Resolving the effects of 2D versus 3D grain measurements on apatite (U-Th)/He age data and reproducibility

Emily H. G. Cooperdock[1,2,†], Richard A. Ketcham[1], Daniel F. Stockli[1]

[1]Department of Geological Sciences, University of Texas at Austin, Austin, 78712, USA
5 [2]Woods Hole Oceanographic Institution, Woods Hole, 02543, USA
[†]Now at Department of Earth Sciences, University of Southern California, Los Angeles, CA, 90089, USA

Correspondence to: Emily H. G. Cooperdock (cooperdo@usc.edu)

Abstract. (U-Th)/He thermochronometry relies on accurate and precise quantification of individual grain volume and surface area, which are used to calculate mass, alpha ejection (F_T) correction, equivalent sphere radius (ESR), and ultimately isotope
10 concentrations and age. The vast majority of studies use 2D or 3D microscope dimension measurements and an idealized grain shape to calculate these parameters, and a long-standing question is how much uncertainty these assumptions contribute to observed intra-sample age dispersion and accuracy. Here we compare the results for volume, surface area, grain mass, ESR, and F_T correction derived from 2D microscope and 3D x-ray computed tomography (CT) length and width data for >100 apatite grains. We analyzed apatite grains from two samples that exhibited a variety of crystal habits, some with inclusions.
15 We also present 83 new apatite (U-Th)/He ages to assess the influence of 2D versus 3D F_T correction on sample age precision and effective uranium (eU). The data illustrate that the 2D approach systematically overestimates grain volumes and surface areas by 20-25%, impacting the estimates for mass, eU, and ESR – important parameters with implications for interpreting age scatter and inverse modeling. F_T factors calculated from 2D and 3D measurements differ by ~2%. This variation, however, has effectively no impact on reducing intra-sample age reproducibility, even on small aliquot samples (e.g., 4 grains). We also
20 present a grain mounting procedure for x-ray CT scanning that can allow 100's of grains to be scanned in a single session, and new software capabilities for 3D F_T and F_T-based ESR calculations that are robust for relatively low-resolution CT data, that together enable efficient and cost-effective CT-based characterization.

1 Introduction

25   (U-Th)/He thermochronometry of accessory phases, such as apatite and zircon, has been widely applied to study tectonic, volcanic, and surface processes (e.g., Zeitler et al., 1987; Stockli et al., 2000; Ehlers and Farley, 2003; Reiners and Brandon, 2006). The method is based on the radiogenic accumulation of He from the alpha decay of U, Th, and Sm isotopes and the diffusive loss of He via thermal processes. In addition, He is lost due to 'long alpha stopping distances' associated with the kinetic energy of alpha decay (~5 MeV), requiring a shape-based alpha ejection correction (F_T correction) (Farley et al.,
30 1996). This correction as traditionally applied includes several simplifications and assumptions, such as an idealized grain

geometry and homogenous parent nuclide concentrations (Farley et al., 1996, Farley; 2002; Ketcham et al. 2011). It has been shown that due to uncertainties in grain geometry, stopping distances, and parent nuclide zonation/variability, this correction can contribute >50% of the total analytical uncertainty (Farley and Stockli, 2002). Similarly, low error, highly dispersed apatite (U-Th)/He ages are problematic for robust interpretation and time-temperature modeling (e.g., Fox et al., 2019). The

5   observation that the scatter of measured ages in even well-understood samples exceeds expectation based on analytical errors, combined with the knowledge that the above simplifications will not always hold, has led to the practice of reporting errors derived from the reproducibility of standards rather than propagated analytical uncertainties in He dating. While the effect and mitigation of parent nuclide zonation in apatite and zircon to improve the accuracy and precision of (U-Th)/He ages has been studied, (e.g., Farley et al., 1996; Hourigan et al. 2005; Ketcham et al. 2011, Gautheron et al., 2011; Bargnesi et al., 2016;

10   Danisik et al., 2017; McDannell et al., 2018), the effects of grain morphology and measurement on age, uncertainty, and intra-sample variability are less known, with only a few previous studies on improvements to grain measurement (Herman et al., 2007; Evans et al., 2008; Glotzbach et al., 2019).

        In practice, for the determination of a correct He age, the grain dimensions and shape must be measured to compute an F_T correction factor prior to He and U, Th, and Sm analysis, assuming either parent nuclide homogeneity or prescribing an

15   assumed or measured 1D or 2D parent nuclide zonation (Farley et al., 1996; Farley, 2002). While not directly related to the computation of He ages, these same grain dimensions are also used to calculate grain size parameters for the purpose of calculating isotopic/elemental concentrations and for age interpretation and diffusion or thermal history modeling (Shuster et al., 2006; Flowers et al., 2007; Flowers et al., 2009; Flowers, 2009; Flowers and Kelley, 2011; Gautheron et al., 2009). For example, the grain mass, which is used to calculate the grain U, Th, Sm and He concentrations, is often derived from the grain

20   volume and an assumed density. Similarly, correlation between grain size (ESR) and He aliquot age has been used for qualitative and quantitative thermal history reconstruction using He diffusivity models (Reiners and Farley, 2001; Flowers and Kelley, 2011). Thus, the ability to measure accurate and precise grain dimensions, volumes, and surface areas for mineral grains has cascading effects for the determination, reporting, and interpretation of (U-Th)/He data.

        Most commonly, F_T, volume, and surface area are calculated using 2 or 3 grain dimensions (length + width 1 ± width

25   2) measured in 2D on an optical microscope using imaging software with a micrometer-based calibration. This approach requires the assumption of an idealized grain shape that most closely matches the mineral habit, such as a hexagonal prism for apatite or tetragonal prism for zircon, while simplifying (or ignoring) the more complex grain terminations (Farley et al., 1996; Farley, 2002). Hence, it has been best practice to select euhedral mineral grains to most closely match assumed, idealized grain shapes and large grains to minimize the amplification of uncertainties related to the F_T correction. However, even in felsic

30   magmatic samples with high-quality apatite, grains are often characterized by a wide range of grain shapes, variations in grain terminations, and the potential for broken or chipped surfaces that cause deviations from the idealized hexagonal prism. Furthermore, apatite grains often do not represent symmetric or equidimensional hexagonal prisms and are characterized by varying face widths, commonly, but also possibly inconsistently, lying on their largest and flattest face on the microscope slide and thus potentially introducing systematic biases during the selection of the clearest, inclusion-free grains.

Recognizing that this optical-microscopy approach is both limiting and may be an important source for error or bias in (U-Th)/He ages and their interpretation, more sophisticated approaches have been proposed to determine grain dimensions, namely methods that do not require assuming a grain shape (Herman et al., 2007; Evans et al., 2008; Glotzbach et al., 2019). One approach presented by Glotzbach and others (2019), called "3DHe", is an openly available software that uses orthogonal 2D grain photos to model accurate 3D grain shapes. Another approach is to employ x-ray computed tomography (CT) to determine accurate grain shapes in an effort to improve precision and accuracy in F_T and (U-Th)/He age determinations (Herman et al., 2007; Evans et al., 2008; Glotzbach et al., 2019). Herman and others (2007) used 3D CT grain dimensions to calculate F_T factors and present a production-diffusion model to extract thermal histories for detrital apatite grains. Evans and others (2008) and Glotzbach and others (2019) both tested the efficacy of 2D microscope measurements against 3D CT data of zircon and apatite grain shape and size, arriving at quite different estimated discrepancies between microscope measurements and the CT data (1-24% and <1-6%, respectively).

This new study investigates the effect of 2D versus 3D grain geometry measurement techniques on grain dimension, volume, surface area, ESR, mass, F_T and the corrected age as well as effective uranium (eU) concentrations. In contrast to previous studies, which used 5-24 grains, we characterized >100 apatite grains from two granitic samples for a more statistically robust comparison and in an effort to more systematically capture variations in apatite morphologies, sizes, and screen for inclusions. We chose samples from crystalline basement that experienced fast-cooling histories in order to target the impact of grain measurement techniques and minimize the effects of cooling history and transport on the (U-Th)/He age and dispersion. The apatite grains were picked and measured by a single analyst using 2D optical techniques and then CT-scanned. Building on previous work, we present a method for relatively rapid scans of >100 grains at 4-5 μm resolution, enabling affordable and efficient 3D screening. We introduce the capabilities of an updated version of Blob3D (Ketcham 2005; freely distributed software) that allows efficient batch processing of CT-scanned grains and outputs parameters such as grain volume and 3D F_T. We further develop an approach for calculating ESR on the basis of equivalent-F_T rather than an equivalent surface-to-volume ratio as a more direct and accurate means of approximating the diffusional domain as a sphere. Finally, in contrast to previous studies, we use the results of >80 apatite (U-Th)/He ages to evaluate the reliability of the 2D measurements as well as the impact on the (U-Th)/He age and uncertainty.

1.1 Geologic background of the samples

For this study, we selected two plutonic samples from the Cretaceous Cordilleran magmatic arc in the western USA that yielded abundant, high-quality apatite and have been part of previous thermochronometric studies. Sample 97BS-CR8 is from a granodiorite in the Carson Range in the eastern Sierra Nevada along the Nevada-California border. The sample yielded an apatite fission track age of 68 ± 2 Ma ($P(X^2) = 75.4\%$, 25 grains, $N_s = 1341$) (Surpless et al., 2002). The second sample, 95BS-11.3 is from a quartz monzonite exposed in the Wassuk Range in western Nevada, exhumed during Basin and Range normal faulting. The sample has a reported apatite fission track age of 16.3 ± 1.4 Ma ($P(X^2) = 76.1\%$, 30 grains, $N_s = 158$) and apatite (U-Th)/He age of 9.9 ± 1.9 Ma (Stockli et al., 2002). These samples were chosen for their abundant apatite and relatively

simple cooling histories. Their geologic histories are relevant to the present study in that the apatite grains derive from plutonic rocks and did not experience complex metamorphic or magmatic histories, nor natural abrasion during sedimentary transport. Furthermore, both are plutonic samples that experienced rapid post-magmatic cooling or fault-related exhumation and are expected to have spent little time in the apatite He partial retention zone.

