A STOCHASTIC SAMPLING APPROACH TO ZIRCON ERUPTION AGE INTERPRETATION

PREPRINT, COMPILED OCTOBER 1, 2018

C. Brenhin Keller[1,2*], Blair Schoene[3], and Kyle M. Samperton[4]

[1]Berkeley Geochronology Center, Berkeley, CA 94709
[2]Department of Earth and Planetary Science, University of California, Berkeley, CA 94720
[3]Department of Geosciences, Guyot Hall, Princeton University, Princeton, NJ 08544
[4]Nuclear and Chemical Sciences Division, Lawrence Livermore National Laboratory, Livermore, CA 94550

ABSTRACT

The accessory mineral zircon is widely used to constrain the timing of igneous processes such as magma crystallization or eruption. However, zircon U-Pb ages record zircon crystallization, which is not an instantaneous process. Zircon saturation calculations link zircon crystallization, temperature, and melt fraction, allowing for the estimation of zircon crystallization distributions as a function of time or temperature. Such distributions provide valuable prior information, enabling Bayesian estimates of magma eruption time and allowing for comparison of the relative accuracy of common weighted-mean and youngest-zircon age interpretations with synthetic datasets. We find that both traditional interpretations carry a risk of underestimating the uncertainty in eruption age; a low mean square of weighted deviates (MSWD) does not guarantee the accuracy of weighted mean interpretations. In the absence of independent confirmation that crystallization timescale is short relative to analytical uncertainties, a Bayesian approach frequently provides the most accurate results and is least likely to underestimate uncertainty. Since U-Pb zircon studies now routinely resolve geological age dispersion due to increasing analytical precision, such considerations are increasingly critical to future progress in resolving rates and dates of Earth processes.

Geochem. Persp. Let. 10.7185/geochemlet.1826

INTRODUCTION

Absolute time constraints are critical for establishing temporal correlations, testing casual relationships, and quantifying rates and durations throughout the Earth sciences (Reiners et al., 2018). However, the time of radioisotopic closure may not directly date the geological events or processes of interest. Throughout the first century of geochronology, this potential mismatch was frequently a minor concern compared to analytical uncertainties at the percent level or greater. Recently, however, continual improvements in analytical precision and accuracy have fundamentally altered longstanding assumptions of geochronological age interpretation (Schoene, 2014).

One chronometer of particular interest is the U-Pb system in zircon, thanks to zircon's ubiquity, resilience, and tendency to exclude initial daughter isotopes. Throughout the geologic record, zircon provides crucial time constraints for processes ranging from evolution and mass extinction to magmatism and crustal differentiation (Bowring et al., 1993; Mundil et al., 2004; Harrison, 2009; Schoene et al., 2015; Samperton et al., 2017). Due to extremely slow parent and daughter isotope diffusion (Cherniak, 2003), zircon U-Pb ages record zircon crystallization, if not compromised by metamictization and subsequent Pb-loss. However, the crystallization of a suite of zircons in a single igneous rock sample has often been assumed to occur rapidly relative to analytical uncertainty, justifying the use of statistical approaches such as the weighted mean (*e.g.*, Bowring et al., 1993). Moreover, even though zircon saturation in magmas is empirically well understood and distinct from whole-rock crystallization to the solidus (Boehnke et al., 2013), U-Pb zircon ages are traditionally interpreted within uncertainty as reflecting bulk crystallization or eruption.

While such assumptions may be justified for sufficiently ancient or homogeneous samples, crystallization timescales may span 200-700 kyr for magmatic zircons (Lissenberg et al., 2009; Wotzlaw et al., 2013; Samperton et al., 2017). Consequently, diachronous crystallization or recrystallization must be considered before calculating a weighted mean of zircon ages derived from either in-situ (SIMS, LA-ICPMS) or bulk (TIMS) analytical techniques. For instance, modern U-Pb Chemical Abrasion – Isotope Dilution TIMS (CA-ID-TIMS) ages (Mattinson, 2005) on single zircons and zircon fragments may surpass 0.05% (2σ) accuracy and precision (*e.g.* Samperton et al., 2017; Schoene et al., 2015) – equivalent to 50 kyr in a 100 Ma sample. Consequently, zircon crystallisation age heterogeneity is increasingly clearly resolved in a wide range of magmatic contexts (Wotzlaw et al., 2013; Samperton et al., 2017).

Analogous issues appear in other geochronological applications ranging from the interpretation of anomalously dispersed Ar-Ar ages (e.g. Ellis et al., 2017) to the estimation of sedimentary depositional ages from detrital mineral geochronology. We consider here the case study of eruption age estimation by CA-ID-TIMS zircon geochronology, where analytical precision is high and confounding open-system behavior is relatively well controlled. Here, a plethora of competing age interpretations have developed in the literature, falling into three broad categories (e.g., Samperton et al., 2015) shown in Fig. 1f.

(1) Weighted mean In cases where the variance of the dataset is plausibly consistent with analytical uncertainty alone, then some authors may calculate a weighted mean of the entire dataset (*e.g.*, Crowley et al., 2007).

(2) Youngest zircon In contrast, where there is an expectation of slow crystallization relative to analytical un-

certainty, or ages are highly dispersed, the youngest single analysis may be considered a better estimate of eruption age (*e.g.*, Wotzlaw et al., 2013).

(3) Low-MSWD weighted mean As an intermediate between (1) and (2), one may calculate a weighted mean of only the N youngest analyses such that, given the acceptance distribution of the MSWD (Wendt and Carl, 1991), the MSWD of this subpopulation does not exceed a value deemed acceptable for N analyses (*e.g.*, Schoene et al., 2015).

The possibility for residual lead loss, even following chemical abrasion, substantially complicates each of these three approaches, as does the common practice of excluding outliers subjectively identified as antecrysts. While interpretation (1) is likely to systematically predate the true eruption age, the accuracy of (2) and (3) has not been well tested. Moreover, while each interpretation has advantages, it is not clear that any of the three yields a statistically robust estimate of eruption age.

MODEL CONFIGURATION

In order to address these problems, we investigate the performance of common weighted-mean, youngest-zircon, and MSWD-test age interpretations, as well as that of an alternative likelihood-based Bayesian approach. To this end, we consider two dimensionless variables which, together with the pre-eruptive zircon crystallization distribution $f(t_r)$, determine the behavior of all possible volcanic zircon age interpretations. The first is $\boldsymbol{\Delta t/\sigma}$, the ratio of the true crystallization timescale Δt to analytical uncertainty σ, while the second is simply $\boldsymbol{N}$, the number of analyses (Fig. 1).

For instantaneous crystallization ($\Delta t/\sigma = 0$) with Gaussian analytical uncertainty, both the mean and variance of an analytical dataset are constant as a function of N, and a weighted mean interpretation is fully justified. However, these assumptions fail for nontrivial Δt, leading to systematic bias and potentially major overestimation of accuracy and precision at high N. In contrast, at high $\Delta t/\sigma$, a youngest zircon estimate is likely to outperform a weighted mean interpretation, but may systematically pre- or post-date the true eruption age as a function of N.

The effectiveness of each approach will depend on $f(t_r)$. Fortunately, magmatic zircon crystallization behavior is understood via empirical saturation equations (Boehnke et al., 2013), kinetic models (Watson, 1996), and observation of natural systems (*e.g.*, Samperton et al., 2017). In particular, we consider the mass of zircon crystallized per unit time or temperature *per unit mass of magma*. This *intensive* distribution should not be confused with the zircon populations considered by Caricchi et al. (2014, 2016), who assume constant zircon crystallization rate per unit magma in the saturation interval (i.e., a flat line in Fig. 1a) in their attempt to estimate pluton-scale magma fluxes, which we do not consider here.

