# Author's response

A. Lipp, P. Vermeesch

March 13, 2023

## Associate Editor

**Regarding comments by referee 1, the change of title seems justified. I also see that the manuscript aligns well with the "Short Communication" scope of the journal. The suggested expansion of contextual information would be helpful as short "onboarding" in the introduction or, maybe, the discussion. This needs to be evaluated after it has been implemented. The suggested actions to the line by line comments appear appropriate to me and I look forward to seeing the updated version of the manuscript in due course.**

We have resubmitted the manuscript as a 'Short Communication' as proposed. Additionally, we have revised the Introduction to better provide context on MDS, as well as the importance of dissimilarity metrics more generally. Additionally we have modified the manuscript according to the reviewer's line-by-line comments.

**Regarding comments by referee 2, I see that a large portion of the discourse arises from different perspectives on the same topic, which usually means that the clarity of the text needs to be improved in order to resolve misunderstandings from the beginning. I suggest to add clear statements about boundary conditions, assumptions and propositions early on (e.g., proposed subjectivity of visual inspection, non-unique sources for similar distribution patterns, etc.), to set the right expectation space for future readers. I would welcome a slightly more profound reevaluation of pros and cons as suggested by referee(s) and authors. In addition, it may help to clearly discuss the extreme endmember states of data and how the metrics behave for those, but also to discuss how the space in between such endmember states affects the KS and W2 results. I agree that the table 1 is of limited value in its current form. Perhaps replacing it by a 1-3 sentence verbal (and referenced) description of the geological context would be a useful way to provide the required background. I would encourage the authors to briefly add the the information on minimum sample size explicitly, to clarify that this has not been overlooked.**

In response to the constructive comments from reviewer 2 we have made substantial revisions to the manuscript. First, we now emphasise that there is no 'correct' dissimilarity metric for all scenarios. We have added the following to the Introduction, and similar text is added in the Discussion section:

> For all uses, the choice of which dissimilarity metric to use is vital as different metrics result in different numerical results and thus different geological interpretations. In general, the most appropriate metric will depend on the data being analysed and the scientific question under investigation.

Second, more substantially, we have added a new Discussion section (Section 3) which carefully weighs up and advantages and disadvantages of $W_2$ in various geochronological scenarios. In this section we present four different case studies, (three real, one synthetic) discussing whether the $W_2$ is preferable to the KS in each. We identify a 'rule-of-thumb' that if absolute age differences are deemed important information for the problem, $W_2$ is to be preferred. Elsewise, the KS distance may be preferable and the KS distance can be

unintuitive. This final scenario applies in cases of mixing of discrete endmembers, as identified by Reviewer 2. This new section is copied below.
Finally, due to the addition of these new examples, we decided that the example from Morton et al. (2008) was redundant, and has been removed from the manuscript.

The new Discussion section:

