# ~~Comparing~~ Short Communication: The Wasserstein distance as a dissimilarity metric for comparing detrital age spectra, and other geological distributions~~, using the Wasserstein distance~~

Alex Lipp[1] and Pieter Vermeesch[2]

[1]Merton College, University of Oxford, Oxford, UK
[2]Department of Earth Sciences, University College London, London, UK

**Correspondence:** Alex Lipp (alexander.lipp@merton.ox.ac.uk)

**Abstract.**

Distributional data such as detrital age populations or grain size distributions are common in the geological sciences. As analytical techniques become more sophisticated, increasingly large amounts of distributional data are being gathered. These advances require quantitative and objective methods, such as multidimensional scaling (MDS), to analyse large numbers of samples. Crucial to such methods is choosing a sensible measure of dissimilarity between samples. At present, the Kolmogorov-Smirnov (KS) statistic is the most widely used of these dissimilarity measures. However, the KS statistic has some limitations ~~. It is very sensitive~~ such as high sensitivity to differences between the modes of two distributions, and ~~relatively insensitive to differences between~~ insensitivity to their tails. Here we ~~introduce~~ propose the Wasserstein-2 distance ($W_2$) as an alternative ~~to address this issue~~metric for use in geochronology. Whereas the KS-distance is defined as the maximum vertical distance between two empirical cumulative distribution functions, the $W_2$-distance is a function of the horizontal distances (i.e., age differences) between ~~individual~~ observations. Using a ~~combination of synthetic examples and a published zircon U-Pb dataset, we show that the~~ variety of synthetic and real datasets we explore scenarios where $W_2$ ~~distance produces similar MDS results to the KS-distance in most cases, but significantly different results in some cases . Where the results differ, the~~ may provide greater geological insight than the KS statistic. We find that in cases where absolute time differences are not relevant (e.g., mixing of known, discrete age peaks), the KS statistic can be more intuitive. However, in scenarios where absolute age differences are important (e.g., temporally/spatially evolving sources, thermochronology, and overcoming laboratory biases) $W_2$ ~~results are geologically more sensible.For the case study, we find that the MDS map that is produced using $W_2$ can be readily interpreted in terms of the shape and average age of the age spectra~~is preferable. The $W_2$-distance has been added to the `R` package `IsoplotR`, for immediate use in detrital geochronology and other applications. The $W_2$ distance can be generalised to multiple dimensions, which opens opportunities beyond distributional data.

## 1 Introduction

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 ~~we require objective and quantitative ways to analyse~~ increasingly large amounts of distributional data ~~. Qualitative comparison becomes infeasible when even modest numbers of samples are being analysed. For example, the dimension reducing technique of~~ 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) ~~has~~ 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).This method, and others, require a dissimilarity metric between samples to be specified (Vermeesch, 2018a).Such a metric corresponds to how 'different' two distributional samples are.The choice of metric~~ (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; D ).

For all uses, the choice of which dissimilarity metric to use is vital as different metrics ~~can~~ result in different ~~MDS 'maps' and potentially~~ 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. The Kolmogorov-Smirnov (KS) distance, calculated as the maximum vertical distance between two empirical cumulative distribution funtions (ECDFs) has emerged as a 'canonical' distance metric between mineral age distributions (Berry et al., 2001; Vermeesch, 2018a). However, the KS-distance has a number of drawbacks, chiefly that as only the *maximum* vertical difference between ECDFs is important, it is insensitive to variability in the tails of distributions. A number of alternative dissimilarity measures have previously been proposed to address this issue, including established methods such as the Kuiper statistic, and ad-hoc dissimilarity measures such as the 'likeness' and 'cross-correlation' coefficients ~~(Satkoski et al., 2013; Saylor et al., 2012; Sharman~~ (Satkoski et al., 2013; Saylor et al., 2012). Unfortunately, ~~all~~ these alternatives have drawbacks, including a propensity for the ad-hoc dissimilarity measures to produce unintuitive results when applied to extremely large and/or precise datasets (Vermeesch, 2018a).

In this paper we present an alternative to the KS-distance that does not suffer from ~~these drawbacks~~some of these limitations: the Wasserstein distance (also known as the Earth-mover's or Kantorovich–Rubinstein distance). To introduce the chief principle behind this measure, let us consider a simple toy example. Table 1 contains four samples ($A$ through $D$), each of which contains exactly one single grain analysis:

As the KS distance is the vertical difference between ECDFs, it is insensitive to the absolute, 'horizontal' age differences between individual observations. Thus, the KS-distances between $A$ and the other three samples are $KS(A,B) = 0$, $KS(A,C) = 1$ and $KS(A,D) = 1$. Counter to our expectation, the KS-distance cannot 'see' the relative age difference be-

**Table 1. A toy, single-grain per sample dataset**

| Sample | A | B | C | D |
|---|---|---|---|---|
| Age, Ma | 1 | 1 | 2 | 11 |

tween sample $A$ and samples $C$ and $D$. For the toy example, the Wasserstein distance simply corresponds to the horizontal distance between the four samples. Thus, $W(A,B) = 0$, $W(A,C) = 1$, and $W(A,D) = 10$, which is a more sensible result than that achieved with the KS-distance.

In the following sections, we first introduce the Wasserstein distance in a more realistic setting, and formally define it. Next we discuss how it can be decomposed into intuitive terms that accord with how qualitatively, as geologists, we might compare distributions. We then proceed to compare the Wasserstein distance to the KS distance using a simple yet realistic synthetic example. Finally, we ~~perform a case study, analysing eight real zircon U-Pb age spectra from Scandinavian river sediments using MDS with the Wasserstein distance~~analyse a series of case studies, analysing real datasets using both the Wasserstein and KS distances. We thus evaluate the benefits and drawbacks of both metrics, identifying scenarios when one metric may be preferred to the other. Whilst we focus primarily on detrital age distributions, we emphasise that much of the following discussion applies equally to any form of distributional data.

