# A Tool for the Ages: The Probabilistic Cosmogenic Age Analysis Tool (P-CAAT)

Jason M. Dortch[1*], Matt D. Tomkins[2,3], Sourav Saha[1], Madhav K. Murari[4], Lindsay M. Schoenbohm[5], Doug Curl[1]

[1] University of Kentucky, Kentucky Geological Survey, Lexington, KY 40506, USA
[2] Department of Geography, School of Environment, Education and Development, University of Manchester, Manchester, M13 9PL, U.K.
[3] Cryosphere Research at Manchester, Manchester, U.K.
[4] National Geochronology Facilities, Inter-University Accelerator Centre, New Delhi 110067, India
[5] University of Toronto Mississauga, Mississauga, ON, M6S 3N6, Canada

**Keywords:**

Terrestrial cosmogenic nuclide dating, Uncertainty, Outlier detection, Clustering, Landform ages, Probability density estimate

**Abstract**

While revolutionary to the geomorphic community, the application of terrestrial cosmogenic nuclide (TCN) dating is complicated by geological uncertainties, which often lead to skewed or poorly clustered TCN age distributions. Although a range of statistical approaches are typically used to detect and remove outliers, few are optimized for analysis of TCN datasets. Many are mean- or median-based and therefore explicitly assume a single probability distribution (e.g., Mean Squared Weighted Deviates, Chauvenet's Criterion, etc.). Given the ubiquity of pre- and post-depositional modification of rock surfaces, which occur at different rates in different geomorphic settings, these approaches struggle with multimodal distributions which often characterize TCN datasets. In addition, most statistical approaches do not propagate measurement or production rate uncertainties, which become increasingly important as dataset size or clustering increases. Finally, most approaches provide arithmetic single solutions, irrespective of geologic context.

To address these limitations, we present the Probabilistic Cosmogenic Age Analysis Tool (P-CAAT), a new approach for outlier detection and landform age analysis. This tool incorporates both sample age and geologic uncertainties and uses Monte Carlo simulations to eliminate dataset skewness by isolating component normal distributions from a cumulative probability density estimate for datasets with three or more samples. This approach allows geologic context to inform post-analysis interpretations, as researchers can assign landform ages based upon statistically distinct subpopulations, informed by the characteristics of geomorphic systems (e.g., exhumation of boulders as moraines degrade through time). To evaluate the effectiveness of P-CAAT, we analyzed a range of synthetic TCN datasets and compared the results to commonly used statistical approaches for outlier detection. Irrespective of dataset size or clustering, P-CAAT outperformed other approaches and returned accurate solutions that improve in precision as sample size increases. To enable more comprehensive utilization of our approach, P-CAAT is packaged with a GUI interface and is available for download at kgs.uky.edu/anorthite/PCAAT.

## 1. Introduction

Terrestrial cosmogenic nuclide (TCN) dating has enabled quantification of the frequencies, magnitudes, and timescales of geomorphic processes by permitting direct age analysis of erosional and depositional landforms (Nishiizumi et al., 1986, 1989; Phillips et al., 1990; Cerling et al., 1994; Molnar et al., 1994; Burbank et al., 1996). The application of this method is complicated by “geologic uncertainty,” however, in which pre- or post-depositional modification of rock surfaces results in scattered or poorly clustered TCN age distributions for individual landforms (Putkonen and Swanson, 2003; Heyman et al., 2011; Dortch et al., 2010b, c, 2011b; Balco, 2011; Hein et al., 2014).

Common examples of geologic processes that influence TCN concentrations include erosion of an exposed surface or post-depositional exhumation of boulders (Gosse and Phillips, 2001), denudation of landforms such as moraines (Hallet and Putkonen, 1994; Putkonen and O'Neal, 2006; Putkonen et al., 2008; Tomkins et al., 2021), nuclide inheritance caused by prior exposure (Putkonen and Swanson, 2003), and post-depositional shielding (Dehnert and Schlüchter, 2008) and reworking (D'arcy et al., 2019). Moreover, these processes operate and vary in significance at the sub-landform scale, as individual boulders can be eroded, toppled, shielded, and exhumed as landforms degrade through time (Hallet and Putkonen, 1994; Briner et al., 2005; Ivy-Ochs et al., 2007). Bedrock can also be differentially eroded through abrasion or quarrying (Hallet, 1996; Briner and Swanson, 1998; Dühnforth et al., 2010; Iverson, 2012; Ugelvig et al., 2018), and snow can be reworked and redeposited, leading to differential shielding (Schildgen et al., 2005). Although there are criteria for field identification and exclusion of surfaces compromised by geologic uncertainty (Gosse et al., 1995; Ivy-Ochs et al., 2007; Dortch et al., 2010a; Akçar et al., 2011; Heyman et al., 2016; Tomkins et al., 2021), the magnitude and direction of this influence is often difficult to predict based on observable geomorphic evidence alone (Dortch et al., 2013; Murari et al., 2014).

Quantitative analysis of exposure-age clustering on an individual landform is typically used to account for these processes. Standard statistical methods for identifying geologic outliers include iterative reduced chi-squared (Small and Fabel, 2016), mean square weighted deviates (Douglass et al., 2006), probability density estimates (Dortch et al., 2013; Stübner et al., 2021), generalized extreme Studentized deviates (Jones et al., 2019), Chauvenet's criterion (Rinterknecht et al., 2006), and 1σ or 2σ uncertainty overlap (Chevalier et al., 2011). Few of these methods are optimized to analyze TCN datasets, however, as such datasets are characterized by unique sample age and geologic uncertainties. Moreover, rigorous statistical assessment is often hindered by small sample sizes resulting from the expense of TCN dating, rarely meeting typical ($n = 10^2 – 10^3$) or even minimum sample sizes ($n = 30$) following the central limit theorem (Borradaile, 2003; Balco, 2011; Kwak and Kim, 2017).

Landform ages based on poorly clustered, skewed, or statistically few ($n < 30$) TCN exposure ages are extremely sensitive to the choice of statistical test, which can significantly affect age interpretation (Applegate et al., 2010, 2012; Chevalier et al., 2005a, b; Brown et al., 2005). This variability can encourage qualitative outlier identification and removal, without statistical justification *(see* Balco, 2011, for further discussion). As a result, a statistically robust standardized approach for TCN age outlier identification is necessary to enable accordant comparison between studies, reproducibility, and to minimize uncertainty in landform-age analysis introduced by different outlier identification techniques (Barrows et al., 2007, 2008; Applegate et al., 2008).

To this end, we present a new probability-based approach for outlier identification and landform-age analysis: The Probabilistic Cosmogenic Age Analysis Tool. P-CAAT is a standalone program, coded in MATLAB, and is freely available for Windows and Mac OS users at kgs.uky.edu/anorthite/PCAAT. The P-CAAT approach improves on previous statistical methods by:

i. Addressing "sample age" uncertainty by incorporating internal measurement uncertainties in a numerically generated composite probability density estimate (PDE) and propagating external measurement uncertainties, the latter of which are typically excluded from commonly utilized statistical approaches.

ii. Addressing "geologic" uncertainty by analyzing the clustering of ages on individual landforms by separating a series of normal distributions (component Gaussians) from the composite PDE to isolate skew.

iii. Quantifying uncertainty for component Gaussians, a step that improves upon the previous skewness-based approach of Applegate et al. (2010, 2012).

iv. Allowing geologic context to inform landform-age analysis through component Gaussian selection. Component Gaussians represent statistically distinct subpopulations and can be used to assign landform ages based upon the characteristics of the studied geomorphic system and likelihood of pre- or post-depositional modification of sampled surfaces. For example, the youngest TCN subpopulation (component Gaussian) may be preferred for TCN dating of alluvial fan surfaces where reworking is dominant (e.g., D'arcy et al., 2019; Saha et al., 2021).

More broadly, P-CAAT improves upon qualitative-only (subjective) approaches, which lack statistical justification (Balco, 2011), and quantitative-only approaches, which typically provide arithmetic single solutions, irrespective of geologic context. To encourage a broader application of our approach, we have included a description of the P-CAAT program, instructions for use, comparative testing against other methods, and rigorous testing using synthetic exposure-age datasets.

## 2. Common statistical approaches for outlier identification

A range of quantitative methods are available to identify geologic outliers in TCN exposure-age datasets. To evaluate P-CAAT performance, several methods were utilized in the analysis of synthetic-age datasets (Section 5.1) for comparison. These include:

- Mean Squared Weighted Deviates (MSWD) and Weighted Mean Squared Weighted Deviates (W-MSWD), which are methods based on an iterative reduced chi-squared approach in which outliers with the highest deviation are removed sequentially until the statistical indicator approximates a value of 1. Generally, TCN-based studies that utilize (W)MSWD methods use the (weighted) standard deviation for uncertainty estimates without calculating asymmetric sigma bounds to validate the statistical indicator (Kaplan and Miller., 2003; Douglass et al., 2006; Dortch et al., 2010a, b, c, 2011a, b; Heyman et al., 2011). We calculated both asymmetrical sigma bounds above and below the test statistic value, following the methods of Wendt and Carl (1991) for validation.

- Generalized Extreme Students Deviates (gESD), which is an iterative test that assumes a normal distribution and eliminates outliers (the most extreme data points with respect to the mean) to reduce Rosner's test statistic (Rosner, 1983). The maximum number of outliers is set at $n - 1$ (where n is the number of samples). The best results are typically obtained from larger sample sizes ($n \geq 15$). At the same time, the weighted mean and weighted standard deviation of the remaining ages are typically used to represent the age of the landform (Jones et al., 2019).

- Chauvenet's Criterion, which calculates a t-value for each exposure age as a function of the difference between the age and the mean divided by the standard deviation. Outliers are identified by comparing t-values to a maximum allowable deviation (e.g., at a 95 percent confidence interval) (Putnam et al., 2013a, b). This process is iterated until all t-values fall within the maximum allowable deviation (Taylor, 1997; Rinterknecht et al., 2006; Dunai, 2010; Saha et al., 2018, 2019).

- Two Standard Deviations from the Mean (henceforth referred to as 2-SD), which is calculated by taking the deviation of all ages in the dataset. Any age that falls outside two-sigma from the mean is considered an outlier and is removed from the dataset before the final mean and standard deviation are calculated (Putnam et al., 2013a). A weighted 2-SD approach can also be undertaken (Blisniuk et al., 2010), although we did not undertake this variation because of intrinsic collinearity between exposure ages and their uncertainties (i.e., as exposure age increases, uncertainty increases), a situation that biases landform ages toward younger ages with smaller uncertainties (Ivy-Ochs et al., 2007).

- Two Mean Absolute Deviations from the Median (henceforth referred to as 2-MAD), which is similar to 2-SD but uses the median as the cluster center and the mean absolute deviation (MAD) as the outlier detection limit. This approach is optimal for skewed datasets because both the median and the MAD are less sensitive to outlier bias (Leys et al., 2013). In contrast, standard deviation-based methods are more effective for datasets that initially conform to a normal distribution. After outliers are removed, the median and MAD of the remaining ages can be used to represent the age of the landform. Although not used widely in TCN studies (Menounos et al., 2017; Darvill et al., 2018), it has been demonstrated to be effective in removing outliers in other quantitative studies (Leys et al., 2013).

- 2σ Overlap of Age Uncertainty (henceforth referred to as 2σ-overlap), which identifies outliers as an age that does not overlap with any other age at 2σ uncertainty limits (Davies et al., 2020). The mean and standard deviation of the remaining ages represent the age of the landform. Although this method is attractive because of its computational ease, the results are typically conservative, and its use is limited to identifying extreme outliers.

- The Press (1997) method (henceforth referred to as Press), as implemented by Muzikar et al. (2017) and Goehring et al. (2018), utilizes a probability-based (Bayesian) approach to assign weights to individual data points (exposure ages or concentrations) based on their likelihood of being correct and returns a weighted mean age and finding sigma form the distribution. This approach bypasses the common problem of collinearity between age and uncertainty, which can unfairly bias the mean towards younger ages. The selection of the standard deviation parameter (s) is crucial (Muzikar et al., 2017); this is modified until the final probability stabilizes (β). Further, β should not be too small as this would indicate that broad Gaussians are making a major contribution to the results (Muzikar et al., 2017), indicating poor handling of overdispersion.

- Probability Density Estimates (henceforth referred to as PDE), which use a smoothing window defined by a numeric bandwidth (Silverman, 1986) to generate a composite PDE through the summation of individual age-uncertainty distributions. The resulting PDE is typically a multimodal curve in which the highest peak can be interpreted to identify the landform deposition event (Kelly et al., 2008). PDEs can be calculated before and after outliers are removed using alternative methods such as MSWD (Douglass et al., 2006) or gESD (Jones et al., 2019) to obtain a density estimate that conforms to or approaches a normal distribution. There are several limitations to this approach, however. For example, interpretation of the PDE is typically subjective and not formal, uncertainties are not estimated directly from the density distribution, numeric bandwidths are typically not discussed or reported, and the normality of the age estimate is not discussed or quantified. These limitations have been addressed over the decade long development period of P-CAAT, based on preliminary work by Dortch et al. (2013) and Murari et al. (2014). Recently, Stübner et al. (2021) developed a Python implementation (henceforth referred to as S-PDE) inspired by Dortch et al. (2013) PDE approach; this method is included for comparison.

### 2.1. Alternative approaches for outlier identification

In addition to the methods listed above, we did not consider some alternative approaches for outlier identification. These include:

- The skewness approach of Applegate et al. (2010, 2012), is a Monte Carlo–based estimator that assesses exposure-age distributions to determine if the dominant form of scatter is a result of post-depositional modification of rock surfaces (e.g., negative skew caused by degradation or exhumation) or pre-depositional processes (e.g., positive skew caused by nuclide inheritance). Although this approach has been shown to be effective at predicting the “true” age of synthetic datasets, it does not allow for quantification of uncertainty. This precludes the widespread use of this approach for landform-age analysis.

- Extreme estimator approaches, such as the oldest-boulder method for degraded landforms (Putkonen and Swanson, 2003; Briner et al., 2005; Delmas et al., 2008; Allard et al., 2020) and the youngest-boulder method (Gosse, 2005; Benson et al., 2005) for landscapes characterized by minimal erosion during glacial cycles (< 3 m) or in which there are long ice-free periods between brief glacial maxima (Briner et al., 2016); such scenarios increase the probability of nuclide inheritance. These methods require the exclusion of obvious outliers *(see* Benson et al., 2005), but the qualitative threshold for determining an “obvious” outlier varies between users. Similarly, the selection of an extreme estimator is linked to the relative probability of pre- or post-depositional modification in the geomorphic system and identifying the frequency of these processes from preserved geomorphic evidence alone is a significant challenge. Although extreme estimators may be appropriate under specific circumstances (Applegate et al., 2010, 2012), we did not consider these approaches.

- Researchers routinely used pre-screening based on landform (e.g., moraine sedimentology; Zreda et al., 1994; Putkonen and O'Neal, 2006; Pallàs et al., 2010; Tomkins et al., 2021) or surface characteristics (e.g., boulder height – Heyman et al. 2016; boulder weathering – Tylmann et al., 2018) to identify and exclude possible geologic outliers. Although many of the applied criteria are theoretically sound, few have been tested quantitatively. One exception is the positive correlation between boulder height and TCN clustering (Heyman et al., 2016), although the overall effect was minor, as a dominant fraction (> 50 percent) of tall boulder groups were sufficiently scattered to fail a reduced chi-square test ($\chi^2 \leq 2$). Because pre-screening is applied prior to sample collection and analysis of the resulting TCN ages, comparison with other statistical approaches was not possible here. More effective pre-screening of geologic outliers could, however, play an essential role in simplifying subsequent statistical analysis.