2 Methods

2.1 Grain selection and 2D measurements

Apatite grains were picked from two samples, 97BS-CR8 (n=50) and BS95-11.3 (n=62), using a Nikon SMZ-U/100 optical microscope at a total magnification of 180x. Apatite grains were selected to include the range of grain morphologies present in the sample (e.g., broken, flat, and prismatic ends). Intentionally, several grains with visible inclusions were also selected to evaluate how well these inclusions showed up in the CT scans. All apatite grains were photographed using a Nikon digital ColorView camera connected to the microscope. The short and long axes were measured manually using AnalySIS® imaging software (Figure 1 and 3). We chose to measure a single width and did not flip the apatite 90° because this is still common practice in many labs and would allow us to compare the 'simplest' 2D measurement approach with the 3D CT data. For sample BS95-11.3, grains were imaged and measured on double-sided sticky tape in preparation for the CT mount (Figure 1). However, we determined that this can cause grains to sit in upright orientations, which is fine for CT scanning, but not for 2D measurements. For sample 97BS-CR8 each apatite grain was placed on a glass slide for 2D measurements and then transferred to the sticky tape for the CT mount to remedy this issue (Figure 1).
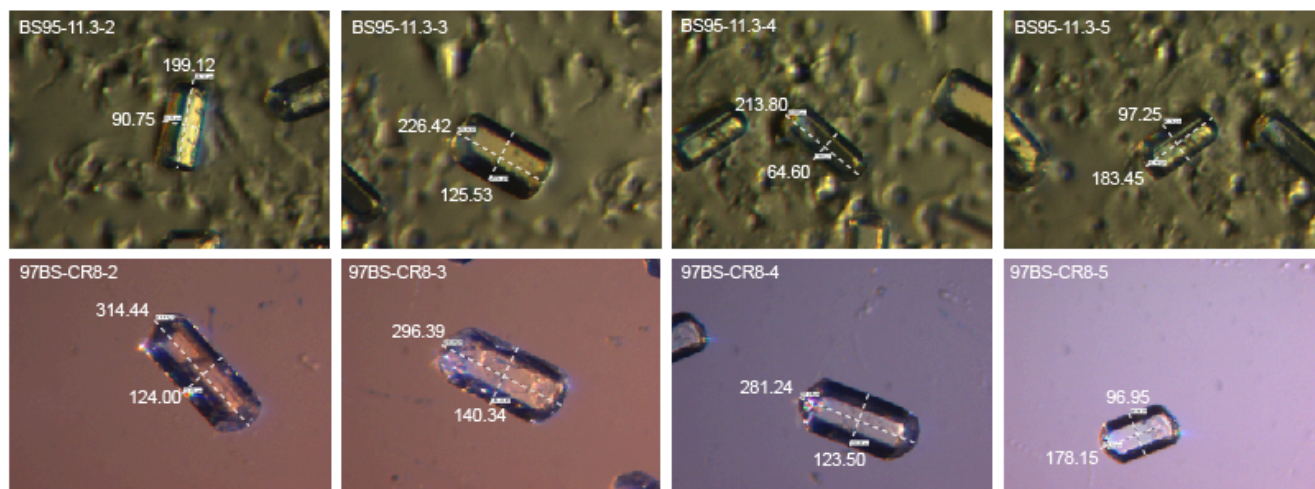
Figure 1: Apatite grain photos with 2D measurements taken on an optical microscope. Dimensions are reported in μm and the grain aliquot name is in the top left corner of each photo. The top row is photographed on double sided sticky tape, and the bottom row is photographed on a glass slide.

2.2 Grain mounting procedure for CT

Once the grains were measured optically in 2D, they were mounted for CT scanning by orienting several tens of grains on a plastic disc and stacking multiple discs (Figure 2). The procedure to create a single-layer mount for multi-grain scanning entails covering a flat top of a pushpin with double-sided sticky tape that can be pre-cut using a standard hole punch.
Apatite grains are then picked directly onto the tape in a grid-like pattern. The pushpin surface is ~5 mm in diameter, which easily allows for ≥ 50 apatite grains to be mounted in one layer, tightly spaced, without touching. Grains could be packed more densely as long as they can be reliably identified after scanning; they can even be touching, although this leads to a small increase in processing time to separate them using functions in the Blob3D software.

To utilize the total scanned volume, at least five multi-grain layers can be stacked for a single scan (up to 5 mm tall). To create stackable layers, sturdy plastic discs are made using a standard hole punch, with one side of the disc covered with double-sided sticky tape and apatite grains mounted in the procedure outlined above. Once all the layers are mounted and all excess tape is trimmed, the discs are stacked on top of the push pin. The arrangement is secured by a thin wrap of parafilm. The parafilm and sticky tape are critical to ensure the crystals and layers do not move during scanning. This mount can be easily disassembled after scanning to retrieve the grains for further analysis.
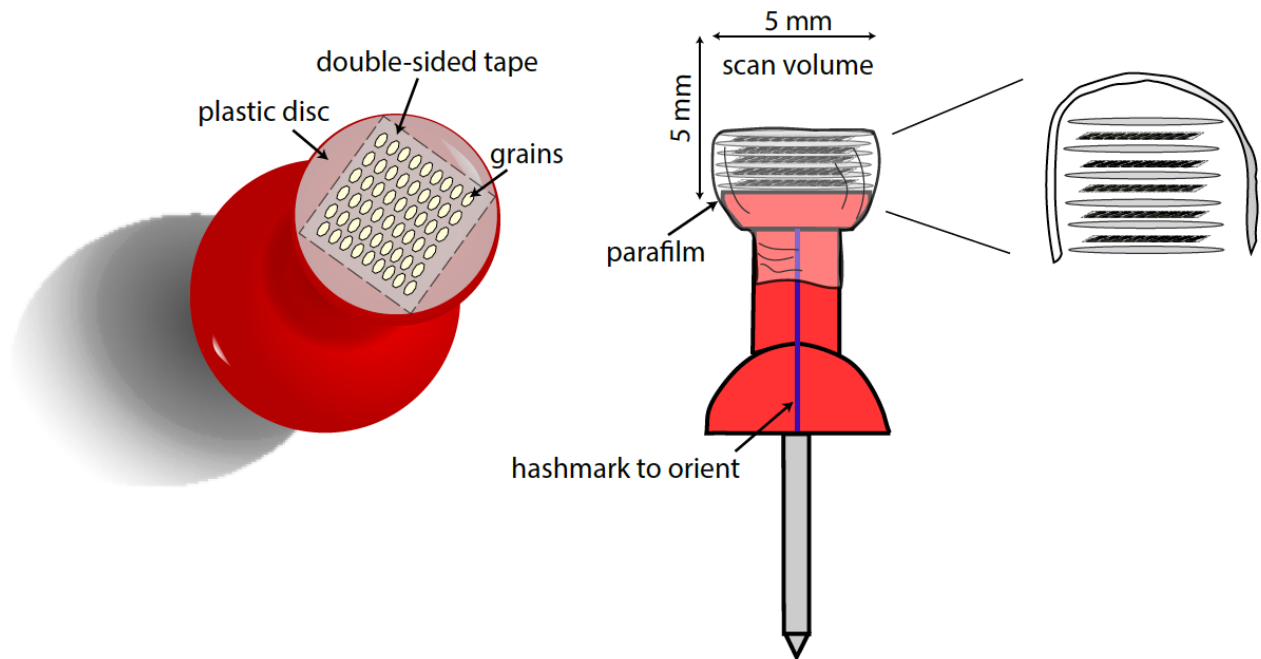
Figure 2: Schematic rendering of CT mounting procedure. Grains are adhered to the top of a plastic disc using double sided sticky tape, with multiple grains placed onto a 5x5 mm surface. Multiple plastic disk layers with grains may be assembled and then stacked to take full advantage of the height of the scan. These layers are held together using parafilm, and a hashmark on the pushpin enables further orientation of the scan in order to retrieve the grains afterwards for further analysis.

2.3 X-ray CT Scanning

The multi-grain mounts were scanned with a Zeiss Xradia MicroXCT scanner at the University of Texas High-Resolution X-ray CT Facility. Optimal scanning parameters will vary with the instrument being used, with top priorities being to minimize scanning artifacts and noise, while also minimizing time and cost. Lower X-ray energies are more sensitive to compositional variations, but more prone to beam-hardening artifacts. We experimented with various settings in this study. The grain mount for sample 97BS-CR8 was scanned with X-rays set at 100 kV and 10W, with a 1.0 mm SiO_2 filter. 1153 views were gathered at 1.5s per view, for an acquisition time of 28.9 minutes. Source-mount distance was 37.7 mm, and mount-detector distance was 12.8 mm. The 2048x2048 camera data were binned by 2, and the lower-energy X-rays and weaker filtering necessitated application of a beam hardening correction during reconstruction. The reconstructed data had a voxel (3D pixel) size of 5.03 μm.

The grain mount for sample BS95-11.3 was scanned with X-rays set at 150 kV and 10 W with a 1.6 mm CaF_2 beam filter, acquiring 571 views at 1.5s per view, for an acquisition time of 14.3 minutes, not including calibration. Source-mount distance was 37.7 mm, and mount-detector distance was 17.8 mm. The camera data were binned by 2, and no beam hardening correction was applied during reconstruction. The resulting data had a voxel size of 4.58 μm.
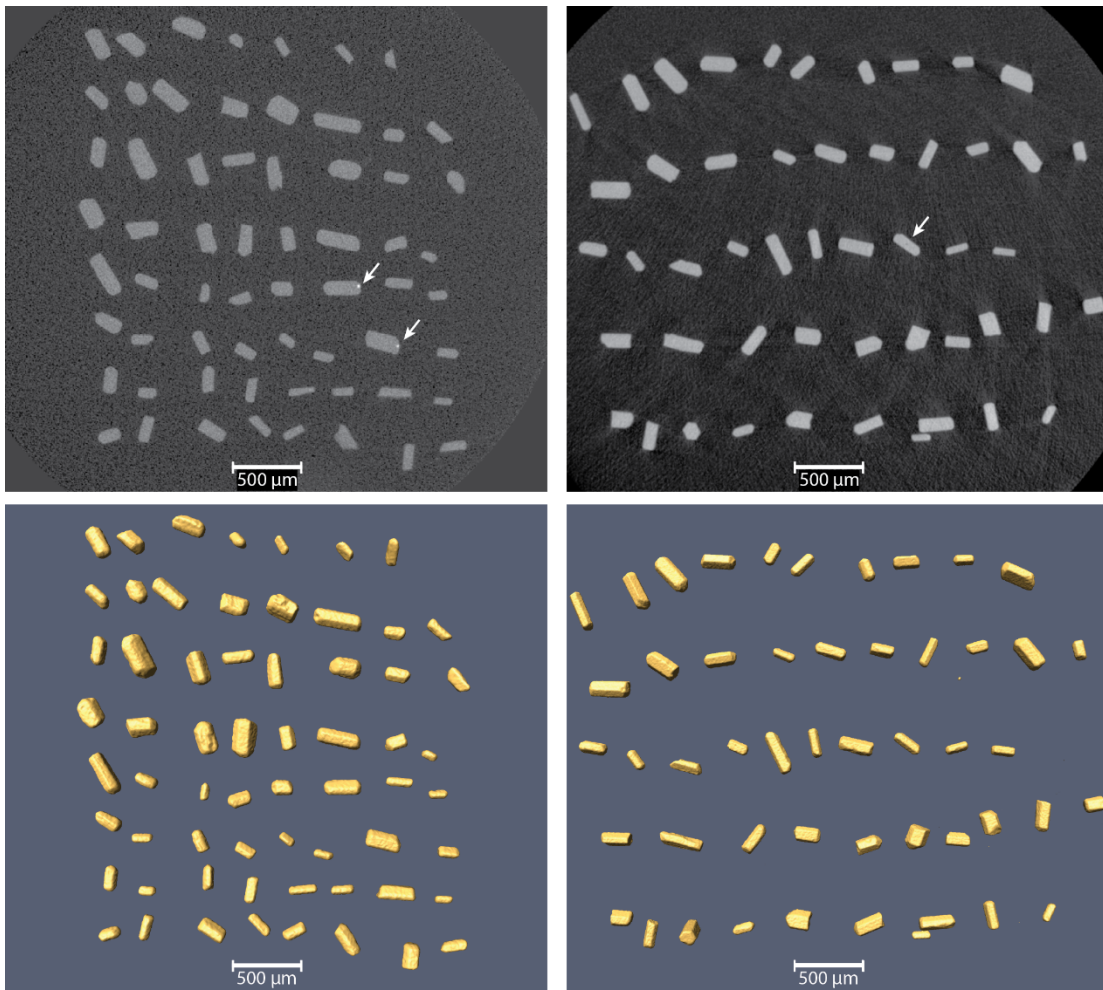
Figure 3: Example CT slices (upper row) and 3D renderings (lower row) of apatite grain mounts for BS95 (left) and 97BS (right). Arrows indicate two grains with high-attenuation mineral inclusions in BS95, and a fluid inclusion in 97BS. CT slice for 97BS is actually an oblique slice through the original data, to allow all grains to appear in the same image.

