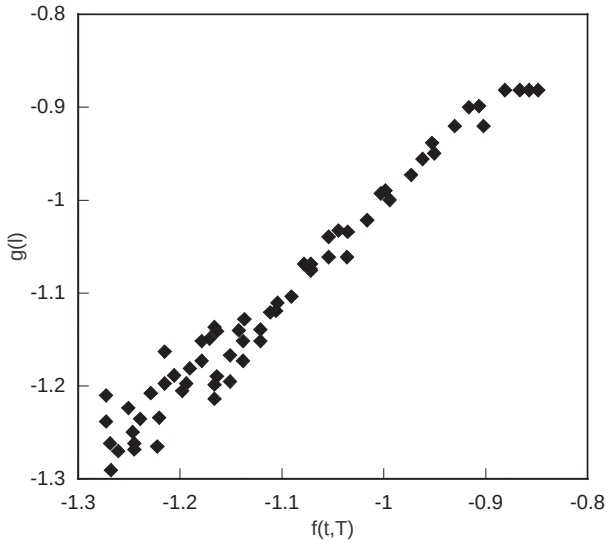
APPENDIX A: FITTING METHODS FOR EMPIRICAL ANNEALING EQUATIONS

As discussed in the text, initial studies that modeled fission-track length annealing data used the least-likelihood formulation provided by Box and Cox (1964), which was derived under the assumption of constant variance after the transform function g is applied. This assumption is inappropriate as a general technique for fitting apatite annealing data, as is clearly demonstrated by a reassessment of the fluorapatite data set of Crowley et al. (1991). Figure A1 shows the relationship between f and g in their final fitted model. There is considerably more scatter in points with low g values, which correspond to experiments with r values near to 1. This is to be expected, because at low levels of annealing a very wide range of times and temperatures can only be differentiated by a narrow range of r values (0.95-1.0) in which the experimental errors span a large fraction of that range. As a result, there is an inescapable lack of definition in this area of time-temperature space, and the r values contain relatively little information. Thus, even though experimental r values near 1 tend to have the smallest measurement errors in and of themselves, they provide the least resolution of model parameters. Laslett and Galbraith (1996) realized these shortcomings, and derived a customized log-likelihood approach that is well-suited for fission-track annealing data.

In order to take errors into account explicitly, we used a χ^2 merit function to describe the misfit between f and g:

$$\chi^2 = \sum \left\{ \left[g(r) - f(t,T) \right] / \sigma_{[g(r)-f(t,T)]} \right\}^2 . \quad \text{(A1)}$$



APPENDIX FIGURE 1. Relation between $f(t,T)$ and $g(l)$ as modeled by Crowley et al. (1991). The Box and Cox (1964) log-likelihood function used to fit the data assumes that after the transform function $g(l)$ is applied, error variance is constant. In fact, there is a higher degree of error in the experiments with long track lengths, as relatively large intervals in experimental temperature produce small changes in track length. As a result, use of the Box and Cox (1964) function for fitting fission-track data is inadvisable.

The error σ can then be estimated from the various experimental errors, which are combined using the standard equation for error propagation (e.g., Bevington and Robinson, 1992, equation 3.14). If we take $g(r)$ as in equation (6), then the error due to imprecision in the measurement of length is:

$$\sigma_{g(r)} \cong \sigma_r r^{\beta-1} \left[\left(1 - r^{\beta}\right) / \beta \right]^{\alpha-1} \quad \text{(A2)}$$

where σ_r is the reduced standard error of the mean track length. It is also possible to cast the function in terms of r instead of $g(r)$. Such an approach would have the advantage of being able to include zero-length experiments among the data being fitted. However, in this case numerical fitting techniques become much more unstable, and require hard-wired intervention to avoid computer errors from exponentiation of negative or very large numbers. Additionally, the error of a zero-length experiment is unknown or zero; either choice complicates incorporating the measurement errors of non-zero points in the analysis.

