Response to reviews of "A new consistent and high-precision earthquake catalogue for the Taupō Volcanic Zone, New Zealand" Submitted to Seismica, 2024

REVIEWER A: LAVAYSSIÈRE

My main comment is the absence of information on what was the state of the catalog prior to this one. It is not clear, i.e. numbers/statistics/errors, how this catalog compares with previous catalog(s) ? How much improvement do you bring ? You touch on this quickly at the start, but it should be discussed in more details. The description of the catalog (paragraph 3) could also be improved by comparing with previous studies, at the moment there is only a mention of consistency with the authors' own previous studies on Taupo.

This is a good point, and we have added an approximate comparison between the GeoNet catalogue and our absolute location catalogue. However, it is challenging to compare this catalogue in detail with previous work because of the varied location methodologies, velocity models, approaches to calculating magnitudes, etc.. We have added the following paragraph to the Discussion section on lines 186-194:

"The differences in location methodology, velocity model, magnitude calculation, and quality control procedure between our catalogue and the GeoNet catalogue (GNS Science, 2024) mean it is not possible to meaningfully compare the errors in earthquake source parameters. However, we can make some basic comparisons of the two catalogues for the same time period and approximate location. From 2007--2023, GeoNet detected 58,317 earthquakes in the TVZ ("Tongariro & Bay of Plenty" region) compared to the 86,579 earthquakes in the absolute earthquake catalogue presented here. The majority of these additional earthquakes are smaller magnitude earthquakes not included in the GeoNet catalogue. The GeoNet catalogue for this region has a magnitude of completeness of 2.0 and a minimum magnitude of -0.3. In comparison, the magnitude of completeness for our catalogue is 1.2 with minimum magnitude -1.2."

The authors mention several temporary deployments in the introduction but use only two of them for creating their catalog. Can you explain why ? Were they not publicly available ? By curiosity, are there ever been OBS deployments to improve your offshore Bay of Plenty earthquakes ?

Yes the other temporary deployment data was not publicly available at the time of analysis. There have been OBS deployments in the Bay of Plenty, and if this data becomes available they would be a welcome addition.

Also, I understand you only want to show the level of details of your relative relocation, but figure 7 asks for a little more discussion. Moreover, you mention this figure only to say that the catalog can be used for finer scale study but the relocation is worth discussing in more details. How improved is the relocation compared to the absolute locations ? You could also mention it during your description of the catalog (paragraph 3).
We have now included further details on the example illustrated in Figure 7 in lines 198–204: "We illustrate the detail available from the relocated earthquakes in Figure 7, with an example from the Hipaua-Waihī-Tokaanu geothermal field. This geothermal field is expressed at the surface by hot springs, fumaroles, and the Hipaua Steaming Cliffs (Severne, 1999; Risk et al., 2002; Soto et al., 2019), however there has been very little research into earthquake activity here. In the region shown in Figure 7, 12,254 earthquakes are detected in the absolute earthquake catalogue, 8,785 (72%) of which are relocated. The majority of these earthquakes occurred in distinct swarms in 2009--2010, 2015, and 2017 (Figure 5). The relocated earthquake catalogue illustrates that these earthquakes occur in discrete clusters at depths 3–7 km (Figure 7)."

In figure 4 and 5, some values on the axes would be good, even just 2 on each. Particularly to re-use them in future studies.
We have added axes labels for Figure 4, for Figure 5 because there are so many panels it quickly becomes very cluttered with any axes labels, so we prefer not to include them. We hope that by using the exact same projection in other figures which are labelled (e.g., Figure 4, 6) that it is clear enough.

For figures 5, 6 and 7, I suggest removing the black contours around the circles so the EQ are more visible (at the moment some points fade with the topography in B&W). It could be worth checking another colorscale that "pops" more.

Removing the outline of the earthquakes completely makes the individual earthquakes very difficult to distinguish. We have made the outline of the earthquakes thinner and slightly increased the size of the earthquakes.

For figure 4, consider removing a) as it is already repeated in 5. For b) you mention in the legend that you use subregions with more than 20 EQ but the color scale in b) starts at 0. Are yellow regions with 20 or 0 events ?
We prefer to keep panel a as its gives context to panels b-f. We have now changed the color palette for b so that it starts at 20.