## 2.2. Good practice for calculating landform-age uncertainty

Irrespective of the choice of statistical approach for outlier identification, there is a clear need to standardize uncertainty reporting. This is particularly important for regional- or global-scale reanalysis of exposure-age datasets, as overestimation of uncertainty can result in inappropriate grouping of distinct events, and underestimation of uncertainty can result in separation of synchronous events or overinterpretation of seemingly correlative events.

We distinguish two quantitatively different and independent forms of uncertainty (Fig. 1). They are:

i. Sample age uncertainty (SAU), which incorporates two distinct forms of uncertainty. These include:

    a. Internal uncertainty, which incorporates errors in sample processing or the ability of an accelerator mass spectrometer (AMS) to reproduce a standard (Jull et al., 2015).

    b. External uncertainty, which incorporates TCN production rate (Lal, 1991; Stone 2000; Stroeven et al., 2015; Borchers et al., 2016; Marrero et al., 2016), scaling scheme (Lifton et al., 2014; 2016) and atmospheric model uncertainty (Uppala et al., 2005).

ii. Geologic uncertainty (GU), which incorporates a range of pre- and post-depositional processes that modify rock-surface TCN concentrations, typically expressed by scatter around a "true" landform age.

Uncertainty reporting for the methods discussed above, with the exception of PDE approaches, primarily use deviation from the mean to quantify geologic uncertainty (GU; Fig. 1):

$$GU = Standard\ Deviation \mid Mean\ Absolute\ Deviation$$

Most TCN studies do not propagate sample age uncertainty, however there are exceptions (Martin et al., 2020. SAU can be quantified by calculating the root-mean-square-error (Taylor and Kuyatt, 1994) of reported internal age uncertainties as follows:

$$SAU = \frac{\sqrt{Sum\ of\ the\ squared\ errors}}{Number\ of\ observations}$$

Total landform-age uncertainty (*t*) can be calculated through summation in quadrature as follows:

$$t = \sqrt{SAU^2 + GU^2}$$

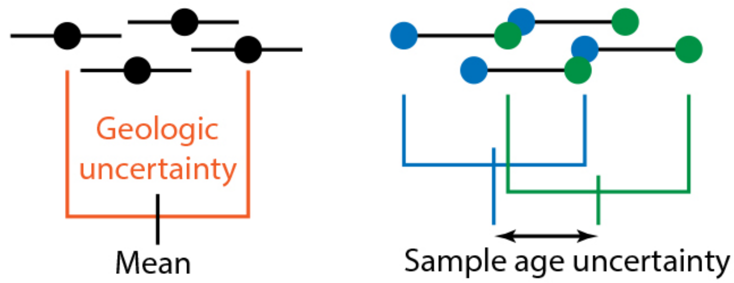


**Fig. 1.** Schematic diagram to illustrate the difference between geologic (GU) and sample age uncertainties (SAU). GU is typically expressed by scatter around a "true" landform age (e.g., distribution around the mean), although the magnitude and direction of this influence is often difficult to predict based on geomorphic evidence alone. In contrast, the magnitude of SAU is typically characterized as normal and consistent at the landform scale. Illustrated here is a worst-case scenario in which all ages have been shifted older (blue → green), for example due to a change in calculated TCN production rate (Borchers et al., 2016) or a systematic difference between instrument results (Small and Fabel, 2016; Putnam et al., 2019).

Typically, SAU is significantly smaller than GU, but SAU becomes increasingly important as clustering improves or sample size increases. The quadratic approach presented here would be suitable for propagating SAU into the results given from (W)MSWD, gESD, Chauvenet's Criterion, 2-SD, 2-MAD, 2σ-overlap, and P-CAAT when internal uncertainties are used to calculate the cumulative PDE. To our knowledge, a PDE based on external uncertainties is the only approach that inherently accounts for both the distribution of TCN ages (GU) and sample age uncertainty (SAU) simultaneously. If SAU is not propagated, it is possible to calculate a total landform-age uncertainty ($t$) that is lower than AMS precision for very tightly clustered datasets *(see* Jull et al., 2015).

When analyzing data from a single landform or site, outlier identification should be based on internal uncertainties only, regardless of the statistical method chosen. After statistical analysis and outlier removal (Balco et al., 2008; *see* Section 3.3), external uncertainties should be propagated to compare landforms in distinct regions and latitudes (Clark et al., 2009) and to integrate results with other geochronological methods (e.g., luminescence or $^{14}C$ dating). In contrast, external uncertainties should be used for outlier identification only for ages on a single landform that were measured using different instruments (Small and Fabel, 2016; Putnam et al., 2019) or nuclides (e.g., $^{10}Be$ and $^{36}Cl$; Wilson et al., 2013).

## 3. P-CAAT tool description

Our Probabilistic Cosmogenic Age Analysis Tool is a new statistical approach for outlier detection and landform-age analysis. P-CAAT generates a composite PDE based upon individual age-uncertainty distributions and then undertakes a series of modelling steps to isolate component normal (Gaussian) distributions. P-CAAT differentiates itself from previous PDE approaches by incorporating robust bandwidth estimators (*see* Section 3.1), breaking down density estimates into true Gaussian components and analyzing them quantitatively to estimate uncertainty. This approach is attractive because many processes in nature follow a normal distribution, and numerous observations of a process generally follow the central limit theorem (Kwak and Kim, 2017). In turn, if a viable sample size is achieved, components that are younger (e.g., from erosion, exhumation, or shielding) or older (e.g., from inheritance) than the "true" age of the landform (Heyman et al., 2011) can be isolated. Most importantly, P-CAAT enables Gaussian choice based on evidence provided by geologic context, which can provide more consistent results than arithmetic single-solution approaches (e.g., MSWD, gESD, etc.). P-CAAT is distinct from the typical approach of qualitative identification and removal of outliers (*see* Balco, 2011), as component Gaussians represent statistically distinct, normally distributed (single event) subpopulations.

P-CAAT was developed in the MATLAB environment and uses a weighted ksdensity kernel smoothing function in MATLAB to generate a PDE based on input exposure ages, their clustering, and their uncertainties. Weights (w) are based on inverse age precision:

$$w = \left(\frac{internal\ uncertainty}{exposure\ age}\right)^{-1}$$

which eliminates collinearity issues, as noted by Ivy-Ochs et al. (2007), while incorporating sample age measurement uncertainties into PDE generation.

### 3.1. Bandwidth estimation

Bandwidth estimation for PDE generation is critical; a bandwidth that is too wide will over-smooth the PDE and one that is too narrow will over-fit peaks to the noise in the data (Fig. 2). A numeric bandwidth is typically calculated using a bandwidth estimator, which returns distinct PDEs as a function of the size and clustering of the input data. Widely used examples can be found in Silverman (1986), Sheather and Jones (1991), and Scott (2015).

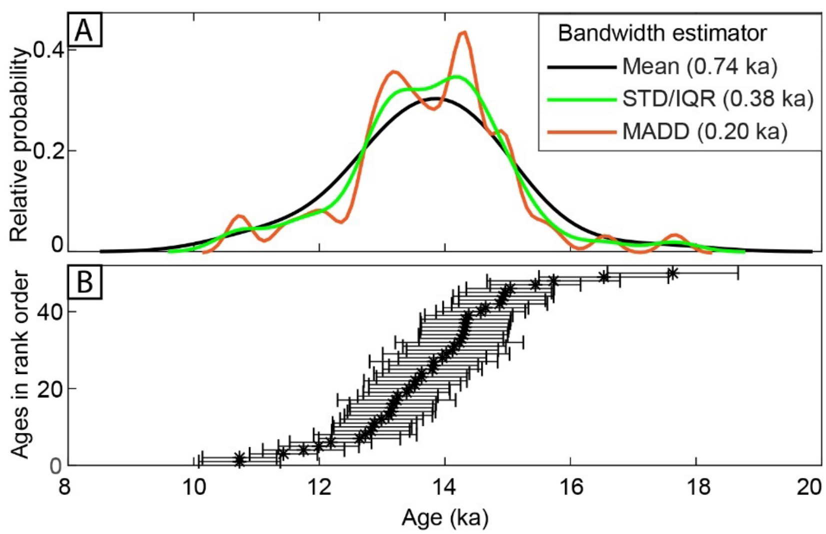


**Fig. 2.** An example of how bandwidth estimator choice affects probability density estimates assuming ages are from a single landform. (A) Probability density estimates using the mean (black line), STD/IQR (green line), and MADD (mean absolute Dortch deviants; red line) bandwidth estimators with numerical solutions in parentheses (*see* below for detailed descriptions). (B) The distribution of exposure ages (ka ± internal uncertainty; n = 50; sorted in rank order; Dortch et al., 2013). The dataset mean and standard deviation is 13.7 ± 1.3 ka, with a range of 6.9 ka and an interquartile range of 1.2 ka. The large number of samples and high degree of uncertainty overlap causes the Mean bandwidth estimator to over-smooth the data into a single peak. In contrast, the MADD (mean absolute Dortch deviants) bandwidth estimator under-smooths the data due to the very tight clustering 12.5-14.5 ka, leading to an overly complex multimodal PDE. The STD/IQR estimator separates the main body of data centered on ~14 ka into two peaks, fitting with prior knowledge of how glacial moraines degrade through time (Dortch et al., 2013).

There is no single agreed-upon theorem or method for bandwidth estimation, however. Moreover, because the performance of an individual bandwidth estimator is intimately linked to the input data, the actual numeric bandwidth used can be highly variable. Variability in age clustering and skewness, in combination with individual age uncertainties, make automatically determining a best-fit bandwidth under changing scenarios difficult. Additionally, the choice of bandwidth estimator is critical, because some datasets cannot be solved with all estimators (i.e., the model fails to converge on a solution).

To address these issues, three bandwidth estimators were selected and incorporated into P-CAAT. They are:

**Option 1** Mean: uses the arithmetic mean of the internal age uncertainties (*iu*) as the numeric bandwidth (*bw*) as follows:

$$bw = \overline{iu}$$

This approach typically returns more accurate solutions than Options 2 and 3 on smaller sample sizes (n = 5 – 6). The narrow numeric bandwidth can prevent model convergence on poorly clustered or small datasets, however.

**Option 2** STD/IQR: follows Silverman's (1986) rule of thumb as follows:

$$bw = 0.9n^{-\frac{1}{5}} \cdot min\left(sd(x), \frac{IQR(x)}{1.34}\right)$$

where *sd* (standard deviation) and *IQR* (interquartile range) represent the clustering of the input data (*x*), which is multiplied by scaled dataset size (*n*). Selection between standard deviation and interquartile range methods is automatic; the former typically returns the smaller bandwidth for poorly clustered or smaller datasets (n ≤ 10). The interquartile range method generally produces the lower estimate for both tightly clustered and larger datasets (n ≥ 10).

**Option 3** MADD: mean absolute Dortch deviants uses a variation of Silverman's (1986) rule of thumb as follows:

$$bw = 0.9n^{-\frac{1}{5}} \cdot \left(\frac{MAD(x)}{2}\right)$$

where *MAD* is the mean absolute deviation of the input ages (*x*), which is multiplied by scaled dataset size (*n)*. This estimator returns more consistent numeric bandwidths with variable sample sizes compared to Option 2 and provides a higher correct solution rate than Options 1 and 2 for larger sample sizes (n ≥ 8).

### 3.2. Natural logarithm mode

Although the three estimators provide a range of numeric bandwidths that work effectively across the full range of data size-clustering scenarios, all statistical methods assessed in this study struggle to distinguish landforms with large age differences within a single dataset (e.g., Holocene vs. MIS 6; *see* Fig. 3). This is particularly important for regional- or global-scale analyses (Clark et al., 2009; Dortch et al., 2013; Murari et al., 2014), which analyze exposure-age distributions on individual landforms before compiling regional landform ages to define glacial stages (Saha et al., 2018). For P-CAAT specifically, bandwidth estimators operating in linear age space can simultaneously over-smooth younger landform ages and under-smooth older landform ages (Fig. 3A).

To account for this, P-CAAT incorporates a natural logarithm (LN) mode ("LN Data" button), which transforms landform ages and relative uncertainties into LN space (Fig. 3C) and enables bandwidth estimators to isolate component Gaussians across a wider range of ages. While the LN mode typically results in larger uncertainties, this functionality is particularly useful for analysis of very old landforms and regional compilations. Ages should be parsed into datasets roughly characterized by magnitude-of-order age ranges (e.g., 100 ka–1 Ma) while following natural breaks in age clustering. Middle age ranges (e.g., 10 – 100 ka) can be assessed in either linear or LN age space. Young datasets (e.g., 0–1 ka, 1–10 ka) are best analyzed in a linear age space because of small absolute uncertainties and the possibility of negative LN returns, which complicates Gaussian behavior.

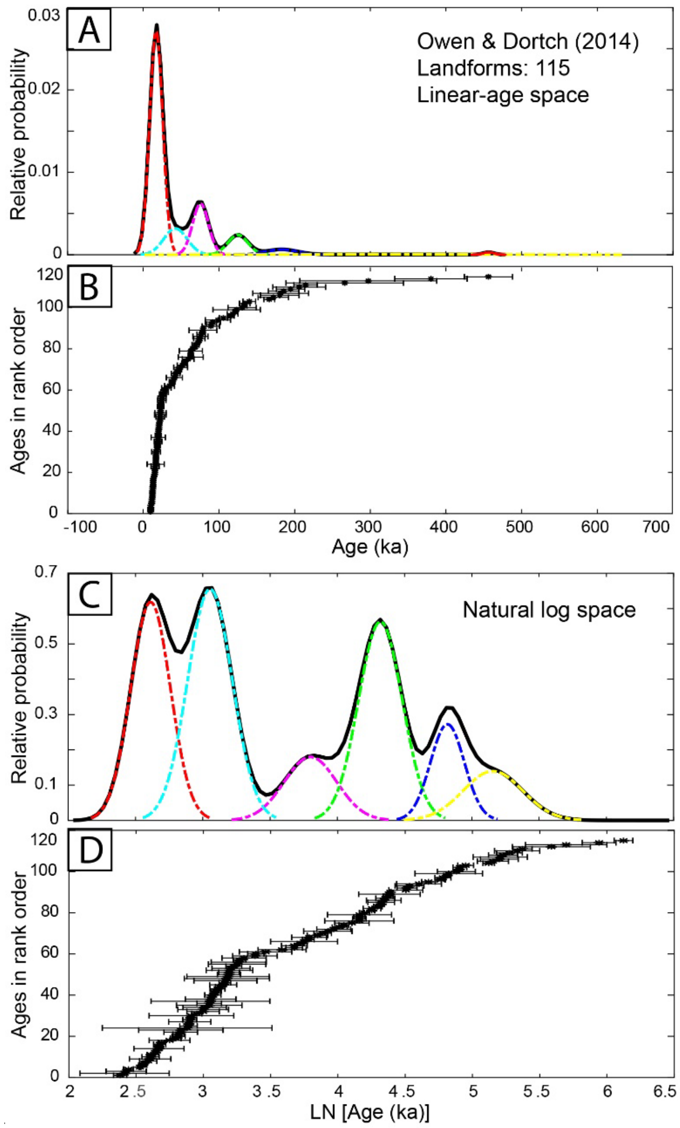


**Fig. 3.** Compilation of 115 Himalayan moraine landform ages (Owen and Dortch, 2014) analyzed in P-CAAT using linear (A–B) and natural logarithm modes (C–D). In (A), the red component Gaussian is centered on 17.1 ± 8.6 ka, whereas (C) distinguishes two events at 14.1 ± 2.8 ka (red) and 22.3 ± 4.4 ka (turquoise). Moreover, LN scale produces more intelligible plotting: higher relative probability and visually narrower sigma for older component Gaussians despite numeric similarity (e.g., yellow component Gaussians are (A) 182.7 ± 21.0 ka and (C) 186.1 ± 24.8 ka or 5.23 ± 0.13 LN (ka)).