Watson (1996) was the first to consider the form of the relative zircon crystallization distribution as a function of temperature, calculating a theoretical distribution on the basis of kinetic constraints, characterized by a rapid onset of zircon crystallization followed by a gradual decline. We also consider

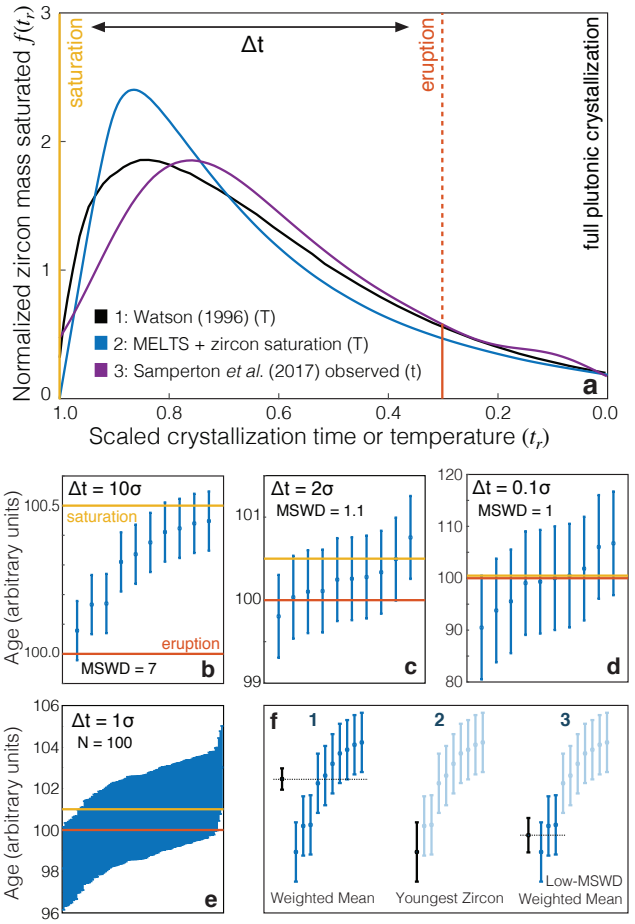


Figure 1: Zircon distributions. **a:** Theoretical and empirical relative zircon crystallization distributions $f(t_r)$, scaled from initiation to termination of zircon crystallization. 1: Kinetic model of Watson (1996), based on zirconium diffusion constraints. 2: Thermodynamic model of Keller et al. (2017) using MELTS calculations. 3: Observed zircon crystallization distributions of Samperton et al. (2017), shown as a kernel density estimate for all autocrystic zircons, truncated at +/- 1 kernel bandwidth. **b-d:** Representative synthetic zircon age datasets for a variety of $\Delta t/\sigma$ at $N = 10$. **e:** Example dataset with N=100 at $\Delta t = 1\sigma$; note the range is greater than in **c** despite lower Δt. **f:** Schematic illustration of the three most common volcanic zircon age interpretations.

a thermodynamic model integrating major and trace element evolution with empirical zircon saturation equations (Boehnke et al., 2013; Keller et al., 2017), as well as an observed average plutonic zircon distribution (as a function of time) derived from CA-ID-TIMS of both single zircons and sub-grain zircon fragments (Samperton et al., 2017). All three approaches yield similar distributions, (Fig. 1a) – a consistency that extends in thermodynamic models to a wide range of whole-rock compositions despite greatly varying saturation conditions (Supp. Fig. 1). Distributions #1 and #2 assume linear cooling of a single magma batch; interaction of multiple magma batches and variable cooling rates may distort the distribution, though an abrupt truncation at eruption must feature in all volcanic zircon age spectra.

Given such a prior expectation of the form of $f(t_r)$, we may quantitatively test the performance of each common age interpretation as a function of N and $\Delta t/\sigma$ by drawing N zircons from a crystallization distribution with arbitrary saturation and eruption ages, adding analytical uncertainty as a Gaussian random variable with variance σ^2 relative to the Δt of the distribution, applying each age interpretation to the resulting synthetic dataset, and repeating the process very many times for each N and $\Delta t/\sigma$ of interest. However, due to the consistency of theoretical and empirical zircon crystallization distributions, we may also use such a distribution as prior information to constrain a likelihood-based Bayesian eruption age estimator as follows. Given an observed dataset and an accurate $f(t_r)$, one may readily calculate the likelihood of obtaining the observed dataset from the crystallization distribution for any given saturation time and eruption time.

While a maximum-likelihood solution might be found by systematically varying both the saturation and eruption age to produce a two-dimensional likelihood surface (*e.g.*, Supp. Fig. 2), such an exhaustive search would be inefficient. Instead, we follow the Metropolis algorithm to estimate the distribution of the eruption age, exploring the likelihood space by moving from its current position to a proposed position in the likelihood space with probability equal to the ratio of proposed and current likelihoods (maximum 1), with each proposal deviating from the previous position in only one dimension at a time (Gelman et al., 2013). After an initial period of equilibration, the series of accepted proposals takes the form of the stationary distribution of a Markov chain (Supp. Fig. 3), which provides both the mean and variance of estimated zircon saturation and eruption ages. In order to test the sensitivity of this approach to the choice of $f(t_r)$, we calculate Bayesian eruption age estimates using (1) the MELTS crystallization distribution (Fig. 1a) from which the synthetic data were drawn, (2) a uniform relative crystallization distribution (i.e., a flat line in Fig. 1a), and (3) a "bootstrapped" distribution, a truncated kernel density estimate of each synthetic dataset (Methods).

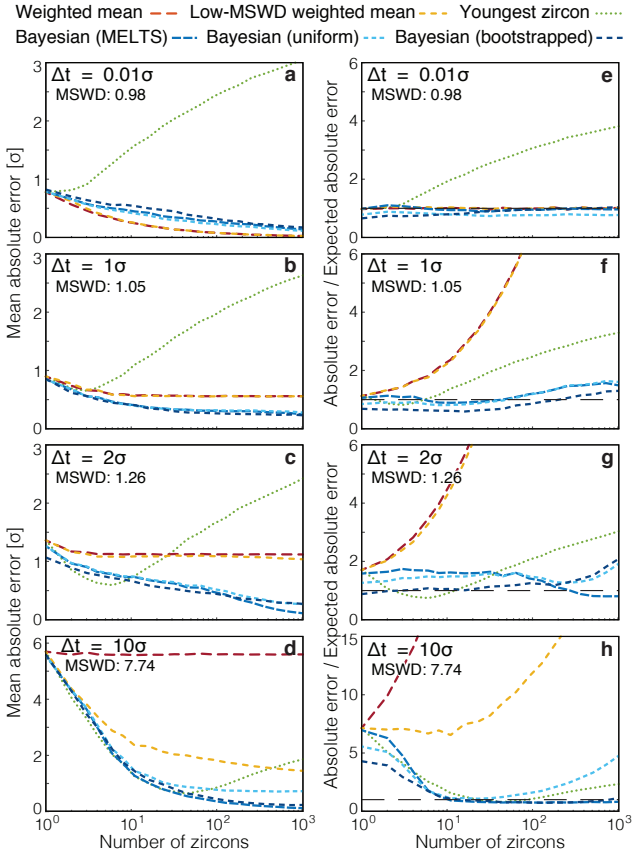


Figure 2: Performance of each age interpretation as a function of N and $\Delta t/\sigma$. **a-d:** Mean absolute error is the mean absolute deviation of the model result from the true value; lower absolute errors are better. **e-h:** Accuracy of the model uncertainty for each age interpretation. A value greater than 1.0 indicates an underestimation of the model uncertainty (*i.e.*, overprecision), while a value lower than 1.0 indicates an overestimation of the model uncertainty. MSWD in each panel is the average mean square of weighted deviation (also known as the reduced chi-squared statistic) for that $\Delta t/\sigma$ over all N. Each datum reflects the mean of 1200 synthetic dataset tests; standard error of the mean is on the order of the line width.

RESULTS

We explore the parameter space from $\Delta t/\sigma$ of 0.01 to 10 and N of 1 to 1000, which includes ranges applicable to both ID-TIMS and in-situ geochronological techniques. As expected, weighted means are accurate at very low $\Delta t/\sigma$, with the lowest absolute error and accurate reported uncertainty at $\Delta t = 0.01\sigma$ (Fig. 2a,e), but fail at high $\Delta t/\sigma$, with absolute error not lower than $\Delta t/2$ and highly inaccurate reported uncertainty (Fig. 2d,h). Conversely, the youngest zircon approach performs poorly at $\Delta t = 0.01\sigma$ with high absolute error and substantial overprecision, but comparatively well at $\Delta t = 10\sigma$. Such interpretations might remain useful if $\Delta t/\sigma$ were readily determinable for natural datasets.

Problems emerge at intermediate levels of age dispersion. At $\Delta t = 1\sigma$, all three traditional interpretations begin to fail visibly above $N = 3$, with high absolute error in youngest-zircon interpretations, and underestimated uncertainty in both weighted mean and low-MSWD weighted mean interpretations, for instance by a factor of two at $N = 10$ (Fig. 2b,f). At $\Delta t = 2\sigma$, the problems with weighted mean interpretations are accentuated, while youngest zircon interpretations coincidentally perform well at moderate N (due to competing biases which happen to cancel at $N = 5$ and $\Delta t = 2\sigma$), but ultimately still fails at high N due to analytical outliers.

