This work presents numerical simulations of the temporal evolution of an instantaneously emplaced sill with the aim to reproduce temperature gradient observed in a natural drill hole.

I find this work to be a very nice and important follow-up study of the work of Borisova et al. (2023) that brings a very significant increment in physical accuracy. In this well-written manuscript that I enjoyed reviewing, the authors make a compelling case for this type of simulations. I have only two minor reservations. The first is that, although I understand that the authors would like very much like to use the results to discard one of the two explored scenarios (rhyolite sill vs. basalt sill), the first part of the discussion is too curt towards the rhyolitic sill scenario. I suggest ways to soften their assessment and honor the complexity of comparing temperature gradient across a notoriously complex interface. The second reservation concerns the limits of the model, which are very clearly stated, except for the reasons behind model crashes and the assumption of the absence of water saturation. Here also, I list below specific (and hopefully constructive) questions that should clarify these two points.

As all my other comments are directed towards clarifications and no additional runs are needed to complete this elegant and topical study, I recommend acceptance with minor revisions.

Detailed comments.

l. 14. This formulation about the choice of rhyolite vs. basalt sill is well-balanced and does not suffer from the same limitations as the beginning of the discussion (cf. comment on l. 285).

l. 16. I strongly suggest rephrasing this sentence, because the approach can only be qualified of "suitable" if it did not crash inexplicably after a few years.

l. 34-37. First, state the observation *According to Eichelberger (2020)*…, then mention the interpretations *Such a high temperature gradient* … and the *presence of this sharp gradient* […] *indicates convective transfer*…

l. 42. The names are confusing. IDDP-1 is the name of a drilling site (or project?) containing a well also numbered IDDP-1 and another well numbered KJ-39? Please clarify for the audience unfamiliar with the Krafla site.

l. 65 *setting* -> *setup*

l. 71 *constant due to* -> *constant in space due to*

l. 80 It seems that there is an unmentioned assumption of no fluid saturation. This is important because the rhyolite has 1.9 wt% H_2O (see comment on l. 460). I suggest that you add that assumption here, with, if relevant, possible justification from previous works that neglecting fluids is a reasonable hypothesis.

Section 2.3 I appreciated the very nice preliminary assessment of the relevant scaling relationships of this system.

l. 163. The text mentions a no-slip velocity condition, whereas Fig. 3 mentions a free-slip condition. Please clarify.

l. 179-180. *temperature gradient* -> *temperature difference*. In fact, it would be much clearer to systematically compare temperature gradients explicitly, just as done in Scenario 2. It was hard to follow and gather information to see that Scenario 1 has a 15±0.25 °C/m gradient after 35 yrs, which is slightly lower (and mostly likely undistinguishable from, more on that later) than the observed gradient of ≥16 °C/m.

l. 186 I do not understand why the grid size selection is evaluated against the melt zone thickness, whereas the Section 2.3 states on l. 155 that the best measure is the Nusselt number. Please clarify why the finding of Stevens et al. (2013) is set aside here.

l. 190 Please give a quantified range of temperature gradient here for consistency (as stated currently, the range is ≥11–33 °C/m, which is compatible with the rhyolite sill gradient).

l.212 *°C/m/m -> °C/m*

l. 214 (and also l. 290) It is important to state the reason(s) behind the crash: vanishing time step, stalling of the residuals, temperature runaway, instability of Eq (4), … Currently, I can only infer that the crash probably does not occur because of too high a viscosity (see next comment).

l. 229 The maximum viscosity is said to have a minor effect on the results. Does it mean that it also does not affect crash time (see comment above)?

l. 265. *"although the 1D model shows temperature gradients within convective regions"*. This is a major drawback as the variable of interest is the temperature gradient. Just looking at Fig. 9, I guess that the 1D gradients across the melt zone thickness are basically meaningless because they are so far away from the 2D gradients. If my guess is correct, I suggest adding a few sentences about this issue as it shows clearly the added value of this 2D study.

l. 283. I did not understand the end of the sentence: by a comparable amount to what?

l. 285. I appreciate the discussion and evaluation of the effect of the basalt sill thickness. I wonder why the rhyolite sill thickness is not mentioned, and thus I suggest adding a few sentences about it. Borisova et al. (2023) tried with a sill thicker that the 300 m used here and found larger temperature gradients after the dreaded 35 yrs. I am not a fan of highlighting every parameter and asking to explore it further, but here the conclusion of the paper hinges on that single sill thickness. Presumably a sill of 350 m would bump the gradient from 15 to 16 °C/m, suddenly making Scenario 1 unquestionably valid?

l. 317 “best matches” is a simplistic assessment of the results. I suggest rephrasing that first § in terms of degrees of freedom and parameter ranges. As an example, here is what I have in mind: *An extrapolation of our results suggests that Scenario 1 likely fits the observed values within a very narrow range of parameters, whereas our results for Scenario 2 covers a wider range of parameters that yield gradient comparable to those observed. […] For all these reasons we prefer Scenario 2 over Scenario 1.*

l. 330-344. The petrology § are not linked to the framework of this study. Please ensure that it is the case. For instance, fractional crystallization is incompatible with the model assumptions, which needs to be mentioned. Another example is that hydrothermal fluids were needed to produce the partial rhyolitic melts, but 1) the model ignores hydrothermal fluids and 2) the whole § on l. 301-309 is dedicated to show that hydrothermal fluids played no *direct role in* […] *partial melting and the following reaction of the rock with* [putative] *basaltic magma*. Finally, the duration of 33 yrs is chosen here (and also in the Conclusions l. 354), whereas the whole work (starting l. 60) and all model results were evaluated at 35 yrs. This would be a detail if I were not tempted to wonder how much higher the 15 °C/m gradient of Scenario 1 is at 95% of the simulation time.

Appendices. I appreciated the clever selection parameter sweeps that gave me confidence in the numerical outputs.

l. 460. I was trying to find the answer to the question: how much, if any, total/dissolved water were assumed to be present in the rhyolite/basalt. Borisova et a., (2023) reports rhyolite composition with 1.9 wt% H_2O, and, unless I got lost, no basaltic composition. This is a confusing issue as this water-bearing rhyolite is the result of partial melting but the injected sill is a different (source) rhyolite in Scenario 1. To clarify this issue, could you simply add a Supplementary/Appendix Table with the initial MELTS composition for both Scenarios?

Alain Burgisser