## How many grains are needed for quantifying catchment erosion from tracer thermochronology?

Andrea Madella<sup>1</sup>, Christoph Glotzbach<sup>1</sup>, Todd A. Ehlers<sup>1</sup>

<sup>1</sup>Department of Geosciences, University of Tuebingen, Schnarrenbergstr. 94-96, 72076 Tuebingen, Germany

5 Correspondence to: Andrea Madella (andrea.madella@uni-tuebingen.de)

Abstract. Detrital tracer thermochronology <u>utilizes</u> the relationship between bedrock thermochronometric age-elevation profiles and a distribution of detrital grain-ages collected from river, glacial, or other sediment to study spatial <u>variations</u> in the distribution of catchment erosion. If <u>bedrock</u> ages increase linearly with elevation, spatially uniform erosion is expected to yield a detrital age distribution that <u>mimics</u> the <u>shape of a</u> catchment's hypsometric curve. Alternatively, a mismatch between

- 10 detrital and hypsometric distributions may indicate <u>spatial variability of sediment production</u> within the source area. For studies / seeking to identify the pattern of <u>sediment production</u>, detrital samples rarely exceed 100 grains due to the time and costs / related to individual measurements. With sample sizes of this order, <u>detecting the effect size</u> between two detrital age / distributions produced by different catchment erosion scenarios can be difficult at a high statistical confidence level. However, there is no established tool to quantify the uncertainty inherent to detrital tracer thermochronology as a function of sample size
- 15 and spatial pattern of sediment production. As a result, practitioners are often left wondering 'how many grains is enough to detect a certain signal?'. Here, we investigate how sample size affects the uncertainty of detrital age distributions and how such uncertainty affects the ability to infer a pattern of sediment production of the upstream area. We do this using the Kolmogorov-Smirnov statistic as a metric of dissimilarity among distributions. From this, we perform statistical hypothesis testing by means of Monte Carlo sampling. These techniques are implemented in a new tool (ESD thermotrace) to: (i).
- 20 consistently report the confidence level allowed by the sample size, as a function of application-specific variables, and given a set of user-defined hypothetical erosion scenarios; (ii) analyse the statistical power to discern each scenario from the uniform erosion hypothesis; and (iii) identify the erosion scenario that is least dissimilar to the observed detrital sample (if available). ESD thermotrace, is made available as a new open-source Python-based script along with test data. Testing between different hypothesized erosion scenarios with this tool provides thermochronologists with the minimum sample size (i.e. number of
- 25 bedrock and detrital grain-ages) required to answer their specific scientific question, at their desired level of statistical confidence.

#### **1** Introduction

Tracer thermochronology uses the distribution of single grain thermochronometric ages from detritus to infer the spatial pattern of erosion in the source area (e.g. Stock et al., 2006; Vermeesch, 2007). This approach is typically applied where bedrock

Formatted	( [1]
Deleted: exploits	
Formatted	( [2]
Deleted: changes	
Formatted	( [3]
Deleted: mirrors	
Formatted	( [4]
Deleted: non-uniform erosion	
Formatted	( [5]
Deleted: a catchment.	
Formatted	( [6]
Deleted: erosion, measured grain-age population	ons
Formatted	( [7]
Deleted: largely	
Formatted	( [8]
Deleted: discerning	
Formatted	( [9]
Deleted: method	
Formatted	( [10]
Deleted: sample-size-dependent	
Formatted	( [11]
Deleted: , and	
Formatted	( [12]
Deleted: uniquely	
Formatted	( [13]
Deleted: the erosional	
Formatted	( [14]
Deleted: , based on which the	
Formatted	( [15]
Deleted: confidence of detecting an erosional p	attern is det [16]
Deleted: The	
Formatted	( [17]
Formatted	( [18]
Deleted: levels	
Formatted	( [19]
Deleted: sample size and	
Formatted	( [20]
Deleted: . The proposed tool	
Formatted	( [21]
Deleted: Furthermore, in cases of unavoidably	small samp [22]
Formatted	( [23]
Deleted:grain thermochronometric ages of	( [24]
Formatted	( [25]

thermochronometric age data exhibit a clear age-elevation relationship, allowing inference of the relative contribution of source elevations from the detrital grain-age distribution. A detrital grain-age distribution that closely follows the catchment's

- 85 hypsometric curve (i.e., the cumulative distribution function of elevation area), is generally interpreted as indicative for spatially uniform erosion. Conversely, a detrital age distribution skewed towards younger (or older) ages may be the consequence of focused erosion at lower (or higher) elevations (Brewer et al., 2003). Tracer thermochronology has been shown to be a powerful tool to investigate the sub-catchment-scale variability of denudation. Geomorphologists have been able to infer changes in climatic parameters (Nibourel et al., 2015; Riebe et al., 2015), glacial erosional processes (Ehlers et al., 2015;
- 90 Enkelmann and Ehlers, 2015; Clinger et al., 2020), sediment dynamics (Lang et al., 2018), relief evolution (McPhillips and Brandon, 2010; Whipp et al., 2009), occurrence of mass-wasting (Vermeesch, 2007; Whipp and Ehlers, 2019) and differences in rock uplift (McPhillips and Brandon, 2010; Glotzbach et al., 2013, 2018; Brewer et al., 2003; Ruhl and Hodges, 2005). Other work has noted that neglecting mineral fertility variations with catchment lithologies may challenge the conclusions of some of these studies (Malusà et al., 2016). Unfortunately, the number of measured detrital ages for tracer thermochronology
- 95 is often dictated by inherent limitations of the sampled material and/or by available finances, rather than a science-based choice. Detrital sample sizes often range between 40-120 grains (e.g. Stock et al., 2006; Vermeesch, 2007; McPhillips and Brandon, 2010; Ehlers et al., 2015; Riebe et al., 2015; Glotzbach et al., 2018; Lang et al., 2018; Clinger et al., 2020), and are considered to yield high-confidence results when surpassing ~100 grains based on previous work on sediment provenance analysis (Vermeesch, 2004), However, in unfortunate cases, two measured distributions generated from different erosional
- 100 patterns <u>cannot</u> be statistically <u>discerned</u> at a high confidence level even with more than 100 grains. Although this issue is well-known to the community (e.g. Avdeev et al., 2011) since the early days of such detrital studies <u>(Brewer et al., 2003)</u>, there is no established <u>tool</u> to quantify the uncertainty inherent to detrital tracer thermochronology as a function of sample size and <u>upstream pattern of sediment production. Moreover</u>, the number of measured grains may often be based on convenience and/or habit.

105

Here, we complement previous work by investigating how sample size affects the uncertainty of detrital cooling age distributions and the related confidence in addressing the pattern of sediment production in the upstream area. We discuss the approaches used in previous case studies, upon which we develop a tool (Earth System Dynamics - *ESD thermotrace*) to consistently report confidence levels as a function of sample size and case-specific variables. We illustrate our approach using

110 a published dataset from the Sierra Nevada Mountains, California (Stock et al., 2006). The proposed tool is made available as a new open-source Python-based script along with test data. We demonstrate how ESD\_thermotrace can assist future tracer thermochronology studies in defining the necessary sample size to answer their specific scientific question. In cases where larger sample sizes are impossible to achieve, the <u>statistical power of a tracer thermochronology analysis can be studied using</u> our script.

Delet	ed:	mirrors

Deleted:

Deleted: (Spiegel et al., 2004).

#### Formatted: English (US)

**Deleted:** (Clinger et al., 2020; Ehlers et al., 2015; Enkelmann and Ehlers, 2015), sediment dynamics (Lang et al., 2018)

Deleted: (McPhillips and Brandon, 2010)

Deleted: (Vermeesch, 2007; Whipp and Ehlers, 2019)

Deleted: (Glotzbach et al., 2013, 2018; McPhillips and Brandon, 2010)....

Deleted: (Clinger et al., 2020; Ehlers et al., 2015; Glotzbach et al., 2018; Lang et al., 2018; McPhillips and Brandon, 2010; Riebe et al., 2015; e.g. Stock et al., 2006; Vermeesch, 2007)

(	Formatted: English (US)
~(	Deleted: discerning between
(	Deleted: can
(	Deleted: impossible
(	Formatted: English (US)
Y	Deleted: (Brewer et al., 2003)
X	Deleted: method
$\langle \rangle$	Deleted: sample-size-dependent
) (	Deleted: ,
Ì	Deleted: chosen
(	Deleted: erosional
(	Deleted: based on

Deleted:	study's confidence level can be quantified
Formatte	d: English (US)

#### 2 Background information

Single grain detrital age distributions are extensively applied in classical detrital geochronology studies (Hurford and Carter, 1991), where U/Pb crystallization ages of zircon constitute by far the most used tool (Spiegel et al., 2004; Andersen, 2005; Malusà et al., 2013). In this type of application, the aim is to obtain the spectrum of all age components (i.e. age peaks) that characterize a siliciclastic sediment. If a range of assumptions hold (Malusà et al., 2013; Malusà and Fitzgerald, 2020), the provenance of a sediment sample's source area can be inferred by matching the detrital age components to those of known

#### Formatted: English (US)

Deleted: (and possibly overused) tool (Andersen, 2005; Malusà et al., 2013; Spiegel et al., 2004). Deleted: (Malusà et al., 2013)

A. Classical Detrital B. Tracer Geochronology Thermochronology 100 Ma source area SO Ma detritus detrital age distribution source elevation [m] 1000 2000 frequency requency 100 50 young old source age [Ma] source age

Figure 1 - Sketch of the difference between classical detrital geochronology (A) and tracer thermochronology (B). A: discrete age components are found in the detritus and refer to different upstream geological units. B: a continuous detrital age distribution informs the relative abundance of material sourced from different elevations, based on a known ageelevation relationship.

upstream geological units and/or events. For that purpose, the number of measured detrital grains determines the confidence of detecting minor/small age components. An exhaustive probabilistic method to

- 150 report such confidence exists (Vermeesch, 2004) and is not the focus of this study. The absolute age components of the source area are in fact unimportant in detrital tracer thermochronology (Avdeev et al., 2011), for which monolithological catchments are best suited, in order to minimize mineral fertility issues in the source rock (Fig.1). The focus here is the dissimilarity between the distribution of ages found in the
- source area and in the fluvial/glacial/hillslope sediment derived therefrom, regardless of their absolute age components. For this purpose, the uncertainty caused by a small sample size strongly limits the least significant dissimilarity that can be resolved between two distributions. 160 This minimum resolution directly affects the power of our inferences.

