Earth Surf. Dynam. Discuss.,
https://doi.org/10.5194/esurf-2020-59-AC9, 2021

Earth Surface Dynamics Discussions

Interactive comment on "Bias and error in modelling thermochronometric data: resolving a potential increase in Plio-Pleistocene erosion rate" *by* Sean D. Willett et al.

Sean D. Willett et al.

swillett@erdw.ethz.ch

Received and published: 6 January 2021

Anonymous Referee #1

This manuscript seems to have been written as a detailed rebuttal of the conclusions of the paper by Schildgen et al. (2018) which questioned the accuracy of the linearised inversion method of Fox et al. (2014) of thermochronological data. Specifically Schildgen et al. (2018) suggested that the methodology encouraged the incorrect interpretation of spatial variability as temporal variability, and thereby creating a systematic bias towards an apparent temporal acceleration in exhumation. No doubt the authors are motivated by the fact that the earlier paper, published in Nature I see, brings into ques-

tion much published work by the present authors. My overall reaction is that the present manuscript contains some useful information in that it sets out all of the resolution and uncertainty issues associated with the method of Fox et al. (2014).

The present manuscript quantitatively describes through theory and example the sources of uncertainty and bias in the Fox et al. (2014) approach, including those arising from theory errors (due to forward modelling) data errors and the limited resolution caused by data sampling and the physics of the data-model relationship. Again this is largely in the context of combatting assertions by Schildgen et al. (2018), nevertheless I think this exposition is in principal useful to have in the literature. At the same time the manuscript is excessively long (my version is 78 pages) and an arduous and at times repetitive read in parts. I question how many readers will understand it all in its present form. There is a good argument to move a lot of this material to an appendix or supplementary material without loss of the main points. I realize that the authors are probably intent on getting this full detail into the literature, but I think having it all in the main body of a paper will certainly diminish its likely impact. I suggest the Bayesian Inference tutorial material could be moved to an Appendix or left out and the results cited as they are required. In addition do we really need to see all numerical tests? again use of appendices would help. Much of the underlying theory of Bayesian inference appears in the literature cited, but there are also many books outside of the Earth Sciences literature, admittedly two popular ones, Menke and Tarantola are cited here and so perhaps better use of specific results in those books could be used here.

RESPONSE: This is a fair assessment of both content and motivation. We will give this section of the paper a hard edit. We will shift some of the models to an appendix, but prefer not to move the theory out of the main paper as it is specific to the Fox et al methodology and therefore essential to the points of the paper. Also, it isn't that long. We have consolidated some of the models, for example combining 3 figures into 1, where the resolution and variance plots were identical. We are less concerned with readability and impact as thoroughness. As any perusal of the comments shows, there is a great deal of misunderstanding and lack of familiarity with concepts that are perhaps well established in geophysics, but not in thermochronometry. If we were writing for a different community, we would not include the inference description or discussion of error sources, but for this paper and this target audience we prefer to lose casual readers put off by the length, but help interested readers understand better how parameter inference is used and what the various analysis tools are, in order to better understand the outcome, how to interpret it and to avoid the sort of "black-box" approach we too often see with modeling tools that enter the community without full awareness of limitations.

Some of the writing is clumsy, an example is on line 251 '...an inversion scheme based on inverse problem theory which minimizes Gaussian errors in observations, and a Bayesian parameter model in which model parameters are assumed to have Gaussian distributions about a prior value'. Some more conciseness would be a help here and in many other places. Note that in Bayesian inference we don't actually 'Minimize Gaussian errors in observations' rather a Likelihood function p(d | m) expresses the probability of observing the data given a model. It may or may not take a Gaussian form, indicating all departures from fitting the data perfectly are expected to be described by a Gaussian PDF. Its maximization occurs in Maximum Likelihood techniques not Bayesian Inference. One might consider Maximizing the Posterior PDF as one way of characterising that PDF but this differs from Likelihood maximisation when the prior is not uniform. Of course when the data model relationship is linear and Likelihood and Prior are Gaussian then the posterior PDF is also Gaussian.

RESPONSE: We clarified the text at this point. We have developed the method, largely assuming Gaussian distributions where maximum likelihood methods and Bayesian Methods are effectively identical, but it is correct that this is not true in general and we should state it correctly.

While I have not studied the Fox et al. (2014) paper in any detail it is clear from the account here that it is a linearised Bayesian inversion method combining a prior with

a Likelihood to estimate a Gaussian approximated posterior probability distribution. This approach is well known and understood, especially in the geophysical literature for many decades. The works of Tarantola and Valette being a prime embodiment of the approach. The present manuscript gives a summary of the linearised theoretical results as they pertain to quantification of resolution and model covariance matrices and their interpretation. It appears to me to be a competent summary of linearised inversion within a Bayesian framework, my comments above notwithstanding. I do feel that some references to the latest editions of some well established books such as Menke (2012, rather than 1984) or Aster, Borchers and Thurber (2018, 3rd edition) could assist the reader in this regard, as all of the relevant theory is set out there. The paper references relevant to linear discrete parameter estimation, or which this is an example, are, correctly, to the original works from the 1970s-1980s which is good to See.