### 3.3. Model iterations and precision limits

P-CAAT uses MATLAB's nonlinear least-squares regression function (nlinfit) and a Monte Carlo–style approach (based on a combination of peak probability, residuals, quantile estimates, and parameter test) to perturbate Gaussian starting points for model runs until it converges on a single solution. The dynamic nature of this process makes it impossible to estimate the number of cycles undertaken before completion, but this typically ranges from 1–20 and under challenging cases, > 200. In each cycle, component Gaussians are deconvolved from the cumulative PDE using chi-squared minimization based on the

Levenberg Marquadt algorithm and nonlinear curve fitting with a maximum of 1,000 iterations (Levenberg, 1944; Marquardt, 1963; Moré and Sorensen, 1983).

As component Gaussians are added, the starting points of all other component Gaussians are affected; final positions move according to optimization iterations. The sequential addition of component Gaussian(s) occurs in each cycle until the model converges with the fewest components possible. The optimal correlation threshold for convergence is set at 3σ ($R^2 \geq 0.997$) and is defined based upon a linear regression derived from the model and data PDE. Results that reach 2σ are also considered valid ($R^2 \geq 0.95$) but lower correlation values are not ($R^2 < 0.95$). Although the Kolmogorov-Smirnov test is a standard method for comparing probability distributions, we found that this approach does not provide a strict enough correlation threshold to ensure convergence on a global solution because of lower test sensitivity to variance between the tails of the compared distributions. By comparison, our approach limits variance between the Gaussian mean and sigma estimates to ~$1 \times 10^{-15}$ order of magnitude.

Following model convergence, isolated component Gaussians can be assessed based on geologic context, and the appropriate component can be selected to represent the age of the landform (± 1$\sigma$). This is critical because the geologic context of statistically valid populations should take precedence over pure probability. While P-CAAT does not incorporate a-priori information, like a Bayesian approach (c.f., Martin et al., 2020 and references therein), researchers should consider alternative information when choosing a preferred component Gaussian. Common examples may include morphostratigraphic order, stratigraphic limitations provided by $^{14}C$ ages, evidence of erosion or landform instability, and landform type (e.g., alluvial fans commonly rework debris leading to significant prior exposure). In contrast, common statistical approaches for outlier identification (see Section 2) converge on a single solution based on internal biases without geologic context. This situation leads researchers to search for a statistical test that fits the context of the studied landform. In turn, we argue that P-CAAT addresses these limitations by allowing for choice of bandwidth estimator and component Gaussian based on geologic context, allowing P-CAAT to be highly versatile and applicable, regardless of dataset size or clustering or landform propensity for inheritance (e.g., alluvial fans) or degradation (e.g., moraines).

Although the 1$\sigma$ bounds of the isolated component Gaussian incorporates both internal and geologic uncertainties *(see* Section 2.2), external uncertainties are necessary for comparison with landforms in different regions, or which have been analyzed using alternative instruments, nuclides, or geochronological techniques (Balco et al., 2008). External uncertainties are inherently accounted for when the cumulative PDE is based on external uncertainties. However, for individual landforms, it is best to use internal uncertainties, which requires external uncertainties to be added after PDE analysis. To address this, external uncertainties are propagated into total landform-age uncertainty (*t*) as follows:

$$t = \sigma + P * mean(\frac{E - I}{A})$$

where $\sigma$ is the 1σ sigma bounds (~ 68 percent) of the component Gaussian, *P* is the component Gaussian peak (exposure age), and *E*, *I,* and *A* represent the external and internal uncertainties and exposure ages of the TCN data enclosed by the component Gaussian, respectively.

## 4. Using P-CAAT

To calculate landform ages using P-CAAT, exposure-age data are imported in comma-separated (.csv) or tab-delimited format (.txt). These data should be based on appropriate production rates, scaling schemes, and calculation methods (Balco et al., 2008; Marrero et al., 2016; Martin et al., 2017; Fenton et al., 2019; Jones et al., 2019). In turn, it is worth noting that recalibration or refinement of TCN data will necessitate reanalysis using P-CAAT, because of the dynamic nature of the Monte Carlo approach and the fact that these refinements may affect individual exposure ages to varying degrees (e.g., the magnitude of change increases with absolute age for exposure ages recalculated with a new production rate). Input data include:

i. A distinct landform name to allow P-CAAT to distinguish between separate landforms and to enable users to compile data from multiple landforms in a single file.
ii. The exposure age of each sample (in ka).
iii. The internal (analytical) uncertainty (in ka) for analysis and outlier identification.
iv. The external uncertainty, which incorporates production rate, scaling, and atmospheric model uncertainties, in ka for analysis and outlier identification or simple error propagation.

Formatted example datasets from Barnard et al. (2004), Seong et al. (2007), Schaefer et al. (2008), Hedrick et al. (2011) and Pratt-Sitaula (2011) are included in the download package for test runs and exploration of P-CAAT outputs (*see* "Instillation and functions of P-CAAT tutorial).

### 4.1. P-CAAT outputs

Following model convergence, P-CAAT generates four plots and two data tables. Three of these plots are used to provide insight into the fitting process and the distribution of the exposure ages *(see* Section 4.1.1). The fourth plot and data tables provide information on the model fit and the age and sigma of isolated component Gaussians *(see* Section 4.1.2). To illustrate this functionality, we used exposure-age datasets from Seong et al. (2007), Applegate et al. (2010), Hendrick et al. (2011), and Owen and Dortch (2014); corresponding outputs are shown in Figs. 4–7.

#### 4.1.1. Insight plots

The first insight plot is a quantile-quantile (QQ) plot (PDE vs. Model Fit), which is a standard approach for identifying patterns in correlations (Fig. 4). The $R^2$ value represents the overall degree of explained variance between the PDE, derived from the exposure-age data, and the model, derived from the summed component Gaussians, but the QQ plot provides greater insight into model accuracy with respect to the distribution of the underlying exposure ages. A good model fit will show minimal deviation from the one-to-one line (Fig. 4A), but a poor model fit may show significant nonlinear divergence (Fig. 4B) or dispersion (Fig. 4C), indicating that part of the data PDE is not accounted for by the model. Deviation in the mid- to high-probability range is particularly important, as this often indicates a mismatch between the PDE and model (e.g., a primary distribution is not accounted for). By comparison, deviation near zero is typically less important, as this indicates part of the extreme ends of the tails have been missed.

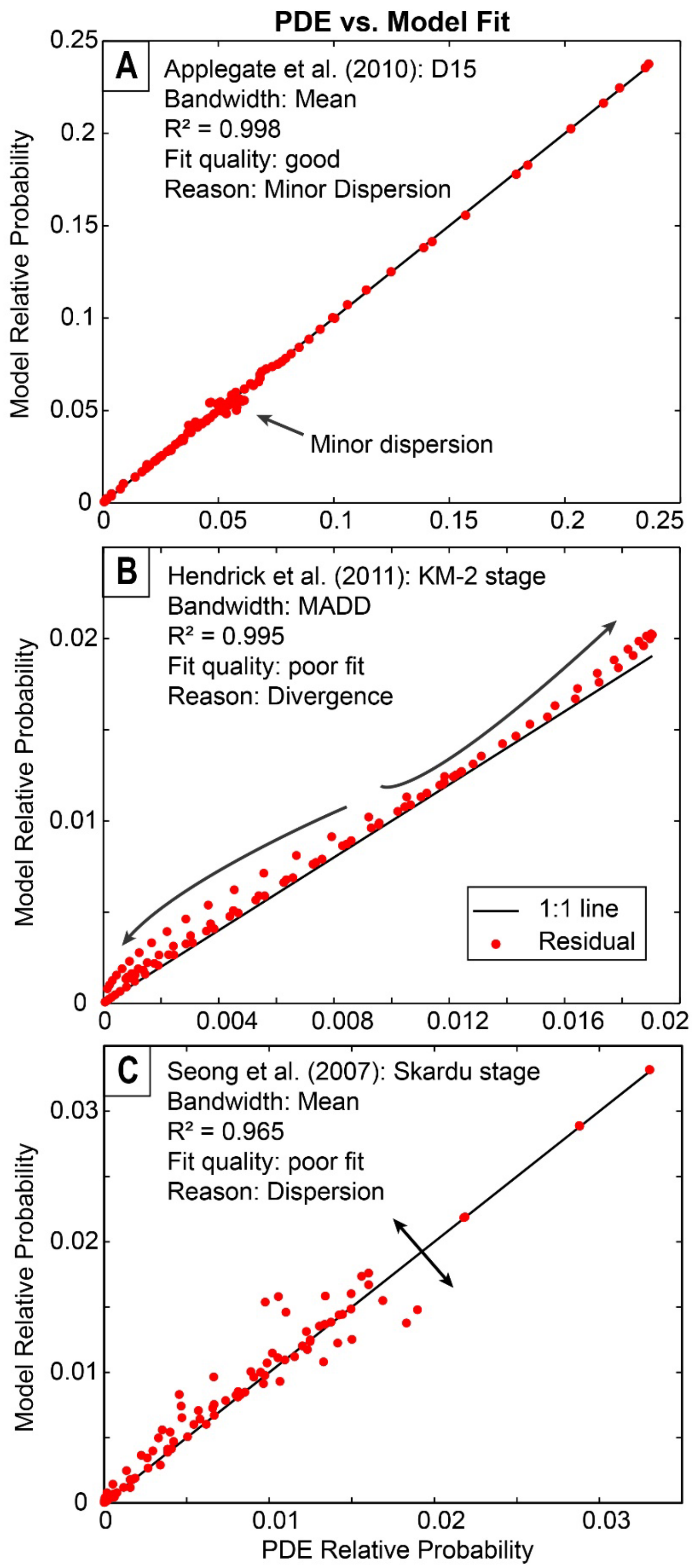


**Fig. 4.** Quantile-quantile plot between the data PDE and the model, evaluated at 100 evenly spaced intervals within the age range of the data, using exposure ages from (A) Applegate et al. (2010), (B) Hendrick et al. (2011), and (C) Seong et al. (2007). Black line represents 1:1 plot for reference.

The second insight plot is a P-CAAT Diagnostic Plot (Fig. 5), which shows the data PDE alongside the results of successive model-cycle PDEs and associated residuals. The number of displayed model cycles reflects the computational ease of the calculation, with more model cycles required for complicated datasets. In Figure 5A, the correlation threshold is reached in just two cycles (Applegate et al., 2010), which reflects the size of the dataset (n = 15) and the absence of positive skew (i.e., pre-depositional exposure). By comparison, analysis of landform ages from Owen and Dortch (2014) required six cycles before model convergence (Fig. 5B), which reflects the size and complexity of the underlying dataset and the use of natural logarithm (LN Data) mode *(see* Section 3.2).

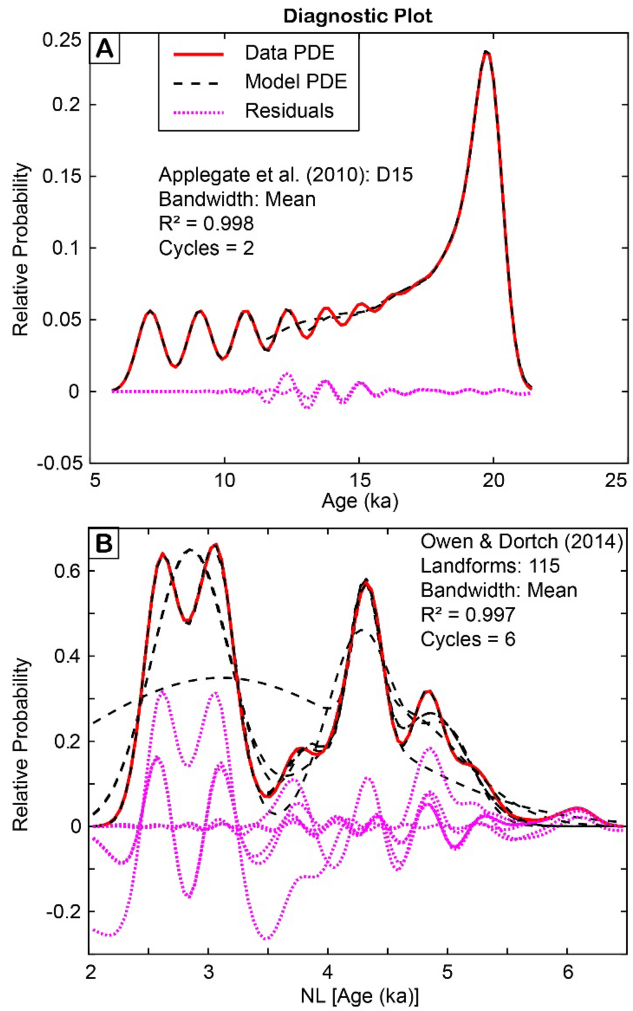


**Fig. 5.** Relative probability plots showing successive P-CAAT model runs. The model continues until a $3\sigma$ correlation threshold is reached ($R^2 \geq 0.997$) or all possible perturbations have been tested. The red line (brown for those with Protanopia and Deuteranopia color-blindness) represents the data PDE and dashed black lines represent the end model of successive P-CAAT cycles; each cycle represents up to 1,000 iterations. The dashed pink lines (blue for colorblind readers) represent residuals between each model fit and the data PDE.

The third insight plot provides an overview of the input data in various formats and is designed to aid the user in choosing a component Gaussian that reflects the distribution of the exposure age data and the geologic context (Fig. 6). Typical choices may include the youngest, highest-probability, or oldest-component Gaussian, but this should be justified based on the characteristics of the geomorphic system, the likelihood of pre- or post-depositional modification of rock surfaces, and the number of exposure ages enclosed by the selected Gaussian *(see* Section 5.3).