These problems are compounded by the fact that the average MSWD at $\Delta t = 2\sigma$ is only 1.26, statistically indistinguishable from the near-unity MSWD of a dataset with $\Delta t = 0.01\sigma$ until one has characterized more than ~700 individual zircon analyses (Supp. Fig. 4). Even magmatic age heterogeneity as high as $\Delta t = 5\sigma$ is not clearly distinguishable from instantaneous crystallization on the basis of MSWD for datasets smaller than $N \approx 50$, and datasets with $\Delta t/\sigma$ less than two are generally indistinguishable from instantaneous crystallization at any practical N (Supp. Fig. 4).

In contrast, the Bayesian eruption age estimate yields slightly higher absolute error than the weighted mean at $\Delta t = 0.01\sigma$, but otherwise equals or outperforms all other approaches across a

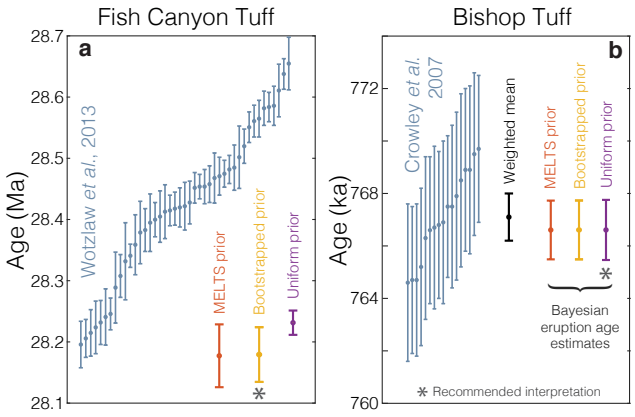


Figure 3: Bayesian eruption age estimates for two well-known volcanic zircon populations. For age spectra with well-resolved dispersion (**a**), a kernel density estimate may recover a close approximation of $f(t_r)$, obviating the need to identify and reject antecrysts. However, for age spectra where igneous dispersion is not well resolved (**b**), assuming a uniform $f(t_r)$ may be more parsimonious.

wide range of N and $\Delta t/\sigma$, with the closest to consistently accurate reported uncertainties. Notably, this result is not highly sensitive to the exact choice of $f(t_r)$, as the Bayesian estimate assuming a uniform $f(t_r)$ only diverges from the equivalent estimate assuming the true prior at high $\Delta t/\sigma$, and only then for N greater than typical in TIMS studies (ca. 5-20).

To explore the practical application of Bayesian eruption age estimation, we consider ID-TIMS datasets from two well-known supereruptions with contrasting zircon age spectra: the ~28 Ma Fish Canyon Tuff, with ~500 ka of continuous zircon age dispersion (Wotzlaw et al., 2013), and the more homogeneous ~767 ka Bishop Tuff (Crowley et al., 2007). As seen in Fig. 3, the results are suggestive of a youngest zircon interpretation for the Fish Canyon Tuff and a weighted mean interpretation for the Bishop Tuff – congruous with the dramatic difference in dispersion between the two datasets. If all zircons were strictly autocrystic, the presence of older outliers would suggest that we are incompletely sampling the zircon saturation distribution, and thus overestimating the eruption age. Including xenocrystic outliers in the Bayesian age interpretation thus counterintuitively leads to *under*estimation of the eruption age and divergence between Bayesian and weighted mean ages for the Bishop Tuff (*e.g.*, Supp. Fig. 5).

DISCUSSION

Considering the results of Fig. 2 in context of the variance of the MSWD (Supp. Fig. 4), for most natural datasets we cannot rely on sufficiently low $\Delta t/\sigma$ to justify a weighted mean interpretation even at low MSWD, nor in general can we justify a youngest-zircon interpretation except at low N and high MSWD. In the absence of reliable external evidence for instantaneous crystallization, the greater precision obtained by a weighted mean approach is illusory. A likelihood-based Bayesian estimate appears to perform competitively under all scenarios, and is the least likely to underestimate the reported uncertainty.

The problem of minimum age estimation may also be considered from the perspective of mixture modelling, which can be approached either numerically or analytically (Galbraith, 2005; Jasra et al., 2006), and is also likely to outperform the "traditional" interpretations. In this context, the main advantage of our approach compared to an analytical equivalent is merely the ability to numerically specify an arbitrary $f(t_r)$ derived from a physics-based model.

While our approach may decrease the impact of subjective interpretational decisions, it does not eliminate them entirely: we must still choose a method by which to estimate $f(t_r)$. This relative crystallization distribution is well-determined for a single magma batch (Fig. 1a #1-2), and may be empirically estimated ("bootstrapped") by a kernel density estimate in datasets with highly-resolved dispersion (Fig. 1a #3). Even when data are inverted with an $f(t_r)$ that does not match the distribution from which they were drawn, a Bayesian approach still significantly outperforms traditional interpretations (Fig. 2, Supp. Figs. 6 -7). Nonetheless, distortions may occur, particularly in datasets featuring extreme outliers, or if our critical assumption that $f(t_r)$ falls to zero at $t_r = 0$ (i.e., no crystallization after eruption) is violated by contamination, unrecognized lead loss, or other open-system behavior.

Finally, while the similarity in form between the observed and theoretical relative crystallization distributions $f(t_r)$ (Fig. 1a) suggests that single-zircon and zircon-fragment TIMS ages are sampling (and not simply integrating) the true $f(t_r)$, further systematic analysis is required. The ideal U-Pb zircon technique to resolve a such geochronological problems must provide high analytical accuracy, high spatial resolution, and the ability to mitigate lead loss. Consequently, we recommend sub-grain microsampling or microfracturing wherever possible in TIMS analyses (e.g. Samperton et al., 2017), and emphasize further study of the effects and detection of lead loss for all U-Pb techniques.

ACKNOWLEDGEMENTS

Thanks to B. Dyer and J.-F. Wotzlaw for discussion and to P. Vermeesch, K. Gallagher, U. Schaltegger, and T. Sheldrake for valuable reviews. C.B.K. was supported in part by the Department of Energy Computational Science Graduate Fellowship Program of the Office of Science and National Nuclear Security Administration under contract DE-FG02-97ER25308. Computational support was provided by the Princeton Institute for Computational Science and Engineering. Lawrence Livermore National Laboratory is operated by Lawrence Livermore National Security, LLC, for the U.S. Department of Energy, National Nuclear Security Administration under Contract DE-AC52-07NA27344. All source code is available at https://github.com/brenhinkeller/BayeZirChron.c

SUPPLEMENTARY METHODS

Eruption age estimation

Given a dataset of mass-spectrometrically determined closed-system mineral crystallization ages from a given volcanic unit, we wish to determine the time of eruption (or deposition) subject the one-sided *a priori* constraint that no such ages may postdate eruption. We represent this prior knowledge in the form of a crystallization distribution that is sharply truncated at eruption. In the first (and simplest) case of a single magma batch, the remarkable convergence of kinetic (Watson, 1996), thermodynamic (MELTS + zircon saturation), and empirical (Samperton et al., 2015) results seen in Figure 1a provides a relative zircon crystallization density function $f_{xtal}(t_r)$ where t_r is relative time, scaled from zircon saturation (t_{sat}) to eruptive truncation (t_{erupt}), that is:

$$t_r = (t - t_{erupt})/\Delta t \tag{1}$$

where

$$\Delta t = t_{sat} - t_{erupt} \tag{2}$$

When thus scaled, the form of this zircon density function remains consistent across a wide range of rock types, as seen in Supplementary Figure 1. More generally, for any system where we can independently determine $f_{xtal}(t_r)$, we may then define a mineral crystallization distribution $\mathcal{D}_{xtal}(t_{sat}, t_{erupt})$ with a normalized probability density function $p_{xtal}(t \mid t_{sat}, t_{erupt})$ given by:

$$p_{xtal}(t \mid t_{sat}, t_{erupt}) = \begin{cases} 0 & t < t_{eruption} \\ 0 & t > t_{saturation} \\ f_{xtal}(t_r)/\Delta t & t_{erupt} \leq t \leq t_{sat} \end{cases} \tag{3}$$

We then approach the estimation of t_{erupt} as a Bayesian parameter estimation problem. Central to this approach is the ability to calculate the likelihood that an observed zircon age was drawn from a given crystallization distribution, accounting for analytical uncertainty. For a single zircon i of age x_i and Gaussian analytical uncertainty with variance σ^2, this likelihood $\mathcal{L}$ is given by an integral over all time:

$$\mathcal{L}(x_i \mid t_{sat}, t_{erupt}) = \int_{-\infty}^{\infty} p_{xtal}(t \mid t_{sat}, t_{erupt}) * \frac{1}{\sqrt{2\,\pi\,\sigma^2}} \exp\left(-\frac{(x_i - t)^2}{2\sigma^2}\right) dt \tag{4}$$

