

# A Bayesian nonparametric approach to unmixing detrital geochronologic data

John R. Tipton · Glenn R. Sharman ·  
Samuel A. Johnstone

Received: date / Accepted: date

1 **Abstract** Sedimentary deposits constitute the primary record of changing  
2 environmental conditions that have acted on Earth's surface over geologic  
3 time. Clastic material is eroded from source locations (parents) in sediment  
4 routing systems and deposited at sink locations (children). Both parents and  
5 children have characteristics that vary across many different dimensions, in-  
6 cluding grain size, chemical composition, and the geochronologic age of con-  
7 stituent detrital minerals. During transport, sediment from different parents  
8 is mixed together to form a child, which in turn may serve as the parent for  
9 other sediment farther down system or later in time when buried sediment  
10 is exhumed. The distribution of detrital mineral ages observed in parent and  
11 child sediments allows for investigation of the proportion of each parent in the  
12 child sediment which reflects the properties of the sediment routing system. To  
13 model the proportion of dates in a child sample that comes from each of the  
14 parent distributions, we use a Bayesian mixture of Dirichlet processes. This  
15 model allows for estimation of the mixing proportions with associated uncer-  
16 tainty while making minimal assumptions. We also present an extension to the  
17 model whereby we reconstruct unobserved parent distributions from multiple  
18 observed child distributions using mixtures of Dirichlet processes. The model  
19 accounts for uncertainty in both the number of mineral formation events that  
20 compose each parent distribution and the mixing proportions of each parent  
21 distribution that composes a child distribution. To demonstrate the model,  
22 we perform analyses using simulated data where the true age distribution is

---

John R. Tipton  
Department of Mathematical Sciences, University of Arkansas, Fayetteville, AR, USA  
E-mail: jrtipton@uark.edu

Glenn R. Sharman  
Department of Geosciences, University of Arkansas, Fayetteville, AR, USA

Samuel A. Johnstone  
U.S. Geological Survey, Geosciences and Environmental Change Science Center, Denver,  
CO, USA

known as well as using a real world case study from the central California, USA coast.

## 1 Introduction

To understand the origins of modern and ancient landscapes, one must understand how erosional processes and associated sedimentary basins evolve through time (Romans et al., 2016). Clastic material is generated by weathering and erosion and subsequently transported downstream, mixed, and ultimately deposited into a depositional sink. Statistical modeling of sediment mixing allows for inference about processes that generated the modern landscape. The ability to decipher the relative proportions of sources that eroded to produce sediment informs understanding of the underlying geologic processes controlling the evolution of the Earth's surface (Stock et al., 2006; Kimbrough et al., 2015; Mason et al., 2017; Sharman et al., 2019).

One of the most common ways to characterize the provenance of sediment is detrital geochronology—dating the time at which the individual minerals that make up a sedimentary rock formed or cooled. These mineralization events typically reflect the timing of igneous rock forming events or metamorphic alteration of source rocks (Gehrels, 2014). In other cases mineral ages reflect the history of mineral cooling (e.g., ‘thermochronology,’ (Reiners and Brandon, 2006)). Detrital geochronologic ages are most commonly determined from measurements of radiogenic isotopes contained within individual mineral crystals. The decay of uranium (U) to lead (Pb) within zircon, a relatively robust mineral, makes this approach ideally suited for tracking sedimentary mixing (Amidon et al., 2005b; Sharman and Johnstone, 2017; Sundell and Saylor, 2017). While the date of a geochronometer from a particular source may be insensitive to the post-formation history of the mineral because of the high closure temperature of geochronometers (e.g., on the order of 1000 °C for zircon U-Pb (Cherniak and Watson, 2001)), the same can not be expected of low-temperature thermochronometers (e.g., at a typical cooling rate apatite U-Th/He has a closure temperature of ~60 °C (Flowers et al., 2009)). Detrital minerals with low-closure temperatures derived from a single source may have ages that were impacted by the burial of sediments (Fosdick et al., 2015) or progressive exhumation of sources (Garver et al., 1999). In addition, using multiple isotropic approaches to characterize individual minerals, such as measuring zircon chemistry in addition to age (Gehrels, 2014) or both the U-Pb and U-Th/He age of a zircon (Reiners et al., 2005), is increasingly recognized as critical for discriminating between sources with indistinct U-Pb ages. While new methods can provide increasing specificity to provenance studies, many past and ongoing studies rely on measurements of age alone; thus we focus on developing an approach to better characterize the uncertainty in inferences made from those data, recognizing that the method we present here can be extended to multiple dimensions as the need arises.

We will follow the convention that sediment sources are called *parents* and sink locations are called *children*. Because the statistical model is not defined mechanistically, the definition of the parent source is flexible enough to accommodate different scales. For example, a river could be considered the parent source sediment to a marine basin (as in the natural case study presented later in the manuscript). If rivers are considered the children, then the parent sources may be the upstream tributaries or distinctive bedrock domains within the upstream catchment. Using this language, the manuscript aims to address two questions. First, can we estimate the proportion of each parent age distribution in a child age distribution with associated uncertainty? Second, can we estimate the age distributions for unobserved parents given a set of child age distributions? These questions are answered using “top-down” and “bottom-up” approaches to sediment unmixing, respectively (see Fig. 1; (Sharman and Johnstone, 2017)). The top-down approach (Fig. 1a) models one or more child samples as a mixture of specified parent samples. The bottom-up approach (Fig. 1b) uses multiple child samples to model likely parents which are more generally referred to as end-members in mixture modeling efforts (Weltje, 1997; Paterson and Heslop, 2015).

Bayesian mixture modeling of geologic data, including detrital data, has the advantage of explicitly quantifying uncertainty (Ward et al., 2010; Cooper and Krueger, 2017; Blake et al., 2018). For geochronologic data, Bayesian methods have been used to estimate the age distribution when addressing single samples with probabilistic estimates (Jasra et al., 2006). We extend the work of Jasra et al. (2006) to multiple sample locations by modeling the geologic mixing of sediments derived from source areas containing minerals recording different crystallization events. A key feature of these data is that the age distributions are multimodal due to a range of mineral formation events for a given mineral phase. Thus, an implicit assumption is that each mode in the age distribution is the result of a distinct mineral formation event.

We demonstrate the utility of the Bayesian approach using both a synthetic dataset and a well-constrained, natural case study in central California, USA (Sickmann et al., 2016). In the simulated data, the model is shown empirically to recover the simulated age distributions. After better understanding model performance with simulated data, the model is applied to the case study dataset from central California. The top-down mixing approach is able to successfully reconstruct parent contributions in both the synthetic and natural datasets. Although the bottom-up unmixing model is able to successfully reconstruct parents in the synthetic dataset, there is evidence of non-identifiability—where parents cannot be uniquely characterized from the children—when applied to the natural dataset. Although the focus of this work is on sediment age distributions, the framework presented here can also give guidance about other scientific questions that relate to mixing of non-parametric sum-to-one data in Earth sciences and other disciplines (e.g., unmixing sediment grain size distribution; Weltje (1997) and references within).

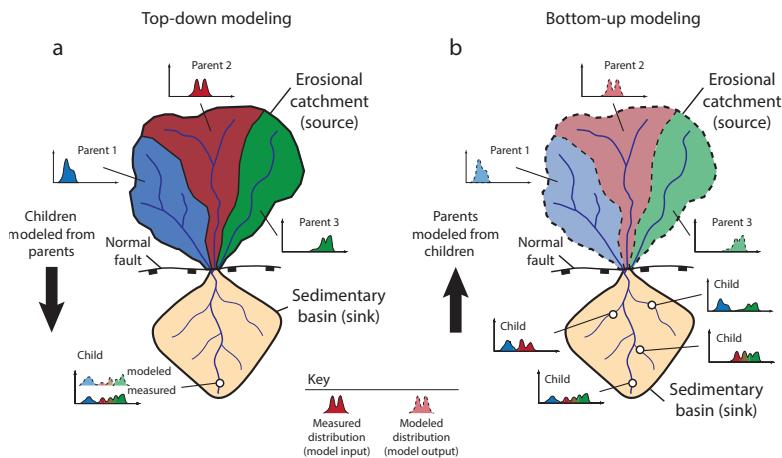


Fig. 1: Schematic depiction of a hypothetical sediment routing system with an erosional source region characterized by three parents (black, red, and green) and an associated sedimentary basin (yellow). The histograms represent the age distributions of detrital minerals from the respective parents and children with lines mapping these distributions to their respective locations in the landscape. Subfigure (a) shows the top-down modeling framework (*sensu* Sharman and Johnstone (2017)) where one or more children are modeled as a mixture of two or more parents. Subfigure (b) shows the bottom-up modeling framework where multiple children are used to reconstruct unobserved end-member sources (parents).

## 110 2 Model Overview

111 To define the statistical model, we follow the convention that letters represent  
 112 data and Greek symbols represent parameters. A plaintext symbol ( $y/\theta$ )  
 113 represents a scalar, a bold lowercase symbol ( $\mathbf{y}/\boldsymbol{\theta}$ ) represents a vector, and a  
 114 bold uppercase symbol ( $\mathbf{Z}/\boldsymbol{\Theta}$ ) is a matrix whose columns are vectors written  
 115 as  $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_p)$  or  $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p)'$  where the appropriate dimensions  
 116 are implied. We use the notation  $[y]$  to represent the probability distribu-  
 117 tion/mass function (pdf/pmf) and let  $[y|\theta]$  represent the conditional pdf/pmf  
 118 of the random variable  $y$  given  $\theta$ .

119 Following Berliner (2003), the statistical model described below is divided  
 120 into three components: the data model, the process model, and the prior  
 121 model. In general, the data model defines probability distributions that de-  
 122 scribe the variability in the data due to the measurement process. The data  
 123 model can be modified to account for non-Gaussian measurement processes  
 124 like counts, outliers, spatial/temporal correlations, etc (Hefley et al., 2017;  
 125 Tipton et al., 2017). Process models describe the best scientific understand-

ing of the process of interest. For example, process models have been used to describe the monthly response of trees to climate (Tipton et al., 2016), the relationship between climate and pollen in sediments (Tipton et al., 2019), and the movement of ice sheets in Antarctica (Chang et al., 2016; Guan et al., 2018). For the sediment mixing model, the two processes we attempt to capture with the statistical model are 1) the mineral formation events and 2) the erosional/weathering, transport, and filtering of grain ages from parent sources to children sinks. The prior model describes the range of parameter values that are plausible and completes the formal mathematical definition of the model by guaranteeing that that posterior distribution is proper (i.e., integrates to 1).

### 3 Top-down mixing model

The model framework presented below, which is appropriate for situations where the parent and children sediment have been independently characterized, will answer the first research question: can one estimate the proportion of each parent that composes the child sediment with appropriate estimates of uncertainty?

