Further discussion of Bias and Error in modelling thermochronometric data by Willett et al.

Peter van der Beek[1], Taylor Schildgen[1,2], Rasmus Thiede[3], Hugh Sinclair[4]
[1]Institute for Geoscience, Potsdam University, Potsdam, 14476, Germany
[2]GFZ German Research Centre for Geosciences, Telegrafenberg, Potsdam, 14473, Germany
[3]Institute for Geoscience, Christian Albrecht University Kiel, Kiel, 24118, Germany
[4]School of Geosciences, The University of Edinburgh, Edinburgh, EH8 9XP, UK

Willett et al. focus their response on two aspects of our initial comment: (1) the role of the different geotherms in the forward and inverse models and (2) the resolution and data density of the synthetic models. We will provide some additional clarification on these aspects below. Before we do so, however, we feel it is important to underline once more that Schildgen et al. (2018) and Willett et al. pursue fundamentally different goals. Schildgen et al. (2018) posed a *geological* question: whether the "worldwide acceleration of mountain erosion under a cooling climate" since 6 Ma inferred by Herman et al. (2013) from thermochronologic datasets was robust. In contrast, Willett et al. aim to identify different sources of error in the GLIDE inversions of thermochronologic ages for erosion rates, thereby posing a *methodological*, model-framed question.

As stated in our initial comment, a detailed analysis of the potential sources of error within the GLIDE inversion procedure used by Herman et al. (2013) was outside the scope of the Schildgen et al. (2018) analysis. Schildgen et al. (2018) did not set out to test whether the model *could* make robust predictions of erosion history, but whether it *had done so* in the Herman et al. (2013) analysis. While Willett et al.'s new analysis is a useful exercise, it does not address the question of whether the model application to real-world data leads to accurate predictions of erosion histories. As Willett et al. state in their response: "*we are attempting to learn something about the method, not trying to simulate the way in which the model is applied in normal circumstances*". Therefore, it is unclear (and Willett et al. do not clarify) how their new analyses validate the Herman et al. (2013) results. In fact, validating the application of such models to natural systems is inherently impossible (Oreskes et al., 1994): even if Willett et al. would have shown that the inversion is capable of retrieving accurate exhumation histories in the conditions they provided (and the reanalysis of some of Willett et al.'s synthetic tests in our initial comment shows that it does not), these results do not address the analysis performed by Herman et al. (2013).

1 Variable geotherm calculation

While the discussion on the choice of thermal boundary conditions in numerical models in Willett et al.'s response is interesting and itself worthy of debate, the issue at stake in this exchange is whether the difference in geotherms as calculated in Pecube and GLIDE causes the spurious accelerations in the synthetic models of Schildgen et al. (2018) and our initial comment. Before we go into this, one point needs to be corrected: in their response, Willett et al. appear to imply that Pecube systematically uses a "steady-state" geotherm to calculate thermochronologic ages, whereas GLIDE would use a "transient" geotherm. However, all of the synthetic models that are being discussed here, whether they were designed by Schildgen et al. (2018) or Willett et al., have temporally constant exhumation rates and *should* therefore *tend towards* steady-state geotherms, in both Pecube and GLIDE. The difference between the models lies in the steady-state shape of the geotherm and the time taken to reach steady state (see Willett et al.'s Fig. 3).

Willett et al. dismiss the synthetic models in Schildgen et al. (2018) and in our initial comment because of the difference in boundary conditions when calculating the geotherm, writing: "*all error in the van der Beek et al. comment and Schildgen et al. (2018) models is due to these geotherm errors*". But Willett et al. provide no support for the assertion that differences in boundary conditions produce spurious accelerations, neither in their manuscript nor in the response to our comment. Nor do they address why spurious accelerations are present in inversions using the high-density synthetic data generated by Willett et al., where there are no differences in geotherm calculation between the forward and the inverse models. We reiterate the observations on which we based our rebuttal of Willett et al.'s dismissal:

- Figure 1 in our initial comment shows that when the inferred cause of accelerations in our synthetic models (what we term the spatial correlation bias) is removed, the models do not predict accelerating erosion rates in the last few Myr, despite the difference in geotherm calculation between the forward and inverse models;
- Figure 3 in our initial comment shows that Willett et al.'s models with a transient geotherm (but ages calculated with the same thermal boundary conditions in the forward and inverse model) show spurious accelerations in regions of spatially variable exhumation rates, either at earlier times than shown in Willett et al. or in the last few Myr when the high-temperature ages are removed;
- Figure 4 in our initial comment shows that Willett et al.'s models with a *fixed* geotherm similarly produce spurious accelerations in regions of spatially variable exhumation rates.