This convolution integral is calculated numerically given a scaled and normalized vector for $\vec{f}_{xtal}(t_r)$ which discretizes $f_{xtal}(t_r)$ between t_{erupt} and t_{sat}. We then calculate the log likelihood of a given proposal for a dataset of N_z zircons as:

$$LL(t_{erupt}, t_{sat}) = \sum_{i=1}^{N_z} \log(\mathcal{L}(x_i \mid t_{sat}, t_{erupt})) \tag{5}$$

Given this log likelihood, t_{erupt} may now be estimated by Markov Chain Monte Carlo methods. We implement the standard Metropolis algorithm (Metropolis et al., 1953) with a symmetric Gaussian proposal distribution for both t_{erupt} and t_{sat}, as follows:

1. Begin with initial proposals $t_{erupt} = \min(\vec{t}_{obs})$ and $t_{sat} = \max(\vec{t}_{obs})$ where $\vec{t}_{obs}$ is the array of observed mineral ages
2. Draw one value from a continuous uniform distribution $u \sim unif(0, 1)$
3. Adjust either t_{erupt} or t_{sat} with a symmetric Gaussian proposal as follows

$$t_{erupt_{prop}} = \begin{cases} t_{erupt} + X & u \leq 0.5 \\ t_{erupt} & u > 0.5 \end{cases} \tag{6}$$

$$t_{sat_{prop}} = \begin{cases} t_{sat} & u \leq 0.5 \\ t_{sat} + X & u > 0.5 \end{cases} \tag{7}$$

   where the random variable $X \sim \mathcal{N}(0, \sigma^2_{prop})$.
4. If $t_{erupt_{prop}} > t_{sat_{prop}}$, reverse the two proposals
5. Calculate the log likelihood of the new proposal

$$LL_{prop} = LL(t_{sat_{prop}}, t_{erupt_{prop}}) \tag{8}$$

6. Accept the proposal with probability $P_{accept} = \min(e^{LL_{prop} - LL_{last}}, 1)$, where LL_{last} is the log likelihood of the last accepted proposal. In the present implementation, any number representable as a 64-bit floating point number is permitted as potential value for t_{sat} and t_{erupt}, providing an exceptionally weak prior which reduces to a constant and thus is eliminated from the acceptance probability function. This prior might reasonably be tightened to e.g., $unif(0, 4.567\text{Ga})$, though the more valuable prior information is contained in p_{xtal}.
7. Repeat steps 2-6 at least 10^4 times, recording a running list of all accepted proposals

In this way our Markov chain explores a likelihood space such as that shown in Supplementary Figure 2. If initial proposals for t_{sat} and t_{erupt} are far from the true value, we may observe an initial period of optimization known as *burn-in*, characterized by systematic variation in t_{sat} and t_{erupt} accompanied by increasing log likelihood. However, initial proposals given by the oldest observed zircon age for t_{sat} and the youngest observed zircon age for t_{erupt} are sufficiently accurate that burn-in is often observed to be negligible (Supplementary Figure 3). After burn-in, our posterior distributions for t_{sat} and t_{erupt} are given by the *stationary distribution* of accepted proposals; for instance, our estimates for the mean and standard deviation of t_{sat} are given by the mean and standard deviation of the stationary distribution of accepted proposals of t_{sat}.

Testing and validation

In order to evaluate the efficacy of the above Bayesian parameter estimation method (with various crystallization distributions) relative to traditional weighted mean, youngest zircon, and low-N weighted mean interpretations, we conducted a range of tests with synthetic datasets drawn from the single-batch crystallization distribution. In particular, we explored synthetic datasets of between 1 and 1024 zircons with $\Delta t/\sigma$ from 0.01 to 10. Each synthetic dataset $\vec{t}_{syn}$ was drawn from the MELTS-derived crystallization distribution scaled over a crystallization timescale Δt between $t_{sat_{syn}}$ and $t_{erupt_{syn}}$, prior to the addition of analytical uncertainty σ as a Gaussian random variable.

$$\vec{t}_{syn} = \vec{t}_{xtal} + \vec{t}_{error} \tag{9}$$

where each element i of $\vec{t}_{xtal}$ and $\vec{t}_{error}$ is distributed as

$$t_{xtal_i} \sim \mathcal{D}_{xtal}(t_{sat_{syn}}, t_{erupt_{syn}})$$

$$t_{error_i} \sim \mathcal{N}(0, \sigma^2)$$

for each of the N synthetic analyses in $\vec{t}_{syn}$.

Using pseudorandom number generators to draw independent and identically distributed samples from $\mathcal{N}(0, \sigma^2)$ and $\mathcal{D}_{xtal}(t_{sat_{syn}}, t_{erupt_{syn}})$, we are able to generate independent synthetic datasets at every N and $\Delta t/\sigma$ of interest. While computationally intensive, the problem of repeatedly testing the weighted-mean, youngest-zircon and Bayesian age interpretations with independent synthetic datasets is inherently highly parallel. Consequently, we are able to use a simple and scalable code written in C and parallelized with MPI to test each interpretation (weighted-mean, Bayesian, etc.) estimation on 1200 independent and identically distributed synthetic datasets for every combination of

$$\Delta t/\sigma \in \{0.01, 1, 2, 10\}$$

with

$$N \in \{1, 2, 3, 4, 6, 8, 11, 16, 23, 32, 45, 64, 91, 128, 181, 256, 362, 512, 724, 1024\}$$

using 320 cores of a Linux cluster at the Princeton Institute for Computational Science and Engineering.

In initial tests, we observed a tendency of the Markov chain to diverge at low N_z. This is perhaps not surprising in the absence of any other imposed prior constraints: to give a concrete example, a single detrital zircon age provides virtually no constraint on the depositional age of a given stratum; the two may differ by hundreds of Myr. The same is not generally true, however, for volcanic zircons in an ash bed. To avoid this problem, we introduce a more informative Bayesian prior on (t_{sat}, t_{erupt}) to slightly favor proposals close to the weighted mean for underdispersed low-N datasets and proposals close to the youngest and oldest observed zircon for overdispersed low-N datasets, adjusting equation 8 as follows:

$$LL_{prop} = LL(t_{sat_{prop}}, t_{erupt_{prop}}) + \frac{(Z_r * A_{wm} + (1 - Z_r) * A_{obs})}{log(1 + N_z)} \quad (10)$$

given

$$A_{wm} = 2\log\left(\frac{|t_{min_{prop}} - \mu_w| + \sigma_w}{\sigma_w} * \frac{|t_{max_{prop}} - \mu_w| + \sigma_w}{\sigma_w}\right) \quad (11)$$

$$A_{obs} = 2\log\left(\frac{|t_{min_{prop}} - t_{yz}| + \sigma_{yz}}{\sigma_{yz}} * \frac{|t_{max_{prop}} - t_{oz}| + \sigma_{oz}}{\sigma_{oz}}\right) \quad (12)$$

where μ_w and σ_w are the value and uncertainty of the weighted mean of the observed (or synthetic) dataset t_{yz} and σ_{yz} are the age and analytical uncertainty of the youngest observed (or synthetic) zircon, t_{oz} and σ_{oz} are the age and analytical uncertainty of the oldest observed (or synthetic) zircon, and Z_r, after Wendt and Carl (1991):

$$Z_r = \exp((N/2 - 1) * \log(\text{MSWD}) - N_z/2 * (\text{MSWD} - 1)) \quad (13)$$

which ranges from 0 to 1, is the relative likelihood of the MSWD of the observed dataset occurring by chance (relative to MSWD = 1) for dataset of N_z observations.

Results of these synthetic dataset tests are shown in Figure 2 and tabulated in the .log files in the `synthetic dataset test directory`. The performance of each interpretational approach is quantified in terms of (1) the mean absolute deviation of the model result from the true answer, in units of analytical uncertainty σ Ma and (2) the mean absolute error of a given interpretation divided by the mean absolute error expected based the reported uncertainty of that interpretation. These units are further explained visually in Supplementary Figure 8.

As with all other computational source code, the resulting program is freely available at `https://github.com/brenhinkeller/BayeZirChron.c`, along with ASCII files containing the vector $\vec{f}_{xtal}(t_r)$ used to draw from $\mathcal{D}_{xtal}(t_{sat_{syn}}, t_{erupt_{syn}})$ and plotted in Figure 1 (`VolcanicZirconDistribution.tsv`) and all other distributions used in the Bayesian eruption age estimation approach.