#### 3.1 Top-down mixing data model

Let  $\mathbf{y}$  be a  $n_y$ -vector of observed age measurements of a single child of interest and let  $\mathbf{z}_b$  be a  $n_b$ -vector of observed date measurements for each of the  $b = 1, \dots, B$  parents. Because the observed ages include measurement uncertainty reported as a standard deviation, we explicitly account for this source of uncertainty in the data model. In the case of U-Pb dating of detrital zircon grains, dates are most commonly determined using laser ablation-inductively coupled plasma-mass spectrometry (Gehrels, 2012). Such date measurements typically have relative  $2\sigma$  analytical precision of 1-4%, with relative uncertainty increasing for younger analyses (Puetz et al., 2018). For each detrital mineral, the analytical uncertainty (in standard deviations) for the child is reported as the  $n_y$ -vector  $\boldsymbol{\sigma}_y$  and for each of the  $b = 1, \dots, B$  parents as a  $n_b$ -vector  $\boldsymbol{\sigma}_b$ . We assume the date measurement uncertainty follows a Gaussian distribution where the observed sediment grain date is

$$\begin{aligned} \mathbf{y}|\tilde{\mathbf{y}}, \boldsymbol{\sigma}_y^2 &\sim N(\mathbf{y}|\tilde{\mathbf{y}}, \text{diag}(\boldsymbol{\sigma}_y^2)), \\ \mathbf{z}_b|\tilde{\mathbf{z}}_b, \boldsymbol{\sigma}_b^2 &\sim N(\mathbf{z}_b|\tilde{\mathbf{z}}_b, \text{diag}(\boldsymbol{\sigma}_b^2)), \end{aligned} \quad (1)$$

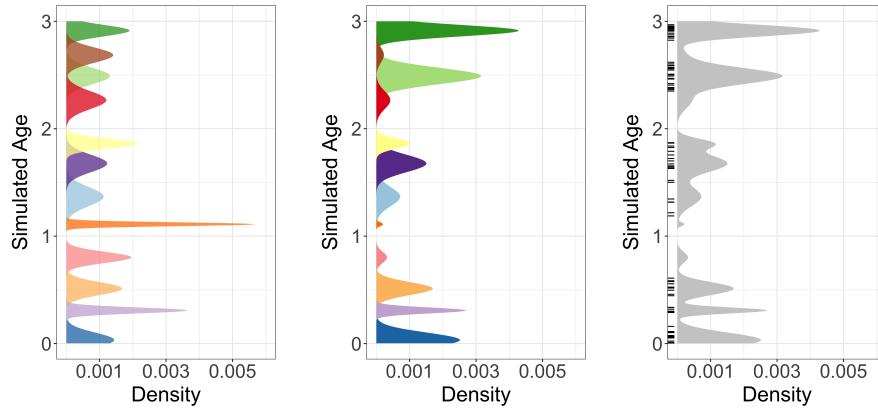
where  $N(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$  is a multivariate normal distribution with data vector  $\mathbf{x}$ , mean vector  $\boldsymbol{\mu}$ , and covariance matrix  $\boldsymbol{\Sigma}$ . The notation  $\text{diag}(\boldsymbol{\sigma}^2)$  represents a diagonal covariance matrix with  $i, i$ th element  $\sigma_i^2$  and off diagonal elements all equal to 0. We break the variable naming convention and let  $\tilde{\mathbf{y}}$  and  $\tilde{\mathbf{z}}_b$  be latent parameters that represent the true, unobserved age of the sediments where  $\mathbf{y}$

( $\mathbf{z}_b$ ) will be close to  $\tilde{\mathbf{y}}$  ( $\tilde{\mathbf{z}}_b$ ) because the measurement uncertainty is small relative to the variability in the data (i.e., the average coefficient of variation of measured dates, defined as the dating standard deviation divided by the estimated date, is about 2-3%). The reason we account for dating uncertainty is twofold. First, in a perfect world we could measure the mineral dates exactly; however, in practice our measurements introduce some uncertainty such that our data are approximations of a true, unknown age. Second, because of this accounting for analytic uncertainty, the model can give more weight to dates with less uncertainty and vice versa. In this way, it is possible within the model framework to combine data from different datasets that have different dating uncertainties in a principled manner. In addition, it is possible to account for more uncertainty in the data or to account for asymmetric measurement uncertainties using a Student's-t, log-normal, or other distribution instead of the normal distribution (Jasra et al., 2006). However, these extensions were not explored in the current work to minimize the number of parameters to estimate in this proof of concept model framework.

### 178 3.2 Top-down mixing process model

The process model addresses two scientific questions. First, what are the estimates of the true, unobserved detrital mineral age distributions at the parent and child locations? Second, what proportions of those detrital minerals did each parent source contribute to the child? Many different methods available to model geochronological age distributions from sample data, including kernel density estimation (Vermeesch, 2012), non-negative matrix factorization (Saylor et al., 2019), and Bayesian nonparametric models of mineral formation event mixing (Jasra et al., 2006; Tye et al., 2019). The following section details how our understanding of the geologic processes that generated the observed data inform the development of the statistical model.

Over geologic time, individual minerals may be repeatedly recycled into sedimentary rocks by erosion, transport, deposition, and exhumation. However, in many cases the dates recorded by individual minerals contained in these deposits are distinctive and unaffected by these recycling processes (e.g., excluding burial reheating of low temperature thermochronometers (Fosdick et al., 2015)). We assume that minerals created by the same geologic event share an age distribution that is relatively homogeneous with only small variability. Furthermore, we characterize episodes of rock and mineral formation as punctuated events (typically lasting  $10^5$  to  $10^7$  years (Chen and Moore, 1982; Irwin and Wooden, 1999; Wotzlaw et al., 2013)), which are nearly discrete events relative to the age of Earth (approximately  $4.5 \times 10^9$  years). While minerals often show overgrowths of different ages, this provides a useful approximation. Under the conceptual model (Fig. 1), sediment at each parent is formed by the decomposition of rocks containing minerals created at different times where the potential mineral creation events are shared across parents. Although all parents share the same potential mineral formation events, this



(a) Distribution of simulated mineral formation events. Each color represents a different formation event. Notice that some formation events have wider standard deviations (i.e., resulting from longer durations of mineral formation), while other formation events are shorter.

(b) The mineral formation events from Fig. 2a are re-weighted to account for relative abundance of potentially replaced by gray because observable mineral ages for the formation events are unknown. The observed data are shown as a rug plot along the y-axis.

(c) Discrete, colored formation events in the parent distribution in Fig. 2b are replaced by gray because observable mineral ages for the formation events are unknown. The observed data are shown as a rug plot along the y-axis.

Fig. 2: A cartoon of the mixing model over hypothetical mineral formation events for a single parent distribution. The y-axis of each plot is the age of formation and the x-axis is the probability density of the hypothetical parent distribution.

205 does not imply that each parent will actually contain minerals from these potential events. The parent sediments are then mixed, producing sediment that  
 206 has the potential to be present at every child location. The model assumes  
 207 the system is closed so that the sediment at each child comes entirely from  
 208 the parent sources. Therefore, each child sediment is composed of minerals  
 209 from the parent sources that are created by an unknown number of mineral  
 210 formation events at source locations.

211 A common choice for modeling a mixture of unknown distributions is the  
 212 Dirichlet process. Figure 2 demonstrates visually an example simulation from  
 213 the Dirichlet process model. Mathematically, a Dirichlet process is constructed  
 214 over an infinite dimensional mixture of base measures  $G(\theta)$  for a family of  
 215 probability measures  $G(\cdot)$  and parameters  $\theta$ . However, despite the Dirichlet  
 216 process being infinite dimensional, the expected number of clusters scales log-  
 217 arithmically with respect to sample size (Ghosal, 2010), which induces sparsity  
 218 and makes it easy to numerically approximate the infinite mixture with a finite  
 219 number of clusters  $K$  much, much less than the sample size  $n$  (Ishwaran and  
 220 James, 2001; Ishwaran and Zarepour, 2002). Thus, even though the true num-

ber of mineral formation events  $K$  recorded by a detrital sample is unknown, the *a priori* expected number of clusters can be determined using the sample size with a fixed  $K$  chosen to be much larger than the expected number of clusters to ensure a good finite approximation.

Figure 2a shows an example simulation from the Dirichlet process prior using a finite approximation with a mixture of  $K$  Gaussian distributions. Each of the shaded colors is a Gaussian distribution where each distinct color shade depicts a discrete mineral formation event. The means and variances differ for each of the distributions reflecting the variety of geologic processes that result in formation events (Fig. 2a). The transition from Fig. 2a to Fig. 2b represents the statistical process model which accounts for geological characteristics and processes including aerial extent, differential erosion, the abundance of minerals of different ages within different rocks, and other factors that influence the proportion of minerals of a given age in sediment at a site (Amidon et al., 2005a,b). The age distributions (Fig. 2a) have been reweighted to account for all of the geologic factors determining the mixing proportions of the possible formation events for the parent (Fig. 2b).

Figure 2c depicts a realization from the process model for a given parent. The labels ( $k = 1, \dots, K$ ) for each mineral formation event are not observed. Thus, the simulated observation distribution only records the age of the detrital minerals in sediment and the color shading is dropped representing this lack of knowledge about the labels. The data are not observations of the mixture density in Fig. 2c but are actually a finite sample taken from the mixture, which is shown as a rug plot where each tick on the y axis represents an observed detrital mineral grain date. Thus, the number of mineral formation events is potentially challenging to extract from the data as neither the true mineral age density nor the labels that identify the underlying formation events are known.

### 3.2.1 Modeling parent sediment ages

Let the  $i = 1, \dots, n_b$  latent detrital sediment grain ages from parent  $b$  be represented by  $\tilde{z}_{ib}$ . The sediment grain from which we estimate the latent age  $\tilde{z}_{ib}$  from the observed age  $z_{ib}$  is assumed to come from a single mineral formation event implying the mixture distribution over mineral formation events

$$\tilde{z}_{ib} | \boldsymbol{\mu}_b, \sigma_b^2, \gamma_{ib} \sim \begin{cases} N(\tilde{z}_{ib} | \mu_{b1}, \sigma_{b1}^2) & \text{if } \gamma_{ib} = 1 \\ \vdots & \vdots \\ N(\tilde{z}_{ib} | \mu_{bK}, \sigma_{bK}^2) & \text{if } \gamma_{ib} = K, \end{cases} \quad (2)$$

where for  $k = 1, \dots, K$ ,  $\mu_{bk}$  is the mixing distribution mean and  $\sigma_{bk}^2$  is the mixing distribution variance for parent  $b$ . For the  $i = 1, \dots, n_b$  observations from parents  $b = 1, \dots, B$ , the variable  $\gamma_{ib}$  is a categorical random variable whose value indicates from which formation event  $k$  the detrital mineral observation comes. We assume the probability of a detrital mineral coming from

260 formation event  $k$  is  $p_{bk} \equiv P(\gamma_{ib} = k)$ . Then, we write the joint distribution  
 261 over all mineral grains from parent  $b$  as

$$\tilde{\mathbf{z}}_b | \boldsymbol{\mu}_b, \boldsymbol{\sigma}_b^2, \gamma_b \sim \prod_{i=1}^{n_b} N(\tilde{z}_{ib} | \mu_{b1}, \sigma_{b1}^2)^{I\{\gamma_{ib}=1\}} \cdots N(\tilde{z}_{ib} | \mu_{bK}, \sigma_{bK}^2)^{I\{\gamma_{ib}=K\}} \quad (3)$$

262 where  $I\{\gamma_{ib} = k\}$  is an indicator function that takes the value 1 if  $\gamma_{ib} = k$   
 263 and 0 otherwise. Because there are a large number of indicator functions, we  
 264 integrate them out of the process model to improve mixing and model fit. The  
 265 integrated age distribution model is

$$\tilde{\mathbf{z}}_b | \boldsymbol{\mu}_b, \boldsymbol{\sigma}_b^2, \mathbf{p}_b \sim \prod_{i=1}^{n_b} \sum_{k=1}^K p_{bk} N(\tilde{z}_{ib} | \mu_{bk}, \sigma_{bk}^2) \quad (4)$$

266 where  $\mathbf{p}_b = (p_{b1}, \dots, p_{bK})'$  is a vector of positive mixing probabilities with  
 267  $\sum_{k=1}^K p_{bk} = 1$ . To account for uncertainty in the number of formation events  
 268  $K$ , the probabilities  $\mathbf{p}_b$  can be modeled using the finite approximation to the  
 269 Dirichlet process as described in detail in Sect. 3.3. In (4), the assumption is  
 270 that there could be a different set of mineral formation events for each parent.

