Earth-Science Reviews xxx (xxxx) xxxx

Contents lists available at ScienceDirect

Earth-Science Reviews

journal homepage: www.elsevier.com/locate/earscirev

Discussion

Comment on the reply to the Comment on "Thermal history modelling: HeFTy vs. QTQt" by Vermeesch and Tian, Earth-Science Reviews (2014), 139, 279–290

Kerry Gallagher[a,*], Richard A. Ketcham[b]

[a] Géosciences Rennes/OSUR, Université de Rennes 1, France
[b] Jackson School of Geosciences, University of Texas at Austin, USA

In the reply of Vermeesch and Tian (2018, hereafter referred to as VT2) to our comment Gallagher and Ketcham (2018; hereafter referred to as GK), they graciously open by expressing appreciation for some of our contributions to thermochronology. We were disappointed, however, that their reply does not acknowledge the clarifications and corrections we presented, nor does it address our more direct criticisms of the errors and misconceptions in their original article. Instead, the reply consolidates the mistakes of the original, cites selectively and provides harmful or contradictory recommendations. We thus see a need for another brief round of clarification. The organisation of our second comment follows the structure of VT2. We refer to the original paper of Vermeesch and Tian (2014) as VT1.

1. VT2 introduction

The statement that GK essentially rephrased the main points of VT1 is false and dismissive. It misconstrues our efforts to show where we agree, and misses the fact that many points made by VT1 without citation have been made previously (e.g. Ketcham, 2005, p 309–311).

The next comment concerns assessment of model fits in HeFTy and QTQt. First, in their point (a) they restate that HeFTy struggles to fit large data sets, but GK demonstrated this is not the case, if a rational and informed approach is taken to dealing with the data.

Their point (b) states that QTQt requires comparison of the model predictions with the observed data. Assessment of any model result should include this, and such assessments can be made statistically and graphically. Nonetheless, VT1 failed to do just that in when assessing the results of fitting an inappropriate order polynomial and thermal history models inferred from inconsistent thermochronology data. In both cases, nothing sophisticated is required to decide that the predictions do not agree with the data. Furthermore, VT2 cites Gallagher (2016) to support their point (a), but the latter does not mention model results from HeFTy. In fact, Gallagher (2016) cited the polynomial example of VT2 as the example of bad practice when not considering the quality of the data fit.

Having been regularly used in presentations by KG for perhaps 10 years, the quote from George Box was not repeated, but cited in a more complete form appropriate to the issue in hand, given we are interested in assessing how useful inversion generated models are.

2. VT2 HeFTy

The primary reason HeFTy sometimes "struggles" is not due to its statistics, but simply that it uses a non-learning Monte Carlo algorithm. If HeFTy utilized a search algorithm, such as Markov Chain Monte Carlo as in QTQt, or Constrained Random Search (Willett, 1997) as in HeFTy's predecessor AFTSolve (Ketcham et al., 2000), it would find solutions more quickly, but at a cost. The design choice to remove the latter functionality when creating HeFTy stemmed from RAK's observation that it tended to collapse into the best-fitting paths while insufficiently exploring the range of other solutions consistent with the data, leading to an improperly rosy picture of the data's resolving power.

GK demonstrated how using HeFTy sensibly with minimal constraints could find thermal histories consistent with even their full-sized data set in fairly short order (as described in the text, ignored in the VT2 response), and how cautious (and far from "complicated") optimization can improve run times. GK also provided advice on how constraints should be defined purposefully, rather than arbitrarily.

VT2's critique of the "tiny" constraint boxes as non-geological reflects another misunderstanding. The simple model embodies the reasonable assumption that the apatite grains were, at some time prior to their measured ages, above their closure temperatures, and then followed some path to the present day. The fact that there is some time after the initial constraint where the thermal history is not constrained by the data is properly represented by wide confidence intervals, making the precise size and positioning of that box unimportant so long as it is not absurd.

We strongly disagree with VT2's claim that their approach to creating constraints is "sensible"; we cannot identify any sense in intentionally arbitrary program input, and GK demonstrated its

* Corresponding author.
E-mail address: kerry.gallagher@univ-rennes1.fr (K. Gallagher).

https://doi.org/10.1016/j.earscirev.2019.102878
Received 27 February 2018; Received in revised form 10 November 2018

shortcomings. Among our motivations for this second comment is to reinforce our warning to readers against using this approach.