> As stated above, the most appropriate dissimilarity metric to use will depend on the data being analysed and the scientific question being answered. In general, the Wasserstein distance is most appropriate when absolute differences along the time axis (or more generally, the x-axis) provide useful information to solving the geologic problem. The KS distance however is more appropriate when the size of the time differences between peaks is not relevant. Both the KS distance and the $W_2$ are calculated in terms of differences between ECDFs. Due to these similarities in construction, in many cases the results from using the KS and $W_2$ are, encouragingly, similar. One exception is whether ages are log transformed prior to analysis. Because the KS distance considers only the order of the ages, it will be the same whether a log transform is used or not. $W_2$ however will be different, and will consider *relative* not absolute age differences. Such an example is discussed below (Figure 3).
>
> Here we discuss a variety of realistic scenarios where the KS and $W_2$ may result in different interpretations. In each, we evaluate the advantages and disadvantages of using $W_2$ or KS. These case-studies can be used to determine which metric is most appropriate for a particular scenario.
>
> ## Discriminating contributions from discrete endmembers
>
> We first consider a scenario where the samples are assumed to be mixtures, in differing proportions, of some known or unknown fixed endmembers. This situation is one where absolute distance along the time-axis is not relevant, as the nature of the endmembers is not sought, simply their relative contributions to a set of mixtures. Instead, it is *vertical* differences in the probability at a given age that is relevant. The KS distance, which is sensitive to such vertical differences in age distributions is better suited for this than $W_2$. Indeed, in such a scenario the $W_2$ can result in some unintuitive behaviour.
>
> For example, let us consider three unimodal potential sediment sources, as shown in Figure 1a. We now consider two mixture samples. The first is an equal mixture of X and Y, and the second an equal mixture of Y and Z (bottom two plots, Figure 1a). Geologically, we would expect these samples to be about half as similar to the two source endmembers. However, a $W_2$ MDS map identifies these samples as being removed from their two endmembers 1b. Additionally, because of the absolute time difference between Source Z and the other sources, Sample 2 is treated as a considerable outlier. The KS distance performs better here, placing the mixtures approximately halfway between the expected endmembers. However, in such a well defined mixing scenario as this, methods such as endmember mixture modelling may be more appropriate than statistical dimension reduction (e.g., Weltje 1997; Sharman et al. 2017; Dietze et al. 2019).
>
> ## Temporally varying source age distributions
>
> In contrast, scenarios where the shape of sediment source age distributions evolves in space and time are well suited to using $W_2$. This is because $W_2$ considers all parts of a distribution, whereas the KS only compares one point, the location of maximum ECDF separation. For example, Figure 2 displays detrital zircon age distributions gathered by DeGraaff-Surpless et al. (2002) from sediments from a section (Cache Creek) across the Great Valley Group in California, USA. The age populations are shown as KDEs and histograms, in stratigraphic order, in Figure 2a.

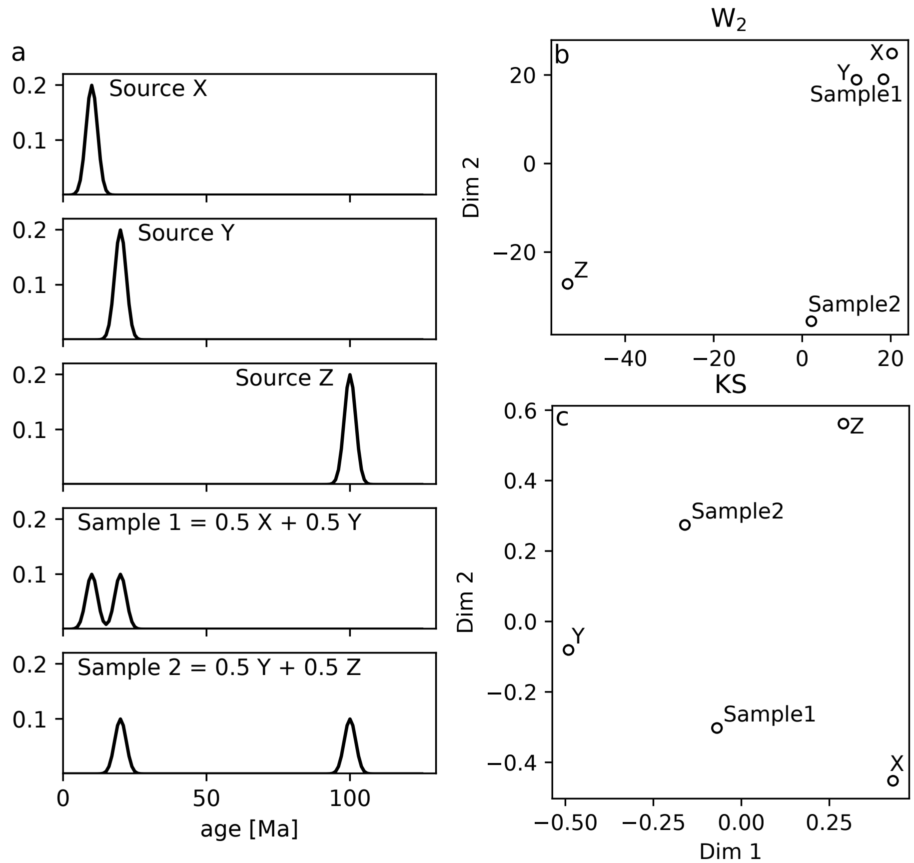


Figure 1: **Mixing of discrete endmembers.** a) Three theoretical, unimodal source age distributions with peaks at 10, 20 and 100 Ma, and two mixture samples. Sample 1 is an equal mixture of X and Y and Sample 2 a mixture of Y and Z. b) Metric MDS map of the three sources and the mixtures using $W_2$ distance. c) Same as panel b for KS distance.

