**“Crustal Seismogenic Thickness and Thermal Structure of NW South America”**

**Author's response to review by Dr. Sam Wimpenny.**

We sincerely thank Dr. Wimpenny for his prompt review and insightful comments. We found them very useful, and we very much appreciate his feedback. Please find below a point-by-point reply to the issues raised by him.

Considering his comments and concerns, we plan to improve the manuscript upon its previous version. We hope that this new version supports in more detail the analysis performed and better frames the contribution that the paper makes to the field.

New references (not already cited in the manuscript) are listed in a separate section at the end.

**Manuscript overview**

**Comment:** …”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.”

**Response:** Thank you for pointing this out. We will add a short paragraph in this regard.

**Comment:** 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.

**Response:**

Most of the papers cited by the reviewer have been in turn already cited and discussed by the references included in the manuscript, such as the reviews by Scholz (2019) and Chen et al. (2021). But, of course, we can add more references to better frame the results. Our findings are not at odds with earlier ones, but are based on a much more detailed thermal modelling than typically achieved. For example, the references cited by the reviewer base their temperature estimates inside the Earth on 1-D or 2-D thermal models, using simplified lithospheric structures, instead of more realistic 3D models as implemented in our work. Nevertheless, this will be specifically addressed in the next version. Please see our response to comment #5, as it is also related to this issue.

**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.

**Response:**

To summarize our reply to this issue: In this research, we have used the ISC Bulletin because it is the most complete, public and homogeneous seismological catalogue available for such a wide region and temporal period. Large-scale relocation of such an extensive catalogue by full-waveform modelling is beyond what can currently be achieved. Nevertheless, for the next manuscript version we will enhance the working catalogue by replacing the ISC locations for the more accurate ones published elsewhere (for a minority of events). This future improvement of the catalogue is not expected to modify much the overall results or conclusions of this work, because they depend on the overall statistics of a large earthquake sample.

Delving into details, some of the reviewer's criticisms about the data refer to an older version of the ISC catalogue, not the newer and improved one used here. The reviewer pointed out the inaccuracies of the ISC location procedures, particularly regarding the hypocentral depths, and cites several research papers noting these limitations, dated from 1983 to 2012. But, as already cited in the manuscript, the ISC Bulletin has been recently completely rebuilt for the period 1964-2010 (Storchak et al., 2020), adding additional earthquakes and relocating hypocenters with the same location procedures used from 2011 onwards (Bondár and Storchak, 2011). The criticisms stated by those papers mentioned in the review refer to the older version of the ISC database, not to the rebuilt and improved one, for which location accuracy is expected to be significantly improved.

For example, currently ISC does not locate seismicity considering only the arrivals of body waves (*P* phases), as mentioned in the review, but the new algorithm also considers diverse *S* phases (Bondár and Storchak, 2011). When ISC cannot relocate an earthquake with teleseismic data (for example, because it has a low magnitude), it considers as prime the location provided by an authoritative regional agency, built from phase arrivals measured in the records of permanent or local temporary seismic stations (Di Giacomo & Storchak, 2016). It is already a great achievement of ISC to build a global catalogue with such a level of detail.

Mapping variations in D90 or other depth percentiles requires as many events as possible for statistical analysis. That is our reason for using the ISC Bulletin as the most extensive earthquake catalogue available in this wide region and temporal period, down to its magnitude of completeness. We end up using the most reliable events, over 1400 earthquakes for the analysis.

In any case, we agree with the reviewer regarding that earthquake locations obtained with full-waveform modelling (or from dense, local, monitoring surveys whose data was not submitted to ISC) may be more reliable than those routinely published by ISC. For this reason, we will improve the earthquake catalogue used, by replacing the locations of the ISC Bulletin with more accurate ones, if available, obtained by waveform modelling (e.g. Whimpenny & Watson, 2021; Wimpenny, 2022) or local surveys (e.g. Dimate et al., 2003, if we can obtain the data from the authors, as they are not public). The work by Alvarado *et al*. (2005), also suggested in the comments, does not cover our study area. These catalogue modifications, nevertheless, are expected to affect only a minority of events. So the results after this extra effort are expected to be similar to the ones already shown in the manuscript.

Unfortunately, relocation of the whole catalogue used (with earthquakes even down to magnitude $M$=3.5), using full-waveform analysis is simply not feasible. While phase arrivals have been reported to ISC by a plethora of agencies in the region, most records of full waveforms are not public. Gathering them would imply contacting each agency, and digitizing the seismograms one by one if they are not in digital form (which is the common case before the year ~2000, when most seismograms are in analog form, either in magnetic tapes or paper records).