Figure 4 again, please move all the colorbars to outside the map.
This has been done.

For figure 5, I would suggest changing the color for the stations to black or a darker color, you want to show off your EQ and the red stations take the stage at the moment.
We have changed the stations to black and made them slightly smaller.

Figure 7, an inset for the position of the subregion would be practical for non Taupo-specialists. For the cross-section, why not show deeper depths ? In the map we can see there are EQ to at least 15 km depth.
We have added an inset to Figure 7, and altered the cross-section to reach down to 15 km. We have also adjusted the depth colour scale (0–30 km) to be consistent with Figures 4–6.

"Event probability was >0.2 and pick probability >0.01" I don't know the algorithm used but those probabilities seem really low, can you explain ? Why not use higher probabilities and less declustering ?
These numbers simply reflect the output probabilities from the EQTransformer model, which can seem quite low. However work by Pita-Silim et al. (2023) shows that there is little relation between output probability and pick quality and the probabilities we use here are acceptable.

"When creating the velocity grid, we allowed for a velocity model tolerance of two seconds." That is a relatively large tolerance, any reason behind this choice ?

As the association stage uses only a 1D velocity model we wanted to have relatively high tolerance to account for the highly three-dimensional nature of the TVZ velocity structure. We are also able to be 'generous' with this tolerance at the association stage as we then apply another stage of earthquake location (NonLinLoc with a 3D model) which excludes erroneously associated picks.

"Finally we set parameters min_node_size to 10, location_split_depth to 10, and location_split_return to 5." If you mention these parameters it would be nice to explain them (and the units associated with those values)

We have added these details to lines 92–94:

"Finally, we set the minimum grid node size (min_node_size) to 10 km, the number of times to subdivide the initial volume (location_split_depth) to 10, and location_split_return value of 5 to determine the size of the new search volume."

"In our area of interest (i.e., within the crust of the TVZ)" did you remove deeper one, why only within the crust ? or where there none ? If you do look only at the crust it should be mentioned and explained earlier. From figure 6 we can see Ruapehu might show deeper events (very interesting).

We have added the definition of our "area of interest" to line 99, we only considered earthquakes shallower than 50 km:

"[…] filtered to preserve only those that were in our area of interest (Longitude: 175◦ to 178◦, Latitude: -40◦ to -37◦, Depth: <50 km) […]"

This was to exclude the underlying subduction zone seismicity (e.g., Mark et al., 2024) as that was not the focus of this study. We have clarified this on lines 37–38:

"Our study is limited to the crustal seismicity of the TVZ (i.e. <50 km depth) and does not consider the underlying subduction zone seismicity (e.g., Mark et al., 2024)."

For the magnitudes calculation/magnitude of completeness, how does it compares with previous values ?

See above for comparison of this catalogue and the GeoNet catalogue including magnitudes.

"At the southern terminus" Southern end would sounds better.
Done.

"Almost the entire thickness of the crust **(**0–30 km**)**"
Done.

"Beneath the andesitic volcanoes Ruapehu and Tongariro**","** seismicity is relatively limited, and when it does occur it tends to **"be"** in the upper 10 km of the crust"
Done.

"Where the majority of the earthquakes are in the mid-crust **(**8–20 km depth**)."**
Done.

"QuakeML [...] ObsPy (Beyreuther et al., 2010)**. It** contains all of the earthquake metadata including phase picks and amplitudes."
Done.

"Alongside this we present the data in a simple CSV format **which** provides [...]"
Done.

REVIEWER B: TRUGMAN

Line 68: “conservative pre-trained model” – pre-trained by whom (Mousavi et al. or Woollam et al.). Are the station types and distance range similar to your dataset?
We have clarified that the conservative pre-trained model refers to a pre-trained model by Mousavi et al., 2020. We now also include a brief discussion of the suitability of the training data for our study on lines 69–71:

"The earthquake dataset used for training is well suited to our study as the majority of earthquakes in the training dataset are shallow (<50 km), have magnitudes across a broad range (-0.5–7.9), and are recorded on three-component seismic data (Mousavi et al., 2019)."