It is also appropriate to take into account errors in experimental temperature determination, insofar as their magnitude can be comparable to or exceed that of measurement errors. The one-sigma temperature error estimated by Carlson et al. (1999, this volume) is 1 °C. The magnitude of the estimated uncertainty in fission-track length can be observed by looking at a sequence of experiments for RN apatite at 100 hours (Table A1). At low levels of annealing, the natural spread in fission-track lengths is far larger than the estimated effect of a 1 °C variation in annealing conditions. However, at higher levels of annealing the errors become comparable. At the highest levels of annealing the errors remain comparable for mean length, while for modeled *c*-axis projected lengths, which have had most of the effects of length anisotropy removed, the estimated error due to temperature predominates.

Errors in experiment duration may also be applicable in some cases. We estimated the 1-sigma error in experiment duration to be 150 s. This primarily represents the uncertainty in the time required for the furnace to regain its original temperature after the apatite sample was introduced. In general, it took from 120 to 300 s after adding the sample before the temperature had come to within 2 °C of its intended value, after which we considered it to be "on temperature". After this point it generally took an additional 120 to 600 s for the final temperature to be reached. This error is small in even 1 hour runs, but if shorter runs are used (e.g., 20 min in Green et al. 1986) the uncertainty may become a significant factor.

Because of the acceleration of annealing, both temperature

APPENDIX TABLE 1. Uncertainty in length measurement vs. effect of uncertainty in temperature conditions, apatite RN, annealing time 100 hours

	mean lengths			*c*-axis-projected lengths		
T (°C)	l_m (μm)	σ_{meas} (μm)	σ_T (μm)	$l_{c,mod}$ (μm)	σ_{meas} (μm)	σ_T (μm)
101	15.89	0.07	0.005	16.30	0.05	0.003
150	15.62	0.07	0.01	16.11	0.04	0.008
202	14.76	0.07	0.03	15.55	0.04	0.02
250	12.11	0.09	0.09	13.69	0.04	0.05
275	8.71	0.26	0.27	11.96	0.05	0.09

T = Experiment temperature. σ_{meas} = Estimated measurement error. σ_T = Estimated effect on length measurement of a 1 °C temperature variation.

and time excursions have an inversely proportional relationship to measured mean or *c*-axis projected length: the shorter the length, the larger the effect of a given temperature or time excursion. The between-experiment variance component used by Laslett and Galbraith (1996, Eq. 8), which they found to be adequately approximated as being independent of the mean, does not agree with this observation.

To incorporate errors in time and temperature determination, the estimated errors must be multiplied by the appropriate derivative of the function $f(t,T)$. For example, errors in temperature determination for a fanning curvilinear model (Equation (4)) are included using the factor:

$$\sigma_T^2 \left\{ \frac{C_1\left[1\mathrm{n}(t) - C_2\right]}{T\left[1\mathrm{n}\left(1/T\right) - C_3\right]^2} \right\}^2, \qquad \text{(A3)}$$

where σ_T is the standard deviation that describes the uncertainty in the temperature determination, t and T are the duration and temperature of the isothermal annealing experiment, and C_1, C_2 and C_3 are the parameters that are being fitted.

In order to help maintain stability in the numerical solver, the fanning curvilinear model was reparameterized using a method analogous to that described by Laslett and Galbraith (1996, Equation 12). Instead of solving directly for C_2 and C_3, which are far outside the vicinity of the data, we used the substitute parameters:

$$b_1 = \frac{12.794 - C_2}{-6.2918 - C_3}, \qquad \text{(A4a)}$$

$$b_2 = \frac{12.794 - C_2}{-6.0334 - C_3} \qquad \text{(A4b)}$$

which correspond to the contour line slopes at 100 hours and 267 °C and 144 °C, respectively.