Full-waveform relocation is typically done only for a few, selected, recent events of moderate to large magnitude (usually $M \geq 5$) for which public digital, teleseismic waveform data (e.g. from IRIS stations) are available. This is actually already illustrated by the catalogue of Whimpenny & Watson (2021, updated online), which contains just 12 earthquakes in our study area (with magnitudes ≥ 5.1, occurred between 1979 and 2021, including those relocated by Whimpenny, (2022).

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.

**Response:**

The cell size of 0.1 X 0.1 degrees in the sampling grid used is indeed chosen to avoid the sampling problem mentioned by the reviewer. The problem would have arised if a larger grid spacing had been used. Namely, the minimum search radius to the nearest 20 epicentres in our region (the minimum value actually represented in Fig. 9b) is ~6.4 km, as can be checked in the data repository for the manuscript. Meanwhile, the maximum half-width of a 0.1 X 0.1 degree cell is ~5.5 km. So for each cell of the used grid, the area sampled is at least as large as the cell itself, as it should (e.g. Wiemer & Wyss, 2000; González, 2017). We will mention this issue in the next version of the manuscript for justifying the grid spacing used.

The mapping approach followed is to use, for each grid node (cell centre) the minimum number of earthquakes (20) which produces stable results of D10 or D90. This allows mapping these percentiles with the best spatial detail possible.

Using a different sample size would inevitably change the results. For example, using more earthquakes in the sample around each grid node would imply using larger sampling radii. This would always lead to different results of D10 or D90, as would imply mixing shallower and deeper earthquakes from different locations farther away from each other, and the spatial variations of D10 or D90 would tend to be inevitably smoothed out. Using smaller samples would produce the opposite effect (yielding an apparently higher variability of the results, but with higher uncertainty and thus less statistically meaningful).

Using a different sample size is indeed not justified. For sample sizes < 20, the uncertainty of the 10th (D10) or 90th (D90) percentile increases significantly, as will be further commented on below. And, given that the results are already stable for 20 events in each sample (as determined by bootstrap), a sample size > 20 would be arbitrary, and would need to be subjectively chosen by checking the appearance of the resulting maps. This kind of subjective choice has been faced elsewhere when sampling earthquakes spatially at different resolutions (e.g. see Schorlemmer *et al*., 2004). The procedure used here (using the minimum sample required for the particular statistical analysis, for each grid node, e.g., in González, 2017) avoids such a subjective choice. This will be emphasized in the next version of the manuscript.

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.

**Response:**

Non-reliable values of D10

After the reviewer's comments, and given the shallow values that we obtain for D10, we reckon that they may be more affected by the uncertainties of hypocentral depths. Even if the absolute depth uncertainties were similar for deeper and shallower events, the relative uncertainty (= uncertainty / depth) is larger for shallower ones.

To accommodate the reviewer's comment, we plan to remove the D10 analysis from the manuscript and only focus on D90, which is more interesting from the point of view of the thermal model and the application, e.g. to help constrain the maximum depth of the base of crustal seismic sources, instead of discussing the crustal seismogenic thickness.

As replied above, we will also include published, improved depth determinations in the catalogue (albeit, given that they will be available for a minority of events, the overall results are not expected to change much).

Justification of the use of D90 instead of other percentiles

Focusing on D90, it seems necessary to recall that the lower depth limit of seismicity is a transition, rather than a sharp boundary, for two reasons. One is physical: as depth and temperature increase, progressively fewer minerals are able to deform in a stick-slip fashion, instead of plastically (hence the term "brittle-ductile transition"). The other reason is observational: uncertainties in hypocentral depths blur this transition even further.

Indeed, in some publications, as noted by the reviewer, the largest hypocentral depths (which would be the percentile 100, i.e. D100) are considered. However this approximation is largely unstable, as it depends only on the local, largest, extreme value of the depth distribution, which could be particularly unreliable, due to the depth uncertainties (the largest, single depth in the sample, if unreliable, would make D100 useless). In fact, contrary to what the reviewer argues, the use of D100 is even more prone to temporal variations, as new earthquakes might potentially nucleate deeper than previous observations. These are the reasons for using a lower percentile, such as the 90% (D90), which is by definition more stable than D100.

For a given sample size, the higher the percentile, the less statistically robust it will be. For example, with a sample of 20 depths, it would make sense to quantify the 90% percentile (2 out of the 20 events, i.e. 20*90/100, are expected to be ≥ than it). It would also be possible to calculate the 95% percentile, but it would be less reliable, as it would depend only on the largest value of the sample (1 out of the 20 events, i.e. 20*95/100, is expected to be ≥ than it), and consequently, the bootstrap uncertainties would be larger than those of D90. Calculating, e.g. a meaningful 99% percentile (D99) would require at least 100 events in the sample and can be attempted only if many small earthquakes are well recorded (e.g. in Taiwan, Wu *et al.*, 2017). And calculating a meaningful 100% (D100) percentile would theoretically require an infinite sample (if a larger observation period is used, there will eventually be deeper earthquakes than observed before, changing D100, but not necessarily D90, for example).

