Geochemical Perspectives Letters

A stochastic sampling approach to zircon eruption age estimation

C.B. Keller, B. Schoene, K.M. Samperton

Supplementary Information

The Supplementary Information includes:

- Supplementary Methods
- Figures S-1 through S-8
- Supplementary Information References

Supplementary Methods

Eruption age estimation

Given a dataset of mass-spectrometrically determined closed-system mineral crystallisation 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 crystallisation 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; Keller *et al.*, 2017), and empirical (Samperton *et al.*, 2015) results seen in Figure 1a provides a relative zircon crystallisation 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:

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

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 Fig. S-1. More generally, for any system where we can independently determine $f_{xtal}(t_r)$, we may then define a mineral crystallisation distribution $D_{xtal}(t_{sat}, t_{erupt})$ with a normalised probability density function $p_{xtal}(t \mid t_{sat}, t_{erupt})$ given by:

$$P_{xtal}\left(t \mid t_{sat}, t_{erupt}\right) = \begin{Bmatrix} 0 & t < t_{eruption} \\ 0 & t > t_{saturation} \\ f_{xtal}\left(t_r\right)/\Delta t & t_{erupt} \le t \le t_{sat} \end{Bmatrix} \tag{3}$$

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 crystallisation distribution, accounting for analytical uncertainty. For a single zircon i of age x_i and Gaussian analytical uncertainty with variance σ^2, this likelihood L is given by an integral over all time:

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

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

$$\mathrm{LL}\left(t_{erupt}, t_{sat}\right) = \sum_{i=1}^{N_z} \log\left(L\left(x_i | t_{sat}, t_{erupt}\right)\right) \quad (5)$$

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 $t_{erupt} = \min(\mathbf{t}_{obs})$ and $t_{sat} = \max(\mathbf{t}_{obs})$ where $\mathbf{t}_{obs}$ 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

$$t_{erupt_{prop}} = \begin{cases} t_{erupt} + X & u \le 0.5 \\ t_{erupt} & u > 0.5 \end{cases} \quad (6)$$

$$t_{sat_{prop}} = \begin{cases} t_{sat} + X & x \le 0.5 \\ t_{sat} & u > 0.5 \end{cases} \quad (7)$$

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

$$\mathrm{LL}_{\mathrm{prop}} = \mathrm{LL}\left(t_{sat_{prop}}, t_{erupt_{prop}}\right) \quad (8)$$

1. Accept the proposal with probability $P_{accept} = \min(\exp(\mathrm{LL}_{\mathrm{prop}} - \mathrm{LL}_{last}), 1)$, where $\mathrm{LL}_{\mathrm{last}}$ 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 Ga), though the more valuable prior information is contained in p_{xtal}.
2. 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 Fig. S-2. If initial proposals for t_{sat} and t_{erupt} are far from the true value, we may observe an initial period of optimisation known as burn-in, characterised 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 (Fig. S-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 crystallisation 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 crystallisation 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 $\mathbf{t}_{\mathrm{syn}}$ was drawn from the MELTS-derived crystallisation distribution scaled over a crystallisation timescale Δt between $t_{sat\ \mathrm{syn}}$ and $t_{erupt\ \mathrm{syn}}$, prior to the addition of analytical uncertainty σ as a Gaussian random variable.

$$\mathbf{t}_{syn} = \mathbf{t}_{xtal} + \mathbf{t}_{error} \quad (9)$$

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

$$t_{xtal_i} \sim D_{xtal}\left(t_{sat_{syn}}, t_{erupt_{syn}}\right)$$

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

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

Using a pseudorandom number generators to draw independent and identically distributed samples from $\mathcal{N}(0,\sigma 2)$ and $D_{xtal}(t_{sat\ \mathrm{syn}}, t_{erupt\ \mathrm{syn}})$, 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:

$$\mathrm{LL}_{\mathrm{prop}} = \mathrm{LL}\left(t_{sat_{prop}}, t_{erupt_{prop}}\right) + \frac{Z_r * A_{wmean} + \left(1 - Z_r\right) * A_{obs}}{\log\left(1 + N_z\right)} \tag{10}$$

given

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

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

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):

$$Z_r = \exp\left(\left(N/2 - 1\right) * \log\left(\mathrm{MSWD}\right) - N_z/2 * \left(\mathrm{MSWD} - 1\right)\right) \tag{13}$$

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 Nz 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 Fig. S-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 $\mathbf{f}(t_r)$ used to draw from $\boldsymbol{D}_{xtal}(t_{sat\ syn}, t_{sat\ syn})$ and plotted in Figure 1 (VolcanicZirconDistribution.tsv) and all other distributions used in the Bayesian eruption age estimation approach.