In this paper we report reduced χ^2 values (χ_υ^2), which are equal to χ^2 divided by the degrees of freedom. The expectation for an appropriate model with well-estimated errors is a χ_υ^2 value of 1. However, this value is seldom achieved in the work presented here. There are a number of reasons for this failure. First, there are undoubtedly sources of error that are not taken into account, including possible errors in etching and undetected systematic errors. The fact that fits of annealing models of *c*-axis projected data produce χ_υ^2 values that are larger than those using mean-length data suggests that the *c*-axis-projection model may have an associated error that we were not able to detect and quantify. In multiple-apatite fits, the asymmetric nature of errors in the parameters r_{mr0} and κ make their inclusion in a χ^2-minimization scheme problematic, although they are certainly non-negligible. Similarly, accounted-for errors, such as uncertainty in temperature determinations, may be incorrectly approximated. Second, there may be aspects of the physical processes underlying the data that the empirical models employed are incapable of reproducing, introducing further errors. Laslett and Galbraith (1996) correctly point out that, as currently measured, unannealed initial spontaneous track lengths are not the correct normalizing quantity (although, as discussed in the text, the correct normalizing value is unknown, and their proposed solution is also subject to argument).

While we view the shortcomings listed above as areas with potential for improvement, we do not consider them debilitating. The primary goal of error propagation in this study was to determine the appropriate weighting for each experiment when finding best-fit parameters, not to provide a complete accounting of the errors.

Because of the apparent effect that Cf-irradiation has on mean lengths, particularly in highly annealed experiments, we did not use any analyses that employed that technique when fitting mean-length models. The primary way that Cf-irradiation affects mean length measurements is apparently to alter the relative frequency of observation of different angular populations. Cf-irradiation effects on modeled *c*-axis projected lengths are thus much smaller, as the influence of anisotropy is removed. We thus included these experiments in fitting modeled *c*-axis projected lengths.

In some cases there were apatites for which the measured mean or *c*-axis projected length of an annealed experiment were slightly longer than for the "unannealed" determination (run 0). In these cases, the longest measurement was used as the normalizing factor, rather than the run 0 measurement.

Two different variations of the simplex method were used to find the best-fitting models. The downhill simplex method described by Press et al. (1988) was used for fitting apatite-apatite pairs and for finding close-fitting models for Monte Carlo estimation of errors. To ensure that the solution found could not be improved, the algorithm was re-executed using the previous solution as one of the points; convergence was not accepted until the routine was called four times without an improvement in the solution. An implementation of the controlled random search algorithm (Willett 1997) was provided to R. Donelick by S. Willett (personal communication) and re-coded and adapted by R. Ketcham to fit the annealing models in this paper. Convergence was accepted when all of the models had χ^2 values that were within a factor of 10^{-8} of each other (i.e., when the difference between them is less than one 100-millionth of their magnitude). Bounds on all variables being fitted (Table A2) were set intentionally wide to ensure that the solution space could be traversed freely by the fitting algorithms. Both methods yielded identical solutions when the parameter space was more tightly bounded, but the looser constraints on the annealing equations necessitated the use of the controlled random search algorithm. Monte Carlo error estimates used the fitted models as initial guesses, which in turn permitted the use of the faster downhill simplex method for calculation.

Errors on all fitted parameters were estimated using a Monte Carlo method. For each solution, 1000 "nearby" solutions were calculated by varying each data point by a Gaussian factor of the standard error of the measurement and the estimated stan-

Table A2. Bounding values used for model fitting

	Fanning Linear		Fanning Curvilinear	
Variable	upper	lower	upper	lower
C_0	20.0	-50.0	0.0	-3000.0
C_1	0.01	0.0	60.0	0.0
C_2	0.0	-50.0	60.0	30.0
C_3	0.002	0.0	60.0	30.0
α	1.0	-1.0	1.0	-1.0
β	10.0	-35.0	5.0	-25.0

dard error of the temperature determination. The reported error bounds contain 95% of all such solutions. For the annealing models used here, the parameters α and β are tightly linked to C_0 and C_1, such that variations in one pair often can be compensated for very closely by variations in the other. Consequently, when all of them were allowed to vary in the Monte Carlo models the result was error bounds that were so large as to be uninformative, often larger than the magnitude of the parameters themselves. We thus opted to follow the convention used by Crowley et al. (1991) and hold α and β constant at the values obtained in the original solution. Even with this measure, the reported errors are not independent of each other, as changes in some parameters can be compensated by changes in others (e.g., Laslett and Galbraith 1996).