271 Because our study site is constrained geographically, the parent and child  
 272 sites likely contain mineral grains derived from common formation events. As  
 273 such, we follow Lock and Dunson (2015) and used shared kernels by letting  
 274  $\mu_{bk} = \mu_k$  and  $\sigma_{bk}^2 = \sigma_k^2$  for all  $b = 1, \dots, B$  and  $k = 1, \dots, K$ . In theory,  
 275 there could be no overlap at all among the formation events, although in this  
 276 situation it would be possible to choose a  $K$  that is large enough such that the  
 277 number of shared kernels is larger than the total number of formation events  
 278 across all the parents and would result in equivalent inference. Doing so, the  
 279 different events at different sites are joined to form a superset of all events  
 280 common across locations with formation events potentially shared among par-  
 281 ents.

### 282 3.2.2 Modeling children sediment ages

283 The process model specifies the proportion of each parent distribution in the  
 284 child distribution. We represent the mixing proportions of the  $B$  parent dis-  
 285 tributions for the child of interest with the parameter  $\phi = (\phi_1, \dots, \phi_B)'$ , with  
 286  $\sum_{b=1}^B \phi_b = 1$  and  $\phi_b > 0$ . The parameter  $\phi_b$  is the proportion of the child  
 287 distribution that comes from parent  $b$  and accounts for differential mixing of  
 288 parents. Because the statistical model is not mechanistic, the specific, geologic  
 289 interpretation of  $\phi$  changes based on the context of the sediment transport sys-  
 290 tem. For example, when the parents are composed of bedrock,  $\phi$  is a function  
 291 of each parent's relative aerial extent in the drainage catchment, average ero-  
 292 sion rate, and average concentration of the detrital mineral of interest (Amidon  
 293 et al., 2005a). If parents are sediment inputs (e.g., rivers), then  $\phi$  is a function

294 of each parent's relative sediment supply and the average concentration of the  
 295 detrital mineral of interest within the sediment.

296 For a single latent child sediment mineral date  $\tilde{y}_i$ , the sediment grain for  
 297 that mineral comes from only one parent. Using a categorical variable, the  
 298 distribution of the child sediment grain can be written as a weighted mixture  
 299 of components (similar to (2)) where each component is a weighted sum of  
 300 the parent distributions (i.e., the sum is over the  $K$  weighted densities in (4)  
 301 but the mixture density is evaluated with child observations rather than the  
 302 parent observations). Then, the latent indicator variables can be integrated  
 303 out of the mixture, similar to (4), giving the integrated child age distribution  
 304 model

$$\tilde{\mathbf{y}}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \{\mathbf{p}_b\}_{b=1}^B, \boldsymbol{\phi} \sim \prod_{i=1}^{n_y} \sum_{b=1}^B \phi_b \sum_{k=1}^K p_{bk} N(\tilde{y}_i | \mu_k, \sigma_k^2),$$

305 where the notation  $\{\mathbf{p}_b\}_{b=1}^B$  denotes the set of parameters  $\{\mathbf{p}_1, \dots, \mathbf{p}_B\}$ .

### 306 3.3 Top-down mixing prior model

307 The conceptual process model assumes the number of mineral formation events  
 308  $K$  is known. In practice, the number of formation events is unknown and is a  
 309 parameter to be estimated. In fact, it is likely that the different parent sites  
 310 will have different numbers of mineral formation events based on site-specific  
 311 history. The prior model addresses the fundamental question of estimating the  
 312 number of mineral formation events.

313 Potential approaches to model the unknown number of formation events  
 314 vary. First, one can treat the number of formation events as a fixed parameter,  
 315 perform a grid search over the different number of formation events, and  
 316 choose the model that best fits the data based on some information theoretic  
 317 criteria (Miller and Harrison, 2018). A second approach is to sample over the  
 318 latent unknown number of formation events using a reversible jump algorithm  
 319 (Green, 1995). The third approach is to assign a Dirichlet process prior over  
 320 the number of formation events. The Dirichlet process estimates an unknown  
 321 number of components without *a priori* specifying the number.

322 The Dirichlet process is an infinite dimensional stochastic process which  
 323 is a distribution over distributions (Ferguson, 1973). We assign the range of  
 324 mineral ages for the  $k$ th formation event the base probability distribution  
 325  $G(\boldsymbol{\theta}_k)$ , which depends on parameters  $\boldsymbol{\theta}_k$ . There are many possible choices for  
 326 the base distribution  $G(\boldsymbol{\theta}_k)$ ; we assume a normal distribution  $N(\mu_k, \sigma_k^2)$  with  
 327 mean  $\mu_k$  and variance  $\sigma_k^2$ , therefore  $\boldsymbol{\theta}_k = (\mu_k, \sigma_k^2)'$ . Other possible choices  
 328 include a log-normal or gamma distribution that enforces a positive support  
 329 on the observed age dates or a skew-t distribution that allows for asymmetry  
 330 in the duration of formation events (Jasra et al., 2006). Although other  
 331 distributions are possible and may better capture the effects of natural disper-  
 332 sion in geochronometers, we rely on normal distributions here to minimize the

number of model parameters and emphasize inclusion of the multiple parent and children sediments within the bottom-up unmixing framework. Because we assume the age distribution of a single mineral formation event is relatively short with respect to the overall time of interest, the variance parameters that model the duration of the mineral formation events  $\sigma_k^2$  will be small relative to the scale of the observed age distribution. Note that the variance  $\sigma_k^2$  represents the process variance due to mineral formation events and is different than the measurement process variances,  $\sigma_y^2$  and  $\{\sigma_b^2\}_{b=1}^B$ , which are fixed and known.

While there are other methods for fitting the Dirichlet process (Richardson and Green, 1997; Ishwaran and James, 2001; Papaspiliopoulos and Roberts, 2008), we use the stick-breaking representation of a Dirichlet process (Sethuraman, 1994)

$$\sum_{k=1}^{\infty} p_{bk} G(\boldsymbol{\theta}_k), \quad (5)$$

where  $p_{bk}$  are the positive mixing weights with  $\sum_{k=1}^{\infty} p_{bk} = 1$ . In practice,  $p_{bk} \approx 0$  for large  $k$ , therefore, the infinite sum is well approximated by the finite sum  $\sum_{k=1}^K p_{bk}$  for a large enough  $K$  (Ishwaran and James, 2001; Ishwaran and Zarepour, 2002). We found in our applications  $K=20$  is sufficiently large. The stick-breaking representation for  $\mathbf{p}_b$  is constructed by transforming auxiliary variables  $\tilde{\mathbf{p}}_b = (\tilde{p}_{b1}, \dots, \tilde{p}_{bK-1})'$  using the stick-breaking representation

$$p_{bk} = \begin{cases} \tilde{p}_{b1} & \text{for } k = 1, \\ \tilde{p}_{bk} \prod_{s=1}^{k-1} (1 - \tilde{p}_{bs}) & \text{for } k = 2, \dots, K-1, \\ \prod_{s=1}^{k-1} (1 - p_{bs}) & \text{for } k = K. \end{cases}$$

Priors on the  $\tilde{p}_{bk}$  are assigned independent Beta( $1, \alpha_b$ ) priors giving rise to the stick-breaking Dirichlet process. The hyperparameters  $\alpha_b$  are given independent gamma( $1, 1$ ) priors that control the Dirichlet process concentration (i.e., smaller  $\alpha_b$  give fewer formation events, larger  $\alpha_b$  give more formation events).

The mixing kernel means  $\mu_k$  are assigned vague, independent  $N(\mu_\mu, \sigma_\mu^2)$  priors with  $\mu_\mu = 150$  million years (Myr) and  $\sigma_\mu^2 = 150^2$  Myr<sup>2</sup>. The standard deviations for the ages of formation are assigned independent truncated half-Cauchy priors  $\sigma_k \sim \text{Cauchy}^+(0, s) I\{0 < \sigma_k < \omega\}$ , where we choose  $s$  to be small relative to the range of dates observed and  $\omega$  provides an upper limit to the duration of formation events. For the case study where the majority of dates span the range of 0 to about 300 Myr, we set  $s$  to be 25 Myr and set  $\omega$  to be 50 Myr. The truncation is important to prevent the Dirichlet process mixture from generating unrealistically long formation events which does not match our *a priori* geologic knowledge.

The mixing parameter  $\phi$  is assigned a Dirichlet( $\alpha_\phi \mathbf{1}$ ) prior where  $\mathbf{1}$  is a vector of ones and the hyperparameter  $\alpha_\phi$  is assigned a gamma( $1, 1$ ) prior. When  $\alpha_\phi$  is small the mixing proportions concentrate with a large probability on a single parent component. When  $\alpha_\phi$  is one,  $\phi$  will be uniformly distributed

<sup>369</sup> over all possible mixing proportions. When  $\alpha_\phi$  is large, the mixing proportion  
<sup>370</sup> will be concentrated at equal mixing proportions ( $\frac{1}{B}, \dots, \frac{1}{B}$ ).

<sup>371</sup> 3.4 Top-down mixing posterior distribution

<sup>372</sup> The top-down mixing model posterior distribution is

$$\begin{aligned} & [\tilde{\mathbf{y}}, \{\tilde{\mathbf{z}}_b\}_{b=1}^B, \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \{\mathbf{p}_b\}_{b=1}^B, \phi, \alpha_\phi, \boldsymbol{\alpha} | \mathbf{y}, \boldsymbol{\sigma}_y, \{\mathbf{z}_b\}_{b=1}^B, \{\sigma_b^2\}_{b=1}^B] \propto \quad (6) \\ & [\mathbf{y} | \tilde{\mathbf{y}}, \boldsymbol{\sigma}_y] \prod_{b=1}^B [\mathbf{z}_b | \tilde{\mathbf{z}}_b, \boldsymbol{\sigma}_b] \times \\ & [\tilde{\mathbf{y}} | \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \{\mathbf{p}_b\}_{b=1}^B, \phi] \prod_{b=1}^B [\tilde{\mathbf{z}}_b | \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \mathbf{p}_b] \times \\ & [\boldsymbol{\mu}] [\boldsymbol{\sigma}^2] [\phi | \alpha_\phi] [\alpha_\phi] \left( \prod_{b=1}^B [\mathbf{p}_b | \alpha_b] [\alpha_b] \right), \end{aligned}$$

<sup>373</sup> where each line on the right-hand side of the proportional symbol is the data,  
<sup>374</sup> process, and prior model, respectively. We estimate the posterior using Markov  
<sup>375</sup> Chain Monte Carlo (MCMC) with the *R* package *NIMBLE* (de Valpine et al.,  
<sup>376</sup> 2017) using an adaptive block Metropolis-Hastings algorithm (Haario et al.,  
<sup>377</sup> 2001).