Empirical crystallization distributions

Notably, the MELTS zircon crystallization distribution is fully accurate only for a single magma batch undergoing monotonic cooling with roughly constant cooling rate; this is not the general case. For highly-dispersed datasets where we cannot assume such magma conditions, we have tested a hierarchical approach in which the form of the relative crystallization distribution $f(t_r)$ is estimated from the data, leading to what may be considered a type of Empirical Bayes approach: first estimate $\vec{f}_{xtal}(t_r)$, then (as usual) use that $\vec{f}_{xtal}(t_r)$ to construct $p_{xtal}(t \mid t_{sat}, t_{erupt})$ and estimate the distribution of t_{sat} and t_{erupt}. In other words, each element of $\vec{f}_{xtal}(t_r)$ is analagous to a hyperparameter which influences the distribution of the parameters t_{sat} and t_{erupt}.

Such an approach, if incautiously applied, may carry with it a significant risk of error. Consequently, it is critical that the performance of our implementation of this hierarchical approach is thoroughly evaluated, particularly in comparison to less informative alternatives such as assuming a uniform crystallization distribution. In order to subject this approach to the same synthetic dataset tests used for the other five interpretation approaches, we have reimplemented our parallel synthetic dataset generation and Bayesian eruption age estimation codes in *Julia*, which allows for scalable parallel calculations in a higher-level programming environment.

In this approach, our key point of prior knowledge is still only that eruption should cause an abrupt cutoff in the crystallization distribution. Consequently, our implementation must reliably produce an estimate of $\vec{f}_{xtal}(t_r)$ that reproduces any broad fluctuations in relative crystallization rate while maintaining an abrupt cutoff at $t_r = 0$. We accomplish this through a truncated kernel density estimate of the scaled crystallization times $\vec{t}_{r_{obs}}$ where

$$\vec{t}_{r_{obs}} = \frac{\vec{t}_{obs} - \min(\vec{t}_{obs})}{\max(\vec{t}_{obs}) - \min(\vec{t}_{obs})} \quad (14)$$

To produce a kernel density estimate of $\vec{f}_{xtal}(t_r)$ from $\vec{t}_{r_{obs}}$, we use the `KernelDensity.jl` package with a Gaussian kernel and bandwidth determined by Silverman's rule. The resulting kernel density estimate is truncated at $t_r = -0.05$. If fewer than

5 analyses are available for a given sample, we default to the $N = 1$ case, which yields a truncated Normal distribution due to the choice of a Gaussian kernel.

As shown in `TestBootstrappedAccuracyParallel.jl`, we again draw synthetic datasets from the MELTS volcanic zircon distribution, for the same range of N and $\Delta t/\sigma$ as above, in parallel on 320 cores of a Linux cluster. For each independent synthetic dataset, $\vec{f}_{xtal}(t_r)$ is then estimated by KDE as described above, and the Bayesian eruption age code run using this relative crystallization distribution. The results, included in Figure 1, allow us to compare the accuracy of this "bootstrapped" estimate of t_{erupt} both (1) in absolute terms, (2) relative to traditional zircon age interpretations, and (3) relative to equivalent Bayesian estimates using either (a) the MELTS prior from which the synthetic data were actually drawn, or (b) assuming a uniform crystallization distribution. As seen in Figure 1, the "bootstrapped" crystallization distribution does not fall to overfitting within the explored parameter space, and significantly outperforms the assumption of a uniform crystallization distribution at high $\Delta t/\sigma$.

REFERENCES

Barboni, M., Annen, C., and Schoene, B., 2015, Evaluating the construction and evolution of upper crustal magma reservoirs with coupled U/Pb zircon geochronology and thermal modeling: A case study from the Mt. Capanne pluton (Elba, Italy): Earth and Planetary Science Letters, v. 432, p. 436–448.

Boehnke, P., Watson, E.B., Trail, D., Harrison, T.M., and Schmitt, A.K., 2013, Zircon saturation re-revisited: Chemical Geology, v. 351, p. 324–334.

Bowring, S.A., Grotzinger, J.P., Isachsen, C., Knoll, A., Pelechaty, S., and Kolosov, P., 1993, Calibrating rates of early Cambrian evolution: Science, v. 261, p. 1293–1298.

Caricchi, L., Simpson, G., and Schaltegger, U., 2014, Zircons reveal magma fluxes in the Earth's crust: Nature, v. 511, p. 457–461.

Caricchi, L., Simpson, G., and Schaltegger, U., 2016, Estimates of volume and magma input in crustal magmatic systems from zircon geochronology: the effect of modelling assumptions and system variables: Frontiers in Earth Science, v. 4, p. 1–15.

Cherniak, D.J., 2003, Diffusion in Zircon: Reviews in Mineralogy and Geochemistry, v. 53, p. 113–143.

Crowley, J.L., Schoene, B., and Bowring, S.A., 2007, U-Pb dating of zircon in the Bishop Tuff at the millennial scale: Geology, v. 35, p. 1123–1126.

Ellis, B.S., Mark, D.F., Troch, J., Bachmann, O., Guillong, M., Kent, A.J.R., and von Quadt, A., 2017, Split-grain 40 Ar/ 39 Ar dating: Integrating temporal and geochemical data from crystal cargoes: Chemical Geology, v. 457, p. 15–23.

Galbraith, R., 2005, *Statistics for Fission Track Analysis*: Chapman and Hall / CRC Press.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B., 2013, Bayesian Data Analysis: CRC Press.

Harrison, T.M., 2009, The Hadean Crust: Evidence from 4 Ga Zircons: Annual Review of Earth and Planetary Sciences, v. 37, p. 479–505.

Jasra, A., Stephens, D.A., Gallagher, K., and Holmes, C.C., 2006, Bayesian Mixture Modelling in Geochronology via Markov Chain Monte Carlo: Mathematical Geology, v. 38, p. 269–300.

Keller, C.B., Boehnke, P., and Schoene, B., 2017, Temporal variation in relative zircon abundance throughout Earth history: Geochemical Perspectives Letters, v. 3, p. 179–189.

Lissenberg, C.J., Rioux, M., Shimizu, N., Bowring, S.A., and Mével, C., 2009, Zircon Dating of Oceanic Crustal Accretion: Science, v. 323, p. 1048–1050.

Mattinson, J.M., 2005, Zircon U–Pb chemical abrasion ("CA-TIMS") method: Combined annealing and multi-step partial dissolution analysis for improved precision and accuracy of zircon ages: Chemical Geology, v. 220, p. 47–66.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E., 1953, Equation of State Calculations by Fast Computing Machines: Journal of Chemical Physics, v. 21, p. 1087–1092.

Mundil, R., Ludwig, K.R., Metcalfe, I., and Renne, P.R., 2004, Age and Timing of the Permian Mass Extinctions: U/Pb Dating of Closed-System Zircons: Science, v. 305, p. 1760–1763.

Reiners, P.W., Carlson, R.W., Renne, P.R., Cooper, K.M., Granger, D.E., McLean, N.M., and Schoene, B., 2018, *Geochronology and Thermochronology*: John Wiley & Sons Ltd.

Samperton, K.M., Bell, E.A., Barboni, M., Keller, C.B., and Schoene, B., 2017, Zircon age-temperature-compositional spectra in plutonic rocks: Geology, v. 45, p. 983–986.

Samperton, K.M., Schoene, B., Cottle, J.M., Keller, C.B., Crowley, J.L., and Schmitz, M.D., 2015, Magma emplacement, differentiation and cooling in the middle crust: Integrated zircon geochronological–geochemical constraints from the Bergell Intrusion, Central Alps: Chemical Geology, v. 417, p. 322–340.