## 2 The Wasserstein distance

The Wasserstein distance is a distance metric between two probability measures from a branch of mathematics called ‘optimal transport’. Optimal transport is often intuited in terms of moving piles of sand from one location to another with no loss or gain of material (e.g., Villani 2003). The problem that optimal transport solves is finding the way to transport the sand such that the least sand is moved the least distance. The Wasserstein distance is the cost associated with this most efficient transportation. The association with moving piles of sand is why the Wasserstein distance is often termed the Earth-mover’s distance. Figure 1a shows an example of how one univariate probability distribution, $\mu$, based on a detrital age spectrum, is transformed into another, $\nu$ according to the optimal transport plan. ~~Like the KS-distance the Wasserstein defines a metric space, satisying the triangle inequality.~~ Elsewhere in the Earth sciences, the Wasserstein distance is increasingly ~~being~~ used for solving non-linear geophysical inverse problems (e.g., Engquist and Froese 2014; Métivier et al. 2016; Sambridge et al. 2022) and has been proposed as a tool for fitting hydrographs (Magyar and Sambridge, 2023). Full mathematical treatments of the Wasserstein distance and optimal transport are beyond the scope of this paper, but interested readers are referred to Villani (2003) or Peyré and Cuturi (2019). A geophysical perspective is given in Sambridge et al. (2022).

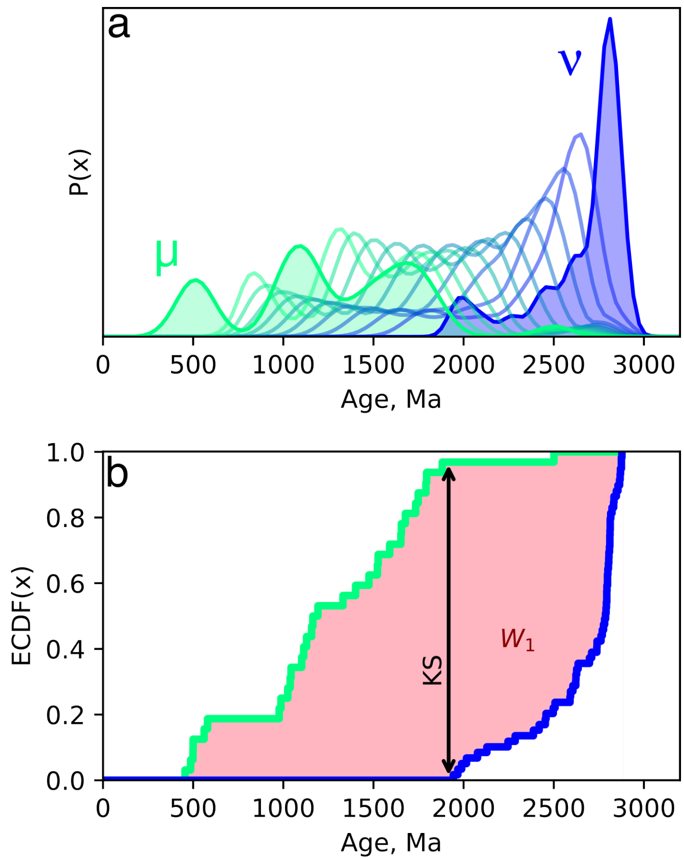


**Figure 1. Intuition of the Wasserstein distance.** a) Green and blue filled polygons show two example ~~probability distributions~~ kernel density estimates of mineral ages from two samples ~~.~~ (based on data from Morton et al. 2008) The distributions are labelled $\mu$ and $\nu$ for consistency with Equation 1. Semi-transparent coloured lines are probability distributions spaced equally in Wasserstein space between $\mu$ and $\nu$ (termed 'barycentres'; Benamou et al. 2015). b) Empirical Cumulative Distribution Functions (ECDFs) of the detrital ages used to calculate the distributions shown in panel a, same colours. The first Wasserstein ($W_1$) distance corresponds to the total area between the two ECDFs (shaded pink). The Kolmogorov-Smirnov (KS) distance is the maximum distance between the two ECDFs (black double-headed arrow).~~The data used to generate these distributions is taken from the 'Byskealven' and 'Vefsna' samples of Morton et al. (2008), but modified to aid illustration.~~

### 2.1 Formal definition

We consider two univariate probability distributions $\mu$ and $\nu$ which have cumulative distribution functions (CDFs) $M$ and $N$ respectively. The $p^{\mathrm{th}}$ Wasserstein distance between $\mu$ and $\nu$ is given by:

$$W_p(\mu,\nu) = \left( \int_0^1 |M^{-1} - N^{-1}|^p \mathrm{d}t \right)^{1/p}. \tag{1}$$

where $M^{-1}$ indicates the inverse of the CDF $M$ and $0 \leq t \leq 1$ (Villani, 2003). Note that this definition of $W_p$ assumes that the cost-function is given by $|x-y|^p$ (e.g., the Euclidean distance where $p = 2$), which is the case for most distributional data in geology. In the further special case of $p = 1$ (i.e., the *first* Wasserstein distance, $W_1$), Equation 1 can be re-written simply as:

$$W_1(\mu,\nu) = \int_X |M - N| \mathrm{d}x, \tag{2}$$

which is the area between two CDFs (e.g., Figure 1b). Recall that the KS-distance between two distributions is the maximum distance between the two corresponding CDFs. Whilst the $W_1$ is easily visualised, we actually use the $W_2$ going forwards as the *squared* distance (i.e., $p = 2$) between observations is the standard distance metric in most statistical analyses (e.g., least squares regression). Additionally, $W_2$ decomposes into readily interpretable terms, as discussed below.

We focus on these univariate instances as they apply to the most common geological distributional data including detrital age distributions and grain size distributions. However, we note that the Wasserstein distance is, in general, multivariate. As a result, some form of the Wasserstein distance could prove useful for analysing a number of other geological datasets such as the geochemical compositions of detrital minerals, or joint U-Pb and Lu-Hf isotope analysis (see Vermeesch et al. 2023). Statistics for comparing distributional data in multiple dimensions are increasingly needed (Sundell and Saylor, 2021).