VT2 conclude this section criticising HeFTy's "so-called" algorithms as not being used elsewhere in capital-S "Science", and thus lacking "bona fide" credentials. Both Monte Carlo algorithms and frequentist use of *p*-values to define confidence intervals are widespread, error bars being an obvious example of the latter. HeFTy's approach is a variant on and descendant of methods described by Willett (1997). The use of goodness of fit metrics for selecting acceptable models and contouring results is another design choice, in this case based on the opinion that, when one cannot find a model that fits the data to within uncertainties, the appropriate thing to do is revisit one's data and assumptions. A specific intention of HeFTy's design is to nudge users toward measuring and including information that is still frequently omitted from apatite fission-track analyses, such as kinetic indicators and length calibrations. As demonstrated in GK, HeFTy's approach also provides intuitive feedback on the benefits of increasing sample sizes.

3. VT2 QTQt

In contrast to VT2, we suggest that fitting the data is not missing the point. It is surely a desirable objective of an inverse modelling approach. The tightly constrained solutions shown in VT1 do not fit the data, neither for the appropriate polynomial solution, nor for thermal histories using inconsistent data. Therefore these are not appropriate solutions but VT2 gloss over this aspect as if irrelevant. We return to the issue of fitting data below.

The typographical error in Ketcham et al. (2007) in the formula for the projected lengths has an insignificant effect on the projected length distributions for the data from fig. 8 in VT1 (Fig. 1a,b). So it is far from obvious that this is a good example of QTQt finding tightly constrained solutions for an inappropriate model with a bug that remained "undetected for 5 years".[1] Any effect is most significant for short lengths at high angles to the c-axis. While short tracks are important for the inference of thermal histories, they are relatively rare. The distributions of thermal histories (Fig. 1c,d) inferred with the data in Fig. 1 differ considerably more for the data sets with different numbers of track lengths than different projection calculations. So let us ignore all the natural uncertainties associated with real fission track data and annealing models (e.g. Ketcham et al., 2009, 2015), and perhaps our colleagues could provide a real data example where this difference in calculating 100 or more projected lengths is actually important rather than merely making an unsupported statement that it is.

The second point concerning problems with marginal distributions is similarly unjustified and misleading. VT2 state "*a tremendous amount of information is lost*", without specifying what the tremendous amount of information actually is. We can guess perhaps it concerns the true thermal history. In fact the relevant information (which is not lost) is that contained in the data concerning the ability to recover the true thermal history. In this case, the ability of the data to provide precise details of multiple reheating-cooling episodes is limited. The algorithms implemented in QTQt try to exploit information in the data to infer the thermal history. In particular this information is used to assess simpler thermal histories relative to complex ones. The sampling adopted in the version of QTQt used in VT2 will explicitly prefer the simpler model over a more complex one if they both fit the data to more or less the same degree.

The apparent lack of resolution of a true model can also be explained in the context of fitting a polynomial $y = f(x)$, relevant to the examples in Figs. 2–6 from VT1. If the coefficients of the higher order terms are small enough, they have minimal influence on the predicted values, depending on the range of the variable x where predictions are sought. In that case, we would prefer polynomials of lower order than the true solution, but we would still fit the observed data.

The results of modelling synthetic data are poorly presented in VT2 and have been edited relative to the default output of QTQt, but we return to that below. However, if we assume the predicted length distributions are for the two thermal histories in Fig. 1 of VT2, then it is difficult to be convinced that the predictions are significantly different in terms of their agreement with the input distribution. Although these data were generated with the specified thermal history, there is not enough information even in these near perfect data to recover the true thermal history with high probability, i.e. it is not well resolved. Fig. 14 of Green and Duddy (2012), cited in the conclusions of VT2, clearly demonstrates this behaviour with a set of forward models of varying complexity, stating explicitly:

> "*The fission-track parameters resulting from these histories are indistinguishable, despite the time of cooling below 110° C varying from 250 Ma to 1000 Ma, and for histories ranging from progressive cooling to various more complex heating and cooling histories*".

This is not referred in the context of Fig. 1 in VT2. Additionally, VT2 does not comment on the quality of the data fit, despite citing Gallagher (2016) in the introduction.

When data can be satisfied with a simple model, it is surely important and desirable to know this. While more complex models will be also consistent with the data, additional complexity needs to be justified independently. For example, Gallagher (2012), Fig. 2) provides an example of the effect of trying to fit data as well as possible. This leads to complex structure in the inferred thermal history that is not in the true solution at all. For these reasons, the sampling algorithms in the version of QTQt used in VT2 are such that more complex models that do not improve the data fit are not overly sampled.