In the following we summarize the conceptual model concerning this matter and the approaches that have been used to address it thus far. Let us consider a monolithological catchment and a set of detrital grain-ages

165 measured at its outlet. The observed grain-age distribution should match a predicted distribution that is forward-modelled stacking the following layers of geologic information about the upstream area (Fig.2):

170

175

180

- I. The catchment hypsometry, or the distribution of the catchment's cumulative area as a function of elevation. The hypsometric curve is constructed based on the digital elevation model (DEM) of the study area, and it has a negligible uncertainty.
- II. The bedrock age-elevation data, <u>In the simplest case</u>, a linear function based on a dataset of ages with associated uncertainty, measured at known elevations.
- III. The mineral fertility, which informs the original abundance of the target mineral in the different elevation ranges and is mostly a function of lithologic variability. This is a critical parameter that can lead to gross misinterpretations if ignored (Malusà et al., 2016),

A. Catchment Data Age Fertility Erosion -old low Deleted: minoritarian Formatted: English (US) young -high Formatted: English (US) Deleted: avoid Deleted: is Deleted: from Deleted: component Deleted: minimum Deleted: about the upstream erosional pattern <object>commonly the ultimate scope of tracer thermochronology Deleted: <object> uniform



Figure 2 – Qualitative sketch to illustrate the effect of mineral fertility and erosion on the detrital distribution. A: The catchment of Fig.1B with known bedrock age (shades of green) is subject to 3 scenarios of spatially varying fertility and erosion. The box outlines refer to the curves below. B: detrital distributions obtained from the different scenarios in above. The green curve refers to spatially uniform erosion and fertility.

- IV. Information on how the sediment-size distribution varies between the headwaters and the detritus. This is to make sure that grains in the detritus are representative of erosion in the catchment (e.g. Vermeesch, 2007; Riebe et al., 2015; Lukens et al., 2020).
- V A pattern of erosion across the catchment, which is the dependent variable of the system. This is any map of erosion probability, a scenario to be tested against uniform erosion, (or other reference scenarios).

Depending on the scope of the application, tracer thermochronology ultimately aims to quantify the mismatch between observed and predicted distributions, where the predicted distributions vary depending on assumed models for catchment erosion, sediment dynamics and tectonics.

#### 3 Comparing predicted and observed distributions

- 200 Predicted distributions should be constructed accounting for all above information and related uncertainties, such that the confidence of the fit to the observed distribution also accounts for them. Brewer et al. (2003), and successively Ruhl and Hodges (2005), were the first to compare the distribution of detrital thermochronometers to that of age-elevation data. Although the scope of their work differed from more recent tracer thermochronology, the evaluation of the dissimilarity between predicted/observed distributions remains the main object of their studies. These authors constructed synoptic probability 205 density functions (SPDF) of the observed data by "stacking" the gaussian distributions of all measured grain-ages, each with their analytical error (this is equivalent to the SPDF<sub>t</sub> in Table 1 of Vermeesch (2007)). In addition to the observed distribution, a predicted SPDF was also constructed with the same method, where the predicted grain-ages are a random subsample of the hypsometric curve and are each given an arbitrary average uncertainty. Brewer et al. (2003) define the mismatch P<sub>diff</sub> between observed and predicted SPDF with equation (1):
- 210  $P_{diff} = \frac{\sum_{t=0}^{t=\infty} |P_1(t) P_2(t)|}{2} \times 100,$

195

(1)

(2)

where  $(P_1 \text{ and } P_2)$  are the probabilities of the two distributions calculated at each age step (t).  $P_{diff}$  relates to the area comprised between two SPDF in the age-frequency space. Brewer et al. (2003) calculate the 95% confidence mismatch between observed and predicted SPDF through a Monte Carlo simulation.

- 215 Vermeesch (2007) has shown that, for the purpose of comparison, observed and predicted distributions are best expressed as a cumulative age distribution (CAD) rather than SPDF. A CAD is a step-function with the sorted mean ages on the x axis and the related quantiles on the y axis (Vermeesch, 2007), This method is preferred because it avoids the possible sources of bias introduced by: (i) choice of the smoothing parameter in the kernel density estimations (KDE); (ii) binning in histograms; and (iii) uncertainty-based weighting in the SPDF curves (Vermeesch, 2012). To evaluate the goodness of fit between observed 220 and predicted CADs, Vermeesch (2007), uses the Kolmogorov-Smirnov (KS) statistic, which informs the maximum distance
- $d_{KS}$  between two cumulative distribution functions as follows (Massey, 1951 and references therein);  $d_{KS} = maximum |CAD_{observed}(t) - CAD_{predicted}(t)|$ .

### Deleted:

-(	Deleted: The ultimate aim
~(	Deleted: is
Y	Deleted: find the best-fit pattern of erosion by minimizing
)(	Deleted: distribution
-(	Formatted: English (US)

-(	Deleted: ,
-(	Formatted: English (US)
	Deleted: differs
-(	Deleted: construct a
-(	Deleted: function
-(	Formatted: English (US)
1	Deleted: one
°(	Deleted: is
-(1	Deleted: (2003)

Deleted: (2003) calculate the 95% confidence mismatch between observed and predicted SPDF through a Monte Carlo simulation. The same approach was complemented by Stock et al. (2006), who performed a Kuiper equality test (Kuiper, 1960) for each Monte Carlo iteration and used the percentage of successful tests as confidence level...

Formatted: English (US)

-(	Formatted: English (US)
-(	Deleted: ),
-(	Deleted: (Vermeesch, 2012)
-(	Formatted: English (US)
(	Formatted: English (US)

Given an observed CAD with k observations (i.e., ages),  $d_{KS}$  is calculated for several n=k sub-samples of the predicted CAD. The 95<sup>th</sup> percentile of all sorted  $d_{KS}$  ( $d_{KS}$  95) is used as the least significant dissimilarity to reject the null hypothesis that the observations are drawn from the predicted CAD. In other words, an observed CAD that plots entirely within the range  $CAD_{predicted} \pm d_{KS}$  (Fig.3) cannot be discerned from the predicted age population at the 95% confidence level. As an alternative to this iterative method, the confidence region for a predicted CAD can be calculated with the analytical solution of the Dvoretzky-Kiefer-Wolfowitz (DKW) inequality as follows:

$$d_{KS_{2}95} \approx \varepsilon = \sqrt{\frac{\ln^2}{a}},$$

255

250 where the DKW distance  $\varepsilon$  approximates well the  $d_{KS \ 95}$  as a function of the confidence level  $(1 - \alpha)$  and the sample size k (Fig.3) (Massart, 1990; Massey, 1951).

Riebe et al. (2015) further developed the bootstrapping approach described above to age distributions (in SPDF form). Instead of basing their analysis on the KS statistic, the 95% confidence envelope of the prediction is iteratively estimated at each age step  $t_i$ . For each  $t_i$ , the distribution of 10,000 predicted age frequencies  $SPDF_{predicted}(t_i)$  is used to draw 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles at each age step. Because the age steps  $t_i$  relate to the elevation steps in the catchment through the bedrock



#### sorted observations

Figure 3 - Example of metrics to compare two cumulative distributions drawn from different sample sizes. Here, an observed Cumulative Age Distribution (CAD) for n=k (orange stepwise dashed line) is compared to a predicted CAD drawn for n>k (solid black line). The Kolmogorov-Smirnov statistic d KS equals the maximum absolute vertical distance between distributions in the observations-frequency space. The Kuiper statistic d Kui equals the sum of both positive and negative maxima. For  $\alpha=0.05$  (i.e. 95% confidence) and n=k, a Dvoretzky-Kiefer-Wolfowitz confidence region can be calculated (grav shading). In this example, observed and predicted distributions are drawn from two statistically different populations at the 95% confidence level, because the orange curve exceeds the grav region.

Deleted: .
Deleted: (d <sub>KS_95</sub> )
<b>Deleted:</b> to define a region of significance around
Deleted: continuous
Deleted: An
Deleted: is considered to
Deleted: drawn
Deleted: same
Deleted: ,
Deleted: approximated calculating
Deleted: bounds
<b>Deleted:</b> where the DKW distance $\varepsilon$ approximates well $d_{KS, 95}$ as a function of the confidence level $(1 - \alpha)$ and the number of observed detrial grain-ages $k$ (Fig.3) (Massart, 1990).
Formatted: English (US)

Deleted: quantiles

(3)

thermochronometric data, those elevation steps exhibiting excess (>97.5<sup>th</sup>) or deficit (<2.5<sup>th</sup>) frequency are interpreted to produce sediment in excess or deficit with respect to the reference scenario of uniform erosion, at the 95% confidence level.

- The approaches summarized above are well-suited to <u>test</u> observed detrital age distributions <u>against</u> the <u>null hypothesis of</u> spatially uniform erosion. However, even if the uniform erosion hypothesis is rejected, such a test does not yield further
- information about the spatial variability of sediment production. One way to gain additional information would be to test the observed distribution against predicted age distributions from erosion scenarios other than spatially uniform, thereby quantifying the likelihood that the measured grain-ages could be produced by the tested scenarios. This approach would help practitioners in deciding the sampling strategy, calculating the appropriate number of measurements and the resources to be allocated for them. In the next paragraphs we introduce the new *ESD thermotrace* software, an open-source tool built on the
- cited previous work to consistently perform the following: (i) determine the confidence level allowed by the detrital sample size in rejecting the uniform erosion hypothesis; (ii) analyze the statistical power of detecting the effect size caused by alternative erosion hypotheses, as a function of the number of grains; and (iii) test observed detrital distributions against all input erosion hypotheses (both uniform and alternative scenarios) and calculate the likelihood that the detrital sample may have been drawn from either of them. The software is introduced in the paragraphs below,

#### 4 ESD\_thermotrace: A tool to study the uncertainty of tracer thermochronology datasets

The software ESD thermotrace performs the steps briefly outlined here below. For additional details, the reader is referred to the illustrative case study <u>farther</u> below and the well-commented code itself (Madella et al., 2021)[Please note that a temporal preview access to the cited future DOI landing page is available: <u>https://dataservices.gfz-</u>
 290 potsdam.de/panmetaworks/review/c375e8eb059bb94ace622t2075b99baf6c513d3b560989a6b307bbe5d5a20217/\_]:

- 1. Bedrock age map interpolation.
  - Input: bedrock age-elevation dataset and digital elevation model.
  - Output: Bedrock age map.
- Alternatively, an externally produced age map can be imported.

#### 295

300

#### 2. Bedrock age uncertainty map interpolation.

- o Input: bedrock age-elevation dataset (point data) and bedrock age map (grid data).
- o Output: Bedrock age uncertainty map.
- Method: The uncertainty of the age map is estimated through bootstrapping. An externally produced uncertainty map is required if the age map was imported.

Method: users can choose among 1D linear regression, 3D linear interpolation, 3D radial basis function.

- 3. Extract catchment bedrock age, coordinates, mineral fertility and erosion data.
  - Input: catchment outline, bedrock age and age uncertainty maps, mineral fertility map, one or more erosion maps.

7

#### Deleted:

#### Deleted: determine whether an

**Deleted:** distribution is significantly dissimilar from a predicted distribution. Such predicted

Deleted: have so far been constructed based on

Deleted: source bedrock age-elevation data and assuming

#### Deleted: Because

**Deleted:** uniform erosion have not been used to make predictions, the inferred pattern of denudation in the cited studies is qualitative to semi-quantitative. Moreover, the confidence level of

**Deleted:** results in terms of dissimilarity is commonly assessed only after samples have been processed and dated.

**Deleted:** , though very convenient and rather widespread in geologic studies, may undermine the statistical rigor of a study. A preferable approach is to define the confidence level for rejecting or confirming...