Empirical crystallisation distributions

Notably, the MELTS zircon crystallisation 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 crystallisation distribution $f(t_r)$ is estimated from the data, leading to what may be considered a type of Empirical Bayes approach: first estimate $\mathbf{f}_{xtal}(t_r)$, then (as usual) use that $\mathbf{f}_{xtal}(t_r)$ 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 the array $\mathbf{f}_{xtal}(t_r)$ 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 crystallisation 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 that eruption should cause an abrupt cutoff in the crystallisation distribution. Consequently, our implementation must reliably produce an estimate of $\mathbf{f}_{xtal}(t_r)$ that reproduces any broad fluctuations in xtal r relative crystallisation rate while maintaining an abrupt cutoff at $t_r = 0$. We accomplish this through a truncated kernel density estimate of the scaled crystallisation times $\mathbf{t}_{r\ obs}$ where

$$t_{r_{obs}} = \frac{t_{obs} - \min\left(t_{obs}\right)}{\max\left(t_{obs}\right) - \min\left(t_{obs}\right)} \tag{14}$$

To produce a kernel density estimate of $\mathbf{f}_{xtal}(t_r)$ from $\mathbf{t}_{r\ obs}$, 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, $\mathbf{f}_{xtal}(t_r)$ is then estimated by KDE as described above, and the Bayesian eruption age code run using this relative crystallisation distribution. The results, included in Figure 2, 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 crystallisation distribution. As seen in Figure 1, the "bootstrapped" crystallisation distribution does not fall to overfitting within the explored parameter space, and significantly outperforms the assumption of a uniform crystallisation distribution at high $\Delta t/\sigma$.