Example images from the two data sets are shown in Figure 3, illustrating some of the trade-offs. The scan data for BS95 are noisier, due primarily to the faster acquisition, higher X-ray energy, and more severe filtering. Even with this level of noise, high-attenuation inclusions are evident. The scan data for 97BS are less noisy, allowing detection of a fluid inclusion, but beam hardening due to the lower-energy X-ray spectrum has caused faint streaks to emanate from or connect some grains. These subtle artifacts have a negligible effect on measurements, but may be expected to increase in severity with more or higher-density grains.

2.4 Grain size and shape, F_T, mass calculations

2.4.1 2D measurement calculations

The microscope length and width measurements are used to calculate volume and surface area, which are then used to calculate mass, ESR, and UF_T and ThF_T for each apatite grain, following methods laid out in Farley et al. (1996), Farley and Stockli (2002), and Farley (2002) (Figure 4). An equidimensional hexagonal prism geometry was assumed with the length (L) measurement for height of the prism, and the half-width (r) for the radius of the prism. All equations used for calculating these parameters are included below or in the Appendix.

EQ. 1 Volume (V):

$$V = \frac{3 \times \sqrt{3}}{2} \times r^2 \times L,$$ where L is height and r is the half width.

EQ. 2 Surface Area (SA):

$$SA = 6 \times r \times L + 3\sqrt{3} \times r^2,$$ where L is height and r is the half width

EQ. 3 Equivalent Spherical Radius (ESR):

$$ESR = 3 * \frac{V}{SA}$$

EQ. 4 Mass:

$$mass = 3.2(\frac{g}{cm^3}) \times V(mm^3) \times 1000$$

EQ. 5 UF_T and ThF_T (2D case, e.g., Farley, 2002):

$$UFt = 1 - \left(5.13 \times \frac{SA}{V}\right) + (6.78 \times \frac{SA^2}{V})$$

$$ThFt = 1 - \left(5.9 \times \frac{SA}{V}\right) + (8.99 \times \frac{SA^2}{V})$$

EQ. 6 Mean F_T (see Appendix for explanation)

    a)   From Farley et al., 1996 (used here for 2D calculations)

$$F_T = a_{238} \times UF_T + (1 - a_{238}) \times ThF_T,$$

where $a_{238} = (1.04 + 0.245 \times \frac{Th}{U})^{-1}$

b) From Blob3D, this study (used here for 3D calculations):

$$\overline{F_T} = A_{238}F_{T,238} + A_{232}F_{T,232} + (1 - A_{238} - A_{232})F_{T,235},$$

where $A_{238} = (1.04 + 0.247 \left[\frac{Th}{U}\right])^{-1}$ and $A_{232} = (1 + 4.21/\left[\frac{Th}{U}\right])^{-1}$

5 EQ. 7 Effective Uranium Concentration (eU) (see Appendix for explanation):

$$eU = [U] + 0.238[Th] + 0.0012[Sm]$$
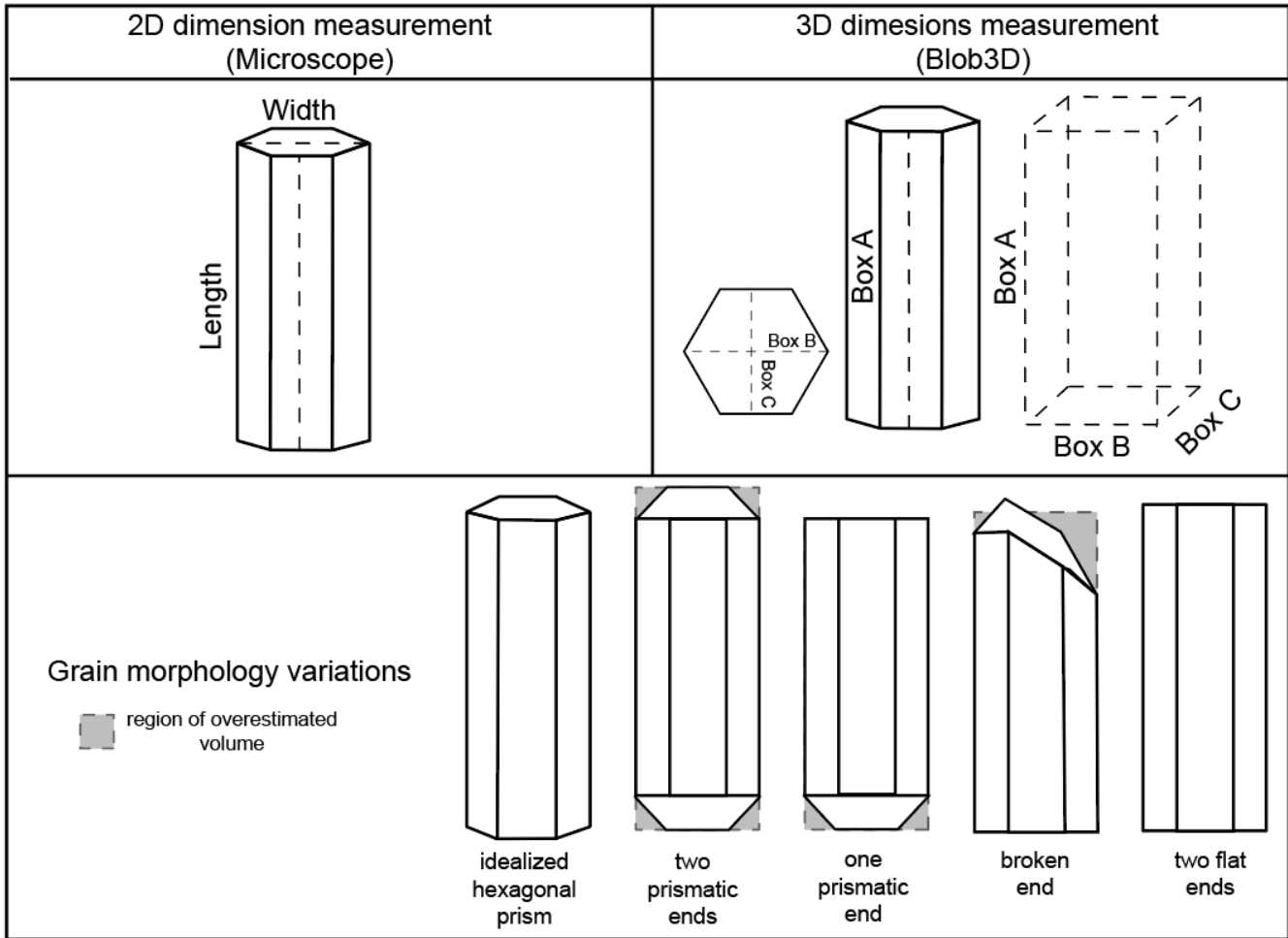
Figure 4: (Top) Rendering of dimension data collected by 2D and 3D methods. Length and width are measured in 2D using an optical microscope measuring the long and wide axis of a grain. Blob3D reports the length, width and height (Box A, B and C) based on the
10 **best-fit rectangle for the grain dimensions. (Bottom) Rendering of the full range of variations in grain terminations exhibited by the apatite in this study. Highlighted in gray are potential areas of over-estimated volume if an ideal hexagonal prism is assumed and calculated with 2D length and width data.**

2.4.2 3D calculations

Our principal 3D calculations were implemented in Blob3D (Ketcham 2005), a program written in the IDL programming environment for efficient measurement of the dimensions, shape, and orientation of discrete features in volumetric data sets. The typical Blob3D method for calculating volume is to segment the grains based a threshold set at 50% of the CT number (grayscale) difference between apatite and the surrounding air. If grains are touching, or close enough to touching that their selected regions are connected, the software provides several separation methods, the simplest being an erode/dilate procedure. Volume is calculated as the number of voxels in a grain multiplied by the voxel volume, and surface area is calculated by summing the areas of the triangular facets of an isosurface surrounding the grain, which is smoothed to reduce excess roughness from the cubic voxel edges. The shape parameters BoxA, BoxB, and BoxC are respectively the length (L), width (W), and height, corresponding to the dimensions of the smallest rectangular box that will enclose the grain (Ketcham and Mote, 2019). BoxC is calculated as the shortest 3D caliper length, BoxB is the shortest caliper length orthogonal to BoxC, and BoxA is the caliper length perpendicular to BoxC and BoxB (Figure 4; Appendix C).

A Monte Carlo method was implemented to measure F_T, probably similar in many, but not all, respects to previous work (Herman et al., 2007; Glotzbach et al., 2019). Stopping distances for ^{238}U, ^{235}U, ^{232}Th, and ^{147}Sm for the set of minerals reported in Ketcham et al. (2011) are included in the software. Taking the set of selected voxels for a grain, the origin point for each alpha particle is selected by first randomizing from which voxel to start, and then randomizing an (x,y,z) location within that voxel. The direction for each particle is obtained by sequentially stepping through a list of near-uniformly distributed orientations calculated by starting with an octahedron and subdividing each triangular face four times until there are 1026 vertices, which are then scaled to lie on a unit sphere (Ketcham and Ryan, 2004). This approach provides slightly better precision than randomizing orientations, and 200,000 Monte Carlo samples are sufficient to get precision to within 0.1% in all tests reported below. Separate F_T factors for each decay chain ($F_{T,238}$, $F_{T,235}$, $F_{T,232}$, $F_{T,147}$) are calculated, and a revised method for calculating mean F_T that more precisely accounts for ^{235}U is provided in EQ 6 (explanation in Appendix A).

If the resolution of the scan is low with respect to the stopping distance (^{238}U stopping distance/voxel size < 4), excess surface roughness effects from voxelation are reduced by super-sampling. The voxels for each grain, and surrounding voxels, are subdivided into 27 (3^3) elements, and the super-sampled volume is smoothed with a 5-voxel-wide cubic kernel. The result is then thresholded using a value that maintains the original volume as closely as possible.

These methods were tested on ideal spheres and cylinders, with radii of 63 and 31.5 μm and the latter with an aspect ratio of 4 (Appendix B). At voxel sizes up to 8 and 4 μm for the respective radii, mean $F_{T,238}$ values averaged within 0.2% of the ideal-shape values for spheres; further doubling the voxel sizes raised the mean error to 0.5%. Cylinders performed better, with a mean error of 0.3% when voxel sizes were ¼ of the radius.

In their Monte Carlo F_T implementation, Herman et al. (2007) report poor precision for small spheres when their centers are not centered in a voxel, with errors rising to several percent for a 40-μm radius sphere with 6.3-μm voxels across a range of center locations (calculated F_T range ~0.58-0.67). Errors of this magnitude correspond to the effect of getting the

radius wrong by plus or minus almost an entire voxel (~15% of the radius), too large to be reasonable and probably caused by a problem with their test. We tested our segmentation method by running 100,000 trials randomizing the location of the sphere center using the same radius and voxel size and got maximum radius errors of +0.8/-1.1% and a standard deviation of 0.2% (Appendix B). We are thus confident that our implementation provides a high degree of accuracy and precision on even very

5   small grains at low resolutions where voxel sizes are up to 25% of the radius.