~~A property of the KS-distance is that it is insensitive to whether the data are presented as 'raw' or log-transformed ages~~Like the KS distance, $W_2$ satisfies the triangle inequality, and as such is a true metric. This property ~~arises as the KS-distance is only sensitive to the relative ordering of observations in a distribution, which is insensitive to a log transformation. The~~ means that classical, as well as metric & non-metric MDS can be used with a $W_2$ ~~however will give different results depending on whether the data are transformed or not~~defined dissimilarity matrix. As $W_2$ is sensitive to absolute time differences, metric (or classical) MDS, which seek to preserve absolute distances, may be preferable to non-metric MDS. For the ~~remainder of this study we consider only raw ages, focussing as a result on *absolute* age differences. However, we can conceive of situations in which it is *relative* age differences which are of interest, in which case a logarithmic transformation would be applied prior to calculating $W_2$~~rest of this manuscript, metric MDS is used.

### 2.2 Decomposition

A particularly useful property of $W_2$ between two univariate distributions is that it can be decomposed in terms of the differences between the two distributions' location, spread and shape. Irpino and Romano (2007) show that:

$$W_2^2(\mu,\nu) = \overbrace{(\bar{\mu}-\bar{\nu})^2}^{Location} + \overbrace{(\sigma_\mu - \sigma_\nu)^2}^{Spread} + \overbrace{2\sigma_\mu\sigma_\nu(1-\rho^{\mu\nu})}^{Shape}, \tag{3}$$

where $\bar{\mu}$ is the mean of $\mu$, $\sigma_\mu$ is the standard deviation of $\mu$ and $\rho^{\mu\nu}$ is the Pearson correlation coefficient between the quantiles of the distributions $\mu$ and $\nu$. These three terms also accord with, qualitatively, how as geologists we might compare two distributions.

### 2.3 Discrete data

Most distributional data in the Earth sciences do not, in raw form, follow continuous probability distributions. Instead, samples may be discrete sets of observations, e.g., lists of individual mineral ages. The above formulations can be easily applied to such cases by describing the probability functions $\mu$ and $\nu$ as weighted sums of $\delta$ functions. For example, let us consider two samples $x_m$ and $x_n$ with $p$ and $q$ numbers of observations respectively:

$$\mu = \sum_i^p m_i \delta_{x_m}, \quad \nu = \sum_i^q n_i \delta_{x_n} \tag{4}$$

where $m$ and $n$ are weight vectors, such that ~~$\sum m = \sum n = 1$~~$\sum m_i = \sum n_i = 1$. In most geological cases these weights would be uniform, $m_i = 1/p$; $n_i = 1/q$, giving each observation within a sample equal weight. In this scenario, $M$ and $N$ are the familiar empirical cumulative distribution functions (ECDF), given as a series of step functions (e.g., Figure 1b).

## 3 ~~Synthetic data~~

### 2.1 A synthetic example

To demonstrate the intuition of $W_2$ we explore a simple synthetic example. We consider two probability density functions of mineral ages: a bimodal distribution and a unimodal distribution, both constructed from Gaussians with the same scale (Figure 2a). We fix the bimodal distribution at 1000 Ma, but translate the unimodal distribution along the time axis. For each translated distribution we calculate both the KS-distance and $W_2$. Figure 2b displays the behaviour of both distances under this scenario. The KS-distance shows an unexpectedly complex response containing a series of steps, as the peaks of the distributions align and misalign. At around $\pm$400 Ma, once the distributions stop overlapping, the KS-distance plateaus at its maximum value of 1. By contrast, $W_2$ increases monotonically with increasing distance. Away from the origin, $W_2$ approximates a linear function of the amount of translation, as is predicted from Equation 3. At the origin, the non-zero value of $W_2$ is the cost of turning the unimodal distribution into the bimodal distribution without translation.

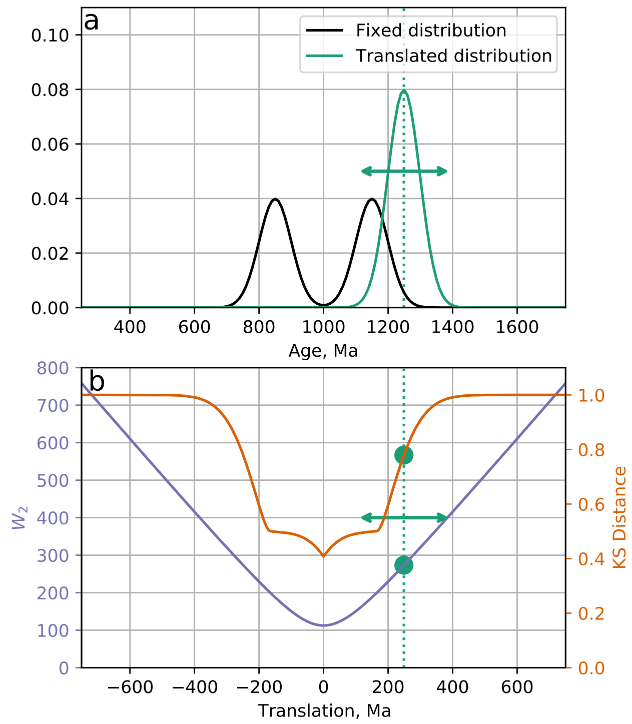


**Figure 2. Comparing the Wasserstein distance to the Kolmogorov-Smirnov distance.** a) Two synthetic probability density functions, modelled on U-Pb age spectra. The black bimodal distribution is fixed at 1000 Ma, and the green unimodal distribution is translated along the time axis. b) For each translated distribution, we calculate the KS-distance (red line) and $W_2$ (blue line). The green dashed line and circles indicate values associated with the location of the green distribution shown in panel a.