Supplementary Figures

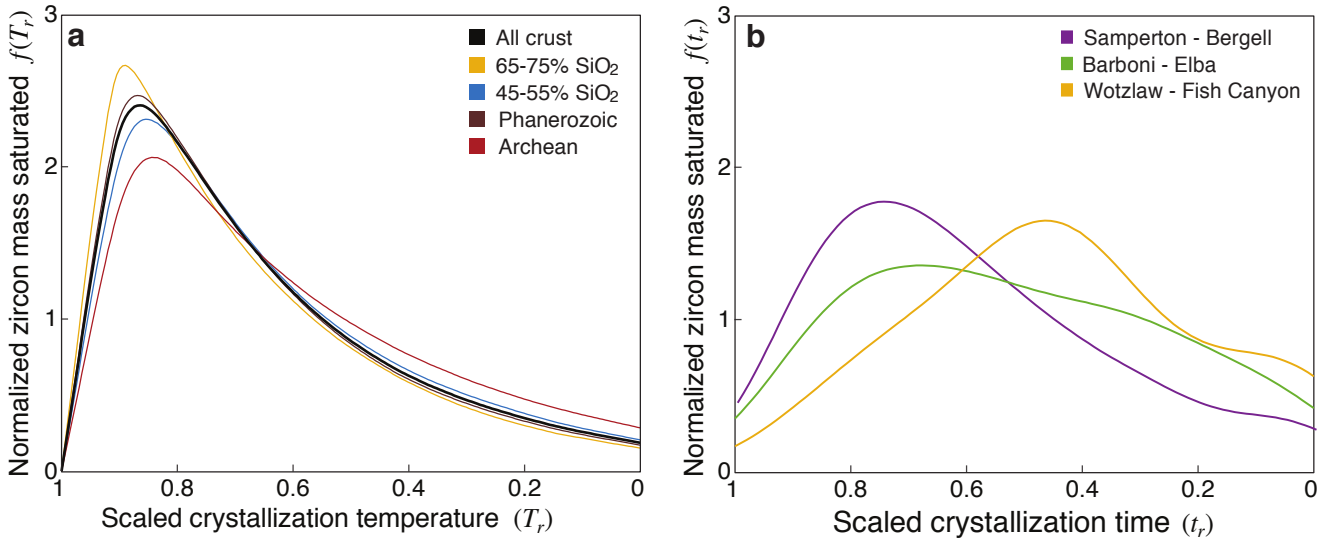

Figure S-1 Relative zircon crystallisation distributions. **(a)** Zircon crystallisation distributions derived from MELTS major element calculations, trace Zr partitioning, and the zircon saturation model of Boehnke *et al.* (2013) for a wide range of whole-rock compositions. **(b)** Empirical "bootstrapped" zircon crystallisation 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* crystallisation 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 crystallisation 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.

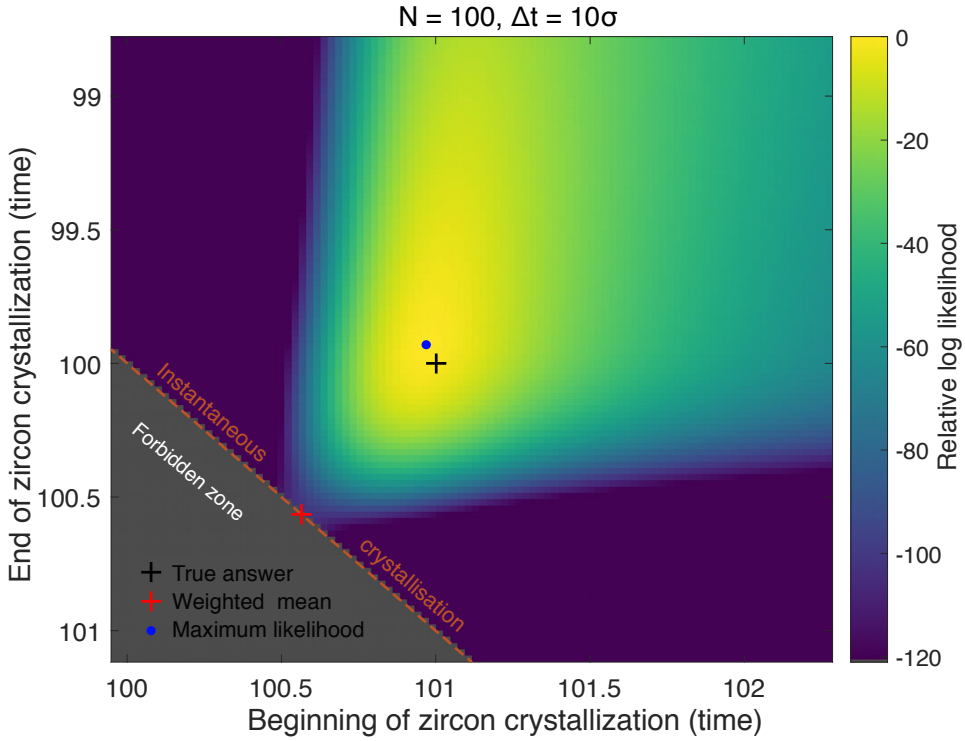

Figure S-2 Likelihood space for dataset with 100 zircons and Δt = 10σ. Warmer colors denote greater likelihood, with the highest likelihood observed near the true answer (denoted by black +).