Relative to the default format of QTQt, the output represented in Fig. 1 of VT2 has been edited. In particular, the credible intervals, the expected and maximum posterior models having been removed, reducing our ability to assess the results. To reassess this result, we generated synthetic data (20 single grain ages, and 100 projected track lengths) with QTQt using the thermal history in VT2, and used these data with uniform proposal functions in QTQt 5.6.0 (the version currently available on the SourceSup website dating from January 2017). A summary of the relevant results are shown in Fig. 2. We can see the maximum posterior model is similar to the simpler model in VT2[2] and predicts the length distribution and age pretty much as well as the maximum likelihood model. The predicted mean length and the length distribution for the expected model replicate the observed data well, although the predicted age is a little older that we might be happy with, given the error for the synthetic data central age is 3.85 m.y. The expected model implies changes in the rate of cooling at the 2 times of maximum temperature for the two reheating events specified by VT2, demonstrating there is some information on these events in the accepted model distribution. Similarly, the credible intervals, not included in Fig. 1 of VT2, show the thermal history is relatively better constrained at those times.

The thermal histories described above can be explained by considering a subset of the sampled thermal histories, an option available in version 5.6.0 of QTQt. These are shown in Fig. 1c and d, the thermal histories being colour coded relative to the maximum likelihood and maximum posterior probability values. Fig. 1c highlights there are many two event reheating-cooling models that give slightly higher likelihoods (better fits to the data). The range in timing of these events captures those in the true model, but the time of maximum temperature

[1] In fact it would be more like 9 years at the time of writing.

[2] VT refer to one thermal history as the expected model but it seems to be the maximum mode model. The expected model should always be relatively smooth and should not lie along the peaks of the marginal distribution when this is asymmetrical at any given time, as is generally the case in their Fig. 1. The peaks of the marginal distribution define the maximum mode distribution.