In particular, D90 was initially devised for characterizing spatial, rather than temporal, changes in the hypocentral depth distribution (Sibson, 1982, Marone & Scholz, 1988). It is a robust metric, because it can be calculated even with small samples, and it is the most commonly used proxy to map spatial variations of lower seismogenic depths and to compare with thermal parameters (e.g. Tanaka, 2004; Omuralieva et al., 2012).

Moreover, in Probablilistic Seismic Hazard Analysis (PSHA), the use of D90 is very extended as a reasonable measure of the depth of seismicity, or the maximum depth of seismic sources (e.g. U.S. Nuclear Regulatory Commission, 2012, their chapter 5; Bommer *et al*., 2023).

Following these arguments, we will briefly justify the use of D90 instead of other percentiles, in the next version of the manuscript.

<u>Mapping  D90 with a sparse catalogue</u>

When mapping the or D90 (or D10) values, we have explicitly represented the mapping resolution (Figure 9b), and considered it in the interpretation, so that we do not over-interpret the results. Of course, in the few regions (such as in Japan or Southern California) with catalogues complete down to much lower magnitude thresholds, and with precise hypocentral depths, it is possible to use many more earthquakes and map these percentiles in greater spatial detail. In any case ,it is necessary to quantify the depth distribution of seismicity elsewhere too (not only in the very best monitored regions), and those limitations, which are accounted for, do not lower the merit of our mapping, which is the first done in the analyzed region.

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 τ =  l2/π2κ 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.

**Response:**

Steady-state assumption:

Sensitivity studies (Meeßen, 2019) have shown that the transient effects causing the temperature changes in the upper 50 km of a subducting system are not far from equilibrium. Likewise, Rodriguez Piceda et al. (2022) have explored the effects of the non-steady state and came to the same conclusion.

Although it is true that mountain building processes imply that the system might not be fully in equilibrium, it is generally accepted that the first-order thermal field within the lithosphere is driven by conduction (e.g.: Liu et al., 2021). See response to comment Line 77 for more details.

Observational constraints:

We already acknowledged in the manuscript that there are limited heat flow and temperature measurements in the study area. Precisely for this reason, we made the effort to provide the best thermal model considering a fully 3D thermal approach, by integrating all the available data.

Indeed, the sensitivity test we performed with 25 different models (see Text S2 in the supplements) shows that the fit to the few temperature measurements available is already highly sensitive to the thermal parameters we explored. Figure S1 depicts the residual temperature for the different models we tested. The residuals of the first 19 models are particularly large, which implies that there is a limited range of the parameter values which can properly fit the observations.

In any case, even in an area with abundant high-quality temperature and heat flow observations (e.g. Southern California), those will be always restricted to the uppermost kilometers of the crust. See response to comment Section 4.1 for more details.

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.

**Response:**

The limiting temperature for seismogenesis in mantle forming materials has been a matter of debate. Some authors define a rather strict limit at 600°C (e.g.: McKenzie et al., 2005; Craig et al., 2014), but new evidence suggests that this limit might occur at higher temperatures. For example King and Marone (2012) explored the deformation of olivine fault gouges in the temperature range of 400–1000°C. Their results suggest that the velocity weakening (negative a-b) occurs at temperatures between 600 and 1000°C, depending on the strain rates applied to the sample (Figure 1).

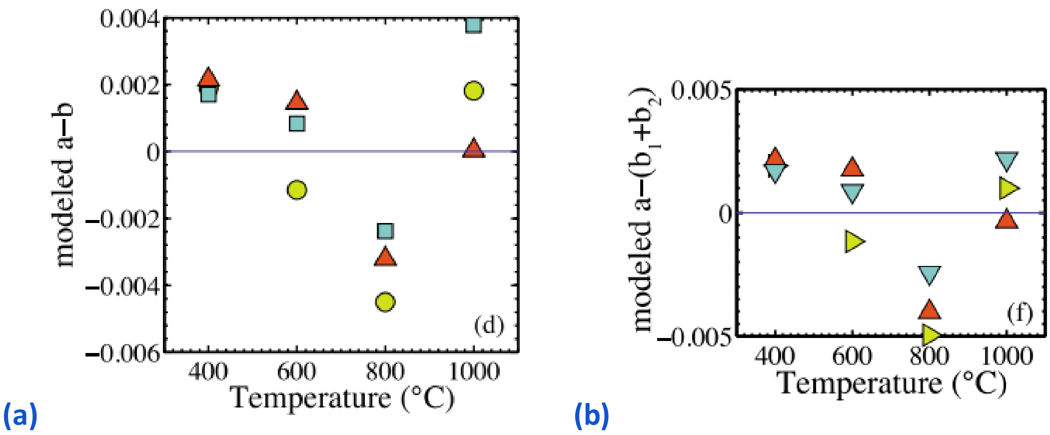


**Figure 1.** Stability parameter, a – b, for different velocity steps: green 1-10µm/s, orange 10-50µm/s and blue 50-1µm/s. (a) Rate-state model. (b) Two state variable approach. Figures taken from Figures 6 and 8 in King and Marone (2012).