We argue that the behaviour of $W_2$ is more geologically intuitive than the KS-distance under this scenario. It is useful geological information if two distributions differ in their means by 400, 500 or 1000 Ma, but if the distributions do not overlap, the KS-distance is insensitive to this. The Wasserstein distance is, by contrast, sensitive to the absolute offset between non-overlapping distributions. Additionally, the stepped response of the KS-distance under translation is undesirable. Under the simple operation of translating a unimodal distribution, we would expect our dissimilarity to increase at a constant, or at least predictable (e.g., quadratic) rate. The change of the KS-distance with translation is, unintuitively, non-linear. By contrast, the $W_2$ increases linearly with respect to translation.

## 3 ~~Use of Wasserstein distance in multi-dimensional scaling~~

~~**Analysing detrital zircon U-Pb ages using the Wasserstein distance.** a–h) Kernel Density Estimates (KDEs) of zircon U-Pb ages from modern sand gathered in Scandinavian rivers by Morton et al. (2008). Sampled river names are indicated in the upper left corners of the plots. 'Ranealven' and 'Ljusnan' samples are filled in and highlighted in panels i–k. KDEs generated using a Gaussian kernel with adaptive bandwidth (Shimazaki and Shinomoto, 2010). i) Empirical Cumulative Distribution Functions (ECDFs) for each zircon age population. j) Multi-Dimensional Scaling (MDS) map for zircon populations calculated using $W_2$ as a dissimilarity metric. Note the closeness of 'Ranealven' and 'Ljusnan'. k) MDS map of same samples but using KS-distance. Note the distance separating 'Ranealven' and 'Ljusnan'.~~

~~We now use $W_2$ to analyse a real dataset of zircon U-Pb ages from Scandinavian river sediments gathered by Morton et al. (2008). This dataset contains eight samples displayed as kernel density estimates in Figure **??**a–h and ECDFs in Figure **??**i. The data required to reproduce our results is provided in `.csv` format at the code repository () . The primary geological province in each sample's drainage basin is shown in Table **??**. Here, we use MDS to jointly compare all samples (Vermeesch, 2013). One of the desirable properties of the Wasserstein distance is that it fulfils the metric requirements, just like the KS-distance (Villani, 2003). Therefore,~~

## 3 Discussion

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 $W_2$ ~~dissimilarity measures can be analysed by classical as well as non-metric MDS algorithms (Vermeesch, 2013). The MDS 'maps' calculated by non-metric MDS, using both $W_2$ and the KS-distance, are shown in Figure **??**j–k. We investigate whether the KS map or the~~ are calculated in terms of differences between ECDFs. Due to these similarities in construction, in many cases the results from using the KS and $W_2$ ~~map shows greater geological meaning~~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 5).

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.

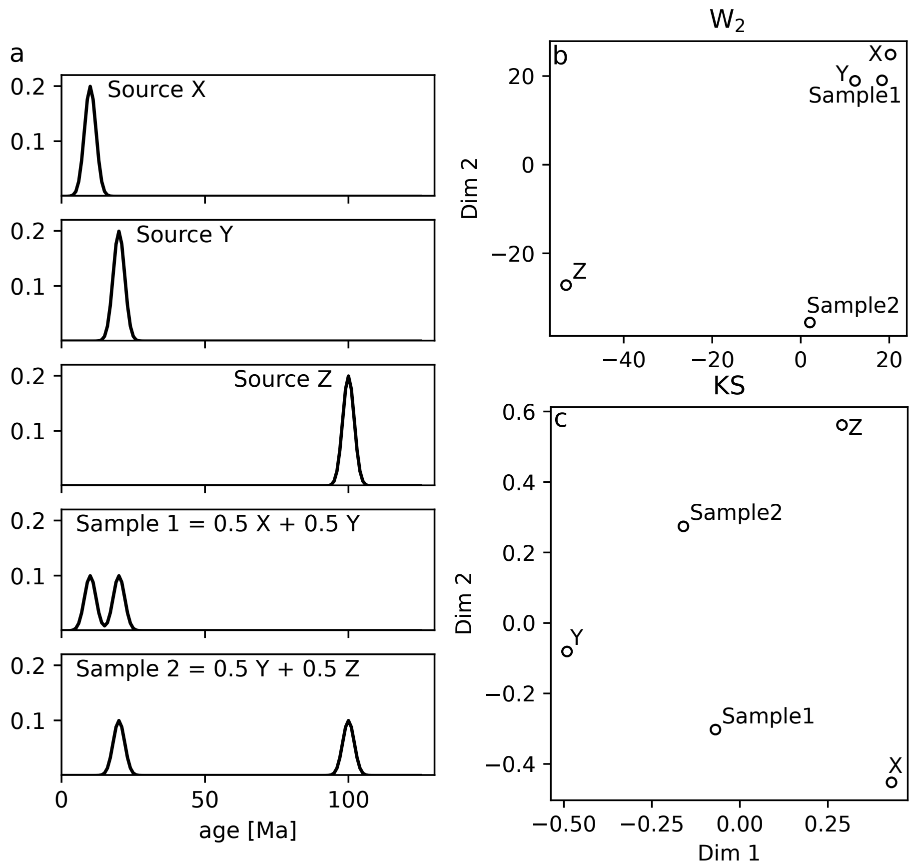


**Figure 3.** ~~The geological provinces drained by each of the rivers sampled in Morton et al. (2008)~~**Mixing of discrete endmembers** a) Three theoretical, ~~reproduced from Table~~ unimodal source age distributions with peaks at 10, 20 and 100 Ma, and two mixture samples. Sample 1 is an equal mixture of X and Y and Sample 2 a mixture of Y and Z. b) Metric MDS map of the ~~original study~~three sources and the mixtures using $W_2$ distance. c) Same as panel b for KS distance. This is a scenario where KS distance may be preferable to $W_2$.