<sup>378</sup> 4 Bottom-up unmixing model

<sup>379</sup> The second research question is: can we reconstruct unobserved parent age  
<sup>380</sup> distributions from multiple child observations? In previous work, this anal-  
<sup>381</sup> ysis has been variably termed “end-member mixing analysis,” “end-member  
<sup>382</sup> modeling,” or “end-member analysis” as applied to unmixing grain size or  
<sup>383</sup> detrital age distributions (Sharman and Johnstone, 2017; Saylor et al., 2019).  
<sup>384</sup> The end-member unmixing model has two components. First, the number of  
<sup>385</sup> parents  $B$  is unknown and needs to be estimated. Second, given the number  
<sup>386</sup> of parents  $B$ , what are the unobserved mineral formation age distributions  
<sup>387</sup> for the  $B$  parents? For this paper, we assume the number of parents  $B$  is  
<sup>388</sup> known. A number of methods for selecting the number of parents include us-  
<sup>389</sup> ing Bayesian information criteria, reversible jump MCMC (Jasra et al., 2006),  
<sup>390</sup> assuming a Dirichlet process over the number of parents, or fitting a mixture  
<sup>391</sup> of finite mixtures (Miller and Harrison, 2018). Rather than explore these ideas,  
<sup>392</sup> we devote our effort on developing the unmixing model for a fixed number of  
<sup>393</sup> parents. The end-member model uses the same general framework presented  
<sup>394</sup> in the mixture of Gaussians model (6) with some modifications.

---

395 4.1 Bottom-up unmixing data model

396 Let  $d = 1, \dots, D$  index the  $D$  child sediments that are each composed of  $i =$   
 397  $1, \dots, n_d$  samples. As before, we assume a Gaussian measurement distribution  
 398 for child  $d$  given by

$$\mathbf{y}_d | \tilde{\mathbf{y}}_d, \boldsymbol{\sigma}_d^2 \sim N(\mathbf{y}_d | \tilde{\mathbf{y}}_d, \text{diag}(\boldsymbol{\sigma}_d^2)),$$

399 where  $\tilde{\mathbf{y}}_d$  is the true, unobserved  $n_d$ -vector of sediment dates and  $\boldsymbol{\sigma}_d$  is a fixed  
 400 and known  $n_d$ -vector of reported dating standard deviations.

401 Unlike in the top-down mixing model above, none of the parent **zs** are  
 402 observed. Hence, the parent distributions are estimated entirely using child  
 403 sediment observations. Assuming a fixed and known number of parents  $B$ , the  
 404 bottom-up process model for the  $d$ th child is

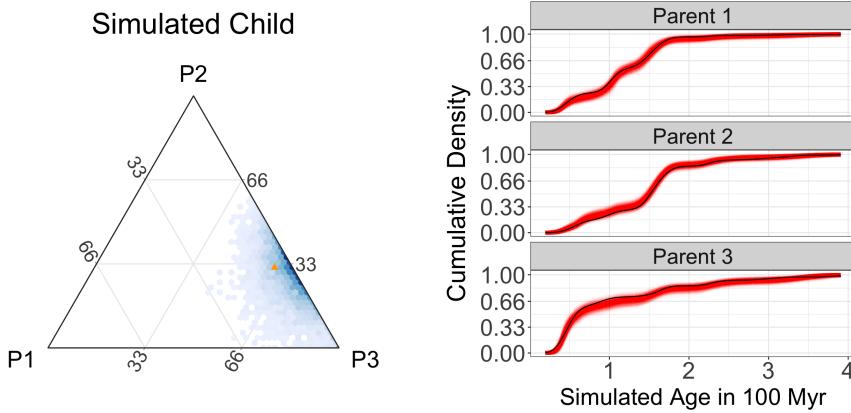
$$\tilde{\mathbf{y}}_d | \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \{\mathbf{p}_b\}_{b=1}^B, \boldsymbol{\phi}_d \sim \prod_{i=1}^{n_d} \sum_{b=1}^B \phi_{di} \sum_{k=1}^K p_{bk} N(\tilde{y}_{id} | \mu_k, \sigma_k^2), \quad (7)$$

405 where, like before, we assume a Gaussian mixing distribution using shared  
 406 kernels across the  $B$  parents. The  $B$ -dimensional mixture proportion  $\boldsymbol{\phi}_d =$   
 407  $(\phi_{d1}, \dots, \phi_{dB})'$  models the proportion of the  $d$ th child sediment that can be  
 408 attributed to each of the  $B$  parents. For each of the  $d = 1, \dots, D$  children,  
 409 the mixing proportions  $\boldsymbol{\phi}_d$  are assigned independent  $Dirichlet(\alpha_d \mathbf{1})$  prior  
 410 where  $\mathbf{1}$  is a vector of ones of length  $B$  and each  $\alpha_d$  is assigned a  $gamma(1, 1)$   
 411 prior. The priors for  $\boldsymbol{\mu}$ ,  $\boldsymbol{\sigma}^2$ , and  $\{\mathbf{p}_b\}_{b=1}^B$  (and their respective hyperparameters  
 412  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_B)'$ ) are the same as in the top-down mixing model.

413 Like the top-down mixing model, these bottom-up unmixing equations can  
 414 be derived by introducing categorical random variables for every observation  
 415 and then marginalizing out the latent indicator variables from the model.  
 416 Thus, conditional on the values of the parameters, the  $b$ th unknown parent age  
 417 distribution for the top-down unmixing model is given by  $\sum_{k=1}^K p_{bk} N(\mu_k, \sigma_k^2)$ .

418 4.2 Bottom-up unmixing posterior distribution

419 The posterior distribution that we estimate with the end member unmixing  
 420 model is



(a) Ternary plot showing posterior mixing proportion estimates shaded relative to posterior density in blue and the simulated true mixing proportions as an orange triangle.

(b) Plot of simulated parents with fitted posterior CDF estimates in red and the simulated true CDF in black. Each red line represents a posterior sample of the cumulative parent age density.

Fig. 3: Results for the top-down mixture modeling approach based on simulated data. As can be seen in the figures, the top-down mixture model is performing well in estimating the simulated mixing distributions.

$$\begin{aligned} & [\{\tilde{\mathbf{y}}_d\}_{d=1}^D, \boldsymbol{\mu}, \sigma^2, \{\mathbf{p}_b\}_{b=1}^B, \{\phi_d\}_{d=1}^D, \boldsymbol{\alpha}_\phi, \boldsymbol{\alpha} | \{\mathbf{y}_d\}_{d=1}^D, \{\sigma_d^2\}_{d=1}^D] \propto \quad (8) \\ & \prod_{d=1}^D [\mathbf{y}_d | \tilde{\mathbf{y}}_d, \sigma_d^2] \times \\ & \prod_{d=1}^D [\tilde{\mathbf{y}}_d | \boldsymbol{\mu}, \sigma^2, \phi_d, \{\mathbf{p}_b\}_{b=1}^B] \times \\ & [\boldsymbol{\mu}] [\sigma^2] \left( \prod_{b=1}^B [\mathbf{p}_b | \alpha_b] [\alpha_b] \right) \left( \prod_{d=1}^D [\phi_d | \alpha_d] [\alpha_d] \right), \end{aligned}$$

where the priors are the same as those in (6) except that there are now  $D$  children which implies there are now  $D$   $\alpha_d$ s. Like the top-down mixing model, the bottom-up unmixing model is estimated using MCMC with the *R* package *NIMBLE* (de Valpine et al., 2017).

## 425 5 Simulation of synthetic detrital age distributions

426 We explore the performance of the model using a synthetic detrital age dis-  
427 tribution dataset. The aim of the simulation study is to understand how the

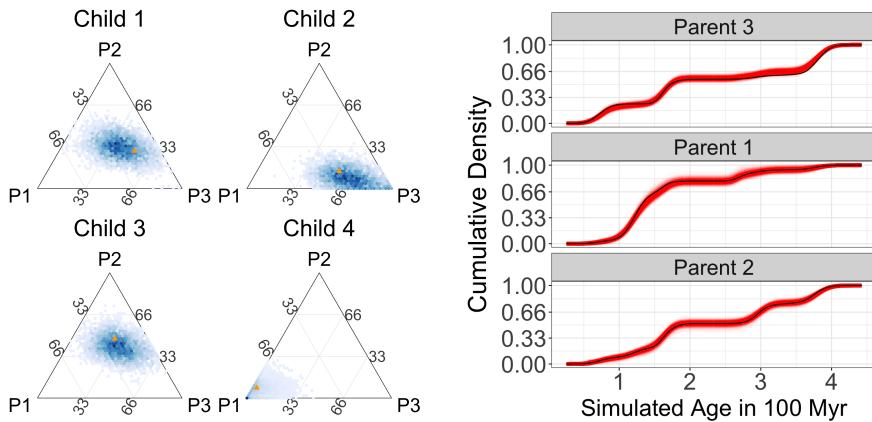
428 model performs using realistic data and verify the model is capable of recovering  
429 the simulated parameters. The simulation study framework can also be used to understand how uncertainty in estimation varies with respect to sample  
430 size, variability in the data, and consequences of prior assumptions, although these details are not explored in this work (Vehtari et al., 2017). For example,  
431 by simulating data with smaller sample sizes than observed, the impact on the uncertainty estimates can be explored and perhaps used to guide sample size  
432 recommendations for the collection of new data. Similarly, simulations with different observation errors can be used to quantify the amount of analytic  
433 precision needed for an inferential question analogous to how sample size and power calculations can be used in experimental design.

434 For the simulation study, a synthetic dataset is created using the top-down  
435 mixing model in (6) for  $B = 3$  parents and a single child (Fig. 1a). The parent  
436 distributions were composed of 200, 250, and 150 simulated sediment grains,  
437 respectively, and the child distribution was composed of 150 sediment grains.  
438 In simulation, we used dating uncertainties ( $\sigma_y, \sigma_z$ ) that were about 1-3%  
439 of the total range of the simulated age distribution of 0-400 Myr. These are  
440 similar to measurement uncertainties in the case study and demonstrate the  
441 model is capable of accounting for measurement uncertainty.

442 The posterior samples for the mixing proportion  $\phi$  are shown in Fig. 3a in  
443 hexagonal bins with blue shading proportional to the posterior density (Hamilton,  
444 2018), and the simulated mixing proportion is represented by the orange  
445 triangle. The simulated mixing proportion (orange triangle) lies within the  
446 region of high posterior density demonstrating that the model is accurately  
447 estimating the mixing proportions. Figure 3b shows the estimated cumulative  
448 distribution functions (CDFs) with posterior samples in red and the simulated  
449 CDF in black. The results in Fig. 3 demonstrate that the model is accurately  
450 estimating the simulated mixing proportions  $\phi$  as well as the parent age distri-  
451 butions, validating the effectiveness of the top-down mixing model to recover  
452 simulated parameters of interest.

453 The second simulation generated data from the bottom-up, end-member  
454 unmixing model (Fig. 1b) to test how well the proposed framework can re-  
455 construct unobserved parent distributions from a set of child observations.  
456 For the simulation, we used  $B = 3$  parents and  $D = 20$  children where each  
457 child consisted of 250 measured sediment grains following the model in (8).  
458 The range of simulated mixing proportions was simulated uniformly over the  
459 three-dimensional simplex resulting in some end members being close to pure  
460 end members (e.g., close to 90% of sediment grains coming from a single par-  
461 ent). The dating uncertainties ( $\sigma_y$ ) were set at about 1-3% of the total range  
462 of the simulated age distribution of 0-400 Myr.

