Tracing Sedimentary Origins in Multivariate Geochronology via Constrained Tensor Factorization

Naomi Graham[1], Nicholas Richardson[2], Michael P. Friedlander[3], and Joel Saylor[4]

[1]*Department of Computer Science, University of British Columbia,* naomigra@cs.ubc.ca
[2]*Department of Mathematics, University of British Columbia,* njericha@math.ubc.ca
[3]*Department of Computer Science, Department of Mathematics, University of British Columbia,* michael.friedlander@ubc.ca *(corresponding author)*
[4]*Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia,* jsaylor@eoas.ubc.ca

October 12, 2024

Abstract

We devise a novel statistical method for deconvolving multivariate geochronology and geochemistry datasets to characterize sediment sources. The approach employs a third-order constrained Tucker-1 tensor decomposition to estimate the probability distributions of multiple features in sediment samples. By integrating kernel density estimation with matrix-tensor factorization, the model quantitatively determines the distributions and mixing proportions of sediment sources. The methodology introduces a numerical test for rank estimation to define the number of latent sources, and uses a maximum-likelihood approach to correlate individual detrital grains to these latent sources based on an arbitrary number of features. The method's efficacy is validated through a numerical experiment with detrital zircon data that captures natural variability associated with temporal changes in crustal thickness in the Andes. The findings hold potential implications for resolving sediment sources, determining sediment mixing, enhancing the understanding of sediment transport processes, characterizing the lithology, tectonic motion, or metallogenic potential of sediment sources. This method is adaptable to other data de-mixing problems and is available in a publicly released software package.

Contents

1 Introduction

Eroded clastic material, once deposited in a sedimentary sink, preserves a complex record of its source and the various processes it underwent during transport, deposition, and diagenesis [10, 19, 60]. In this context, a *sink* refers to any sample that can be considered a mixture of detritus derived from multiple sediment *sources* (Fig. 1.1). Detrital zircon grains, in particular, serve as discrete carriers of information about their respective sources. They capture a record that includes the eroded material's mineralogical and geochemical characteristics, as well as changes resulting from selective deposition, weathering, or mixing. A statistical model can provide insight into the sediment mixture process and help to trace sediment provenance. The accuracy with which the actual proportions of the various contributing sources can be estimated necessarily degrades as the material undergoes changes during transport [14, 20]. For this reason, the refractory nature of detrital zircons makes uranium-lead (U-Pb) geochronology an especially powerful and versatile tool, widely used to study the distribution of sediments, the development of sedimentary basins, tectonic plate motion, and chronostratigraphy [9, 16, 23, 27, 41, 46, 47].

While zircons have the potential to identify sources based on a single feature such as U-Pb age, this method is limited when sources have similar crystallization ages [15, 45, 48]. Various methods have been developed to discriminate sources that have similar detrital zircon U-Pb spectra [15, 38]. These methods often include additional features such as thermochronology, secondary isotopic analyses (e.g., Lu-Hf), or trace and rare earth element (TREE) analyses to discriminate between otherwise similar sources [7, 17, 31, 40, 45].

Beyond source discrimination, trace element analysis of zircons yields valuable petrogenetic information about the magmas from which the zircons originated. This information is crucial for tracking crustal evolution and identifying ore-bearing lithologies [25]. Multiple features can be used to trace the petrogenetic characteristics of source magmas, including, for example, titanium concentrations to track crystallization temperatures [55, 56]; europium anomalies to trace plagioclase crystalization; strontium-yttrium (Sr/Y) ratios to infer crustal thickness [12, 51, 52]; and a suite of 26 trace element concentrations to determine the parent rock type [4, 51].

The low diffusion rates of trace elements such as U, hafnium (Hf), and thorium (Th) in zircon, combined with its physical and chemical resilience, make these trace elements reliable fingerprints of sediment provenance. Our ability to acquire these multivariate datasets, however, has exceeded the capability of the available analytical tools that can fully exploit and quantitatively model zircon geochronology and geochemistry data. For example, methods such as forward Monte Carlo mixture modelling and nonnegative matrix factorization [40, 48] are constrained to one or two features and therefore limited in their effectiveness in sediment provenance or crustal evolution analyses, or economic geology applications.

1.1 Contributions

This paper describes a statistical approach to deconvolving detrital geochronology and geochemistry datasets into their constituent sources using an arbitrary number of measured features. Our approach is based on a statistical model (Sec. 2) representing sediment sinks as product probability distributions on their features. The decomposition approach estimates these probability distributions through the mixed sedimentary data. Section 1.2 describes the geology of this model.

We design a computational workflow that begins by using a density estimation technique to quantize the data into a three-dimensional tensor with nonnegative entries, derived from samples of the estimated densities (Sec. 3). This tensor is then decomposed using a model-fitting algorithm (Sec. 4), which factorizes it into a tensor predicting the distributions of materials contributed by each latent source, and a matrix providing their relative proportions. Because the number of latent sources is unknown, we have developed a statistical test (Sec. 4.5) to estimate the optimal rank for the decomposition. Additionally, we introduce a maximum-likelihood method for correlating individual detrital grains to latent sources based on their multivariate features (Sec. 5). This computational toolchain (Sec. 6) is implemented in a fully reproducible software package featuring efficient algorithms whose computational costs scale linearly with the number of features.

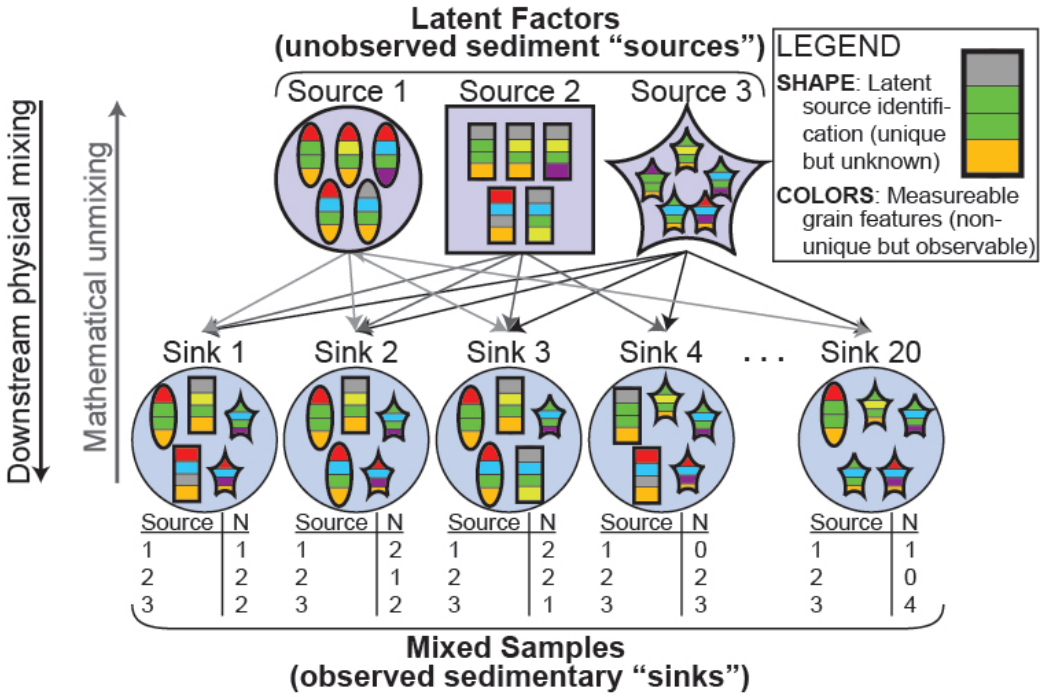

Figure 1.1: Schematic of the geological model depicting the physical downstream mixing from multiple sources (3 in total) into multiple sinks (20 in total). The model measures multiple features (represented by different colors) of the sedimentary grains (represented by different shapes). The aim is to extract three key elements from the sink data: (1) the distribution of source features, (2) the mixing proportions (shown in the table below each sink), and (3) the source label (oval, rectangle, or star) for each grain.

We demonstrate in Sec. 7 the accuracy of this method by applying it to the Sundell et al. [51] data set of grain samples with known sinks and sources. To evaluate our approach, we apply it to data from the twenty sinks, and correctly determine the unknown source distributions and their mixing proportions to within 10 and 5 percent, respectively, of their known counterparts. Furthermore, correlation of individual detrital grains to their factorized sources based on their multivariate features is approximately 90% accurate, which is significantly more accurate than a similar approach applied to univariate or bivariate data sets.

We provide in Table 1 a glossary of the mathematical symbols used throughout the paper. Section 8 concludes with a discussion of possible refinements and extensions of our method.

1.2 Geological framework and prior approaches

Most prior work has focused on quantifying the similarity between observed samples based on one or two measured features rather than on sediment mixing or sediment source characterization. However, quantitative methods like principal component analysis [44], Bhattacharya distance [6, 18], "mismatch" or "likeness" [2, 34], cross-correlation [37], or Kolmogorov-Smirnov or Kuiper distances [13, 39], often fail to discriminate between sediment sources with similar feature distributions. Even bivariate approaches [40, 47], though they provide some improvement, are still limited to two features, reducing their effectiveness for analyzing complex, multivariate detrital datasets such as those from petrochronological or triple-dating studies [8, 11]. Other multivariate techniques, such as 3-way multidimensional scaling (MDS) or generalized Procrustes analysis (GPA), help qualitatively visualize sample similarities [54]. Our approach advances the available quantitative tools for identifying potential sediment sources, which in combination with recent semi-qualitative analyses such as [53],

facilitates a more comprehensive analysis.

Studies that have addressed sediment mixing or characterization of unknown sediment sources have also done so on the basis of just one or two features. The available methods include deterministic or Monte Carlo forward mixing [2, 26, 49, 59], and inversion via nonnegative matrix factorization [24, 38, 40, 42, 50, 58], which simplifies the data into a lower-dimensional representation. There are two main objectives of these studies. First, to recover the proportions in which individual sources contribute to each sink sample, and second, to infer the features of the sediment sources when they are unknown or no longer exist. The identified sources must match the features measured in the detrital samples, and the analysis must generate mixing coefficients that accurately represent the linear combinations of these source features.

Our approach significantly diverges from previous methods by incorporating multiple features simultaneously from zircon grains. By considering multiple features simultaneously, our approach yields more a more accurate decomposition into sources, which can enhance the accuracy with which individual grains can be identified with sources. Additionally, the model and the attendant algorithm scale effectively with additional features, enhancing its applicability to complex geological datasets.

2 A statistical mixing model

The main assumption of our model is that the sink data can be represented as a mixture of a small number of distinct sources. By "small", we mean a number that is much less than the total number of input sinks [36]. The sink data is modeled as a mixture of these sources with unknown mixing proportions, and the goal is to estimate both the source distributions and these proportions.

Given these assumptions, the input data sets must be restricted to features that can be represented as distributions, following the logic outlined by Vermeesch et al. [53, Section 2]. Additionally, features should be derived from transport invariant subpopulations [59, 60]. Features that undergo post-depositional modification, such as the chemical index of alteration [29] based on bulk geochemistry, should not be included because their proportions may change during diagenesis, making them inconsistent with the mixing model's assumptions. Similarly, features significantly altered during transport should also be excluded from the data set. For instance, bulk clay mineralogy should not be modeled alongside zircon geochemistry data, as they may be subject to different processes.

We formalize this model as follows. Let β_{rj} represent the distribution of the quantity of feature j within source r, and let B_r denote the overall feature distribution for source r. For now, we consider β_{rj} and B_r as abstract distributions. In Sec. 3.2, as part of the tensor decomposition, we associate these with discrete counterparts. We define B_r as the product of the individual feature distributions, that is,

$$B_r = \bigtimes_{j\in[J]} \beta_{rj} \quad \text{for each source} \quad r \in [R]. \tag{2.1}$$

Our second key assumption is that the distribution of features within each sink is a mixture of these source distributions. Let S_i represent the distribution of features within sink i. Each sink distribution S_i is therefore a convex combination of the R source distributions:

$$S_i = \sum_{r\in[R]} \alpha_{ir} B_r, \quad \sum_{r\in[R]} \alpha_{ir} = 1,\ \alpha_{ir} \geq 0 \quad \text{for each sink} \quad i \in [I], \tag{2.2}$$

where the scalars α_{ir} represent the (unknown) convex weights of contributions for source r in sink i. Equations (2.1) and (2.2) together imply that the sink distributions can also be decomposed into product distributions:

$$S_i = \sum_{r\in[R]} \bigtimes_{j\in[J]} \alpha_{ir}\beta_{rj} \quad \text{for each sink} \quad i \in [I]. \tag{2.3}$$

In summary, the unknown sources $B_1, \ldots, B_R$, which we refer to as *latent sources*, each consist of different continuous product distributions of the J features. We assume that individual zircon grains collected from sink i originate from a single source r according to the proportion α_{ir}, and that the grains are indivisible. The overarching goal is to estimate the source distributions β_{rj} and the mixing proportions α_{ir} from the sink data.