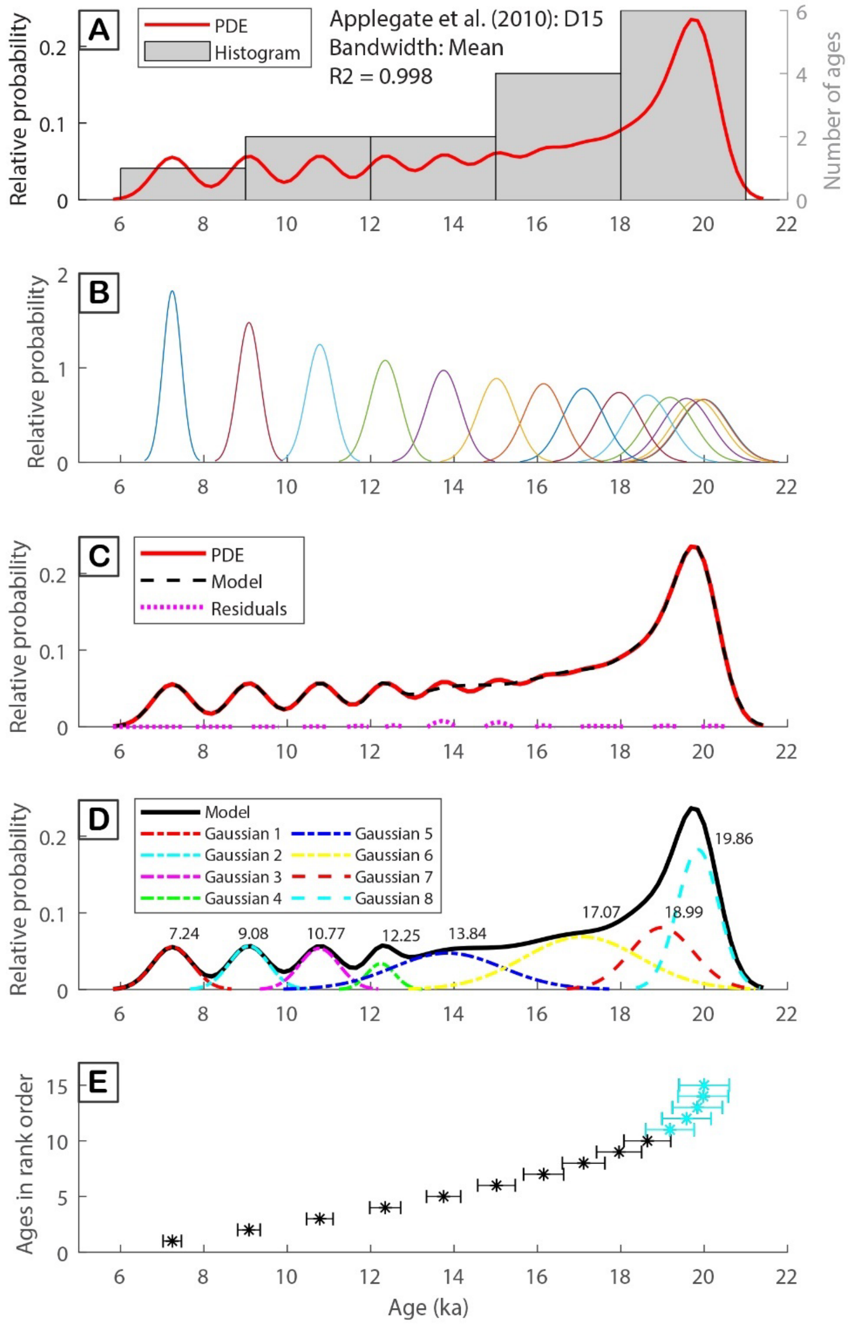


**Fig. 6.** Third insight plot showing the distribution of exposure-age data (Applegate et al., 2010) and P-CAAT model outputs. Results are only displayed if correlation is ≥ 2σ. (A) Histogram of exposure ages calculated using MATLAB's automatic binning algorithm, and the corresponding data PDE. (B) Individual exposure age relative probabilities calculated using internal measurement uncertainties. (C) Data PDE (red line; brown for color-blind), final model PDE (black dashed line), and corresponding residuals (pink dotted line). The minor deviation between the data and model PDE at ~14 ka (relative probability = ~0.05) accounts for the imperfect $R^2$ correlation value of 0.998 and minor dispersion highlighted in Figure-4A. (D) Model PDE (black line) and component Gaussian distributions (other colored lines). (E) Exposure ages in rank order (± internal uncertainties). We identified the oldest component Gaussian (far right - Gaussian 8, dashed turquoise line; light blue for color-blind readers) as correct, with ages completely enclosed at 2σ (internal uncertainty), highlighted in turquoise for reference.

#### 4.1.2. Output tables and plots

To allow future researchers to evaluate the distribution of the input data and the suitability of the analytical choices (i.e., bandwidth estimator and component Gaussian), it is critical that users report all necessary information. To facilitate this, P-CAAT generates two data tables that provide information on:

i. The model fit ($R^2$, $p$ value, bandwidth method, numeric bandwidth).
ii. The characteristics of each component Gaussian, including probability height, the corresponding exposure age (ka), internal and external uncertainties, and the number of ages enclosed by component Gaussians at 2σ.

Although reporting the entirety of these data is not necessary, some core elements must be reported for reproducibility. These are:

i. The P-CAAT version number.
ii. The age and internal and external uncertainties of the selected component Gaussian.
iii. The bandwidth estimator used, along with the model fit, $p$ value, and numeric bandwidth.
iv. A clear rationale for the choice of component Gaussian.

Using information from Figure 6 and the associated data tables, the user can select which component Gaussian best approximates the age of the landform. Entering the associated Gaussian number will produce a Publication Plot (Fig. 7), highlighting the selected Gaussian (in either red or black) and its component ages and visualizes the propagated external uncertainty *(see* Section 3.3). For regional datasets, enter an uppercase 'R' to highlight all component Gaussians.

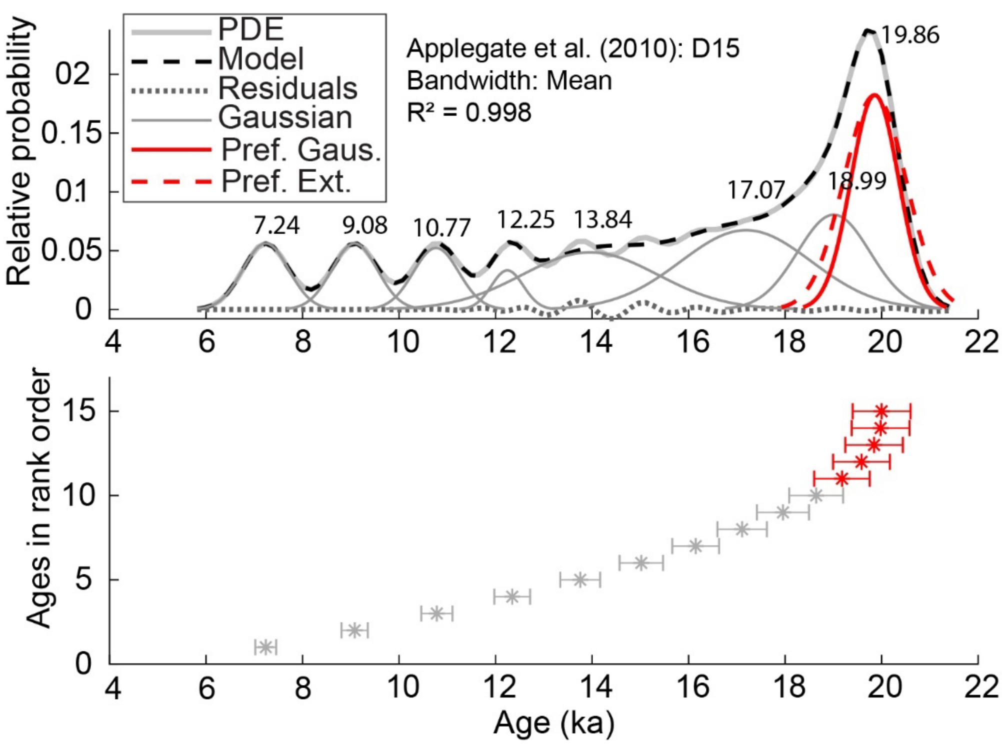


**Fig. 7.** A Publication Plot generated using P-CAAT's red option based on data from Applegate et al. (2010). The upper subplot is a simplified publication-ready version of the P-CAAT model results *(see* Fig. 6), which includes the data PDE (thick gray line), the model PDE (black dashed line), model residuals (gray dotted line), and the individual component Gaussians (thin gray lines). Solid and dashed red lines represent internal and external uncertainties for the selected component Gaussian, respectively *(see* Section 3.2). The lower subplot shows TCN exposure ages in rank order (± internal uncertainties), with ages completely

enclosed by the selected component Gaussian at 2σ highlighted in red (brown for color-blind readers). There are no external uncertainties for the Applegate et al. (2010) test data, thus they were set at 125 percent of internal uncertainties for illustration purposes.

## 5. Testing P-CAAT

Early versions of P-CAAT were extensively modified and tested during reanalysis of >1,500 TCN ages from across the Himalayan-Tibetan Orogen (Dortch et al., 2013; Murari et al., 2014). Although this process provided a consistent tool for regional-scale landform-age analysis, the absence of independent age controls for the vast majority of sites (e.g., minimum or maximum $^{14}C$ ages; Briner et al., 2005) precluded the possibility of rigorously assessing the accuracy of this approach with respect to a known or “true” age.

To address this limitation, we assessed P-CAAT performance through analysis of synthetic datasets with an assigned “true” age. First, we used the synthetic datasets of Applegate et al. (2012) to compare P-CAAT performance against common statistical approaches for outlier identification and assessed model accuracy for datasets compromised by either pre-depositional or post-depositional skew *(see* Section 5.1). Second, we constructed new synthetic datasets to evaluate bandwidth performance and the effects of component Gaussian selection for datasets influenced by both pre- and post-depositional skew *(see* Section 5.2). Synthetic dataset testing is extensive and far exceeds the number of sites with independent age controls.

### 5.1. Synthetic datasets with unidirectional skew

Applegate et al. (2012) took great care in developing skewed synthetic-age datasets based on models of moraine degradation or inheritance (Applegate et al., 2010). These synthetic datasets represent end-member scenarios with unidirectional skew, reflecting the influence of either pre-depositional or post-depositional processes. Representative samples were obtained from the complete datasets based on quantile sampling to ensure consistency with the parent distribution (P. Applegate, pers. comm., 2012) and to produce new datasets with sizes that encompass a realistic range of typical TCN sampling approaches (n = 5–25). They assumed that the negative skew on degraded datasets (D#) represents exhumation of boulders as moraines degrade through time (Putkonen et al., 2008), and positive skew on inherited datasets (I#) represents prior exposure (Putkonen and Swanson, 2003), where “#” represents the number of ages within each dataset.

All 10 datasets (D5–D25; I5–I25) were processed in P-CAAT; combined results are shown in Figure 8. For each dataset, results are based on the narrowest numeric bandwidth that P-CAAT could solve for and that exceeded the 3σ correlation threshold ($R^2 \geq 0.997$; *see* Supplementary Table S1). For comparison, each dataset was also evaluated using MSWD, W-MSWD, Chauvenet's criterion, gESD, 2-SD, 2-MAD, 2σ-overlap, Press and S-PDE methods *(see* Section 2), the results of which are shown in Figure 9 and Table 1.

We applied a three-sample minimum for all statistical methods to ensure an age cluster is adequate to quantify statistical performance. This threshold is based upon the probability of nuclide inheritance and the morphology of sampled rock surfaces *(see* Appendix 1 for further information). In turn, methods that converge on a solution with less than three ages remaining in the calculation pool, or in which the selected component Gaussian encloses less than three ages at 2σ, were considered failed runs. Instances in which methods could not identify an outlier, identified the entire dataset as outliers, or failed to meet their internal test statistic threshold were also considered failed runs.

P-CAAT was able to isolate a component Gaussian that overlapped with the “true” age (20 ka) within 1$\sigma$ (internal uncertainty) for all datasets (Fig. 8). The selection of end-member Gaussian components for both assumed degraded (oldest component) and inherited datasets (youngest component) is justified based on the observable underlying skewness of the data. Although 1$\sigma$ uncertainties of the selected Gaussians are large with small sample sizes (n = 5), uncertainties are reduced rapidly in the degraded datasets as sample size (n = 10) and the degree of overlap between exposure ages increases. A similar pattern holds true for the inherited datasets, but the reduction in component Gaussian uncertainty is delayed to larger sample sizes (n ≥ 10), which reflects the minimal overlap of uncertainties for young synthetic ages. Overall, as sample size increases, P-CAAT returns more accurate and precise solutions.

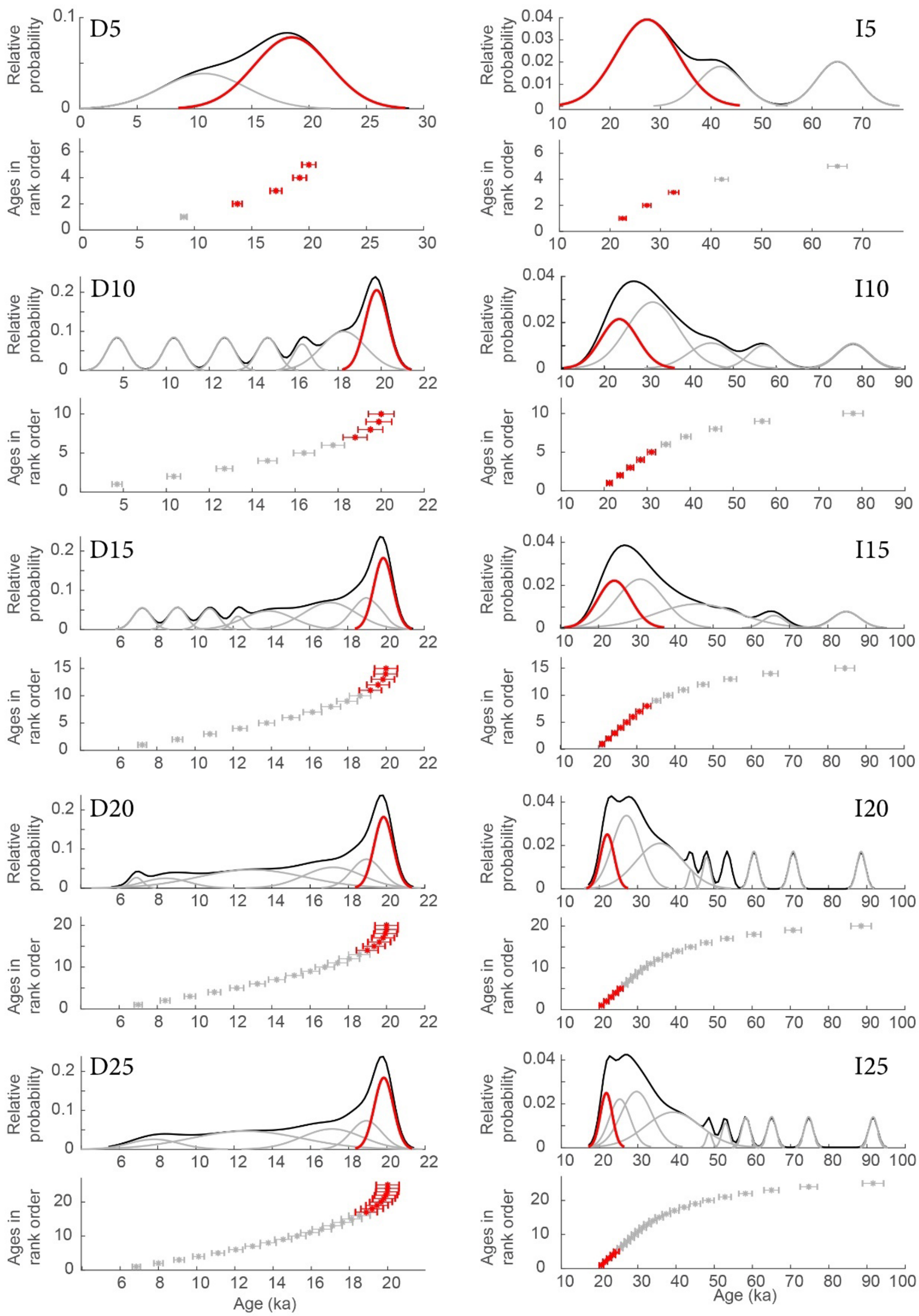


**Fig. 8.** P-CAAT model results for degraded (D#) and inherited datasets (I#), constructed based on quantile sampling of synthetic datasets developed by Applegate et al. (2012). Upper subplots are simplified

publication-ready versions of the P-CAAT model results. They include the model PDE (thick black line), individual component Gaussians (thin gray lines), and the chosen component Gaussian (thick red line; brown for color-blind readers). The lower subplots show TCN exposure ages in rank order (± internal uncertainties); ages enclosed by the selected component Gaussian at 2σ are highlighted in red. True age is 20 ka.