RESULTS

The result of using a χ^2 fit of the F-apatite data of Crowley et al. (1991) is illustrated in Tables 1 and 2. Predicted index temperatures are substantially lower, because greater emphasis is placed on fitting the more highly annealed points, rather than the relatively unannealed points that represent the majority of the data.

It should be noted that the choice of merit function does not have as large an effect on analysis of the Durango apatite data of Green et al. (1986), probably because that data set is much more balanced in terms of the relative amounts of annealing among the experiments. For example, the Green et al. (1986) data set has 4% of its experiments with reduced lengths of >0.975, and 13% with >0.95, whereas the Crowley et al. (1991) data set has 11% with >0.975 and 31% with >0.95. This overabundance of experiments with very low levels of annealing, brought about because of the choice to use regular temperature intervals in the experiments, makes the Crowley et al. (1991) data set more susceptible to problems caused by overweighting of these points. The data set presented by Carlson et al. (1999, this volume) was also constructed based on constant temperature intervals.

The statistical approach used here differs conceptually in a number of ways from the method of Laslett and Galbraith (1996). First, it uses measured length errors on an experiment-by-experiment basis, whereas the Laslett and Galbraith (1996) analysis used a single fitted function to characterize the average length error as a function of length. Second, their treatment takes length measurements of zero into account through the introduction of a probability function that predicts a diminishing probability of seeing tracks in an experiment at high levels of annealing, even when tracks may in fact be present. This is an interesting concept, but it appears to assign a constant value to what may be an operator-specific quantity: how thoroughly does the investigator search for tracks in highly annealed experiments? This probability is also certainly affected by Cf-irradiation, as employed in many cases in this study, which helps to reveal highly shortened tracks. In the approach used here, experiments that produce total annealing can instead be used as a check for the fits based on non-zero data; if the fits predict non-zero track lengths where only zero lengths are observed, it may indicate an error, whereas if there is no contradiction, the veracity of the model is supported and no information is lost.

Fitting multi-apatite models

Fits that included multiple types of simultaneously annealed apatites utilized the above procedures and assumed constant values of r_{mr0} and κ. We note that there are two shortcomings to this approach. First, uncertainties in r_{mr0} and κ for many of the apatites studied represent potentially substantial components of error. As a result the uncertainty in the index temperatures of some apatites (primarily the highest-resistance ones) are underreported. Exclusion of these error terms is probably also partly responsible for inflated χ_υ^2 values. Second, proper calculation of χ_υ^2 should distinguish between errors that affect all apatites in a single annealing run identically, such as temperature excursions, and those that affect each apatite separately, such as measurement uncertainty. Separation of components of error is complicated by the circumstance that different experimental errors will have different effects on measured mean lengths. For example, a given temperature excursion will cause a different change in annealing for each apatite being studied. While more rigorous statistical treatments that overcome these difficulties would be preferable, the level of agreement between the single-apatite models and the multiple-apatite ones suggests that the inaccuracies introduced are minor.