3 Sink distribution estimation

To estimate the sink distributions S_i from unprocessed zircon grain samples, we employ a two-step method explained in the sections that follow. First, we apply kernel density estimation (KDE) [43] to each feature across all sinks, which turns the raw data into continuous probability density functions that mirror the underlying distributional characteristics of the features. Next, we discretize these continuous density estimates into a new batch of samples, which are gathered into a tensor structure covering all sink-feature pairs. This approach of using samples from continuous density estimates rather than the original raw data improves our ability to compare similar features across various sinks, without being affected by differences in the number of grains each sink might have.

3.1 Feature sampling and continuous density estimation

Let $\{\boldsymbol{g}_i^n[j]\}_{n\in[N_i]}$ denote the N_i observations obtained for feature j from grains sampled from sink i. Thus, $\boldsymbol{g}_i^n$ is a vector of length J where each coordinate corresponds to a measurement from a different feature. Consequently, it follows from the sink product model Eq. (2.2) that each coordinate $\boldsymbol{g}_i^n[j]$ is a random variable distributed according to a mixture of source features j represented by the sink i. In particular, each measurement $\boldsymbol{g}_i^n[j]$ is distributed according to the mixture

$$\boldsymbol{g}_i^n[j] \sim \sum_{r\in[R]} \alpha_{ir}\beta_{rj}.$$

We estimate the continuous density of each feature j for each sink i using a KDE built from the standard Gaussian kernel $\kappa(x) = \exp(-x^2/2)/\sqrt{2\pi}$. The corresponding KDE of the sink-feature pair (i, j) is the smooth univariate density function

$$f_{ij}(x) = \frac{1}{N_i}\sum_{n\in[N_i]} \frac{1}{h_j}\kappa\left(\frac{x - \boldsymbol{g}_i^n[j]}{h_j}\right),$$

where the positive bandwidths scalars h_j for each feature j are chosen to minimize the mean integrated square error of the KDE. (Section 6.2 outlines the methodology used to estimate these bandwidths.) The KDE f_{ij} approximates the probability

$$\mathbb{P}\left(X \in [a, b] \;\middle|\; X \sim \textstyle\sum_{r\in[R]} \alpha_{ir}\beta_{rj}\right) = \displaystyle\int_a^b f_{ij}(x)\,dx.$$

Figure 3.1 shows an example of a KDE produced from age measurements of zircon grains from a single sink and the KDEs for all sinks for the same feature.

3.2 Discretization and tensor construction

We discretize the continuous density estimates on a uniform grid across all sink-feature pairs. Specifically, for each feature j, we sample the KDEs f_{ij} at K points $\{x_{jk}\}_{k\in[K]}$. The grid spacing Δx_j is uniform for each feature but varies between features due to differences in their value ranges. Although a more complex nonuniform spacing is possible, we opt for uniform spacing for simplicity. Details of the discretization, including the number of samples K, bandwidths h_j, grid spacing Δx_j, and the sampling domain of the KDEs, are crucial in practice and are discussed in Sec. 6.

We obtain a discrete approximation at each sample point x_{jk} as

$$f_{ij}(x_{jk})\Delta x_j \approx \mathbb{P}\left(X \in [x_{jk}, x_{jk} + \Delta x_j] \mid X \sim \textstyle\sum_r \alpha_{ir}\beta_{jr},\right),$$

which is based on the assumption that sink distributions are mixtures of the source distributions, as detailed in Eq. (2.3). These resampled points are indexed by a source-feature-sample triple index (i, j, k) and assembled into the elements of a third-order tensor as

$$\mathcal{Y}[i, j, k] = f_{ij}(x_{jk})\Delta x_j \;\text{ for each }\; (i, j, k) \in [I] \times [J] \times [K]. \tag{3.1}$$

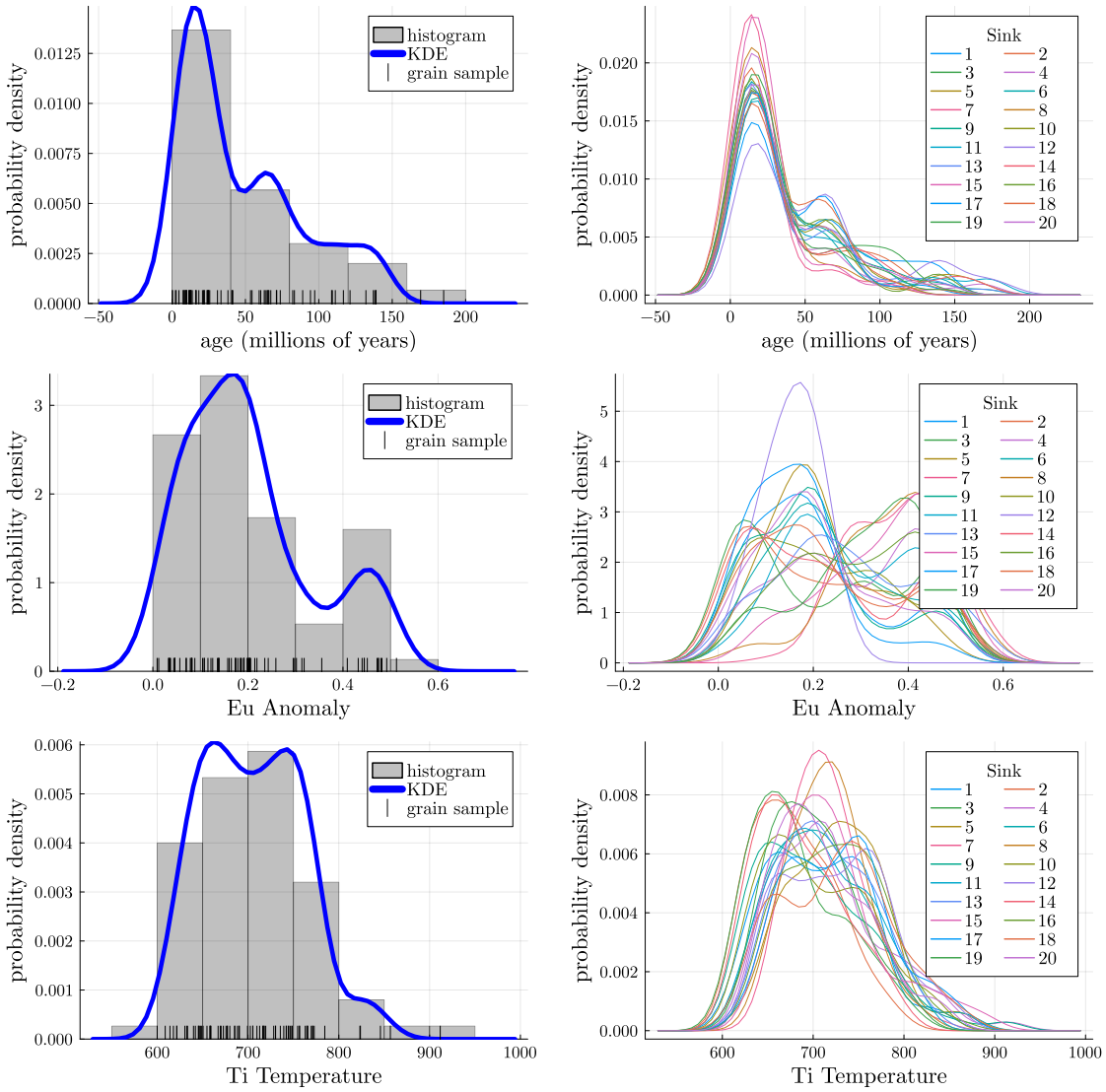

Figure 3.1: (Left) Histogram with overlaid kernel density estimates (KDEs) and raw grain data (rug plot) for three features (Age, Eu Anomaly, Ti Temperature) for a single sink ($i = 1$). (Right) KDEs for these same features across all 20 sinks, illustrating variations in distributions. Although the age and Eu anomaly measurements are necessarily positive, the Gaussian kernel—and thus also the KDE—has unbounded support over real numbers.

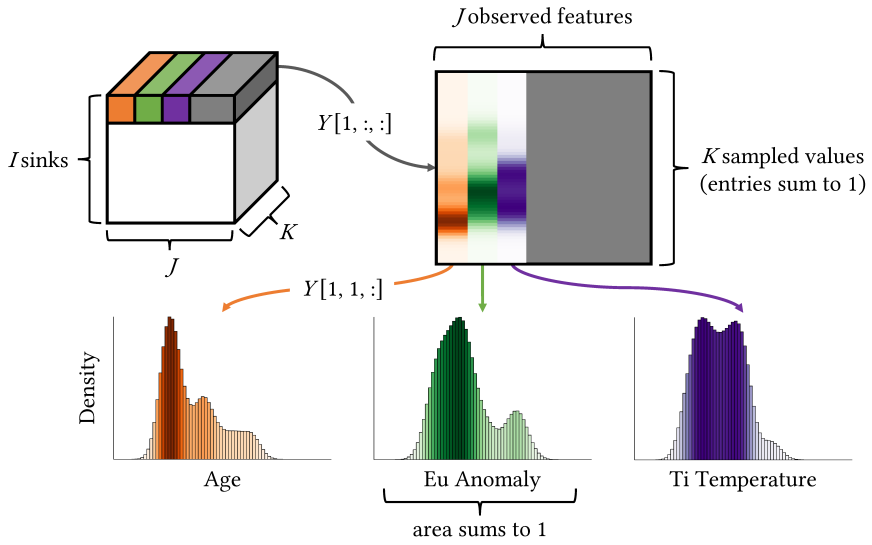

Figure 3.2: Schematic of the KDE discretization and embedding into the input tensor $\mathcal{Y}$. Every horizontal slice of $\mathcal{Y}$, as defined by Eq. (3.1), corresponds to a sink. For each sink there are J feature distributions that are collected into a matrix. The first three features of sink 1 are highlighted here.

Thus, each element of the nonnegative tensor $\mathcal{Y}[i, j, k]$ quantifies the probability that a grain from sink i exhibits feature j within the kth interval $[x_{kj}, x_{kj} + \Delta x_j]$. Figure 3.2 depicts a schematic of this tensor model. The slice $\mathcal{Y}[i, :, :]$ for each sink i is a matrix that collects the discretized distributions for each feature. The fibers of each slice $\mathcal{Y}[i, j, :]$ are the discretized distributions for each feature j in sink i, and $\sum_{k\in[K]} \mathcal{Y}[i, j, k] \approx 1$ for each $i \in [I]$ and $j \in [J]$. The tensor $\mathcal{Y}$ serves as the input for our tensor decomposition algorithm, which is elaborated in Sec. 4.

4 Tensor decomposition over probability simplices

This section outlines the decomposition of the 3-way empirical discrete distribution tensor, $\mathcal{Y}$ (cf. Sec. 3.2), into a smaller latent-distribution tensor and proportion matrix that quantifies the contributions of latent sources to sinks. We use a Tucker decomposition variant, tailored to specific tensor structures, where each mode-3 fiber of the core tensor and each row of the proportion matrix is nonnegative and sum to one, thus fitting within probability simplices. This model aligns with our hypothesis that sinks are mixtures of several latent sources; cf. Eq. (2.2). The decomposition process uses a block coordinate-descent algorithm, as detailed in Sec. 4.4. For a comprehensive introduction to tensor notation and decompositions, consult Kolda and Bader [22].