Similarly, Ueda et al., (2020) studied the brittle-ductile transition in peridotites based on a fault system developed in the Balmuccia peridotite body. Their results indicate that the ductile-to-brittle transition in the peridotite occurs at ~720 °C. Lastly, Grose and Afonso (2013) studied the evolution of the oceanic lithospheric using more realistic thermal models than McKenzie et al. (2005), yielding a brittle-ductile transition closer to the 700-800°C isotherms depending on the mantle potential temperature (see Figure 10 in Grose and Afonso, 2013). This implies a higher temperature than a purely thermal, simplified approach, as their workflow takes into account not only hydrothermal circulation, but also P-T-dependent thermal properties.

Regarding the citation of Scholz (2019), we will modify the text adding the original references compiled by the textbook, and will add some additional works pointing to a higher temperature for seismogenesis of mantle forming materials, as described above.

The original references cited by Scholz (2019) are:

> King, D. S. H., & Marone, C. (2012). Frictional properties of olivine at high temperature with applications to the strength and dynamics of the oceanic lithosphere. *Journal of Geophysical Research: Solid Earth*, *117*(12), 1–16. https://doi.org/10.1029/2012JB009511
>
> Boettcher, M. S., Hirth, G., & Evans, B. (2007). Olivine friction at the base of oceanic seismogenic zones. *Journal of Geophysical Research: Solid Earth*, *112*(1), 1–13. https://doi.org/10.1029/2006JB004301

**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.

**Response:**

We had used the term "crustal seismogenic thickness" because we are targeting the depth distribution of seismicity at crustal levels, which implies removing from the analysis the sub-crustal seismicity (as done by others before, e.g.: Wu *et al*., 2017 in Taiwan, or Tsuda *et al*. 2019 in Japan).

Our study area includes two subducting flat-slabs, and therefore, there is intermediate to deep seismicity associated with them (see Figure 2 below). In order to disregard events that might be associated with the slab dynamics, we used the Moho from the GEMMA model (Reguzzoni and Sampietro, 2015) to filter out these events.

As can be seen in Figure 2, if we had not disregarded these deep events, the computed seismogenic thickness would reach up to 200 km (or more) in some parts of the study area, which is unrealistic for the purposes of analyzing crustal seismogenesis.

Nevertheless, in the next manuscript version, following the reviewer's comments, we will not analyze D10 and will not use the term "crustal seismogenic thickness". We will rather mention, as Wu et al. (2017) the "seismogenic depths of crustal earthquakes".