By comparison, the standard approaches for outlier identification in the cosmogenic community demonstrate inconsistency across the skewed datasets, with numerous failures (Fig. 9). This includes MSWD, W-MSWD, Chauvenet's criterion, and 2-SD (n = 4–6), with gESD consistently failing to meet its internal statistical “k” indicator threshold for valid results. Performance varied across the degraded and inherited datasets (Table 1); W-MSWD performed moderately well on negatively skewed (degraded) datasets, but poorly on positively skewed datasets (inherited), and the reverse was true for Chauvenet's criterion and 2-SD. The 2σ-overlap method is more consistent than other standard approaches, with only a single failure (I5; Fig. 9), but solutions trend away from the “true” age and uncertainties increase with larger sample sizes for both degraded and inherited datasets (Table 1), a result of less stringent exclusion criteria. For datasets devoid of geologic uncertainty, mean-based outlier detection could prove effective. These methods struggle to identify outliers when geologic uncertainty is present, however, because they explicitly assume a single probability distribution.

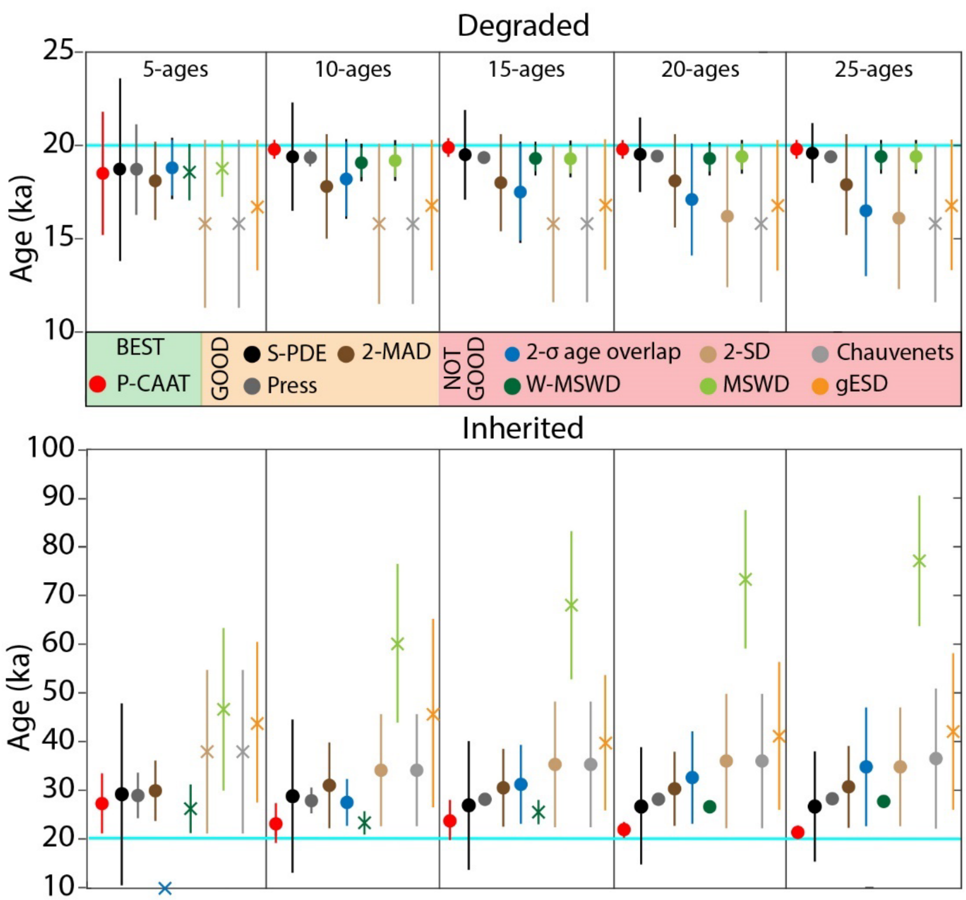


**Fig. 9.** Comparison plot testing the nine methods against skewed datasets from Applegate et al. (2012). Turquoise (light blue for color-blind readers) horizontal bar represents true age (20 ka). Note that P-CAAT is consistently closer to the “true age” (more accurate), with smaller average uncertainties (more precise) than other methods. × = failed test and circles = successful test. All vertical uncertainty bars are 1σ. Quadratically propagated internal uncertainty is visible as vertical black extension lines for (W)MSWD results on degraded datasets. P-CAAT is represented by the left most points in each subplot.

Of the remaining methods, 2-MAD, (Muzikar et al., 2017), and S-PDE (Stübner et al., 2021) successfully converge on solutions with no failures and provide consistent ages for both degraded and inherited datasets. The superior performance of 2-MAD, with respect to common statistical approaches (e.g., 2-SD), is consistent with previous research (Leys et al., 2013). If researchers insist on using traditional non-PDE approaches, we recommend adopting 2-MAD as a new standard over typical mean/std reporting in the cosmogenic community. P-CAAT consistently outperforms these methods, however, with a smaller average deviation from the true age for both degraded (P-CAAT = 0.4 ka; S-PDE = 0.7 ka; Press = 0.8 ka; 2-MAD = 2.0 ka) and inherited datasets (P-CAAT = 3.7 ka; S-PDE = 7.7 ka; Press = 8.3 ka; 2-MAD = 10.5 ka) and with markedly reduced uncertainties *(see* Table 1).

Differences between the results of P-CAAT and S-PDE (Stübner et al., 2021) largely reflect the performance of the chosen bandwidth estimator and collinearity between age and uncertainty. These approaches also differ in other subtle but important features, including the propagation of external uncertainties, error normalization, and the viability threshold for component Gaussian selection (three enclosed ages vs. ≥ 5 percent relative probability). The Press approach appears to model the highest probability Gaussian, which leads it to overestimate the age of inherited datasets. Overall, P-CAAT outperforms both S-PDE and Press, most notably for inherited datasets *(see* Fig. 9), but all three techniques represent substantial improvements upon mean- or median-based approaches and are particularly effective at eliminating the negative skew associated with degrading landforms.

**Table 1.** Results of analysis of degraded and inherited datasets that represent a 20 ka old landform from Applegate et al. (2012), reporting the landform age ± internal uncertainty (1σ) for each method (n = 10) and each sample size (n = 5–25). Columns are ordered by the number of failed runs (italic red text), which occurred where methods identified the entire dataset as outliers (a), could not identify an outlier (b), or failed to meet their internal test statistic threshold (c). (*see* Section 2 for definitions of statistical names and details of calculations).

| Dataset | Sample size | Statistical method (ka) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | P-CAAT | S-PDE | Press | 2-MAD | 2σ-overlap | W-MSWD | 2-SD | MSWD | Chauvenet's | gESD |
| Degraded | 5 | 18.5 ± 3.3 | 18.7 ± 4.9 | 18.7 ± 2.4 | 18.1 ± 2.1 | 18.8 ± 1.5 | *18.6 ± 1.2* [c] | *15.8 ± 4.5* [b] | *18.8 ± 1.5* [c] | *15.8 ± 4.5* [b] | *16.8 ± 3.5* [b, c] |
| | 10 | 19.8 ± 0.5 | 19.4 ± 2.9 | 19.3 ± 0.4 | 17.8 ± 2.8 | 18.2 ± 2.0 | 19.1 ± 0.8 | *15.8 ± 4.3* [b] | 19.2 ± 0.9 | *15.8 ± 4.3* [b] | *16.8 ± 3.5* [b, c] |
| | 15 | 19.9 ± 0.5 | 19.5 ± 2.4 | 19.3 ± 0.3 | 18.0 ± 2.6 | 17.5 ± 2.6 | 19.3 ± 0.7 | *15.8 ± 4.2* [b] | 19.3 ± 0.8 | *15.8 ± 4.2* [b] | *16.8 ± 3.5* [b, c] |
| | 20 | 19.9 ± 0.5 | 19.5 ± 2.0 | 19.4 ± 0.3 | 18.1 ± 2.5 | 17.1 ± 3.0 | 19.3 ± 0.7 | 16.2 ± 3.8 | 19.4 ± 0.7 | *15.8 ± 4.2* [b] | *16.8 ± 3.5* [b, c] |
| | 25 | 19.8 ± 0.5 | 19.6 ± 1.6 | 19.4 ± 0.2 | 17.9 ± 2.7 | 16.5 ± 3.5 | 19.4 ± 0.6 | 16.1 ± 3.8 | 19.4 ± 0.7 | *15.8 ± 4.2* [b] | *16.8 ± 3.5* [b, c] |
| Mean deviation ± mean uncertainty | | 0.4 ± 1.1 | 0.7 ± 2.8 | 0.8 ± 0.7 | 2.0 ± 2.5 | 2.4 ± 2.5 | 0.9 ± 0.8 | 4.1 ± 4.1 | 0.8 ± 0.9 | 4.2 ± 4.3 | 3.2 ± 3.5 |
| Inherited | 5 | 27.3 ± 6.1 | 29.2 ± 18.4 | 28.9 ± 4.4 | 29.9 ± 6.2 | *N/A* [a] | *26.2 ± 4.5* [c] | *37.9 ± 16.8* [b] | *46.6 ± 16.7* [c] | *37.9 ± 16.8* [b] | *43.9 ± 16.4* [b, c] |
| | 10 | 23.4 ± 4.3 | 28.8 ± 15.5 | 27.8 ± 2.4 | 31.0 ± 8.8 | 27.5 ± 4.8 | *23.3 ± 2.0* [c] | 34.1 ± 11.5 | *60.2 ± 16.3* [c] | 34.1 ± 11.5 | *45.7 ± 19.3* [b, c] |
| | 15 | 24.0 ± 4.3 | 26.9 ± 13.0 | 28.2 ± 1.3 | 30.5 ± 8.0 | 31.2 ± 8.1 | *25.5 ± 1.3* [c] | 35.3 ± 12.9 | *68.0 ± 15.2* [c] | 35.3 ± 12.9 | *39.7 ± 13.8* [c] |
| | 20 | 22.1 ± 1.8 | 26.7 ± 11.9 | 28.1 ± 0.9 | 30.3 ± 7.6 | 32.6 ± 9.5 | 26.6 ± 1.0 | 36.0 ± 13.8 | *73.3 ± 14.4* [c] | 36.0 ± 13.8 | *41.1 ± 15.2* [c] |
| | 25 | 21.7 ± 1.6 | 26.7 ± 11.3 | 28.2 ± 1.0 | 30.7 ± 8.4 | 34.8 ± 12.2 | 27.7 ± 1.1 | 34.8 ± 12.2 | *77.1 ± 13.4* [c] | 36.5 ± 14.4 | *42.0 ± 16.1* [c] |
| Mean deviation ± mean uncertainty | | 3.7 ± 3.6 | 7.7 ± 14.0 | 8.3 ± 2.0 | 10.5 ± 7.8 | 11.5 ± 8.7 | 5.9 ± 2.0 | 15.6 ± 13.5 | 45.0 ± 15.2 | 16.0 ± 13.9 | 22.5 ± 16.2 |
| Number of failures | | **0** | **0** | **0** | **0** | **1** | **4** | **4** | **6** | **6** | **10** |

### 5.2. Synthetic datasets with bidirectional skew

The synthetic datasets developed by Applegate et al. (2012) provide useful end-member scenarios for analyzing statistical performance. TCN datasets often incorporate both positive (e.g., inheritance) and negative skew (e.g., erosion, exhumation, shielding), however, reflecting the characteristics of the geomorphic system and the genetic history of the landform (Ivy-Ochs et al., 2007; Pallàs et al., 2010; Tomkins et al., 2021). As a result, combining the degraded and inherited components of the Applegate et al. (2012) datasets would not produce TCN age distributions that accurately reflect the relative frequencies and magnitudes of pre- and post-depositional processes. To address this, we constructed two new synthetic data pools with more realistic bidirectional skew, based on statistics from glacial compilations *(see* Dortch et al., 2013; Murari et al., 2014; and references within).

Based on analysis of exposure-age distributions from across the Himalayan-Tibetan Orogen, Dortch et al. (2013; n = 595 ages) and Murari et al. (2014; n = 934 ages) concluded that for recent to Last Glacial Maximum (gLGM; ~18–26 ka; *see* Hughes et al., 2013) glacial landforms, approximately 73 percent of exposure ages matched the calculated age of deposition, and the remaining ages were either younger (~17 percent) or older (~10 percent). By comparison, for landforms deposited prior to the gLGM, only about 42 percent of exposure ages matched the calculated deposition age. Of the remaining exposure ages, most were younger than the age of the landform (~48 percent) with only 10 percent being too old. For both datasets, the average magnitude of under- and overestimation was approximately 45 percent and approximately 175 percent, respectively. Using these observations as a reasonable first-order approximation of a typical TCN dataset, we developed two synthetic data pools ($n = 10^4$ exposure ages) with assigned landform ages of 22 ka and 60 ka to represent gLGM and pre-gLGM stages *(see* Fig. 10, Appendix S1).

This new synthetic data pool served as a basis to evaluate P-CAAT bandwidth performance and the effects of component Gaussian selection across a range of sampling resolutions. Random ages were drawn from the data pool without replication to generate 10,000 individual datasets at several sample sizes (n = 5, 6, 7, 8, 9, 10, 15, 20, 25). The 60 ka five-sample dataset was generated twice to ensure meaningful variation was limited to sample size. This range was chosen to encompass typical TCN sampling approaches in which collecting five to six TCN samples from a single landform is common (Pallàs et al., 2010), and datasets comprising ≥ 20 samples for a single landform are rare but not unheard of (Rinterknecht et al., 2006). To analyze these data, we processed each dataset ($n = 19 \times 10^4$) and recorded the highest-probability Gaussian and the oldest-component Gaussian (peak age ± 1σ) for each convergent model run ($R^2 \geq 0.95$), applying a three-sample minimum to reduce the probability of selecting inherited or non-representative components *(see* Section 5.1). Successful model runs were those that produced component Gaussians with peak ages within 10 percent of the “true age” (22 ± 2.2 ka and 60 ± 6 ka). Because the input datasets varied markedly in size and clustering, results are based on the narrowest numeric bandwidth that P-CAAT could solve for. Full results are presented in Supplementary Table S2 for brevity.