463 Figure 4a shows the posterior samples for the mixing proportion of four  
464 representative child samples in hexagonal bins with shading in blue propor-  
465 tional to posterior density and the orange triangle at the simulated mixing  
466 proportion. Based on the simulation study, the model can recover the mixing  
467 proportions in this simulation example with high accuracy because the or-  
468 ange triangle is within regions of high posterior mass. Even though the model



(a) Posterior estimates of mixing proportions for 4 of the 20 children from the unmixing model shown. The blue shading is relative to posterior density and the simulated true mixing proportions are shown as orange triangles.

(b) Posterior estimates of the unobserved parent CDFs in red. The simulated parent CDF is shown in black.

Fig. 4: Results for the bottom-up, end-member unmixing model using simulated data. The bottom-up unmixing model does a good job of estimating the true, unobserved parent age CDFs despite the model not using any simulated parent data.

474 uses none of the data from the parents, the bottom-up unmixing model pro-  
 475 duces reasonable end-member parent age distribution estimates. Figure 4b  
 476 shows the estimated CDF produced by the bottom-up unmixing model, which  
 477 shows the model is estimating the unobserved parent distributions. However,  
 478 the precision for the bottom-up unmixing model is lower in the bottom-up  
 479 unmixing simulation relative to the the top-down mixture simulation despite  
 480 the bottom-up umixing simulated data containing many more sediment grain  
 481 samples.

## 482 6 Application to a Natural Case Study

483 We apply the top-down mixing and bottom-up unmixing models to a well-  
 484 constrained modern dataset from the central California coast (Sickmann et al.,  
 485 2016) shown in Fig. 5. Following the same mixing framework presented in Shar-  
 486 man and Johnstone (2017), five samples (river and beach sediment) are used  
 487 to characterize three distinct sediment inputs (parents) to the region, each  
 488 with a distinct detrital age distribution. Parents 1 and 2 (P1 and P2) are  
 489 composed of river samples (CAR and SAL) that represent sediment sources  
 490 along the Big Sur coastline and Salinas River drainage, respectively. Parent 3  
 491 (P3) is composed of two river samples (SNR and PAR) and one beach sample

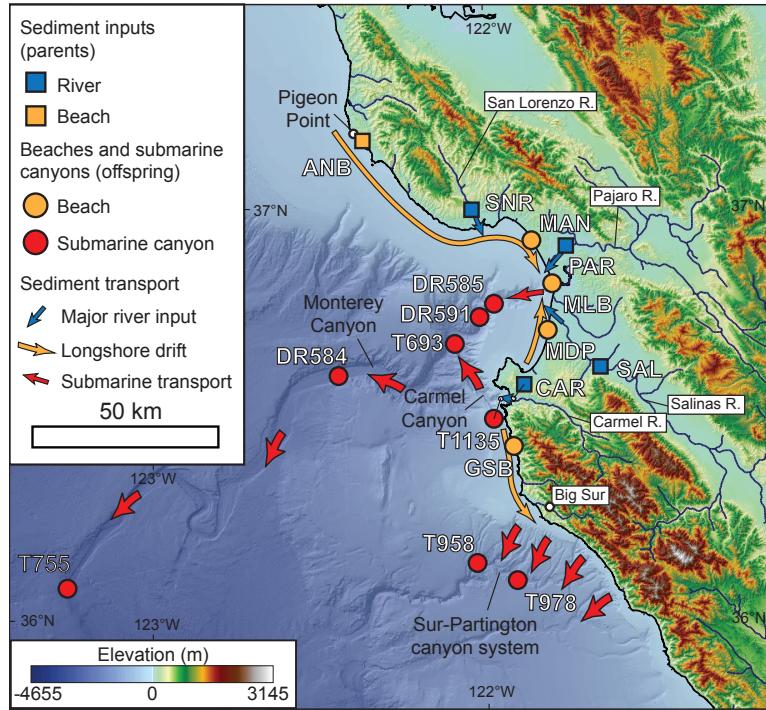


Fig. 5: Locations of the parent and children data for the study region in California, USA (Sharman and Johnstone, 2017).

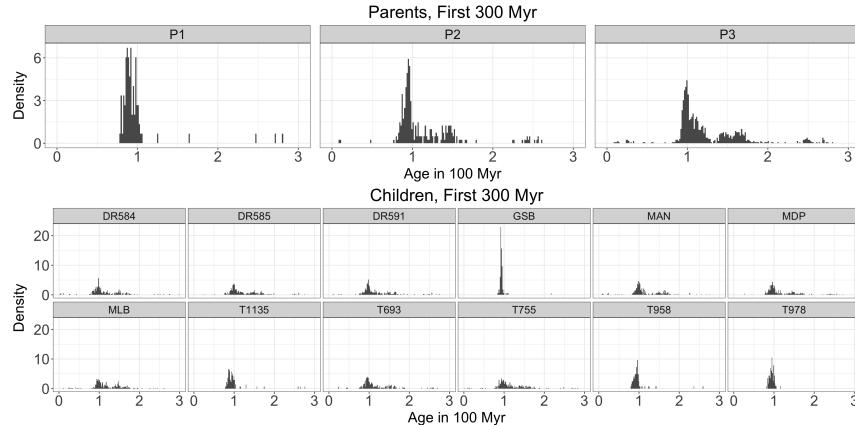
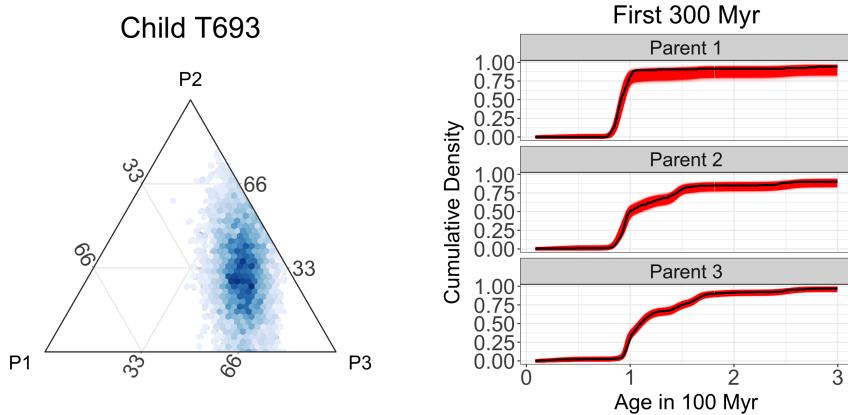


Fig. 6: The sediment age data for the study region in California, USA, used for the mixing and unmixing models (Sickmann et al., 2016). The three parent age distributions are shown in the top plot and the 12 child age distributions are shown in the bottom plot. The  $x$ -axes represent the measured age in 100 Myr and the  $y$ -axes show the empirical density.



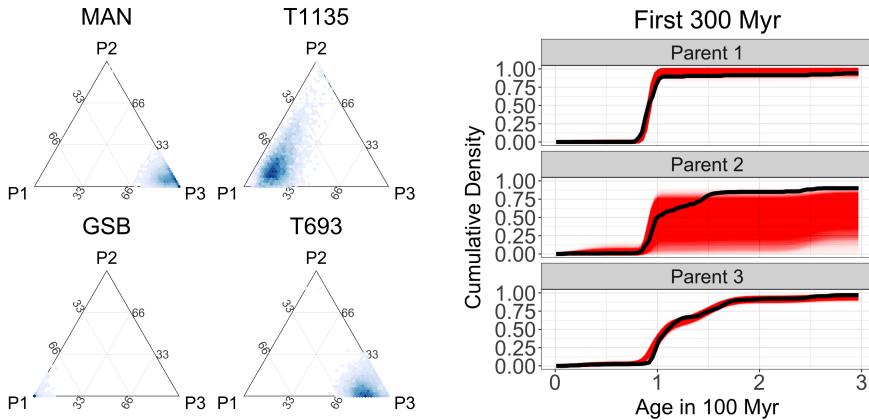
(a) Ternary plot showing posterior density estimates of mixing proportions. The relative posterior density is shown in blue shading.

(b) Posterior estimates of the parent and child CDFs shown in red. The empirical CDFs calculated from the raw data are shown in black.

Fig. 7: Results from the top-down mixing model using data from the study region in California, USA. The model results, when applied to sample T693, demonstrate that the top-down mixing model is able to accurately reconstruct the parent and child distributions and produce estimates of the mixing proportions with associated uncertainty.

(ANB) that represent northern sediment sources in the Santa Cruz Mountains and western Diablo Range (Sickmann et al., 2016; Sharman and Johnstone, 2017). Twelve child samples (beach and submarine canyon sediment) are used to characterize how these parents are mixed in littoral and marine environments. In total, this dataset (Fig. 6) consists of 4,026 individual detrital zircon U-Pb analyses, with individual samples having 82 to 316 analyses each (median of 290 analyses per sample) (Sickmann et al., 2016). The age range of the case study data covered approximately 80–3000 Myr; however, there were very few observations older than 300 Myr (approximately 6% of the data), so even though the model was fit to the entire age range, the presentation of results focuses on the period 0–300 Myr to better interpret the results.

We first examine the top-down mixture model (Fig. 1a). Figure 7a shows the reconstruction of the mixing proportions for a sample from a submarine canyon (T693) modeled as a mixture of the three specified parent distributions (P1–P3). Visual inspection of the histograms of the data (Fig. 6) would suggest that this child sample is a mixture composed mostly of P3 with some contribution from P2. The posterior estimates of the mixing proportion of each parent for child T693 confirms that the primary component of the mixture is from the P3 with parent P2 as the secondary component (Fig. 7a). Figure 7b shows the model is capturing the basic patterns in the parent CDFs as the estimated CDFs are very close to the empirical CDFs for the parents.



(a) Posterior estimates for the mixing proportions of each parent for four child sediments. Notice that without observing the parents, the posterior distribution of mixing proportions for child T693 is generally similar to the top-down mixing model in Fig. 7a but has a slightly different shape.

(b) Posterior estimates for the unobserved parent cumulative distribution functions shown in red over 0–300 Myr. The black lines show the empirical cumulative distribution functions.

Fig. 8: Results of the end-member unmixing model fit to data from the study region in California, USA. These figures show that the end-member unmixing model is estimating the parameters of interest, but with some inaccuracies due to a lack of identifiability. However, these issues are easily identified by the end user due to the large amount of posterior uncertainty which provides a check on overly strong inferential claims.

513 The bottom-up, end-member unmixing model results for the case study  
 514 data are shown in Fig. 8. The posterior density for the mixing proportions  
 515 are shown for four children in Fig. 8a as hexagonal bins with shading in blue  
 516 proportional to posterior density. The posterior distribution of CDFs for the  
 517 reconstructed parents in Fig. 8b show that the model is very uncertain about  
 518 the age distribution for parent P2. This is not a totally unsurprising result  
 519 because the age distributions for parent P2 and parent P3 are very similar  
 520 and thus are difficult to tease apart in an unsupervised learning situation like  
 521 that in the bottom-up unmixing model. In Fig. 8a, the mixture probability for  
 522 child T693 is more concentrated at P3 when compared to the top-down mixing  
 523 proportion estimates in Fig. 7a. Because the model expresses uncertainty about  
 524 the distribution of parent P2, the difference in estimation between the top-  
 525 down and bottom-up mixing proportions for child T693 can be attributed  
 526 to this lack of identifiability among parents P2 and P3. Similar to how the  
 527 important variables in a linear regression can change with small perturbations  
 528 of the data when the covariates are highly co-linear (i.e., in models with a  
 529 high variance inflation factor), the attribution of the child sediment to a given  
 530 parent would change which parent distribution (P2 or P3) is estimated with

accuracy and which parent shows evidence of non-identifiability. Thus, the non-identifiability manifests as large posterior uncertainty in the CDF for parent P2 and provides a caution about being overly confident in the inference about the bottom-up unmixing model.