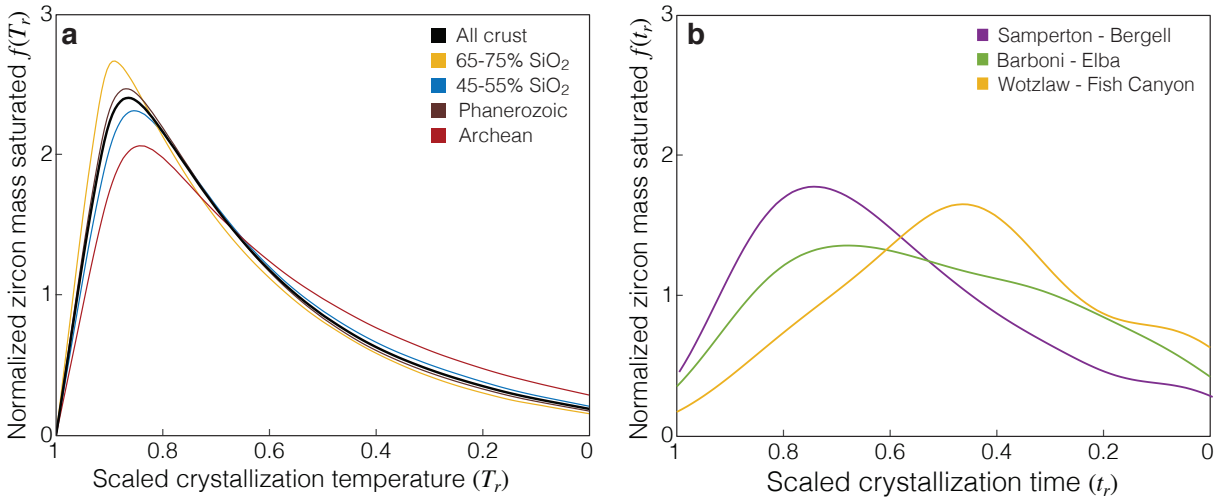
Schoene, B., 2014, 4.10 U–Th–Pb Geochronology: In R.L. Rudnick (ed.), *Treatise on Geochemistry*, p. 341–378, Elsevier.

Schoene, B., Samperton, K.M., Eddy, M.P., Keller, G., Adatte, T., Bowring, S.A., Khadri, S.F.R., and Gertsch, B., 2015, U-Pb geochronology of the Deccan Traps and relation to the end-Cretaceous mass extinction: Science, v. 347, p. 182–184.

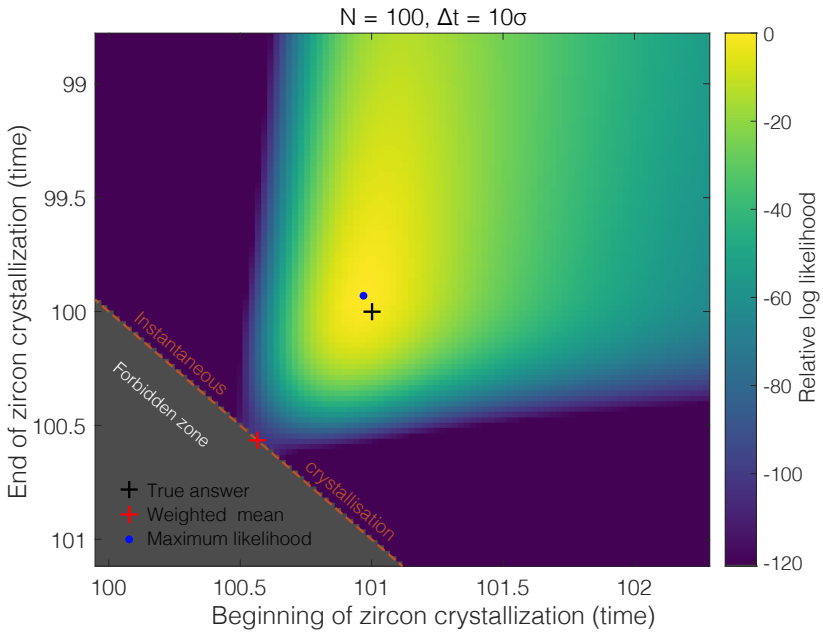
Watson, E.B., 1996, Dissolution, growth and survival of zircons during crustal fusion: kinetic principles, geological models and implications for isotopic inheritance: Transactions of the Royal Society of Edinburgh: Earth Sciences, v. 87, p. 43–56.

Wendt, I. and Carl, C., 1991, The statistical distribution of the mean squared weighted deviation: Chemical Geology, v. 86, p. 275–285.

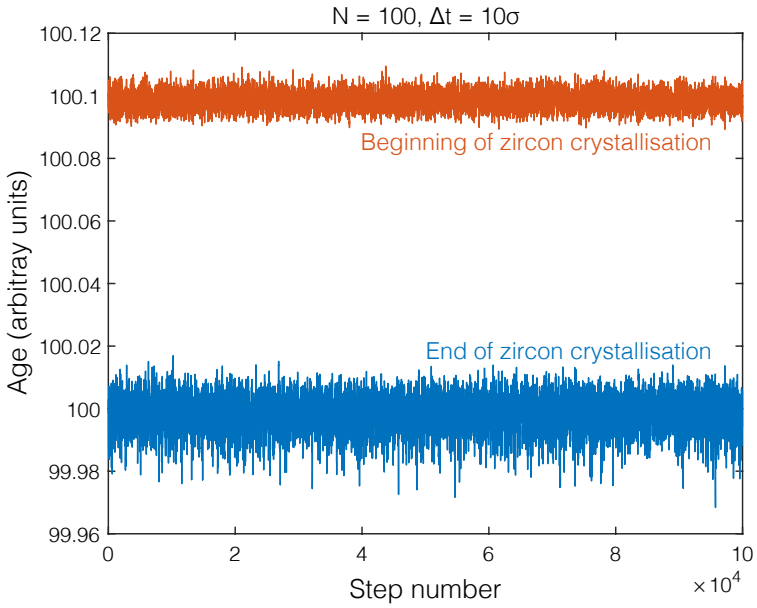
Wotzlaw, J.F., Schaltegger, U., Frick, D.A., Dungan, M.A., Gerdes, A., and Günther, D., 2013, Tracking the evolution of large-volume silicic magma reservoirs from assembly to supereruption: Geology, v. 41, p. 867–870.



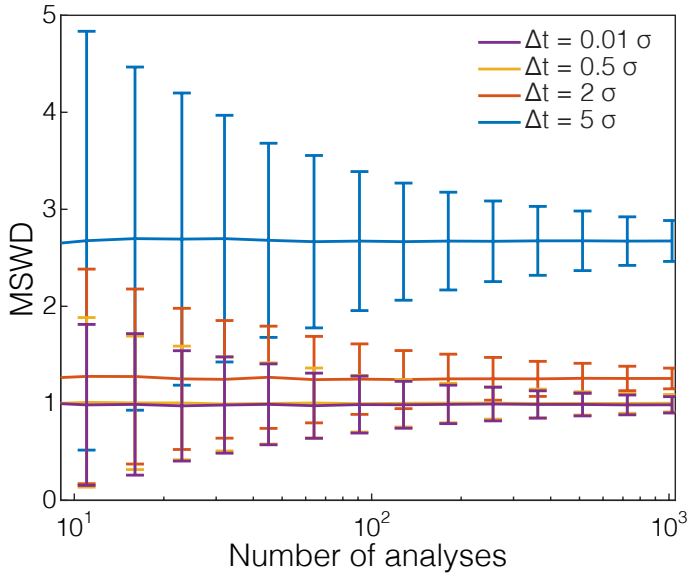
Supplementary Figure 1: Relative zircon crystallization distributions. **a**: Zircon crystallization distributions derived from MELTS major element calculations, trace Zr partitioning, and the zircon saturation model of Boehnke et al. (2013) for a wide range of whole-rock compositions. **b**: Empirical "bootstrapped" zircon crystallization distributions, kernel density estimates of published datasets from Samperton et al. (2015), Barboni et al. (2015), and Wotzlaw et al. (2013). The simple in-situ crystallization distribution of (**a**) seen in the Bergell case becomes increasingly distorted in the Elba and Fish Canyon datasets, which may be attributed to a combination of (1) potentially complicated thermal histories, (2) truncation of the long tail of plutonic crystallization by eruption (Fish Canyon) or hypabyssal porphyry intrusion (Elba), or (3) a lack of sub-grain microsampling, which has been conducted at scale only in the Bergell dataset.