Deleted: study's hypothesis in the first place. The

#### Deleted: and

**Deleted:** to test the hypothesis should be decided therefrom.

Deleted:

Deleted: help

Deleted: number

Deleted: grain-ages necessary to produce statistically sound results

**Deleted:** bedrock data, catchment hypsometry, erosion scenario and the chosen confidence level. We will also show the use of our tool to evaluate the fit between predicted and observed distributions

#### Formatted: English (US)

Deleted: Python-based Jupyter Notebook

#### Formatted: English (US)

**Deleted:** We developed a new tool to predict detrital age distributions, calculate their dissimilarity to the observations and determine the confidence level of the results. The software

#### Deleted: down

Deleted: ESD_thermotrace	
Formatted: English (US)	
Deleted: mode	

0	Output: a table of all the listed	catchment properties necessa	ary to predict detrital age distributions

- Method: for each cell of the age map contained in the catchment outline, the local coordinates, age, fertility and erosional weight(s) are extracted.
- 4. Predict detrital grain age distributions for each erosion scenario.
  - Input: table of catchment data.

345

350

355

360

365

370

- o Output: a predicted detrital age population for each erosion scenario and related cumulative age distribution.
- Method: A number of ages proportional to erosional weight and fertility is drawn from each cell, for each scenario. The ages from all catchment cells collectively represent a predicted population, from which cumulative age distributions are constructed.
- <u>Calculate:</u> (i) the likelihood of rejecting the uniform erosion hypothesis with the observed sample size; and (ii) the statistical power of discerning (predicted) erosion scenarios as a function of sample size.
  - o Input: one or more sets of observed grain-ages, the predicted detrital populations and distributions.
  - Output: A graph displaying: (i) the confidence of rejecting uniform erosion with the observed sample size; and (ii) the statistical power curve of discerning the scenarios from uniform erosion varying sample size.
  - Method: first, the d<sub>ks</sub> 25 for the available sample size k is calculated with Equation 3; then, the likelihood that the n=k observed CAD is more dissimilar than the d<sub>ks</sub> 25 is calculated through bootstrapping. The same operation is also repeated for a range of sample sizes (20<k<130), in order to estimate the rise in confidence level caused by the increasing sample size (if the observed distribution and associated uncertainty remained identical despite the changing sample size). To calculate the statistical power of discerning the tested erosion hypotheses, the same approach is applied. In this case, however, the software draws a number of distributions from each erosion scenario, instead of the observed grain-ages.

 Given a set of observed detrital grain-ages, calculate the plausibility of each erosion scenario (i.e. the likelihood that hypotheses cannot be rejected with high confidence).

• Input: same as point 5).

o Output: two plots to visualize how plausibly the observed grain-ages have been drawn from the predictions

0	Method: in the first plot, dissimilarities calculated between predictions and observations are sorted and their
	distribution is plotted in the form of a violin plot. Second, a two-dimensional Multi-Dimensional Scaling
	(MDS) model is fitted to the dissimilarities among the predicted and observed distributions and plotted
	following Vermeesch (2013).

All of the above operations are embedded in an open-source Jupyter Notebook (jupyter.org), a software that allows integrating
text, Python code, and visualizations within the same document for maximum editability and transparency. All plots are
produced with Matplotlib (Hunter, 2007) and Seaborn (Waskom et al., 2020), Python libraries and are colored using the color
blind-friendly and perceptually uniform ScientificColourMaps6 (Crameri et al., 2020), In the following paragraph, for

Deleted: Evaluate
Deleted: confidence
Deleted: detecting
Deleted: from
Deleted: set
Deleted: measured detrital ages
Deleted: informing
Deleted: in discerning the tested
Deleted: <#>Method: predicted age distributions from erosion scenarios are compared to predicted uniform erosion distributions using the KS statistic as metric for dissimilarity. Confidence intervals are determined through Monte Carlo subsampling as a function of varying sample sizes.¶ Evaluate the fit of the erosion scenarios to the observed detrital ages¶

Deleted	the fit of predictions to
Deleted	age distribution(s).

Deleted: Secondly

Formatted: English (US)

(	Formatted: English (US)	)
~(	Formatted: English (US)	)
~(	Formatted: English (US)	)



Figure 4 – Input bedrock data from Stock et al. (2006). (A) Raster images of the study area's digital elevation model, resampled to the user-specific cell size, and (B) bedrock surface AHe age interpolated based on the linear regression of age-elevation data. In both plots, point data inform bedrock sample locations and related AHe ages. Polygons shows the location

of the Inyo Creek catchment.

#### 5 Application of ESD\_thermotrace to the Inyo Creek case study

We apply *ESD\_thermotrace* to the bedrock and detrital apatite (U-Th(-Sm)/He (AHe thereafter) age datasets of Stock et al. (2006). In that study, 9 bedrock AHe ages from the Inyo Creek and adjacent Lone Pine Creek catchments (eastern Sierra Nevada, California, USA) are used to constrain the age-elevation relationship of the source area. The authors compared these bedrock data to the AHe age distribution of river sand samples from both catchment outlets. For the sake of simplicity, here

we only consider the fine sand sample (k=52) from the Inyo Creek catchment. Based on their analysis Stock et al. (2006) <u>inferred</u> no significant variability of sediment production (erosion) across all elevations within the catchment.

#### 5.1 Data import and bedrock age interpolation

400

405

395

The first few steps of *ESD\_thermotrace* allow importing: (i) the digital elevation model of the study area; (ii) the polygon of the catchment of interest; (iii) a table of bedrock age and elevation, data; (iv) the (optional) erosion scenarios to be tested; and (v) a (optional) fertility map. Next, the surface bedrock ages are calculated with the preferred method (see above and README file in Madella et al., 2021). The bedrock AHe ages of Stock et al. (2016), recalculated after Riebe et al. (2015), exhibit a ~60 My age increase from ca. 2 to 4.5 km of elevation range that is well-described by <u>an inverse variance-weighted</u> linear regression ( $q^2 = 0.03$ ). The input data are plotted in Figure 4a. In Figure 4b, every cell is assigned an AHe age as a linear function of elevation (Fig.5 inset), so as to map the bedrock cooling age on the topographic surface. The error of this interpolation is also mapped through bootstrapping as a function of elevation (Fig.5).



Figure 5— The raster image displays the bedrock age interpolation error mapped to the topographic surface of the study area. The inset shows the inverse variance-weighed linear regression employed as age-elevation function, as well as the associated  $1\sigma$  uncertainty.

10

Deleted: this

Deleted: infer

 Deleted: ,

 Deleted: ,

 Deleted: age is

 Deleted: age is

 Deleted: a

 Deleted: R<sup>2</sup>

 Deleted: 94

**Deleted:** This informs the elevation-dependent standard deviation of 1,000 regressions, in which the *<object>*bedrock ages are each drawn from a normal distribution that account for the input analytical uncertainty.

#### 420 5.2 Extraction of catchment data

Next, *ESD\_thermotrace* extracts the x,y,z coordinates, the bedrock cooling age and the related error for all the cells bound by the catchment outline. These data are written in a table, to which a column informing erosional weights is added for each desired erosion scenario and for mineral fertility. In addition to the user-defined erosion maps, by default the software considers the uniform erosion scenario '*Euni*' (spatially constant erosional weight). Two further example scenarios can be toggled to test

- 425 for an exponential increase of erosion with elevation <u>ot</u> an exponential decrease of erosion with elevation, <u>(not considered here)</u>. We note that any other spatial variation in erosion can be defined by a user such that 'erosion maps' for the catchment following stream power, slope dependent, glacial slide velocity, or other approaches from geomorphic transport laws can be input. <u>For this case study, bulk geochemistry and point-counting analyses of Hirt (2007)</u> indicate that apatite fertility does not significantly vary within the three lithologies found in the Inyo Creek catchment (Lone Pine granodiorite, Paradise
- 430 granodiorite, Whitney granodiorites). For illustrative purpose, we test uniform erosion against two opposite step functions of elevation: a scenario with F-times higher erosion efficiency above the median catchment elevation and one with F-times higher erosion below the median elevation. The factor F equals twice the ratio between the most frequent and the least frequent elevation. For this calculation, the hypsometric histogram is constructed with a number of bins equal to the maximum difference in bedrock cooling ages, divided by the mean age uncertainty (rounded up to the next integer). Here, the Inyo Creek catchment is binned in 7 elevation ranges, of which the most frequent is 5-times the least frequent one, resulting in F = 10. In other words, we test for an increase in erosion twice as prominent as the hypsometric peak (10-fold), once below ('*E*<*Zmed*')
- and once above ('*E*>Zmed') the median catchment elevation.

#### 5.3 Prediction of age populations and <u>detrital</u> age distributions

The erosional weights are used to forward model sediment production at each position in the catchment. To do so, a number of grain-ages proportional to the erosional weight and to the local mineral fertility are randomly chosen for each cell. These grain-ages are randomly drawn from a normal distribution constructed using the local interpolated bedrock age and the related error. The randomly picked grain-ages of all cells are stored in one suite of grain-ages, which represents the predicted detrital grain-age population of a well-mixed fluvial sediment at the catchment outlet. Such a detrital population is predicted for each erosion scenario, and related predicted CADs are constructed by sorting the age populations and plotted as a step function (Fig.cb). For the sake of quick visual comparisons, *ESD thermotrace* plots a kernel density estimation with arbitrary

(right), for the sake of quick visual comparisons, ESD intermotice prois a kenter density estimation with abilitary smoothing (Fig.6a). However, the dissimilarity among distributions will be evaluated exclusively based on the cumulative form, in order to minimize bias (Vermeesch, 2007). In Figure 6b, all predicted and observed cumulative age distributions are plotted. For a first order impression of the uncertainty due to sample size, <u>the 95% confidence envelope</u> of the reference

Deleted:	_
Deleted: and	
Deleted: .	5
1	Deleted:

**Deleted:** For this application, bulk geochemistry and pointcounting analyses of Riebe et al. (2015)

**Deleted:** simple step-wise erosion functions of elevation: a scenario of 2-fold decrease in erosion below 3000 m ('E < 3000') and the opposite scenario with 2-fold increase in erosion above 3000 m ('E > 3000')....

De	leted:	x.v

Deleted: forward modeled	
leteu. Ioi ward modeled	

Deleted:	100 n=k	random	subsamples
----------	---------	--------	------------



as a function of sample size. This is the main step to define how many grain-ages are necessary to detect a hypothetical erosion pattern. Let us consider the observed CAD constructed with the sorted mean grain-ages. The mean standard deviation of the

- 475 observed grain-ages is used to generate 10'000 Monte Carlo samples of this observed CAD, which account for age dispersion due to analytical error and/or reproducibility. Then the confidence in rejecting uniform erosion equals the fraction of sampled CADs that is more dissimilar to 'Euni' than the least significant dissimilarity (*d<sub>KS</sub>* 95) allowed by the sample size *k* (see Eq.3). The so-calculated confidence can be read in the scatterplot of Figure 7 (light blue circle), and it equals 55% for the Stock et al. (2006) data. Figure 7 also shows that the 95% level of significance could be achieved with more than 128 grain-ages. Here,
- 480 we clarify that the latter estimate assumes that the observed distribution would not change in shape while varying the sample size, and as such it can only be treated as an indicative number. If the confidence allowed by the actual sample size is lower than the previously chosen level of significance, an analysis of statistical power can help identifying how likely the available number of grain-ages could detect a known effect size (i.e. the dissimilarity caused by a known erosion scenario). This analysis of statistical power is based on the iterative comparison between the reference scenario and all other <u>erosion scenario</u>.

