Summary

The authors introduce a new metric to the world of detrital geochronology: the Wasserstein Distance. The authors assert that this metric overcomes shortcomings of the KS distance. The authors define the metric and compare its use to the KS distance in a simple toy example. They then apply the WS Distance and the KS Distance to compare a number of river samples from Scandinavia. They conclude that the WS Distance produces more geologically meaningful comparisons than the KS Distance.

Recommendation

The authors have not demonstrated that the WS Distance produces geologically meaningful results. At the very least they need to address the concerns raised below before publication. However, the comments below suggest that the WS Distance may not be an appropriate metric to compare geochronological distributions, because of the unique geological implications of minorly distinct age modes (distinct sources) versus multimodal distributions which include some shared age modes (potentially shared sources).

General comments

The WS Distance may be more intuitive than the KS distance in many cases, but whether it is more "sensible" depends on the application. While the toy dataset clearly shows the advantage of WS over KS for assessing simple dissimilarity between sample ages, in many DZ studies it is the degree to which samples share the same sources that is of interest; the absolute difference in age is not directly relevant. For example, if we assume the sources for the samples in the toy dataset each have distinct ages, A and C are no more similar in terms of their sources than A and D, and the equal KS value of 1 (i.e., complete dissimilarity) for (A, C) and (A, D) is actually more informative than the WS values W(A, C) = 1 and W(A, D) = 10.

The authors need to explore the behavior of WS Distance for mult-modal data sets. For example, what is the interpretation of high versus low WS Distances when comparing multi-modal data sets? This may seem intuitive, however, I suspect that the results will not be intuitive given how the WS Distance is calculated.

The geological context of the samples needs to be provided. Even a reproduction of the Morton et al. (2008) Figure 1 would be helpful in terms of understanding which samples would be predicted to be derived from similar sources.

I am fairly surprised that the authors chose samples with n<=60 grains to showcase a new statistical comparison metric. Multiple studies (including one by the co-author) have shown that n>>100 (ideally n>300) are needed for robust statistical comparison. Although it can be argued that samples of that size are unnecessary for simple age distributions such as those presented here, this raises the question of why choose the simplest possible scenario to showcase a new application of a statistical metric. Rather, it seems that a true breakthrough would be demonstrating that the WS Distance can deal with a previously unresolved problem or elegantly deals with an intensely complex data set.

Figure 2 demonstrates problems with both the KS and WS distances. First, the WS distance increases linearly with displacement away from 1000 Ma. However, once the two distributions no longer overlap, they are no longer any more or less similar because the x-axis is age, not distance.

Two age distributions that share no age modes are equally dissimilar regardless of the age difference between their modes due to the geological implications of sharing versus not sharing age modes. Hence, the increasing WS Distance with displacement beyond zero overlap is an undesirable trait. Second, both the WS and KS distances indicate that the distributions are most alike (minimum WS and KS distance) when both are centered at 1000 Ma. However, at that point they share no age modes. The green distribution would have an age mode at 1000 Ma, but the black distribution would have age modes at ~900 and ~1100 Ma. Rather, they should be most alike when the green distribution overlaps with one of the age modes of the black distribution. As an aside, this is the behaviour that cross-correlation of KDEs of these distributions provides (below).

