 0.5$)".*

   What is meant here is that, in the mind of the cherry picker, the p-value is not a metric but a goal[1]. Their aim is to ensure that all p-values in a study are as high as possible. So psychologically, their "target" p-value is not $\alpha = 0.05$ but $\alpha = 0.5$, say. The text will be rewritten to clarify this point.

6. *the α values corresponding to a MSWD cutoff of 1 would be more clearly derived by reminding that, as $MSWD = \chi^2_\nu/\nu$, then $MSWD \leq 1$ corresponds to finding the probability that $\chi^2_\nu \leq \nu$".*

   Line 283 will be changed to "Then this MSWD-cutoff would be equivalent to $\alpha = P(\chi^2 > 4|\nu = 4) = 0.41$ for $n = 5$; to $\alpha = P(\chi^2 > 19|\nu = 19) = 0.46$ for $n = 20$; and $\alpha = P(\chi^2 > 99|\nu = 99) = 0.49$ for $n = 100$."

7. *Regarding the fourth point in the Discussion (section 8, lines 454-459): excessive outlier removal not only increases bias, it also decreases precision as the sample size is reduced.*

   That is true in theory. However, in practice cherry picking nearly always improves precision as it reduces the scatter of the data. A note will be added to draw attention to this point.

The other suggestions will also be followed in the revised manuscript.

[1] My colleague Murat Tamer pointed out that this behaviour is a manifestation of "Goodhart's Law": when a metric becomes a goal, it ceases to be a good metric.