We took three approaches to calculating ESR from the 3D data. The first two are based on the equivalent surface-to-volume ratio (SV) approach (Meesters and Dunai, 2002). The model-based value ESR_{SVm} uses the BoxA and BoxB caliper dimensions as L and W for equations (1) through (3), while the 3D CT-based value ESR_{SV3D} uses the 3D-measured volume and surface area for equation (3). Because of the unsupported assumptions of the model approach and the shortcomings of

10   surface area measurements, both discussed below, neither of these solutions are ideal. An alternative ESR is based on the equivalent-F_T approach; Ketcham et al. (2011) demonstrated than an equivalent-F_T sphere provides a more accurate conversion for diffusion calculations than an equivalent-SV one. The set of calculations to determine the F_T-equivalent sphere radius ESR_{FT} are provided in Appendix A.

2.5 (U-Th)/He procedure

15   The apatite (U-Th)/He ages were analyzed in the UTChron Thermochronology Laboratory at the University of Texas at Austin. Individual grains were measured, wrapped into platinum tubes, loaded into a 42-hole sample holder, and pumped to ultra-high vacuum. Each aliquot was heated to ~1070°C for 5 minutes using a Fusions Diode laser system. The released gas was spiked with a [3]He tracer, and purified by a Janis cryogenic cold trap at 40K and SAES NP-10 getter prior to measurement of the [4]He/[3]He on a Blazers Prisma QMS-200 quadrupole mass spectrometer. Final [4]He contents were calculated using a

20   manometrically-calibrated [4]He standard of known concentration measured during the analytical run. All apatite aliquots were reheated once under the same conditions to ensure full gas release.

After degassing, the platinum packets containing the apatite grains were placed into plastic vials and dissolved in a 100 μl 30% HNO_3 [235]U-[230]Th-[149]Sm spike solution for 90 minutes at 90°C. After acid digestion, 500 μl of Mili-Q ultra-pure H_2O was added to dilute the solutions to ~5% HNO_3 and equilibrated for ≥ 24 hours prior to analysis. The solutions were

25   analyzed using a Thermo Element2 HR-ICP-MS equipped with a 50μl/min micro-concentric nebulizer. Final [238]U, [232]Th, and [147]Sm values were blank corrected and calibrated using a spiked, gravimetrically calibrated ~ 1 ppb standard solution. Final (uncorrected) ages were calculated by solving the He age equation by means of Taylor Series expansion and reported with a 6% standard error, based on long-term intra-laboratory analysis of apatite age standards. Corrected final ages are determined by dividing the uncorrected age by the mean F_T factor (EQ. 5). U, Th, and Sm concentration, although not used in the age

30   calculations, were determined for reporting purposes using the grain volumes and a nominal apatite density (e.g., Figure 4, EQ. 4).

3 Results

CT scanning combined with Blob3D analysis provides 3D grain-specific volume, surface area, dimensions, and F_T factors for each decay chain. The 2D optical measurements provide dimension information, which are used to calculate volume, surface area, UF_T and ThF_T based on an assumed grain geometry of an equidimensional hexagonal prism (all results are reported in the appendix). We assume that the 3D-measured volume and F_T values are sufficiently accurate to benchmark the 2D data (all comparisons reported in Table 1 and Figure 5). Surface area is more problematic to benchmark due to a number of factors, such as fractal roughness, CT data blurring and voxelation effects, as discussed below, and thus 2D and 3D results can only be compared in a relative sense for surface area.

2D and 3D data are compared for each sample and as an entire population in Tables 1 and 2. The average 3D/2D ratio of each parameter is reported with its 1σ standard deviation. This average ratio shows whether the 2D measurements on average overestimates (ratio <1) or underestimates (ratio >1) the 3D measurements. Also reported is the absolute percent difference between the 2D and 3D measurements to illustrate the magnitude of deviation between the measurements. While comparing 2D and 3D results, it became apparent that one 2D grain measurement was made at an incorrect microscope magnification setting, causing the length and width to be off by 2x, far greater than every other grain measured. Hence, this grain measurement (97BS-CR8-1) was not included when calculating the average differences between 3D and 2D measuring techniques.

3.1 Grain factors

Grains from both samples display a range of habits typical for apatite, including two flat ends, two prismatic ends, one flat and one prismatic end, and one or two broken or chipped ends (Figures 1 & 4). The grain morphology and the presence of any visible inclusions were recorded during handpicking (Table 2). Surprisingly, there are no clear systematic relationships between the presence of inclusions and grain age, or grain shape and ESR, volume, or surface area. The 2D length measurements are on average ~2% smaller than the 3D BoxA dimension. On the other hand, the 2D width dimension is on average ~3% greater than the 3D BoxB dimension (Table 1).

One inevitable source of uncertainty in 2D length and width measurements is analyst judgment and error. For example, if a grain has uneven terminations, it is at the analyst's discretion to measure the longest axis or split the difference, whereas the CT analysis always reflect the longest axis. Similarly, CT scanning is also not subject to any user error introduced by measuring the apatite grain not lying on its widest face, or at an incorrect magnification. In our dataset, a couple of grains have very large deviations from the CT-derived volume, which may be caused by the microscope magnification setting being slightly off during measuring. Of course, the degree of analyst error is subject to many factors (e.g., experience of the analyst, the age and type of microscope, measuring software, etc.) and must be addressed on a lab-by-lab basis. In this study we found that human error may lead to 'outliers' in the results, and therefore it is a factor that we consider.

3.2 Volume and Surface Area

Volumes and surface areas calculated using the 2D microscope dimensions both average ~20% larger than the 3D calculations for both volumes and surface areas (3D/2D$_{VOL}$ = 0.82, 3D/2D$_{SA}$ = 0.81) (Table 1, Figure 5). Specifically, 2D volumes and surface areas calculated from length and width data assuming a hexagonal prism shape have an absolute average
5   difference of 23 ± 32% (2σ) and 22 ± 18% (2σ), respectively, from 3D Blob3D calculated volumes and surface areas.

3.3 ESR and Mass

The 2D ESR is calculated using the surface area to volume ratio (SA/V), which is derived assuming a hexagonal prism with the length and width dimensions measured on the microscope (EQ. 2, Figure 6). The 3D data had the ESR calculated based on SA/V in three ways. First, the SA/V for ESR$_{SVm}$ is calculated using the BoxA and BoxB values provided by Blob3D
10   and assuming a hexagonal prism, mimicking the 2D approach. The variation between 2D and 3D ESR$_{SVm}$ measurements has a 2σ spread of ± 12%, but the variability is fairly evenly split in over- and under-estimating the ESR, such that the average 3D/2D ratio is 1.02. Second, the 3D SA/V is calculated using the surface area and volume measurements output by Blob3D (ESR$_{SV3D}$). The variation between 2D and 3D ESR$_{SV3D}$ is even larger at ±18% (2σ), with an average 3D/2D ratio of 1.01 (Table 1, Figure 5).

15   The F$_T$-based ESR was on average similar to the SV-based one (ESR$_{FT}$/ESR$_{SVm}$ = 1.0), but the variation was ±9% for the two samples, and extreme values were 9% higher and 21% lower. The relative variation of the ESR$_{FT}$ value with the 2D data is ±14%, similar to that for the other 3D ESR calculations (Table 1, Figure 5).

The grain mass is calculated from the volume data using a nominal apatite density, and therefore 2D and 3D mass determination directly reflect the variability in the 2D and 3D volume data. The 2D approach consistently overestimates the
20   mass, with a high degree of scatter (3D/2D = 0.82 ± 0.44 (2σ)) (Table 1, Figure 5).

3.4 F$_T$ corrections

UF$_T$ and ThF$_T$ correction factors calculated from the 2D data are generally 1-2% lower than the Blob3D U and Th F$_T$ factors. To combine the F$_T$ factors into a single term that is applied to the (U-Th)/He age, a mean F$_T$ was calculated in two ways using EQ. 6 (see methods). This results in mean F$_T$ factors that vary by an average of 2% between the 2D and 3D datasets.
25   The 1σ scatter in 3D/2D FT factors is 1.8%, though individual differences can reach up to 9% (Table 1, Figure 5).

3.5 (U-Th)/He age and effective uranium

We calculated the apatite (U-Th)/He corrected age by dividing the raw (U-Th)/He age by the mean F$_T$ factor. The 2DF$_T$ (U-Th)/He ages tend to be slightly older than the 3DF$_T$ (U-Th)/He ages (3D/2D = 0.99), owing to the fact that the 2DF$_T$ values are slightly lower, leading to a larger correction (Table 1, Figure 5). The average difference between the 2D and 3D F$_T$
30   corrected ages is 2%, mimicking that of the variation between 2D and 3D F$_T$ (full range is <1 to 9%). This has an insignificant

impact on the mean age and uncertainty for both samples. Sample 97BS-CR8 has a $2DF_T$ mean age of 56.8 ± 2.9 Ma, and a $3DF_T$ mean age of 56.0 ± 2.9 Ma (Table 2, Figure 5). Sample BS95-11.3 has a $2DF_T$ age of 12.2 ± 4.0 Ma and a $3DF_T$ mean age of 12.1 ± 4.0 Ma (Table 2, Figure 5).

