Review of "Crustal Seismogenic Thickness and Thermal Structure of NW South America" by Gómez-Garciá et al., submitted to Solid Earth.

Review by: Dr Sam Wimpenny, University of Leeds, UK.

Manuscript Overview

Gómez-Garciá et al., present an analysis of the spatial variability in the depth of seismicity within the NW South America and its relationship with the thermal structure of the lithosphere. Their work touches on a long-running research question regarding the controls on the depth of seismicity within the continents [e.g. Chen and Molnar 1983; McKenzie et al., 2005; Jackson et al., 2008]. Gómez-Garciá et al., argue that the majority of seismicity in the NW Andes nucleates in rocks at temperatures <350 degrees Celsius, with a smaller number of earthquakes nucleating in crustal or mantle rocks at temperatures higher than 350 degrees Celsius. They argue that earthquakes nucleating in rocks at temperatures >350 degrees Celsius indicate that they may have more mafic lithologies, or that the depth estimates of these events may be inaccurate. The authors also find that there is a strong correlation between areas of elevated geothermal gradient and seismicity, but do not provide a physical interpretation for the correlation.

I found the manuscript relatively easy to read and well annotated, though in a number of places the text needs adjusting to remove some ambiguous statements and definitions (I have outlined these in the line-by-line comments).

I do have methodological concerns with the analyses presented in the manuscript, including with the accuracy of the routine depth estimates used, the method used to compute the seismogenic thickness, and the assumptions made in computing the temperature structure. These concerns make it difficult (at least in my view) to assess how robust the major conclusions in the manuscript currently are. Nevertheless, I have outlined how the authors might address these concerns.

A criticism I have with this manuscript is that it has mostly overlooked the extensive literature on the relationship between earthquake depths, geology, and temperature within the continents that is directly relevant to the arguments they present [Jackson 2002; Jackson et al., 2004; McKenzie et al., 2005; Jackson et al., 2008; Sloan et al., 2011; Craig et al., 2012]. I recommend the authors more explicitly address how their conclusions advance on the findings made in the aforementioned papers, as this will help contextualise the contribution that this paper makes to the field.

Overall, because of the technical concerns I have with this manuscript, I would recommend that the manuscript be returned for **major corrections** in order for the authors to address these concerns.

General Comments

1. **Inaccuracies in the ISC earthquake hypocentral depths limit the earthquake depth analyses presented:** The authors use the ISC catalogue's hypocentral depth estimates to map out the depth distribution of seismicity. The ISC locates earthquakes using the reported travel times of body waves, and for events less than ~40 km deep the ISC location procedure can be incorrect by tens of kilometers in depth. These errors are not necessarily reflected in the formal uncertainties in the catalogue, which is a well-known limitation associated with using the ISC earthquake catalogue to study shallow seismicity [Maggi et al., 2000; Chen et al., 2009; Weston et al., 2011] and is something the authors acknowledge. Most studies use waveform modelling of the earthquakes to accurately determine earthquake depths and thereby draw robust conclusions regarding the depth of seismicity in the continents [see examples for South America in: Suarez et al., 1983; Chinn and Isacks 1983; Devlin et al., 2012]. Without accurate earthquake depth estimates, it is unclear to me to what extent variability in the calculated seismogenic thickness, D10 or D90 represent errors in the earthquake depths in the ISC catalogue, or real spatial variability in the depth of earthquake generation.

To address this comment, I would recommend that the authors re-analyse the depths of earthquakes in their study area using waveform modelling of teleseismic [e.g. Devlin et al., 2012; Wimpenny 2022] or regional [e.g. Alvarado et al., 2005] seismic data. Alternatively, if the author's believe that the depths of crustal earthquakes in the ISC catalogue are accurate enough for their purposes, then they need to demonstrate this by comparing the ISC event depths with an independent source of earthquake depths (e.g. from local seismic deployment), because in most other settings the ISC event depths are not accurate enough for these purposes.