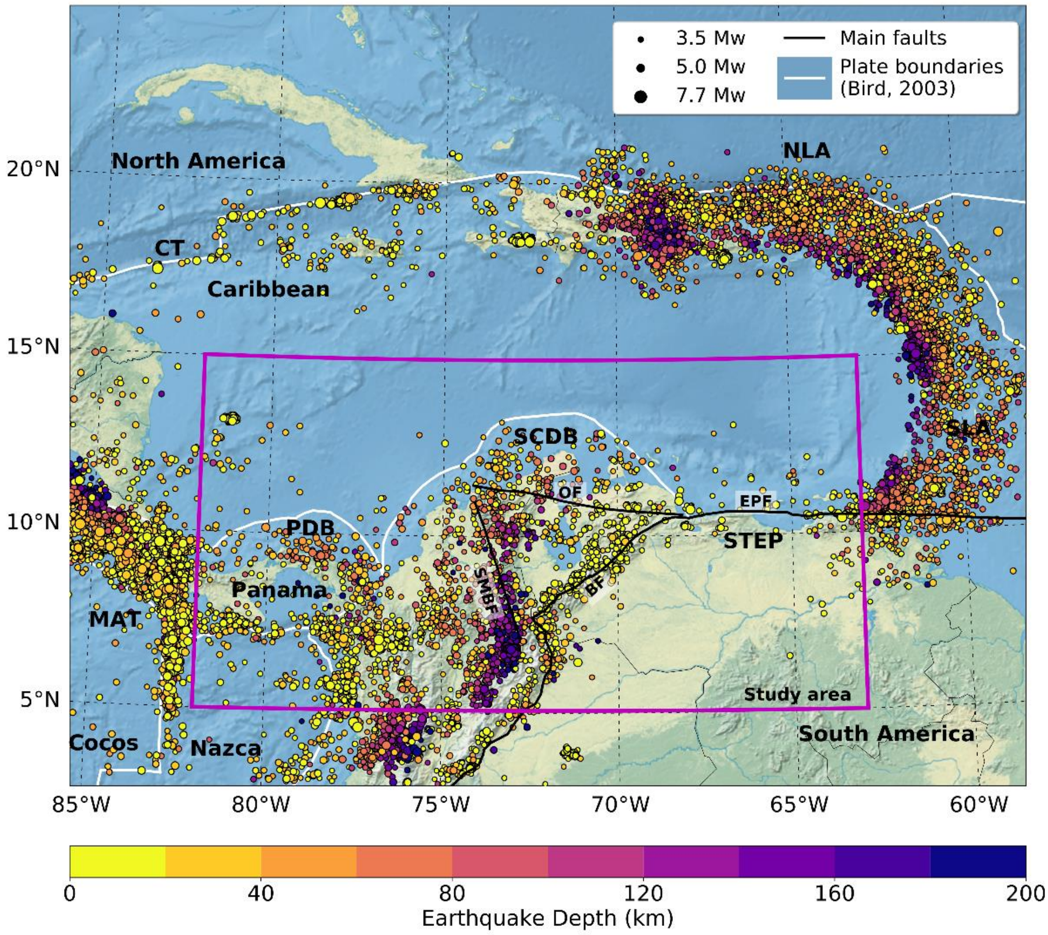


**Figure 2.** Seismicity in the Caribbean and northwestern South America (ISC, 2020).

**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".

**Response:** Thanks for pointing this out. We will modify the text accordingly.

**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.

**Response:** Please see response to comment regarding Line 11.

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

**Response:** Please see response to comment #3.

**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].

**Response:** Please see response to comment #5.

**Line 73**: Typo “as the extent of the CST” not “extend”.

**Response:** Fixed, thanks.

**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?

**Response:**

The present-day flat slab geometry has been established since about 6 Ma, when the Nazca tear developed separating the north (flat) and south (step) segments (Wagner et al., 2017). As a result, the volcanic activity has ceased in the continental crust of the overriding plate of the north segment, which spatially corresponds to our study area. This allows us to consider that the propagation of heat within the crust is mainly driven by conduction (e.g.: Liu et al., 2021).

Geodynamic models (e.g.: Schellart and Strak, 2021; Currie and Copeland, 2022) suggest that the subducting velocities are potentially reduced during the evolution from step to flab slab (see Fig 16 in Schellart and Strak, 2021 as an example -Figure 3 below). Considering that we are targeting the thermal field of the uppermost 75 km of the lithosphere, and that our applications are limited to crustal realms, we believe that a steady-state assumption can be applied in the study area to have a first-order estimate on the 3D feedback of the system heterogeneities and their imprint in the resulting thermal field. Moreover, given that the timeframe of the earthquakes we are studying is instantaneous in the geological timescales, our goal is to take a “snapshot” of the present-day thermal configuration considering the 3D system heterogeneities at crustal scales.