The effective uranium concentrations (eU = [U] + [Th]*0.238 + [Sm]*0.0012) for the apatite are normalized to the mass of the grain. Since 2D and 3D grain masses calculations varied by ~25%, the eU concentration measurements vary by a similar degree (3D/2D = 1.29 ± 0.24 (2σ)) (Table 1, Figure 5). Note, not all grains were analysed for U, Th, and Sm and so there are less data for eU comparison than mass.
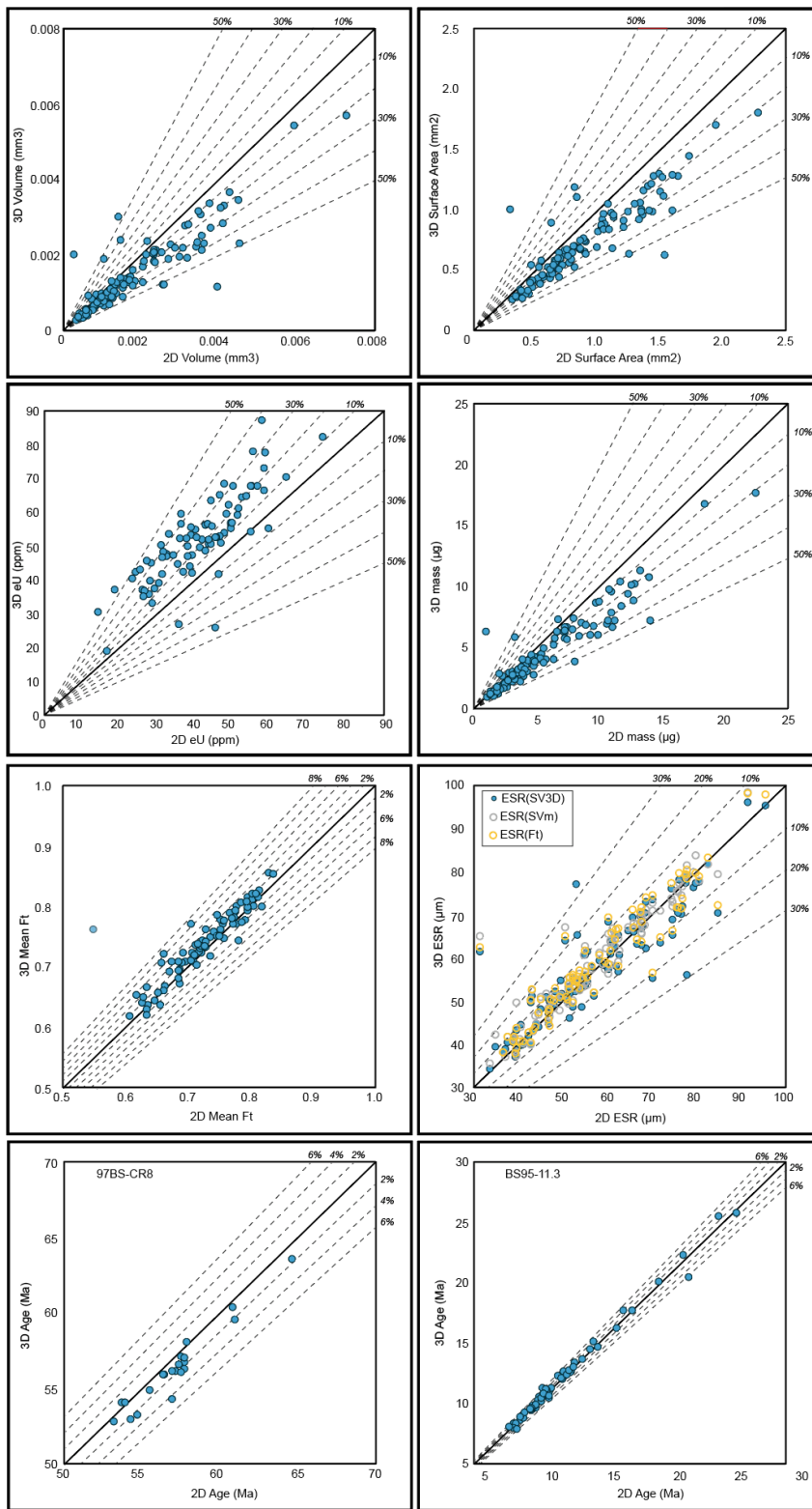
Figure 5: XY plots for Volume, Surface Area, eU, Mass, Mean F_T, ESR, and Age for both samples. Both samples are plotted together, unless otherwise noted. Each data point represents a single apatite aliquot. Dashed lines represent percent difference from 1:1 line (black line). 3D data measurements plotted on Y-axis in all plots. 2D measurements overestimate volume, surface area and mass, and underestimate eU and mean F_T.

4 Discussion

4.1 Accuracy of 2D vs 3D grain measurements

4.1.1 Volume and surface area

One of the main motivations behind this study was to assess the accuracy of 2D grain measurements and using an assumed grain geometry for calculating grain parameters (volume, ESR, mass, F_T) and the impact on the accuracy of the final (U-Th)/He age and eU. For this reason, we selected two samples from crystalline basement rocks that experienced relatively fast exhumation and no significant subsequent reheating, in order to reduce the impact of geologic or kinetic factors that could lead to age dispersion.

The most striking deviations between 2D and 3D measurements are in the volume and surface area, which 2D measurements consistently overestimated by 20-25% in our study, with a large degree of scatter (1σ = 22% and 14%, respectively). These results are in line with previous work. Evans and others (2008) observed a similar discrepancy in the 5 apatite grains they measured, where their 2D-based volumes were 30% greater than the 3D volumes (Table 3). Our dataset contains >100 apatite grains, implying that the 2D overestimation of volume (and therefore mass) may be systematic in the 2D measurement approach. In contrast, Glotzbach and others (2019) analyzed 24 apatite grains and found that the 2D volume measurements varied by a similar magnitude (~15%), but did not systematically overestimate the volume as in our study and Evans et al (2008) (Table 3). This is likely due in large part to their procedure of measuring three dimensions and selecting the appropriate shape model on a grain by grain basis, including ellipsoids for anhedral grains and accounting for terminations using the functions provided in Ketcham et al. (2011), rather than assuming exclusively flat-terminated hexagonal prisms.

There are multiple factors that can contribute to overestimating the volume of a given apatite crystal. First, the assumption of a hexagonal prism crystal shape with flat terminations, in which the length of the grain is used as the height of the prism, has the potential to overestimate the volume if the crystal has tapered ends (Figure 4). However, our data suggest this can only account for about a third of the volume difference, because even crystals with two flat (or broken) ends still had an average volume difference of 13%. Second, the ideal-prism model also presumes a perfect, equal-sided hexagonal cross section perpendicular to the c-axis, for which the ratio of width to height should be 2/sqrt(3), or 1.1547. The 3D shape measurements give mean ratios of 1.25(02) and 1.23(01) for our two samples, indicating that the cross sections are on average flatter than ideal hexagonal prisms. The non-ideality of this cross section was also noted by Glotzbach et al. (2019), and can result in either an underestimate or overestimate of volume, depending on which face the grain is lying on when measured in 2D. The systematic bias we observe is not surprising as apatites commonly come to rest on their flatter side, whereas some of our observed scatter comes from this not always being the case. We estimate that this shape divergence explains about a quarter of the departure between 2D and 3D volume in our data. The remaining deviation may be due to chipped crystals, surface roughness, or other deviations from a perfect prism that the 2D calculation cannot account for.

A number of factors will directly impact surface area calculations. Surface area is calculated from the 2D measurements by assuming a perfectly smooth prism. CT has the potential to capture irregular surfaces present in natural

apatite grains, which if present and resolution is sufficient, should lead to higher surface area calculations in the 3D data. However, surface area is problematic to measure in CT data, regardless of resolution. Irregular surfaces are to some degree fractal entities, making their measured areas dependent on measurement scale, and the "correct" answer is not straightforward to define. All CT images are naturally blurry to some extent, smoothing out both irregularities and also sharp corners and edges. Conversely, the 3D measurement process of segmentation by thresholding can lead to artificial enhancement of surface area due to "voxelation" effects (the 3D equivalent of pixilation).

In our data, the 2D measurements consistently result in a higher surface area than the 3D measurements. This is probably partly due to the ~5-μm resolution of our CT data, and also to the flat-terminated hexagonal prism model leading to an overestimate. Evans et al. (2008) observe a similar discrepancy in surface area measurements between 2D and 3D data (2D ~23% higher) with a 3.77 μm resolution scan (Table 3). On the other hand, Glotzbach et al., (2019) scanned their grains at a 1.2 μm resolution and their 2D measurements gave surface areas on average 8% lower than 3D (Table 3). As with volume, a large part of the difference is probably due to their using a more accurate shape model than an ideal equal-sided hexagonal prism. The overshoot may be in part due to their higher CT data resolution capturing roughness better, but their 3D images also show voxelation effects such as ridge sets on flat surfaces that likely increased their surface areas to an unknown extent.

We note that nature of the alpha stopping process, both in reality and as simulated, makes it essentially a ~20-μm smoothing filter, and so short-length-scale roughness has a negligible effect on alpha particle retention and F_T calculation. This point is demonstrated by our sensitivity analysis (Appendix B) which shows that a bumpy, voxelated sphere has the same F_T correction as a perfect, smooth one. Thus, while surface area is difficult to measure precisely in general, it is unimportant to measure precisely for this application.

4.1.2 Mass and eU

The discrepancy in volume between 2D and 3D measurements directly impacts the mass calculation, causing the grain masses derived from the 2D measurements to be ~25% higher than the 3D grain mass determinations (Figure 6). Evans and others (2008) found similar deviations, with their masses calculated from 2D volumes ~30% greater than their masses for 3D volumes (Table 3). Both of these divergences stem from using the assumption of a flat-ended hexagonal prism, whereas an approach that takes grain shape into account when choosing the F_T formula (Ketcham et al., 2011; Glotzbach et al., 2019) avoids this systematic bias. However, in all cases that use perfect shape models, the relative scatter is on the order of 20% (1σ), which is high enough to be worth fixing.

Although the age equation does not require knowledge of the grain volume or mass, both are necessary to calculate reported concentrations for U, Th, Sm and He (Figure 6). The U, Th, Sm concentrations, often combined into a single term, 'effective uranium' (eU), has been used a proxy for radiation damage within a crystal, and age versus eU correlations are commonly used for interpretation of age scatter, and thermal history inverse modeling (e.g., Flowers et al., 2009; Guenthner et al., 2013; Fox et al. 2019). Therefore, accurate knowledge of volume has cascading effects from mass to eU concentration and age interpretation (Figure 6). Comparison between eU calculated for the 3D mass data and 2D mass data show that the 2D

masses underestimate the bulk eU concentrations by ~20-30%. This is consistent with the 2D mass data being ~25% higher than the 3D mass data, which would have the effect of 'diluting' any eU signal; moreover, the much higher degree of scatter in the mass data caused by 2D analysis (±44% (2σ)) can be expected to muddy any age-eU correlation that may be present.

4.1.3 ESR

5   The various ESR calculations all yielded similar results on average, but high degrees of variation between measurement and calculation modes (5-6%). In addition to being more accurate for simplifying complex shapes to spheres for diffusion calculations, the ESR_{FT} method is also likely more robust than others that presume or measure surface area. Surface area, beyond being difficult to define and measure for irregular natural objects in a resolution-resistant way, has only secondary importance for diffusion and F_T calculations when it varies on a fine scale compared to the grain (i.e. μm-scale roughness). Analogously with mass, excess variation in ESR (±14% (2σ)) can degrade age-size correlations.

4.1.4 F_T

        A somewhat surprising result of our study is that, despite volume and surface areas being very different between the 2D and 3D methods, these differences largely cancelled each other out in S/V-based F_T calculations. This is in large part because volume and surface area co-vary, both in the assumed models and the actual measurements, so an error in one leads to a similar-magnitude error in the other (Figure 6).

        A result that more closely conformed to expectation is that, as grain size fell, dispersion between 2D and 3D F_T values increased, although it remained modest. The standard deviation of 3D/2D U_{FT} was 2.7% for grains with U_{FT} values from 0.6-0.7, 2.4% from 0.7-0.8, and 1.3% for grains above 0.8.

        While the above comparison takes into account 24 to 53 grains per sample, most applications of (U-Th)/He analyse 3-5 grains per sample. As a more practical comparison of the difference between 2D and 3D Mean F_T, we randomly subsampled the average of four grains from our results 1000 times (Fig. 7). We found that even when sub-sampling four grains, ~90% of runs had a mean deviation in 3D/2D F_T less than 3%.
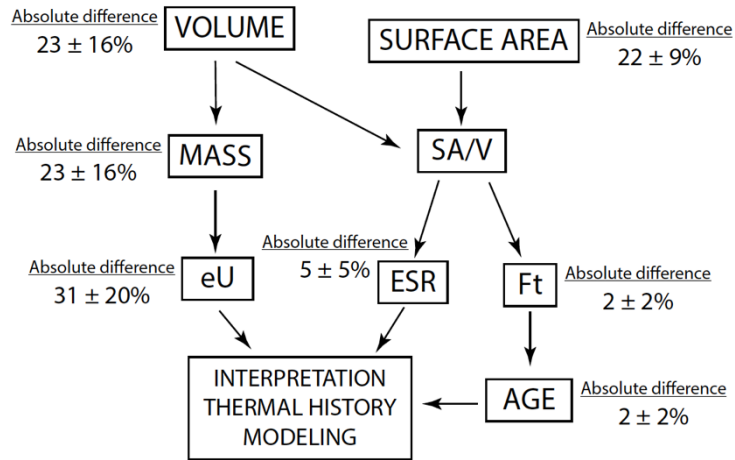
Figure 6: Workflow diagram showing the effect of volume and surface area measurements on other parameters used for (U-Th)/He age calculation and interpretation. The average absolute difference between 2D and 3D measurements for each of the parameters are reported with their 1σ uncertainties (reported in Table 1). Note the greatest deviations are in volume and surface area, and those parameters that rely on volume alone. ESR, F_T and age deviate less because they use the SA/V, which is ~1 between 2D and 3D measurements.