Analysis of confidence and statistical power

100 95% ikelihood to reject "Euni" [%] 80 68% 60 0 40 20 E<Zmed E>Zmed Inyo Creek Detrital 0+ 20 40 60 80 100 120 number of grains

Figure 7— Confidence level at which the observed CAD (n=52) can be discerned from the uniform erosion prediction (cvan circle). The dashed cvan line shows how the allowed confidence level would vary, if the same observed CAD relied on 20-130 grain-ages. Statistical power of discerning the tested erosion scenarios from "Euni' as function of sample size (solid lines). In this case study, the observed CAD of the Invo Creek detrital sample allows rejecting the uniform erosion hypothesis with 55% confidence at best. An analysis to detect the scenarios 'E<Zmed' and 'E>Zmed' would have a statistical power greater than 95% in rejecting 'Euni' with more than 37 grain-ages. Euni: uniform erosion; E<Zmed: 10-fold erosion below median elevation; E>Zmed: 10-fold erosion above median elevation.

Deleted: predicted CADs.

Deleted: The confidence in discerning scenarios

predictions. This is achieved through random subsampling of the predicted distributions for a range of possible sample sizes Q0≤n≤130 in this case). For each n<sub>i</sub>, the least significant dissimilarity d<sub>KX,QS</sub>(n<sub>i</sub>)\* of the reference scenario is calculated for a=0.05 and n=n<sub>i</sub> using Equation 3. In the observations-cumulative frequency space (Fig.6b), d<sub>KX,QS</sub>(n<sub>i</sub>)\* represents the vertical distance that any n=n<sub>i</sub> distribution needs to exceed to be discerned from a n=n<sub>i</sub> subsample of the reference scenario, with 95% confidence (purple shading in Fig.6b). Next, 1,000 d<sub>KX,QS</sub>(n<sub>i</sub>) are calculated between n=n<sub>i</sub> subsamples of each erosion scenario and the reference scenario. The probability of d<sub>KX,QS</sub>(n<sub>i</sub>) exceeding d<sub>KX,QS</sub>(n<sub>i</sub>)<sup>\*</sup> for each n<sub>i</sub> can be read from the curves in Figure 7. For the Inyo Creek case study (Stock et al., 2006) both tested erosion scenarios yield predicted CADs that are very dissimilar from uniform erosion. Consequently, the probability to detect such scenarios yould exceed 95% even with only 40 grains.

495

500

Figure 7 shows the use of *ESD\_thermotrace* as a tool to explore the feasibility of a tracer thermochronology study. Based on their research question, users can <u>apply</u> a number of possible erosion maps and test the likelihood with which the respective <u>detrital distributions could</u> be<u>discerned from uniform erosion</u> through detrital tracer thermochronology, as a <u>function of sample</u> <u>size</u>. Such analysis of feasibility is not only beneficial in terms of <u>better quantifying uncertainties</u>, but it can also assist investigators in defining the budget for measurements at an early stage of proposal writing. Alternatively, in cases where the number of datable grains is limited by material properties, budget, or other logistic reasons, Figure 7 informs the <u>maximum</u> confidence level of the results.

#### 5.5 Evaluating the plausibility of each scenario,

The final steps of *ESD\_thermotrace* assist users in finding the <u>erosion</u> scenario that <u>is most likely to generate a predicted CAD</u>
that resembles the observed <u>CAD</u>. For each erosion scenario, *d<sub>KS</sub>* and *d<sub>Kut</sub>* (Fig.3) are calculated between 10,000 n=k
subsamples of the respective predicted CAD and the observed CAD. The distribution of these *d<sub>KS</sub>* and *d<sub>Kut</sub>* values is shown in the form of a split violin plot in Figure 8a and it is to be compared to the range of values shown by the violin in Figure 8b. The latter shows the distribution of *d<sub>KS</sub>* and *d<sub>Kut</sub>* calculated between random subsamples of the observed CAD that account for <u>the mean analytical error</u> and the observed CAD itself (constructed only with the mean ages). In other words, Fig. 8b <u>displays how</u>
the dispersion of a predicted CAD (accounting for analytical error and sample size) is distributed. The "plausibility" of each scenario is plotted beneath each violin and <u>it equals</u> the probability that the values plotted in Figure 8a fall within the <u>one-sided</u> lower 95<sup>th</sup> percentile of those shown in Figure 8b. More specifically, there is <u>51.6% probability that</u> the observed CAD <u>could</u> <u>be drawn from</u> the uniform erosion scenario, and <u>13.7%</u> and <u>0.0%</u> probability that scenarios 'E<Zmed' and <u>t</u>E>Zmed' generate predicted CAD stat fall within the spread of the observed CAD, respectively.

515

	subsamples of the reference scenario are randomly drawn and the KS distance $d_{kS}(n_i)$ from the original $n > n_i$ distribution is calculated for each iteration. The 95 <sup>th</sup> percentile of all sortedhe least significant dissimilarity $d_{kS}(n_i)$ is referred to as $d_{kS}(n_i)^*$ , the inherent noise (i.e. dissimilarity)[26]
1	Formatted: Not Superscript/ Subscript
	<b>Deleted:</b> in ordero be discerned from a $n=n_i$ subsample of the reference scenario, with 95% confidence. ( [27])
1	Formatted: Not Superscript/ Subscript
	<b>Deleted:</b> is plottedor each $n_i$ (Figan be read from the curves in Figure 7) For the Inyo Creek case study (Stock et al., 2006) theoth tested erosion scenarios $F \leq 3000^{\circ}$ and $t \geq 5000^{\circ}$ ield predicted CADs that are not extremelyery dissimilar from uniform erosion. As a consequence, even 140 grain-ages would not sufficeonsequently, the probability to detect theseuch scenarios the 95% confidence level
	<b>Deleted:</b> importpply a number of possible erosion maps (or calculate them as functions of elevation)nd test whether or not canhe likelihood with which the respective detrital distributions could be detectediscerned from uniform erosion through detrital tracer thermochronology. In other words, Figure 7 displays an assessment of the number of erain-ages required to risorrously test

**Deleted:**  $30...0 \le n \le 140...30$  in this case). For each  $n_i$ , 1,000  $n = n_i$ 

**Deleted:** fit...lausibility of each scenario to the detrital observations

not only beneficial in terms of scientific rigor

as a hypothesis....unction of sample size. Such analysis of feasibility

.. [29]

(... [30])

Deleted: best fits ... s most likely to generate a predicted CAD that resembles the observed distribution of detrital grain-ages... AD. For each erosion scenario, dKS and dKui (Fig.3) are calculated between 10,000 n=k subsamples of the respective predicted CAD and the observed CAD. The distribution of these  $d_{KS}$  and  $d_{Kui}$  values is shown in the form of a split violin plot in Figure 8a and it is to be compared to the range of values shown by the violin in Figure 8b. The latter shows the distribution of  $d_{KS}$  and  $d_{Kui}$  calculated between random subsamples of the observed CAD that account for 10 uncertainty,...he mean analytical error and the observed CAD itself (constructed only with the mean ages). In other words, Fig. 8b visualizes...isplays how the dispersion of a detrital distribution's noise due to ... redicted CAD (accounting for analytical error and sample size .... is distributed. The "plausibility" of each scenario is plotted beneath each violin and is computed as ... t equals the probability that the values plotted in Figure 8a fall within the on...ne sided lower 95th percentile of those shown in Figure 8b. More specifically, there is 62.9% fit to ...1.6% probability that the observed CAD for ...ould be drawn from the uniform erosion and a 87.6...3.7% and 3.3% fit for the factor of 2 increase in erosion below 3,000 m ('....0% probability that scenarios 'E<3000')...med' and above 3,000 m ('... E>3000'),...med' generate predicted CADs that fall within the spread of the observed CAD, respectively. Regardless of the best-fit erosion scenario, the 52 available detrital grain-ages provide a (limited) confidence level of 45-50% for the interpretation. We refrain from making an improved interpretation of the erosion distribution in this catchment from that of Stock et al. (2006), because more grains would be needed, and as recent (and in progress) work has demonstrated, a larger range of different grain sizes should preferably be considered than that of Stock et al. (2006) to better understand the pattern of erosion (Lukens et al., 2019; see also discussion below; Riebe et al., 2015). ... [31]



Figure 8 – Violin plot showing the KS (and Kuiper) statistics between the predicted CAD of each erosion scenario and observed CAD, calculated for 1000 n=k=52 subsamples (A). KS (and Kuiper) statistics between the observed CAD and 1000 random n=k=52 distributions, drawn from the same detrial age population but including uncertainty (B). Every violin is split in two halves with equal area, each representing the probability distribution function of 10000 KS (or Kuiper) statistics. The widest point of each semi-violin informs the median KS (or Kuiper) statistic; The closer to zero, the more similar is a scenario to the observed detrital age distribution. The most unlikely scenario is 'b>Zmed'. Both other scenarios are plausible, because their dissimilarity to CAD<sub>observed</sub> largely overlaps the range of dissimilarity due to analytical error of the detrital ages.

Lastly, as suggested by Vermeesch (2013), the *ESD thermotrace* program applies a two-component Multi-Dimensional Scaling model (MDS) to all predicted and observed distributions. This algorithm fits a 2-dimensional coordinate system to the measured dissimilarities  $d_{KS}$  among all considered distributions (both predicted and observed). In a well-fit MDS model, distances among points in the modelled 2D-space are a good approximation of the actual dissimilarities among the input elements (Fig.9a). To reach a satisfactory fit, modelled dissimilarities are plotted against input dissimilarities (Fig.9b) and the

sum of all distances to the 1:1 line is minimized (a procedure commonly referred to as *stress* minimization). The MDS plot renders an immediate visualization of the dissimilarity among distributions, where more similar distributions plot closer to each other. Moreover, with the addition of the 68% and 95% confidence ellipses, the degree of overlap among distributions is also easily visually assessed (Fig.9a).

10 also easily visually assessed (Fig. 9a).

705

The application of the ESD\_thermotrace software to Stock et al.'s (2006) Inyo Creek data shows the following:

# Deleted: <object> Deleted: also Deleted: components Deleted: . Deleted: calculated Deleted: .

Deleted: We have illustrated the functioning of the ESD\_thermotrace by evaluating Stock et al.'s (2006) Inyo Creek data. Despite the good fit of scenario 'E<3000', our analysis suggests that based on the used set of 52 grain-ages it is not possible to uniquely detect any of the tested scenarios at with confidence (Fig.7). This outcome stems from the considerable overlap between the range of possible predicted distributions 'Euni' and the detrital observations (Fig.9). Nevertheless, using our approach, we add <object>a quantitative assessment of both the confidence due to sample size (45-50% confidence) and the scenarios fit. We note that Riebe et al. (2015), in contrast, have shown that erosion in the Invo Creek catchment is most likely to increase exponentially with elevation. These authors have fitted an erosional curve that additionally accounts for 10Be-derived denudation rates, as well as 73 additional AHe ages from coarse sand-sized sediment. They conclude that grain size bias is a viable explanation of the observed discrepancy. While we will address the possible issues concerning grain size below, the best-fit erosional pattern of the Inyo Creek is beyond the scope of this paper and will not be further discussed.