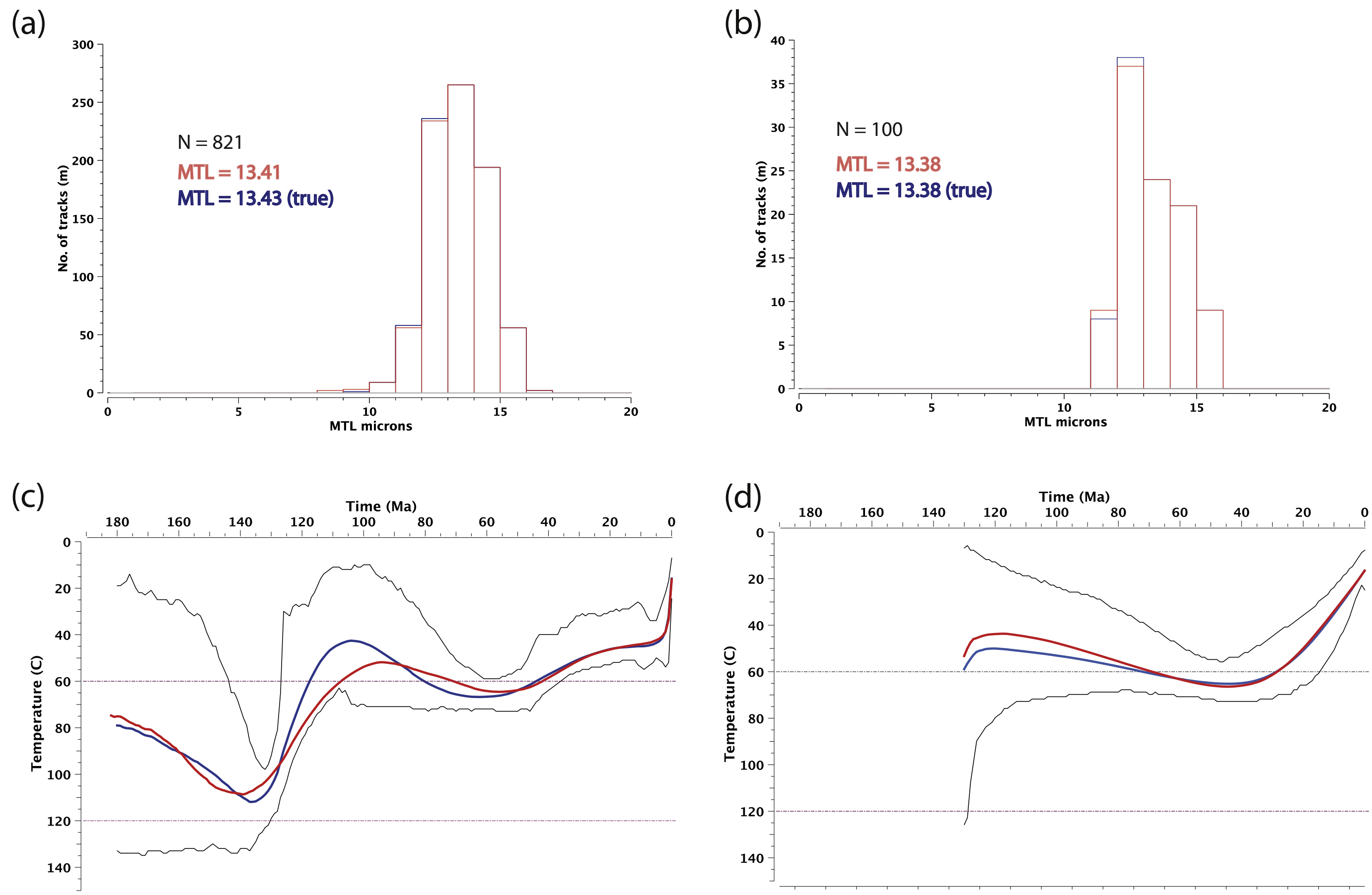

Fig. 1. (a) Projected length distributions using the 821 track length data set from fig. 8 in VT1. The red histograms are produced using the erroneous projection equation while the blue histograms represent the correct equation.
(b) As (a) but using the 100 track length data set in VT1.
(c) The expected thermal histories using the two sets of projected track length data from (a), and the count data from VT1 (corrected for the typographical error mentioned in GK). The same colour coding as (a) applies. For clarity we show only the 95% credible bounds for the blue curve. Note the thermal histories starting in the partial annealing zone (approximately defined by the two horizontal lines) implies rapid cooling from above total annealing temperatures just before the start of the thermal history as illustrated.
(d) As (c) but for the data in (b). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

for the younger event being less well resolved that the earlier event. Also, we can see that the rate of cooling-reheating prior to the second event is quite variable, and not really constrained at all as fission track data are not sensitive to the magnitude of cooling if a reheating event occurs subsequently. This is also clearly demonstrated in fig. 13 of Green and Duddy (2012).

Now consider Fig. 1d, in which the same group of models as 1c is used, but now colour coded in terms of posterior probability, relative to the maximum value equated to 100%. In the Bayesian approach implemented in QTQt, acceptance of proposed thermal histories is strongly based around the posterior probability, which can be thought of as a combination of the data fit and the model complexity. We see these higher posterior probability models fall into a relatively well defined band similar to non-blue parts of the marginal distribution, and which contain the maximum posterior and maximum mode model in Fig. 1a. Relative to Fig. 1c, many models with high likelihoods translate to lower relative posterior probability as a result of their complexity and so lower prior probability. This behaviour is briefly explained in the appendix.

In the final lines concerning QTQt, there are statements in VT2 concerning the marginal distributions that give "*false confidence in this over-simplified thermal history*", reflecting "*the tight clustering of the simple models which graphically dominate the more accurate but less precise models*". Here more accurate is being taken to mean closer to the true solution (which we never know in practice), and less precise presumably meaning more scattered, less well resolved or with greater uncertainty. As mentioned above, if simpler models can explain the data adequately they are preferred. Any graphical dominance then comes from the fact that proportionally more simple models are represented in the final distribution. This distribution will be more restricted than that containing all models, irrespectively of their complexity, that can fit the data. It is worth emphasising again that the credible intervals reflect the preference for simpler models that can fit the data adequately, rather than the sampling of all possible (including very complex) models that could fit the data similarly. In this sense, confidence may be false if we forget that the credible intervals are conditional on preferring simpler models.

As we never know the true solution in practice, the motivation for preferring simpler models to more complex ones, provided both can explain the data, is that we have a lower bound on the complexity of the thermal history. In the example of VT2, this could also be expressed as: the "true thermal history" is not ruled out, but the potential for recovering it is low as there is relatively little information in the data about the details. Other sampling algorithms, such as the Metropolis algorithm which uses just the likelihood ratio to accept or reject models, will always accept more complex models if they fit the data equally or better than a simpler model. The distribution of accepted models in this case will generally be broader than that in Fig. 1, but will