A large overlap in the distribution of parent ages is a feature that often occurs in detrital zircon geochronology studies. The preservation of zircons through multiple cycles of erosion and re-sedimentation means that overlapping zircon ages will be present in many rocks. For parent age distributions that are quite similar to one another, the reconstruction of the unknown parent distributions suffers from weak identifiability. In these situations, the estimated parents jointly contain all of the formation events, but the model is unable to attribute the formation events to the correct parents. In other words, while the model identifies the correct age components, the model sometimes struggles to correctly group these components into the correct parent distributions. This aliasing effect is not an unexpected result because Bayesian nonparametric models are well understood to suffer from non-identifiability issues (Ferguson, 1983; Diebolt and Robert, 1994; Richardson and Green, 1997; Frühwirth-Schnatter, 2006). However, the non-identifiability is not a result of the common label switching phenomenon (Stephens, 2000) that is common in mixing model as the multiple children in the bottom-up unmixing model provides a soft constraint on the label switching behavior.

Non-identifiability is inherent in all end-member unmixing models (Weltje and Prins, 2007). To overcome the non-identifiability in other modeling frameworks, a potential solution is to impose constraints on the end-members and/or provide informative initial conditions for maximum likelihood optimization algorithms (Donoho and Stodden, 2004; Miao and Qi, 2007; Chen and Guillaume, 2012). Therefore, any end-member unmixing model that uses only child age distributions will have issues in accurately reconstructing the parent distributions if the assumption of the constraints is not met (i.e., the parent age distributions are structurally similar). Bottom-up unmixing models provide a useful way to explore large detrital datasets with unknown sedimentary sources. Our proposed model framework provides a way to identify those datasets that either are or are not susceptible to non-identifiability by producing uncertainty estimates that are larger when the model is weakly identifiable (Fig. 8b, parent 2). Thus, the uncertainty intervals are a useful diagnostic check for identifiability.

Direct, probabilistic estimates of uncertainty and the ability to calculate derived quantities with uncertainty is a benefit of the proposed method and of Bayesian methods in general. Thus, we can answer questions like what is the probability that at least 50% of child sediment T693 comes from parent P3 using the top-down mixing model applied to the case study data? The answer is calculated directly from the posterior samples using the Monte Carlo approximation  $\frac{1}{L} \sum_{\ell=1}^L I\{\phi_3^{(\ell)} \geq 0.5\} = 0.672$ , where  $\ell = 1, \dots, L$  are the indices of the MCMC samples and  $\phi_3^{(\ell)}$  is the estimated mixing proportion for the  $\ell$ th MCMC iteration. The probability that at least 50% of the child

576 sediment comes from parent P3 and at least 25% comes from parent P2 is  
 577  $\frac{1}{L} \sum_{\ell=1}^L I\{\phi_{P3}^{(\ell)} \geq 0.5\} \times I\{\phi_{P2}^{(\ell)} \geq 0.25\} = 0.283$ . Another question of interest  
 578 that can be expressed in term of the model is “How many mineral formation  
 579 events occurred for each parent?” We assumed that there was an upper bound  
 580 of  $K$  possible mineral formation events for each parent distribution modeled by  
 581 the  $K$ -dimensional probability vector  $\mathbf{p}_b$  for each of the  $B$  parents. For a given  
 582 parent  $b$  and a fixed threshold  $\tau$  where we conclude that a potential formation  
 583 event is realized (say  $\tau = 0.01$  or  $\tau = 0.05$ ), the estimated number of formation  
 584 events is  $\frac{1}{L} \sum_{\ell=1}^L \sum_{k=1}^K I\{p_{bk}^{(\ell)} \geq \tau\}$ . For example, the posterior mean estimate  
 585 for the number of formation events for parent P1 under the top-down mixing  
 586 model is 4.37 (95% CI 4-6) when the threshold  $\tau = 0.05$  and 8.34 (95% CI  
 587 7-10) when  $\tau = 0.01$ . Because the model produces a posterior probability, any  
 588 other such probabilistic questions like those above can be calculated as derived  
 589 quantities. For example, we can ask questions like: what proportion of a given  
 590 sample contains grains older than a given age? or what is the probability that  
 591 an unobserved parent contains grains within a particular age range. Once the  
 592 posterior samples have been calculated, any questions that can be evaluated  
 593 using derived quantities can be answered probabilistically.

594 In addition, the ability to include prior information in the Bayesian frame-  
 595 work is a useful tool that can be used to improve estimation and test geologic  
 596 hypotheses. For example, certain geologic events, such as the Grenville orogeny,  
 597 produced large amounts of zircon that have since been broadly dispersed and  
 598 recycled in sedimentary rocks (Moecher and Samson, 2006). Priors that ac-  
 599 count for the likelihood of observing zircons of Grenville-age (or other known  
 600 zircon-producing events) can be introduced into this model framework to im-  
 601 prove performance. In addition, our framework can accommodate a variety of  
 602 detrital data with different magnitudes of uncertainty. As analytic techniques  
 603 for dating minerals improve, it is important to account for dating uncertainties  
 604 that might vary widely, making our method more robust to future improve-  
 605 ments in analytic laboratory techniques.

## 606 7 Conclusion

607 Starting from a conceptual model of how sediments mix over a landscape,  
 608 we developed a generative Bayesian nonparametric statistical model for detri-  
 609 tal mineral age data. This model allows us to characterize the uncertainty in  
 610 the age distributions of parents and children and the mixing proportions for  
 611 sediments while explicitly accounting for the uncertainties in measured dates  
 612 (Jasra et al., 2006; Tye et al., 2019). Because the model can generate sedi-  
 613 ment age distributions, we can directly explore the assumptions of the model  
 614 by simulating synthetic data. Running a simulation experiment demonstrated  
 615 the model is capable of recovering simulated distributions which supports the  
 616 usefulness of the framework when applied to observed data.

617 We proposed two frameworks to model the sediment mixing mechanisms:  
 618 the top-down mixing model where mineral dates are measured for both par-

ent and child sediments and a bottom-up unmixing framework where mineral dates are only measured for the children. The top-down model estimated the parent and child distributions and the mixing proportions with high precision and accuracy. The bottom-up model occasionally showed evidence of non-identifiability in the simulation experiments and showed non-identifiability in the case study, suggesting the inference for the bottom-up model is less precise than for the top-down mixing model. Because the variances of these estimates are larger in our bottom-up unmixing model, the user is provided with feedback about the potential pitfalls in being overly confident about the reconstructed parent distributions.

Obtaining correct inference is vitally important for any statistical model. However, many models make identifying when inference is suspect challenging. The explicit modeling of uncertainty presented in this manuscript provides a check on overly confident inference. As such, the inference in the model presented herein provides a useful diagnostic on the quality of model fit. In the case study, the posterior distribution for one of the parent distributions was estimated with a large amount of uncertainty. Therefore, the inference based on the model fit is less reliable than the model that includes the parent data. Thus, explicit modeling of uncertainty is critical in providing information about the quality of the inference.

While this manuscript focuses on estimating the mixture of geochronological measurements from sediments, the methods discussed can be applied to mixtures of any univariate variable of interest. For example, the top-down mixing and bottom-up unmixing models can be applied to mixtures of sediment grain size (Weltje and Prins, 2007). In addition to applying the model frameworks to other variables, the extension of the mixing and unmixing models to multivariate data would allow for the inclusion of more nuanced and detailed data. In the cases where the distributions are only weakly identified based on one variable, the mixing distributions might be identifiable using other variables. Thus, the results presented in this manuscript can provide a roadmap for future development and extension to better characterize geologic landscapes.

## 650 8 Declarations

651 The authors declare that they have no known competing financial interests or  
652 personal relationships that could have appeared to influence the work reported  
653 in this paper. Support for SAJ came from the FEDMAP component of the  
654 U.S. Geological Survey National Cooperative Geologic Mapping Program. Any  
655 use of trade, firm, or product names is for descriptive purposes only and does  
656 not imply endorsement by the U.S. Government.

657 Code and data for replication of results presented in this manuscript can be  
658 found freely available under the permissive MIT license on GitHub at <https://github.com/jtipton25/mixing-manuscript>.

660 **Keywords** Detrital sediment age distributions · Sediment unmixing ·  
661 Bayesian nonparametrics · Uncertainty quantification

---

**662 References**

- 663 Amidon WH, Burbank DW, Gehrels GE (2005a) Construction of detrital mineral populations: Insights from mixing of U-Pb zircon ages in Himalayan  
664 rivers. *Basin Research* 17(4):463–485
- 665 Amidon WH, Burbank DW, Gehrels GE (2005b) U–Pb zircon ages as a sediment mixing tracer in the Nepal Himalaya. *Earth and Planetary Science Letters* 235(1-2):244–260
- 666 Berliner LM (2003) Physical-statistical modeling in geophysics. *Journal of Geophysical Research: Atmospheres* 108(D24)
- 667 Blake WH, Boeckx P, Stock BC, Smith HG, Bodé S, Upadhyay HR, Gaspar L, Goddard R, Lennard AT, Lizaga I, et al. (2018) A deconvolutional  
668 Bayesian mixing model approach for river basin sediment source apportionment. *Scientific Reports* 8(1):1–12
- 669 Chang W, Haran M, Applegate P, Pollard D (2016) Calibrating an ice sheet  
670 model using high-dimensional binary spatial data. *Journal of the American Statistical Association* 111(513):57–72
- 671 Chen JH, Moore JG (1982) Uranium-lead isotopic ages from the Sierra  
672 Nevada batholith, California. *Journal of Geophysical Research: Solid Earth*  
673 87(B6):4761–4784
- 674 Chen W, Guillaume M (2012) HALS-based NMF with flexible constraints for  
675 hyperspectral unmixing. *EURASIP Journal on Advances in Signal Processing* 2012(1):54
- 676 Cherniak D, Watson E (2001) Pb diffusion in zircon. *Chemical Geology* 172(1-  
677 2):5–24
- 678 Cooper RJ, Krueger T (2017) An extended Bayesian sediment fingerprinting  
679 mixing model for the full bayes treatment of geochemical uncertainties.  
680 *Hydrological Processes* 31(10):1900–1912
- 681 de Valpine P, Turek D, Paciorek C, Anderson-Bergman C, Temple Lang D,  
682 Bodik R (2017) Programming with models: Writing statistical algorithms  
683 for general model structures with NIMBLE. *Journal of Computational and  
684 Graphical Statistics* 26:403–413, DOI 10.1080/10618600.2016.1172487
- 685 Diebolt J, Robert CP (1994) Estimation of finite mixture distributions  
686 through Bayesian sampling. *Journal of the Royal Statistical Society Series B (Methodological)* pp 363–375
- 687 Donoho D, Stodden V (2004) When does non-negative matrix factorization  
688 give a correct decomposition into parts? In: *Advances in Neural Information  
689 Processing Systems*, pp 1141–1148
- 690 Ferguson TS (1973) A Bayesian analysis of some nonparametric problems. *The  
691 Annals of Statistics* pp 209–230
- 692 Ferguson TS (1983) Bayesian density estimation by mixtures of normal distri-  
693 butions. In: *Recent Advances in Statistics*, Elsevier, pp 287–302
- 694 Flowers RM, Ketcham RA, Shuster DL, Farley KA (2009) Apatite (U–Th)/He  
695 thermochronometry using a radiation damage accumulation and annealing  
696 model. *Geochimica et Cosmochimica Acta* 73(8):2347–2365