~~**Sample Geological Province** Byskealven Fennoscandian Shield Ranealven Fennoscandian Shield Lainioalven Archaean Ljusnan Trans Scandinavian Igneous Belt Salteva Norwegian Caledonides Vefsna Norwegian Caledonides Vindelalven Swedish Caledonides Ljungan Swedish Caledonides~~

### 3.1 Discriminating contributions from discrete endmembers

We ~~initially focus on two samples: Ranealven (Figure ??a ) and Ljusnan (Figure ??d). These two samples have very similar distribution shapes with one prominent peak, and a smaller younger peak. However, the prominent peak is slightly offset between the two distributions such that there is little overlap. This feature is well shown in the ECDF plot in Figure ??h. As there is limited overlap between them, the KS-distance between these two visually similar distributions approaches its maximum value of 1. As aresult, when MDS is applied to the dataset using the KS-distance these two samples are, counter-intuitively, widely separated (Figure ??k). By contrast, when~~ 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 3a. 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 3a). 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 3b. 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 and Johnstone 2017; Dietze and Dietze 2019).

### 3.2 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 $W_2$~~, the samples are close together (Figure ??j)~~. 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 4 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 4a. 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 4b–c display MDS maps calculated using $W_2$ and KS respectively. The $W_2$ ~~MDS projection also accords well with the actual geological provenance of these samples (Table ??), with samples of the same provenance being grouped together. Whilst the axes of an MDS plot hold no inherent meaning, we can interpret relative positions on the map in terms of distributions' shapes and average ages. The horizontal axis, in this case, appears approximately coincident with the average age of the samples, with the samples to~~ map clearly identifies the stratigraphic order of the samples by the changing distribution

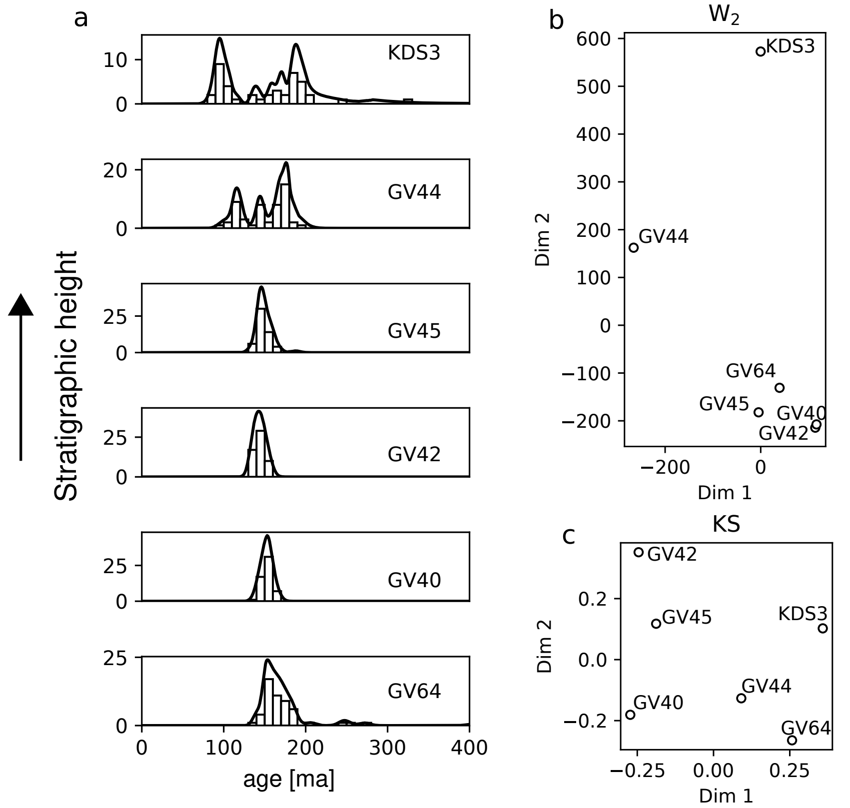


**Figure 4. Temporally evolving source distributions.** a) KDEs and histograms for zircon age distributions for samples from Cache Creek section across Great Valley Group, arranged in stratigraphic order (DeGraaff-Surpless et al., 2002). b) MDS map using $W_2$ for data shown in panel a. c) Same as b using KS distance. In this scenario, the results from $W_2$ are preferable.

shape. Additionally, it clusters the four unimodal samples together. By contrast, the ~~left being generally older than those on the right~~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.

### 3.3 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 5a 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 5b–c respectively. As there is limited overlap between the samples, the ~~peak of Ljusnan is younger than that of Ranealven. In addition, the sample containing the most recent grains, Vefsna, is the furthest to the right. Contrastingly, Lainioalven, which uniquely drains Archean rocks, is the furthest to the left. Similarly, the vertical axis correlates approximately with distribution shape. Salteva & Vefsna have a broad, multimodal distributionand are placed towards the bottom of the map. Conversely, Ranealven & Ljusnan are largely unimodal. Byskealven & Vindelalven lie between these two endmembers and this is reflected in~~ 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 ~~MDS map. Given that~~ $W_2$ ~~can be deconvolved into interpretable statistics (Equation 3) it is not surprising that the MDS maps produced can also be discussed in these terms~~map correctly identifies the simple temporal and geographical trend of the samples from south to north.

### 3.4 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 6 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.

## 4 Implementation

We provide example code (`github.com/AlexLipp/detrital-wasserstein/`) in both python and R that ~~calculates~~ demonstrates how to calculate the $W_2$ between two univariate distributions (U-Pb zircon ages)~~using the analytical expression above (Equation 1 ). In~~. For these examples we make use of the the POT and transport packages in `python` and `R` respectively which implement solutions to Equation 1 (Flamary et al., 2021; Schuhmacher et al., 2022).