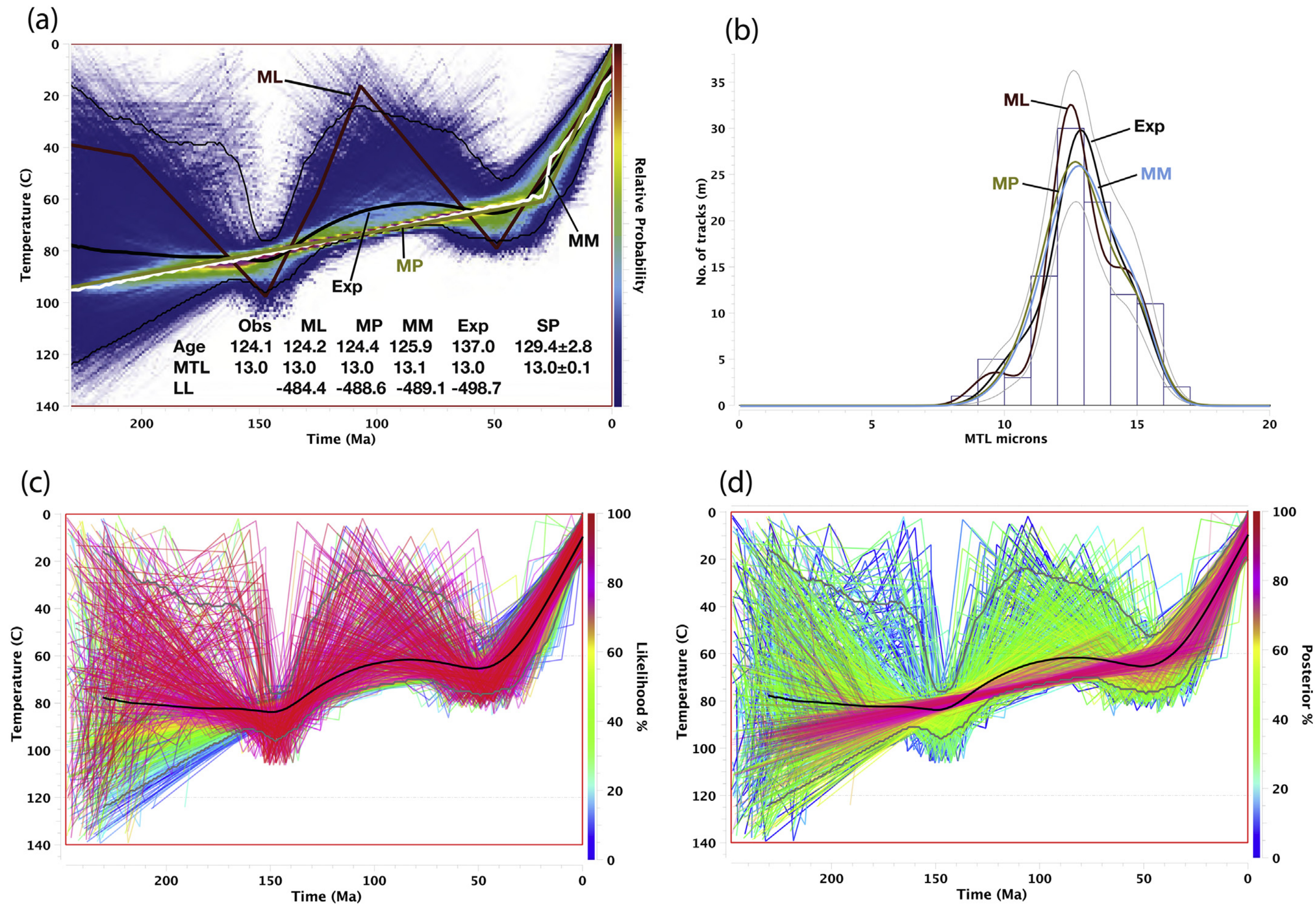

Fig. 2. (a) Output from QTQt showing the maximum likelihood (ML), maximum posterior (MP), maximum mode (MM) and expected (Exp) models, together with the 95% credible interval (the region between thinner black lines) and the marginal posterior distribution. The maximum likelihood solution is that which has the highest likelihood, equivalent to the thermal history that best fits the data. The maximum posterior solution is the thermal history that has the maximum posterior probability, and is sometimes considered the best model in the Bayesian sense. The posterior probability combines the likelihoods and prior probabilities for each model, attempting to balance fitting the data with model complexity. The maximum mode solution is constructed at 1 m.y. intervals by running along the peak of the marginal distribution, while the expected model is the average of the marginal distribution. Consequently, these two do not represent individual thermal histories sampled during the inversion process, but summaries of the posterior distribution.
We also show the values of the observed fission track age and mean track length and their predictions for each of the 4 individual models. SP represents sampled predicted, which summarises the mean and standard deviation of the predicted ages and mean track lengths for all accepted models (from a total of 100,000 post-burn-in iterations, after 200,000 burn-in). LL is the log likelihood for each of the 2 models.
(b) The predicted length distributions for the 4 models described for Fig. 1a, while the 2 grey lines define are the 95% credible interval on the predicted distributions for 100,000 models.
(c) Representative thermal histories from the posterior distribution, colour coded by the relative likelihood, calibrated against the maximum likelihood with a value of 100%. The black line is the expected thermal history model with the 95% credible interval defined by the two grey lines. The higher likelihood heating-cooling thermal histories are relatively complex and the timings of reheating events are dispersed, as are the pre-reheating temperatures. Simpler models similar to the maximum posterior model in 1a, also have relatively high likelihoods. Most of the accepted models fall into one of these two groups, effectively defining two local maxima in the likelihood.
(d) Representative thermal histories from the posterior distribution, colour coded by the relative posterior, calibrated against the maximum posterior probability with a value of 100%. The black line is the expected thermal history model with the 95% credible interval defined by the two grey lines. In comparison with Fig. 1c, we see the simpler models define the higher posterior probability, while the complex models have moderate to low posterior probability. This reflects limited information in the data on the form of the true thermal history.