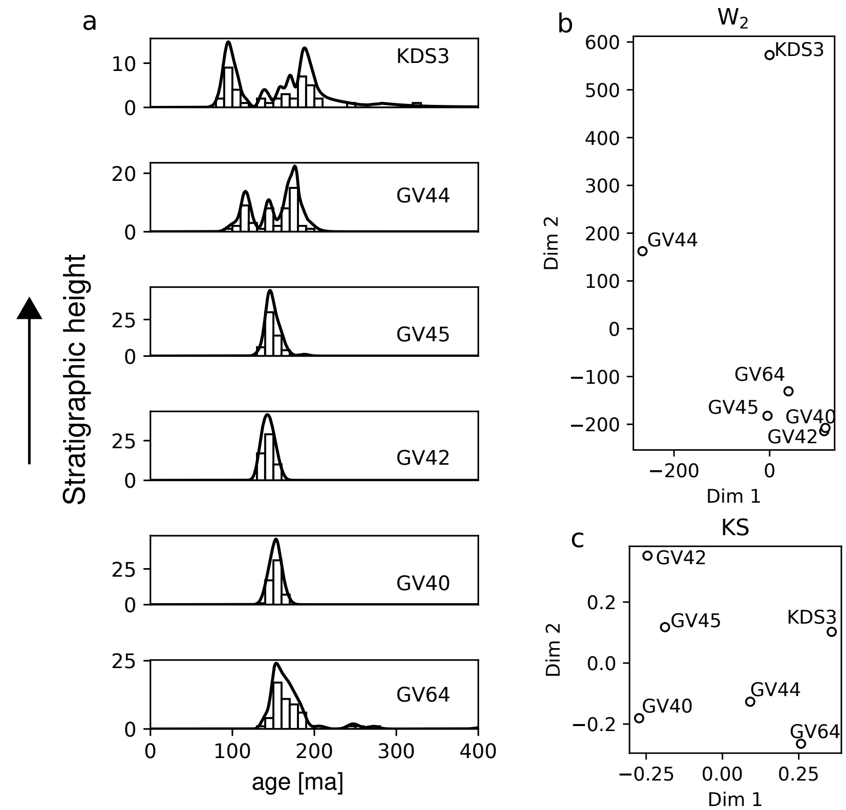


Figure 2: **Temporally evolving source distributions.** a) KDEs and histograms for zircon age distributions for samples from Cache Creek section across Great Valley Group, arranged in stratigraphic order (DeGraaff-Surpless et al. 2002). b) MDS map using $W_2$ for data shown in panel a. c) Same as b using KS distance.

The uppermost samples show an increasingly broad distribution than the lower four unimodal samples. DeGraaff-Surpless et al. (2002) attribute this trend, *inter alia*, to expanding sediment source areas.

Figures 2b–c display MDS maps calculated using $W_2$ and KS respectively. The $W_2$ map clearly identifies the stratigraphic order of the samples by the changing distribution shape. Additionally, it clusters the four unimodal samples together. By contrast, the KS map does not identify the stratigraphic trend, locating the lowermost stratigraphic sample GV64 with the uppermost samples KDS3 and GV44. We conclude then that the $W_2$ has better captured the geological information in this scenario.

## Thermochronology

In thermochronology, age distributions shift along the time-axis according to thermal signals (e.g., exhumation). In many thermochronological studies, we may seek to characterise how such a signal evolves in space and time. For this question absolute distance along the time-axis is useful information and so the $W_2$ may be more effective than the KS distance. For example, Wobus et al. (2003) use $^{40}$Ar/$^{39}$Ar detrital mica thermochronometry to explore spatially varying exhumation along a spatial transect in the Himalaya. The KDEs of the samples are shown in Figure 3a arranged south to north. The southern samples (WBS1, WBS2, WBS3, WBS8) show old exhumation signals, but a dramatic shift to younger ages is observed north of a distinct physiographic transition. MDS maps of these samples are shown using the KS distance and $W_2$ in Figures 3b–c respectively. As there is limited overlap between the samples, the KS distance struggles to capture the NS progression in exhumation age. Whilst the physiographic division is found, it weights it equally to variation within one cluster. By contrast, the $W_2$ map correctly identifies the simple temporal and geographical trend of the samples from south to north.