### 4.1 IsoplotR

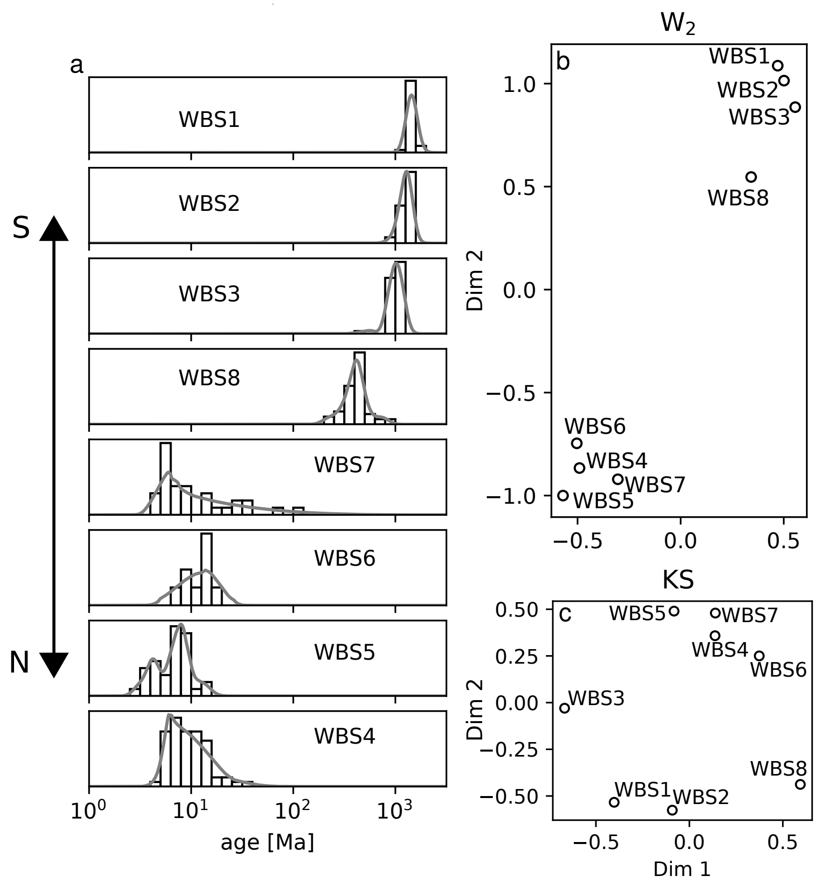


**Figure 5.** **Analysing thermochronological data using $W_2$ and KS distances.** a)KDEs for a detrital mica $^{40}Ar/^{39}Ar$ dataset of Wobus et al. (2003) arranged from south to south across a physiographic transition of the central Himalaya in Nepal. Note the logarithmic scale. b) The MDS configuration 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.

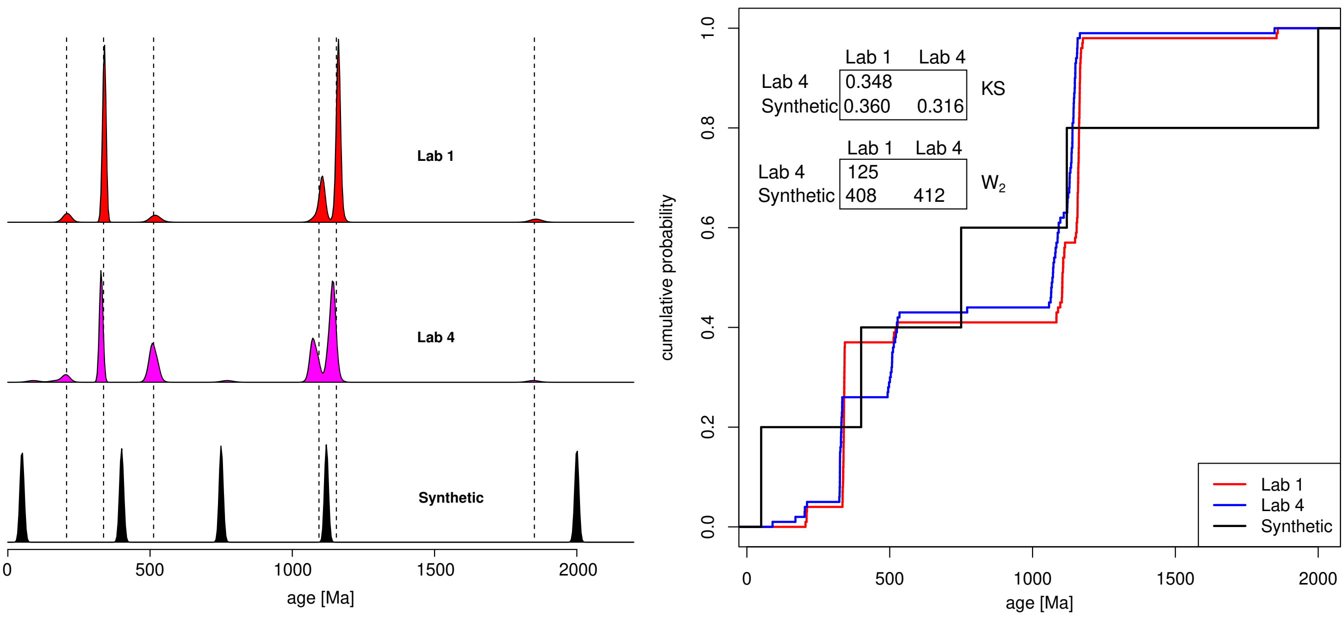


**Figure 6.** **Comparing samples from an inter-laboratory calibration study**. KDEs (left) and ECDFs (right) of two samples from the inter-laboratory comparison study of (Košler et al., 2013), plus a purposefully misaligned synthetic sample. Dashed lines mark the true ages of the detrital mixture. According 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.