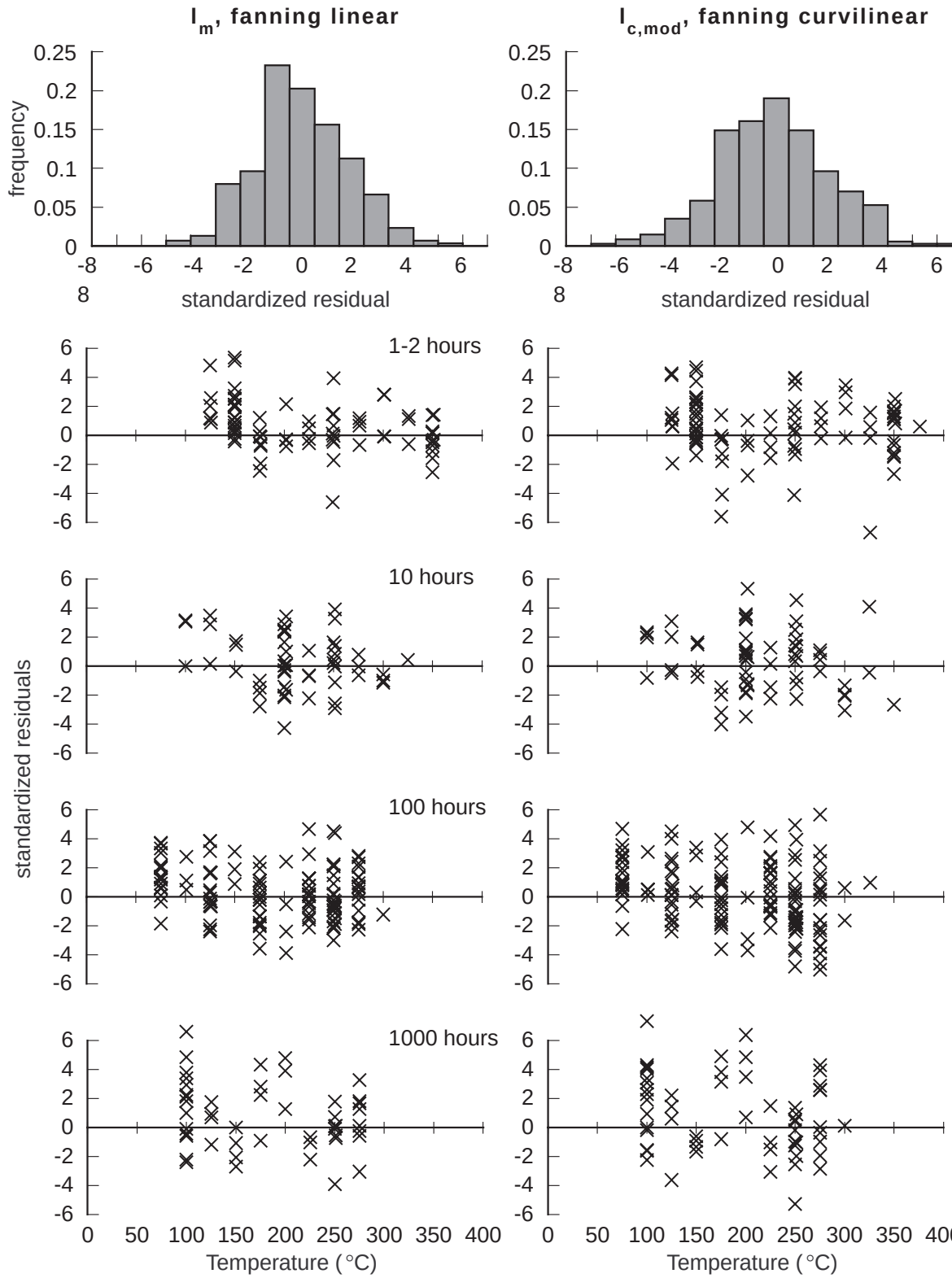
Residuals

Standardized residuals for two multiple-apatite models are shown in Figure A2. Standardized residuals are calculated here using the equation:

$$\frac{r_1 - \hat{r}_i}{\sqrt{\sigma_{\text{meas}}^2 + \sigma_T^2}}, \qquad \text{(A5)}$$

where r_i is a reduced length measurement, r_i is the reduced length predicted by the model, σ_{meas} is the standard error of the length measurement, and σ_T is the length error caused by uncertainty in the temperature determination. The left column shows residuals for the fanning linear model of mean rack lengths, and the right column shows residuals for the fanning curvilinear model of *c*-axis projected track lengths.