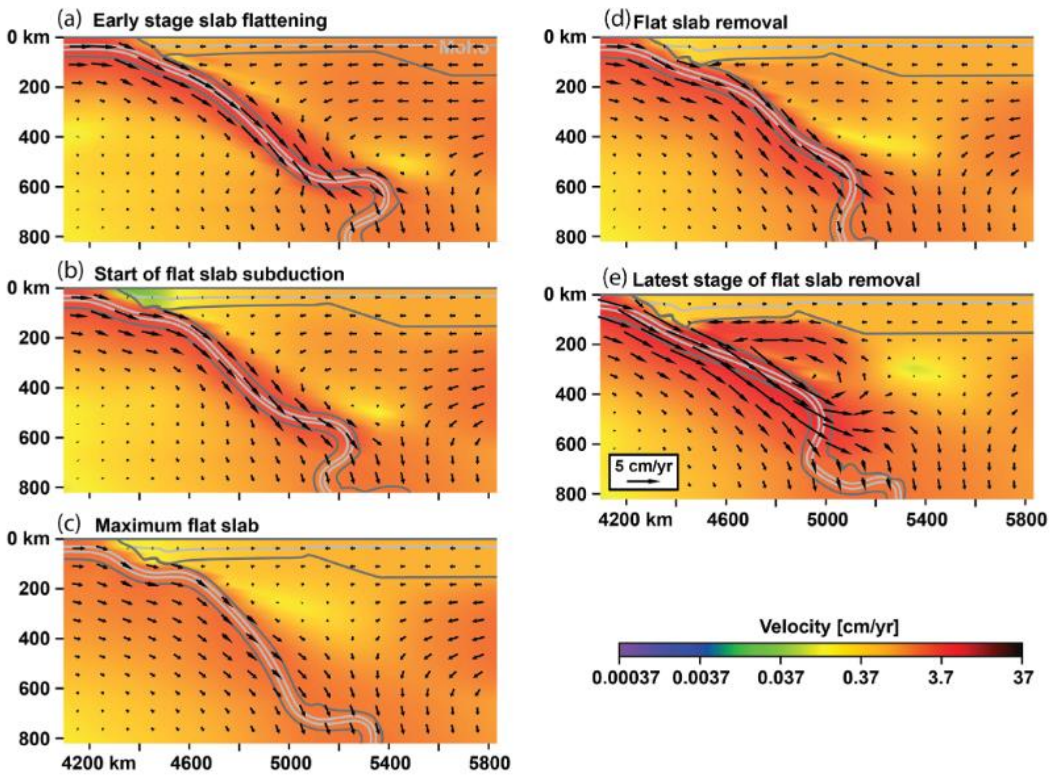


**Figure 3.** Resulting velocity field during the evolution of slab flattening (Schellart and Strak, 2021).

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

**Response:**

The Moho uncertainty map is provided in the supplementary Figure S8, as computed by the authors of the GEMMA model (Reguzzoni & Sampietro, 2015) used in this research (see line 500 of published preprint).

**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.

**Response:** Thank you. Yes, we will carefully check this issue and improve the figures.

**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.

**Response:** Please see response to comment regarding Line 11.

**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.

**Response:**

Small events below Mc are not useful for defining the depth distribution of seismicity, because their depths are biased. The deeper an earthquake is, the less likely it is to detect, that is, Mc increases with depth (e.g. Fig. 5 of Schorlemmer *et al*., 2010). If we used an incomplete sample, considering earthquakes below Mc, small deep earthquakes would be preferentially missing and the overall apparent depth distribution would be biased towards shallower values. We will mention these issues explicitly in the next manuscript version.

**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).

**Response:**

Despite the efforts taken at quantifying the uncertainties in our analysis, certainly, non-formal depth errors, (such as systematic errors due to the velocity model used in the location), cannot be directly quantified. As mentioned earlier, we will consider better earthquake locations, if available, to try to improve the reliability of the results. We will also discuss the additional uncertainties in the next manuscript version.

Low uncertainties estimated from bootstrapping indeed show that the results are stable. If the sample size were reduced, the results of the percentiles (e.g. D90) would be unstable: removing one earthquake from the sample and substituting it with another (i.e. bootstrapping the sample) would significantly change the percentile. The only exception would be the very unlikely case that all the earthquakes in each sample had the same or almost identical depths, so removing one or another would not change the percentile.

We find the comment about using a hypothetical sample of 1 earthquake inadequate, given that the very minimum sample to calculate a 10% or 90% percentile would be 10 cases. Repeating the calculations with samples of 10 earthquakes, the bootstrap uncertainties increase significantly, so such small samples are not as reliable as the value of 20 used here.

As discussed before, modifying the sample size would actually not be justified, and the results would change. Fewer than 20 events would yield more spatially variable but more uncertain D10 or D90 results, and more than 20 events would smooth out spatial variations, as earthquakes farther from each grid node would have to be considered in the corresponding sample.

**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.

**Response:** We agree, thanks. We will modify both the text and the figure accordingly.

**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).

**Response:**

See response to comment #4. Even if the temperature observations at depth are not located at those regions, they are able to constrain the thermal model parameters. Choosing other values for the thermal parameters results in a worse fit, as explored in the sensitivity analysis performed (Text S2).

Thinking about the study area as a 3D fully coupled system, we believe that being able to reproduce the observations is a good sign of the model first-order robustness for the applications we currently consider in our manuscript. A main reason for our chain of arguments is that we not only model temperatures (for example by interpolating those) but consider the structural complexity of the heterogeneous distribution of thermal properties together with the physics of conductive heat transport to predict the 3D thermal field that is consistent with the few observed temperatures. The 3D heterogeneity in physical properties is constrained by other methods, such as seismic data or 3D gravity modelling, as certain seismic velocity- density pairs can be interpreted as lithologies, from where we can conclude on the lithology-dependent thermal properties. In fact, as expected, we do get different temperature structures all over the study area, as the model integrates the complexity of the region.

We should note that the procedure followed, of building a 3D model integrating all available information, is a major improvement with respect to earlier approaches. Moreover, if eventually more data of heat flow or thermal measurements become available, they could be integrated in future models following this philosophy.

**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.

**Response:**

The histogram of the hypocentral depths of the selected earthquake catalogue (Figure 7 c) shows that the majority of earthquakes in the study area nucleate at less than 20 km depth. The computed D90 of the entire catalogue is 20.9 km (see line 457). Therefore, we selected the temperature difference between the surface and 20 km as a proxy for estimating the geothermal gradient in the seismogenic zone at regional scale.