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 ~~either~~ be accessed using an (online) graphical user interface, at [1] `isoplotr.es.ucl.ac.uk`. Alternatively, the function can also be accessed from the ~~command line~~[1]~~:~~

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

[1] ~~Note to reviewers: This is a temporary URL pointing to the beta version of the software. This will be replaced with a link to the public `IsoplotR` mirror once the review process has been completed.~~

[1] ~~Note to reviewers: To install the beta version of the `IsoplotR` package, enter `remotes::install_github("pvermees/IsoplotR")` in the R console~~

~~In python, we make use of the POTpackage to calculate $W_2$ (Flamary et al., 2021)~~R command line. The following snippet ~~calculates~~ uses $W_2$ ~~between the Byskealven and Vefsna age distribution from the example above~~to calculate an MDS map for the dataset from Wobus et al. (2003) discussed in the manuscript (Figure 5). The data required ~~, and a python script, is provided at :~~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 in the packages
import numpy as np
import ot
# Load data
vefsna = np.loadtxt("vefsna.csv",delimiter=",",skiprows=1)
byskealven = np.loadtxt("byskealven.csv",delimiter=",",skiprows=1)
# Calculate W_p^p between vefsna and byskealven samples, for p=2
W2_2 = ot.wasserstein_1d(vefsna,byskealven,p=2)
# Calculate W2 using square root
W2 = np.sqrt(W2_2)
```