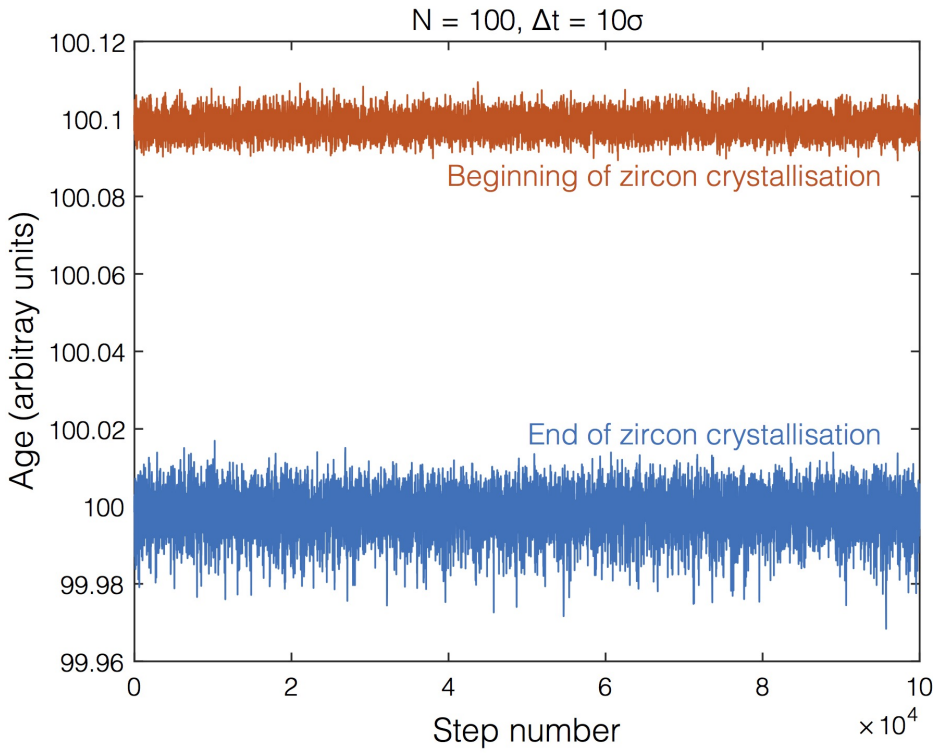

Figure S-3 Saturation and eruption ages of the first 10^5 steps of of the Markov chain for an MCMC inversion of a synthetic dataset with 100 zircons and Δt = 10σ. Due to the simple monotonic nature of the likelihood function for such an age inversion, and the availability of accurate initial guesses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

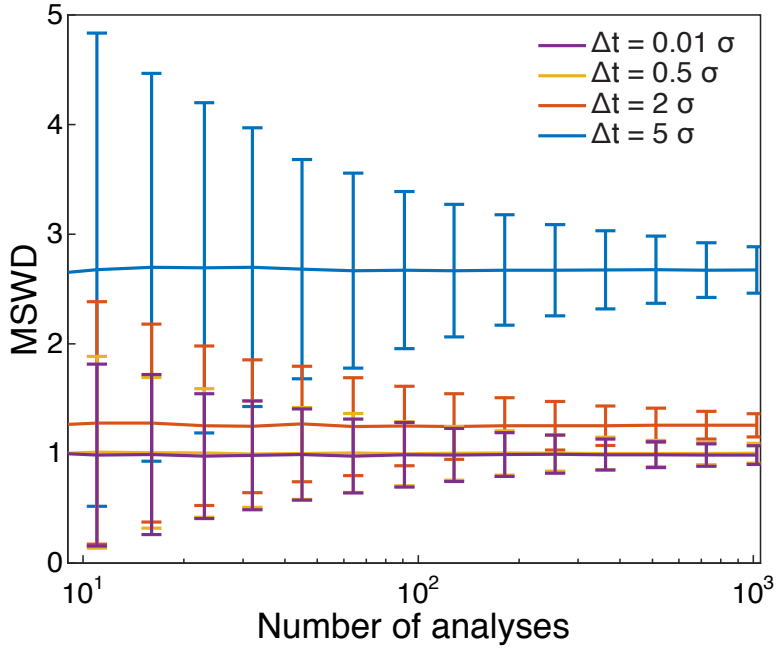

Figure S-4 The uncertainty of the MSWD for analytical datasets as a function of N for a range of Δt/σ. The case of Δt = 5σ is not distinguishable from instantaneous crystallisation with N less than ~50, and Δt = 2σ is not distinguishable with N less than ~700.