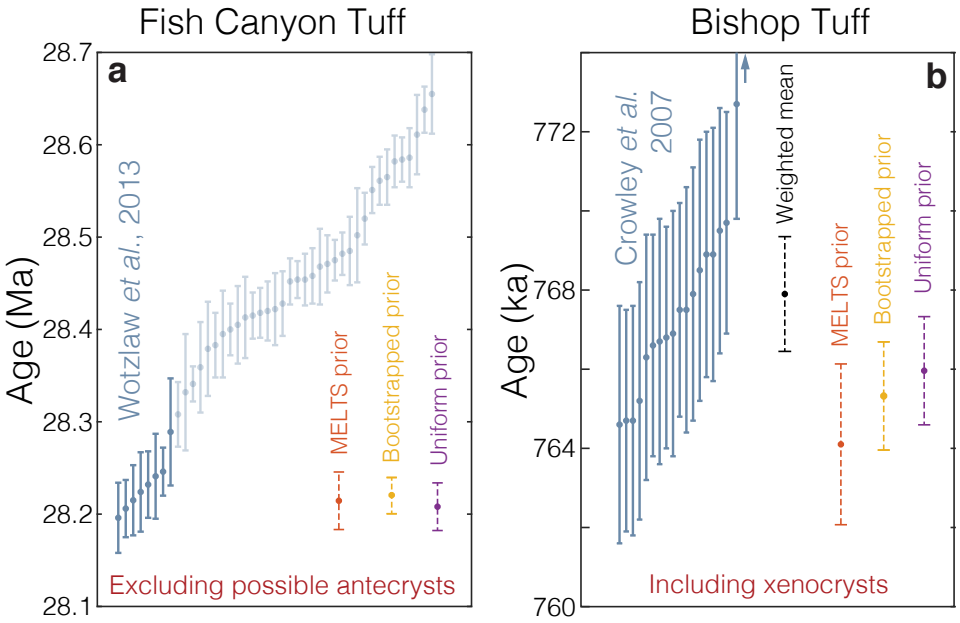
Supplementary Figure 2: Likelihood space for dataset with 100 zircons and $\Delta t = 10\sigma$. Warmer colors denote greater likelihood, with the highest likelihood observed near the true answer (denoted by black +)



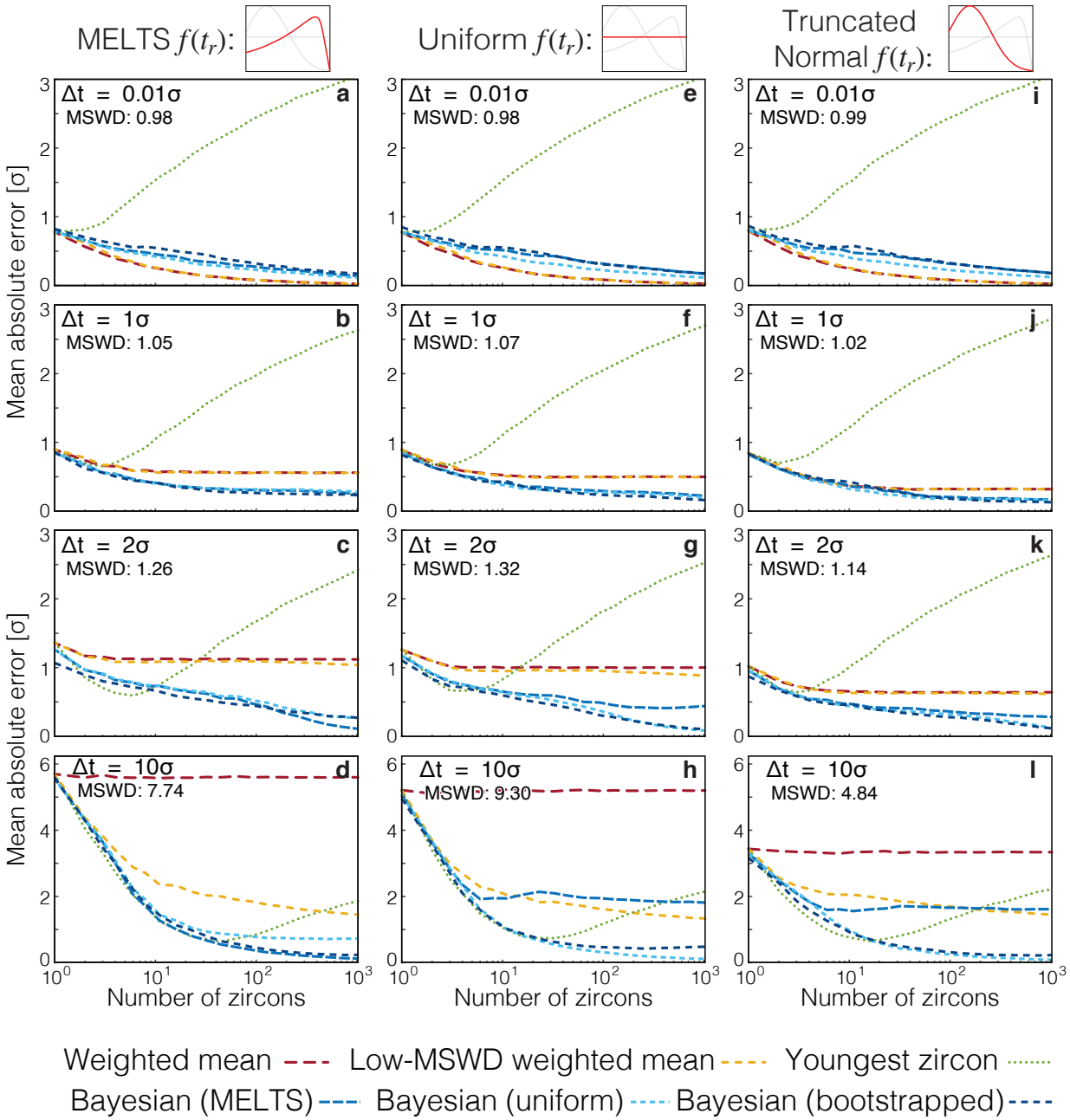
Supplementary Figure 3: Saturation and eruption ages of the first 10^5 steps of of the Markov chain for an MCMC inversion of a synthetic dataset with 100 zircons and $\Delta t = 10\sigma$. Due to the simple monotonic nature of the likelihood function for such an age inversion, and the availability of accurate initial guesses (*i.e.*, the oldest and youngest zircon), the distribution is immediately stationary. A weighted mean age must always plot on the line of instantaneous crystallization.



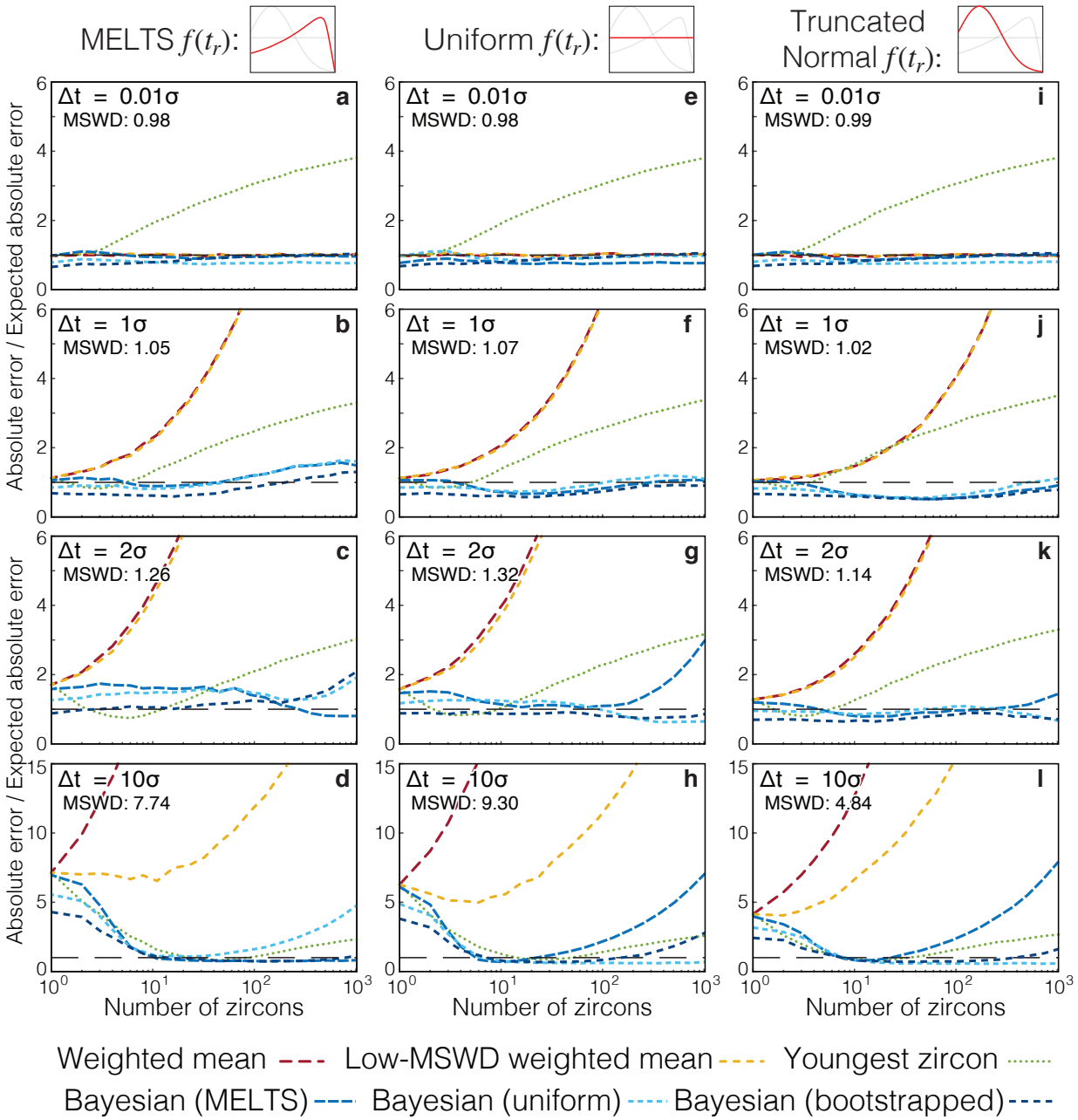
Supplementary Figure 4: The uncertainty of the MSWD for analytical datasets as a function of N for a range of $\Delta t/\sigma$. The case of $\Delta t = 5\sigma$ is not distinguishable from instantaneous crystallization with N less than ~50, and $\Delta t = 2\sigma$ is not distinguishable with N less than ~700.