```
# 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")
```

~~The above code returns a $W_2$ of 490.01.~~

## 5 Conclusions

The second Wasserstein distance, $W_2$, is an effective metric for comparing distributional data in the geological sciences such as detrital age spectra or grain size. ~~The metric is particularly useful for univariate data, but~~ Unlike the KS distance, $W_2$ can be extended to further dimensions. $W_2$ is a function of the horizontal distances between observations, in contrast to the KS distance, which corresponds to vertical differences between ECDFs. ~~Consequently, unlike the KS-distance, $W_2$ is sensitive to variability in the tails of distributions, not just the modes. Under synthetic tests we find that~~ Using a variety of case studies we explore scenarios where the $W_2$ ~~metric behaves more intuitively in comparing distributions relative to the KS-distance. We performed a case study in which eight zircon U-Pb age distributions from Scandinavian river sediments were analysed by MDS using~~ may or may not be preferable to the KS distance. In scenarios where discrete, known age peaks are mixed, the KS distance may be preferable. However, in other scenarios where absolute differences along the time axis are useful information, $W_2$ ~~. We showed that the resulting MDS map accurately clusters samples with the same provenance together. Additionally, the relative positions of samples on the map coincide with trends in interpretable qualities such as distribution shape and average age. The univariate $W_2$ distance has an analytical solution, which we provide implementations of in `R` and~~ is preferable. Example scenarios include spatially/temporally evolving source distributions, thermochronological data, and combining detrital samples from different laboratories. The Wasserstein distance has been added to the `IsoplotR` software, and example scripts are provided in `python` ~~for detrital geochronology and other Earth science applications~~and `R`.

*Code availability.* The code and data repository is found at `https://github.com/AlexLipp/detrital-wasserstein`

*Author contributions.* AGL conceived the project, both authors contributed to development, writing, and software production.

*Competing interests.* PV is an Associate Editor of Geochronology

*Acknowledgements.* AGL is funded by a Junior Research Fellowship from Merton College, Oxford. PV is supported by NERC Standard Grant #NE/T001518/1. This work benefited from discussions with Malcolm Sambridge & Kerry Gallagher. We thank two anonymous reviewers, and the associate editor Michael Dietze for their constructive feedback.

## References


Amidon, W. H., Burbank, D. W., and Gehrels, G. E.: Construction of detrital mineral populations: insights from mixing of U–Pb zircon ages in Himalayan rivers, Basin Research, 17, 463–485, https://doi.org/10.1111/j.1365-2117.2005.00279.x, 2005.

Benamou, J.-D., Carlier, G., Cuturi, M., Nenna, L., and Peyré, G.: Iterative Bregman Projections for Regularized Transportation Problems, SIAM Journal on Scientific Computing, 2, A1111–A1138, https://doi.org/10.1137/141000439, publisher: Society for Industrial and Applied Mathematics, 2015.

Berry, R. F., Jenner, G. A., Meffre, S., and Tubrett, M. N.: A North American provenance for Neoproterozoic to Cambrian sandstones in Tasmania?, Earth and Planetary Science Letters, 192, 207–222, https://doi.org/10.1016/S0012-821X(01)00436-8, 2001.

Cawood, P., Hawkesworth, C., and Dhuime, B.: Detrital zircon record and tectonic setting, Geology, 40, 875–878, https://doi.org/10.1130/G32945.1, 2012.

Condie, K. C., Belousova, E., Griffin, W. L., and Sircombe, K. N.: Granitoid events in space and time: Constraints from igneous and detrital zircon age spectra, Gondwana Research, 15, 228–242, https://doi.org/10.1016/j.gr.2008.06.001, 2009.

De Doncker, F., Herman, F., and Fox, M.: Inversion of provenance data and sediment load into spatially varying erosion rates, Earth Surface Processes and Landforms, 45, 3879–3901, https://doi.org/https://doi.org/10.1002/esp.5008, 2020.

DeGraaff-Surpless, K., Graham, S. A., Wooden, J. L., and McWilliams, M. O.: Detrital zircon provenance analysis of the Great Valley Group, California: Evolution of an arc-forearc system, GSA Bulletin, 114, 1564–1580, https://doi.org/10.1130/0016-7606(2002)114<1564:DZPAOT>2.0.CO;2, 2002.

Dietze, E. and Dietze, M.: Grain-size distribution unmixing using the R package EMMAgeo, E&G Quaternary Science Journal, 68, 29–46, https://doi.org/10.5194/egqsj-68-29-2019, 2019.

Engquist, B. and Froese, B. D.: Application of the Wasserstein metric to seismic signals, Communications in Mathematical Sciences, 12, 979–988, https://doi.org/10.4310/CMS.2014.v12.n5.a7, 2014.

Flamary, R., Courty, N., Gramfort, A., Alaya, M. Z., Boisbunon, A., Chambon, S., Chapel, L., Corenflos, A., Fatras, K., Fournier, N., Gautheron, L., Gayraud, N. T. H., Janati, H., Rakotomamonjy, A., Redko, I., Rolet, A., Schutz, A., Seguy, V., Sutherland, D. J., Tavenard, R., Tong, A., and Vayer, T.: POT: Python Optimal Transport, Journal of Machine Learning Research, 22, 1–8, 2021.

Irpino, A. and Romano, E.: Optimal histogram representation of large data sets: Fisher vs piecewise linear approximation, in: Actes des cinquièmes journées Extraction et Gestion des Connaissances, vol. E-9, pp. 99–110, Namur, Belgium, 2007.

Košler, J., Sláma, J., Belousova, E., Corfu, F., Gehrels, G. E., Gerdes, A., Horstwood, M. S. A., Sircombe, K. N., Sylvester, P. J., Tiepolo, M., Whitehouse, M. J., and Woodhead, J. D.: U-Pb Detrital Zircon Analysis – Results of an Inter-laboratory Comparison, Geostandards and Geoanalytical Research, 37, 243–259, https://doi.org/10.1111/j.1751-908X.2013.00245.x, 2013.

Magyar, J. C. and Sambridge, M.: Hydrological objective functions and ensemble averaging with the Wasserstein distance, Hydrology and Earth System Sciences, 27, 991–1010, https://doi.org/10.5194/hess-27-991-2023, 2023.

Morton, A., Fanning, M., and Milner, P.: Provenance characteristics of Scandinavian basement terrains: Constraints from detrital zircon ages in modern river sediments, Sedimentary Geology, 210, 61–85, https://doi.org/10.1016/j.sedgeo.2008.07.001, 2008.

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J.: An optimal transport approach for seismic tomography: application to 3D full waveform inversion, Inverse Problems, 32, 115 008, https://doi.org/10.1088/0266-5611/32/11/115008, 2016.

Peyré, G. and Cuturi, M.: Computational Optimal Transport, Foundations and Trends in Machine Learning, 11, 355–607, 2019.

Reimink, J. R., Davies, J. H. F. L., and Ielpi, A.: Global zircon analysis records a gradual rise of continental crust throughout the Neoarchean, Earth and Planetary Science Letters, 554, 116 654, https://doi.org/10.1016/j.epsl.2020.116654, 2021.

Sambridge, M., Jackson, A., and Valentine, A. P.: Geophysical inversion and optimal transport, Geophysical Journal International, 231, 172–198, https://doi.org/10.1093/gji/ggac151, 2022.

Satkoski, A. M., Wilkinson, B. H., Hietpas, J., and Samson, S. D.: Likeness among detrital zircon populations—An approach to the comparison of age frequency data in time and space, GSA Bulletin, 125, 1783–1799, https://doi.org/10.1130/B30888.1, 2013.

Saylor, J., Stockli, D., Horton, B., Nie, J., and Mora, A.: Discriminating rapid exhumation from syndepositional volcanism using detrital zircon double dating: Implications for the tectonic history of the Eastern Cordillera, Colombia, Bulletin of the Geological Society of America, 124, 762–779, https://doi.org/10.1130/B30534.1, 2012.

Schuhmacher, D., diagrams), B. B. a. a. p., Bonneel (networkflow), N., shortlist), C. G. s., , Hartmann (semidiscrete1), V., integration), F. H. t. a. n., Schmitzer (shielding), B., Schrieber (subsampling), J., and Wilm (wpp), T.: transport: Computation of Optimal Transport Plans and Wasserstein Distances, `https://CRAN.R-project.org/package=transport`, 2022.

Sharman, G. R. and Johnstone, S. A.: Sediment unmixing using detrital geochronology, Earth and Planetary Science Letters, 477, 183–194, https://doi.org/10.1016/j.epsl.2017.07.044, 2017.

Sharman, G. R., Sharman, J. P., and Sylvester, Z.: detritalPy: A Python-based toolset for visualizing and analysing detrital geothermochronologic data, The Depositional Record, 4, 202–215, https://doi.org/10.1002/dep2.45, 2018.

Shimazaki, H. and Shinomoto, S.: Kernel bandwidth optimization in spike rate estimation, Journal of Computational Neuroscience, 29, 171–182, https://doi.org/10.1007/s10827-009-0180-4, 2010.

Sundell, K. E. and Saylor, J. E.: Two-Dimensional Quantitative Comparison of Density Distributions in Detrital Geochronology and Geochemistry, Geochemistry, Geophysics, Geosystems, 22, e2020GC009 559, https://doi.org/10.1029/2020GC009559, 2021.

Vermeesch, P.: Multi-sample comparison of detrital age distributions, Chemical Geology, 341, 140–146, https://doi.org/10.1016/j.chemgeo.2013.01.010, 2013.

Vermeesch, P.: Dissimilarity measures in detrital geochronology, Earth-Science Reviews, 178, 310–321, https://doi.org/10.1016/j.earscirev.2017.11.027, 2018a.

Vermeesch, P.: IsoplotR: A free and open toolbox for geochronology, Geoscience Frontiers, 9, 1479–1493, https://doi.org/10.1016/j.gsf.2018.04.001, 2018b.

Vermeesch, P., Lipp, A. G., Hatzenbühler, D., Caracciolo, L., and Chew, D.: Multidimensional scaling of varietal data in sedimentary provenance analysis, Journal of Geophysical Research: Earth Surface, p. e2022JF006992, https://doi.org/10.1029/2022JF006992, 2023.

Villani, C.: Topics in Optimal Transportation, American Mathematical Soc., 2003.

Weltje, G. J.: End-member modeling of compositional data: Numerical-statistical algorithms for solving the explicit mixing problem, Mathematical Geology, 29, 503–549, https://doi.org/10.1007/BF02775085, 1997.

Wobus, C. W., Hodges, K. V., and Whipple, K. X.: Has focused denudation sustained active thrusting at the Himalayan topographic front?, Geology, 31, 861–864, https://doi.org/10.1130/G19730.1, 2003.