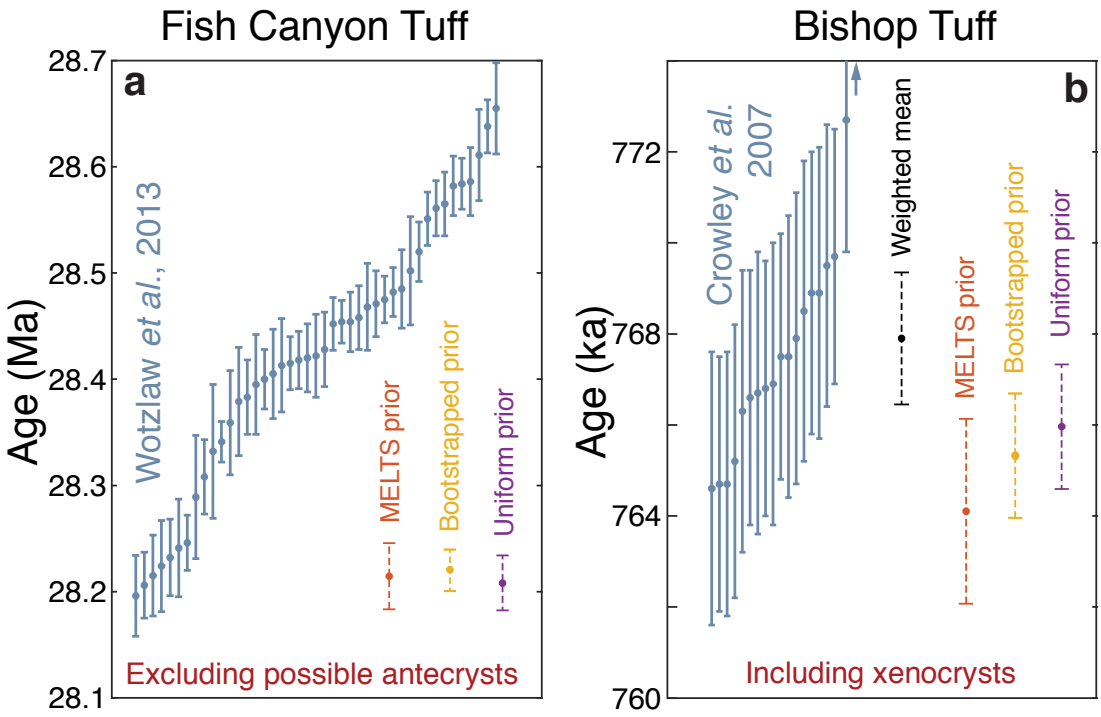

Figure S-5 Bayesian eruption age estimates for alternate Bishop Tuff and Fish Canyon Tuff datasets: **(a)** arbitrarily excluding all zircon ages older than 28.3 Ma in the Fish Canyon dataset, and **(b)** including two xenocrysts (one off-scale) in the Bishop Tuff dataset. Compared to a uniform distribution, empirical estimates and MELTS calculations provide more informative relative crystallisation 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 underestimation 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* crystallisation of the erupted magma are of particular utility.