4.1 Tucker decomposition

The Tucker decomposition factorizes a 3-way real-valued tensor $\mathcal{T}$ (dimensions $t_1 \times t_2 \times t_3$) into a core tensor $\mathcal{B}$ (dimensions $m_1 \times m_2 \times m_3$) and three factor matrices $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, and $\boldsymbol{A}_3$ (with corresponding dimensions $t_1 \times m_1$, $t_2 \times m_2$, and $t_3 \times m_3$, respectively). This decomposition involves n-mode multiplication, defined through the matricization (or unfolding) of the tensors. Specifically, the n-mode matricization, denoted as $\mathcal{B}_{(n)}$ (similarly for $\mathcal{T}$), reorganizes the elements of $\mathcal{B}$ into a $p_1 \times p_2$ matrix, where $p_1 = m_n$ and p_2 is the product of the other dimensions $\prod_{i\neq n} m_i$. The order of unfolding is arbitrary so long as it is consistent throughout all computations.

The n-mode product, defined as $\mathcal{B} \times_{(n)} \boldsymbol{A}_n$, involves matrix multiplication between the matricized version of $\mathcal{B}$ and the matrix $\boldsymbol{A}$, i.e.,

$$\mathcal{R} = \mathcal{B} \times_{(n)} \boldsymbol{A}_n \quad \Longleftrightarrow \quad \mathcal{R}_{(n)} = \boldsymbol{A}_n \mathcal{B}_{(n)}, \tag{4.1}$$

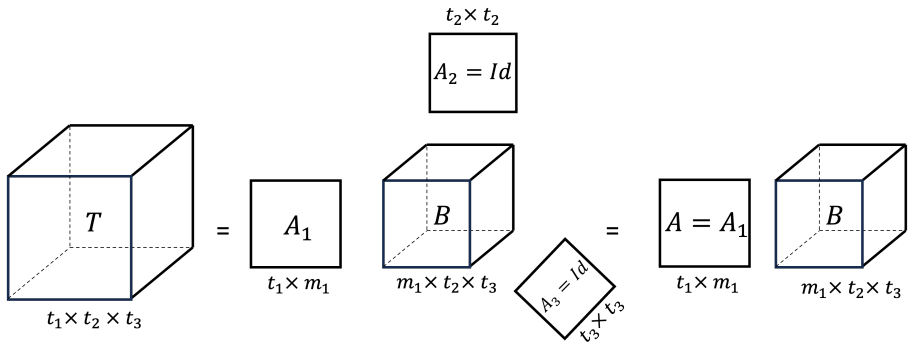

Figure 4.1: An illustration of Tucker-1; see Eq. (4.3). By fixing two of the factors $\boldsymbol{A}_2$ and $\boldsymbol{A}_3$ in the full Tucker decomposition Eq. (4.2), we obtain the Tucker-1 decomposition model.

which defines the transformation process for each mode.

The full Tucker decomposition of $\boldsymbol{\mathcal{T}}$ is then expressed as

$$\boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{B}} \times_1 \boldsymbol{A}_1 \times_2 \boldsymbol{A}_2 \times_3 \boldsymbol{A}_3, \tag{4.2}$$

where $\boldsymbol{\mathcal{B}}$ acts as a compressed representation of $\boldsymbol{\mathcal{T}}$. If the core tensor is chosen small, then the decomposition may only approximate the original tensor, and in that case $\boldsymbol{\mathcal{B}}$ is a lossy compression of $\boldsymbol{\mathcal{T}}$.

4.2 Tucker-1 variant over the probability simplex

Our model employs the Tucker-1 variant, which simplifies the standard Tucker model by setting the second and third factor matrices, $\boldsymbol{A}_2$ and $\boldsymbol{A}_3$ in Eq. (4.2), to identity matrices. For simplicity, we drop the unneeded subscripts on the factor matrix $\boldsymbol{A}$:

$$\boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{B}} \times_1 \boldsymbol{A}.$$

This approach, depicted in Fig. 4.1, focuses compression along a single mode, reducing complexity and improving interpretability, especially when dimensionality reduction is primarily desired in one dimension. We adopt the notation

$$\boldsymbol{\mathcal{T}} = \boldsymbol{A}\boldsymbol{\mathcal{B}}. \tag{4.3}$$

This simplified model aligns with our goal of identifying the latent sources and their mixing proportions. We approximate the empirical density tensor $\boldsymbol{\mathcal{Y}}$ as a product of a nonnegative proportion matrix $\boldsymbol{A}$ and a nonnegative tensor $\boldsymbol{\mathcal{B}}$, with constraints that ensure all elements remain nonnegative and that the distributions sum to one along the designated fibers and rows. These constraints aid in maintaining the probabilistic interpretations of the tensor entries. The computed decomposition of $\boldsymbol{\mathcal{Y}}$ is then

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{A}\boldsymbol{\mathcal{B}} + \boldsymbol{\mathcal{E}},$$

where $\boldsymbol{\mathcal{E}}$ captures noise and other unmodelled effects. It follows from Eq. (4.1) that this decomposition is equivalent simplex-constrained matrix factorization under any flattening of $\boldsymbol{\mathcal{Y}}$ and $\boldsymbol{\mathcal{B}}$. However, the tensor formulation maintains the inherent relationships among the various quantities.

Within the framework of Eq. (2.2), the $I \times R$ proportion matrix $\boldsymbol{A}$ provides mixture coefficients for each latent source, while the mode-3 fibers of the $R \times J \times K$ core tensor $\boldsymbol{\mathcal{B}}$ represent the feature distributions of latent sources. Elementwise, this relationship is modeled as

$$\boldsymbol{\mathcal{Y}}[i,j,k] \approx \sum_{r=1}^{R} \boldsymbol{A}[i,r] \cdot \boldsymbol{\mathcal{B}}[r,j,k]. \tag{4.4}$$

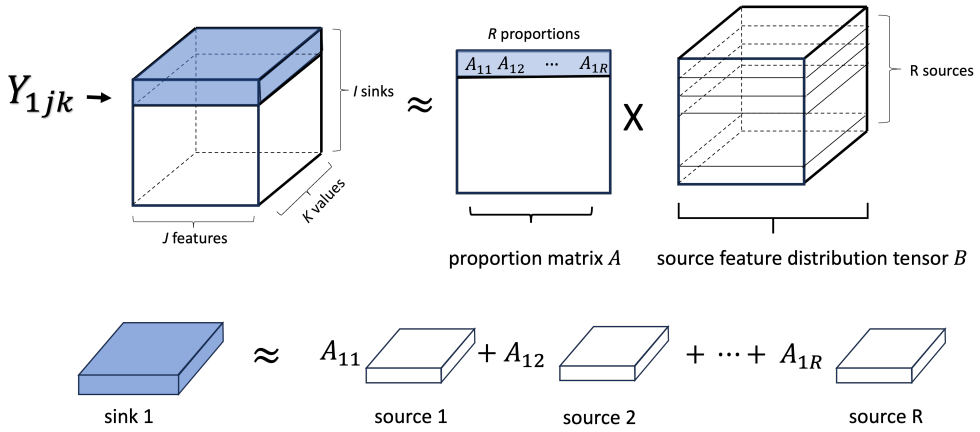

Figure 4.2: The factorization Eq. (4.4) reveals how the sources, given by slices of $\mathcal{B}$ contribute to each sink in different proportions, given by the entries of $\boldsymbol{A}$ in Eq. (4.5).

Each element

$$\boldsymbol{A}[i, r] = \alpha_{ir} \tag{4.5}$$

gives the proportion of each source r in sink i as defined by Eq. (2.2); see also Fig. 4.2. Similarly, we interpret the elements of the latent source distribution tensor $\mathcal{B}$ analogously to Eq. (3.1):

$$\mathcal{B}[r, j, k] = \beta_{rj}(x_{jk})\Delta x_j \approx \mathbb{P}\left(X \in \left[x_{jk}, x_{j(k+1)}\right] \mid X \sim \beta_{rj}\right). \tag{4.6}$$

Here the first dimension of $\mathcal{B}$ is indexed over sources, while the first index in $\mathcal{Y}$ is indexed over sinks, thus ensuring a consistent interpretation across both tensors.

4.3 The constrained matrix-tensor factorization model

We employ a block coordinate descent algorithm to compute the approximation Eq. (4.4) as the solution of the constrained least-squares problem

$$\min_{\boldsymbol{A},\mathcal{B}} \left\{ \ell(\boldsymbol{A}, \mathcal{B}) := \tfrac{1}{2} \left\| \mathcal{Y} - \boldsymbol{A}\mathcal{B} \right\|_F^2 \,\middle|\, \boldsymbol{A} \in \Delta_{\boldsymbol{A}},\ \mathcal{B} \in \Delta_{\mathcal{B}} \right\}, \tag{4.7}$$

where the norm $\|\cdot\|_F$ defines the root of sum-of-squares objective and the constraints

$$\Delta_{\boldsymbol{A}} = \left\{ \boldsymbol{A} \in \mathbb{R}_+^{I \times R} \,\middle|\, \textstyle\sum_{r \in [R]} \boldsymbol{A}[i, r] = 1 \ \forall i \in [I] \right\}, \tag{4.8a}$$

$$\Delta_{\mathcal{B}} = \left\{ \mathcal{B} \in \mathbb{R}_+^{R \times J \times K} \,\middle|\, \textstyle\sum_{k \in [K]} \mathcal{B}[r, j, k] = 1 \ \forall (r, j) \in [R] \times [J] \right\}, \tag{4.8b}$$

preserve the probabilistic interpretations of the proportion matrix $\boldsymbol{A}$ and the source-feature tensor $\mathcal{B}$.

A benefit of transforming the original recorded data values into probabilities is the implicit normalization of the measurements obtained across the various features, rendering them unit-independent. This normalization is crucial because it prevents biases that might otherwise favour features with larger scale units over those with smaller ones. Without this transformation, the model's fit could disproportionately reflect the influence of features with larger numerical values.

4.4 The block coordinate descent algorithm

This section describes our implementation of the block coordinate descent algorithm for nonnegative tensor factorization, as developed by Xu and Yin [61]. The core of this algorithm involves the concept of *block convexity*: the objective function ℓ in Eq. (4.7) is convex with respect to each block of

Algorithm 1: Simplex constrained Tucker-1 decomposition

Input: $\mathcal{Y} \in \mathbb{R}_+^{I \times J \times K}$ s.t. $\sum_k \mathcal{Y}_{ijk} = 1 \ \forall i, j$, rank $R \in [I]$

Initialize $\boldsymbol{A}^1 \in \Delta_{\boldsymbol{A}}$ and $\mathcal{B}^1 \in \Delta_{\mathcal{B}}$

for $t = 1, 2, \ldots$ **do**

  if *converged* **then** break

  $\boldsymbol{A}^{t+1} = P_{\boldsymbol{A}}\left(\boldsymbol{A}^t - \frac{1}{L_{\boldsymbol{A}}} \nabla_A \ell(\boldsymbol{A}^t, \mathcal{B}^t)\right)$ `[projected gradient update to A]`

  $\mathcal{B}^{t+1} = P_{\mathcal{B}}\left(\mathcal{B}^t - \frac{1}{L_{\mathcal{B}}} \nabla_{\mathcal{B}} \ell(\boldsymbol{A}^{t+1}, \mathcal{B}^t)\right)$ `[projected gradient update to B]`

Output: $\boldsymbol{A}^t$, $\mathcal{B}^t$

variables, $\boldsymbol{A}$ or $\mathcal{B}$, when the other block is fixed. Thus, while ℓ is not convex jointly over $(\boldsymbol{A}, \mathcal{B})$, each block update can be approached as a convex optimization problem.

Each iteration of the algorithm improves the fit by alternately fixing each block and updating the other. For each block, we employ a projected-gradient step, where the step sizes $1/L_{\boldsymbol{A}}$ and $1/L_{\mathcal{B}}$ are derived, respectively, from the Lipchitz constants of the gradients of the functions $\ell(\cdot, \mathcal{B})$ and $\ell(\boldsymbol{A}, \cdot)$. The Euclidean projections $P_{\boldsymbol{A}}$ and $P_{\mathcal{B}}$ enforce the constraints by ensuring that the updates remain within the feasible sets defined by $\mathcal{C}_{\Delta \boldsymbol{A}}$ and $\mathcal{C}_{\Delta \mathcal{B}}$. This procedure is outlined in Algorithm 1.

A significant adaptation in our implementation, relative to that of Xu and Yin, is our choice to preserve the tensor and matrix structures of the data throughout the computation rather than flattening them into matrices for updates. This approach, while mathematically equivalent to the flattening approach, offers conceptual clarity and aligns more naturally with the tensor structure of the data. This choice simplifies the implementation and enhances the interpretability of the algorithm.