The upper graph in each column shows the distribution of standardized residuals; each appears to be a rough but non-ideal normal distribution. The subsequent graphs show the change in residuals as a function of temperature for the various annealing intervals. The primary obvious trend is that the low-temperature residuals for many of the classes appear preferentially positive. This apparent bias is primarily caused by the maximum *r* value being 1.0, and the normalizing values being close to many of the other measurements from low-temperature annealing runs. Because the models cannot predict *r* values above 1.0, substantial divergences must have positive residuals. If the lower two temperatures from each time interval are disregarded, there appears to be little structure in the remaining residuals.



APPENDIX FIGURE 2. Standardized residuals for multiple apatite models: the fanning linear fit of mean lengths in the left column, and the fanning curvilinear fit of *c*-axis-projected lengths in the right column. Top-most graphs show histograms of overall distribution of residuals, while lower graphs show residuals vs. temperature for various annealing times.

APPENDIX B: APATITE-APATITE FITTING

The most robust way of fitting Equation 8 to a set of paired apatite experiments in which one is less resistant to annealing than the other is to minimize a straightforward χ^2 formulation:

$$\chi^2 = \sum_i \left(\frac{r_{lr,i} - \hat{r}_{lr,i}}{\sigma_i} \right)^2 = \sum_i \left[\frac{r_{lr,i} - \left(\frac{r_{mr,i} - r_{mr0}}{1 - r_{mr0}} \right)^{\kappa}}{\sigma_i} \right]^2 \quad \text{(B1)}$$

where $r_{lr,i}$ and $r_{mr,i}$ are the reduced lengths of the less-resistant and more-resistant apatite, respectively, for a single experiment *i*. The standard deviation can be estimated from the individual length measurements:

$$\sigma_i^2 = \sigma_{r_{lr,i}}^2 + \sigma_{r_{mr,i}}^2 \left[\left(\frac{\kappa}{1 - r_{mr0}} \right) \left(\frac{r_{mr,i} - r_{mr0}}{1 - r_{mr0}} \right)^{\kappa - 1} \right]^2 . \quad \text{(B2)}$$

Because all apatites have undergone exactly the same high-temperature annealing history, there is no need for error terms to account for uncertainties in experimental conditions.

When a suite of more than two apatites is being studied, it is more general to assume that neither apatite in a random pair is necessarily the most resistant one for the entire system. In this case, any two apatites can be compared by mapping each one back to the most resistant apatite:

$$\chi_{1,2}^2 = \sum_1 \left(\frac{\hat{r}_{mr,1,i} - \hat{r}_{mr,2,i}}{\sigma_{1,2,i}} \right)^2 \quad \text{(B3)}$$

where $\hat{r}_{mr,1,i}$ is the estimated reduced length of the most resistant apatite for the conditions in experiment *i* based on the reduced length of the first apatite:

$$\hat{r}_{mr,1,i} = r_{1,i}^{1/\kappa_1} \left(1 - r_{mr0,1} \right) + r_{mr,0,1} \quad \text{(B4)}$$

where $r_{1,i}$ is the reduced length of the first apatite in experiment *i*, and $r_{mr0,1}$ and κ_1 are the equation (8) parameters for the first apatite. The estimate based on the second apatite is analogous. The error term is then:

$$\sigma_{1,2,i}^2 \cong \sigma_{r_{lr,1,i}}^2 \left[\frac{1}{\kappa_1} r_{1,i}^{1/\kappa_1 - 1} \left(1 - r_{mr0,1} \right) \right]^2 + \sigma_{r_{lr,2,i}}^2 \left[\frac{1}{\kappa_2} r_{2,i}^{1/\kappa_2 - 1} \left(1 - r_{mr0,2} \right) \right]. \quad \text{(B5)}$$

For the most resistant apatite in the suite, r_{mr0} and k are held constant at 0 and 1, respectively. Because of the inverse exponent, this method is less stable for k values near 0, which can be encountered during the fitting process; however, in practice the minimum κ value observed has always been >0.1, permitting us the reasonable step of restricting k values to be >0.05.