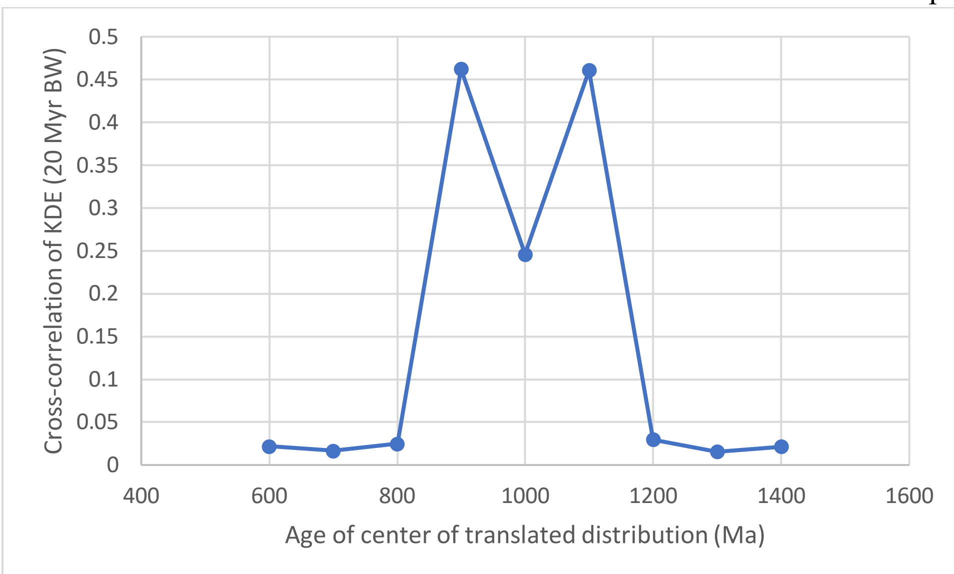
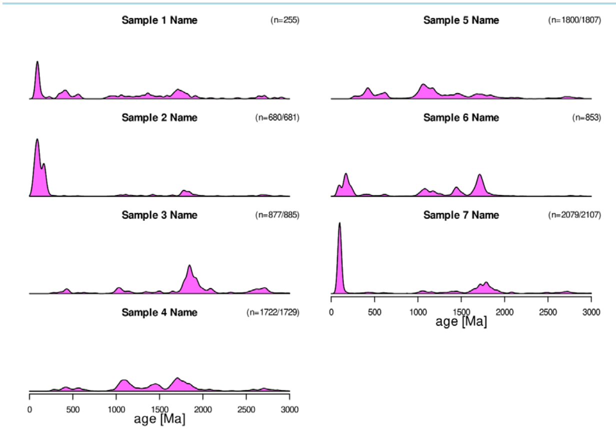
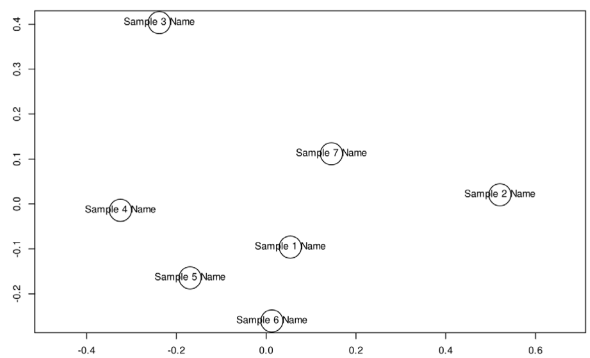
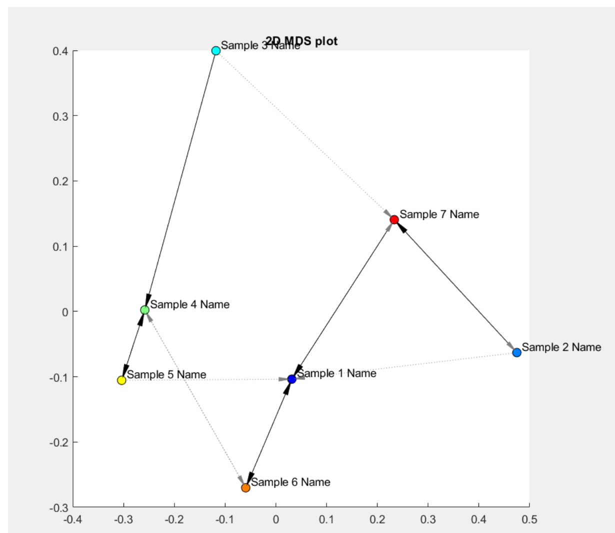


Figure 3 seems problematic for application of the WS distance to detrital geochronology. For example, Ranealven, Lainioalven, and Bysealven all share peaks at ~1800 Ma. The KS distance correctly locates these samples closest to each other. The WS distance in contrast locates the Ljusnan closer to the Ranealven than the Lainioalven even though the major age modes between the Ranealven and Lainioalven (~100 Myr offset) are closer than those of the Ranealvan and Ljusnan (~200 Myr offset). The WS Distance also locates the Vindelalven and Lainioalven equidistant from the Ranealven. The problem is that distinct source areas may have similar but non-overlapping age modes, and the WS distance is insensitive to these minor differences and therefore unable to discriminate them as distinct sources. A detrital geochronology distribution that has a mode at 1800 Ma and 2800 Ma is more likely to share a source with a sample with a distribution at 1800 Ma (i.e., one overlapping age mode) than one at 1700–1750 Ma (i.e., no overlapping age modes). The authors may be able to address this concern, but it seems to me to be a fatal flaw in this metric.