4.5 Rank estimation

The effectiveness of the tensor decomposition approach depends heavily on choosing an appropriate rank R, which represents the number of latent sources in the model. Determining the optimal rank is crucial because it balances the model's complexity against its ability to accurately fit the observed data [3, 30, 35, 38].

In situations where we have no prior information about the number of sources, we use a method that identifies the rank at the point of maximum curvature in the relationship between the model's rank and its misfit with the observed data [35]. This method involves numerically estimating the second derivative of the loss function with respect to the rank using 5-point centered finite difference estimates, except at the boundaries, which are one-sided [1, Table 25.2]. The optimal rank $\hat{R}$ is selected by maximizing the curvature of the loss function $\ell(R) := \ell(\boldsymbol{A}_R, \mathcal{B}_R)$, which measures the discrepancy between the empirical distribution tensor $\mathcal{Y}$ and its factorized approximation at rank R, i.e.,

$$\hat{R} = \underset{R \in [I]}{\operatorname{argmax}} \frac{\ell''(R)}{(1 + \ell'(R)^2)^{1.5}}. \tag{4.9}$$

Here, $\ell'(R)$ and $\ell''(R)$ are the first and second derivatives of the loss function with respect to the rank, evaluated numerically. This formula captures the most significant change in the rate of improvement in the model fit as rank increases, suggesting the point beyond which additional complexity (higher rank) does not yield proportionate gains in fitting the data. As suggested by Satopää et al. [35], we can first normalize the input-output pairs $(R, \ell(R))$ to the unit square to ensure that the curvature metric is scale-invariant before estimating the first and second derivatives and applying the formula Eq. (4.9).

The maximum rank considered, I, corresponds to the number of sinks in the data set and is the size of the first dimension of $\mathcal{Y}$. This choice represents a scenario where each sink is modeled as a unique source, and the mixing proportions are trivial. A practical approach to rank selection is to start with $R = 1$ and incrementally test higher ranks until the rate of improvement in model fit

diminishes, as indicated by the curvature metric. This methodology ensures that the chosen rank is justified based on the model's performance, and also keeps the computational burden manageable. Depending on the outcome of these tests and the available computational resources, it may be feasible to explore ranks up to I, or to stop at a lower rank if the gains in model fit are minimal. Figure 7.2 illustrates the rank estimation process for our experimental data set. In practice, we find the estimated rank to be robust to the maximum rank considered. For this reason, we only consider $R \in [7]$ in Sec. 7.2.1 to avoid computing more expensive decompositions for larger ranks.

5 Grain source identification

After identifying the latent source distributions, we systematically categorize each grain according to its most likely origin. We use the features of the individual zircon grains to determine the likelihood that their multivariate feature combinations originate from one of the inferred multivariate source distributions, and assign each grain to the source with the highest likelihood.

Although the features are not necessarily independent we approximate the probability that a grain feature vector $\boldsymbol{g}$ originates from a particular source r by the product of the probabilities that each feature j of $\boldsymbol{g}$ originates from the corresponding source distribution β_{rj}. This probability is computed as

$$p_r = \mathbb{P}(\boldsymbol{g} \in \mathcal{V}_{\boldsymbol{g}} \mid \boldsymbol{g} \sim B_r) = \prod_{j \in [J]} \mathbb{P}\left(x_{jk_j} \leq \boldsymbol{g}[j] \leq x_{j(k_j+1)} \;\middle|\; \boldsymbol{g}[j] \sim \beta_{rj} \right),$$

where $\mathcal{V}_{\boldsymbol{g}} = [x_{1k_1}, x_{1(k_1+1)}] \times \cdots \times [x_{Jk_J}, x_{J(k_J+1)}]$ forms a J-array of tuples within which the features of the grain lie. The product distribution B_r, defined by (2.1), encapsulates the joint distribution of the features for source r. The index k_j indicates the interval $[x_{jk_1}, x_{j(k_1+1)}]$ that $\boldsymbol{g}[j]$ falls within. These sampling points reference the discretization detailed in Sec. 3.2.

The grain $\boldsymbol{g}$ is assigned a label from the source distribution $B_{\hat{r}}$ where $\hat{r} = \operatorname{argmax}_r p_r$, and p_r is approximated using the estimated source distributions contained in the core tensor $\boldsymbol{\mathcal{B}}$, i.e.,

$$p_r = \prod_{j=1}^{J} \beta_{rj}(x_{j\hat{k}_j}) \approx \prod_{j=1}^{J} \frac{1}{\Delta x_j} \boldsymbol{\mathcal{B}}[r, j, \hat{k}_j],$$

where we set the index

$$\hat{k}_j = \begin{cases} 1 & \text{if } \boldsymbol{g}[j] < x_{j1} \\ k_j & \text{if } x_{jk_j} \leq \boldsymbol{g}[j] < x_{j(k_j+1)} \\ K & \text{if } \boldsymbol{g}[j] \geq x_{jK}. \end{cases}$$

See Eq. (4.6).

In practice, we can enhance the accuracy of estimating the individual feature probabilities shown in the left-hand side of (5) by averaging the value $\boldsymbol{\mathcal{B}}[r, j, \hat{k}_j]$ with its neighboring value $\boldsymbol{\mathcal{B}}[r, j, \hat{k}_j + 1]$. This approach is analogous to using a trapezoidal estimate for calculating the area under the density function, as opposed to using just the value at the left end point. For even more precise estimates, we could interpolate the estimated densities $\boldsymbol{\mathcal{B}}[r, j, :]$ into a continuous function $\tilde{\beta}_{rj}$ using techniques like splines or a weighted moving average, which would provide a smoother and more accurate representation of the density functions.

6 Software implementation

Our computational approach is implemented in the Julia programming language [5] and is available as a fully-reproducible package from the GitHub repository `SedimentSourceAnalysis.jl` [33]. The tensor factorization code is based on `MatrixTensorFactor.jl` [32]. Here, we detail the key aspects of the software implementation, including the discretization of the kernel density estimates, the optimization algorithm, and the rank estimation procedure. We also discuss the grain source identification process and the visualization of the results.

Algorithm 2: Relaxed constraint Tucker-1 Decomposition

Input: $\mathcal{Y} \in \mathbb{R}_+^{I\times J\times K}$ s.t. $\sum_k \mathcal{Y}_{ijk} = 1\ \forall i,j$, and rank $R \in [I]$
Initialize $\boldsymbol{A}^1 \in \Delta_{\boldsymbol{A}}$ and $\mathcal{B}^1 \in \Delta_{\mathcal{B}}$
for $t = 1, 2, \ldots$ **do**
  if *converged* **then** break
  $\boldsymbol{A}^{t+\frac{1}{2}} = \left[\boldsymbol{A}^t - \frac{1}{L_A}\nabla_A \ell(\boldsymbol{A}^t, \mathcal{B}^t)\right]_+$ `[Projected gradient update to A]`
  $\mathcal{B}^{t+\frac{1}{2}} = \left[\mathcal{B}^t - \frac{1}{L_B}\nabla_{\mathcal{B}} \ell(\boldsymbol{A}^{t+\frac{1}{2}}, \mathcal{B}^t)\right]_+$ `[Projected gradient update to B]`
  $\boldsymbol{C}[r,r] = \frac{1}{J}\sum_{jk} \mathcal{B}^{t+\frac{1}{2}}[r,j,k]\ \forall r \in [R]$ `[Update diagonal renormalization matrix]`
  $\boldsymbol{A}^{t+1} = \boldsymbol{A}^{t+\frac{1}{2}}\boldsymbol{C}$
  $\mathcal{B}^{t+1} = (\boldsymbol{C}^{t+1})^{-1}\mathcal{B}^{t+\frac{1}{2}}$ `[Renormalize into the set` $\tilde{\Delta}_{\mathcal{B}}$`; see Eq. (6.1)]`
Output: $\boldsymbol{A}^t$, $\mathcal{B}^t$

6.1 Constraint relaxation

To improve the computational efficiency of the decomposition step, we modified the constrained Tucker-1 decomposition method described by Algorithm 1. Instead of cyclically applying gradient steps followed by projections onto the probability simplices described by the constraints in Eq. (4.8), we use a dynamic strategy that includes nonnegative projections coupled with a renormalization process. This method is described below and summarized in Algorithm 2.

Each iteration of the algorithm begins with gradient update steps followed by projections of the variables into the nonnegative orthant. Immediately after, a renormalization step adjusts $\boldsymbol{A}$ and $\mathcal{B}$ to satisfy relaxed constraint conditions. Specifically, the renormalization ensures that $\mathcal{B}$ satisfies the constraint set

$$\tilde{\Delta}_{\mathcal{B}} = \left\{\mathcal{B} \in \mathbb{R}_+^{R\times J\times K} \;\middle|\; \textstyle\sum_{j\in[J]}\sum_{k\in[K]} \mathcal{B}[r,j,k] = J\ \forall r \in [R]\right\}, \tag{6.1}$$

which relaxes the original constraint set $\Delta_{\mathcal{B}}$ in Eq. (4.8b). This constraint is enforced by rescaling the tensor $\mathcal{B}$, rather than by a Euclidean projection onto the constraint. Subsequently, these weights are moved to the matrix $\boldsymbol{A}$, which ensures that the objective value $\ell(\boldsymbol{A}, \mathcal{B})$ does not increase after renormalization. This is not necessarily the case when using the simplex projection in Algorithm 1, where projection onto the simplex-like sets in Eqs. (4.8a) and (4.8b) could increase the objective.

6.2 KDE bandwidth selection

To calculate the bandwidth h_j, we separately calculate a bandwidth h_{ij} for the data $\{\boldsymbol{g}_i^n[j]\}_{n=1}^N$, and take h_j to be the median over $\{h_{ij}\}_{i=I}^I$. We calculate h_{ij} according to Silverman's rule of thumb [43] on the trimmed data. We use 2.5% trimming (inner 95 percentile) to minimize the chance that the bandwidth is made arbitrarily large from a few outliers. The bandwidth calculation is given by Algorithm 3. Other bandwidth selection methods, such as cross-validation [57] or the improved Sheather-Jones method [21], could be used instead of Silverman's rule of thumb. We have not detected significant differences in the learned mixing coefficient matrix $\boldsymbol{A}$, the estimated rank R, and the grain source identification when using these alternative methods, although the learned densities $\mathcal{B}$ are more or less smoothed.

6.3 Implementation of KDE discretization

Our method for turning each continuous probability density function $f_{ij}(x)$ into K samples $f_{ij}(x_{ik})$, $k \in [K]$ has three steps:

1. Fix the number of samples K from each KDE f_{ij}. For the experiments in Sec. 6, we set $K = 64$. Numerical testing with different numbers of samples ($K = 16, 32, 64, 128, 256$) show that the

Algorithm 3: Bandwidth selection

Input: Data $\boldsymbol{g}_i^n \in \mathbb{R}^J$, inner percentile filter $p \in (0, 100]$
for $j \in [J]$ **do**
  for $i \in [I]$ **do**
    Collect $\boldsymbol{g}_i^n[j]$ for $n = 1, \ldots, N$
    Keep only $\boldsymbol{g}_i^n[j]$ within the middle p percentile
    Compute $h_{ij} = 0.9 \min(\hat{\sigma}, \frac{IQR}{1.34})$ `[Silverman's rule of thumb]`
  $h_j = \texttt{median}\{h_{ij}\}_{i=I}^{I}$ `[average the bandwiths across all sinks i]`

results are robust to the value of K. Powers of 2 are preferred because the kernel density procedure we use for our implementation is based on a fast Fourier transform.

2. Choose the domain over which we sample the KDE for each feature $j \in [J]$: for each j, the left end point is the smallest sample across all sinks $\min_{i,n} \boldsymbol{g}_i^n[j]$, and the right end point is the largest $\max_{i,n} \boldsymbol{g}_i^n[j]$. Because the shifted kernel κ is nonzero slightly to the left and right of the grain samples, we move the left endpoint an additional 3 bandwidths, $-3h_j$, and the right end, $+3h_j$ to capture most of the density of our density estimate f_{ij}.