4.2 Reproducibility of (U-Th)/He ages

In addition to assessing the accuracy of using the 2D measurements, this study aimed to quantify the uncertainties that may be introduced by such measurements, particularly in F_T, as a means to potentially improve age accuracy, precision, and intra-sample dispersion. Previous studies have estimated that uncertainties in F_T calculation can account for 1-5% of sample age uncertainty (Evans et al., 2008; Glotzbach et al., 2019). Our results are consistent with this range, and suggest that uncertainties in the U and Th F_T calculation are on the order of 1-3%, and mean F_T varies by 2% (Table 1). We find the greatest deviations are likely caused by user error for our samples and not the assumed grain geometry. In samples with less euhedral apatite grains, the effects of F_T and an assumed grain geometry can increase.

Our data also show that the 3D F_T correction does not increase the overall sample age precision for the samples in this study. For sample 97BS-CR8, 24 apatite grains were analyzed, two of which are outliers. Of the two outliers, one (97BS-CR8-1) was clearly caused by a user error during microscope measurement, leading to an incorrect F_T correction (0.55) and old age (78.8 Ma). This was discovered during 3D image processing, in which the same grain was identified, measured correctly, and produced an F_T of 0.76 and a more congruent corrected age of 57.2 Ma. In contrast, for a second outlier (97BS-CR8-24), the 2D and 3D F_T corrected ages both produced anomalous ages of 101.2 and 98.4 Ma, respectively. An unusually high He concentration the likely culprit for the old age for this grain, but its cause is not evident from our data. Excluding these two outliers, the average age and uncertainty for the sample population (n=22 grains) calculated based on the 2D and 3D measurements are indistinguishable (56.8 ± 2.9 Ma and 56.0 ± 2.9 Ma); relative errors are 5.1% in both cases.

Similarly, the sample ages calculated with 3D and 2D data for 95BS-11.3 (n=59 aliquots) are indistinguishable, 12.2 ± 4.0 and 12.1 ± 4.0 Ma, respectively. Unlike sample 97BS-CR8, there was no clear-cut evidence of user error, and the relatively high age uncertainty (33%) is reproducible between the 2D and 3DF$_T$ corrected ages. Five aliquots produced ages > 20 Ma, which skews the mean age older (the median age is 10.2 Ma, within error of the previous reported age in Stockli et al., 2002). The apatite ages do not correlate with factors such as ESR (grain size) or eU. The >20 Ma aliquots all have high He concentrations (nmol/g) compared with the bulk of the sample, suggesting that excess He, possibly due to implantation from high U-Th neighbors, or the presence of undetected and insoluble high-eU inclusions.

In addition to the above calculations, we randomly sub-sampled four grains 1000 times to assess the variability in F$_T$ corrected age for a number of grains that is more comparable to other studies. The results are plotted on Fig. 7 and reported in Table 2. The mean of the 1000 trials is indistinguishable from the entire analyzed population.

Overall, these data suggest that although the 3DF$_T$ can provide a more accurate F$_T$ correction and varies from 2D estimations by ~2%, it has a minimal effect on the calculated sample age (1-2%) and no effect on the reproducibility for these two samples. This is not surprising, as a ~2% error would constitute a negligible proportion of the often cited 6% dispersion derived from analyzing age standards; error propagation indicates that removing a source of 2% error would only reduce an overall 6% error to 5.7%. This points to the importance of other factors in intra-sample dispersion, such as U-Th zonation, and/or excess He from nano-inclusions or high U-Th neighbors.
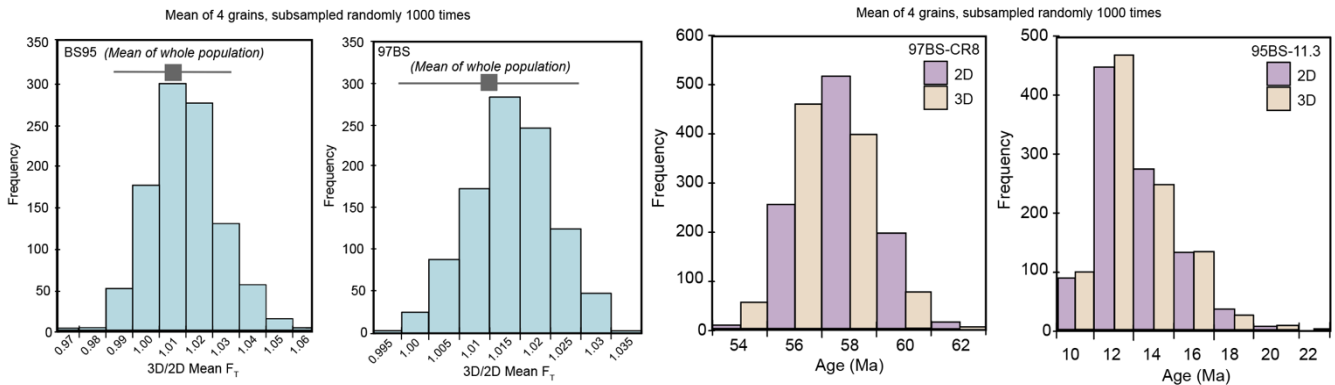
Figure 7: Histograms showing the 3D/2D Mean F$_T$ and 2D and 3D (U-Th)/He age. Histograms show the results of randomly subsampling 4 grains from each sample 1000 times. For Mean F$_T$ and sample age, the subsampling is indistinguishable from the mean of the whole population analyzed. Note: numbers on x-axis refer to the bars (bins) and not the tick marks.

4.3 Effects of inclusions or broken grains

It is widely regarded that inclusions and broken grains are both contributors to intra-sample dispersion and inaccurate He ages, particularly anomalously old ages. Inclusions in apatite can act as He traps or a source for excess He, particularly mineral inclusions that do not dissolve during apatite HNO$_3$ digestion (e.g., Ehlers and Farley, 2003). Both apatite samples

had multiple grains with high density and low-density inclusions detectable by microscope during picking and/or the CT scan (Figure 2). In both samples, the presence of inclusions did not have any discernable effect on the (U-Th)/He age (Table 2). While inclusions are certainly a source for error and dispersion in many samples and should be avoided, at least the easily visible ones do not appear be relevant in these samples, which suggests they are likely also not U-Th bearing inclusions. For future studies, an added benefit of CT is the detection of high- and low-density mineral and fluid inclusions.

Similarly, broken grains can be a source of dispersion if they were broken after the sample passed through the He partial retention zone, e.g., after the grain began to accumulate He (see Beucher et al., 2013; Brown et al., 2013). Typically, this may occur during erosional transport or during mineral separations. Brown and others (2013) assess that broken grains can contribute from 7 to > 50% dispersion from the sample age, depending on cooling history. In our samples, grain terminations varied from doubly prismatic to flat, and in some cases appeared chipped or broken. However, there is no clear correlation between the chipped or broken grains and He age (See Table 2). One possibility is that the grains broke prior to cooling through the He retention zone. This seems somewhat unlikely, given that both samples come from crystalline rocks. Alternatively, and perhaps more plausibly, the variety of crystal habits may reflect how the crystals grew in the host rock. In any case, the grains in these samples that appear to be chipped or broken are not obvious sources for the age dispersion observed in the samples.

4.4 Benefits and limitations of X-ray CT over microscope measurements

This study purposefully selected 'high quality' apatite from fast-cooled plutonic samples to quantify the base uncertainty introduced by 2D measurements and grain shape assumptions on F_T and (U-Th)/He age factors. Although we found that 3D grain characterization techniques did not reduce intrasample age dispersion in our samples, it is still highly probable that the 3D approach can improve dispersion in samples with less euhedral apatite and more complicated geologic histories. Furthermore, CT scanning mineral grains for (U-Th)/He chronometry has both analytical and practical benefits that go beyond grain measurement. CT provides more accurate grain volume measurements, which becomes increasingly important as grain shapes deviate from idealized forms (e.g., abraded or broken grains). CT data are able to highlight inclusions or other internal heterogeneities based on contrasts in density in the x-ray data, which may not be visible by the naked eye. Furthermore, the CT-mounting method and scanning conditions outlined in this study allows for the scanning of up to 250 grains in a single session, and potentially many more, making it cost and time effective. Different mineral phases can be scanned together, and data can be processed in a batch so that from a single scan, one can gather volume, surface area, caliper dimensions, F_T, mass, and ESR at once for several samples and phases. Furthermore, the 3DF_T and F_T-based ESR capabilities of Blob3D software introduced in this study makes batch processing the CT data straightforward. Thus, an analyst will be able to image, characterize, and quantify 100's of mineral grains in significantly less time than conventional microscope measuring. We anticipate that more volume-based shape measurements can and will be developed to automatically and quantitatively evaluate grains for euhedrality, rounding, broken faces, and a wealth of other potentially informative data.

CT scanning mineral grains used for (U-Th)/He dating also has the benefit of removing many possible sources of 'user error' during the grain measurement step. Unlike with microscope measurements, the orientation of the apatite grain on the CT mount does not matter, and there is no need to set a magnification or trace the dimensions of the grain by hand, reducing potential for mistakes. CT also eliminates variability that may arise from different microscopes, lighting conditions, and imaging software, and creates a digital archive of 3D grain shapes, densities and internal structures that a microscope photo cannot capture.

The one required user input to our method is specifying the threshold CT number for grain measurement, for which we recommend using the midpoint value between the mineral and the surrounding medium (e.g., air, epoxy). When scan resolution is low in terms of both voxel size and sharpness, additional care is required; if edge blurring approaches the center of a grain, an alternative thresholding or segmentation procedure may be necessary to obtain accurate volumes (Ketcham and Mote, 2019). We thus do not recommend pushing resolution limits too far; voxel sizes generally should not exceed 1/8 to 1/10 of the shortest dimension of a grain. CT measurement accuracy also requires that the scans be as free as possible from artifacts that cause local changes in CT numbers, such as beam hardening, photon starvation, or rings. We further note that software artifact corrections can sometimes introduce secondary artifacts that may be harder to recognize but still affect calculations (Ketcham and Carlson, 2001), and so care is required in the scanning process.

The main limitation of using CT is access to the instrumentation and cost for sample analysis. However, CT scanners are becoming more common as desktop instruments in earth science departments, and many universities have imaging facilities that include micro-CT. As CT instruments continue to proliferate and costs continue to fall, we anticipate that measuring, screening, and documenting grains used for thermo-geochronology will become a widely-used practice.