## Combining data from multiple laboratories

A final scenario where the $W_2$ could be preferable is when comparing samples from different laboratories which are affected by inter-laboratory bias. Košler et al. (2013) provided ten different laboratories with identical synthetic zircon samples with a known age distribution. Different instruments introduced small differences in the ages of each peak. For example, in Figure 4 we display the results from Lab 1 (red) and Lab 4 (pink) as KDEs. The expected peak at $\sim$ 1200 Ma (dashed line) is offset between the two samples. As it is the maximum distance between two ECDFs, the KS distance is very sensitive to minor offsets in sharply defined peaks. In this case, the KS distance between these theoretically identical samples is large at 0.348, which is over one third of the maximum possible distance between samples. Indeed, the KS distance considers a synthetic, purposefully misaligned series of peaks (black KDE) to be more similar to the Lab 4 results than the results from Lab 1. The $W_2$ distance, does not suffer from this oversensitivity to minorly offset peaks and correctly identifies the samples from Lab 1 and Lab 4 as being much more similar than the random synthetic distribution.

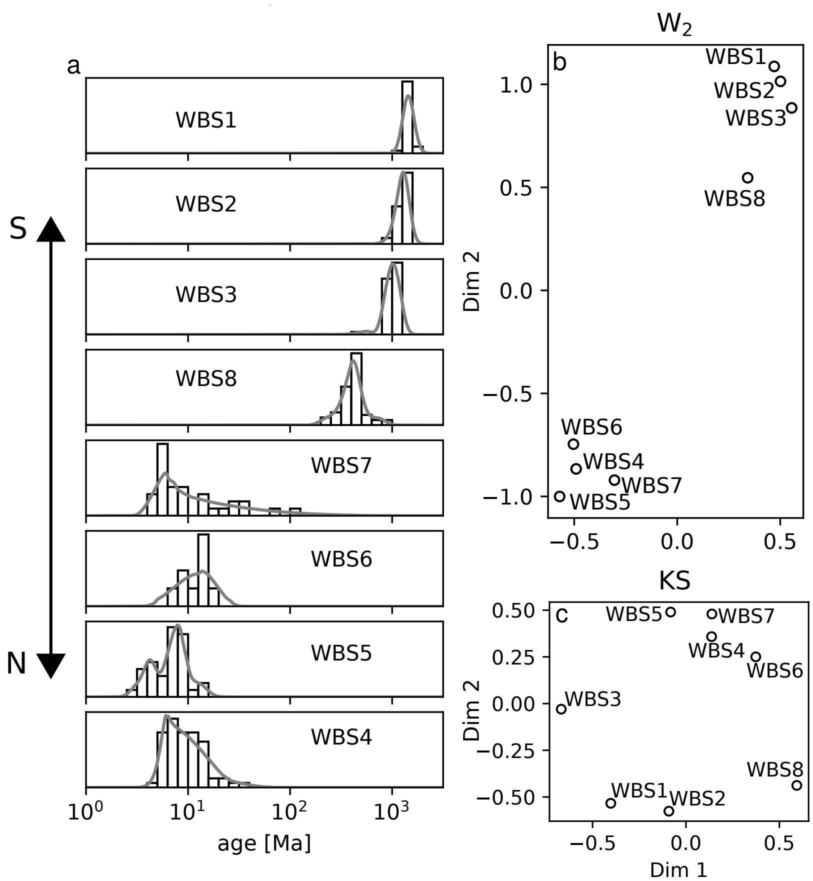


Figure 3: **Analysing thermochronological data using $W_2$ and KS distances.** a)KDEs for a detrital mica $^{40}$Ar/$^{39}$Ar dataset of Wobus et al. (2003) arranged from south to south across a physiographic transition of the central Himalaya in Nepal. Note the logarithmic scale. b) The MDS configuration using $W_2$, following a log transform. c) MDS map using KS statistic. In this example, $W_2$ performs better than the KS distance at identifying the geographic trend.