   In summary, for each feature $j \in [J]$, uniformly sample the domain

$$\left[\min_{i\in[I],n\in[N_j]} \boldsymbol{g}_i^n[j] - 3h_j,\ \max_{i\in[I],n\in[N_j]} \boldsymbol{g}_i^n[j] + 3h_j\right]$$

   to obtain K samples $x_{j1}, \ldots, x_{jK}$.

3. Finally, sample all continuous KDEs f_{ij}, $i \in [I]$, $j \in [J]$ to obtain samples $f_{ij}(x_{ik})$, $k \in [K]$.

6.4 Grain label confidence score

We calculate a confidence score for the grain labels using the probabilities estimated according Sec. 5. We define the score to be

$$\text{confidence score} = \min\left(1, \log_{10}\left(\frac{p_{(1)}}{\max(p_{(2)}, \epsilon(p_{(1)}))}\right)\right) \tag{6.2}$$

where $p_{(1)} = p_{\hat{r}}$ and $p_{(2)}$ are the top two probabilities estimated, and $\epsilon(p_{(1)})$ (machine epsilon at $p_{(1)}$) is compared against $p_{(2)}$ to avoid division by zero. We interpret a score of 1 to indicate it is at least 10 times as likely the grain came from source $r = \hat{r}$ than any other source, and a score near 0 to indicate there is a similar probability the grain came from a difference source.

6.5 Evaluation Metrics

Many evaluation quantities are used in practice to assess closeness of data [28]. We primarily use the L_2 error, mean absolute error (MAE), and a version of mean relative error (MRE) suitable for our problem.

To evaluate the quality of the factors $\boldsymbol{A}$ and $\boldsymbol{\mathcal{B}}$, we permute the R learned proportions $\hat{\boldsymbol{A}}[:, r]$ and source densities $\hat{\boldsymbol{\mathcal{B}}}[r, :, :]$, $r \in [R]$, by greedily minimizing the L_2 error between the learned proportions $\hat{\boldsymbol{A}}[:, r]$ and true proportions $\boldsymbol{A}[:, r]$. This allows a fair and consistent comparison between the algorithm output and the ground truth because the metrics used are mainly computed entrywise.

L2 error When permuting the R learned sources, we need to compare the closeness of these sources $\hat{r} \in [R]$ to the ground truth sources $r \in [R]$. To do this, we use the L_2 error defined by

$$L_2 \text{ error} := \left\| \hat{\boldsymbol{A}}[:, \hat{r}] - \boldsymbol{A}[:, r] \right\|_2.$$

Mean absolute error We calculate the mean absolute error to evaluate the closeness of two $I \times R$ matrices $\boldsymbol{M}$ and $\hat{\boldsymbol{M}}$ by

$$\mathrm{MAE} := \frac{1}{IR} \sum_{i \in [I]} \sum_{r \in [R]} \left| \boldsymbol{M}[i, r] - \hat{\boldsymbol{M}}[i, r] \right|.$$

Mean relative error We calculate the mean relative error (MRE) of the 3-fibres between two arbitrary $I \times J \times K$ tensors $\boldsymbol{\mathcal{T}}$ and $\hat{\boldsymbol{\mathcal{T}}}$ by

$$\mathrm{MRE} := \frac{1}{IJ} \sum_{i \in [I]} \sum_{j \in [J]} \frac{\left\| \boldsymbol{\mathcal{T}}[i, j, :] - \hat{\boldsymbol{\mathcal{T}}}[i, j, :] \right\|_2}{\left\| \boldsymbol{\mathcal{T}}[i, j, :] \right\|_2}.$$

7 Empirical evaluation

We describe the setup and methodology of a numerical experiment conducted using a dataset collected by Sundell et al. [51]. The dataset consists of sediment samples (sinks) collected from various locations in a basin. Each sample was analyzed to identify seven distinct features, including the age of zircon grains and several geochemical markers. The objectives of this experiment are to identify the number of latent sources contributing to these samples, determine the proportion of each source within the samples, learn the distributions of these sources, accurately label each grain by its source, and validate these findings against established ground truths.

7.1 Experimental setup

The dataset from Sundell et al. was originally used to reconstruct continental crustal thicknesses over time in the central Andes. It includes U-Pb ages and concentrations of 23 trace and rare-earth elements: P, Sc, Ti, Y, Nb, La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Hf, Ta, Th, and U. Three artificial source ($R = 3$) were constructed by randomly drawing 75, 140, and 399 grains, without replacement, from the original dataset of 710 grains. For each source, seven key variables were selected for analysis: age, Eu anomaly, Ti-based crystallization temperature, Th-U ratio, sum of light rare-earth elements over heavy rare-earth elements (ΣLREE/ΣHREE), Dy-Yb ratio, and normalized (Ce/ND)/Y ratio. These variables were chosen based on their variability, as explored by Sundell et al. [51]. This variability is crucial because if all sources are identical, their mixtures will also be identical, providing no basis for the Tucker-1 method to discriminate. Ultimately, the choice of variables was guided by the need for distinct and informative features.

From these defined sources, $I = 20$ sinks were created by randomly selecting $N_1 = N_2 = N_3 = 75$ grains, without replacement, for each sink, ensuring each sink proportionally represents the $R = 3$ sources. This procedure resulted in a ground truth proportion matrix with entries $\boldsymbol{A}^\natural[i, r]$ indicating the fraction of grains in each sink i originating from source r, normalized such that the sum of proportions for each sink sums to one, i.e., $\sum_{r \in [R]} \boldsymbol{A}^\natural[i, r] = 1$.

The raw data is converted into the empirical density tensor $\boldsymbol{\mathcal{Y}}$ following the procedure of Sec. 1.2, and subsequently factorized using Algorithm 2 with ranks ranging from $R = 1$ to $R = 20$ (up to the number of sinks I). The factorization process was terminated when the norm of the projected gradient (a standard measure of optimality) reached the tolerance 10^{-5}.

7.2 Results

The experimental outcomes were promising, demonstrating a mean relative error (MRE) of 11.2% between the empirical distribution tensor $\boldsymbol{\mathcal{Y}}$ and the decomposed tensor $\hat{\boldsymbol{\mathcal{Y}}} = \hat{\boldsymbol{A}}\hat{\boldsymbol{\mathcal{B}}}$, with an optimal rank correctly estimated as $R = 3$. This MRE suggests that the decomposition model can explain approximately 88.8% of the data. The fitting of the learned proportion matrix $\hat{\boldsymbol{A}}$ to the ground truth $\boldsymbol{A}^\natural$ revealed a mean absolute error (MAE) of 4.8%. Additionally, comparing the learned density tensor $\hat{\boldsymbol{\mathcal{B}}}$ to the hypothetical ground-truth density tensor $\boldsymbol{\mathcal{B}}^\natural$ yielded an MRE of 9.0%.

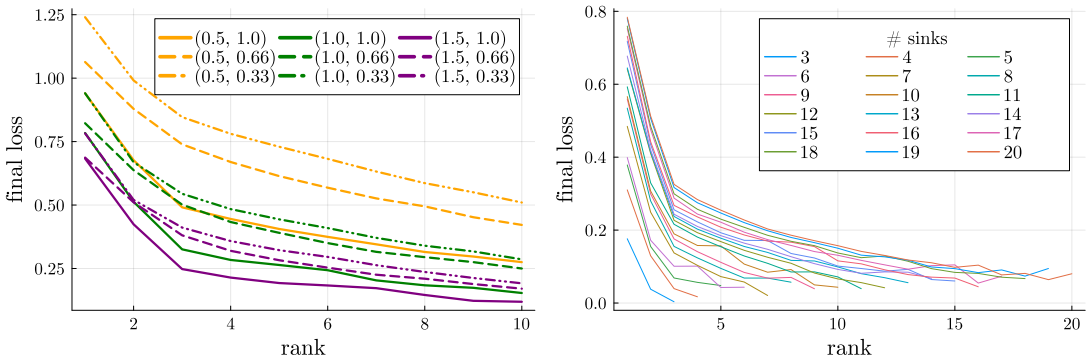

Figure 7.1: The final objective loss versus rank. Each line in the left panel varies the bandwidth relative to that obtained by Silverman's rule, and varies the fraction of grains used, e.g., $(0.5, 0.66)$ indicates the results obtained by halving Silverman's bandwidth and using 66% of the overall available grains. The right panel shows results for varying number of sinks I.

These findings are elaborated in Sec. 7.2.2 and Sec. 7.2.3. See Sec. 6.5 for details on the error metrics used. Grain labeling results, detailed in Sec. 7.2.4, demonstrated an accuracy of 88.5% across all sinks, affirming the effectiveness of the labeling approach based on the learned source densities.

7.2.1 Rank estimation

The optimal rank for this dataset was estimated using the procedure outlined in Sec. 4.5. This method can be seen as a refinement of the approach by Saylor et al. [36], which selects the rank based on segmented linear regression. We only check up to a maximum rank of 7 in our main experiment, rather than the theoretical maximum $I = 20$ to avoid the computational expense of decompositions for large ranks. We find the estimated rank is robust to selecting different maximum ranks, allowing for computational savings by restricting the search to smaller, more likely optimal ranks. Additionally, as shown in Fig. 7.1, the rank estimation is robust to variations in several parameters, including the KDE bandwidths h_j, the number of grain samples in each sink N, the number of sinks I. This can be seen by observing a noticeable "elbow" or "knee" at rank $R = 3$ in all tests. Additional tests show rank is also robust to the number of features used J, and number of discretization points K.

Figure 7.2 shows the error as a function of rank in our main experiment, with the standard curvature of the loss function plotted on the right. The optimal rank is the argument that maximizes the curvature, which in this case is $R = 3$. The error plot in Fig. 7.2 (left) exhibits a significant reduction in loss from ranks 1 to 3, after which the improvements plateau, indicating diminishing returns with higher ranks. This observation is corroborated by the standardized curvature plot in Fig. 7.2 (right), which displays a clear maximum at $R = 3$. This point marks where the increase in rank no longer justifies the marginal improvement in fitting error.

7.2.2 Proportion matrix accuracy

The comparison between the true and learned proportion matrices $\boldsymbol{A}$ and $\hat{\boldsymbol{A}}$ is shown in Fig. 7.3. The MAE of 4.8% reflects a modest deviation, averaging 5%, between the estimated and actual source proportions.

7.2.3 Source feature density tensor accuracy

In our experimental framework, the source of each grain is known, but the actual densities of these tensor sources are not directly available. Therefore, to make a meaningful comparison between the learned density tensor $\hat{\boldsymbol{\mathcal{B}}}$ and a hypothetical true density tensor $\boldsymbol{\mathcal{B}}^\natural$, we estimate the true densities using a KDE approach similar to Sections 3.1 and 3.2, leveraging the known labels of each grain.

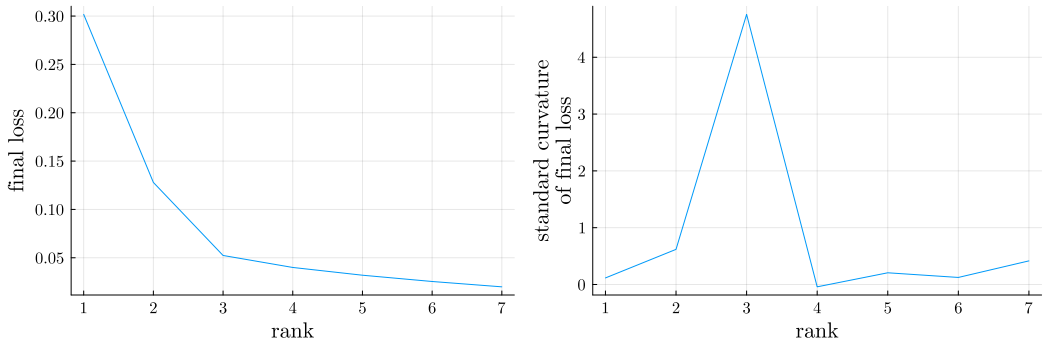

Figure 7.2: (Left) The final objective loss (cf. Eq. (4.7)) obtained by Algorithm 2 for the experiment described in Sec. 7. (Right) standard curvature of the loss curve, indicating a "knee" at rank $R = 3$, which represents the optimal balance between low rank (simpler model) and low error (expressive model). This rank was identified by the maximum standard curvature, as shown on the right; see Eq. (4.9).