contain many models with structure unconstrained by the data. The question then is how to choose a preferred thermal history and which parts of it are actually constrained by the observations.

When using QTQt, the four individual thermal history models shown in Fig. 2a, together with the credible intervals and the marginal distributions, allows a user explicitly to assess the consistency of the results. Here we note that only the maximum likelihood and maximum posterior models are sampled as part of the inversion process. The expected and maximum mode models are constructed from the population of accepted models.

When the 4 models are similar and the data are adequately predicted, then we can assume that we have solutions that are coherent and well constrained (but always subject to all the uncertainties in the predictive models for annealing and diffusion, and in the data themselves). When these individual models are different, then the results need to examined more closely. In some cases, such as Fig. 1 of VT2 and Fig. 2 here, we see there are effectively two groups of solutions. Typically, these are represented by the general form of the maximum likelihood (with multiple reheating events) and a group of simpler solutions, often represented by the form of the simpler maximum posterior model. The two groups of models can generally predict the data more or less equally well. However, depending on the relative proportions of these two groups in the overall population, the expected model, being the average of all accepted models, can fall between the two groups. Similarly, the maximum mode model, representing the peak of the marginal distributions at typically 1 m.y. intervals, can oscillate between the two groups. In such cases, the predictions for the expected and mode models can be relatively poor, as they do not represent either group well.

Overall, it is recommended to show at least the 4 individual solutions described above, together with the credible intervals and the marginal posterior, to give the most complete summary of the solutions. In general cases, it may be that we are happy to recognise there are multiple groups of possible solutions. In this case, independent information can help in the choice of model. It is up to the user to choose a particular solution, if so desired, but this choice should be justified in the geological context.

Finally, it is not clear quite what is implied by the phrase "inherited from Sambridge et al. (2006)". Firstly, Sambridge et al. (2006) show distributions of the predictions for different dimensional polynomial models (known as posterior predictive densities), rather than marginal distributions on the actual model parameters. These are very different things. However, in an earlier paper, Stephenson et al. (2006) do show marginal distributions on model parameters in 2D (their Figs. 1 and 4). Moreover, similar approaches had been used in geophysics for some time (Malinverno and Briggs, 2004; Mosegaard and Tarantola, 1995, 2002). The approach of extracting marginal density distributions is general, having well established analytical solutions (e.g. Lee, 1989, as used in fig. 6 of GK). The sampling approach to estimate marginal distributions in 1-D goes back to at least Gelfand and Smith (1990) and the 2-D case follows from this.

4. VT2 conclusions

We agree it is important to examine (visually and statistically) the data before any modelling exercise is undertaken. However, this is followed by a statement that contradicts what is implied in the section on QTQt. Arriving at the conclusions, one paragraph later, it now seems that a typical thermochronological data set[3] "*does not contain enough information to reliably constrain all but the simplest thermal histories*". Yet, just one paragraph earlier, recovering these simple thermal histories was deemed to be misleading and imparting false confidence.

VT2 continues with a statement that QTQt supports only crude V shaped thermal histories, yet the earlier section criticises QTQt for producing linear cooling histories. Hopefully, any modelling approach will produce thermal histories that are consistent with the observed data, but will always be conditional on the model assumptions. These may be variably complex, but if the data are consistent with a simple V-shaped or linear thermal history, this is what we would hope to obtain with QTQt. As mentioned previously, this may not be as complex as the true thermal history, but if the data do not justify more complexity, what would we expect?

It is not clear to us why thermal history modelling with more data is "*hampered by the non-uniqueness of forward models*", presumably meaning different thermal histories can predict much the same thing. In fact, more data may well increase the possibility of recovering a more complex thermal history, but this will depend on how much independent information is provided by additional data. A corollary of this is that inconsistent data may lead to simple thermal histories but poor data fits, as increasing complexity can not resolve the inconsistency. For example, data from multiple samples and/or multiple thermochronometers may lead to conflicting interpretations depending on how the data are dealt with. Again looking at how well the data are predicted by resulting thermal history models can help identify if such problems exist.