**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.

**Response:** Thank you, we will improve this sentence.

**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.

**Response:**

We did not explore the skillfulness of other geodynamic variables in our study area because testing this systematically would require a whole new paper (as done, e.g. by Becker et al., 2015). It would require dedicated modelling efforts, to check if, for example, strain rates and rates of topography change are skillful geodynamic variables at forecasting the spatial distribution of seismicity, as in the region analyzed by Becker et al. (2015). It is important to note that it is not only the thermal gradient per se, but the contrast between domains of differing gradients that matters. We will add a sentence on these points to the new manuscript.

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

**Response:**

The original sources cited by Scholz (2019) are:

King, D. S. H., & Marone, C. (2012). Frictional properties of olivine at high temperature with applications to the strength and dynamics of the oceanic lithosphere. *Journal of Geophysical Research: Solid Earth*, *117*(12), 1–16. https://doi.org/10.1029/2012JB009511

Boettcher, M. S., Hirth, G., & Evans, B. (2007). Olivine friction at the base of oceanic seismogenic zones. *Journal of Geophysical Research: Solid Earth*, *112*(1), 1–13. https://doi.org/10.1029/2006JB004301

Nevertheless, we will modify the text to make it clear that the upper boundary of this temperature range is highly debated, including additional references we have been citing throughout this document.

**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?

**Response:**

Considering the uncertainties discussed above for shallow hypocentres, we will remove this comment. This interpretation was explained by Scholz (2019, p. 149 and references therein), based on the seismogenic depths observed in different faults worldwide:

*"The presence of this upper transition was first proposed by Scholz, Wyss, and Smith (1969) and is based on the observation that there are both upper and lower seismicity cutoffs on well-developed faults (Figure 3.45). The upper transition is most likely due to the presence of unconsolidated and phyllosilicate-rich gouge in the fault zone at shallow depth – material that is generally found to be velocity strengthening (Section 2.3.4). <u>The observation of an upper cutoff in seismicity at about 2–4 km seems to be limited to well-developed fault zones where a thick gouge layer is likely to be present</u> (Marone and Scholz, 1988). <u>In regions where there are no well-developed faults</u>, such as the central Adirondacks of New York (Blue Mountain Lake), Miramichi, New Brunswick, and Meckering, Australia, <u>earthquakes occur up to very shallow depths</u> (Figure 3.45). In contrast, in the Imperial Valley, California, earthquakes occur only deeper than 5 km, below an unusually thick sequence of unconsolidated sediments (Doser and Kanamori, 1986)." [Underlining is ours.]*

**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.

**Response:**

As we have already discussed above, it is not possible to avoid those depth uncertainties, as a complete relocation of the catalogue is beyond reach. We already considered the measurable, formal, depth uncertainties (which are, on average, 7 km in the selected catalogue, see Figure S7). And we will incorporate better depth determinations if available and take care, as we already did before, of not interpreting the results beyond their limitations. Future updates of the earthquake catalogue will eventually enable successive improvements of this kind of analysis in the region, for which our present manuscript is proposed as a necessary first step.

**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).

**Response:**

What we mean is rather obvious: D90 at the fault location is almost 35 km, so its seismicity is distributed along the crust down to such depths, and the fault seems to be a crustal-scale feature. If D90 were, instead, e.g. 5 km, we could not argue that the fault is of crustal scale, as it might be shallow only.

We find this remark important because there are no seismic profiles or detailed tomographic models which could provide alternative evidence of the fault extent at depth.

**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?

**Response:** Thank you; we will improve the figure in the next version of the manuscript.

**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.

**Response:**

Thank you for this recommendation, we will consider it for the next version of the manuscript.

---

**New references cited (not already included in the manuscript)**


Alvarado, P., Beck, S., Zandt, G., Araujo, M. & Triep, E. (2005): Crustal deformation in the south-central Andes backarc terranes as viewed from regional broad-band seismic waveform modelling. *Geophysical Journal International*, 163: 580-598. doi:10.1111/j.1365-246X.2005.02759.x

Boettcher, M. S., Hirth, G., & Evans, B. (2007). Olivine friction at the base of oceanic seismogenic zones. *Journal of Geophysical Research: Solid Earth*, *112*(1), 1–13. doi:10.1029/2006JB004301

Bommer, J.J.; Ake; J.P. & Munson, C.G. (2023): Seismic source zones for site-specific probabilistic seismic hazard analysis: The very real questions raised by virtual fault ruptures. *Seismological Reseach Letters*, 94, 1900–1911, doi:10.1785/0220230037

Currie, C. A., & Copeland, P. (2022). Numerical models of Farallon plate subduction: Creating and removing a flat slab. *Geosphere*, 18(2), 476–502. doi:10.1130/GES02393.1