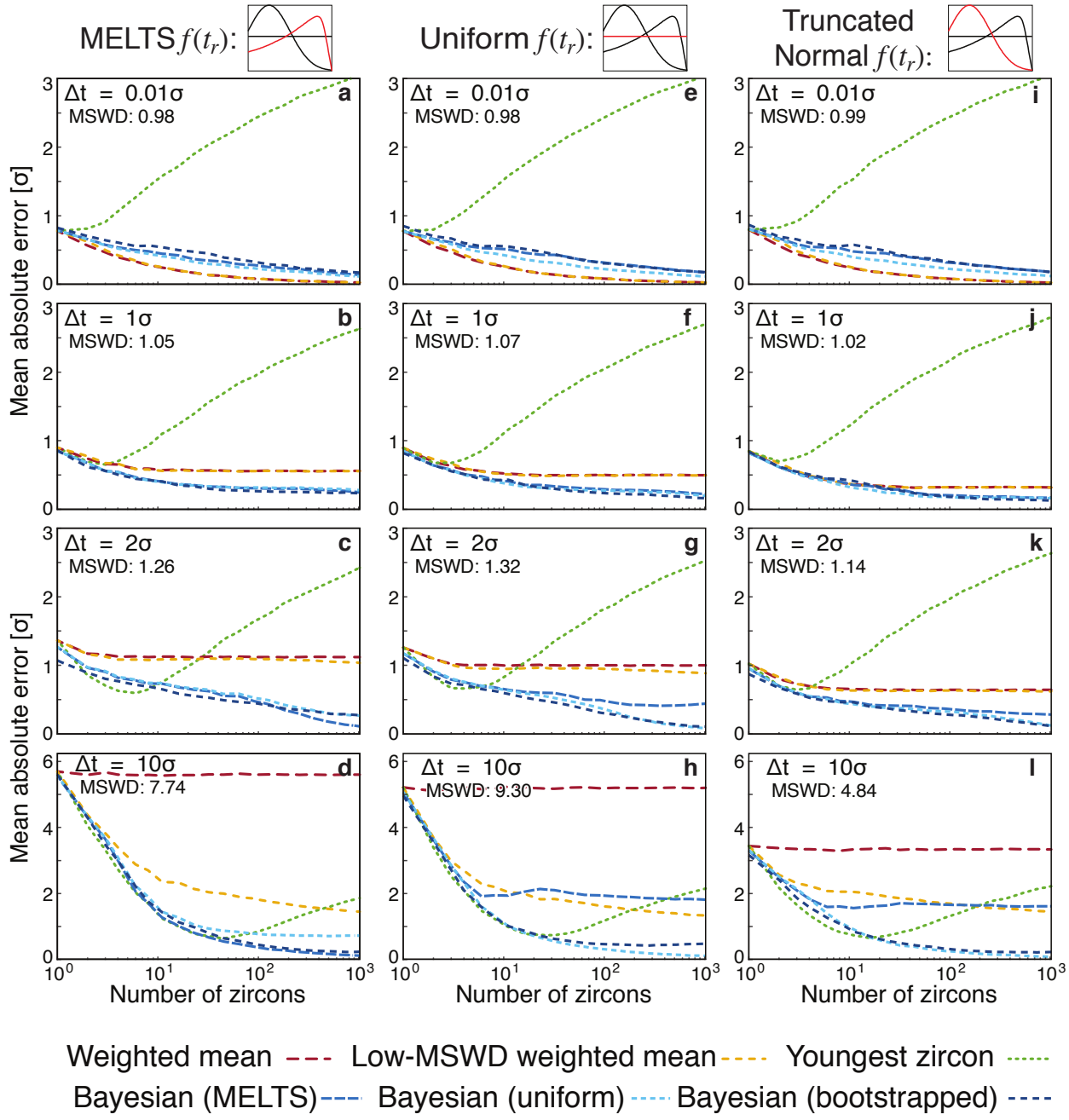

Figure S-6 A comparison of the absolute error of each age interpretation for synthetic data drawn from different relative crystallisation distributions $f(t_r)$. **(a-d)** MELTS crystallisation distribution as in Figure 2. **(e-h)** Uniform crystallisation distribution. **(i-l)** Truncated Normal crystallisation distribution. Assuming a uniform crystallisation distribution provides the most consistently accurate results at low Δt/σ, while the "bootstrapped" distribution interpretation (base on a truncated kernel density estimate for each synthetic dataset) consistently performs well at high Δt/σ. 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.