2. **The method used to compute the seismogenic thickness does not necessarily account for the real depth-distribution of events in each grid area:** The authors say that the D10/D90 statistics are calculated on a 0.1x0.1 degree grid by considering the nearest 20 events to each grid centre as the sample from which D10/D90 are computed. This method appears biased to me. Consider a case where there are 100 events within any particular 0.1x0.1 degree grid area, and the nearest 20 events to the grid centre are all <10 km depth but the remaining 80 are all at 50 km depth. The current method for estimating the D90 would yield a value <10 km for the whole grid area, despite the fact that earthquakes are occurring down to 50 km. Therefore, I suspect the method used to compute the seismogenic thickness could yield misleading results. This might not be an issue if the typical earthquake spacing in lat/lon is larger than ~0.1x0.1 degree, but the authors should add new results demonstrating that their interpretations are independent of the gridding scale and sample size used in calculating the D10/D90/seismogenic thickness.

3. **D10 and D90 may not be the relevant metrics for understanding the absolute depth range of earthquake nucleation:** The D10 and D90 parameters have been developed to study temporal variations in the depth of seismicity in regions with dense earthquake catalogues (e.g. California, Japan; see Rolandone et al., [2004]). These metrics do not seem suitable when studying the depth-extent of seismicity in the ISC catalogue. Firstly, the ISC catalogue is relatively sparse, and so choosing D10/D90 may not be robust as it might vary with longer observation intervals. Secondly, the 10% and 90% cut-off are arbitrary values – the more common and logical definition is that the seismogenic thickness is the depth of the deepest observed earthquake in an area [like used in Maggi et al., 2000; Jackson et al., 2004]. Could the authors add some explanation as to why they use D10 and D90 and why it is a relevant metric?

   I will also suggest that D10 is unlikely to be a robust metric when calculated from the ISC catalogue, because hypocentral locations derived using travel time data have particularly poor depth resolution within the top 10 km of the crust. Weston et al., [2011] and Wimpenny and Watson, [2021] have demonstrated the poor resolution of ISC-EHB event depths in the upper crust by comparing ISC hypocentral depth estimates with both finite-fault slip solutions derived from modelling InSAR data and more accurate earthquake centroid depths from waveform modelling. Both studies found that there can be differences on the order of 5-10 km, and that the ISC systematically overestimates the depth-range of slip in shallow earthquakes.

4. **The temperature field within the mountains is most likely not in steady state:** The authors have assumed that the temperature field throughout NW South America can be modelled as being in steady-state. The time taken for a system to approach thermal steady state is given by the thermal time-constant $\tau = l^2/\pi^2\kappa$ where *l* is the thickness of the lithosphere and κ is the thermal diffusivity of the lithosphere (~10–100 Ma for normal lithosphere). Given that mountain building will advect heat and move the lithosphere out of thermal steady-state [England and Thompson, 1984], and that mountain building in the NW Andes has taken place more recently than 10-100 Ma, then I would assume that the temperature field within the NW Andes cannot be modelled as being in thermal steady state. I agree that the steady-state assumption is more reasonable within the forelands.

   The heat flow and downhole temperature data the author's use to validate the models come from the top 4 km of the crust, and are mostly located offshore and not within the continental lithosphere, therefore are not a good test of whether the model is representative of the temperature field at depth in Region 1, 2 or 3. To address this comment, the authors need to justify why they think the steady-state assumption is valid for the mid/lower crust.