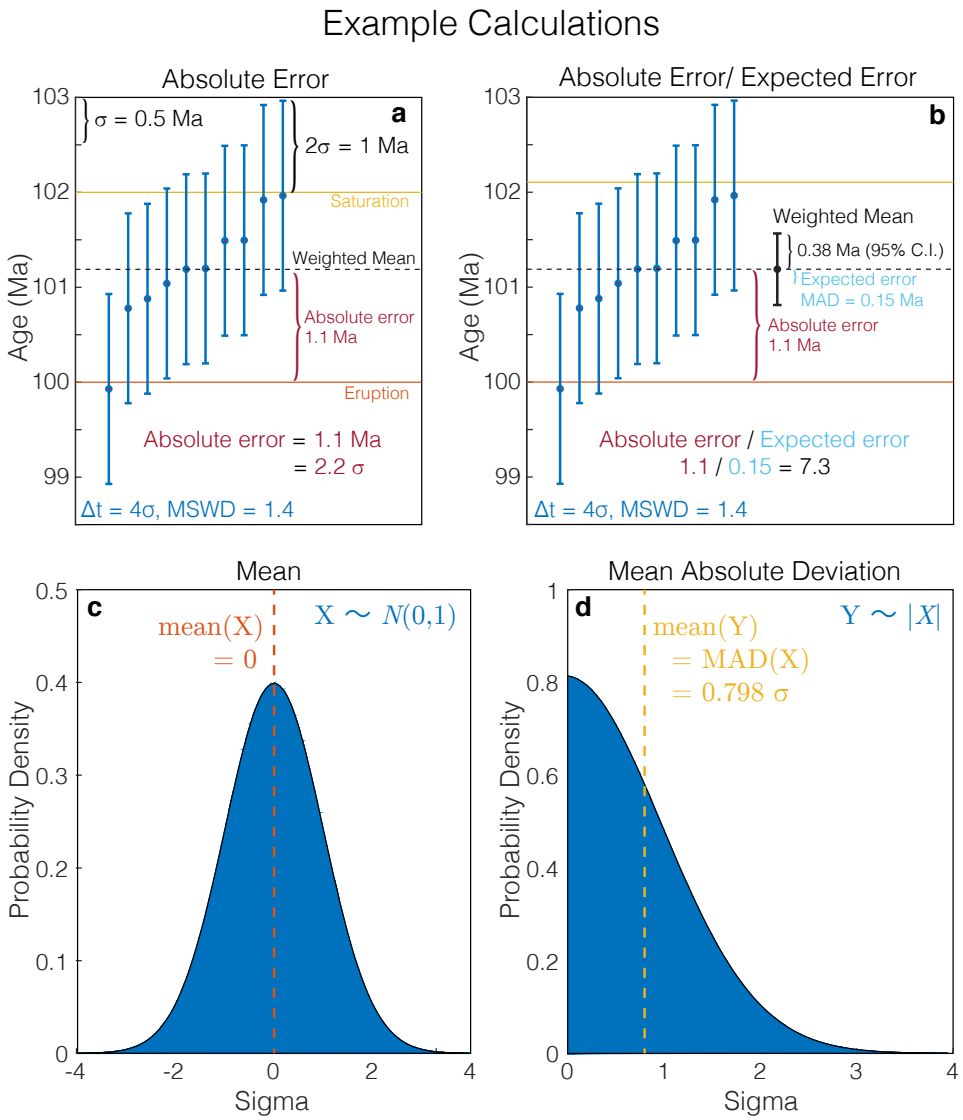
Supplementary Figure 5: Bayesian eruption age estimates for alternate Bishop Tuff and Fish Canyon Tuff datasets: **a** arbitrarily excluding all zircon ages older than 28.3 Ma in the Fish Canyon dataset, and **b** including two xenocrysts (one off-scale) in the Bishop Tuff dataset. Compared to a uniform distribution, empirical estimates and MELTS calculations provide more informative relative crystallization distributions – yielding more accurate results in an ideal system, but with increased risk of overfitting. If all zircons were strictly autocrystic, the presence of outliers would suggest that we are incompletely sampling the zircon saturation distribution, and thus overestimating the eruption age. Including xenocrystic outliers in the Bayesian age interpretation thus leads to *under*estimation of the eruption age and divergence between Bayesian and weighted mean ages for the Bishop Tuff. Consequently, methods for quantitatively identifying xenocrystic grains unrelated to in-situ crystallization of the erupted magma are of particular utility.



Supplementary Figure 6: A comparison of the absolute error of each age interpretation for synthetic data drawn from different relative crystallization distributions $f(t_r)$. **a**-**d**: MELTS crystallization distribution as in Figure 2. **e**-**h**: uniform crystallization distribution. **i**-**l**: truncated Normal crystallization distribution. Assuming a uniform crystallization distribution provides the most consistently accurate results at low $\Delta t/\sigma$, while the "bootstrapped" distribution interpretation (base on a truncated kernel density estimate for each synthetic dataset) consistently performs well at high $\Delta t/\sigma$. As in Figure 2, mean absolute error is the mean absolute deviation of the model result from the true value, reported in units of analytical uncertainty, σ; lower absolute errors are better. Each datum reflects the mean of 1200 synthetic dataset tests; standard error of the mean is on the order of the line width.



Supplementary Figure 7: A comparison of the relative error of each age interpretation for synthetic data drawn from different relative crystallization distributions $f(t_r)$. **a**-**d**: MELTS crystallization distribution as in Figure 2. **e**-**h**: uniform crystallization distribution. **i**-**l**: truncated Normal crystallization distribution. Assuming a uniform crystallization distribution provides the most consistently accurate results at low $\Delta t/\sigma$, while the "bootstrapped" distribution interpretation (based on a truncated kernel density estimate for each synthetic dataset) consistently performs well at high $\Delta t/\sigma$. As in Figure 2, "error / expected error" quantifies the accuracy of the model uncertainty for each age interpretation. A value greater than one indicates an underestimation of the model uncertainty (*i.e.*, overprecision), while a value lower than one indicates an overestimation of the model uncertainty. Each datum reflects the mean of 1200 synthetic dataset tests; standard error of the mean is on the order of the line width.



Supplementary Figure 8: Explanation of some of the terms used in Fig. 2. **a** Illustrates the calculation of absolute error (c.f. Fig. 2a-d) for an example dataset with $\Delta t = 4\sigma$ and average analytical error $\sigma = 0.5$ Ma. For a weighted mean age of 101.1 Ma and a true eruption age of 100.0 Ma, we find an absolute error of 1.1 Ma, equal to 2.2 σ. For the same example dataset, the ratio of absolute error to expected error (c.f. Fig. 2e-h) is calculated in **b**: absolute error is unchanged, while expected error is equal to the mean absolute deviation (MAD) of the resulting weighted mean. Mean absolute deviation is further explained in **c** and **d**. The familiar probability density function (PDF) of a standard normal random variable X with mean of 0 and variance 1 is illustrated in panel **c**. The distribution is symmetric about the mean. The PDF of a corresponding half-normal random variable $Y = |X|$ is shown in **d**; the mean of Y is the mean absolute deviation of X. In general, the mean absolute deviation of any Gaussian random variable is equal to 0.798 times the standard deviation.