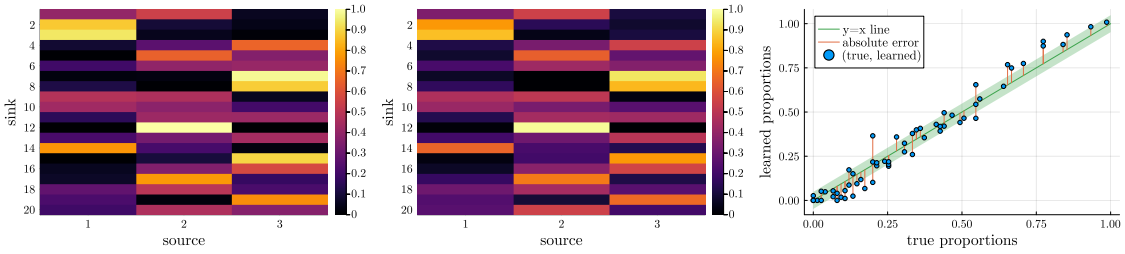

Figure 7.3: (Left) Learned proportions $\hat{\boldsymbol{A}}$, (middle) ground truth $\boldsymbol{A}$, and (right) learned entries of $\hat{\boldsymbol{A}}$ plotted against true entries of $\boldsymbol{A}$ from the experiment in Sec. 7. The vertical bars represent the absolute error between entries, and the shaded region shows a deviation of ± 4.8 percentage points from equality. The points deviate on average 4.8% from equality, which is the same amount as the MAE between the learned and true proportions.

This approach constructs $\mathcal{B}^\natural$ as our best representation of the source distributions from which the grains are assumed to originate.

It is important that the bandwidth h_j and the sample points $x_{j1}, \ldots, x_{jK}$ used for the KDE of the true densities match those used to create the input tensor $\mathcal{Y}$. This consistency ensures that any comparison between $\mathcal{B}^\natural$ and $\hat{\mathcal{B}}$ is based solely on the differences in density estimation, without confounding discrepancies in the KDE parameters.

The MRE between $\hat{\mathcal{B}}$ and $\mathcal{B}^\natural$ is 9.0%, indicating an average relative deviation of 9% in the learned distributions from those estimated as true. This quantity underscores the accuracy of the learned densities in approximating the true source distributions. We note that our algorithm is successful even though these sources are relatively similar to each other, which points to the stability of the approach.

7.2.4 Grain labels

In our methodology, grains are labeled according to their most probable source based on the learned densities (cf. Sec. 5). Ideal labeling would result in distinct clusters of grains assigned to their respective sources without error. Figure 7.5 demonstrates the labeling performance in the first sink, where the grains are organized by their source of origin and colour-coded based on the confidence of their labeling, as described in Sec. 6.4. High-confidence grains are generally labeled correctly, whereas grains with lower confidence scores are more prone to mislabeling, though some are correctly identified by chance rather than accurate modeling.

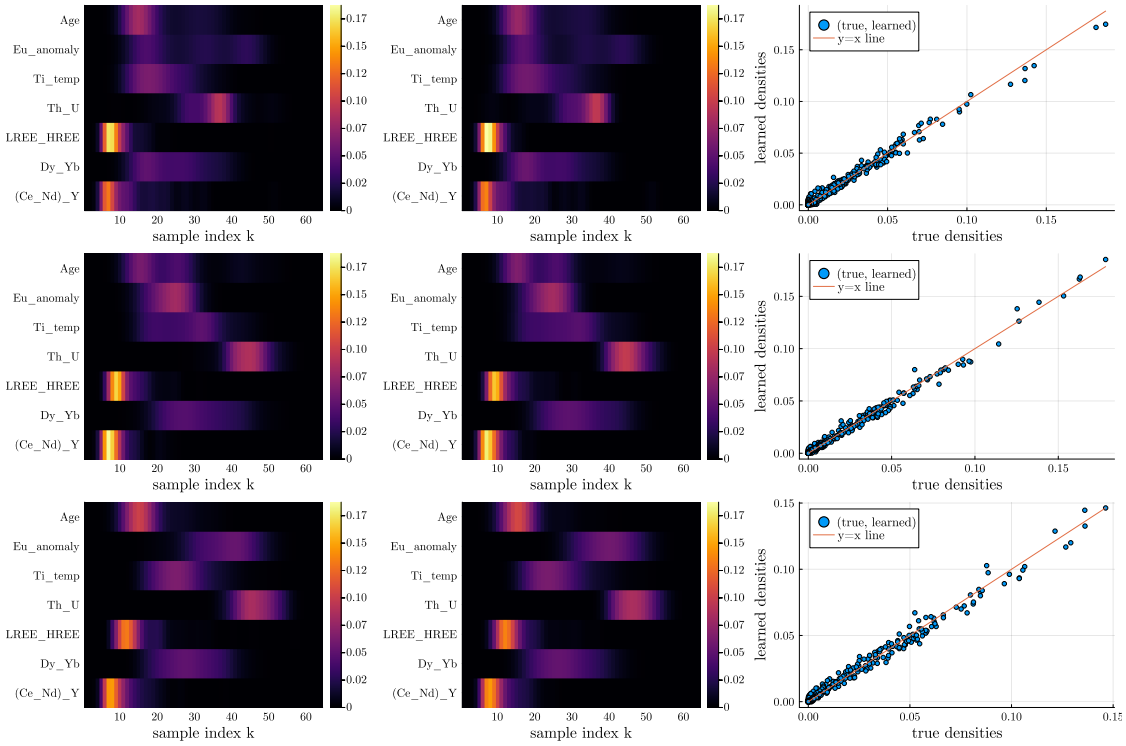

Figure 7.4: For sources $r = 1$ (top row), $r = 2$ (middle row), and $r = 3$ (bottom row), we display (left column) learned source densities $\hat{\mathcal{B}}[r, :, :]$, (middle column) true source densities $\mathcal{B}^{\natural}[r, :, :]$, and (right column) a comparison of learned entries of $\hat{\mathcal{B}}$ against hypothetical true entries of $\mathcal{B}^{\natural}$ for the experiment in Sec. 7. The plots show a close match, with most points lying near the diagonal line, indicating that the learned densities closely mirror the true densities.

The overall labeling accuracy across all grains from every sink is $1327/1500 = 88.5\%$. This high success rate suggests that the model effectively captures and applies the distinguishing features of each source.

To further validate the model's labeling, we compare the proportions of labeled grains in each sink to the expected proportions derived from the true source proportions given by $\boldsymbol{A}$. Specifically, we count grains labeled as belonging to source r in each sink i, normalize these counts by the total grains in sink i, and compare to the corresponding entry $\boldsymbol{A}[i, r]$. This comparison is visualized in Fig. 7.6, where we find an MAE of 4.1% between the learned and true proportions, indicating close alignment. Because in practice the true source proportions $\boldsymbol{A}$ are not available, we also compare the grain labels to the learned proportions $\hat{\boldsymbol{A}}$, resulting in an MAE of 6.8%.

7.2.5 Independence and multiple features

As highlighted in the introduction, current approaches are limited to data with only two features. Rather than characterizing sources and sinks via the multidimensional joint distributions, we choose to model the product distribution by constructing an array of J 1-dimensional feature distributions, allowing us to contain the data in a 3rd order tensor. We make this choice because multidimensional joint discretized KDEs would require a tensor of order $J + 1$, which could lead to intractable computation when using many features. If the features were independent, the product distribution would be equivalent to the joint distribution, but this may not be the case for an arbitrary set of features. Nonetheless, two experiments described below are meant to validate the use of a product distribution in place of a J-dimensional joint distribution.

In the first experiment, we consider only two features (Eu anomaly and Ti-based crystallization

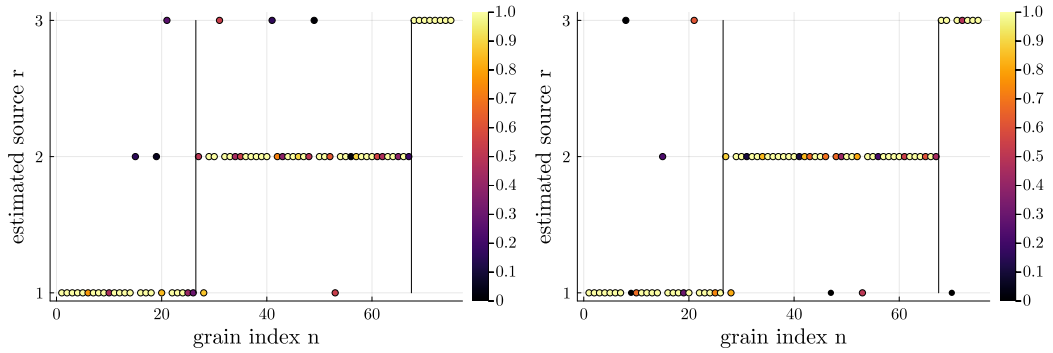

Figure 7.5: (Left) Estimated grain labels for sink $i = 1$ using the model densities from Sec. 7. (Right) Labels for the same grains using the true densities; see Sec. 7.2.3. Labeling accuracy for this sink was $67/75 = 89.3\%$ using model densities and $68/75 = 90.6\%$ using true densities. Grains are sorted by their original source for clearer visualization, with markers colored according to their confidence scores as defined in Sec. 6.4.

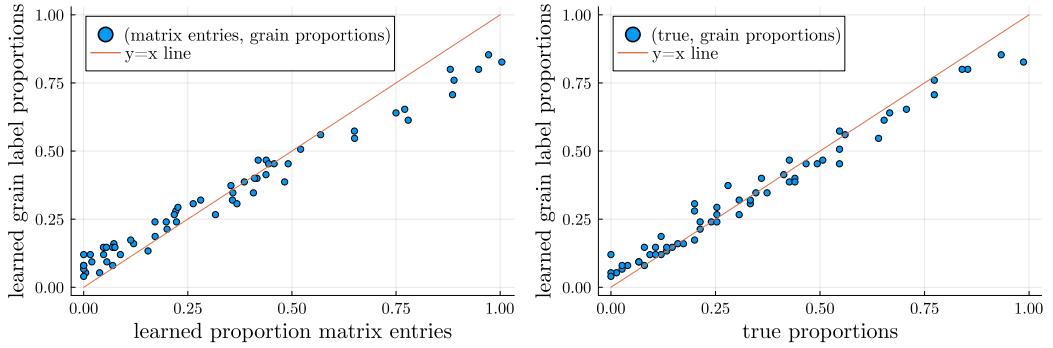

Figure 7.6: (Left) Comparison between the proportion of learned grain labels in each sink with the learned proportion matrix $\hat{\boldsymbol{A}}$ from Sec. 7. (Right) Comparison between the proportion of learned grain labels and the ground truth proportions $\boldsymbol{A}$. We observe an MAE between the corresponding matrices of 6.8% and 4.1% respectively. See Sec. 7.2.4 for more details.

temperature) using our method with 1-dimensional feature KDEs. Our methodology achieved a labeling accuracy of 76.1%, which is marginally lower than 79.5% accuracy obtained by labeling based on known source distributions. This result underscores that incorporating a greater number of features enhances labeling accuracy, as detailed in Sec. 7.2.4.

The second experiment explores the trade-off between estimating joint distributions of two features and against using more features under the assumption of independence. Here, the joint distributions are modeled with 2-dimensional KDEs. The optimal scenario, using known joint-feature densities, achieved a labeling accuracy of 79.5%, which happens to coincide with the first experiment's results. A variant of our model that estimates joint-feature distributions achieved an accuracy of 80.4%, which is close to the theoretical optimum, yet still less than the accuracy attained when all seven features are considered.

8 Concluding remarks

Detrital provenance research continues to expand the variety of features that can be extracted from individual detrital grains. However, the ability to collect this data has surpassed the scientific community's ability to interpret and model it quantitatively. This study introduces a novel approach to multivariate sediment source identification using the Tucker-1 decomposition applied to various

sediment samples.

We have demonstrated that the Tucker-1 decomposition method can effectively recover both the source feature distributions and the proportion of each source in the sink samples. The success of this method hinges on precise rank estimation, for which we have introduced and validated a novel technique based on maximizing the curvature of the residual function. Additionally, we have shown that individual zircon grains can be confidently attributed to specific latent sediment sources with approximately 90% accuracy by maximizing the likelihood of their feature distribution given the estimated latent source distributions. All methodologies discussed are available as fully reproducible open-source software.