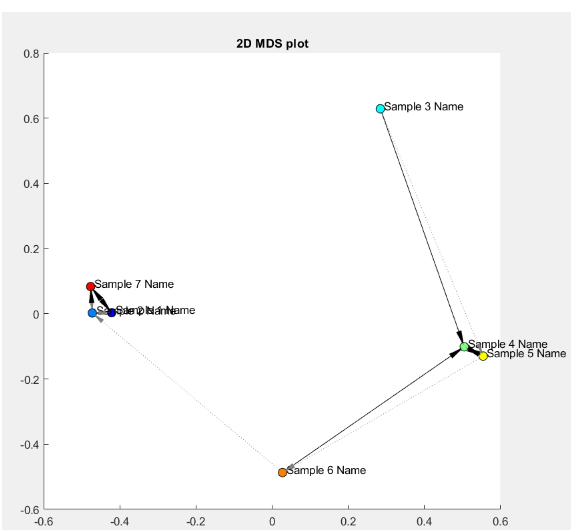
The problems with this metric can be seen when comparing some more complex age distributions. Simple visual inspection indicates that Samples 1, 2, and 7 should be quite similar.



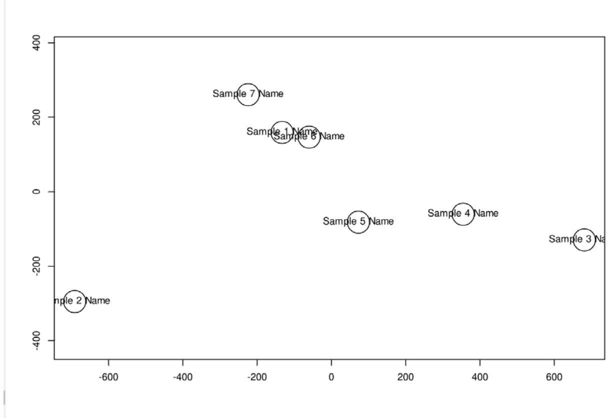




This is confirmed by a KS-based MDS (above left), or Kuiper-based MDS (above right).

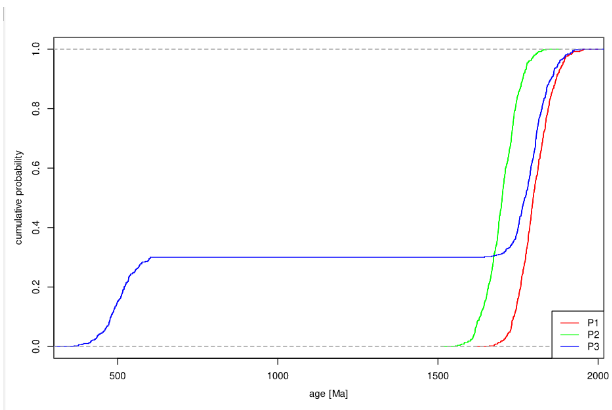


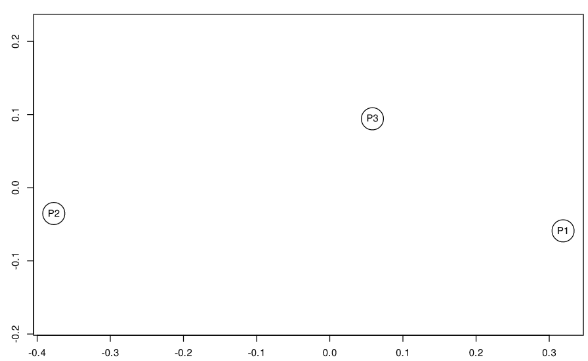
It is also confirmed by a cross-correlation-based MDS (above).