Finally, VT2 states that estimating unconstrained thermal histories may not always be the right question. We were surprised to read this, given that VT1 explicitly proposes that constraints should not be used with QTQt (section 5 VT1, *"we would urge the user to refrain from using this facility"*). According to VT1, the rationale for not using constraints is that QTQt cannot be used to disprove geological constraints. If the constraints do not allow thermal histories that can explain the data then surely that should be enough proof the constraints are not valid (or there is a problem with the data or predictive models for annealing or diffusion). However, if the constraints do not contradict the observed data, then all we know is that the data are consistent with the constraints, but not that they prove them. Assuming here that unconstrained means adopting a general prior in QTQt and using no additional (geologically based) constraints, asking for the form of such unconstrained thermal histories is just one question that should be asked. As explained above, QTQt should provide thermal histories with simple structure and any additional complexity needs to be justified. If reliable geological constraints are available, they should be incorporated, and perhaps even less reliable constraints can be incorporated. In both cases, however, the significance of their incorporation or not should be explored as part of the modelling process. This is perhaps best achieved by a combination of inverse model results and forward modelling (e.g. Ketcham, 2005; Cogné et al., 2012), incorporating appropriate geological information. Irrespective of the modelling approach and assumptions, the results should always be assessed. As in the case with data, this may be addressed both statistically and by visual inspection.

5. Conclusion

We agree that there are differences between the algorithms and outputs of HeFTy and QTQt, and that the thermochronological community may benefit from a well-constructed independent comparison and assessment. In our opinion, VT1 and VT2 fall short of that. However these contributions do serve to reinforce the fact that modelling software should not be treated as a black box and the results should not be blindly accepted at face value.

Appendix

Why simpler models are preferred over more complex ones in transdimensional Bayesian Markov chain Monte Carlo

As shown in Fig. 1 of Sambridge et al. (2006), which is actually inherited (from MacKay, 2003), more complex models have the potential to make a wider range of predictions relative to a simpler model. That is the potential posterior predictive data distribution is wider and so flatter, or lower amplitude (as probability distributions must integrate to 1 by definition). Then, if we consider a complex model and a simple model that both explain the observed data to the same degree, the posterior probability on the more complex model is lower than the simpler one in the region around the actual observed data. This means that, given similar likelihoods or data fits, simpler models will be naturally preferred over more complex ones.

Another way to consider this is that in Bayes' theorem, the posterior probability is proportional to product of the prior and the likelihood. If we consider uniform independent priors on time and temperature, given as

[3] Actually more of just a typical apatite fission track data set.

$$p(t) = \frac{1}{(t_{max} - t_{min})} = \frac{1}{\Delta t}$$

and

$p(T) = \frac{1}{(T_{max} - T_{min})} = \frac{1}{\Delta T}$ so the prior probability for a model with n time-temperature points is then given as

$$p(m) = \prod_{i=1}^{n} p(t)p(T) = \left(\frac{1}{\Delta t \Delta T}\right)^n$$

and the prior term will become smaller as n increases, or as the model has more time-temperature points.

Now consider two thermal history models, m_s (simple, with ns time-temperature points) and m_c (complex with nc time-temperature points, and $nc > ns$) with likelihoods, (or fit the data), given as $L(m_s)$ and $L(m_c)$. The ratio of the posterior probability, given the observed data, d_{obs}, is given from Bayes' theorem as

$$\frac{p(m_c|\ d_{obs})}{p(m_s|\ d_{obs})} = \frac{p(m_c)L(m_c)}{p(m_s)\,L(m_s)}$$
$$= \left(\frac{1}{\Delta t\Delta T}\right)^{nc}\left(\frac{1}{\Delta t\Delta T}\right)^{-ns}\frac{L(m_c)}{L(m_s)}$$
$$= \left(\frac{1}{\Delta t\Delta T}\right)^{\Delta n}\frac{L(m_c)}{L(m_s)}$$

where $\Delta n \geq 1$ is the difference in number of time temperature points between the complex and the simpler models. If we take $\Delta t = 300$ Ma, and $\Delta T = 140$ °C, then the value of the prior for 1 time-temperature point is 2.4×10^{-5}. Therefore to be sure of accepting a model with an extra time-temperature point ($\Delta n = 1$), the likelihood needs to increase by at least this much, relative to the simpler model. Note that the calculations are generally made in using the logarithm of all the terms above, so this corresponds to an increase in log likelihood of about 10.65. Alternatively, if the likelihoods of the two models are equal, this implies a probability of accepting the complex model over the simple one of 2.4×10^{-5}, i.e. rarely.