5. **Lab and nature suggest olivine remains seismogenic up to ~600 degrees – not 600-1000 degrees:** There are two strong lines of evidence that suggest olivine-rich rocks can only remain able to nucleate earthquakes up to temperatures of ~600 degrees Celsius. The first comes from the depths of well-constrained earthquakes within the oceanic lithosphere, where the temperature structure is well known. Here the maximum centroid depths of large earthquakes can be seen to deepen with the age of the lithosphere and remain consistently shallower than the 600 degree isotherm [McKenzie et al., 2005; Craig et al., 2014]. Secondly, laboratory experiments show that olivine aggregates switch from deforming through stick-slip, to deforming through stable creep, at temperatures equivalent to ~600 degrees at the strain rates expected within the oceanic lithosphere [Boettcher et al., 2007]. The authors cite Scholz [2019] (a textbook) regarding the seismogenic range of temperatures for olivine being 600-1000 degrees, but do not describe the evidence for this. I would suggest that the author's cite the original papers that came to this conclusion, as I have not been able to check where their logic has come from.

Line-by-Line Comments

Line 11: "Crustal seismogenic thickness" is a misleading term in my view, because the seismogenic layer can include parts of the crust and upper mantle. I would suggest that just the depth extent of earthquakes in the lithosphere is the more logical description of the seismogenic thickness, and the depth most likely to correlate with a "stability transition" from seismogenic to aseismic deformation processes.

Line 18: "Potential temperature" means specifically the temperature at which a rock would be if moved from a particular depth along an adiabat, which I think is different from what the authors mean here. I think they mean just the possible absolute temperature, and would recommend removing the word "potential".

Line 32: Why is the crustal seismogenic thickness not just the layer where all earthquakes occur? I'm not sure why it would be defined as where 'most', but not all, earthquakes occur.

Line 39: Is 90% of events based on a measure of statistical significance, or is it just defined as an arbitrary value?

Line 46-48: I would argue that it is generally accepted that earthquakes *mostly* nucleate within the crust at temperature <350 degrees Celsius, but *can* nucleate in lower crustal rocks at temperatures up to 600 degrees Celsius within certain areas where the lower crust might be dry [see Jackson et al., 2008; Craig et al., 2011; Sloan et al., 2011; Craig et al., 2012]. Examples of lower-crustal seismicity are now extensive, and there are plenty within South America [Assumpcao et al., 1992; Devlin et al., 2012; Wimpenny 2022].

Line 73: Typo "as the extent of the CST" not "extend".

Line 77: It is not clear to me why flat slab subduction means that you can assume thermal steady state. If anything, I would assume that flat slab subduction would suggest that you need to account for the heat advection in the modelling of the temperature field, because the slab has not always been flat and therefore the temperature boundary condition on the base of the lithosphere has changed through time. Please could the authors elaborate on why a flat slab configuration allows them to make an assumption of thermal steady state?

Line 92: What is meant by "large uncertainties" in the Moho? Could the authors give quantitative estimates of the typical Moho uncertainty?

Figure 1: There are a number of references to different geographic places (e.g. Panama-Chaco Block) in the text that are not defined on Figure 1. Please could the authors add these place names to the figure to help the reader.

Line 242-245: The authors "disregard earthquakes below the Moho…", and use only events above the Moho to compute "… the upper and lower stability transitions". By ignoring earthquakes below the Moho, then the maximum depth of earthquakes the authors calculate is limited by the crustal thickness, and doesn't necessarily correspond to any "stability transition" or mechanical condition in the lithosphere. I would recommend changing the phrasing the authors are using to describe what they are actually measuring, which is more like the fraction of the crust that is seismogenic.

Line 263: Why is removing events that are below the magnitude of completeness (Mc) relevant for this analysis, because the authors are not necessarily studying earthquake frequency or moment release through time? Small events below Mc could still be useful in defining the depth distribution of seismicity.

Line 274-276: The bootstrapping method of estimating the uncertainties in D10/D90 yield what I would consider unrealistically small values of <1 km. This is presumably because the 'formal' uncertainties in earthquake depths included in the ISC catalogue that are propagated through the bootstrapping are not realistic representations of the true depth errors. For example, comparing earthquake hypocentral depths in the ISC-EHB catalogue with more accurate centroid depths calculated from body-waveform modelling yields a mean difference on the order of ~5±10 km with errors of up to 50 km [Wimpenny and Watson, 2021]. Therefore, I expect a more realistic uncertainty is at least 5-10 km.