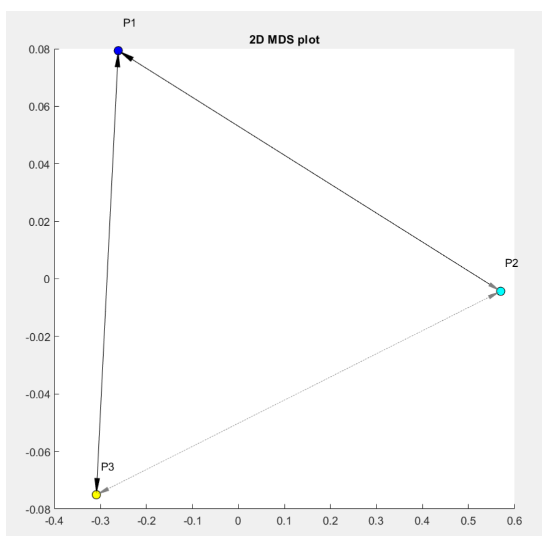


However, a WS Distance-based MDS (above) non-intuitively locates samples 1 and 6 quite close, and locates 1, 6, and 7 closer to 5 than to 2.

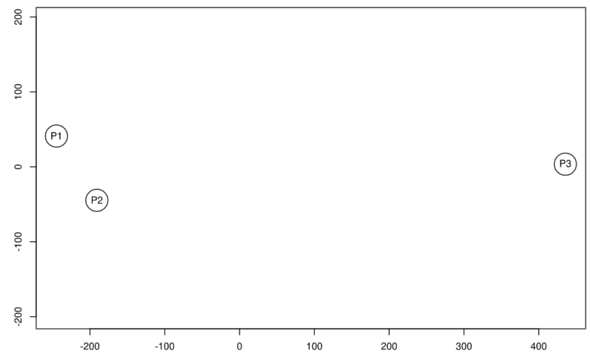
A second example highlights the disproportionate impact of misalignment of detrital age modes on the WS Distance. P1 and P3 below share age modes at 1800 Ma. P3 has an additional mode at 500 Ma. P2 shares no age modes with either P1 or P3 and rather has an age mode at 1700 Ma.







This is reflected in their KS MDS plot (above left) and an MDS plot based on cross-correlation (above right, note difference in x- and y-axis scale).



However, the WS Distance MDS (above) does not reflect the true relationship between age modes, but rather reflects the anomalous area between P1 and P3 (i.e., the horizontal distance between 500 Ma and 1800 Ma in the ECDF plot).

Detailed comments

The authors assert that cross-correlation is an ad hoc method, yet the Pearson coefficient which is the basis for the cross-correlation coefficient is widely used in seismic analysis, waveform analysis, and image analysis. As such, it is unclear what is meant by "ad hoc" in this context. Although Vermeesch (2018) correctly points out limitations of cross-correlation as applied to probability density plots of extremely precise data, these caveats do not apply to its application to kernel density estimates or kernel functional estimates. Similarly, the charge that Likeness is ad hoc is unfounded. Likeness is an adaptation of the L^1 norm applied, usually, to a 1D geochronology distribution. (However, see Sundell et al. (2021) for application of the L^1 norm to 2D distributions.) Finally, like the cross-correlation coefficient and Likeness, the Sircombe-Hazelton distance (L^2 norm, Sircombe and Hazelton, 2004; Vermeesch, 2018) also requires discretization of continuous functions for its calculation. I guess the take away from this is that it may be useful to compare the performance of cross-correlation and Likeness in addition to the KS Distance to newly applied metrics like the WS Distance.

It is ironic that after railing against use of the cross-correlation coefficient, the authors reintroduce it in section 2.2.

I am confused by the "Unimodal" vs "Multimodal" and "Older" vs "Younger" labels in Figure 3. To my eye the Ljungan is more prominently bimodal than the Byskealven, yet it plots closer to the Unimodal side of the figure than the Byskealven. Similarly, what portion of the distribution is "Younger" or "Older" when comparing multimodal distributions?

What are the stress values for the MDS plots in figure 3?