This formulation has the advantage of allowing any two apatites from a suite to be evaluated against each other. Thus, to find the most internally consistent set of r_{mr0}-κ parameters to characterize a suite, χ^2 is calculated by comparing each apatite to each other apatite:

$$\chi^2 = \sum_k \sum_{l \neq k} \chi_{k,l}^2 . \quad \text{(B6)}$$

This method also makes it possible to use experimental runs that do not encompass the entire suite of apatites, as long as each apatite can be linked to the others through some common experiments. For example, in the Carlson et al. (1999, this volume) data set, all experiments for all 15 apatites can be used for a fit of this type because of the 11 experiments they all had in common, even though four apatites had up to 60 more experiments than the others. Fitting the entire suite of 408 experiments with what amounts to an equation with 28 parameters is a computationally challenging task. In addition to the fitting methods described in Appendix A, we also employed Powell's method as implemented in Press et al. (1988), with minor modifications. Monte Carlo estimation of parameter errors used only 100 simulations because of the computational time required.

Paired apatite relations were also examined to see if they are able to provide an improved estimate of the initial track length. This was done using the equation:

$$r_{lr} \delta_{lr} = \left(\frac{r_{mr} \delta_{mr} - r_{mr0}}{1 - r_{mr0}} \right)^{\kappa} \quad \text{(B7)}$$

where δ_{lr} and δ_{mr} are correction factors for the reduced length that are allowed to vary in the range 0.7–1.0, in effect permitting the initial track length for each apatite to be revised upward by up to 30%, or to about 21.5 µm. Table B1 shows the results obtained from fitting both mean and modeled *c*-axis projected lengths using a simultaneous solution. Whereas allowing the initial length to vary evidently results in a statistically better fit to the data, the fact that the range of implied adjustments to the initial length (4–14%) far exceeds that seen in the experimental data collected thus far leads us to conclude that this approach is not advisable.

REFERENCES CITED

Bevington, P.R. and Robinson, D.K. (1992) Data Reduction and Error Analysis for the Physical Sciences, 328 p. McGraw-Hill, New York.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vettering, W.T. (1988) Numerical Recipes in C, 735 p. Cambridge University Press, Cambridge.

Other references cited with paper Ketcham et al. 1999, 1235–1255.

TABLE B1. Results of fitting r_{mr0}-κ parameters while allowing initial track length to vary

	l_m (χ_u^2 = 1.89)			$l_{c,mod}$ (χ_u^2 = 2.28)		
apatite	r_{mr0}	κ	δ	r_{mr0}	κ	δ
AY	0.6889	0.3218	0.9026	0.7158	0.1830	0.8838
B2	0.0000	1.0000	0.9269	0.0000	1.0000	0.8682
B3	0.4444	0.6524	0.9064	0.6754	0.2186	0.8959
DR	0.7041	0.2732	0.9137	0.7208	0.1615	0.8969
FC	0.5938	0.3888	0.9219	0.6903	0.1725	0.9111
HS	0.7435	0.3391	0.8673	0.7461	0.1990	0.8649
KP	0.5832	0.3614	0.9323	0.6932	0.1529	0.9213
OL	0.7219	0.3128	0.9001	0.7287	0.1885	0.8788
PC	0.5930	0.2729	0.9622	0.6661	0.1538	0.9369
PQ	0.7036	0.3832	0.8733	0.7199	0.2017	0.8719
RN	0.7434	0.2622	0.9057	0.7357	0.1785	0.8780
SC	0.7402	0.2754	0.8946	0.7299	0.1903	0.8714
TI	0.4308	0.5535	0.9337	0.6482	0.2008	0.9197
UN	0.7149	0.3405	0.8879	0.7266	0.1915	0.8765
WK	0.7177	0.3244	0.8878	0.7275	0.1852	0.8764