- 706 Fosdick JC, Grove M, Graham SA, Hourigan JK, Lovera O, Romans BW  
707 (2015) Detrital thermochronologic record of burial heating and sediment re-  
708 cycling in the Magallanes foreland basin, Patagonian Andes. *Basin Research*  
709 27(4):546–572
- 710 Frühwirth-Schnatter S (2006) Finite mixture and Markov switching models.  
711 Springer Science & Business Media
- 712 Garver JI, Brandon MT, Roden-Tice M, Kamp PJ (1999) Exhumation history  
713 of orogenic highlands determined by detrital fission-track thermochronology.  
714 Geological Society, London, Special Publications 154(1):283–304
- 715 Gehrels G (2012) Detrital zircon U-Pb geochronology: Current methods and  
716 new opportunities. *Tectonics of Sedimentary Basins: Recent Advances* pp  
717 45–62
- 718 Gehrels G (2014) Detrital zircon U-Pb geochronology applied to tectonics.  
719 *Annual Review of Earth and Planetary Sciences* 42:127–149
- 720 Ghosal S (2010) The Dirichlet process, related priors and posterior asymptotic  
721 *Bayesian Nonparametrics* 28:35
- 722 Green PJ (1995) Reversible jump Markov chain Monte Carlo computation and  
723 Bayesian model determination. *Biometrika* 82(4):711–732
- 724 Guan Y, Haran M, Pollard D (2018) Inferring ice thickness from a glacier dy-  
725 namics model and multiple surface data sets. *Environmetrics* 29(5-6):e2460
- 726 Haario H, Saksman E, Tamminen J (2001) An adaptive Metropolis algorithm.  
727 *Bernoulli* 7(2):223–242
- 728 Hamilton N (2018) ggtern: An extension to ggplot2, for the creation of ternary  
729 diagrams. URL <https://CRAN.R-project.org/package=ggtern>, R pack-  
730 age version 2.2.2
- 731 Hefley TJ, Brost BM, Hooten MB (2017) Bias correction of bounded location  
732 errors in presence-only data. *Methods in Ecology and Evolution* 8(11):1566–  
733 1573
- 734 Irwin WP, Wooden JL (1999) Plutons and accretionary episodes of the Kla-  
735 math Mountains, California and Oregon. US Geological Survey Open-File  
736 Report 99-374 URL <https://pubs.usgs.gov/of/1999/0374/>
- 737 Ishwaran H, James LF (2001) Gibbs sampling methods for stick-breaking pri-  
738 ors. *Journal of the American Statistical Association* 96(453):161–173
- 739 Ishwaran H, Zarepour M (2002) Exact and approximate sum representations  
740 for the Dirichlet process. *Canadian Journal of Statistics* 30(2):269–283
- 741 Jasra A, Stephens DA, Gallagher K, Holmes CC (2006) Bayesian mixture  
742 modelling in geochronology via Markov chain Monte Carlo. *Mathematical  
743 Geology* 38(3):269–300
- 744 Kimbrough DL, Grove M, Gehrels GE, Dorsey RJ, Howard KA, Lovera O,  
745 Aslan A, House PK, Pearthree PA (2015) Detrital zircon U-Pb provenance  
746 of the Colorado River: A 5 my record of incision into cover strata overlying  
747 the Colorado Plateau and adjacent regions. *Geosphere* 11(6):1719–1748
- 748 Lock EF, Dunson DB (2015) Shared kernel Bayesian screening. *Biometrika*  
749 102(4):829–842
- 750 Mason CC, Fildani A, Gerber T, Blum MD, Clark JD, Dykstra M (2017)  
751 Climatic and anthropogenic influences on sediment mixing in the Mississ-

- 752 sippi source-to-sink system using detrital zircons: Late Pleistocene to recent.  
753 Earth and Planetary Science Letters 466:70–79
- 754 Miao L, Qi H (2007) Endmember extraction from highly mixed data using min-  
755 imum volume constrained nonnegative matrix factorization. IEEE Transac-  
756 tions on Geoscience and Remote Sensing 45(3):765–777
- 757 Miller JW, Harrison MT (2018) Mixture models with a prior on the number of  
758 components. Journal of the American Statistical Association 113(521):340–  
759 356
- 760 Moecher DP, Samson SD (2006) Differential zircon fertility of source terranes  
761 and natural bias in the detrital zircon record: Implications for sedimentary  
762 provenance analysis. Earth and Planetary Science Letters 247(3-4):252–266
- 763 Papaspiliopoulos O, Roberts GO (2008) Retrospective Markov chain Monte  
764 Carlo methods for Dirichlet process hierarchical models. Biometrika  
765 95(1):169–186
- 766 Paterson GA, Heslop D (2015) New methods for unmixing sediment grain size  
767 data. Geochemistry, Geophysics, Geosystems 16(12):4494–4506
- 768 Puetz SJ, Ganade CE, Zimmermann U, Borchardt G (2018) Statistical anal-  
769 yses of global U-Pb database 2017. Geoscience Frontiers 9(1):121–145
- 770 Reiners PW, Brandon MT (2006) Using thermochronology to understand oro-  
771 genic erosion. Annual Review Earth Planetary Sciences 34:419–466
- 772 Reiners PW, Campbell I, Niculescu S, Allen CM, Hourigan J, Garver J, Mat-  
773 tinson J, Cowan D (2005) (U-Th)/(He-Pb) double dating of detrital zircons.  
774 American Journal of Science 305(4):259–311
- 775 Richardson S, Green PJ (1997) On Bayesian analysis of mixtures with an  
776 unknown number of components (with discussion). Journal of the Royal  
777 Statistical Society: Series B (Statistical Methodology) 59(4):731–792
- 778 Romans BW, Castelltort S, Covault JA, Fildani A, Walsh J (2016) Environ-  
779 mental signal propagation in sedimentary systems across timescales. Earth-  
780 Science Reviews 153:7–29
- 781 Saylor JE, Sundell K, Sharman G (2019) Characterizing sediment sources by  
782 non-negative matrix factorization of detrital geochronological data. Earth  
783 and Planetary Science Letters 512:46–58
- 784 Sethuraman J (1994) A constructive definition of dirichlet priors. Statistica  
785 sinica pp 639–650
- 786 Sharman GR, Johnstone SA (2017) Sediment unmixing using detrital  
787 geochronology. Earth and Planetary Science Letters 477:183–194
- 788 Sharman GR, Sylvester Z, Covault JA (2019) Conversion of tectonic and cli-  
789 matic forcings into records of sediment supply and provenance. Scientific  
790 Reports 9(1):4115
- 791 Sickmann ZT, Paull CK, Graham SA (2016) Detrital-zircon mixing and parti-  
792 tioning in fluvial to deep marine systems, central California, USA. Journal  
793 of Sedimentary Research 86(11):1298–1307
- 794 Stephens M (2000) Dealing with label switching in mixture models. Journal of  
795 the Royal Statistical Society: Series B (Statistical Methodology) 62(4):795–  
796 809

- 797 Stock GM, Ehlers TA, Farley KA (2006) Where does sediment come from?  
798 Quantifying catchment erosion with detrital apatite (U-Th)/He ther-  
799 mochronometry. *Geology* 34(9):725–728
- 800 Sundell KE, Saylor JE (2017) Unmixing detrital geochronology age distribu-  
801 tions. *Geochemistry, Geophysics, Geosystems* 18(8):2872–2886
- 802 Tipton J, Hooten M, Goring S (2017) Reconstruction of spatio-temporal tem-  
803 perature from sparse historical records using robust probabilistic principal  
804 component regression. *Advances in Statistical Climatology, Meteorology and*  
805 *Oceanography* 3(1):1–16
- 806 Tipton JR, Hooten MB, Pederson N, Tingley MP, Bishop D (2016) Recon-  
807 struction of late Holocene climate based on tree growth and mechanistic  
808 hierarchical models. *Environmetrics* 27(1):42–54
- 809 Tipton JR, Hooten MB, Nolan C, Booth RK, McLachlan J (2019) Predicting  
810 paleoclimate from compositional data using multivariate Gaussian process  
811 inverse prediction. *Annals of Applied Statistics* 13(4):2363–2388, DOI 10.  
812 1214/19-AOAS1281
- 813 Tye A, Wolf A, Niemi N (2019) Bayesian population correlation: A probabilis-  
814 tic approach to inferring and comparing population distributions for detrital  
815 zircon ages. *Chemical Geology* 518:67–78
- 816 Vehtari A, Gelman A, Gabry J (2017) Practical Bayesian model evaluation  
817 using leave-one-out cross-validation and WAIC. *Statistics and Computing*  
818 27:1413–1432, DOI 10.1007/s11222-016-9696-4
- 819 Vermeesch P (2012) On the visualisation of detrital age distributions. *Chemical*  
820 *Geology* 312:190–194
- 821 Ward EJ, Semmens BX, Schindler DE (2010) Including source uncertainty  
822 and prior information in the analysis of stable isotope mixing models. *En-  
823 vironmental Science & Technology* 44(12):4645–4650
- 824 Weltje GJ (1997) End-member modeling of compositional data: Numerical-  
825 statistical algorithms for solving the explicit mixing problem. *Mathematical*  
826 *Geology* 29(4):503–549
- 827 Weltje GJ, Prins MA (2007) Genetically meaningful decomposition of grain-  
828 size distributions. *Sedimentary Geology* 202(3):409–424
- 829 Wotzlaw JF, Schaltegger U, Frick DA, Dungan MA, Gerdes A, Günther D  
830 (2013) Tracking the evolution of large-volume silicic magma reservoirs from  
831 assembly to supereruption. *Geology* 41(8):867–870

---

**832 A Appendix**

833 In the body of the text, we explain in detail how the scientific understanding of the geological  
 834 processes that generated the data led to the development of the statistical model. Here in  
 835 the appendix, we define the statistical models directly without the scientific motivation for  
 836 clarity of model definition.

837 **A.1 Top-down mixing model**

In the top-down mixing model, observations are made on both the child sediments and parent sediments. The  $n_y$  observed age measurements for the child sediment are given by the vector  $\mathbf{y} = (y_1, \dots, y_{n_y})'$  and are reported with an observed analytic measurement standard deviation  $\sigma_y = (\sigma_{y1}, \dots, \sigma_{yn_y})'$  associated with each observation. Assuming a normal distribution for the measurement process, the latent true ages are defined as  $\tilde{\mathbf{y}} = (\tilde{y}_1, \dots, \tilde{y}_{n_y})'$  and are modeled by

$$\mathbf{y}|\tilde{\mathbf{y}}, \sigma_y^2 \sim N(\mathbf{y}|\tilde{\mathbf{y}}, \text{diag}(\sigma_y^2)),$$

838 where  $N(\mathbf{x}|\mu, \Sigma)$  is a multivariate normal distribution with data vector  $\mathbf{x}$ , mean vector  $\mu$ ,  
 839 and covariance matrix  $\Sigma$ . The notation  $\text{diag}(\sigma^2)$  represents a diagonal covariance matrix  
 840 with  $i, i$ th element  $\sigma_i^2$  and off diagonal elements all equal to 0.