Related comment / question, knowing very little personally about the nature of the seismicity in this area, I could imagine that due to differences in source (nontectonic mechanisms) or path/site (high attenuation and/or extreme topographic and soil effects) the P and S arrivals from these events may be quite distinct from those used to train the ML models. Can you comment on this possibility, and how it could potentially hinder model performance? Perhaps in the Discussion (this could extend your current last paragraph)? Maybe this is not an issue at all, but I am probably not the only reader wondering about this…

This is a good suggestion, we have expanded on our existing discussion of this point on lines 238–245:

"Finally, the *EQTransformer* model we use was trained on a global earthquake dataset that does not include low-frequency seismic events which are known to occur in volcanic regions and have been documented in the TVZ (Hurst and Sherburn, 1993; Park et al., 2019; Steinke et al., 2024). This means that our earthquake catalog will not include these earthquakes. Future work that seeks to detect and catalogue low-frequency seismicity (e.g., Zhong and Tan, 2024) would greatly improve our understanding of these signals in the TVZ, particularly alongside the earthquake catalogue presented here. It is also possible that the geological setting of the TVZ, in particular the high rates of seismic attenuation (Illsley-Kemp et al., 2022), may hinder the performance of the pre-trained model. Future work could train earthquake picking models on an earthquake dataset that is specifically designed for volcanic regions."

Line 57: I assume also that accounting for station elevations explicitly as in NonLinLoc is helpful?

We're not too sure which line this comment is referring to, but we do account for station elevation in the NonLinLoc location, we have added this detail to lines 115-116.

Line 87-88: For those unfamiliar with PyOcto, it would be useful to very briefly describe what these parameters mean.

We have now included a description of these parameters (see above).

Line 111: “fulfil” → “fulfill”

We prefer to use British English throughout, hence fulfil.

Lines 111-113: I imagine the total event count is sensitive to the values used in these criteria. Can you comment on how these criteria were selected and how they impact the resulting catalog? In my experience there are some users that want a highly complete catalog and are willing to tolerate a reduction in location quality, while others require greater accuracy or precision. You could consider publishing two versions of the catalog with greater and lesser QC, or else just one catalog with some categorization about the quality? Totally up to you, but that is my experience…

This is a good point, although something we should have detailed is that we found that this quality control stage was also important for removing any mis-associated earthquakes from large distant earthquakes that had made it through the location filter we applied to PyOcto. Therefore it could be misleading to provide the ‘unfiltered’ catalogue as it would contain these ‘ghost quakes’. We have added these details to lines 121–124:

“These quality filters were selected based on an iterative process which sought to; remove earthquakes which were far outside of the TVZ network, remove any remaining falsely associated earthquakes which had been mis-located inside the network, and retain all well-constrained TVZ earthquakes. This resulted in a total of 86,579 earthquakes, 57% of the earthquakes which were associated by *PyOcto*.”

Line 113: The absolute residual will depend on the distance from source to station: for a very close station a residual of 0.5s is very poor indeed while for a distant station a residual of 0.5s may actually be a fine solution. Have you considered making a distance adjustment for this criterion?

This is a good point, although we emphasise that the constraint we apply is based on the mean residual at all stations used for a location. We found 0.5 s to be suitable for removing very poorly located earthquakes, which often had been misassociated.

Line 122: Does this approach use or require station terms? I could imagine that some of the stations used have extreme site amplification / attenuation effects that would effect the magnitude estimate. It's probably fine for now if not, but maybe just note this limitation for the record here and/or in the Discussion. Alternatively, this could be a nice dataset to generalize the magnitude relation by estimating the station terms.

We didn't solve for or use station correction terms for this calculation. This would be a useful future study when combined with a 3D inversion of attenuation. We have added the following to lines 226–228:

We also did not solve for or use individual station correction terms for this calculation. This could be a useful future addition, potentially combined with a three-dimensional inversion of attenuation.

Line 132: I assume 10km horizontal distance?

No this is true hypocentral distance, i.e. taking earthquake depth into account, we have clarified this in the text.

Figure 2: Consider making this bigger, especially the tick labels but the whole plot is generally hard to read.

Agreed, we have made this figure larger so that the detail can be seen.