RESPONSE: We have updated some references, including the Menke book.

From L430 onward the authors' analyse the Normalised erosion rate metric, NR, eqn. (14) (lines 425-475). This is a nonlinear function of the model parameters recovered in an inversion. I note that there is a detailed critique of this term, as it was important to Schildgen et al. (2018) to support some of their arguments. A discussion is given of Expectation of a ratio (Line 440). This expectation identity relies on the terms x and x being independent which they are clearly not in eqn. (14), since both involve e1 and e2. There must be a cross term affecting the Expectation of theFrom L430 onward ratio which is ignored here. This is an error.

RESPONSE: This is a good point - we concur. In fact this paragraph is not really needed as we don't use this result for anything, so we will simply remove it as part of the shortening and editing process.

The authors continue on and obtain an approximation of the variance of NR using a Taylor expansion, as well as discussing its bias and its complicated behaviour. Quantities such as variance and expected value of some property of the model parameters while analytical, are just moments of what is in reality a probability distribution over NR constrained by the data. Since the point of Bayesian inversion is to determine a Posterior probability distribution of the model parameters then by drawing from that (Gaussian) distribution and calculating NR provides a posterior ensemble of NR values from which a complete (non Gaussian) PDF can be inferred. A better (more complete) approach would be to map out that PDF in full rather than drawing inferences from estimates such as variance, which are at best only partially characterise its behaviour. How well can a variance describe a high skewed distribution anyway? As I understand it this is the objective of the authors, i.e. trying to quantify the uncertainty and bias in the NR term, in which case it is better to examine the actual distribution, if its accessible, and I think here it must surely be. The PDFs give the full picture, low order moments such as variance and expectation do not.

RESPONSE: We agree that this could be done - it would require numerical sampling of posterior distributions given the covariances, and would be limited in that we have only Gaussian posteriors, i.e. only moments, as the inversion is analytical, but it would solve the inverse Gausssian distribution problems. However, the purpose of this section is mostly to argue that one should not use ratio quantities such as the NR. We hope that the case is made and we will not see this quantity in the literature again, in which case a more sophisticated analysis of the error is not needed, so we prefer to not put in additional work for an analysis tool that we hope will now disappear.

The authors are correct in their assertions (L1305-1310) that the Bayesian solution will tend to go to the prior when data influence is poor, which means have large data covariances. This is well known and a fundamental property of Bayesian inference. However it is not the case that the posterior will tend to the prior when data, or averaged data (in this case), are inconsistent. It is not entirely clear to me that this is what the authors are suggesting, but I got the impression it might be implied by their discussion of resolution and the influence of the prior (multiple sections across the manuscript).

This seems to be a pertinent point in their rebuttal of the Schildgen results. Inconsistent data (e.g. tight constraints on a subset of model parameters which do not agree) is quite a different thing from weak data constraints (e.g. due to large data noise). Only the weak constraint case does the posterior tend to the prior, not the inconsistent data case.

A toy example makes this clear. Consider a single prior variable with Gaussian uncertainty. Imagine that new data are a direct observation of this variable. The weak data case is when measurements occur with large errors relative to the Gaussian prior. They will have little affect on the prior mean, i.e. posterior tends to the prior. The inconsistent data case is when measurements with errors comparable to the Gaussian prior are widely separated. Here the posterior mean may be far from the prior mean. As I say its not entirely clear to me whether the authors have mistook one for the other in their arguments, but even if they haven't, I think it best to explicitly clarify this difference in the manuscript so that there is no confusion in the readers mind.

RESPONSE: We agree with this point. However, we are almost always referring to the sparse data case, not the inconsistent data case. This is most important for the numerous cases of very sparse data where one wants to differentiate between bias to the prior and interpolation of very distal data. The inconsistent data case could come up, but this is more the "how are data averaged" question as discussed in our Figure 2. We will add a comment to this effect and discuss how data residuals can be used to test for data-model consistency.

How is the closure temperature estimation from Dodson (1973) applicable here? Isn't this a rather approximate method in itself? How relevant is it to low temperature thermochronological data such as fission track and He? As I understand it there is a continuous response, in terms of annealing of diffusion over a range of temperatures, which would be a significant proportion (say 50%) of the estimated closure temperature. In this case the measured age is not directly related to the time that the system became closed (i.e. started to accumulate fission tracks or He). Shouldn't one ideally solve the full FT annealing and He diffusion equations to find when and at what depth the system became closed? This does not appear to be done here. No doubt if it were this would create a more nonlinear inversion problem which may be more difficult to handle with the linear/linearized inversion approaches used here? These issues should probably be discussed and justified.