5 Conclusions

The shape and size of 109 apatite grains from two rapidly-cooled plutonic samples were analysed by 2D and 3D methods. 2D length and width measurements made on an optical microscope were used to calculate surface area, volume, ESR, mass, and F_T assuming an ideal equal-sided, flat-terminated hexagonal prism grain shape. The same apatite crystals were scanned using x-ray computed tomography at a 4-5 µm resolution, and the same factors were calculated using Blob3D software, which does not require assuming a grain shape. 83 new apatite (U-Th)/He ages were collected to resolve the influence of 2D versus 3D F_T correction factors on final (U-Th)/He age and reproducibility. With these data, we derive the following conclusions:

1. Deviations between 2D and 3D measurements were greatest in volume and surface area (~25%), which caused mass and eU calculations to deviate by a similar magnitude. Volume and surface area measurements also showed high dispersion of 44% and 28% (2σ), respectively. These sources of scatter weaken the ability to use age-eU and age-size correlations to help interpret age distributions.

2. 2D F_T measurements only contribute ~ 2% error on average, even with the erroneous assumption of an ideal grain shape.

3. Inclusions and broken or chipped ends did not have a discernible impact on the (U-Th)/He age dispersion in these samples.

4. The combined (U-Th)/He ages for each sample were indistinguishable for 2D and 3D F_T corrections. Similarly, the amount of intra-sample dispersion was identical (both >5%). This implies that factors other than F_T dominate the intra-sample age uncertainty.

In addition, we present a bulk scanning method that easily allows analysis of > 250 grains in a single session, new Blob3D software 3D F_T and shape measurement functions, and new calculations for eU and ESR_{FT}.

Conflict of interest: The authors declare that they have no conflict of interest.

Author contributions: EHGC collected and processed data, made figures, and contributed to the writing of the manuscript. RAK initiated the study, processed data, and contributed to the writing of the manuscript. DFS initiated the study with RAK and contributed to the writing of the manuscript.

Code/Data availability: Code and data are available in the supplement to this manuscript. CT data is archived at https://www.digitalrocksportal.org/projects/216, DOI: 10.17612/CZYH-KC13.

Acknowledgements

We thank Jessie Maisano for acquiring and reconstructing the CT data. These data were collected at the UTCT NSF Multi-User Facility. This work was funded through Jackson School of Geosciences funds to Stockli and UTCT facility support (EAR-1762458) to Ketcham, and an NSF GRF and WHOI Postdoc Scholarship to Cooperdock. This manuscript was improved by helpful reviews from C. Glotzbach and two anonymous reviewers.

References

ASTM: E1441-11: Standard Guide for Computed Tomography (CT) Imaging, ASTM International, West Conshohocken, PA, https://doi.org/10.1520/E1441-11, 2011.

Bargnesi, E.A., Stockli, D.F., Hourigan, J.K., and Hager, C.: Improved accuracy of zircon (U – Th)/ He ages by rectifying parent nuclide zonation with practical methods, Chem. Geol. 426, 158–169, https://doi.org/10.1016/j.chemgeo.2016.01.017, 2016.

Blott, S. J. and Pye, K.: Particle shape: a review and new methods of characterization and classification, Sedimentology, 55, 31-63, https://doi.org/10.1111/j.1365-3091.2007.00892.x, 2008.

5   Beucher, R., Brown, R. W., Roper, S. Stuart, F., and Persano, C.: Natural age dispersion arising from the analysis of broken cystals: Part II. Practical application to apatite (U-Th)/He thermochronometry, Geochim. Cosmochim. Acta, 120, 395-416, http://dx.doi.org/10.1016/j.gca.2013.05.042, 2013.

Brown, R. W., Beucher, R., Roper, S. Persano, C., Stuart, F., and Fitzgerald, P.: Natural age dispersion arising from the analysis of broken crystals. Part I: Theoretical basis and implications for the apatite (U-Th)/He thermochronometer,
10   Geochim. Cosmochim. Acta, 122, 478-497, https://doi.org/10.1016/j.gca.2013.05.041, 2013.

Browne, E.: Nuclear Data Sheets for A = 232, Nuclear Data Sheets, 107, 2579-2648, doi:10.1016/j.nds.2006.09.001, 2006.

Danisik, M., McInnes, B. I. A., Kirkland, C. L., McDonald, B. J., Evans, N. J., and Becker, T.: Seeing is believing:
15   Visualization of He distribution in zircon and implications for thermal history reconstruction on single crystals, Sci. Adv., 3, 2, DOI: 10.1126/sciadv.1601121, 2017.

Ehlers, T. A., and Farley, K. A.: Apatite (U-Th)/He Thermochronometry: Methods and Applications to Problems in Tectonic and Surface Processes, Earth Planet. Sci. Lett., 206, 1-2, 1-14. https://doi.org/10.1016/S0012-821X(02)01069-5, 2003.
20
Evans, N.J., McInnes, B.I.A., Squelch, A.P., Austin, P.J., McDonald, B.J., and Wu, Q.: Application of X-ray micro-computed tomography in (U-Th)/He thermochronology, Chem. Geol., 257, 101–113, https://doi.org/10.1016/j.chemgeo.2008.08.02, 2008.

25   Farley, K.A.: (U-Th)/He Dating: Techniques, Calibrations, and Applications, Rev. Mineral. Geochemistry, 47, 819–844, https://doi.org/10.2138/rmg.2002.47.18, 2002.

Farley, K.A., and Stockli, D.F.: (U-Th)/ He Dating of Phosphates: Apatite, Monazite, and Xenotime, Rev. Mineral. Geochemistry, 15, 559–577, https://doi.org/10.2138/rmg.2002.48.15, 2002.
30
Farley, K.A., Wolf, R.A., and Silver, L.T.: The effects of long alpha-stopping distances on (U - Th)/He age, Geochim. Cosmochim. Acta, 60, 4223–4229, https://doi.org/10.1016/S0016-7037(96)00193-7, 1996.

Flowers, R. M.: Exploiting radiation damage control on apatite (U–Th)/He dates in cratonic regions, Earth Planet. Sci. Lett., 277, 1-2, 148-155, https://doi.org/10.1016/j.epsl.2008.10.005, 2009.

Flowers, R. M., and Kelley, S. A.: Interpreting data dispersion and "inverted" dates in apatite (U–Th)/He and fission-track datasets: An example from the US midcontinent, Geochim. Cosmochim. Acta, 75, 18, 5169-5186,

5   https://doi.org/10.1016/j.gca.2011.06.016, 2011.

Flowers, R.M., Ketcham, R.A., Shuster, D.L., and Farley, K.A.: Apatite (U-Th)/He thermochronometry using a radiation damage accumulation and annealing model, Geochim. Cosmochim. Acta 73, 2347–2365, https://doi.org/10.1016/j.gca.2009.01.015, 2009.

Flowers, R. M., Shuster, D. L., Wernicke, B. P, and Farley, K. A.: Radiation damage control on apatite (U-Th)/He dates from the Grand Canyon region, Colorado Plateau. Geology, 35 (5), 447-450, https://doi.org/10.1130/G23471A.1, 2007.

Fox, M., Dai, J.-G., and Carter, A.: Badly behaved detrital (U-Th)/He ages: Problems with He diffusion models or geological

15   models?, Geochem. Geophys. Geosys., 20, 2418-2432, https://doi.org/10.1029/2018GC008102, 2019.

Gautheron, C., Tassan-Got, L., Barbarand, J., and Pagel, M.: Effect of alpha-damage annealing on apatite (U–Th)/He thermochronology, Chem. Geol., 266, 3-4, 157-170, https://doi.org/10.1016/j.chemgeo.2009.06.001, 2009.

20   Gautheron, C., Tassan-Got, L., Ketcham, R.A., and Dobson, K.J.: Accounting for long alpha-particle stopping distances in (U-Th-Sm)/He geochronology: 3D modeling of diffusion, zoning, implantation, and abrasion, Geochim. Cosmochim. Acta 96, 44–56, https://doi.org/10.1016/j.gca.2012.08.016, 2012.

Glotzbach, C., Lang, K.A., Avdievitch, N.N., and Ehlers, T.A.: Increasing the accuracy of (U-Th(-Sm))/He dating with 3D

25   grain modelling, Chem. Geol, 506, 113–125, https://doi.org/10.1016/j.chemgeo.2018.12.032, 2019.

Guenthner, W.R., Reiners, P.W., Ketcham, R.A., Nasdala, L., and Giester, G.: Helium diffusion in natural zircon: radiation damage, anisotropy, and the interpretation of zircon (U-TH)/He thermochronology, Am. J. Sci., 313, 145–198, https://doi.org/10.2475/03.2013.01, 2013.

Herman, F., Braun, J., Senden, T.J., and Dunlap, W.J.: (U-Th)/He thermochronometry: Mapping 3D geometry using micro-X-ray tomography and solving the associated production-diffusion equation, Chem. Geol., 242, 126–136, https://doi.org/10.1016/j.chemgeo.2007.03.009, 2007.

Holden, N. E.: Total half-lives for selected nuclides, Pure and Applied Chemistry, 62(5), 941-958, doi:10.1351/pac199062050941, 1990.

5   Hourigan, J.K., Reiners, P.W., and Brandon, M.T.: U-Th zonation-dependent alpha-ejection in (U-Th)/He chronometry, Geochim. Cosmochim. Acta, 69, 3349–3365, https://doi.org/10.1016/j.gca.2005.01.024, 2005.

Ketcham, R.A.: Computational methods for quantitative analysis of three-dimensional features in geological specimens, Geosphere, 1(1), 32-41, https://doi.org/10.1130/GES00001.1, 2005.

Ketcham, R.A., and Carlson, W.D.: Acquisition, optimization and interpretation of X-ray computed tomographic imagery: Applications to the geosciences, Computers and Geosciences, 27, 381-400, https://doi.org/10.1016/S0098-3004(00)00116-3, 2001.

15   Ketcham, R.A., Gautheron, C., and Tassan-Got, L.: Accounting for long alpha-particle stopping distances in (U-Th-Sm)/He geochronology: Refinement of the baseline case, Geochim. Cosmochim. Acta, 75, 7779–7791, https://doi.org/10.1016/j.gca.2011.10.011, 2011.

Ketcham, R.A., and Mote, A.S.: Accurate measurement of small features in X-ray CT data volumes, demonstrated using
20   gold grains, Journal of Geophysical Research, 124, https://doi.org/10.1029/2018JB017083, 2019.

Ketcham, R.A., and Ryan, T.M.: Quantification and visualization of anisotropy in trabecular bone, Journal of Microscopy, 213, 158-171, doi: 10.1111/j.1365-2818.2004.01277.x, 2004.

25   Ketcham, R. A., Slottke, D. T., and Sharp, J. M. J.: Three-dimensional measurement of fractures in heterogeneous materials using high-resolution X-ray CT, Geosphere, 6, 499-514, https://doi.org/10.1130/GES00552.1, 2010.

Le Roux, L. A., and Glendenin , L. E., 1963. Half-life of 232Th, in Proceedings of the National Meeting on Nuclear Energy, Pretoria, South Africa, 83, 94.

McDannell, K. T., Zeitler, P. K., Janes, D. G., Idleman, B. D., and Fayon, A. K.: Screening apatites for (U-Th)/He thermochronometry via continuous ramped heating: He age components and implications for age dispersion, Geochim. Cosmochim. Acta, 223, 90-106. https://doi.org/10.1016/j.gca.2017.11.031, 2018.