- A sample size of 52 grain-ages allows a maximum confidence level of ~54% in rejecting the uniform erosion hypothesis 'Euni' (cyan circle in Fig.7). If the observed CAD had been drawn from ~130 grain-ages, this confidence level would exceed 95% (dashed cyan curve in Fig.7).
  - 52 grain-ages would suffice to detect the effect size caused by 'E<Zmed' and 'E>Zmed' (green and brown curves in Fig. 7). However, the observed detrital sample is extremely unlikely to have been drawn from either of these scenarios.
  - Among the tested scenarios, the uniform erosion hypothesis 'Euni' is the least dissimilar (51.6% likelihood) (Fig.8 and Fig.9).
- 745 Here, while we have shown the functioning of ESD\_thermotrace using the Inyo Creek data from Stock et al. (2006), we refrain from proposing a revised interpretation of the catchment's erosion dynamics. We do note that, in addition to the data from Stock et al. (2006), Riebe et al. (2015) have accounted for <sup>10</sup>Be-derived denudation rates and 73 additional AHe ages from coarse sand-sized sediment. These authors have shown that erosion in the Inyo Creek catchment is best explained by an exponential increase of sediment production with elevation. Thereby, they demonstrate that understanding the pattern of erosion in this catchment requires taking multiple sediment sizes into account (Riebe et al., 2015; Lukens et al., 2020; Sklar et
  - al., 2020; see also discussion below).



Figure 9 – Multi-Dimensional Scaling (MDS) plot. The 2-components MDS model fits a coordinate system to the measured dissimilarities (i.e. KS statistics) among all distributions. If the MDS model has a good fit, distances calculated in the new 2D-space (A) are a good approximation of the input dissimilarities among the elements. The goodness of fit is shown in (B), where modelled dissimilarities are plotted against input dissimilarities. For each distribution in (A) 68% and 95% confidence ellipses are also plotted.

#### 6 How many grains do we need for tracer thermochronology?

- The analysis with ESD thermotrace shows that the appropriate sample size for a tracer thermochronology study cannot be determined a-priori without knowing: (i) the case-specific scientific question (i.e. the spatial pattern of erosion to be tested); 755 (ii) the source area hypsometry; (iii) the desired minimum confidence level; and (iv) the surface bedrock ages. On this note, if possible, it is advisable to explore the feasibility of a study with the already available bedrock ages before sampling. In absence of available published ages, bedrock samples are best processed first to avoid wasting resources on potentially inconclusive analyses of detrital grains. To better illustrate the importance of the initial knowledge of the catchment's geology, we conducted a set of simulations with the same inputs as the Inyo Creek case study (Stock et al., 2006). In these simulations we vary the
- 760 location (in terms of elevation range) of maximal erosion and observe the impact on the statistical power in detecting the imposed pattern. A broad and a narrow gaussian curve of erosion ( $1\sigma = 500$  m, and 100 m, respectively) are applied to the catchment's range of elevations, shifting the position of the peak in 100-m-steps at each simulation. We test two sets of gaussian functions of elevation (2x and 5x), computed in such a way that the prominence of peak erosion equals 2- and 5- times the prominence of the catchment's hypsometric peak, respectively. Therefore, because the most frequent elevation range is ca.
- 765 5-times the least frequent bin (Fig. 10C,F), the "2x" and "5x" curves define an erosion efficiency where peak sediment production is 10-times and 25-times higher than the minimum in the catchment, respectively. Results from this parameter study, example erosional functions and the hypsometric histogram are illustrated in Figure 10.

Figure 10 shows how the statistical power in detecting these gaussian-peak-scenarios is affected by (i) the elevation of the rosional peak; (ii) its amplitude; (iii) its width; and (iv) the detrital sample size. An increase in sample size and/or an increase in amplitude of the erosional peak always correspond to higher statistical power. The elevation of the erosional peak affects the statistical power depending on the position of the erosional peak relative to the peak of hypsometry (Fig.10). In the case of a broad gaussian function of erosion ( $1\sigma = 500$  m), minima (plue areas) are observed where the erosional peak is located at elevations straddling 2,800 m (Fig. 10A,B). These minima in statistical power at ~2,800 m coincide with the broad frequency peak of catchment elevations (Fig, 10C). In this elevation range, the diffused increase in erosion efficiency does accentuate

the peak of the hypsometric curve, but it results in a limited effect size (i.e. a small  $d_{KS}$  dissimilarity to "Euni"). With this configuration, all erosion scenarios peaking within ca. 2,700-2,900 m produce detrital distributions whose effect size is detected in only <60% of the simulations, regardless of sample size and peak amplitude. This implies that certain combinations of erosional pattern and distribution of bedrock ages are poorly suited to be investigated by means of tracer thermochronology

780

In the case of a narrow gaussian function of erosion  $(1\sigma = 100 \text{ m})$  (Fig. 10D,E,F), minima of statistical power are additionally found at peak elevations higher than 3,600 m and lower than 2,200 m (Fig. 10D,E). These occur because peak erosion is applied to a narrow elevation range that is not frequent enough to produce a statistically relevant number of grains. However, in the case of narrow erosional peaks, an increase in sample size substantially increases the statistical power to detect erosion

Deleted: )...; (ii) the source area hypsometry..., (iii) the desired minimum confidence level; and (iv) the surface bedrock ages. On this note, if possible, it is advisable to explore the feasibility of a study with the already available bedrock ages before sampling. In absence of available published ages, bedrock samples are best processed first in order ... o avoid wasting resources on potentially inconclusive analyses of detrital grains. To better illustrate the importance of the initial knowledge of the catchment's geology, we conducted a set of simulations with the same inputs as the Inyo Creek case study (Stock et al., 2006). In these simulations we vary the location (in terms of elevation range) of maximal erosion and observe the impact on the confidence level...tatistical power in detecting the imposed pattern. broad and a narrow gaussian function...urve of erosion (10 = 500 m) is...and 100 m, respectively) are applied to the entire ... atchment's range of catchment ... elevations, shifting the position of the peak erosion ... in 100-m-steps at each simulation. We test two sets of gaussian functions of elevation (2x and 5x), computed in such a way that the prominence of peak erosion equals 2- and 5 -times the prominence of the catchment's hypsometric peak. respectively. Therefore, because the most frequent elevation range is ca. 5-times the least frequent bin (Fig. 10C,F), the "2x" and "5x" curves define an erosion efficiency where peak sediment production is 10-times and 25-times higher than the minimum found ...n the catchment, respectively. These scenarios are meant to simulate realistic geomorphologic situations, where erosion rates might experience a 2- or 5-fold increase within the same catchment. In addition, these scenarios are also tested setting all bedrock age uncertainties to 20%. ... esults from this parameter study and

Deleted: confidence level...tatistical power in discerning...etecting these gaussian-peak-scenarios from spatially constant erosion ('Euni') ...s affected by both ... (i) the elevation of the erosional peak, the ... (ii) its amplitude of the peak, the ... (iii) its width; and the detrital sample size and the uncertainty of bedrock ages... An increase in sample size and/or an increase in amplitude of the erosional neak always correspond to higher success rates. In contrast, an increase in age uncertainty always reduces the success rate....tatistical power. The elevation of the erosional peak affects the confidence level...tatistical power depending on the divergence from position of the erosional peak relative to the catchment's hypsometric hypsometric curve...eak of hypsometry (Fig.10). Confidence...n the case of a broad gaussian function of erosion ( $1\sigma = 500$  m), minima (dark ... lue areas) are observed where the erosional peak is located at elevations greater than 3,500 m and at ca. ...traddling 2,800 m (Fig. 10). The...0A,B). These minima at 3,500 m occur because peak erosion involves an elevation range that is not frequent enough to produce a statistically relevant grain population. In contrast, the region of low confidence at around ...n statistical power at ~2.800 m coincides...oincide with the broad frequency peak of catchment 

. [32]

Deleted: Figure 11 displays an analogous parameter study to that illustrated in Figure 10, with ...n the difference that all...ase of a narrow gaussian peaks...unction of erosion are set... $1\sigma = 100 \text{ m}$ ) (Fig. 10D,E,F), minima of statistical power are additionally found at peak elevations higher than 3,600 m and lower than 2,200 m (Fig. 10D,E). These occur because peak erosion is applied to a narrow elevation range that is not frequent enough to be narrower  $(1\sigma = 100)$ m). While Figure 11 maintains ...roduce a statistically relevant number of grains. However, in the same trends as Figure 10, it also shows that narrow, low-amplitude...ase of narrow erosional peaks remain undetectable regardless of both peak elevation and ... an increase in sample size. However, high-amplitude, narrow erosional peaks (Fig.11C) are likely to be detected ... [34]



#### scenarios, even if centered at the critical elevation of ca. 2,700-2,900 m described just above (Fig.10D,E). This parameter study

demonstrates the importance of analyzing the catchment hypsometry and testing erosion scenarios even before collecting data,

#### (Deleted: in

Deleted: of

**Deleted:** range of peak hypsometry individuated in Figure 10. The application of these simple erosional patterns to the parameter studies above demonstrate

Statistical power to discern gaussian peaks in erosion from "Euni" example erosion curves broad erosional peak: 2x peak hypsometry broad erosional peak: 5x peak hypsometry centered at 2.8 km С в A elevation of erosional peak [km] 3.5 5x 3.0 2x 2.5 hypsometry 2.0 narrow erosional peak: 2x peak hypsometry narrow erosional peak: 5x peak hypsometry F D elevation of erosional peak [km] hypsometry 3.5 3.0 5x 2.5 2.0 40 60 80 100 120 140 40 60 80 100 120 140 0 5 10 number of dated grains number of dated grains frequency ō 20 40 60 80 100

probability [%]

Figure 10 – Analysis of statistical power in discerning between the uniform erosion scenario 'Euni' and scenarios with a broad (A,B) or narrow (D,E) gaussian peak of erosion, as a function of sample size and elevation of the gaussian peak. The color scale informs the statistical power as a function of the number of grains (x axis) and the elevation at which the gaussian peak is centered (y axis). The catchment hypsometry (C,F) and age-elevation-relationship used here are the same as the Inyo Creek data (Stock et al., 2006). (A) Gaussian erosional function with 1 $\sigma$  = 500 m and 2 times the peak prominence of the hypsometric histogram. (B) Gaussian erosional function with 1 $\sigma$  = 100 m and 2 times the peak prominence of the hypsometric histogram. (E) Gaussian erosional function with 1 $\sigma$  = 100 m s times the peak prominence of the hypsometric histogram. (E) Gaussian erosional function with 1 $\sigma$  = 100 m s times the peak prominence of the hypsometric histogram. (E) Gaussian erosional function with 1 $\sigma$  = 100 m s times the peak prominence of the hypsometric histogram. (E) Gaussian erosional function with 1 $\sigma$  = 100 m to the peak prominence of the hypsometric histogram. (E) Gaussian erosional function with 1 $\sigma$  = 100 m to the text for more details.

in order to make an informed choice on the appropriate sample sizes and to identify possible scenarios that are unlikely to be detected with high confidence through detrital tracer thermochronology.