There are several potential improvements and adaptations for Algorithm 2 that could enhance its applicability.

Sampling efficiency Instead of uniform sampling for the KDE, as described in Sec. 6.3, sampling could be weighted by the density values or gradients. This targeted importance sampling might yield a more efficient distribution of samples x_{ij}, ensuring that areas with higher information density are sampled more intensively.

Model extensions for complex distributions If the features are not independent, then their joint distribution is not fully captured by the product of the 1-dimensional distributions. In this case, some insight on the interaction between features may be lost. One could extend our method to one that fully captures the joint distributions by expanding the decomposition model from a third order tensor model to a higher-order tensor model, where both $\mathcal{Y}$ and $\mathcal{B}$ would be tensors of order $(J+1)$, rather than order 3. In this extension, the first-order slices $\mathcal{Y}[i,:,\ldots,:]$ $(i \in [I])$ and $\mathcal{B}[r,:,\ldots,:]$ $(r \in [R])$ represent the normalized densities for sink i and source r, respectively. This would involve a more complex multiplication scheme between $\boldsymbol{A}$ and $\mathcal{B}$, extending across multiple dimensions:

$$\mathcal{Y}[i_0, i_1, \ldots, i_J] = (\boldsymbol{A}\mathcal{B})[i_0, i_1, \ldots, i_J] = \sum_{r \in [R]} \boldsymbol{A}[i_0, r] \cdot \mathcal{B}[r, i_1, \ldots, i_J]. \tag{8.1}$$

Alternative metrics The fitting of the empirical density tensor $\mathcal{Y}$ to the model $\boldsymbol{A}\mathcal{B}$ could be performed using alternative metrics such as the 1-norm, Kullbach-Leibler divergence, or Wasserstein distance, instead of the sum-of-least-squares error currently used. These metrics may provide models that are better suited to the nature of the data and the specific research questions at hand.

We also note that alternate distribution representations could be used to perform the decomposition of sinks into sources. One example is the empirical cumulative distribution function. This approach would eliminate the need to use kernel density estimation to determine the number of sources and the mixture coefficients. However, some form of smoothing would still be required to turn the estimated cumulative distributions into probability density functions in order to perform grain labeling.

These refinements and extensions may enhance the model's utility and accuracy in complex multivariate sediment source identification problems.

Acknowledgments

The authors extend their sincere thanks to Glenn Sharman and Pieter Vermeesch for their detailed and thoughtful comments that helped to improve our presentation.

Statements and Declarations

Funding

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) under their Discovery Grant program.

Symbol	Description
$[I]$	set of integers from 1 to I, $\{1, 2, \ldots, I\}$
I, J, K, R, N_i	dimensions of vectors, matrices, and tensors
i, j, k, r, n	indexes that run from 1 to their corresponding dimension
$\mathbb{R}_+$	set of nonnegative real scalars
$\mathbb{R}_+^{I \times R}$	set of nonnegative matrices with I rows and R columns
$\mathbb{R}_+^{R \times J \times K}$	set of nonnegative $R \times J \times K$ 3-way tensors
I	number of sinks (collections of grains)
J	number of features/types of measurements
K	number of samples from the estimated continuous distributions
R	number of latent sources
N_i	number of grains collected and analysed from the ith sink
$\boldsymbol{g}_i^n \in \mathbb{R}^J$ where $n \in [N_i]$	nth grain in sink i, a vector of measured features
$\boldsymbol{A}$, $\boldsymbol{M}$	matrices
$\boldsymbol{\mathcal{Y}}$, $\boldsymbol{\mathcal{T}}$	tensor / three (or higher) dimensional array
$\boldsymbol{g}_i^n[j]$	the jth entry of the vector $\boldsymbol{g}_i^n$
$\boldsymbol{A}[i, r]$	entry in the ith row & rth column of the matrix $\boldsymbol{A}$
$\boldsymbol{\mathcal{B}}[r, j, k]$	entry in the rth row, jth column, & kth frontal slice of the tensor $\boldsymbol{\mathcal{B}}$
$\boldsymbol{A}[:, r]$	the rth column of the matrix $\boldsymbol{A}$
$\boldsymbol{\mathcal{B}}[r, j, :]$	the 3-fibre in the rth row and jth column of the tensor $\boldsymbol{\mathcal{B}}$
$\boldsymbol{A\mathcal{B}} := \boldsymbol{\mathcal{B}} \times_1 \boldsymbol{A}$	1-mode product/multiplication of a matrix and tensor
S_i	probability distribution of the ith sink
B_r	probability distribution of the rth source
β_{rj}	probability distribution of the jth feature in the rth source
$\alpha_{ir} \in \mathbb{R}_+$	scalar proportion of source r present in sink i
$h_j \in \mathbb{R}_{++}$	kernel density estimate bandwidth for the jth feature
$f_{ij} : \mathbb{R} \to \mathbb{R}_+$	kernel density estimate for the jth feature of sink i
$x_{jk} \in \mathbb{R}$	kth domain sample/input for densities of feature j
$\Delta x_j \in \mathbb{R}_{++}$	uniform step size between domain samples for densities of feature j
$\|\boldsymbol{v}\|_2$, $\|\boldsymbol{M}\|_F$, $\|\boldsymbol{\mathcal{T}}\|_F$	entrywise norm: root of sum-of-squared entries

Table 1: Summary of symbols, notation, and definitions used throughout the paper.

Conflicts of Interest/Competing Interests

The authors declare that they have no financial or non-financial conflicts of interest related to the content of this manuscript.

Financial Conflicts

There are no financial interests or relationships that could be perceived to influence the work reported in this paper.

Availability of Code, Data, and Material

The code used for analysis and the data supporting this study are available at the following repository: https://github.com/njericha/SedimentSourceAnalysis.jl

References

[1] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions: With Formulas, Graphs and Mathematical Tables*. In collab. with Conference on mathematical tables, National science foundation, and Massachusetts institute of technology. Unabridged, unaltered and corr.

republ. of the 1964 ed. Dover Books on Advanced Mathematics. New York: Dover publ, 1972. ISBN: 978-0-486-61272-0.

[2] W.H. Amidon, D.W. Burbank, and G.E. Gehrels. "Construction of detrital mineral populations: insights from mixing of U-Pb zircon ages in Himalayan rivers". In: *Basin Research* 17 (2005), pp. 463–485.

[3] Woody Austin, Tamara G. Kolda, and Todd Plantenga. "Tensor Rank Prediction via Cross Validation" (Sandia National Labs, Livermore, CA). Aug. 13, 2014.

[4] E. Belousova, W. Griffin, Suzanne Y. O'Reilly, and N. Fisher. "Igneous zircon: trace element composition as an indicator of source rock type". In: *Contributions to Mineralogy and Petrology* 143.5 (2002), pp. 602–622. ISSN: 1432-0967. DOI: 10.1007/s00410-002-0364-7.

[5] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B Shah. "Julia: A fresh approach to numerical computing". In: *SIAM review* 59.1 (2017), pp. 65–98.

[6] A. Bhattacharya. "On a measure of divergence between two statistical populations defined by their probability distributions". In: *Bulletin of the Calcutta Mathmatical Society* 35 (1943), pp. 99–109.

[7] I. H. Campbell, P. W. Reiners, C. M. Allen, S. Nicolescu, and R. Upadhyay. "He-Pb double dating of detrital zircons from the Ganges and Indus Rivers: Implication for quantifying sediment recycling and provenance studies". In: *Earth and Planetary Science Letters* 237.3-4 (2005), pp. 402–432. ISSN: 0012-821X. DOI: 10.1016/j.epsl.2005.06.043.

[8] Matthew J Campbell, Gideon Rosenbaum, Charlotte M Allen, and Carl Spandler. "Continental crustal growth processes revealed by detrital zircon petrochronology: Insights from Zealandia". In: *Journal of Geophysical Research: Solid Earth* 125.8 (2020), e2019JB019075. ISSN: 2169-9313.

[9] Tomas N. Capaldi, Sarah W. M. George, Jaime A. Hirtz, Brian K. Horton, and Daniel F. Stockli. "Fluvial and Eolian Sediment Mixing During Changing Climate Conditions Recorded in Holocene Andean Foreland Deposits From Argentina (31–33°S)". In: *Frontiers in Earth Science* 7.298 (2019). ISSN: 2296-6463. DOI: 10.3389/feart.2019.00298.

[10] L Caracciolo. "Sediment generation and sediment routing systems from a quantitative provenance analysis perspective: Review, application and future development". In: *Earth-Science Reviews* 209 (2020), p. 103226. ISSN: 0012-8252.

[11] B Carrapa, PG DeCelles, PW Reiners, GE Gehrels, and M Sudo. "Apatite triple dating and white mica 40Ar/39Ar thermochronology of syntectonic detritus in the Central Andes: A multiphase tectonothermal history". In: *Geology* 37.5 (2009), pp. 407–410. ISSN: 1943-2682.

[12] James B. Chapman, Mihai N. Ducea, Peter G. DeCelles, and Lucia Profeta. "Tracking changes in crustal thickness during orogenic evolution with Sr/Y: An example from the North American Cordillera". In: *Geology* 43.10 (2015), pp. 919–922. ISSN: 0091-7613. DOI: 10.1130/g36996.1.

[13] K. DeGraaff-Surpless, J. B. Mahoney, J. L. Wooden, and M. O. McWilliams. "Lithofacies control in detrital zircon provenance studies: Insights from the Cretaceous Methow basin, southern Canadian Cordillera". In: *Geological Society of America Bulletin* 115.8 (2003), pp. 899–915. ISSN: 0016-7606.

[14] W. R. Dickinson, L. S. Beard, G. R. Brakenridge, J. L. Erjavec, R. C. Ferguson, K. F. Inman, R. A. Knepp, F. A. Lindberg, and P. T. Ryberg. "Provenance of North-American Phanerozoic sandstones in relation to tectonic setting". In: *Geological Society of America Bulletin* 94.2 (1983), pp. 222–235. ISSN: 0016-7606.

[15] W. R. Dickinson, T. F. Lawton, and G. E. Gehrels. "Recycling detrital zircons: A case study from the Cretaceous Bisbee Group of southern Arizona". In: *Geology* 37.6 (2009), pp. 503–506. ISSN: 0091-7613. DOI: 10.1130/g25646a.1.

[16] C. M. Fedo, K. N. Sircombe, and R. H. Rainbird. "Detrital zircon analysis of the sedimentary record". In: *Zircon: Reviews in Mineralogy and Geochemistry*. Ed. by J. M. Hanchar and P. W. O. Hoskin. Vol. 53. Reviews in Mineralogy & Geochemistry. 2003, pp. 277–303. ISBN: 1529-6466 0-93995-065-0.

[17] George Gehrels and Mark Pecha. "Detrital zircon U-Pb geochronology and Hf isotope geochemistry of Paleozoic and Triassic passive margin strata of western North America". In: *Geosphere* 10.1 (2014), pp. 49–65. DOI: 10.1130/ges00889.1.

[18] George E. Gehrels and William R. Dickinson. "Detrital zircon geochronology of the Antler overlap and foreland basin assemblages, Nevada". In: *Paleozoic and Triassic paleogeography and tectonics of western Nevada and Northern California*. Ed. by Michael J. Soreghan and George E. Gehrels. Vol. 347. Geological Society of America, 2000, pp. 57–63. ISBN: 9780813723471. DOI: 10.1130/0-8137-2347-7.57.

[19] Douglas J. Jerolmack and Chris Paola. "Shredding of environmental signals by sediment transport". In: *Geophysical Research Letters* 37.19 (2010). ISSN: 0094-8276. DOI: https://doi.org/10.1029/2010GL044638.

[20] M. J. Johnsson, R. F. Stallard, and N. Lundberg. "Controls on the composition of fluvial sands from a tropical weathering environment - Sands of the Orinoco River drainage-basin, Venezuala and Colombia". In: *Geological Society of America Bulletin* 103.12 (1991), pp. 1622–1647. ISSN: 0016-7606.

[21] MC Jones and SJ Sheather. "Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives". In: *Statistics & Probability Letters* 11.6 (1991), pp. 511–514.

[22] Tamara G. Kolda and Brett W. Bader. "Tensor Decompositions and Applications". In: *SIAM Review* 51.3 (Aug. 6, 2009), pp. 455–500. ISSN: 0036-1445, 1095-7200. DOI: 10.1137/07070111X.