Reiners, P. W., and Brandon, M. T.: Using thermochronology to understand orogenic erosion, Annual Review of Earth and Planetary Sciences, 34, 419-466, https://doi.org/10.1146/annurev.earth.34.031405.125202, 2006.

Reiners, P. W., and Farley, K. A.: Influence of crystal size on apatite (U-Th)/He thermochronology: an example from the
5   Bighorn Mountains, Wyoming, Earth and Planet. Sci. Lett., 188, 3-4, doi: 10.1016/S0012-821X(01)00341-7, 2001.

Shuster, D. L., Flowers, R. M., and Farley, K. A.: The influence of natural radiation damage on helium diffusion kinetics in apatite, Earth Planet. Sci. Lett., 249, 3-4, 148-161, https://doi.org/10.1016/j.epsl.2006.07.028, 2006.

10   Silver, L. T., and Deutsch, S.: Uranium-Lead Isotopic Variations in Zircons: A Case Study, J. Geol., 71(6), 721-758, https://doi.org/doi:10.1086/626951, 1963.

Sneed, E. D. and Folk, R. L.: Pebbles in the lower Colorado River, Texas a study in particle morphogenesis, J. Geol., 66, 114-150, https://doi.org/10.1086/626490, 1958.

Steiger, R. H., and E. Jäger.: Subcomission on geochronology: Convention on the use of decay constants in geo- and cosmochronology, Earth Planet. Sci. Lett., 36, 359-362, doi:10.1016/0012-821X(77)90060-7, 1977.

Stockli, D. F., Farley, K. A., and Dumitru, T. A.: Calibration of the apatite (U-Th)/He thermochronometer on an exhumed
20   fault block, White Mountains, California, Geology, 28, 11, 983-986, https://doi.org/10.1130/0091-7613(2000)28<983:COTAHT>2.0.CO;2, 2000.

Stockli, D.F., Surpless, B.E., Dumitru, T.A., and Farley, K.A.: Thermochronological constraints on the timing and magnitude of Miocene and Pliocene extension in the central Wassuk Range, western Nevada, Tectonics, 21,
25   https://doi.org/10.1029/2001TC001295, 2002.

Surpless, B., Stockli, D.F., Dumitru, T.A., and Miller, E.L.: Two-phase westward encroachment of basin and range extension into the northern Sierra Nevada, Tectonics, 21, 2-1-2–13, https://doi.org/10.1029/2000TC001257, 2002.

30   Wilson, L. and Huang, T. C.: The influence of shape on the atmospheric settling velocity of volcanic ash particles, Earth Planet. Sci. Lett., 44, 311-324, https://doi.org/10.1016/0012-821X(79)90179-1, 1979.

Zeitler, P. K., Herczeg, A. L., McDougall, I., and Honda, M., 1987. U-Th-He dating of apatite: A potential thermochronometer, Geochim. Cosmochim. Acta, 51, 10, 2865-2868, https://doi.org/10.1016/0016-7037(87)90164-5, 1987.

Table 1: Comparison of 2D and 3D measurement data

Sample	n (grains)	Avg 3D/2D	1σ	Absolute Difference				
				Avg	1 σ	Median	Max	Min
LENGTH								
97BS-CR8	*49*	1.00	0.03	2	2	2	7	0.1
BS95-11.3	*59*	0.97	0.1	6	7	5	49	0.1
Combined	*108*	0.98	0.1	4	6	3	49	0.1
WIDTH								
97BS-CR8	*49*	1.04	0.05	15	8	14	32	0.2
BS95-11.3	*59*	1.03	0.09	17	9	17	44	1
Combined	*108*	1.03	0.07	16	8	16	44	0.2
VOLUME								
97BS-CR8	*49*	0.85	0.25	22	19	18	109	2
BS95-11.3	*59*	0.80	0.20	24	14	23	80	5
Combined	*108*	0.82	0.22	23	16	20	109	2
SURFACE AREA								
97BS-CR8	*49*	0.83	0.15	21	10	20	60	4
BS95-11.3	*59*	0.80	0.12	22	8	21	43	3
Combined	*108*	0.81	0.14	22	9	20	60	3
MASS								
97BS-CR8	*49*	0.85	0.25	22	19	18	109	2
BS95-11.3	*59*	0.80	0.20	24	14	23	80	5
Combined	*108*	0.82	0.22	23	16	20	109	2
ESR (SV_m)								
97BS-CR8	*49*	1.03	0.04	4	4	2	21	0.1
BS95-11.3	*59*	1.02	0.08	5	6	4	32	0.03
Combined	*108*	1.02	0.06	5	5	3	32	0.03
ESR (SV_{3D})								
97BS-CR8	*49*	1.01	0.10	6	8	3	45	0.04
BS95-11.3	*59*	1.00	0.08	6	5	4	26	0.02
Combined	*108*	1.01	0.09	6	7	4	45	0.02
ESR (F_T)								
97BS-CR8	*49*	1.03	0.04	4	3	4	12	1
BS95-11.3	*59*	1.02	0.08	6	6	5	28	0.1
Combined	*108*	1.02	0.07	5	5	4	28	0.1

U F_T								
97BS-CR8	*49*	1.02	0.02	2	1	2	8	0.1
BS95-11.3	*59*	1.01	0.03	2	2	2	9	0.1
Combined	*108*	1.01	0.02	2	2	2	9	0.1
TH F_T								
97BS-CR8	*49*	1.01	0.02	2	1	1	7	0.01
BS95-11.3	*59*	1.00	0.03	2	2	2	9	0.1
Combined	*108*	1.00	0.03	2	2	1	9	0.01
MEAN F_T								
97BS-CR8	*49*	1.02	0.02	2	1	1	5	0.2
BS95-11.3	*59*	1.01	0.03	2	2	2	9	0.01
Combined	*108*	1.01	0.02	2	2	2	9	0.01
eU								
97BS-CR8	*49*	1.24	0.12	24	12	22	49	3
BS95-11.3	*59*	1.30	0.27	33	23	28	112	6
Combined	*108*	1.29	0.24	31	20	26	112	3
AGE								
97BS-CR8	*49*	0.99	0.01	2	1	1	5	0.2
BS95-11.3	*59*	0.99	0.03	2	2	2	8	0.01
Combined	*108*	0.99	0.02	2	2	2	8	0.01

ESR(SVm) = BoxA and BoxB assuming hexagonal prism shape, ESR(SV3D) = Blob3D vol and SA measurements, ESR(Ft) = Ft equivalent sphere

Table 2: (U-Th)/He Age results

Aliquot	Grain morph.	3DFt Age (Ma)	2DFt Age (Ma)
Sample 97BS-CR8-			
*1	FL PR	57.2	78.8
2	PR CH, i	56.3	57.7
3	PR PR	54.1	53.8
4	PR PR	56.2	57.2
5	PR PR	58.0	57.9
6	PR PR	54.1	53.9
7	PR PR	60.3	60.8
8	FL PR	53.0	54.3
9	FL PR	57.1	57.5
10	PR PR	56.0	56.4
11	FL FL, i	54.9	55.5
12	FL PR	63.5	64.6
13	FL PR	56.7	57.7
14	PR PR	55.9	56.4
15	PR PR	56.6	57.4
16	FL FL	-	-
17	FL PR	57.0	57.8
18	FL CH	-	-
19	FL PR	52.8	53.2
20	PR CH	49.7	51.2
21	PR CH	59.5	60.9
22	PR PR	53.3	54.7
23	PR CH	56.2	57.0
*24	PR PR	101.2	98.4
25	PR PR	56.1	57.5
26	PR PR	54.3	57.0
Average		**56.0**	**56.8**
Stdev		**2.9**	**2.9**
%RSD		**5.1**	**5.1**
Subsample Average		**56.0**	**56.9**

Sample BS95-11.3-			
1	PR PR	10.5	10.3
2	FL PR	13.6	13.8
3	FL PR	8.6	8.8
4	PR PR	13.3	13.2
5	FL PR	12.6	12.8
6	PR PR	10.8	11.1
7	PR PR	10.6	10.7
8	FL PR	10.2	10.2
9	PR PR	25.5	24.7
10	PR PR	12.3	11.8
11	PR CH	10.4	11.1
14	FL PR	10.8	10.8
15	FL PR	9.5	10.0
16	FL PR	9.7	9.9
17	FL CH	8.7	9.1
18	PR CH	9.2	9.2
19	FL PR	9.4	9.6
20	FL FL	25.7	26.1
21	FL PR	8.8	8.8
22	PR PR	8.3	8.2
23	PR CH, i	12.6	12.2
24	FL PR	10.8	10.9
25	FL PR	9.8	10.0
26	FL FL	10.1	10.5
27	FL CH	15.1	14.7
28	FL CH, i	8.2	8.4
29	PR CH	10.8	10.8
30	PR PR, i	12.5	12.8
31	PR PR	12.0	12.1
32	FL PR	14.6	15.0
33	PR PR	17.6	17.1
34	PR CH	11.0	10.9
35	PR PR	12.1	12.1
36	PR CH, i	9.5	9.7

37	FL FL	10.4	10.5
38	FL CH	9.6	9.8
39	PR PR, i	12.3	12.5
40	PR PR	7.9	8.3
41	PR CH	22.3	21.8
42	PR PR	11.2	11.3
43	FL CH	9.9	10.0
44	PR PR	9.7	9.8
45	FL CH	8.0	7.9
46	PR PR	11.0	11.0
48	PR PR	20.0	19.9
49	PR CH, i	17.6	17.8
50	PR PR, i	11.2	10.6
51	FL PR, i	14.5	14.4
52	FL CH	12.7	12.7
53	FL PR	12.9	13.0
54	PR CH	9.8	10.0
55	PR PR	9.6	9.6
56	FL CH	16.1	16.6
57	PR PR	11.1	10.8
58	PR CH, i	9.5	9.6
59	PR PR	20.4	22.3
60	PR PR	7.8	8.5
61	FL CH	10.8	10.7
62	CH CH	10.6	11.1
Average		**12.1**	**12.2**
Stdev		**4.0**	**4.0**
%RSD		**33.2**	**32.8**
Subsample Average		**12.2**	**12.1**
FL = Flat, PR = Prismatic, CH = Chipped or Broken, i = inclusion *** excluded from average, stdev**			

Table 3: Comparison with other studies

This study: 108 grains, CT pixel 4-5 μm				
	Avg 3D/2D	**1σ**	**% Diff Avg**	**1σ**
Volume	0.82	0.22	23	16
Surface area	0.81	0.14	22	9
ESR	1.02	0.06	5	5
Mass	0.82	0.22	23	16
Ft	1.01	0.02	2	2
Evans et al., 2008: 4 grains, CT pixel 3.77 μm				
	Avg 3D/2D	**1σ**	**% Diff Avg**	**1σ**
Volume	0.68	0.09	32	9
Surface area	0.77	0.08	23	8
ESR	-	-	-	-
Mass	0.66	0.08	34	8
Ft	0.93	0.10	7	10
Glotzbach et al., 2019: 24 grains, CT pixel 1.2 μm				
	Avg 3D/2D	**1σ**	**% Diff Avg**	**1σ**
Volume	1.04	0.2	15	13
Surface area	1.12	0.17	16	14
ESR	0.93	0.06	8	5
Mass	-	-	-	-
Ft	0.99	0.02	2	2