In practice, there are additional terms, such as the proposal functions, that are considered for the acceptance criterion in the transdimensional Markov chain Monte Carlo algorithm implemented in QTQt. These modify the probability of accepting more complex models relative to the description above, but the basic principle is the same. This also demonstrates the well known dependence of the behaviour of any Bayesian approach to the choice of prior. In simple terms, the wider (or weaker) the prior, the more likely we will accept simpler models. This is also dependent on the nature of the data, and the information contained in the data concerning the thermal history. Adding many constraints is equivalent to very strong prior information and can restrict the sampler of a wider model space defined by the general prior of QTQt Therefore, it is always worthwhile exploring the sensitivity of the results to the choice of prior and constraints.

References

Cogné, N., Gallagher, K., Cobbold, P., Riccomini, C., Gautheron, C., 2012. Post-break-up tectonics in southeastern Brazil from thermochronological data and combined forward-inverse thermal history modelling modelling. J. Geophys. Res. 117, B11413. https://doi.org/10.1029/2012jb009340.

Gallagher, K., 2012. Transdimensional inverse thermal history modelling for quantitative thermochronology. J. Geophys. Res. 117, B02408. https://doi.org/10.1029/2011JB00882.

Gallagher, K., 2016. Comment on 'a reporting protocol for thermochronologic modeling illustrated with data from the Grand Canyon' by Flowers, Farley and Ketcham. Earth Planet. Sci. Lett. 441, 211–212.

Gallagher, K., Ketcham, R.A., 2018. Comment on "thermal history modelling: HeFTy vs. QTQt" by Vermeesch and Tian, Earth-Science Reviews (2014), 139, 279–290. Earth Sci. Rev. 176, 387–394.

Gelfand, A.E., Smith, A.F.M., 1990. Sampling-based approaches to calculating marginal densities. J. Am. Stat. Assoc. 85, 398–409.

Green, P.F., Duddy, I.R., 2012. Thermal History Reconstruction in Sedimentary Basins using Apatite Fission-Track Analysis and Related Techniques SEPM Special Publication No. 103. pp. 65–104.

Ketcham, R.A., 2005. Forward and inverse modelling of low-temperature thermochronometry data. Rev. Mineral. Geochem. 58, 275–314.

Ketcham, R.A., Donelick, R.A., Donelick, M.B., 2000. AFTSolve: a program for multi-kinetic modeling of apatite fission-track data. Geol. Mater. Res. 2 (1), 1–32.

Ketcham, R.A., Carter, A., Donelick, R.A., Hurford, A.J., 2007. Improved measurement of fission-track annealing in apatite using C-axis projection. Am. Mineral. 92, 789–798.

Ketcham, R.A., Donelick, R.A., Balestrieri, M.L., Zattin, M., 2009. Reproducibility of apatite fission-track length data and thermal history reconstruction. Earth Planet. Sci. Lett. 284, 504–515.

Ketcham, R.A., Carter, A., Hurford, A.J., 2015. Inter-laboratory comparison of fission track confined length and etch figure measurements in apatite. Am. Mineral. 100, 1452–1468.

Lee, P.M., 1989. Bayesian Statistics: An Introduction. Edward Arnold.

MacKay, D.J.C., 2003. Information Theory, Inference, and Learning Algorithms. C.U.P, Cambridge.

Malinverno, A., Briggs, V.A., 2004. Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes. Geophysics 69, 1005–1016. https://doi.org/10.1190/1.1778243.

Mosegaard, K., Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems. J. Geophys. Res. 100 (B7), 12,431–12,447.

Mosegaard, K., Tarantola, A., 2002. Probabilistic approach to inverse problems. Int. Geophys. Ser. 81 (A), 237–268.

Sambridge, M., Gallagher, K., Jackson, A., Rickwood, P., 2006. Trans-dimensional inverse problems, Model Comparison and the evidence. Geophys. J. Int. 167, 528–542.

Stephenson, J., Gallagher, K., Holmes, C., 2006. Low temperature thermochronology and modelling strategies for multiple samples 2 : Partition modeling for 2D and 3D distributions with discontinuities. Earth Planet Sci. Letts. 241, 557–570.

Vermeesch, P., Tian, Y., 2014. Thermal history modelling : HeFTy vs. QTQt. Earth Sci. Rev. 139, 279–290.

Vermeesch, P., Tian, Y., 2018. Reply to comment "thermal history modelling : HeFTy vs. QTQt". Earth Sci. Rev. 176, 395–396.

Willett, S.D., 1997. Inverse modeling of annealing of fission tracks in apatite 1: a controlled random search method. Am. J. Sci. 297, 939–969.