In addition, the inference that "*low uncertainties [estimating from bootstrapping] indicate that using 20 earthquakes for each node is [sic] already reliable…* " is not necessarily true in my view. If you used a sample size of 1 earthquake, then you would get low uncertainties, but the uncertainty estimate is not necessarily robust in that changing the sample used would give a different answer. The way to check whether 20 events is a reasonable sample size for giving robust results would be to make sure that increasing the sample size does not alter the conclusions (even if the spatial resolution has to decrease as a product of the greater sample size).

Line 292: I recommend quoting the misfit as 5 degrees, not 4.99 degrees, as the latter makes it sound like the model and data have precision to two decimal places, which is unlikely.

Section 4.1: A more general comment is that the thermal model validation relies on comparing measurements of the temperature in the very shallowest part of the crust in areas that are mostly offshore and do not overlap with Regions 1, 2, or 3. Could the authors add some text explaining why they think that validating the model using measurements from outside of the region of interest means that it will be valid in the regions of interest, which may itself have a different geological history and therefore different temperature structure (e.g. the effects of mountain building on the geotherms).

Line 317: Why is the "regional seismogenic zone" defined as 20 km? This seems arbitrary to me, and different from the seismogenic thickness quoted elsewhere.

Line 326-329: I would recommend re-wording this sentence, as on first reading I thought the authors were suggesting that areas with high geothermal gradients were areas with a strongly coupled crust and lithospheric mantle, though they actually mean it is the other way around. Maybe just split it up into a few shorter sentences for clarity.

Line 348-354: This is an interesting test, and it is physically intuitive that areas of higher geothermal gradient will have more seismicity because the lithosphere is hotter, therefore weaker and deforms more rapidly in those areas. However, I would be interested to know whether the thermal modelling adds any more useful information on where earthquakes are expected to occur relative to direct surface observables such as topography or fault traces.

Line 407: Please cite the original source for the constraint on the seismogenic temperatures for olivine gouge as 600-1000 degrees.

Line 415: Why does shallow earthquake hypocentres mean faults are ‘not well developed’, and what is meant by ‘not well developed’? Do you mean ‘immature’ in that they haven’t accommodated a significant amount of relative motion? Is there any evidence from natural fault zones that the hypocentral distribution of earthquakes differs with the ‘development’ of the fault, or are these inferences based off of laboratory experiments?

Line 417-419: This sentence captures much of my concern regarding the results of this study. The authors state: *“However, despite relying only on the best located earthquakes (see Sect. 3.2.1), large errors in the hypocentral depths still remain (up to 30 km, see Fig. S5), and should be considered in the analysis of our results.”* I agree, but it’s not clear to me how the authors have considered this in their own analysis of their results. In my view, the best way to mitigate this issue is to use more accurate methods for determining earthquake depths, which are more time consuming but will yield more robust results.

Line 443: What do the authors mean by “D90 depths of almost 35 km … are in agreement with the crustal-scale structure that these systems likely represent”? Faults can be ‘crustal scale’ but be aseismic still presumably (e.g. segments of the San Andreas Fault).

Line 445: “The D90 values … are clearly bounded by major faults” – I find it hard to see these patterns reflected in Figure 8, mainly because there are a lot of faults that do not seem to bound changes in the colour patterns. Maybe highlight the specific faults thought to bound the changes in D90 in the Figure?

Line 560-565: There are a few points where the authors mention the diverse geology of the northern Andes and how mafic/ultramafic terranes at depth might account for the deep seismicity. These arguments would benefit significantly from a simplified geological map of the study that highlights where these ultramafic/mafic terranes are based on surface outcrop and display these data relative to the seismicity and D90 maps.