Likewise, the  $b = 1, \dots, B$  parent observations are each comprised of  $n_b$  observations and are given by the vector  $\mathbf{z}_b = (z_{b1}, \dots, z_{bn_b})'$  and are reported with an observed analytic measurement standard deviation  $\sigma_b = (\sigma_{b1}, \dots, \sigma_{bn_b})'$  associated with each observation. Assuming a normal distribution for the measurement process, the latent true ages are defined as  $\tilde{\mathbf{z}}_b = (\tilde{z}_{b1}, \dots, \tilde{z}_{bn_b})'$  and are modeled by

$$\mathbf{z}_b|\tilde{\mathbf{z}}_b, \sigma_b^2 \sim N(\mathbf{z}_b|\tilde{\mathbf{z}}_b, \text{diag}(\sigma_b^2))$$

The latent parent age distributions for the  $b = 1, \dots, B$  parents are modeled using a finite mixture of  $K$  Gaussian distributions

$$\tilde{\mathbf{z}}_b|\mu, \sigma^2, p_b \sim \prod_{i=1}^{n_b} \sum_{k=1}^K p_{bk} N(\tilde{z}_{ib}|\mu_k, \sigma_k^2)$$

841 where  $\mu = (\mu_1, \dots, \mu_K)'$  and  $\sigma^2 = (\sigma_1^2, \dots, \sigma_K^2)'$  are the mean and variance of the mixture  
 842 distributions (which are shared across each of the parents) and  $\mathbf{p}_b = (p_{b1}, \dots, p_{bK})'$  are  
 843 mixture weights where for  $k = 1, \dots, K$ ,  $p_{bk} > 0$  and  $\sum_{k=1}^K p_{bk} = 1$ . For  $k = 1, \dots, K$ , the  
 844 mixture distribution means are assigned independent, vague priors  $N(\mu_k = 150 \text{ Myr}, \sigma_k^2 =$   
 845  $150^2 \text{ Myr}^2)$  where Myr represents a million years. To ensure the mixing distributions are rel-  
 846 atively concentrated with respect to geologic time, the mixing kernel standard deviations are  
 847 assigned independent truncated half-Cauchy priors  $\sigma_k \sim \text{Cauchy}^+(0, 25 \text{ Myr}) I\{0 < \sigma_k <$   
 848  $50 \text{ Myr}\}$ , which enforces the mixing distribution scales (which represent geologic mineral  
 849 formation events) to be small relative to the range of dates from about 0 to 300 Myr.

For each parent  $b = 1, \dots, B$ , the mixing probabilities  $\mathbf{p}_b$  are modeled by introducing  
 $k = 1, \dots, K-1$  independent and identically distributed random variables  $\tilde{p}_{bk} \sim \text{Beta}(1, \alpha_b)$   
 random variables and transforming the  $\tilde{p}_{bk}$ s by

$$p_{bk} = \begin{cases} \tilde{p}_{b1} & \text{for } k = 1, \\ \tilde{p}_{bk} \prod_{s=1}^{k-1} (1 - \tilde{p}_{bs}) & \text{for } k = 2, \dots, K-1, \\ \prod_{s=1}^{k-1} (1 - \tilde{p}_{bs}) & \text{for } k = K. \end{cases}$$

850 which induces a finite approximation to the stick-breaking representation of a Dirichlet  
 851 process so long as  $K$  is chosen large enough (Section 3 Ishwaran and James (2001) and  
 852 Ishwaran and Zarepour (2002)). For  $b = 1, \dots, B$ ,  $\alpha_b$  is assigned a  $\text{gamma}(1, 1)$  prior and  
 853 the vector  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_B)'$ .

854 Combining the parent distributions, the unobserved, latent ages are modeled using the  
 855 finite mixture of mixtures

$$\tilde{\mathbf{y}}|\boldsymbol{\mu}, \sigma^2, \{\mathbf{p}_b\}_{b=1}^B, \phi \sim \prod_{i=1}^{n_y} \sum_{b=1}^B \phi_b \sum_{k=1}^K p_{bk} N(\tilde{y}_i|\mu_b, \sigma_b^2),$$

856 where the notation  $\{\mathbf{p}_b\}_{b=1}^B$  denotes the set of parameters  $\{\mathbf{p}_1, \dots, \mathbf{p}_B\}$ . The parameter  
 857  $\phi = (\phi_1, \dots, \phi_B)'$  models the proportion of the child sediment  $\phi_b$  that comes from parent  
 858  $b$  where  $\phi_b > 0$  and  $\sum_{b=1}^B \phi_b = 1$ . The mixing proportion  $\phi$  is assigned a  $Dirichlet(\alpha_\phi \mathbf{1})$   
 859 prior where  $\mathbf{1}$  is a vector of ones of length  $B$  and  $\alpha_\phi$  is assigned a  $gamma(1, 1)$  prior.  
 860 All combined, the top-down mixing model posterior is

$$\begin{aligned} \left[ \tilde{\mathbf{y}}, \{\tilde{\mathbf{z}}_b\}_{b=1}^B, \boldsymbol{\mu}, \sigma^2, \{\mathbf{p}_b\}_{b=1}^B, \phi, \alpha_\phi, \alpha \middle| \mathbf{y}, \sigma_y^2, \{\mathbf{z}_b\}_{b=1}^B, \{\sigma_b^2\}_{b=1}^B \right] \propto \\ [\mathbf{y}|\tilde{\mathbf{y}}, \sigma_y^2] \prod_{b=1}^B [\mathbf{z}_b|\tilde{\mathbf{z}}_b, \sigma_b^2] \times \\ \left[ \tilde{\mathbf{y}} \middle| \boldsymbol{\mu}, \sigma^2, \{\mathbf{p}_b\}_{b=1}^B, \phi \right] \prod_{b=1}^B [\tilde{\mathbf{z}}_b|\boldsymbol{\mu}, \sigma^2, \mathbf{p}_b] \times \\ [\boldsymbol{\mu}] [\sigma^2] [\phi|\alpha_\phi] [\alpha_\phi] \left( \prod_{b=1}^B [\mathbf{p}_b|\alpha_b] [\alpha_b] \right), \end{aligned}$$

861 where each line on the right-hand side of the proportional symbol is the data, process, and  
 862 prior model, respectively.

## 863 A.2 Bottom-up unmixing model

In the bottom-up unmixing model, observations are made on  $d = 1, \dots, D$  child sediments whereas the parent sediments are unobserved. For each of the  $d = 1, \dots, D$  children, the  $n_d$  observed age measurements are given by the vector  $\mathbf{y}_d = (y_{d1}, \dots, y_{dn_d})'$  and are reported with an observed analytic measurement standard deviation  $\sigma_d = (\sigma_{d1}, \dots, \sigma_{dn_d})'$  associated with each observation. Assuming a normal distribution for the measurement process, the latent true ages are defined as  $\tilde{\mathbf{y}}_d = (\tilde{y}_{d1}, \dots, \tilde{y}_{dn_d})'$  and are modeled by

$$\mathbf{y}_d|\tilde{\mathbf{y}}_d, \sigma_d^2 \sim N(\mathbf{y}_d|\tilde{\mathbf{y}}_d, \text{diag}(\sigma_d^2)),$$

As none of the parent ages are observed, the latent parent age distributions for the  $b = 1, \dots, B$  parents are represented as a finite mixture of  $K$  Gaussian distributions

$$\sum_{k=1}^K p_{bk} N(\mu_k, \sigma_k^2)$$

864 where  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_K)'$  and  $\sigma^2 = (\sigma_1^2, \dots, \sigma_K^2)'$  are the mean and variance of the mixture  
 865 distributions (which are shared across each of the parents) and  $\mathbf{p}_b = (p_{b1}, \dots, p_{bK})'$  are  
 866 mixture weights where for  $k = 1, \dots, K$ ,  $p_{bk} > 0$  and  $\sum_{k=1}^K p_{bk} = 1$ . For  $k = 1, \dots, K$ , the  
 867 mixture distribution means are assigned independent, vague priors  $N(\mu_k = 150 \text{ Myr}, \sigma_k^2 =$   
 868  $150^2 \text{ Myr}^2)$  where Myr represents a million years. To ensure the mixing distributions are rel-  
 869 atively concentrated with respect to geologic time, the mixing kernel standard deviations are  
 870 assigned independent truncated half-Cauchy priors  $\sigma_k \sim \text{Cauchy}^+(0, 25 \text{ Myr}) I\{0 < \sigma_k <$   
 871  $50 \text{ Myr}\}$ , which enforces the mixing distribution scales (which represent geologic mineral  
 872 formation events) to be small relative to the range of dates from about 0 to 300 Myr.

Because none of the parent ages are observed, the parent distributions are estimated entirely using child sediment observations. Assuming a fixed and known number of parents  $B$ , the bottom-up process model for the  $d$ th child is

$$\tilde{\mathbf{y}}_d | \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \{\mathbf{p}_b\}_{b=1}^B, \boldsymbol{\phi}_d \sim \prod_{i=1}^{n_d} \sum_{b=1}^B \phi_{db} \sum_{k=1}^K p_{bk} N(\tilde{y}_{id} | \mu_k, \sigma_k^2),$$

where the  $B$ -dimensional vector of mixture proportions  $\boldsymbol{\phi}_d = (\phi_{d1}, \dots, \phi_{dB})'$  models the proportion of the  $d$ th child sediment that can be attributed to each of the  $B$  parents where for  $b = 1, \dots, B$ ,  $\phi_{db} > 0$  and  $\sum_{b=1}^B \phi_{db} = 1$ . For each of the  $d = 1, \dots, D$  children, the mixing proportions  $\boldsymbol{\phi}_d$  are assigned independent  $Dirichlet(\alpha_d \mathbf{1})$  prior where  $\mathbf{1}$  is a vector of ones of length  $B$  and each  $\alpha_d$  is assigned a  $gamma(1, 1)$  prior. The priors for  $\boldsymbol{\mu}$ ,  $\boldsymbol{\sigma}^2$ , and  $\{\mathbf{p}_b\}_{b=1}^B$  (and their respective hyperparameters  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_B)'$ ) are the same as in the top-down mixing model.

Thus, the bottom-up unmixing model posterior distribution is

$$\begin{aligned} \left[ \{\tilde{\mathbf{y}}_d\}_{d=1}^D, \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \{\mathbf{p}_b\}_{b=1}^B, \{\boldsymbol{\phi}_d\}_{d=1}^D, \boldsymbol{\alpha}_\phi, \boldsymbol{\alpha} \middle| \{\mathbf{y}_d\}_{d=1}^D, \{\sigma_d^2\}_{d=1}^D \right] \propto \\ \prod_{d=1}^D [\mathbf{y}_d | \tilde{\mathbf{y}}_d, \sigma_d^2] \times \\ \prod_{d=1}^D [\tilde{\mathbf{y}}_d | \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{\phi}_d, \{\mathbf{p}_b\}_{b=1}^B] \times \\ [\boldsymbol{\mu}] [\boldsymbol{\sigma}^2] \left( \prod_{b=1}^B [\mathbf{p}_b | \alpha_b] [\alpha_b] \right) \left( \prod_{d=1}^D [\boldsymbol{\phi}_d | \alpha_d] [\alpha_d] \right), \end{aligned}$$

where each line on the right-hand side of the proportional symbol is the data, process, and prior model, respectively.