Grose, C. J., & Afonso, J. C. (2013). Comprehensive plate models for the thermal evolution of oceanic lithosphere. Geochemistry, Geophysics, Geosystems, 14(9), 3751–3778. https://doi.org/10.1002/ggge.20232

King, D. S. H., & Marone, C. (2012). Frictional properties of olivine at high temperature with applications to the strength and dynamics of the oceanic lithosphere. *Journal of Geophysical Research: Solid Earth*, *117*(12), 1–16. doi.org/10.1029/2012JB009511

Liu, X., Currie, C. A., & Wagner, L. S. (2021). Cooling of the continental plate during flat-slab subduction. *Geosphere*, 18(1), 49–68. doi.org/10.1130/GES02402.1

McKenzie, D., Jackson, J., & Priestley, K. (2005). Thermal structure of oceanic and continental lithosphere. *Earth and Planetary Science Letters*, 233(3–4), 337–349. doi.org/10.1016/j.epsl.2005.02.005

Meeßen, C. (2019). The thermal and rheological state of the Northern Argentinian foreland basins, PhD Thesis, Potsdam: Universität Potsdam, 151 p. https://doi.org/10.25932/publishup-43994

Mora-Bohórquez, J. A., Ibánez-Mejia, M., Oncken, O., de Freitas, M., Vélez, V., Mesa, A., & Serna, L. (2017). Structure and age of the Lower Magdalena Valley basin basement, northern Colombia: New reflection-seismic and U-Pb-Hf insights into the termination of the central Andes against the Caribbean basin. *Journal of South American Earth Sciences*, 74, 1–26. doi.org/10.1016/j.jsames.2017.01.001

Omuralieva, A.M.; Hasegawa, A.; Matsuzawa, T.; Nakajima, J. & Okada, T. (2012): Lateral variation of the cutoff depth of shallow earthquakes beneath the Japan Islands and its implications for seismogenesis. *Tectonophysics*, 518–521, 93-105. doi:10.1016/j.tecto.2011.11.013

Rodriguez Piceda, C., Scheck-Wenderoth, M., Bott, J., Gomez Dacal, M.L., Cacace, M., Pons, M., Prezzi, C., Strecker, M. (2022). Controls of the Lithospheric Thermal Field of an Ocean-Continent Subduction Zone: The Southern Central Andes. *Lithosphere*. 2022 (1): 2237272. doi: https://doi.org/10.2113/2022/2237272

Schellart, W. P., & Strak, V. (2021). Geodynamic models of short-lived, long-lived and periodic flat slab subduction. *Geophysical Journal International*, 226(3), 1517–1541. doi.org/10.1093/gji/ggab126

Schorlemmer, D., Wiemer, S. & Wyss, M. (2004), Earthquake statistics at Parkfield: 1. Stationarity of *b* values, *J. Geophys. Res.*, 109, B12307, doi:10.1029/2004JB003234.

Schorlemmer, D.; Mele, F. & Marzocchi, W. (2010): A completeness analysis of the National Seismic Network of Italy. *Journal of Geophysiscal Research, Solid Earth*, 115, B04308, doi:10.1029/2008JB006097.

Tanaka A. (2004). Geothermal gradient and heat flow data in and around Japan (II): Crustal thermal structure and its relationship to seismogenic layer. *Earth, Planets, and Space* 56, 1195-1199. doi:10.1186/BF03353340

Ueda, T., Obata, M., Ozawa, K., & Shimizu, I. (2020). The Ductile-to-Brittle Transition Recorded in the Balmuccia Peridotite Body, Italy: Ambient Temperature for the Onset of Seismic Rupture in Mantle Rocks. *Journal of Geophysical Research: Solid Earth*, 125(2). doi.org/10.1029/2019JB017385

U.S. Nuclear Regulatory Commission (2012): *Central and Eastern United States Seismic Source Characterization for Nuclear Facilities.* Technical Report. EPRI, Palo Alto, CA, U.S. DOE, and U.S. NRC. http://www.ceus-ssc.com

Wagner, L. S., Jaramillo, J. S., Ramírez-Hoyos, L. F., Monsalve, G., Cardona, A., & Becker, T. W. (2017). Transient slab flattening beneath Colombia. *Geophysical Research Letters*, 44(13), 6616–6623. doi.org/10.1002/2017GL073981

Wimpenny, S. (2022): Weak, seismogenic faults inherited from Mesozoic rifts control mountain building in the Andean foreland. *Geochemistry, Geophysics, Geosystems*, 23, e2021GC010270. doi:10.1029/2021GC010270

Wimpenny, S. & Watson, C.S. (2020): gWFM: A global catalog of moderate-magnitude earthquakes studied using teleseismic body waves. *Seismological Research Letters*, 92 (1), 212–226. doi:10.1785/0220200218