RESPONSE: This is an important point for those cases where there is a complex cooling history, or where the first-order kinetics are a less-than-optimal approximation (fission track annealing). We present this as a potential model error (lines 405-410, line 593), and this is discussed in earlier publication (Fox et al. 2014), but to do more would require an entire new study and paper, particularly for those aspects that destroy the linearity of the GLIDE model. The same is true for the other model error, temperature, where multidimensional aspects of advection or mantle heat flow differences are not included. We expect this is even more important than the kinetics, and acknowledge this problem under model errors, but investigation would require much new work.

Ultimately any inverse model result is conditional on the assumptions (e.g. model resolution matrix depends on the assumed linear relationship between model and data). I agree that using different forward models to generate synthetic data and perform an inversion introduces unquantified 'theoretical errors' in an inversion test. Quantifying such errors however is usually difficult, especially systematic errors. The question of spatial vs temporal variability in inference results, which is at issue as one of the central arguments between the present authors' and those of Schildgen et al. (2018). From what I can tell the original work of Fox et al. collected and smoothed over data sets spatially by introducing a Gaussian prior imposing a spatial (smoothing) covariance. Much is discussed on this point here and whether this introduces bias and encourages spatial variability to be mis-interpreted as temporal variability. The present authors obviously argue not, however its not clear to me why use of global spatial prior smoothing of Fox et al. is even necessary. It seems a rather backward step in itself, compared to more sophisticated 'data adaptive' spatial smoothing approaches popular in Bayesian

methods these days. I'm thinking here of Bayesian spatial sampling known as Partition modelling which has been around for more than twenty years, See Denison et al. (2002 for a summary). That type of approach allows sharp local variations to be recovered if they are supported by the data, e.g. by sampling data across a discontinuity. This is much more preferable than simply imposing global spatial correlation via a prior. I see Partition Modelling has already been applied to in low temperature Thermo-chronology problems by Stephenson et al. (2006) almost than a decade and a half ago. This paper or any related work is not cited or discussed. In that sense I see a lot of energy here expended here on defending the validity of what appears to be a somewhat outdated framework.

RESPONSE: Other approaches are possible, and we have investigated some, including breaking correlation structure (Stalder et al., 2020). The spatial correlation approach is one of the simplest. The question of its utility or whether some new approach is needed depends somewhat on how common discrete fault offsets are in thermochronometry. Our experience is that they are surprisingly rare. Their importance is often exaggerated by the sparsity of data and the fact that there are mapped faults everywhere, so for any sampled ages that vary in space, there is a fault between them, but that does not prove that that fault was active in order to offset the ages. Any fault that moved prior to closure of thermochronometric ages is irrelevant to those ages, so most mapped faults have no effect on kinematics and age offset. This point is often missed in interpretations. The simplest test for the effect of cross-fault averaging is simply to invert suites of ages independently on each side. We did this for the Alps and in the revised manuscript, for the Nanga Parbat example. In each case there was no effect on the result, indicating no cross-fault averageing.

My impression is that the manuscript is detailed defence of the Fox et al. (2014) inversion approach which may be useful for readers to see in some form. However it is unnecessarily long and could probably start at page 35 after an introduction. A manuscript 50% shorter would be more appropriate and I expect have more impact.

RESPONSE: As discussed, we will edit the intro, but as other reviewers and commenters have indicated, there are still issues arising that need explaining and we don't think that our target audience is going to pick up a book on inverse theory and figure it out on their own. If we can provide a readable introduction for a community with less experience in parameter estimation, it could be seen as helpful and we would rather err on the side of completeness, so as to avoid the sort of criticisms that we are getting now.

Minor points:

I can see no content in eqn. (8)? - It appears blank to me.

RESPONSE: Oops - correct - we lost an equation in formatting somewhere.

References: Menke W. (2012) Geophysical Data Analysis: Discrete Inverse Theory, 3rd ed. Aster, R. Borchers B. & Thurber C. (2018). Parameter Estimation and Inverse Problems - 3rd Edition, Elsevier D.G.T. Denison, N.M. Adams, C.C. Holmes, D.J. Hand, Bayesian partition modelling, r Comput. Stat. Data Anal. 38 (4) (2002) 475– 485. Stephenson, J. and Gallagher, K. and Holmes, C. C., Low temperature thermochronology and strategies for multiple samples 2: Partition Modelling for 2D/3D distributions with discontinuites, EPSL, 2006, 241, 557-570.

Interactive comment on Earth Surf. Dynam. Discuss., https://doi.org/10.5194/esurf-2020-59, 2020.