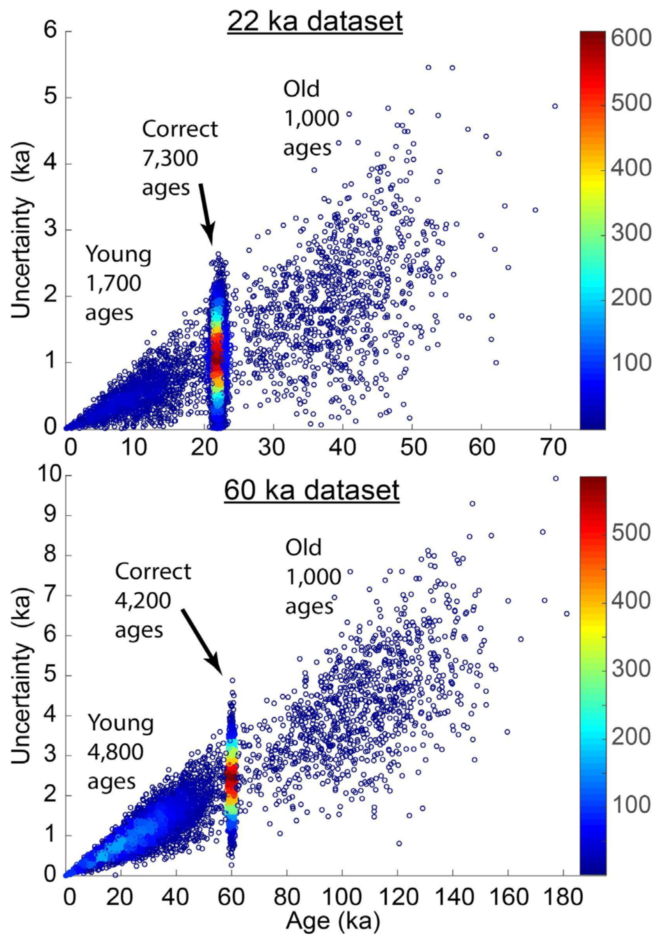


**Fig. 10.** Distribution plots showing all 10,000 ages in the 22 ka and 60 ka data pools, colored by point density (blue → red scale bar; blue to brown for color blind readers). The highest density forms a bullseye pattern in the correct ages scatter. For a full description of the construction of these synthetic datasets, *see* Appendix S1.

### 5.2.1 Bandwidth performance

P-CAAT returned consistent numeric bandwidths across the range of sample sizes *(see* Supplementary Fig. S2). Typically, STD/IQR provides the widest numeric bandwidths, followed by MADD then Mean, but the estimator that yields the smallest numerical bandwidth will vary based on the ratio of age uncertainties, deviation, and skewness of the dataset (*see* Fig. 2 for a contrasting example). The precision of the selected component Gaussian (σ) scales with the numeric bandwidth, with narrow bandwidths yielding smaller uncertainties than wide bandwidths, although the former are generally more difficult for P-CAAT to solve. The choice of bandwidth estimator is less critical on well-clustered data. However, poorly clustered datasets demonstrate complex behavior when scaled against sample size (*see* Supplementary Fig. S4). Generally, the narrowest numeric bandwidth that P-CAAT can solve for ($R^2 \geq 0.95$) is preferred, although each model fit should be carefully assessed to avoid qualitatively poor model fits (e.g., nonlinear divergence or dispersion on a typical QQ plot). Further consideration should be given to wider bandwidths for old (>100 ka) scattered datasets (*see* data analysis tutorial videos).

### 5.2.2 Component Gaussian selection

The number of component Gaussians isolated by P-CAAT varies with sample size, clustering, and bandwidth choice. Although P-CAAT solves > 99 percent of all datasets, obtaining an answer within 10 percent of the “true” age is more difficult. Figure 11 visualizes the number of “correct” solves at each sample size for both the 22 ka and 60 ka datasets, with results subset by the selected component Gaussian (oldest vs. highest).

The 22 ka dataset has a > 90 percent “correct” solve rate with the minimum sample size (n = 5), and the highest-probability component Gaussian is consistently correct (> 99 percent) as sample size increases (n ≥ 8). By comparison, the oldest-component Gaussian outperforms the highest-probability component Gaussian by a significant margin for the 60 ka dataset. The highest-probability Gaussian gives a higher correct solve rate only when sample sizes are large (n ≥ 20).

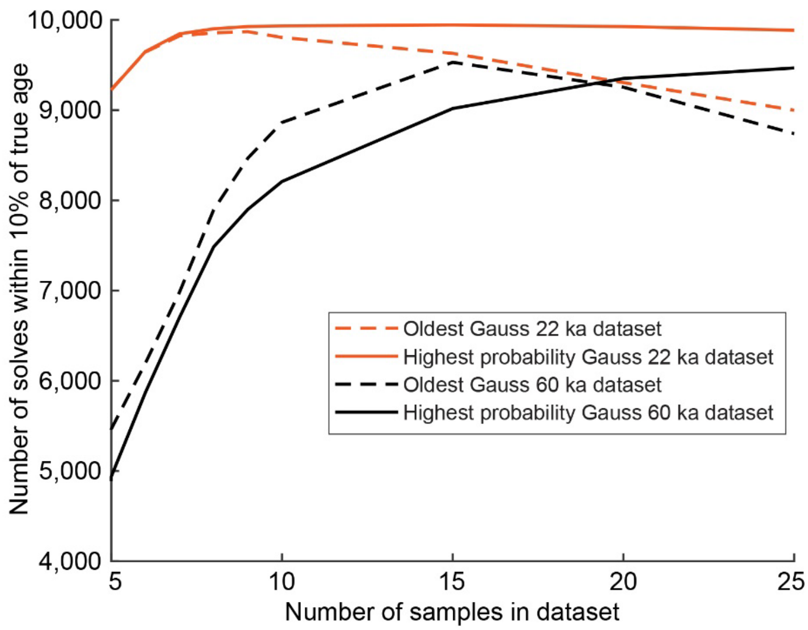


**Fig. 11.** Comparison plot of oldest (dashed lines) and highest-probability component Gaussians (solid lines) for the 22 ka (orange; brown-tan for color-blind readers) and 60 ka (black) datasets, quantified by the number of correct P-CAAT model runs (≤ 10 percent from the “true” age). The highest-probability Gaussian is consistently the most accurate for the 22 ka dataset, whereas the oldest-component Gaussian returns more accurate solutions for the 60 ka dataset at most sample sizes (n < 20).

At all sample sizes, P-CAAT’s correct solve rate is lower for the 60 ka dataset, which primarily reflects the distribution of the underlying data (*see* Section 5.2). We argue, however, that the five-sample solve rate of 55 percent (5,500 correct model runs) is very good considering that P-CAAT outperforms the 42 percent of the population of “correct” ages (*see* Fig. 10). Similarly, the initial solve rate of 92 percent for the 22 ka dataset exceeds the proportion of “correct” ages in the underlying data (73 percent). Outperforming the number of “correct” ages with only five ages demonstrates that P-CAAT is effective at removing significant pre- or post-depositional skew, even with an objective and non-contextual interpretation scheme, with performance improving as sample size increases.

### 5.3. P-CAAT applications in geomorphic systems

Based on the extensive testing described in Sections 5.1 and 5.2 and the interpretation above, we argue that in absence of geologic context, using P-CAAT and a consistent interpretation scheme could play a key role in standardizing TCN age interpretations. Although model assessment should be based on site-specific information, such as independent age control (e.g., $^{14}C$) or geomorphologic evidence, the following interpretation scheme should be used as a guide for component Gaussian selection. This guide is not necessarily prescriptive as alternative approaches may be required to ensure consistency with the geologic context.

Based on analysis of the synthetic datasets above, and for glacial landforms deposited during or following the gLGM, we recommend using the highest-probability component Gaussian that encloses a minimum of three ages at 2σ to represent the age of the landform. In contrast, for landforms older than the gLGM, the oldest-component Gaussian is preferred for small sample sizes ($n < 15$) and the highest-probability Gaussian for large sample sizes ($n \geq 20$).

Although synthetic datasets were not explicitly developed for other landform types, an understanding of the associated geomorphic and TCN systems enables us to make the following initial recommendations. For landslides, and in particular large rock avalanches (Dortch et al., 2009), the distribution of TCN ages appears comparable to glacial deposits, where post-depositional processes prevail over pre-depositional exposure (Heyman et al., 2011). In turn, using the highest-probability component Gaussian on younger deposits and the oldest-component Gaussian on older deposits is theoretically prudent. In contrast, landforms subject to reworking (e.g., alluvial fans, flood deposits, and fluvial terraces) often preserve pre-depositional exposure and incorporate inherited TCNs (Dortch et al., 2011a, b). For younger landforms, selecting the youngest component Gaussian may be required to offset high rates of inheritance (Hancock et al., 1999), whereas the highest-probability component Gaussian may be preferable for older landforms, because inherent landform stability mitigates some post-depositional processes and the relative difference between inherited and "true" ages diminishes with increasing age. Further work is necessary, however, to assess "typical" TCN age distributions for these landforms.

## 7. Conclusion

Geologic uncertainty (e.g., erosion, exhumation, shielding, nuclide inheritance) can profoundly influence the distribution of TCN datasets. Most common statistical approaches for outlier identification, however, assume a single underlying distribution, do not propagate external uncertainties, and provide arithmetic single solutions, irrespective of geologic context. To address these limitations, we developed the Probabilistic Cosmogenic Age Analysis Tool (P-CAAT), which uses a Monte Carlo approach to isolate component normal distributions (Gaussians) to remove pre- and post-depositional skew. Using synthetic datasets developed by Applegate et al. (2012), we demonstrated that P-CAAT consistently outperforms alternative statistical approaches, many of which are characterized by frequent failures and reduced accuracy and precision as sample size increases. Other probabilistic approaches (S-PDE & Press) perform well on degraded datasets, but less well on those influenced by inheritance. If alternatives to PDEs must be considered, we recommend 2-MAD over other more common approaches for detecting outliers and reporting results and stress the need to move beyond standard deviation-based approaches.

New synthetic datasets ($n = 10^4$ each) based on typical TCN distributions observed in the Himalayan-Tibetan Orogen (Dortch et al., 2013; Murari et al., 2014) were developed to guide component Gaussian selection. Results indicate that for glacial and rock avalanches deposited at or following the gLGM, users should typically select the highest-probability component Gaussian that encloses a minimum of three ages at 2σ to represent the age of the landform. By comparison, the oldest-component Gaussian is preferred for landforms older than the gLGM when sample sizes are small ($n < 20$). Further analysis is required to incorporate a wider range of landform types (e.g., flood deposits, alluvial fans, fluvial terraces), but applying a consistent interpretation scheme could aid in standardizing regional- or global-scale analyses, while minimizing uncertainty in landform-age analysis associated with the choice of statistical test. In summary, P-CAAT is optimized for analysis of TCN datasets, as it incorporates both systematic and geologic uncertainty, quantifies uncertainty directly from component Gaussians, enables multimodal distributions to be separated, and allows geologic context to inform landform-age analysis. To encourage wider testing and application of this standalone tool, P-CAAT is available for free, along with tutorial videos, to download at kgs.uky.edu/anorthite/PCAAT.

## Acknowledgments

This paper is dedicated to Jason's cat, Anorthite "Rooget" Waldrop-Dortch, who passed away in June 2019. Anorthite inspired Jason to routinely ponder Dr. Schrödinger's ideas and probabilistic approaches to both science and life. The current version of P-CAAT (2.1) and future versions will be documented and archived through the Kentucky Geologic survey [kgs.uky.edu/anorthite/PCAAT] along with all datasets generated and analyzed during this study. M.K. Murari is supported by Ministry of Earth Science [MoES/P.O.(Seismic) 8(09)-Geochron/2012]. We would like to thank the three anonymous reviewers and Dr. Blard for their constructive comments that made this a more complete manuscript, and Prof. Brent Goehring for kindly sharing a MATLAB implementation of the Press method. Thank to Dr. Applegate for insightful conversations and datasets when P-CAAT was a nascent project.

## Bibliography

Akçar, N., Ivy-Ochs, S., Kubik, P.W., Schlüchter, C., 2011. Post-depositional impacts on “Findlinge” (erratic boulders) and their implications for surface-exposure dating. Swiss Journal of Geosciences 104, 445–453. https://doi.org/10.1007/s00015-011-0088-7

Allard, J.L., Hughes, P.D., Woodward, J.C., Fink, D., Simon, K., Wilcken, K.M., 2020. Late Pleistocene glaciers in Greece: A new $^{36}$Cl chronology. Quaternary Science Reviews 245, 106528. https://doi.org/10.1016/j.quascirev.2020.106528

Applegate, P.J., Lowell, T.V., Alley, R.B., 2008. Comment on “Absence of Cooling in New Zealand and the Adjacent Ocean During the Younger Dryas Chronozone.” Science 320, 746–746. https://doi.org/10.1126/science.1152098

Applegate, P.J., Urban, N.M., Keller, K., Lowell, T.V., Laabs, B.J.C., Kelly, M.A., Alley, R.B., 2012. Improved moraine age interpretations through explicit matching of geomorphic process models to cosmogenic nuclide measurements from single landforms. Quaternary Research 77, 293–304. https://doi.org/10.1016/j.yqres.2011.12.002

Applegate, P.J., Urban, N.M., Laabs, B.J.C., Keller, K., Alley, R.B., 2010. Modeling the statistical distributions of cosmogenic exposure dates from moraines. Geoscientific Model Development 3, 293–307. https://doi.org/10.5194/gmd-3-293-2010

Balco, G., 2011. Contributions and unrealized potential contributions of cosmogenic-nuclide exposure dating to glacier chronology, 1990–2010. Quaternary Science Reviews 30, 3–27. https://doi.org/10.1016/j.quascirev.2010.11.003

Balco, G., Stone, J.O., Lifton, N.A., Dunai, T.J., 2008. A complete and easily accessible means of calculating surface exposure ages or erosion rates from $^{10}$Be and $^{26}$Al measurements. Quaternary Geochronology, Prospects for the New Frontiers of Earth and Environmental Sciences 3, 174–195. https://doi.org/10.1016/j.quageo.2007.12.001

Barnard, P.L., Owen, L.A., Sharma, M.C., Finkel, R.C., 2004. Late Quaternary (Holocene) landscape evolution of a monsoon-influenced high Himalayan valley, Gori Ganga, Nanda Devi, NE Garhwal. Geomorphology 61, 91–110. https://doi.org/10.1016/j.geomorph.2003.12.002

Barrows, T.T., Lehman, S.J., Fifield, L.K., Deckker, P.D., 2007. Absence of cooling in New Zealand and the adjacent ocean during the Younger Dryas Chronozone. Science 318, 86–89. https://doi.org/10.1126/science.1145873

Barrows, T.T., Lehman, S.J., Fifield, L.K., Deckker, P.D., 2008. Response to “Comment on ‘Absence of Cooling in New Zealand and the Adjacent Ocean During the Younger Dryas Chronozone.’” Science 320, 746–746. https://doi.org/10.1126/science.1152216