[23] Timothy F. Lawton, Cody D. Buller, and Todd R. Parr. "Provenance of a Permian erg on the western margin of Pangea: Depositional system of the Kungurian (late Leonardian) Castle Valley and White Rim sandstones and subjacent Cutler Group, Paradox Basin, Utah, USA". In: *Geosphere* (2015). DOI: 10.1130/ges01174.1.

[24] Daniel Lee and H. Sebastian Seung. "Algorithms for Non-negative Matrix Factorization". In: *Advances in Neural Information Processing Systems*. Vol. 13. MIT Press, 2000.

[25] Robert G Lee, Alain Plouffe, Travis Ferbey, Craig JR Hart, Pete Hollings, and Sarah A Gleeson. "Recognizing porphyry copper potential from till zircon composition: A case study from the Highland Valley Porphyry District, south-central British Columbia". In: *Economic Geology* (2021).

[26] A Licht, A Pullen, P Kapp, J Abell, and N Giesler. "Eolian cannibalism: Reworked loess and fluvial sediment as the main sources of the Chinese Loess Plateau". In: *Geological Society of America Bulletin* 128.5-6 (2016), pp. 944–956. ISSN: 0016-7606.

[27] J. Brian Mahoney, James W. Haggart, Marty Grove, David L. Kimbrough, Virginia Isava, Paul K. Link, Mark E. Pecha, and C. Mark Fanning. "Evolution of the Late Cretaceous Nanaimo Basin, British Columbia, Canada: Definitive provenance links to northern latitudes". In: *Geosphere* 17.6 (2021), pp. 2197–2233. ISSN: 1553-040X. DOI: 10.1130/ges02394.1.

[28] M. Z. Naser and Amir H. Alavi. "Error Metrics and Performance Fitness Indicators for Artificial Intelligence and Machine Learning in Engineering and Sciences". In: *Architecture, Structures and Construction* 3.4 (Dec. 2023), pp. 499–517. ISSN: 2730-9894. DOI: 10.1007/s44150-021-00015-8.

[29] H. W. Nesbitt and G. M. Young. "Early Proterozoic Climates and Plate Motions Inferred from Major Element Chemistry of Lutites". In: *Nature* 299.5885 (Oct. 1982), pp. 715–717. ISSN: 1476-4687. DOI: 10.1038/299715a0.

[30] Liqun Qi, Yannan Chen, Mayank Bakshi, and Xinzhen Zhang. *Triple Decomposition and Tensor Recovery of Third Order Tensors*. Mar. 1, 2020. DOI: 10.48550/arXiv.2002.02259. arXiv: 2002.02259 [cs, math]. URL: http://arxiv.org/abs/2002.02259 (visited on 08/01/2023). preprint.

[31] R. M. Rahl, P. W. Reiners, I. H. Campbell, S. Nicolescu, and C. M. Allen. "Combined single-grain (U-Th)/He and U/Pb dating of detrital zircons from the Navajo Sandstone, Utah". In: *Geology* 31.9 (2003), pp. 761–764. ISSN: 0091-7613.

[32] Nicholas Richardson, Naomi Graham, Michael P. Friedlander, and Joel Saylor. *MatrixTensorFactor.jl*. https://github.com/njericha/MatrixTensorFactor.jl. 2024.

[33] Nicholas Richardson, Naomi Graham, Michael P. Friedlander, and Joel Saylor. *SedimentSourceAnalysis.jl*. https://github.com/njericha/SedimentSourceAnalysis.jl. 2024.

[34] A. M. Satkoski, B. H. Wilkinson, J. Hietpas, and S. D. Samson. "Likeness among detrital zircon populations-An approach to the comparison of age frequency data in time and space". In: *Geological Society of America Bulletin* 125.11-12 (2013), pp. 1783–1799. ISSN: 0016-7606. DOI: 10.1130/b30888.1.

[35] Ville Satopaa, Jeannie Albrecht, David Irwin, and Barath Raghavan. "Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior". In: *2011 31st International Conference on Distributed Computing Systems Workshops*. 2011 31st International Conference on Distributed Computing Systems Workshops. June 2011, pp. 166–171. DOI: 10.1109/ICDCSW.2011.20.

[36] J. E. Saylor, K. E. Sundell, and G. R. Sharman. "Characterizing sediment sources by non-negative matrix factorization of detrital geochronological data". In: *Earth and Planetary Science Letters* 512 (2019), pp. 46–58. ISSN: 0012-821X. DOI: 10.1016/j.epsl.2019.01.044.

[37] J.E. Saylor, D.F. Stockli, B. K. Horton, J. Nie, and A Mora. "Discriminating rapid exhumation from syndepositional volcanism using detrital zircon double dating: Implications for the tectonic history of the Eastern Cordillera, Colombia". In: *Geological Society of America Bulletin* 124.5-6 (2012), pp. 762–779. DOI: 10.1130/B30534.1.

[38] J.E. Saylor, K.E. Sundell, and G.R. Sharman. "Characterizing Sediment Sources by Non-Negative Matrix Factorization of Detrital Geochronological Data". In: *Earth and Planetary Science Letters* 512 (Apr. 2019), pp. 46–58. ISSN: 0012821X. DOI: 10.1016/j.epsl.2019.01.044.

[39] Joel E. Saylor and Kurt E. Sundell. "Quantifying comparison of large detrital geochronology data sets". In: *Geosphere* 12 (2016), pp. 203–220. DOI: 10.1130/ges01237.1.

[40] Joel E. Saylor and Kurt E. Sundell. "Tracking Proterozoic–Triassic sediment routing to western Laurentia via bivariate non-negative matrix factorization of detrital provenance data". In: *Journal of the Geological Society* (2021). ISSN: 0016-7649. DOI: 10.1144/jgs2020-215.

[41] Joel E. Saylor, Kurt E. Sundell, Nicholas D. Perez, Jeffrey B. Hensley, Payton McCain, Brook Runyon, Paola Alvarez, José Cárdenas, Whitney P. Usnayo, and Carlos S. Valer. "Basin formation, magmatism, and exhumation document southward migrating flat-slab subduction in the central Andes". In: *Earth and Planetary Science Letters* 606 (2023), p. 118050. ISSN: 0012-821X. DOI: https://doi.org/10.1016/j.epsl.2023.118050.

[42] Glenn R. Sharman and Samuel A. Johnstone. "Sediment unmixing using detrital geochronology". In: *Earth and Planetary Science Letters* 477.Supplement C (2017), pp. 183–194. ISSN: 0012-821X. DOI: https://doi.org/10.1016/j.epsl.2017.07.044.

[43] Bernard W. Silverman. *Density Estimation for Statistics and Data Analysis*. CRC Press, Apr. 1, 1986. 190 pp. ISBN: 978-0-412-24620-3.

[44] K. N. Sircombe. "Quantitative comparison of large sets of geochronological data using multivariate analysis: A provenance study example from Australia". In: *Geochimica et Cosmochimica Acta* 64.9 (2000), pp. 1593–1616. ISSN: 0016-7037. DOI: 10.1016/s0016-7037(99)00388-9.

[45] Tyson M. Smith, Joel E. Saylor, Tom J. Lapen, Kendall Hatfield, and Kurt E. Sundell. "Identifying sources of non-unique detrital age distributions through integrated provenance analysis: An example from the Paleozoic Central Colorado Trough". In: *Geosphere* (2023). ISSN: 1553-040X. DOI: 10.1130/ges02541.1.

[46] Tyson M. Smith, Joel E. Saylor, Tom J. Lapen, Ryan J. Leary, and Kurt E. Sundell. "Large detrital zircon data set investigation and provenance mapping: Local versus regional and continental sediment sources before, during, and after Ancestral Rocky Mountain deformation". In: *GSA Bulletin* (2023). ISSN: 0016-7606. DOI: 10.1130/b36285.1.

[47] K. E. Sundell, G. E. Gehrels, M. D. Blum, J. E. Saylor, M. E. Pecha, and B. P. Hundley. "An exploratory study of "large-n" detrital zircon geochronology of the Book Cliffs, UT via rapid (3 s/analysis) U–Pb dating". In: *Basin Research* (2023). DOI: https://doi.org/10.1111/bre.12840.

[48] K. E. Sundell and J. E. Saylor. "Two-dimensional Quantitative Comparison of Density Distributions in Detrital Geochronology and Geochemistry". In: *Geochemistry, Geophysics, Geosystems* n/a.n/a (2021), e2020GC009559. ISSN: 1525-2027. DOI: https://doi.org/10.1029/2020GC009559.

[49] K.E. Sundell and J. E. Saylor. "Unmixing detrital geochronology age distributions". In: *Geophysics, Geochemistry, Geosystems* 18 (2017). DOI: 10.1002/2016GC006774.

[50] Kurt Sundell, Joel E. Saylor, and Mark Pecha. "Provenance and recycling of detrital zircons from Cenozoic Altiplano strata and the crustal evolution of western South America from combined U-Pb and Lu-Hf isotopic analysis". In: *Andean Tectonics*. Ed. by Brian K. Horton and Andrés Folguera. Elsevier, 2019, pp. 363–397. ISBN: 978-0-12-816009-1. DOI: 10.1016/B978-0-12-816009-1.00014-9.

[51] Kurt E. Sundell, Sarah W.M. George, Barbara Carrapa, George E. Gehrels, Mihai N. Ducea, Joel E. Saylor, and Martin Pepper. "Crustal Thickening of the Northern Central Andean Plateau Inferred From Trace Elements in Zircon". In: *Geophysical Research Letters* 49.3 (2022), pp. 1–9. ISSN: 1944-8007. DOI: 10.1029/2021GL096443.

[52] Ming Tang, Wei-Qiang Ji, Xu Chu, Anbin Wu, and Chen Chen. "Reconstructing crustal thickness evolution from europium anomalies in detrital zircons". In: *Geology* 49.1 (2020), pp. 76–80. ISSN: 0091-7613. DOI: 10.1130/g47745.1.

[53] P. Vermeesch, A. G. Lipp, D. Hatzenbühler, L. Caracciolo, and D. Chew. "Multidimensional Scaling of Varietal Data in Sedimentary Provenance Analysis". In: *Journal of Geophysical Research: Earth Surface* 128.3 (2023), e2022JF006992. ISSN: 2169-9011. DOI: 10.1029/2022JF006992.

[54] Pieter Vermeesch and Eduardo Garzanti. "Making geological sense of 'Big Data' in sedimentary provenance analysis". In: *Chemical Geology* 409 (2015), pp. 20–27. ISSN: 0009-2541. DOI: http://dx.doi.org/10.1016/j.chemgeo.2015.05.004.

[55] E Bruce Watson and TM Harrison. "Zircon thermometer reveals minimum melting conditions on earliest Earth". In: *Science* 308.5723 (2005), pp. 841–844. ISSN: 0036-8075.

[56] EB Watson, DA Wark, and JB Thomas. "Crystallization thermometers for zircon and rutile". In: *Contributions to Mineralogy and Petrology* 151.4 (2006), p. 413. ISSN: 0010-7999.

[57] Stanislaw Weglarczyk. "Kernel density estimation and its application". In: *ITM web of conferences*. Vol. 23. EDP Sciences. 2018, p. 00037.

[58] G. J. Weltje. "End-member modeiling of compositional data: Numerical-statistical algorithms for solving the explicit mixing problem". In: *Mathematical Geology* 29 (1997), pp. 503–549.

[59] G. J. Weltje. "Quantitative models of sediment generation and provenance: State of the art and future developments". In: *Sedimentary Geology* 280 (2012), pp. 4–20. ISSN: 0037-0738. DOI: 10.1016/j.sedgeo.2012.03.010.

[60] Gert Jan Weltje. "A quantitative approach to capturing the compositional variability of modern sands". In: *Sedimentary Geology* 171.1-4 (2004), pp. 59–77. ISSN: 0037-0738.

[61] Yangyang Xu and Wotao Yin. "A Block Coordinate Descent Method for Regularized Multiconvex Optimization with Applications to Nonnegative Tensor Factorization and Completion". In: *SIAM Journal on Imaging Sciences* 6.3 (Jan. 2013), pp. 1758–1789. ISSN: 1936-4954. DOI: 10.1137/120887795.