Our conclusion from these tests is thus the opposite of what is suggested by Willett et al.: the geotherm errors cannot explain the spurious accelerations observed in these synthetic tests. We provided some tentative explanations for why this may be the case, but we did not argue that one or the other model provides a "*more accurate*" solution to the geotherm problem; we acknowledge that the thermal boundary conditions in both Pecube and GLIDE are imperfect.

2 Resolution and Data Density

In their response, Willett et al. frequently invoke "high" versus "low" resolution and data density, but without quantifying these classifications. The first argument in Willett et al.'s response is that the "*commonality of ages*" between the Alpine dataset and their "high-resolution" synthetic datasets A and D makes the results from these synthetic tests applicable to the Alps. We have already refuted this argument in our original comment (lines 210-222).

Second, Willett et al. argue that looking at earlier time-steps of the model is unwarranted, writing: "*Because these timesteps are not the target of the experiment, we did not generate ages over these intervals, they are poorly resolved and there is a correspondingly larger resolution error*". This statement is incorrect: Dataset D includes ~30 mica $^{40}Ar/^{39}Ar$ ages within the 8-10 Ma time-window (Fig. 5 of the Willett et al. manuscript, or Fig. 1 of the response). Moreover, we could have taken the 10-12 Ma time-window as our starting point, which contains >50 mica $^{40}Ar/^{39}Ar$ ages in both datasets A and D. The predicted erosion rates are lower in the 10-12

Ma time-window than in the 8-10 Ma time-window (leading to larger accelerations when compared to later time-windows) but the resolution is similar (Figure 1).

Third, Willett et al. argue that by removing the hypothetical mica $^{40}Ar/^{39}Ar$ data from the datasets A and D, we have "*converted a high-density model into a low-density model and thereby reintroduced resolution errors*". We note, however, that (1) since all the synthetic mica $^{40}Ar/^{39}Ar$ ages are >8 Ma, the synthetic tests in which they are removed contain exactly the same number of data within the critical 6-0 Ma time-windows as the original tests; (2) as already noted in our initial comment, the resolution values for the recent time-windows are nearly identical between the two tests (compare panels a-c and g-i in Figs. 3 and 4 of our initial comment). Given the similarity of data and resolution values within the critical time-windows, we do not understand on what grounds Willett et al. characterize one model as "high-density" and the other as "low-density".

As a final point, we provide some clarification of why Willett et al.'s models that use the true solution as the prior (Figs. 8 and 13 of the Willett et al. manuscript) are irrelevant. In Bayesian theory, if the posterior probability distribution of the parameter values (i.e., in this case, the predicted spatial-temporal distribution of erosion rates) is equal to the prior probability distribution (i.e., the input prior erosion rates), then the available data have not contributed to constraining the solution. In this case, the prior erosion rates are equal to the true erosion rates, which have been used to predict the synthetic data. Therefore, for each point in space and time, the model's "initial guess" (the prior), is correct and perfectly reproduces the data. The data cannot influence this problem, and the only thing we learn from this test is that there are no obvious errors in the code. This is the explanation for why the errors disappear. From a practical viewpoint, this test is irrelevant because we cannot know *a-priori* what the exhumation history of a particular region is (i.e., there is no way to know the prior, it is necessarily a guess) and if we did, there would be no point collecting data in such a region because we wouldn't learn anything new from that data.

References

Herman, F., Seward, D., Valla, P. G., Carter, A., Kohn, B., Willett, S. D. and Ehlers, T. A.: Worldwide acceleration of mountain erosion under a cooling climate, Nature, 504, 423–426, doi: 10.1038/nature12877, 2013.

Oreskes, N., Shrader-Frechette, K. and Belitz, K.: Verification, validation, and confirmation of numerical models in the earth sciences, Science, 263, 641–646, doi: 10.1126/science.263.5147.641, 1994.

Schildgen, T. F., van der Beek, P. A., Sinclair, H. D. and Thiede, R. C.: Spatial correlation bias in late-Cenozoic erosion histories derived from thermochronology, Nature, 559, 89–93, doi: 10.1038/s41586-018-0260-6, 2018.

Willett, S.D., Herman, F., Fox, M., Stalder, N., Ehlers, T.A., Jiao, R., Yang, R.: Bias and error in modelling thermochronometric data: resolving a potential increase in Plio-Pleistocene erosion rate. Earth Surf. Dyn. Discussions, doi: 10.5194/esurf-2020-59.