#### 1120 7 Other sources of uncertainty

The approach presented here for predicting and interpreting grain-age distributions enables quantifying the confidence level in rejecting the uniform erosion hypothesis as well as the statistical power in discerning a prescribed erosion scenario from uniform erosion, through detrital tracer thermochronology. Confidence level and statical power are computed as a function of the samples size, the catchment properties and input erosion scenario. Nevertheless, even if the plausibility of a scenario is 125 inferred to be high, our software, by design, does not suggest a unique solution and interpretations are also subject to additional sources of uncertainty. These include: (1) complex bedrock age-elevation relationship (Malusà and Fitzgerald, 2020); and (2) spatial variability of sediment size resulting from either of transport distance (e.g. Lukens et al., 2020; Malusà and Garzanti, 2019), geomorphic process (e.g. Riebe et al., 2015; van Dongen et al., 2019), lithological differences (von Eynatten et al., 2012), or vegetation effects on weathering and erosion (Starke et al., 2020). Both factors should be taken into account as discussed here below.

1130

In some cases, elevation alone cannot explain the entire variance of bedrock thermochronometric ages. For example, spatial variability of bedrock ages may reflect the proximity to tectonic structures or sub-catchment thermal events (e.g., magmatism, spatial changes in thermal gradients in proximity of major faults, etc). In such cases, an improved sampling of the spatial

- 1135 distribution of bedrock ages is needed, and may require more complex interpolation functions (e.g. Glotzbach et al., 2013). To accommodate these complexities, ESD thermotrace allows for 1D and 3D-linear interpolation as well as radial basis function interpolation. In all these cases, the interpolated age uncertainty is calculated through bootstrapping. In addition, users can opt to import an independently interpolated surface bedrock map such as from a thermokinematic model (e.g. Whipp and Ehlers, 2019). For additional information on bedrock thermochronometric age mapping the reader is referred to the "README" file
- 140 included with the software (Madella et al., 2021). Regardless of which method is used, users should be aware that depending on the quality of the bedrock data, different interpolation methods may yield different predicted distributions. Accordingly, resulting interpolated surfaces should be carefully evaluated, and a preference should be motivated by field observations and/or independent constraints. Here, the map of interpolated age uncertainty can also inform the locations where additional bedrock sampling would help reduce the uncertainty. Lastly, we note that an increase in age uncertainty always determines a decrease 145 in the statistical power of the analysis.

Other possible sources of bias concern the grain size of the analyzed samples. Several issues may modify the original fingerprint of river sand (Malusà and Garzanti, 2019), such as downstream grain abrasion and fracturing, hydraulic sorting and weathering on the hillslope associated with a grain (Attal and Lavé, 2009). For example, grains sourced the farthest from the

19

Deleted: a science-Deleted: size

Deleted: The approach presented here for interpreting grain-age distributions enables the quantification of the confidence level in detecting erosion patterns with detrital tracer thermochronology. Confidence levels are computed as a function of the number of samples, input erosion scenario and catchment properties. Nevertheless, even if a scenario is inferred to be highly probable, it may not be a unique solution or it may only inform the behavior of the one analyzed grain size fraction. Uncertainty in the interpretation can stem from factors such as: (1) complex bedrock age-elevation relationship and (2) spatial variability of sediment size resulting from transport distance (e.g. Lukens et al., 2019)

Deleted: (van Dongen et al., 2019; e.g. Riebe et al., 2015)

#### Deleted:

Deleted:

Deleted: due to structures

#### Deleted: are

Deleted: (e.g. Whipp and Ehlers, 2019). For additional information on bedrock age mapping the reader is referred to the "README" file attached to the code

Deleted: Lastly

Deleted: noise

#### Deleted:

Other possible sources of bias concern the grain size of the analyzed samples. One issue is that downstream sediment abrasion may significantly modify detrital grain-age distributions, as can the weathering and erosion processes associated with a grain. For example, grains sourced the farthest from the sampling spot may be underrepresented in the analyzed grain size fraction (Lukens et al., 2019). A second issue is the initial grain size distribution found on the hillslopes. Depending on the lithology (von Eynatten et al., 2012) and on the locally dominant denudational process (van Dongen et al., 2019), different hillslopes of the same catchment may produce significantly different sediment size distributions (e.g. Riebe et al., 2015). As a consequence, the mixed detrital sample may exhibit a bias in the relative abundance of the different age components. Both these issues can be mitigated through analysis of multiple grain size fractions (Lukens et al., 2019), multiple measures of hillslope sediment size distributions, composite analyses of trunk stream and tributary stream sediment samples, analyses of thermochronometers from different minerals.

sampling spot may be underrepresented in the analyzed grain size fraction (Lukens et al., 2020), as has also been shown for the Inyo Creek catchment (Sklar et al., 2020). Furthermore, in addition to the mineral fertility inherent to the exposed bedrock, the grain size distribution of the material found on the hillslopes (i.e. the material ready for transport) should be taken into account. Depending on the lithology (von Eynatten et al., 2012; Roda-Boluda et al., 2018) and on the locally dominant denudational process (van Dongen et al., 2019), different hillslopes of the same catchment may produce substantially different sediment size distributions (e.g. Riebe et al., 2015; Attal et al., 2015). Consequently, the mixed detrial sample may exhibit a

bias in the relative abundance of the different age components. Both these issues can be mitigated through analysis of multiple grain size fractions (Lukens et al., 2020), multiple measures of hillslope sediment size distributions, composite analyses of

trunk stream and tributary stream sediment samples, and analyses of thermochronometers from different minerals.

#### 8 Conclusion

This study reviewed previous approaches used to compare predicted and observed detrital grain-age distributions in the framework of tracer thermochronology. We have built upon these to develop a new tool (*ESD\_thermotrace*) to investigate the upstream pattern of catchment erosion and the confidence level in uniquely inferring this as a function of sample size and study-site-specific variables. To demonstrate the utility of this approach, we presented an analysis of previously published data from the Inyo Creek Catchment in California. The example highlighted the utility of measuring a large number of grains, and how multiple erosion scenarios are plausible for this catchment with the considered number of grains. The degree of statistical

confidence permitted by this case study has also been quantified. We showed how the use of *ESD\_thermotrace* can increase the statistical rigor of tracer thermochronology studies and how it can also assist investigators in budgeting analytical costs of
a future project. In cases where the number of datable grains is limited by any reason, the confidence level of the results can be quantified, and the statistical power of the analysis can be estimated.

#### Acknowledgements

Research was funded through the DFG Project numbers GL724/9-1 awarded to C.G., and EH329/18-1 awarded to T.A.E.. We thank M. Malusà as well as C. Lukens and C. Riebe for their constructive comments.

#### 215 Code availability

The source code of *ESD\_thermotrace* is freely available for download (Madella et al., 2021), [Please note that a temporal preview access to the cited future DOI landing page is available: <u>https://dataservices.gfz-potsdam.de/panmetaworks/review/c375e8eb059bb94ace622f2075b99baf6c513d3b560989a6b307bbe5d5a20217/]</u>

Deleted:

Formatted: English (US)

-(	Deleted: XXX
-(	Deleted: YYY
•(	Deleted: reviews
•	Formatted: English (US)
(	Formatted: English (US)
•(	Formatted: English (US)
•	Formatted: English (US)
(	Formatted: English (US)
(	Formatted: English (US)
(	Formatted: English (US)
í	Formatted: English (US)

	References	Formatted: English (US)
	Andersen, T.: Detrital zircons as tracers of sedimentary provenance: limiting conditions from statistics and numerical	Formatted: English (US)
225	simulation, Chem. Geol., 216, 249–270, https://doi.org/10.1016/j.chemgeo.2004.11.013, 2005.	<b>Deleted:</b> (3–4),
	Attal, M. and Lavé, J.: Pebble abrasion during fluvial transport: Experimental results and implications for the evolution of the sediment load along rivers, J. Geophys. Res. Earth Surf., 114, https://doi.org/10.1029/2009JF001328, 2009.	Formatted: English (US)
		Formatted: English (US)
	Attal, M., Mudd, S. M., Hurst, M. D., Weinman, B., Yoo, K., and Naylor, M.: Impact of change in erosion rate and landscape	Deleted: .
230	steepness on hillslope and fluvial sediments grain size in the Feather River basin (Sierra Nevada, California), Earth Surf. Dyn., 3 201–222 https://doi.org/10.5194/esurf-3-201-2015 2015	Formatted: English (US)
200		<b>Deleted:</b> (3–4),
	Avdeev, B., Niemi, N. Ana and Clark, M. K.: Doing more with less: Bayesian estimation of erosion models with detrital	Formatted: English (US)
	thermochronometric data, Earth Planet. Sci. Lett., 305, 385–395, https://doi.org/10.1016/j.epsl.2011.03.020, 2011.	Deleted: .
	Brewer, I. D., Burbank, D. W., and Hodges, K. V.: Modelling detrital cooling-age populations: Insights from two Himalavan	Formatted: English (US)
	catchments, Basin Res., 15, 305-320, https://doi.org/10.1046/j.1365-2117.2003.00211.x, 2003.	Deleted: (3),
		Formatted: English (US)
235	Clinger, A. E., Fox, M., Balco, G., Cuffey, K <sub>rea</sub> and Shuster, D. L.: Detrital Thermochronometry Reveals That the Topography Along the Antarctic Peninsula is Not a Pleistocene Landscape L Geophys. Res. Earth Surf. 125, e2019IE005447	Formatted: English (US)
	https://doi.org/10.1029/2019JF005447, 2020.	Deleted:
		Formatted: English (US)
	Crameri, F., Shephard, G. F., and Heron, P. J.: The misuse of colour in science communication, Nat. Commun., 11, 1–10, https://doi.org/10.1028/641457-020.1016/7_2020	Deleted: (6),
	https://doi.org/10.1036/s+1+0/-020-17100-7, 2020.	Formatted: English (US)
240	van Dongen, R., Scherler, D., Wittmann, Haand von Blanckenburg, F.: Cosmogenic 10Be in river sediment: where grain size	Deleted: .
	matters and why, Earth Surf. Dyn., 7, 393–410, https://doi.org/10.5194/esurf-7-393-2019, 2019.	Formatted: English (US)
	Fhlers T. A. Szameitat A. Enkelmann F. Vanites B. I. and Woodsworth G. I. Identifying spatial variations in glacial	Deleted: (1),
	catchment erosion with detrital thermochronology, J. Geophys. Res. Earth Surf., 120, 1023–1039,	Formatted: English (US)
	https://doi.org/10.1002/2014JF003432, 2015.	Deleted: .
245	Enkalmann E and Ehlars T. A. Evaluation of detrival thermospherical and for quantification of classical actohment deputation	Formatted: English (US)
245	and sediment mixing. Chem. Geol., 411. 29–309. https://doi.org/10.1016/j.chemgeo.2015.07.018.2015.	Deleted: .
	······································	Formatted: English (US)
	von Eynatten, H., Tolosana-Delgado, R., and Karius, V.: Sediment generation in modern glacial settings: Grain-size and	Deleted: (6),
	source-rock control on sediment composition, Sediment. Geol., 280, 80–92, https://doi.org/10.1016/j.sedgeo.2012.03.008, 2012	Formatted: English (US)
	2012.	Deleted: .
250	Glotzbach, C., van der Beek, P., Carcaillet, I, and Delunel, R.: Deciphering the driving forces of erosion rates on millennial	Formatted: English (US)
	to million-year timescales in glacially impacted landscapes: An example from the Western Alps, J. Geophys. Res. Earth Surf.,	Deleted: .
	11a 1491-1515, https://doi.org/10.1002/jgft.2010/, 2013.	Formatted: English (US)
	Glotzbach, C., Busschers, F. S., and Winsemann, J.: Detrital thermochronology of Rhine, Elbe and Meuse river sediment	Deleted: (3),
0.5.5	(Central Europe): implications for provenance, erosion and mineral fertility, Int. J. Earth Sci., 107, 459-479,	Formatted: English (US)
255	nttps://doi.org/10.100//s00531-01/-1502-9, 2018.	Deleted:
	Hirt, W. H.: Petrology of the Mount Whitney Intrusive Suite, eastern Sierra Nevada, California: Implications for the	Formatted: English (US)
	emplacement and differentiation of composite felsic intrusions, GSA Bull., 119, 1185–1200, https://doi.org/10/bt7pbr, 2007.	Deleted: (2),
I		Formatted: English (US)