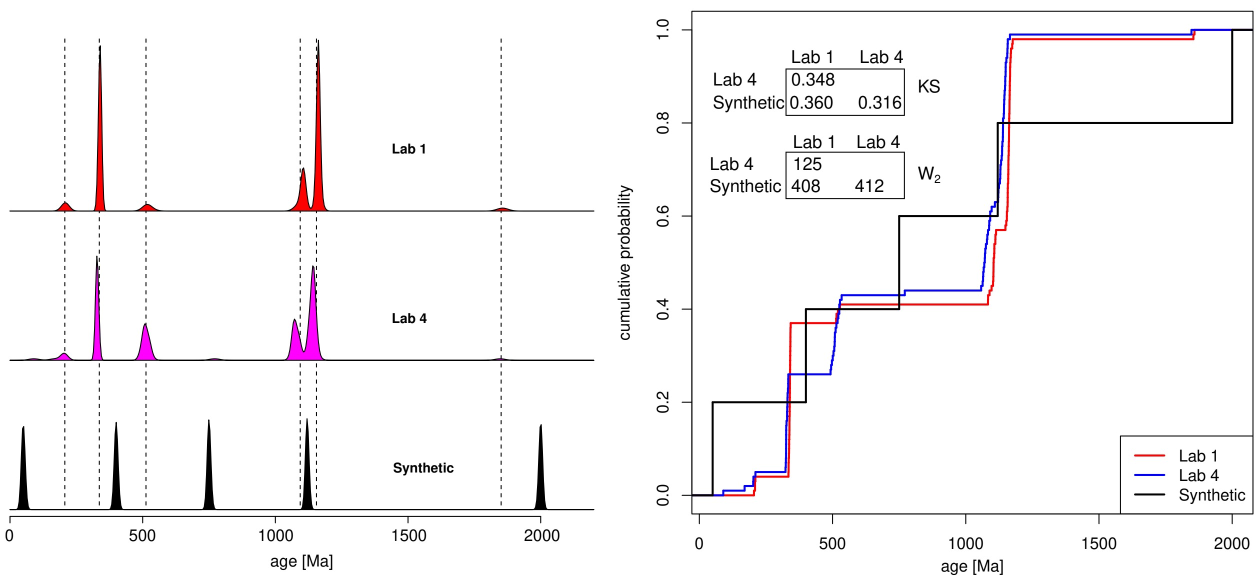


Figure 4: KDEs (left) and ECDFs (right) of two samples from the inter-laboratory comparison study of (Košler et al. 2013), plus a purposefully misaligned synthetic sample. Dashed lines mark the true ages of the detrital mixture. According to the KS-statistic, the age distribution produced by Lab 4 is more similar to the synthetic distribution than it is to the distribution produced by Lab 1, despite the absence of any shared age components. The $W_2$ distance correctly deems the distribution produced by Lab 4 to be closer to that of Lab 1 than to the synthetic mixture.

## Reviewer 1

**Minor point: The title seems to promise more than the manuscript delivers. "Comparing . . . ". The manuscript does not compare something. The presented "comparison" is a performance test of the Wasserstein-2 distance and the Kolmogorov-Smirnov distance. I think the title should reflect better what the manuscript tries to achieve: a presentation of an alternative metric in the realm of multidimensional scaling.**

To make the title more clear we have revised the manuscript title to now read as: *'Short Communication: The Wasserstein distance as a dissimilarity metric for comparing detrital age spectra, and other geological distributions'*

**The general idea of the manuscript fits within the scope of GChron. However, I am in a little bit of doubt about whether it justifies requesting a peer-review procedure and a peer-review publication. The numerical metric presented here is not new, and the manuscript does not (yet?) show significant scientific progress. The implementation in R appears limited to a few code lines. Perhaps under different circumstances, the implementation in R would have remained a single line in a news files, along with a few lines in the package manual or an entry in a science blog.**

As proposed, we are re-submitting the manuscript as a 'Short Communication'.