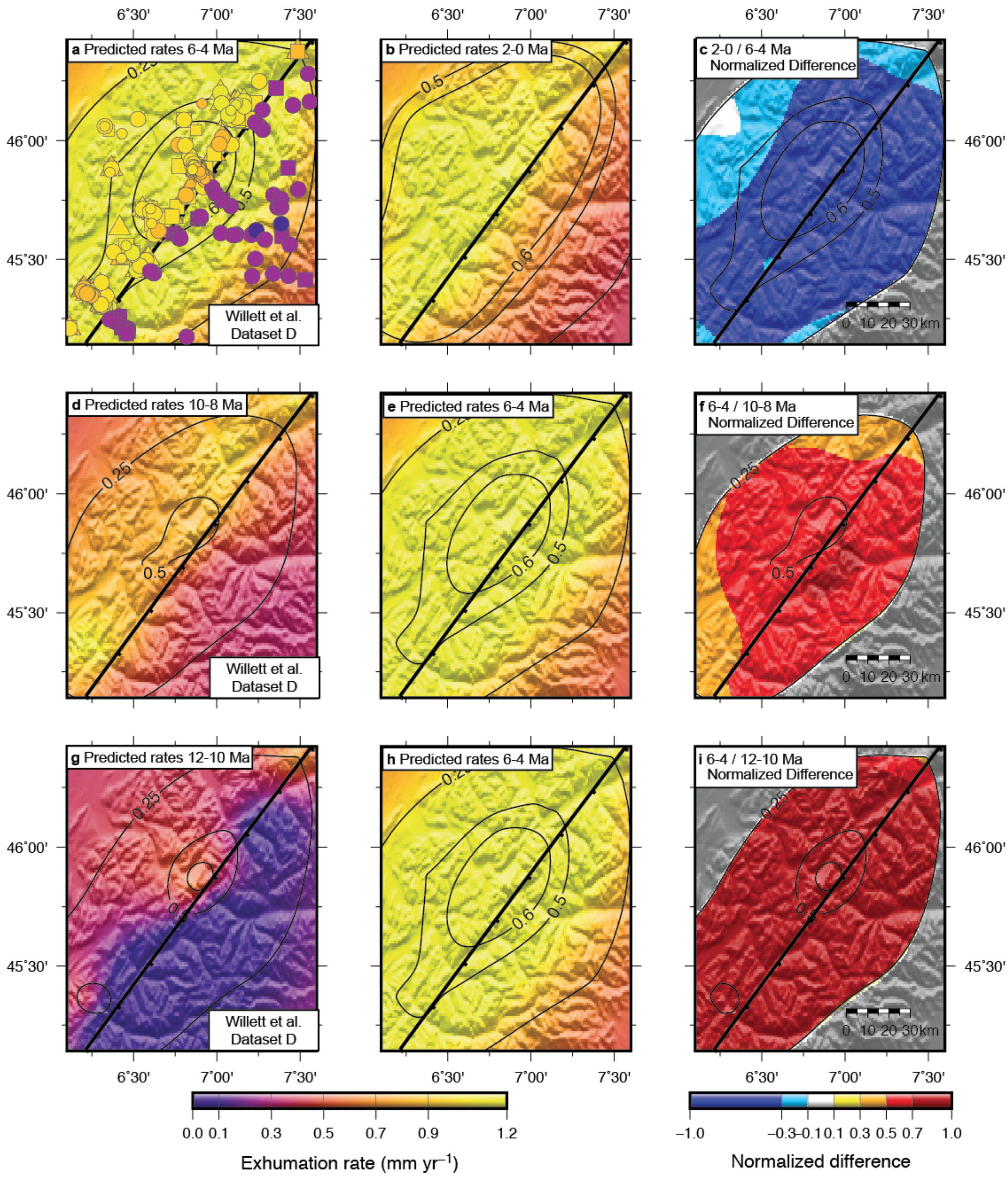

Figure 1. Modified version of Figure 3 in our initial comment, showing synthetic results for earlier time-windows. Input exhumation rates are 1 mm/yr NW of the fault (thick black line) and 0.25 mm/yr SE of the fault. Dataset D of Willett et al. (shown in a; symbols and colours as in Fig. 3 of our initial comment) and a prior erosion rate (e_p) of 0.35 mm/yr were used. (a-c) Predicted temporal evolution in erosion rates using all the data (equivalent to Willett et al.'s Figure 14): (a) predicted erosion rates for the 6-4 Ma time bin; (b) predicted erosion rates for the 2-0 Ma time bin; (c) normalized difference in erosion rates, only shown where the resolution in each time bin is > 0.25. Contours in plots (a-c) show the predicted resolution. (d-i) Predicted temporal evolution during earlier time bins for this inversion: (d) predicted erosion rates for the 10-8 Ma time bin; (e) predicted erosion rates for the 6-4 Ma time bin (same as a); (f) normalized difference in erosion rates, only shown where the resolution in each time bin is > 0.25; (g) (d) predicted erosion rates for the 12-10 Ma time bin; (e) predicted erosion rates for the 6-4 Ma time bin (same as a); (f) normalized difference in erosion rates, only shown where the resolution in each time bin is > 0.25. Note the significant increase in erosion rates that precedes the decrease shown by Willett et al.