Benson, L., Madole, R., Landis, G., Gosse, J., 2005. New data for Late Pleistocene Pinedale alpine glaciation from southwestern Colorado. Quaternary Science Reviews 24, 49–65. https://doi.org/10.1016/j.quascirev.2004.07.018

Blisniuk, K., Rockwell, T., Owen, L.A., Oskin, M., Lippincott, C., Caffee, M.W., Dortch, J., 2010. Late Quaternary slip rate gradient defined using high-resolution topography and $^{10}$Be dating of offset landforms on the southern San Jacinto Fault Zone, California. Journal of Geophysical Research: Solid Earth 115. https://doi.org/10.1029/2009JB006346

Borchers, B., Marrero, S., Balco, G., Caffee, M., Goehring, B., Lifton, N., Nishiizumi, K., Phillips, F., Schaefer, J., Stone, J., 2016. Geological calibration of spallation production rates in the CRONUS-Earth project. Quaternary Geochronology 31, 188–198. https://doi.org/10.1016/j.quageo.2015.01.009

Borradaile, G.J., 2003. Statistics of Earth Science Data: Their Distribution in Time, Space and Orientation. Springer-Verlag, Berlin Heidelberg. https://doi.org/10.1007/978-3-662-05223-5

Briner, J.P., Goehring, B.M., Mangerud, J., Svendsen, J.I., 2016. The deep accumulation of $^{10}$Be at Utsira, southwestern Norway: Implications for cosmogenic nuclide exposure dating in peripheral ice sheet landscapes. Geophysical Research Letters 43, 9121–9129. https://doi.org/10.1002/2016GL070100

Briner, J.P., Kaufman, D.S., Manley, W.F., Finkel, R.C., Caffee, M.W., 2005. Cosmogenic exposure dating of late Pleistocene moraine stabilization in Alaska. Geological Society of America Bulletin 117, 1108–1120. https://doi.org/10.1130/B25649.1

Briner, J.P., Swanson, T.W., 1998. Using inherited cosmogenic $^{36}Cl$ to constrain glacial erosion rates of the Cordilleran ice sheet. Geology 26, 3–6. https://doi.org/10.1130/0091-7613(1998)026<0003:UICCTC>2.3.CO;2

Brown, E.T., Molnar, P., Bourlès, D.L., 2005. Comment on “Slip-Rate Measurements on the Karakorum Fault May Imply Secular Variations in Fault Motion.” Science 309, 1326–1326. https://doi.org/10.1126/science.1112508

Burbank, D.W., Leland, J., Fielding, E., Anderson, R.S., Brozovic, N., Reid, M.R., Duncan, C., 1996. Bedrock incision, rock uplift and threshold hillslopes in the northwestern Himalayas. Nature 379, 505–510. https://doi.org/10.1038/379505a0

Cerling, T.E., Poreda, R.J., Rathburn, S.L., 1994. Cosmogenic $^{3}He$ and $^{21}Ne$ age of the Big Lost River flood, Snake River Plain, Idaho. Geology 22, 227–230. https://doi.org/10.1130/0091-7613(1994)022<0227:CHANAO>2.3.CO;2

Chevalier, M.-L., Hilley, G., Tapponnier, P., Van Der Woerd, J., Liu-Zeng, J., Finkel, R.C., Ryerson, F.J., Li, H., Liu, X., 2011. Constraints on the late Quaternary glaciations in Tibet from cosmogenic exposure ages of moraine surfaces. Quaternary Science Reviews 30, 528–554. https://doi.org/10.1016/j.quascirev.2010.11.005

Chevalier, M.-L., Ryerson, F.J., Tapponnier, P., Finkel, R.C., Woerd, J.V.D., Haibing, L., Qing, L., 2005a. Response to “Comment on ‘Slip-Rate Measurements on the Karakorum Fault May Imply Secular Variations in Fault Motion.’” Science 309, 1326–1326. https://doi.org/10.1126/science.1112629

Chevalier, M.-L., Ryerson, F.J., Tapponnier, P., Finkel, R.C., Woerd, J.V.D., Haibing, L., Qing, L., 2005b. Slip-rate measurements on the Karakorum Fault may imply secular variations in fault motion. Science 307, 411–414. https://doi.org/10.1126/science.1105466

Clark, P.U., Dyke, A.S., Shakun, J.D., Carlson, A.E., Clark, J., Wohlfarth, B., Mitrovica, J.X., Hostetler, S.W., McCabe, A.M., 2009. The Last Glacial Maximum. Science 325, 710–714. https://doi.org/10.1126/science.1172873

Davies, B.J., Darvill, C.M., Lovell, H., Bendle, J.M., Dowdeswell, J.A., Fabel, D., Garcia, J.L., Geiger, A., Glasser, N.F., Gheorghiu, D.M., Harrison, S., Heini, A.S., Kaplan. M.R., Martin, J.R.V., Mendelova, M., Palmer, A., Pelto, M., Rodes, A., Sagredo, E.A., Smedley, R.K., Smellie, J.L., Thorndycraft, V.R., 2020. The evolution of the Patagonian Ice Sheet from 35 ka to the present day (PATICE). Earth Science Reviews 24, 103152. https://doi.org/10.1016/j.earscirev.2020.103152

D'arcy, M. K., Schildgen, T. F., Turowski, J. M., & Dinezio, P.,2019. Inferring the timing of abandonment of aggraded alluvial surfaces dated with cosmogenic nuclides. *Earth Surface Dynamics, 7(3)*, 755–771. https://doi.org/10.5194/esurf-7-755-2019

Darvill, C.M., Menounos, B., Goehring, B.M., Lian, O.B. and Caffee, M.W., 2018. Retreat of the western Cordilleran Ice Sheet margin during the last deglaciation. *Geophysical Research Letters*, 45(18), pp.9710-9720. https://doi.org/10.1029/2018GL079419Dehnert, A., Schlüchter, C., 2008. Sediment burial dating using terrestrial cosmogenic nuclides. E&G Quaternary Science Journal 57, 210–225. https://doi.org/10.3285/eg.57.1-2.8

Delmas, M., Gunnell, Y., Braucher, R., Calvet, M., Bourlès, D., 2008. Exposure age chronology of the last glaciation in the eastern Pyrenees. Quaternary Research 69, 231–241. https://doi.org/10.1016/j.yqres.2007.11.004

Dortch, J., Owen, L.A., Haneberg, W.C., Caffee, M.W., Dietsch, C., Kamp, U., 2009. Nature and timing of large-landslides in northern India. Quaternary Science Reviews 28, 1037–1056. https://doi:10.1016/j.quascirev.2008.05.002Dortch, J.M., Dietsch, C., Owen, L.A., Caffee, M.W., Ruppert, K., 2011a. Episodic fluvial incision of rivers and rock uplift in the Himalaya and Transhimalaya. Journal of the Geological Society 168, 783–804. https://doi.org/10.1144/0016-76492009-158

Dortch, J.M., Owen, L.A., Caffee, M.W., Kamp, U., 2011b. Catastrophic partial drainage of Pangong Tso, northern India and Tibet. Geomorphology 125, 109–121. DOI: 10.1016/j.geomorph.2010.08.017

Dortch, J.M., Owen, L.A., Caffee, M.W., 2010a. Quaternary glaciation in the Nubra and Shyok valley confluence, northernmost Ladakh, India. Quaternary Research 74, 132–144. https://doi.org/10.1016/j.yqres.2010.04.013

Dortch, J.M., Owen, L.A., Caffee, M.W., 2013. Timing and climatic drivers for glaciation across semi-arid western Himalayan–Tibetan Orogen. Quaternary Science Reviews 78, 188–208. https://doi.org/10.1016/j.quascirev.2013.07.025

Dortch, J.M., Owen, L.A., Caffee, M.W., Brease, P., 2010b. Late Quaternary glaciation and equilibrium line altitude variations of the McKinley River region, central Alaska Range. Boreas 39, 233–246. https://doi.org/10.1111/j.1502-3885.2009.00121.x

Dortch, J.M., Owen, L.A., Caffee, M.W., Li, D., Lowell, T.V., 2010c. Beryllium-10 surface exposure dating of glacial successions in the central Alaska Range. Journal of Quaternary Science 25, 1259–1269. https://doi.org/10.1002/jqs.1406

Douglass, D.C., Singer, B.S., Kaplan, M.R., Mickelson, D.M., Caffee, M.W., 2006. Cosmogenic nuclide surface exposure dating of boulders on last-glacial and late-glacial moraines, Lago Buenos Aires, Argentina: Interpretive strategies and paleoclimate implications. Quaternary Geochronology, Terrestrial Cosmogenic Nuclides 1, 43–58. https://doi.org/10.1016/j.quageo.2006.06.001

Dühnforth, M., Anderson, R.S., Ward, D., Stock, G.M., 2010. Bedrock fracture control of glacial erosion processes and rates. Geology 38, 423–426. https://doi.org/10.1130/G30576.1

Dunai, T., 2010. Cosmogenic Nuclides: Principles, Concepts and Applications in the Earth Surface Sciences, 1 edition. Cambridge University Press, Cambridge, UK ; New York.

Fenton, C.R., Niedermann, S., Dunai, T., Binnie, S.A., 2019. The SPICE project: Production rates of cosmogenic $^{21}$Ne, $^{10}$Be, and $^{14}$C in quartz from the 72 ka SP basalt flow, Arizona, USA. Quaternary Geochronology 54, 101019. https://doi.org/10.1016/j.quageo.2019.101019

Goehring, B.M., Muzikar, P., Lifton, N.A. Establishing a Bayesian approach to determining cosmogenic nuclide reference production rates using He-3. Earth and Planetary Science Letters 481, 91–100. https://doi.org/10.1016/j.epsl.2017.10.025

Gosse, J.C., 2005. The Contribution of Cosmogenic Nuclides to Unraveling Alpine Paleoclimate Histories, in: Huber, U.M., Bugmann, H.K.M., Reasoner, M.A. (Eds.), Global Change and Mountain Regions: An Overview of Current Knowledge, Advances in Global Change Research. Springer Netherlands, Dordrecht, pp. 39–49. https://doi.org/10.1007/1-4020-3508-X_5

Gosse, J.C., Evenson, E.B., Klein, J., Lawn, B., Middleton, R., 1995. Precise cosmogenic $^{10}$Be measurements in western North America: Support for a global Younger Dryas cooling event. Geology 23, 877–880. https://doi.org/10.1130/0091-7613(1995)023<0877:PCBMIW>2.3.CO;2

Gosse, J.C., Phillips, F.M., 2001. Terrestrial in situ cosmogenic nuclides: Theory and application. Quaternary Science Reviews 20, 1475–1560. https://doi.org/10.1016/S0277-3791(00)00171-2

Hancock, G. S., Anderson, R.S., Chadwick, O.A., Finkel, R.C., 1999. Dating fluvial terraces with 10Be and 26Al profiles: Application to the Wind River, Wyoming, Geomorphology, 27, 41–60. https://doi.org/10.1016/S0169-555X(98)00089-0

Hedrick, K.A., Seong, Y.B., Owen, L.A., Caffee, M.C., Dietsch, C., 2011. Towards defining the transition in style and timing of Quaternary glaciation between the monsoon-influenced Greater Himalaya and the semi-arid Transhimalaya of Northern India. Quaternary International 236, 21–33. https://doi:10.1016/j.quaint.2010.07.023

Hallet, B., 1996. Glacial quarrying: A simple theoretical model. Annals of Glaciology 22, 1–8. https://doi.org/10.3189/1996AoG22-1-1-8

Hallet, B., Putkonen, J., 1994. Surface dating of dynamic landforms: Young boulders on aging moraines. Science 265, 937–940. https://doi.org/10.1126/science.265.5174.937

Hein, A.S., Fogwill, C.J., Sugden, D.E., Xu, S., 2014. Geological scatter of cosmogenic-nuclide exposure ages in the Shackleton Range, Antarctica: Implications for glacial history. Quaternary Geochronology, Tracking the Pace of Quaternary Landscape Change with Cosmogenic Nuclides 19, 52–66. https://doi.org/10.1016/j.quageo.2013.03.008

Heyman, J., Applegate, P.J., Blomdin, R., Gribenski, N., Harbor, J.M., Stroeven, A.P., 2016. Boulder height—Exposure age relationships from a global glacial $^{10}$Be compilation. Quaternary Geochronology 34, 1–11. https://doi.org/10.1016/j.quageo.2016.03.002

Heyman, J., Stroeven, A.P., Harbor, J.M., Caffee, M.W., 2011. Too young or too old: Evaluating cosmogenic exposure dating based on an analysis of compiled boulder exposure ages. Earth and Planetary Science Letters 302, 71–80. https://doi.org/10.1016/j.epsl.2010.11.040

Hughes, P.D., Gibbard, P.l., Ehlers, J., 2013., Timing of glaciation during the last glacial cycle: evaluating the concept of a global 'Last Glacial Maximum' (LGM), Earth-Science Reviews, 125, 171–198. https://doi.org/10.1016/j.earscirev.2013.07.003

Iverson, N.R., 2012. A theory of glacial quarrying for landscape evolution models. Geology 40, 679–682. https://doi.org/10.1130/G33079.1

Ivy-Ochs, S., Kerschner, H., Kubik, P.W., Schlüchter, C., 2006. Glacier response in the European Alps to Heinrich Event 1 cooling: the Gschnitz stadial. Journal of Quaternary Science 21, 115–130. https://doi.org/10.1002/jqs.955

Ivy-Ochs, S., Kerschner, H., Reuther, A., Preusser, F., Heine, K., Maisch, M., Kubik, P.W., Schlüchter, C., 2008. Chronology of the last glacial cycle in the European Alps. Journal of Quaternary Science 23, 559–573. https://doi.org/10.1002/jqs.1202

Ivy-Ochs, S., Kerschner, H., Schlüchter, C., 2007. Cosmogenic nuclides and the dating of Late glacial and Early Holocene glacier variations: The Alpine perspective. Quaternary International, From the Swiss Alps to the Crimean Mountains—Alpine Quaternary Stratigraphy in a European Context 164–165, 53–63. https://doi.org/10.1016/j.quaint.2006.12.008

Jones, R.S., Small, D., Cahill, N., Bentley, M.J., Whitehouse, P.L., 2019. iceTEA: Tools for plotting and analysing cosmogenic-nuclide surface-exposure data from former ice margins. Quaternary Geochronology 51, 72–86. https://doi.org/10.1016/j.quageo.2019.01.001

Jull, A.J.T., Scott, E.M., Bierman, P., 2015. The CRONUS-Earth inter-comparison for cosmogenic isotope analysis. Quaternary Geochronology 26, 3–10. https://doi.org/10.1016/j.quageo.2013.09.003

Kaplan, M.R., Miller, G.H., 2003. Early Holocene delevelling and deglaciation of the Cumberland Sound region, Baffin Island, Arctic Canada. Geological Society of America Bulletin 115, 445–462. https://doi.org/10.1130/0016-7606(2003)115<0445:EHDADO>2.0.CO;2

Kelly, M.A., Lowell, T.V., Hall, B.L., Schaefer, J.M., Finkel, R.C., Goehring, B.M., Alley, R.B., Denton, G.H., 2008. A $^{10}$Be chronology of lateglacial and Holocene mountain glaciation in the Scoresby Sund region, east Greenland: Implications for seasonality during lateglacial time. Quaternary Science Reviews 27, 2273–2282. https://doi.org/10.1016/j.quascirev.2008.08.004

Kwak, S.G., Kim, J.H., 2017. Central limit theorem: The cornerstone of modern statistics. Korean Journal of Anaesthesiology 70, 144–156. https://doi.org/10.4097/kjae.2017.70.2.144

Lal, D., 1991. Cosmic ray labeling of erosion surfaces: In situ nuclide production rates and erosion models. Earth and Planetary Science Letters 104, 424–439. https://doi.org/10.1016/0012-821X(91)90220-C

Levenberg, K., 1944. A method for the solution of certain non-linear problems in least squares. Quarterly of Applied Mathematics 2, 164–168.