**As a non-expert in multidimensional scaling, I feel the manuscript would benefit from more context. The formal description is sufficient and easy to follow, but the likely impact of this manuscript seems low except for having announced a 'new' feature.**

We have now revised the introduction to improve the background coverage of MDS and also to better emphasise the importance of dissimilarity measures. The new introductory paragraph reads as follows:

> A distributional dataset is one where the information does not lie in individual observations, but in the *distribution* of many observations associated with one sample. Such data are common in the geological sciences, for example, detrital mineral ages or grain size distributions. Zircon U-Pb ages, in igneous and detrital samples, are one particularly widely used class of distributional data, which are used *inter alia* to constrain sediment provenance, global magmatic processes, and the evolution of plate tectonics (e.g., Condie et al. 2009; Cawood et al. 2012; Reimink et al. 2021). Grainsize distributions are another common form of geological distributional data. Analytical advances mean that increasingly large amounts of distributional data are being generated in the Earth sciences meaning that qualitative comparison of samples is becoming infeasible, and objective dissimilarity metrics between samples must be used. Some measure of dissimilarity (or more specifically, distance) is also required for many widely used statistical methods such as clustering, ANOVA, and dimension reduction. Dissimilarity metrics in geochronology at present are most commonly used for dimension reducing techniques such as multi-dimensional scaling (MDS) or principal component analysis (PCA). Such methods have become popular for analysing large numbers of detrital age spectra simultaneuously (Vermeesch 2013; Sharman et al. 2018; Vermeesch 2018a). Fitting models (e.g., sediment source partitioning) using distributional data also requires a definition of dissimilarity for comparing observed and predicted distributions (e.g., Amidon et al. 2005; De Doncker et al. 2020).

**In other words: How does this new measure perform for real samples and their (new) interpretation? Section 4 reads interesting, but was a new conclusion reached? Did it lead to better (e.g., more accurate, more precise) results, or did the geoscientific interpretation essentially**

**remains the same? If the latter is the case, perhaps you can present a real case underlining the point you want to make better.**

In the new Discussion section (Section 3 in the revised manuscript) we prevent three real and one synthetic example where the behaviour of KS and $W_2$ are evaluated in detail. We show that in the examples from DeGraaff-Surpless et al. (2002), Wobus et al. (2003), and Košler et al. (2013), the KS distance provides unsatisfying solutions (and inaccurate in the case of Košler et al. 2013) whereas the $W_2$ distance better captures geological intuition.

**The manuscript comes without a proper discussion. Section 4 is an application example that includes elements of a discussion. However, for a scientific manuscript, I would expect to see more. In particular I would like to see a discussion about the question: Does it likely change the outcome of studies working with this 'new' metric.**

As mentioned in the response to the previous comment we have now added a Discussion section (Section 3) which discusses four different scenarios where the use of the $W_2$ distance may (or may not) improve the outcome of different geochronological studies.

**The synthetic data outlines the general problem you want to address. I suggest leading with an example based on a case study where the Kolmogorov-Smirnov distance did not perform as expected for the reasons you have mentioned.**

We believe this has been resolved by the inclusion of three examples in the Discussion section in which the KS distance performs in an unexpected manner, in particular using the data of Košler et al. (2013), where the KS distance fails to identify two theoretically identical synthetic samples as being similar.

## Line-by-line comments

**L111: I've played a bit with the proposed synthetic data and found that it depends to some extent on the standard deviation. A more narrow standard deviation for the same fixed mean values leads to more complex KS-distance patterns. The higher the degree of overlap (higher standard deviation), the more conclusive the KS distance becomes. Perhaps you can add a few lines about it in the text.**

Whilst we agree that this is a useful observation we could not identify an immediate geological parallel for this specific behaviour, and as such felt it may be beyond the scope of the manuscript.

**L150-L175: I think this paragraph can be improved in order to provide a better experience to readers.**

This section has been revised to now read as follows:

> Additionally, the $W_2$-distance has been added to the `IsoplotR` package in `R`, which calculates dissimilarity matrices and MDS maps (Vermeesch 2018b). This software can be accessed using an (online) graphical user interface, at `isoplotr.es.ucl.ac.uk`. Alternatively, the function can also be accessed from the `R` command line. The following snippet uses $W_2$ to calculate an MDS map for the dataset from Wobus et al. (2003) discussed in the manuscript (Figure 5). The data required is also available at the above repository. Note that the MDS map produced may show slight differences to those in the manuscript due to dependence of metric MDS on a random state variable.

```
# load the package:
library(IsoplotR)
DZ <- read.data("wobus.csv",method="detritals")
# example 1. calculate the W2 distance matrix for the dataset:
d <- diss(DZ,method="W2")
# example 2. apply MDS to the dataset:
mds(DZ,method="W2")
```

**L164: Please consider adding the example data to the manuscript or the R package**

The example data has been added to the data repository.