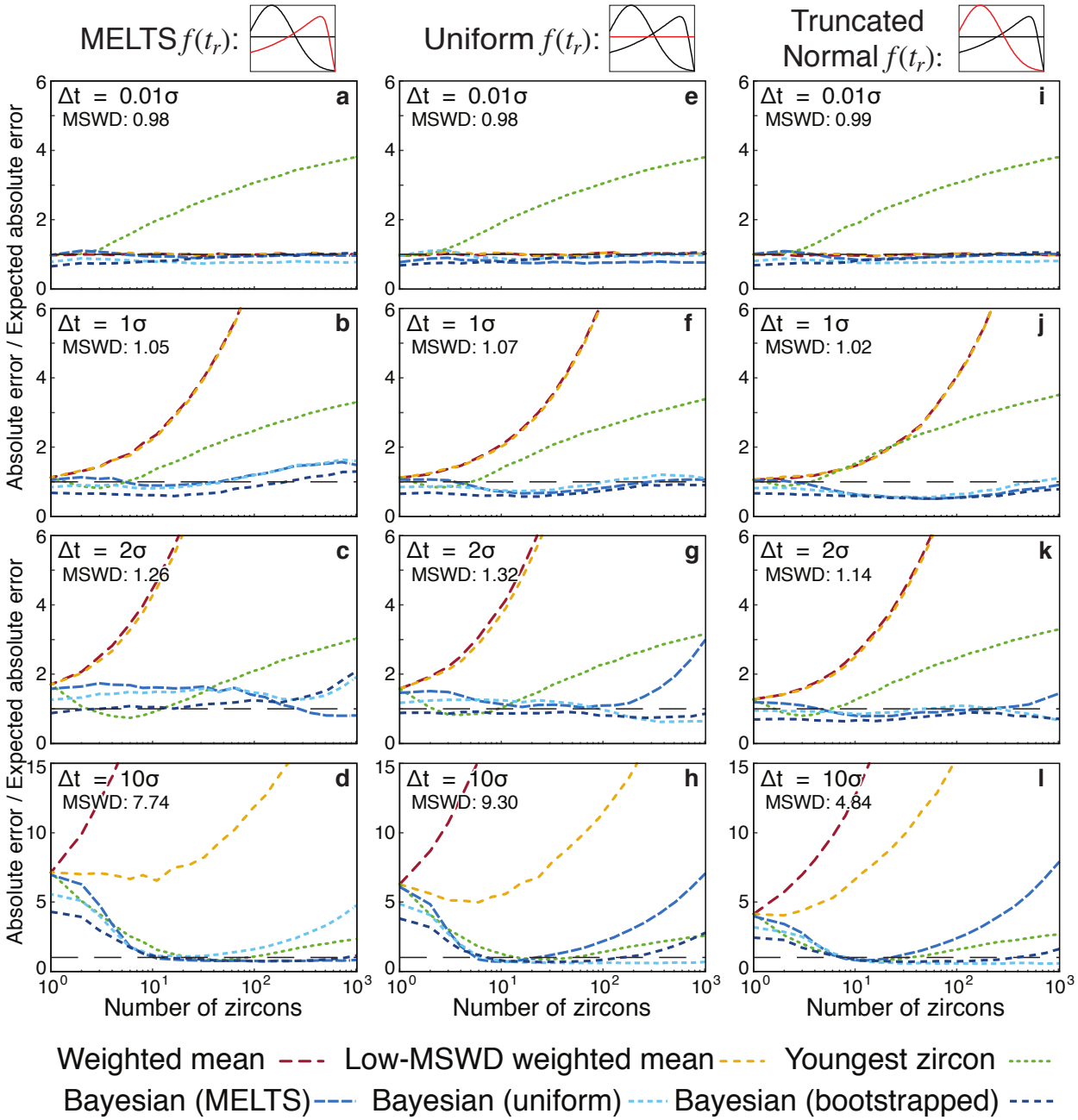

Figure S-7 A comparison of the relative error of each age interpretation for synthetic data drawn from different relative crystallisation distributions $f(t_r)$. **(a-d)** MELTS crystallisation distribution as in Figure 2. **(e-h)** Uniform crystallisation distribution. **(i-l)** Truncated Normal crystallisation distribution. Assuming a uniform crystallisation distribution provides the most consistently accurate results at low Δt/σ, while the "bootstrapped" distribution interpretation (based on a truncated kernel density estimate for each synthetic dataset) consistently performs well at high Δt/σ. 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.