Leys, C., Ley, C., Klein, O., Bernard, P., Licata, L., 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. Journal of Experimental Social Psychology 49, 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Lifton, N., 2016. Implications of two Holocene time-dependent geomagnetic models for cosmogenic nuclide production rate scaling. Earth and Planetary Science Letters 433, 257–268. https://doi.org/10.1016/j.epsl.2015.11.006

Lifton, N., Sato, T., Dunai, T.J., 2014. Scaling in situ cosmogenic nuclide production rates using analytical approximations to atmospheric cosmic-ray fluxes. Earth and Planetary Science Letters 386, 149–160. https://doi.org/10.1016/j.epsl.2013.10.052

Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. Journal of the Society for Industrial and Applied Mathematics 11, 431–441. https://doi.org/10.1137/0111030

Martin, L.C.P., Blard, P.-H., Balco, G., Lavé, J., Delunel, R., Lifton, N., Laurent, V., 2017. The CREp program and the ICE-D production rate calibration database: A fully parameterizable and updated online tool to compute cosmic-ray exposure ages. Quaternary Geochronology 38, 25–49. https://doi.org/10.1016/j.quageo.2016.11.006

Martin, L.C.P., Blard, P.-H., Lavé, J., Jomelli, J., Charreau, J., Condom, T., Lupker, M., ASTER Team. Antartic-like temperature variations in the Tropical Andes recorded by glaciers and lakes during the last deglaciation. Quaternary Science Reviews, 247, 106524. https://doi.org/10.1016/j.quascirev.2020.106542

Menounos, B., Goehring, B.M., Osborn, G., Margold, M., Ward, B., Bond, J., Clarke, G.K., Clague, J.J., Lakeman, T., Koch, J., Caffee, M.W., 2017. Cordilleran Ice Sheet mass loss preceded climate reversals near the Pleistocene Termination. *Science*, 358(6364), pp.781-784. https://doi.org/10.1126/science.aan3001

Molnar, P., Brown, E.T., Burchfiel, B.C., Deng, Q., Feng, X., Li, J., Raisbeck, G.M., Shi, J., Zhangming, W., Yiou, F., You, H., 1994. Quaternary climate change and the formation of river terraces across growing anticlines on the north flank of the Tien Shan, China. The Journal of Geology 102, 583–602. https://doi.org/10.1086/629700

Marrero, S. M., Phillips, F. M., Borchers, B., Lifton, N., Aumer, R., Balco, G., 2016. Cosmogenic nuclide systematics and the CRONUScalc program. Quaternary 1300 Geochronology 31, 160–187. https://doi.org/10.1016/j.quageo.2015.09.005

Moré, J.J., Sorensen, D.C., 1983. Computing a trust region step. SIAM Journal on Scientific and Statistical Computing 4, 553–572. https://doi.org/10.1137/0904038

Murari, M.K., Owen, L.A., Dortch, J.M., Caffee, M.W., Dietsch, C., Fuchs, M., Haneberg, W.C., Sharma, M.C., Townsend-Small, A., 2014. Timing and climatic drivers for glaciation across monsoon-influenced regions of the Himalayan-Tibetan Orogen. Quaternary Science Reviews 88, 159–182. https://doi.org/10.1016/j.quascirev.2014.01.013

Muscheler, R., Beer, J., Kubik, P.W., Synal, H.-A., 2005. Geomagnetic field intensity during the last 60,000 years based on $^{10}$Be and $^{36}$Cl from the Summit ice cores and $^{14}$C. Quaternary Science Reviews 24, 1849–1860. https://doi.org/10.1016/j.quascirev.2005.01.012

Muzikar, P., Goehring, B., Lifton, N. Handling overdispersion in CRONUS-EARTH intercomparison measurements: a Bayesian approach. Radiocarbon 59, 1133–1145. DOI:10.1017/RDC.2017.16

Nishiizumi, K., Lal, D., Klein, J., Middleton, R., Arnold, J.R., 1986. Production of $^{10}$Be and $^{26}$Al by cosmic rays in terrestrial quartz in situ and implications for erosion rates. Nature 319, 134–136. https://doi.org/10.1038/319134a0

Nishiizumi, K., Winterer, E.L., Kohl, C.P., Klein, J., Middleton, R., Lal, D., Arnold, J.R., 1989. Cosmic ray production rates of $^{10}$Be and $^{26}$Al in quartz from glacially polished rocks. Journal of Geophysical Research: Solid Earth 94, 17907–17915. https://doi.org/10.1029/JB094iB12p17907

Owen, L.A., Dortch, J.M., 2014. Nature and timing of Quaternary glaciation in the Himalayan-Tibetan Orogen. Quaternary Science Reviews 88, 14–54. https://doi.org/10.1016/j.quascirev.2013.11.016

Pallàs, R., Rodés, Á., Braucher, R., Bourlès, D., Delmas, M., Calvet, M., Gunnell, Y., 2010. Small, isolated glacial catchments as priority targets for cosmogenic surface exposure dating of Pleistocene climate fluctuations, southeastern Pyrenees. Geology 38, 891–894. https://doi.org/10.1130/G31164.1

Pavón-Carrasco, F.J., Osete, M.L., Torta, J.M., De Santis, A., 2014. A geomagnetic field model for the Holocene based on archaeomagnetic and lava flow data. Earth and Planetary Science Letters 388, 98–109. https://doi.org/10.1016/j.epsl.2013.11.046

Phillips, F.M., Zreda, M.G., Smith, S.S., Elmore, D., Kubik, P.W., Sharma, P., 1990. Cosmogenic chlorine-36 chronology for glacial deposits at Bloody Canyon, eastern Sierra Nevada. Science 248, 1529–1532. https://doi.org/10.1126/science.248.4962.1529

Pratt-Sitaula, B., Burbank, D.W., Heimsath, A.M., Humphrey, N.F., Oskin, M., Putkonen, J., 2011. Topographic control of asynchronous glacial advances: a case study from Annapurna, Nepal. Geophysical Research Letters 38, L24502. http://dx.doi.org/10.1029/2011GL049940.

Putkonen, J., Connolly, J., Orloff, T., 2008. Landscape evolution degrades the geologic signature of past glaciations. Geomorphology, Glacial Landscape Evolution—Implications for Glacial Processes, Patterns and Reconstructions 97, 208–217. https://doi.org/10.1016/j.geomorph.2007.02.043

Putkonen, J., O'Neal, M., 2006. Degradation of unconsolidated Quaternary landforms in the western North America. Geomorphology, Quaternary Landscape Change and Modern Process in Western North America 75, 408–419. https://doi.org/10.1016/j.geomorph.2005.07.024

Putkonen, J., Swanson, T., 2003. Accuracy of cosmogenic ages for moraines. Quaternary Research 59, 255–261. https://doi.org/10.1016/S0033-5894(03)00006-1

Putnam, A.E., Bromley, G.R.M., Rademaker, K., Schaefer, J.M., 2019. In situ $^{10}$Be production-rate calibration from a $^{14}$C-dated late-glacial moraine belt in Rannoch Moor, central Scottish Highlands. Quaternary Geochronology 50, 109–125. https://doi.org/10.1016/j.quageo.2018.11.006

Putnam, A.E., Schaefer, J.M., Denton, G.H., Barrell, D.J.A., Andersen, B.G., Koffman, T.N.B., Rowan, A.V., Finkel, R.C., Rood, D.H., Schwartz, R., Vandergoes, M.J., Plummer, M.A., Brocklehurst, S.H., Kelley, S.E., Ladig, K.L., 2013a. Warming and glacier recession in the Rakaia valley, Southern Alps of New Zealand, during Heinrich Stadial 1. Earth and Planetary Science Letters 382, 98–110. https://doi.org/10.1016/j.epsl.2013.09.005

Putnam, A.E., Schaefer, J.M., Denton, G.H., Barrell, D.J.A., Birkel, S.D., Andersen, B.G., Kaplan, M.R., Finkel, R.C., Schwartz, R., Doughty, A.M., 2013b. The Last Glacial Maximum at 44°S documented by a $^{10}$Be moraine chronology at Lake Ohau, Southern Alps of New Zealand. Quaternary Science Reviews 62, 114–141. https://doi.org/10.1016/j.quascirev.2012.10.034

Rinterknecht, V.R., Clark, P.U., Raisbeck, G.M., Yiou, F., Bitinas, A., Brook, E.J., Marks, L., Zelčs, V., Lunkka, J.-P., Pavlovskaya, I.E., Piotrowski, J.A., Raukas, A., 2006. The last deglaciation of the southeastern sector of the Scandinavian ice sheet. Science 311, 1449–1452. https://doi.org/10.1126/science.1120702

Rosner, B., 1983. Percentage points for a generalized ESD many-outlier procedure. Technometrics 25, 165–172. https://doi.org/10.2307/1268549

Saha, S., Owen, L.A., Orr, E.N., Caffee, M.W., 2019. High-frequency Holocene glacier fluctuations in the Himalayan-Tibetan Orogen. Quaternary Science Reviews 220, 372–400. https://doi.org/10.1016/j.quascirev.2019.07.021

Saha, S., Owen, L.A., Orr, E.N., Caffee, M.W., 2018. Timing and nature of Holocene glacier advances at the northwestern end of the Himalayan-Tibetan Orogen. Quaternary Science Reviews 187, 177–202. https://doi.org/10.1016/j.quascirev.2018.03.009

Seong, Y.B., Owen, L.A., Bishop, M.P., Bush, A., Clendon, P., Copland, P., Finkel, R.C., Kamp, U., Shroder, J.F., 2007. Quaternary glacial history of the Central Karakoram. Quaternary Science Reviews 26, 3384–3405. https://doi:10.1016/j.quascirev.2007.09.015

Schäfer, J.M., Oberholzer, P., Zhao, Z.Z., Ivy-Ochs, S., Wieler, R., Baur, H., Kubik, P.W., Schluchter, C., 2008. Cosmogenic beryllium-10 and neon-21 dating of late Pleistocene glaciations in Nyalam, monsoonal Himalayas. Quaternary Science Reviews 27, 295–311. https://doi.org/10.1016/j.quascirev.2007.10.014

Schildgen, T.F., Phillips, W.M., Purves, R.S., 2005. Simulation of snow shielding corrections for cosmogenic nuclide surface exposure studies. Geomorphology 64, 67–85. https://doi.org/10.1016/j.geomorph.2004.05.003

Scott, D.W., 2015. Kernel Density Estimators, in: Multivariate Density Estimation. John Wiley & Sons, Ltd, pp. 137–216. https://doi.org/10.1002/9781118575574.ch6

Sheather, S.J., Jones, M.C., 1991. A reliable data-based bandwidth selection method for kernel density estimation. Journal of the Royal Statistical Society. Series B (Methodological) 53, 683–690.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Monographs on Statistics and Applied Probability, London: Chapman and Hall/CRC, 1986.

Small, D., Fabel, D., 2016. Was Scotland deglaciated during the Younger Dryas? Quaternary Science Reviews 145, 259–263. https://doi.org/10.1016/j.quascirev.2016.05.031

Stone, J.O., 2000. Air pressure and cosmogenic isotope production. Journal of Geophysical Research: Solid Earth 105, 23753–23759. https://doi.org/10.1029/2000JB900181

Stübner, K., Bookhagen, B., Merchel, S., Lachner, J., Gadoev, M., 2021. Unravelling the Pleistocene glacial history of the Pamir Mountains, Central Asia. Quaternary Science Reviews 257, 106857. https://doi.org/10.1016/j.quascirev.2021.106857

Stroeven, A.P., Heyman, J., Fabel, D., Björck, S., Caffee, M.W., Fredin, O., Harbor, J.M., 2015. A new Scandinavian reference $^{10}$Be production rate. Quaternary Geochronology 29, 104–115. https://doi.org/10.1016/j.quageo.2015.06.011

Taylor, B.N., Kuyatt, C.E., 1994. Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results (NIST Technical Note No. 1297). National Institute of Standards and Technology, Gaithersburg, Md.

Taylor, J.R., 1997. Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements, 2 edition. University Science Books, U.S., Sausalito, Calif.

Tomkins, M.D, Dortch, J.M., Hughes, P.D., Huck, J., Pallàs, R., Rodés, Á., Allard, J.L., Stimson, A., Bourlès, D., Rinterknecht, V.R., Rodríguez-Rodríguez, L., Jomelli, V., Copons, R., Barr, I.D., Darvill, C.M., Bishop, T., 2021. Moraine crest or slope: An analysis on the effects of boulder position on cosmogenic exposure age. Earth and Planetary Science Letters.

Tylmann, K., Woźniak, P.P., Rinterknecht, V.R., 2018. Erratics selection for cosmogenic nuclide exposure dating—An optimization approach. Baltica 31, 100–114. https://doi.org/10.5200/baltica.2018.31.10

Ugelvig, S.V., Egholm, D.L., Anderson, R.S., Iverson, N.R., 2018. Glacial erosion driven by variations in meltwater drainage. Journal of Geophysical Research: Earth Surface 123, 2863–2877. https://doi.org/10.1029/2018JF004680

Uppala, S.M., KÅllberg, P.W., Simmons, A.J., Andrae, U., Bechtold, V.D.C., Fiorino, M., Gibson, J.K., Haseler, J., Hernandez, A., Kelly, G.A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R.P., Andersson, E., Arpe, K., Balmaseda, M.A., Beljaars, A.C.M., Berg, L.V.D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B.J., Isaksen, L., Janssen, P.A.E.M., Jenne, R., Mcnally, A.P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N.A., Saunders, R.W., Simon, P., Sterl, A., Trenberth, K.E., Untch, A., Vasiljevic, D., Viterbo, P., Woollen, J., 2005. The ERA-40 re-analysis. Quarterly Journal of the Royal Meteorological Society 131, 2961–3012. https://doi.org/10.1256/qj.04.176

Wendt, I., Carl, C., 1991. The statistical distribution of the mean squared weighted deviation. Chemical Geology: Isotope Geoscience Section 86, 275–285. https://doi.org/10.1016/0168-9622(91)90010-T

Wilson, P., Schnabel, C., Wilcken, K.M., Vincent, P.J., 2013. Surface exposure dating ($^{36}$Cl and $^{10}$Be) of post-Last Glacial Maximum valley moraines, Lake District, northwest England: Some issues and implications. Journal of Quaternary Science 28, 379–390. https://doi.org/10.1002/jqs.2628

Zreda, M.G., Phillips, F.M., Elmore, D., 1994. Cosmogenic $^{36}$Cl accumulation in unstable landforms: 2. Simulations and measurements on eroding moraines. Water Resources Research Res. 30, 3127–3136. https://doi.org/10.1029/94WR00760