**L167+ (footnote): The repository pvermees/IsoplotRbeta does not exist, but I guess the branch beta was meant and it should read: remotes::install_github('vermeesch/IsoplotR@beta')**

The link has been corrected to: `isoplotr.es.ucl.ac.uk`.

**Figure 1: How did you modify the data to "aid illustration"? It appears that you have shifted the 'Byskealven' dataset by 1 Ma. This should be mentioned. If so, how does it affect the W1 distance? Your pink area becomes considerably smaller if the non-shifted dataset is used. Perhaps you have a better dataset at hand; one that does not need such manipulation.**

As discussed in the previous response to Authors we not believe that modifying this Figure provides greater geological insight.

**Figure 2: The upper plot would benefit from y-axis labelling**

We have made the axis labelling of MDS plots consistent throughout the manuscript.

**Figure 3: Something is at odds with the lower figures (j and k), if I try to reproduce them with IsoplotR::mds() and the example data from GitHub. Figure 3j does not look as in the manuscript, because 'Ljusnan' and 'Byskealven' seem to be no different. Figure 3k is somewhat mirrored. Below my code, I first used IsoplotR():mds(), here more manually to show the calculation steps. The mirrored figure is not a big deal because the interpretation should not change, but it should be presented as the users would see it running the code.**

We have added the following text to the 'Implementation' section (Line 199) to emphasise that results of MDS calculations may vary: '*Note that the MDS map produced may show slight differences to those in the manuscript due to dependence of metric MDS on a random state variable.*'

## Reviewer 2

**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).**

We agree with the reviewer that in this particular scenario (mixing of discrete and known sources) $W_2$ may not be preferable, and now discuss this, with an example, in the new Section 3.1 of the manuscript. It is presented along with three other scenarios where $W_2$ is found to be preferable to the KS distance. This new section is copied above.

**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.**

We agree with the reviewer that $W_2$ may not be relevant for all scenarios. As a result we have now added the new Discussion section to consider cases where $W_2$ may or may not be preferred. We explicitly describe how $W_2$ is preferable when absolute age differences are useful information, and the KS distance, when such information is not relevant. For example on Line 143 we now state:

> In general, the Wasserstein distance is most appropriate when absolute differences along the time axis (or more generally, the x-axis) provide useful information to solving the geologic problem. The KS distance however is more appropriate when the size of the time differences between peaks is not relevant.

**The authors need to explore the behavior of WS Distance for multi-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.**

In Section 3.2 ('Temporally varying source age distributions') we discuss how $W_2$ is able to accurately distinguish multi-modal samples from uni-modal ones. We note that in this particular example, the KS distance is less effective at identifying the change in distribution shape with stratigraphic height. This section is copied above.

**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.**

Given that we have now added four more examples (three with real data), we decided that this original example from Morton et al. (2008) was no longer needed and have removed it from the revised manuscript.

**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$\gg$100 (ideally n$>$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.**

We have removed the example from Morton et al. (2008) from the manuscript. As such, whilst we agree with the reviewer about the importance of sufficient numbers of grains, this comment is no longer relevant.

In the new Discussion section (Section 3), we provide three examples where the $W_2$ distance is better able to extract geological information from distributional data than the KS distance. Note however, that we also include an example (Section 3.1) that discusses where the $W_2$ sample may *not* provide greater insight over the KS distance.

**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).**

We agree with the reviewer that in some scenarios (including the one indicated by the the reviewer) the $W_2$ may not be appropriate. In the new Discussion of the manuscript (see Section 3.1) we now consider a case where absolute distance along the time-axis may not be useful information, and as a result $W_2$ behaves unintuitively.

**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.**

As discussed above, this example has been removed from the manuscript.

**The problems with this metric can be seen when comparing some more complex age distri-**

**butions. 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.**

As discussed in the previous Author's comment we do not believe that in the example provided the $W_2$ behaved unintuitively. We now mention twice (Lines 34 & 129) that there is no one correct dissimilarity metric and that different metrics will sometimes produce different results. We believe the new Discussion section will equip readers to decide whether the $W_2$ is appropriate for any particular scenario.