275	Hunter, J. D.: Matplotlib: A 2D Graphics Environment, Comput. Sci. Eng., 9, 90–95, https://doi.org/10.1109/MCSE.2007.55,		Deleted: (3),
	2007.		Formatted: English (US)
	Hurford, A. J. and Carter, A.: The role of fission track dating in discrimination of provenance, Geol. Soc. Lond. Spec. Publ.	Y	Formatted: English (US)
	57, 67-78, https://doi.org/10.1144/GSL.SP.1991.057.01.07, 1991.	(	Deleted: (1),
		(	Formatted: English (US)
280	bervations from tracer thermochronology, Earth Planet. Sci. Lett., 493, 140–149, https://doi.org/10/gdcf54, 2018.		Deleted: Kuiper, N. H.: Te circle, Indag. Math. Proc., 63 7258(60)50006-0, 1960.¶
	Lukens, C. E., Riebe, C. S., Sklar, L. S., and Shuster, D. L.: Sediment size and abrasion biases in detrital thermochronology,	$\langle \rangle \rangle$	Deleted:
	Earth Planet. Sci. Lett., 115929, https://doi.org/10.1010/j.epsi.2019.115929, 2020	, \\Y	Formatted: English (US)
	Madella, A., Glotzbach, C., and Ehlers, Todd. A.: ESD_thermotrace, A new software to interpret tracer thermochronometry	$\Lambda \gamma$	Formatted: English (US)
	datasets and quantify related confidence levels, GFZ Data Services, 2021.	$\langle \langle \rangle \rangle$	Deleted:
285	Malusà M. G. and Fitzgarald P. G. The geologic interpretation of the detriful thermochropology record within a stratigraphic	. \\ Y	Formatted: English (US)
205	framework, with examples from the European Alps, Taiwan and the Himalayas, Earth-Sci. Rev., 201, 103074,	<u>////</u>	Deleted: 2019
	https://doi.org/10.1016/j.earscirev.2019.103074, 2020.	{///}	Formatted: English (US)
	Maluad M. C. and Conventi E. The Sedimentaless of Detuited Thermoschereneless, in Fiscien Treats Thermoschereneless and		Deleted: .
	its Application to Geology, edited by: Malusa, M. G. and Fitzgerald, P. G., Springer International Publishing, Cham, 123–143,		Formatted: English (US)
290	https://doi.org/10.1007/978-3-319-89421-8_7, 2019.	- /}	Deleted: . https://doi.org/10
		Ì	Formatted: English (US)
	Malusa, M. G., Carter, A., Limoncelli, M., Villa, I. M., and Garzanti, E.: Bias in detrital zircon geochronology and thermochronometry. Chem. Geol. 359, 90–107. https://doi.org/10.1016/j.chem.geo.2013.09.016.2013	(	Formatted: English (US)
		$\bigtriangledown$	Deleted: .
	Malusà, M. G., Resentini, A, and Garzanti, E.: Hydraulic sorting and mineral fertility bias in detrital geochronology,	Y	Formatted: English (US)
	Gondwana Res., 31, 1–19, https://doi.org/10.1016/j.gr.2015.09.002, 2016.	$ \subset $	Deleted:
295	Massart, P.: The Tight Constant in the Dvoretzky-Kiefer-Wolfowitz Inequality, Ann. Probab., 18, 1269-1283,		Formatted: English (US)
	https://doi.org/10.1214/aop/1176990746, 1990.		Deleted: (3),
	Marrie E. L. The Walnussen Continuer Test for Cardinar of Ett. I. Am. Card. Acces. 46 (9.79)	Y	Formatted: English (US)
	Massey, F. J.: The Kolmogorov-Smirnov Test for Goodness of Fit, J. Am. Stat. Assoc., 40, 08–78, https://doi.org/10.1080/01621459.1951.10500769.1951.		Deleted: (253),
		1	Formatted: English (US)
	McPhillips, D. and Brandon, M. T.: Using tracer thermochronology to measure modern relief change in the Sierra Nevada,		
300	California, Earth Planet. Sci. Lett., 296, 373–383, https://doi.org/10.1016/j.epsl.2010.05.022, 2010.		<b>Deleted:</b> (3–4),
	Nibourel, L., Herman, F., Cox, S. C., Bevssac, Q., and Lavé, J.: Provenance analysis using Raman spectroscopy of		Formatted: English (US)
	carbonaceous material: A case study in the Southern Alps of New Zealand, J. Geophys. Res. Earth Surf., 120, 2056-2079,	$\sim$	Deleted:
	https://doi.org/10.1002/2015JF003541, 2015.	$\sum_{i=1}^{n}$	Formatted: English (US)
	Riebe C. S. Sklar, I. S. Lukens, C. F. and Shuster, D. L. Climate and tonography control the size and flux of sediment	1	Deleted: (10),
305	produced on steep mountain slopes, Proc. Natl. Acad. Sci., 201503567, https://doi.org/10.1073/pnas.1503567112, 2015.	× 1	Formatted: English (US)
	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		Deleted:
	Roda-Boluda, D. C., D'Arcy, M., McDonald, J., and Whittaker, A. C.: Lithological controls on hillslope sediment supply: insights from landslide activity and grain size distributions, Earth Surf. Process. Landf., 43, 956–977, https://doi.org/10.1002/esp.4281, 2018.	Y	Formatted: English (US)
1			

eleted: (3),
ormatted: English (US)
ormatted: English (US)
eleted: (1),
ormatted: English (US)
eleted: Kuiper, N. H.: Tests concerning random points on a rcle, Indag. Math. Proc., 63, 38–47, https://doi.org/10.1016/S1385- 258(60)50006-0, 1960.¶
eleted: .
ormatted: English (US)
ormatted: English (US)
eleted: .
ormatted: English (US)
eleted: 2019
ormatted: English (US)
eleted: .
ormatted: English (US)
eleted: . https://doi.org/10.5880/fidgeo.2021.003
ormatted: English (US)
ormatted: English (US)
eleted: .
ormatted: English (US)
eleted: .
ormatted: English (US)
eleted: (3),
ormatted: English (US)
eleted: (253),
ormatted: English (US)
eleted: (3-4),
ormatted: English (US)
erected: .
eleted: (10)
ereceu: (10),

Ruhl, K. W. and Hodges, K. V.: The use of detrital mineral cooling ages to evaluate steady state assumptions in active orogens: An example from the central Nepalese Himalaya: DETRITAL MINERAL COOLING AGES, Tectonics, 24, n/a-n/a, https://doi.org/10.1029/2004TC001712, 2005.

330 Sklar, L. S., Riebe, C. S., Genetti, J., Leclere, S., and Lukens, C. E.: Downvalley fining of hillslope sediment in an alpine catchment: implications for downstream fining of sediment flux in mountain rivers, Earth Surf. Process. Landf., 45, 1828– 1845, https://doi.org/10/gmpt7p, 2020.

Spiegel, C., Siebel, W., Kuhlemann, J. and Frisch, W.: Toward a comprehensive provenance analysis: A multi-method approach and its implications for the evolution of the Central Alps, in: Special Paper 378: Detrital thermochronology Provenance analysis, exhumation, and landscape evolution of mountain belts, Geological Society of America, 37–50, 2004.

Starke, J., Ehlers, T. Ara and Schaller, M.: Latitudinal effect of vegetation on erosion rates identified along western South America, Science, 367, 1358–1361, https://doi.org/10.1126/science.aaz0840, 2020.

Stock, G. M., Ehlers, T. A, and Farley, K. A.: Where does sediment come from? Quantifying catchment erosion with detrital apatite (U-Th)/He thermochronometry, Geology, 34, 725–728, 2006.

340 Vermeesch, P.: How many grains are needed for a provenance study?, Earth Planet. Sci. Lett., 224, 441–451, https://doi.org/10.1016/j.epsl.2004.05.037, 2004.

Vermeesch, P.: Quantitative geomorphology of the White Mountains (California) using detrital apatite fission track thermochronology, J. Geophys. Res., 112, https://doi.org/10.1029/2006JF000671, 2007.

Vermeesch, P.: On the visualisation of detrital age distributions, Chem. Geol., 312–313, 190–194, 345 https://doi.org/10.1016/j.chemgeo.2012.04.021, 2012.

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

Waskom, M., Gelbart, M., Botvinnik, O., Joel Ostblom, J., Hobson, P., Lukauskas, S., Gemperline, D. C., Augspurger, T., Halchenko, Y., Warmenhoven, J., Cole, J. B., de Ruiter, J., Vanderplas, J., Hoyer, S., Pye, C., Miles, A., Swain, C., Meyer,

350 K., Martin, M., Bachant, P., Quintero, E., Kunter, G., Villalba, S., Brian, B., Fitzgerald, C., Evans, C., Williams, M. L., O'Kane, D., Yarkoni, T., and Brunner, T.: mwaskom/seaborn: v0.11.1 (December 2020), Zenodo, https://doi.org/10.5281/zenodo.4379347\_2020.

Whipp, D. M. and Ehlers, T. A.: Quantifying landslide frequency and sediment residence time in the Nepal Himalaya, Sci. Adv., 5, eaav3482, https://doi.org/10.1126/sciadv.aav3482, 2019.

355 Whipp, D. M., Ehlers, T. A., Braun, J., and Spath, C. D.: Effects of exhumation kinematics and topographic evolution on detrital thermochronometer data, J. Geophys. Res., 114, https://doi.org/10.1029/2008JF001195, 2009.