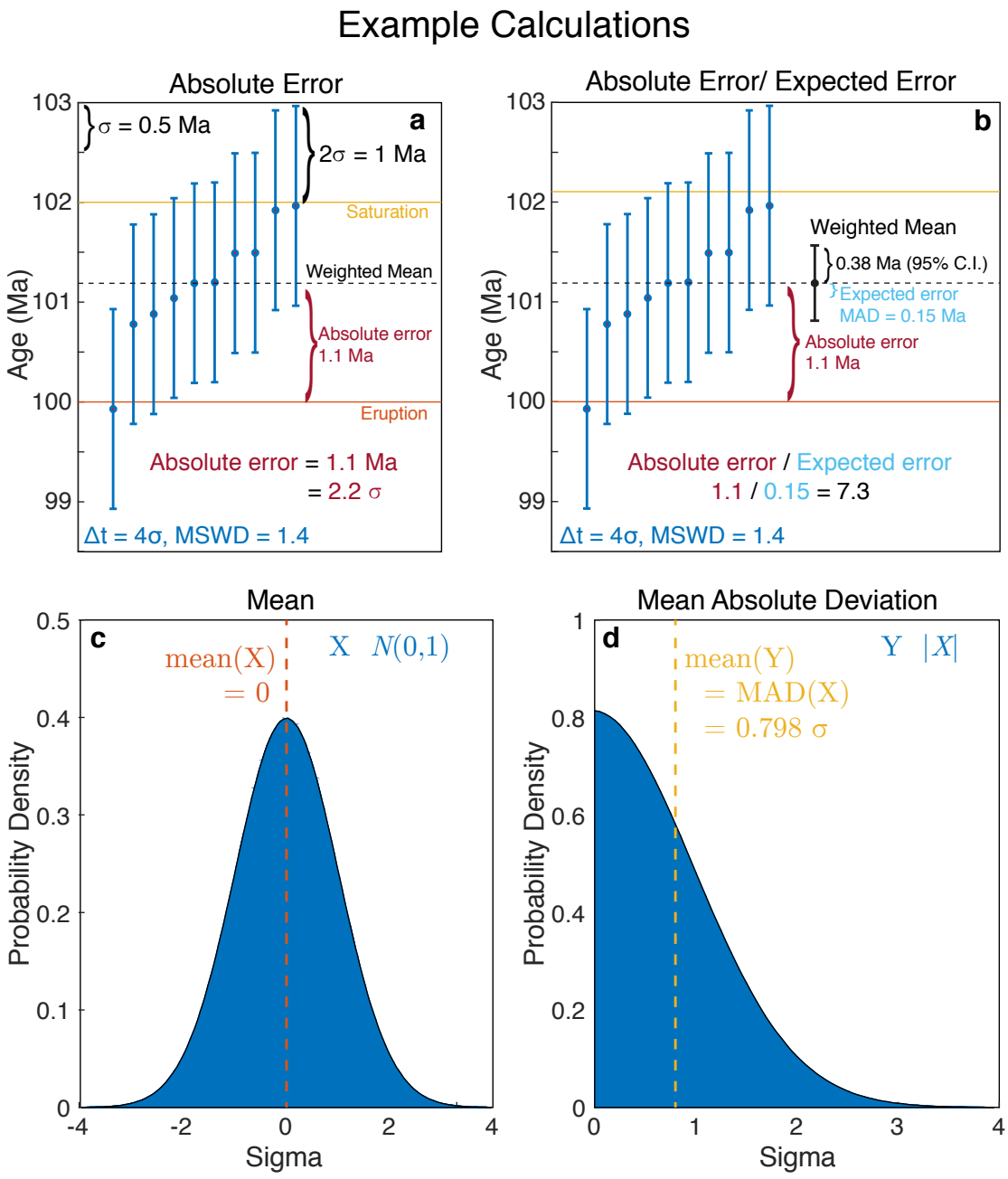

Figure S-8 Explanation of some of the terms used in Fig. 2. Panel **(a)** Illustrates the calculation of absolute error (*c.f.* Fig. 2a-d) for an example dataset with Δt = 4σ and average analytical error σ = 0.5 Ma. For a weighted mean age of 101.1 Ma and a true eruption age of 100.0 Ma, we find an absolute 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.

Supplementary Information References

Barboni, M., Annen, C., 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* 432, 436–448.

Keller, C.B., Boehnke, P., Schoene, B. (2017) Temporal variation in relative zircon abundance throughout Earth history. *Geochemical Perspectives Letters* 3, 179–189.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E. (1953) Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics* 21, 1087–1092.

Samperton, K.M., Schoene, B., Cottle, J.M., Keller, C.B., Crowley, J.L., 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* 417, 322–340.

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* 87, 43–56.

Wendt, I., Carl, C. (1991) The statistical distribution of the mean squared weighted deviation. *Chemical Geology* 86, 275–285.

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