**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).**

We agree with the reviewer that the $W_2$ distance can be unintuitive in scenarios of mixing between well defined age peaks. This limitation is discussed in the new Discussion Section (specifically, Section 3.1) of the manuscript, copied above.

**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 L1 norm applied, usually, to a 1D geochronology distribution. (However, see Sundell et al. (2021) for application of the L1 norm to 2D distributions.) Finally, like the cross-correlation coefficient and Likeness, the Sircombe-Hazelton distance (L2 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.**

Detailed responses to these comments are provided in the previous author comments.

**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?**

This example has been removed the manuscript.

# References


Amidon, W. H., D. W. Burbank, and G. E. Gehrels (2005). "Construction of detrital mineral populations: insights from mixing of U–Pb zircon ages in Himalayan rivers". *Basin Research* 17.4, pp. 463–485.

Cawood, P., C. Hawkesworth, and B. Dhuime (2012). "Detrital zircon record and tectonic setting". *Geology* 40.10, pp. 875–878.

Condie, K. C., E. Belousova, W. L. Griffin, and K. N. Sircombe (2009). "Granitoid events in space and time: Constraints from igneous and detrital zircon age spectra". *Gondwana Research*. Special Issue: Supercontinent Dynamics 15.3, pp. 228–242.

De Doncker, F., F. Herman, and M. Fox (2020). "Inversion of provenance data and sediment load into spatially varying erosion rates". *Earth Surface Processes and Landforms* 45.15, pp. 3879–3901.

DeGraaff-Surpless, K., S. A. Graham, J. L. Wooden, and M. O. McWilliams (2002). "Detrital zircon provenance analysis of the Great Valley Group, California: Evolution of an arc-forearc system". *GSA Bulletin* 114.12, pp. 1564–1580.

Dietze, E. and M. Dietze (2019). "Grain-size distribution unmixing using the R package EMMAgeo". *E&G Quaternary Science Journal* 68.1, pp. 29–46.

Košler, J., J. Sláma, E. Belousova, F. Corfu, G. E. Gehrels, A. Gerdes, M. S. A. Horstwood, K. N. Sircombe, P. J. Sylvester, M. Tiepolo, M. J. Whitehouse, and J. D. Woodhead (2013). "U-Pb Detrital Zircon Analysis – Results of an Inter-laboratory Comparison". *Geostandards and Geoanalytical Research* 37.3, pp. 243–259.

Morton, A., M. Fanning, and P. Milner (2008). "Provenance characteristics of Scandinavian basement terrains: Constraints from detrital zircon ages in modern river sediments". *Sedimentary Geology* 210.1, pp. 61–85.

Reimink, J. R., J. H. F. L. Davies, and A. Ielpi (2021). "Global zircon analysis records a gradual rise of continental crust throughout the Neoarchean". *Earth and Planetary Science Letters* 554, p. 116654.

Sharman, G. R. and S. A. Johnstone (2017). "Sediment unmixing using detrital geochronology". *Earth and Planetary Science Letters* 477, pp. 183–194.

Sharman, G. R., J. P. Sharman, and Z. Sylvester (2018). "detritalPy: A Python-based toolset for visualizing and analysing detrital geo-thermochronologic data". *The Depositional Record* 4.2, pp. 202–215.

Vermeesch, P. (2013). "Multi-sample comparison of detrital age distributions". *Chemical Geology* 341, pp. 140–146.

– (2018a). "Dissimilarity measures in detrital geochronology". *Earth-Science Reviews* 178, pp. 310–321.

– (2018b). "IsoplotR: A free and open toolbox for geochronology". *Geoscience Frontiers*. SPECIAL ISSUE: Frontiers in geoscience:A tribute to Prof. Xuanxue Mo 9.5, pp. 1479–1493.

Weltje, G. J. (1997). "End-member modeling of compositional data: Numerical-statistical algorithms for solving the explicit mixing problem". *Mathematical Geology* 29.4, pp. 503–549.

Wobus, C. W., K. V. Hodges, and K. X. Whipple (2003). "Has focused denudation sustained active thrusting at the Himalayan topographic front?" *Geology* 31.10, pp. 861–864.