Formatted: English (US)	
Deleted: (4),	

Formatted: English (US)

	Formatted: English (US)
	Deleted: .
	Formatted: English (US)
Ì	Formatted: English (US)
$\langle \rangle$	<b>Deleted:</b> vol. 378, pp. 37–50,
()	Formatted: English (US)
	Deleted: http://specialpapers.gsapubs.org/cgi/doi/10.1130/0-8137- 2378-7.37, last access: 7 July 2016
	Formatted: English (US)
	Deleted: .
	Formatted: English (US)
	<b>Deleted:</b> (6484),
	Formatted: English (US)
	Deleted: .
	Formatted: English (US)
	Deleted: (9),
	Formatted: English (US)
	<b>Deleted:</b> (3–4),
	Formatted: English (US)
	Deleted: (F3),
	Formatted: English (US)
	Deleted: .
Ò	Formatted: English (US)
Ò	Deleted: .,
Ì	Formatted: English (US)
$\langle \rangle$	Deleted: (4),
)	Formatted: English (US)
	Formatted: English (US)
	Formatted: English (US)

Page 1: [1] Formatted	AM	9/30/21 8:25:00 PM	
English (US)			
Page 1: [2] Formatted	AM	9/30/21 8:25:00 PM	
English (US)			
Page 1: [3] Formatted	AM	9/30/21 8:25:00 PM	
English (US)			
Page 1: [3] Formatted	AM	9/30/21 8:25:00 PM	
English (US)			
Page 1: [4] Formatted	AM	9/30/21 8:25:00 PM	
English (US)			
Page 1: [4] Formatted	AM	9/30/21 8:25:00 PM	
English (US)			
Page 1: [5] Formatted	AM	9/30/21 8:25:00 PM	
English (US)			
Page 1: [6] Formatted	AM	9/30/21 8:25:00 PM	
English (US)			
Page 1: [7] Formatted	AM	9/30/21 8:25:00 PM	
English (US)			
Page 1: [8] Formatted	AM	9/30/21 8:25:00 PM	
English (US)			
Page 1: [9] Formatted	AM	9/30/21 8:25:00 PM	
English (US)			
Page 1: [10] Formatted	AM	9/30/21 8:25:00 PM	
English (US)			
Page 1: [11] Formatted	AM	9/30/21 8:25:00 PM	
English (US)			
Page 1: [12] Formatted	AM	9/30/21 8:25:00 PM	
English (US)			
Page 1: [12] Formatted	AM	9/30/21 8:25:00 PM	
English (US)			
Page 1: [13] Formatted	AM	9/30/21 8:25:00 PM	

English (US)

Page 1: [14] Formatted	AM	9/30/21 8:25:00 PM
English (US)		
Dago 1. [15] Formattod	A.M.	0/20/21 8-25-00 DM
Findlish (US)	АМ	9/30/21 8:23:00 PM
English (00)		
Page 1: [16] Deleted	AM	9/30/21 8:25:00 PM
<b>V</b>		
Page 1: [17] Formatted	АМ	9/30/21 8:25:00 PM
English (US)		
Page 1: [18] Formatted	AM	9/30/21 8:25:00 PM
English (US)		
Page 1: [18] Formatted	AM	9/30/21 8:25:00 PM
English (US)		
Page 1: [18] Formatted	AM	9/30/21 8:25:00 PM
English (US)		
Page 1: [19] Formatted	AM	9/30/21 8:25:00 PM
English (US)		
D 4. [20] F	4.14	0/20/21 0-25-00 PM
Findlish (US)	АМ	9/30/21 8:25:00 PM
Lightin (05)		
Page 1: [21] Formatted	AM	9/30/21 8:25:00 PM
English (US)		
Page 1: [22] Deleted	۸M	9/30/21 8·25·00 PM
Page I. [22] Deleteu	API	9/30/21 0.23.00 FM
<b>T</b>		
Page 1: [23] Formatted	AM	9/30/21 8:25:00 PM
English (US)		
Page 1: [24] Deleted	AM	9/30/21 8:25:00 PM
V		
Page 1: [24] Deleted	AM	9/30/21 8:25:00 PM
۲		
Page 1: [25] Formatted	AM	9/30/21 8:25:00 PM
English (US)		
Page 14: [26] Deleted	AM	9/30/21 8:25:00 PM

**v**....

Page 14: [26] Deleted	AM	9/30/21 8:25:00 PM	]
▼			
Page 14: [27] Deleted	ΔМ	9/30/21 8·25·00 PM	1
		5/50/21 0.25.00 FM	1
·			_
Page 14: [27] Deleted	AM	9/30/21 8:25:00 PM	
▼			
Page 14: [28] Deleted	AM	9/30/21 8:25:00 PM	1
v			-
			1
Page 14: [28] Deleted	AM	9/30/21 8:25:00 PM	J
۲			
Page 14: [28] Deleted	AM	9/30/21 8:25:00 PM	]
▼			
Dana 14 [20] Dalatad	A. 1.4	0/20/21 8-25-00 PM	1
Page 14: [28] Deleted	AM	9/30/21 8:25:00 PM	]
V			
Page 14: [28] Deleted	AM	9/30/21 8:25:00 PM	]
▼			
Page 14: [28] Deleted	AM	9/30/21 8:25:00 PM	1
V			4
			٦
Page 14: [28] Deleted	AM	9/30/21 8:25:00 PM	J
▼			
Page 14: [28] Deleted	AM	9/30/21 8:25:00 PM	1
۷			
Dago 14: [29] Deleted	A.M.	0/20/21 8-25-00 DM	1
Page 14: [26] Deleted	AM	9/30/21 8:25:00 PM	J
<b>V</b>			
Page 14: [29] Deleted	AM	9/30/21 8:25:00 PM	]
▼			
Page 14: [29] Deleted	АМ	9/30/21 8:25:00 PM	1
			4
			1
Page 14: [29] Deleted	AM	9/30/21 8:25:00 PM	J
▼			
Page 14: [29] Deleted	AM	9/30/21 8:25:00 PM	1

Page 14: [29] Deleted	AM	9/30/21 8:25:00 PM	
▼			 
			-
Page 14: [30] Deleted	AM	9/30/21 8:25:00 PM	
۲			
Page 14: [20] Deleted	AM	0/20/21 9:25:00 DM	1
Page 14: [50] Deleted	АМ	9/30/21 8:25:00 PM	
▼			 
Page 14: [31] Deleted	AM	9/30/21 8:25:00 PM	1
			-
¥			
Page 14: [31] Deleted	АМ	9/30/21 8:25:00 PM	
▼			 
			-
Page 14: [31] Deleted	AM	9/30/21 8:25:00 PM	
▼			
Page 14: [31] Deleted	ΔМ	9/30/21 8:25:00 PM	1
-			-
Υ			
Page 14: [31] Deleted	АМ	9/30/21 8:25:00 PM	
۲			 
			-
Page 14: [31] Deleted	AM	9/30/21 8:25:00 PM	
<b>V</b>			
Page 14: [31] Deleted	ΔМ	9/30/21 8:25:00 PM	1
	A.1	5756721 0125100 TH	-
•			 
Page 14: [31] Deleted	AM	9/30/21 8:25:00 PM	
▼			 
			-
Page 14: [31] Deleted	AM	9/30/21 8:25:00 PM	
۲			 
Page 14: [31] Deleted	ΔΜ	9/30/21 8·25·00 PM	1
-	A''I	5, 50, 22 5125100 T PI	
ν			
Page 14: [31] Deleted	АМ	9/30/21 8:25:00 PM	1
V			 _
			 -
Page 14: [31] Deleted	AM	9/30/21 8:25:00 PM	

Page 14: [31] Deleted	AM	9/30/21 8:25:00 PM	
▼			
Page 14: [31] Deleted	АМ	9/30/21 8:25:00 PM	1
v			
			-
Page 17: [32] Deleted	AM	9/30/21 8:25:00 PM	
۲			
Page 17: [32] Deleted	AM	9/30/21 8:25:00 PM	
▼			
Page 17: [32] Deleted	ΔМ	9/30/21 8·25·00 PM	1
	A.1	5,56,22 6125166 111	
•			
Page 17: [32] Deleted	AM	9/30/21 8:25:00 PM	
▼			
Page 17: [32] Deleted	AM	9/30/21 8:25:00 PM	
۲			
Page 17: [32] Deleted	АМ	9/30/21 8:25:00 PM	1
·		-,,	
Page 17: [32] Deleted	AM	9/30/21 8:25:00 PM	
Υ			
Page 17: [32] Deleted	AM	9/30/21 8:25:00 PM	
۲			
Page 17: [32] Deleted	ΔМ	9/30/21 8:25:00 PM	1
			-
Page 17: [32] Deleted	AM	9/30/21 8:25:00 PM	
۲			
Page 17: [32] Deleted	AM	9/30/21 8:25:00 PM	
▼			
Page 17: [32] Deleted	۵м	9/30/21 8:25:00 PM	
· age 171 [02] Deleted	AI'I		
			_
Page 17: [32] Deleted	AM	9/30/21 8:25:00 PM	
▼			
Page 17: [33] Deleted	AM	9/30/21 8:25:00 PM	1

Page 17: [33] Deleted	AM	9/30/21 8:25:00 PM	
▼			
Page 17: [33] Deleted	AM	9/30/21 8:25:00 PM	
			-
Page 17: [33] Deleted	AM	9/30/21 8:25:00 PM	
۲			
Page 17: [33] Deleted	AM	9/30/21 8:25:00 PM	
▼			
Page 17: [33] Deleted	ΔМ	9/30/21 8·25·00 PM	
		5/56/21 6/26/00 111	
•			
Page 17: [33] Deleted	AM	9/30/21 8:25:00 PM	
▼			
Page 17: [33] Deleted	AM	9/30/21 8:25:00 PM	
۲			
Page 17: [33] Deleted	АМ	9/30/21 8:25:00 PM	
·			
			_
Page 17: [33] Deleted	AM	9/30/21 8:25:00 PM	
۷			
Page 17: [33] Deleted	AM	9/30/21 8:25:00 PM	
۲			
Page 17: [33] Deleted	ΔМ	9/30/21 8:25:00 PM	
Page 17: [33] Deleted	AM	9/30/21 8:25:00 PM	
۲			
Page 17: [33] Deleted	AM	9/30/21 8:25:00 PM	
▼			
Page 17: [33] Deleted	۵м	9/30/21 8:25:00 PM	
· age 171 [00] Deleted	A11		
			_
Page 17: [33] Deleted	AM	9/30/21 8:25:00 PM	
▼			
Page 17: [33] Deleted	AM	9/30/21 8:25:00 PM	

Page 17: [33] Deleted	АМ	9/30/21 8:25:00 PM	
Υ			
Page 17: [33] Deleted	AM	9/30/21 8:25:00 PM	
▼			
Page 17: [33] Deleted	AM	9/30/21 8:25:00 PM	
▼			
Page 17: [34] Deleted	AM	9/30/21 8:25:00 PM	
۲			
Dage 17: [24] Delated	A 14	0/20/21 9-25-00 DM	
Page 17: [54] Deleted	AM	9/30/21 8:25:00 PM	
▼			
Page 17: [34] Deleted	ΔΜ	9/30/21 8·25·00 PM	
rage 17. [54] Deleted		5/50/21 0.25.00 FM	
▼			
Page 17: [34] Deleted	АМ	9/30/21 8:25:00 PM	
τ			
Page 17: [34] Deleted	АМ	9/30/21 8:25:00 PM	
<b>V</b>			•
Page 17: [34] Deleted	AM	9/30/21 8:25:00 PM	
▼			
Page 17: [34] Deleted	AM	9/30/21 8:25:00 PM	
▼			
Page 17: [34] Deleted	AM	9/30/21 8:25:00 PM	
τ			