

Interactive comment on “TephraNZ: a major and trace element reference dataset for prominent Quaternary rhyolitic tephra in New Zealand and implications for correlation” by Jenni L. Hopkins et al.

Stephen Kuehn (Referee) sckuehn@concord.edu

Received and published: 8 January 2021

The core purpose of this paper is to present and evaluate a set of New Zealand tephra reference data to provide an openly-available framework and database to support tephra-using multidisciplinary research. To facilitate unambiguous eruption identification, samples from key reference locations have been characterized using a consistent, well-described methodology. This dataset is then examined to outline where tephra identifications may still be challenging, even with such consistent data. The paper follows many recommended best practices in tephra data reporting (e.g. including detailed methods, reporting H₂O₂, reporting results on quality control standards), but it could go further. With additional documentation, better data filtering, and better exploration of deposit heterogeneity it could become outstanding example.

Our thanks to Assoc. Prof Kuehn for his thoughtful and valuable comments on our manuscript. The insight into “community best practise” will be invaluable in our revised manuscript.

The authors also discuss the importance of data repositories, yet this dataset appears to exist only as Excel tables currently. Are there specific plans to house the TephraNZ dataset in an open database system that supports data search functions (e.g. Petlab)? The authors discuss the potential for this to be a living data set that expands over time. What plans are there for long-term curation of such an expanding reference dataset? Additionally, interoperability across and linking between data systems is a goal that has been expressed by multiple scientific communities. Careful documentation of data provenance that is recorded in a consistent fashion is one part of achieving this. Unique identifiers like IGSNs for samples and DOIs for papers are another. In volcanology, globally unique IDs exist for volcanic systems and are maintained by the Smithsonian GVP, and these could be incorporated into the dataset. For the Holocene, volcanic eruption and episode IDs also exist.

In its current “in review” status, the data linked to this manuscript are indeed only available through the supplementary file. However, once this manuscript is published, we have in place an online data repository. In addition, all these data will be available through PetLab. We also have a website set up to disseminate information and highlight when new data are added to the set. At present our plan is to update the PetLab data base as we continue to grow the data, but our idea is that this original foundation glass-shard analyses dataset of known tephra act as the “primary” dataset. These will be highlighted on PetLab, and any additional samples added to the set by ourselves or other studies will be labelled as “secondary” if a correlation is made of an unknown/uncorrelated tephra deposit.

Thank you for bringing the Smithsonian GVP number system to our attention. This information has been added into Table 1 for Taupō = 241070 and Okataina = 241050; no others are listed.

Analytical and sample processing methodologies are relatively well-described, but the paper could go further. Be sure to include the detection limits for all elements and how these were determined. Additionally, the supplemental spreadsheet file contains a table outlining the LA-ICP-MS method. Please add a similar table for EPMA. Even better, review Abbott et al. (2020, <https://zenodo.org/record/4075613>) “Community Established Best Practice Recommendations for Tephra Studies-from Collection through Analysis” for the documents

on methodology reporting. Version 1 of these recommendations was released in May, 2020. Version 2 and some worked examples (e.g. <https://doi.org/10.5281/zenodo.4074289> with EPMA methodology detailed) were released in early October.

We thank Assoc. Prof Kuehn for bringing these studies to our attention and have included the guidance given in Abbott et al. (2020a&b).

For EPMA, this includes details such as spectrometer assignments, diffracting crystals, and sequence. As there is not yet a worked example for LA-ICP-MS, the Tephra NZ project could supply the first one for this international tephra community effort. Such a stand-alone methods document with DOI could then also be re-usable and could be linked with all papers and datasets that use that same methodology now and in the future.

Table 2 will be updated for more information as to the EPMA setup and running protocols.

The paper states that a robust data set with geochemical consistency resulting from consistent methodology is necessary for the optimal use of statistical techniques to correlate tephtras. This is because these methods generally do not account for variations resulting from changes in analytical methods over time within a single lab or from biases between different laboratories. Although perhaps ideal, this is not the only approach. Routine use of common reference materials across laboratories can be used to remove most laboratory bias and harmonize datasets, provided that alkali element migration effects (e.g. sodium loss) are sufficiently limited. This is a key reason why the International Focus Group on Tephrochronology and Volcanism (INTAV, recently renamed Commission on Tephrochronology or COT) distributed a set of four reference glasses to more than 25 laboratories for analysis and why a similar effort is currently underway with additional glasses for LA-ICP-MS.

We (Victoria Univ. of Wellington) took part in, and look forward to the publication of the Commission on Tephrochronology (COT) reference glass study, and hope that it can provide a more appropriate reference glass standard for our analyses.

When using the reference glass ATHO-G, it is important to be aware that the published preferred Na₂O concentration of 3.75 wt% (Jochum et al., 2006, Table 13h) is too low. Several analyses with significant sodium loss were not screened out prior to calculating the preferred composition. A majority of newer analyses in the GeoReM database exceed the above reference concentration as do XRF and INAA analyses which are immune to alkali element migration. Kimura et al (2018) also report a concentration of 4.6 by FS-LA-ICP-MS. Together, these data suggest that a Na₂O concentration of 4.3 to 4.6 wt% for ATHO-G is probably more accurate. Using the 3.75 concentration during EPMA standardization or offline correction will bias all results to too low values. Using the same 3.75 concentration when assessing whether Na-loss is sufficiently controlled by an analytical method may also produce an inaccurate interpretation.

We were unaware of this problem with the ATHO-G standard values for Na₂O, and appreciate this issue being brought to our attention. The discrepancy will be in the order of ~ 1 wt% higher than our current values measured for ATHO-G, and we will further investigate this issue. We also do not hang any interpretations on the Na₂O data in these rhyolitic glasses.

Careful removal (but not deletion) of outliers is important before computing summary statistics. The data tables need to be further screened for these. For example, Table 2 contains obviously problematic trace element analyses like line Kaharoa-P_22 which has elevated Zr.

Such points should be set aside and labeled as outliers within the dataset, following recommended practice. Analyses with very low totals should be considered as outliers, labeled, and set aside. Often, a 90% total (10% water and unanalyzed elements) is used as the threshold for this, but that is not appropriate in all cases. Some maximally hydrated tephra yield mean totals of 90-91 wt%, and require a somewhat lower cutoff to allow for the range of analytical precision. If a Cl-bearing epoxy is used, elevated Cl concentrations on points that also have low totals indicate that the beam spot is not entirely on tephra glass. Often, these points will compare well to others when normalized to 100% totals, but the Cl will be inaccurate.

We agree that there remain some outliers in the data that need to be removed, but appreciate Assoc. Prof Kuehn's suggestion they should be "set aside and labelled". Once this further data reduction is undertaken, we will revise the statistical analyses and the figures and tables presented.

We are also aware, as also highlighted by both R1 and R2, that some of the H₂O_b values are 'too high' in the major element data and these will be removed. We agree that in some circumstances a standard cut-off is not necessarily applicable, however, because only ~1.6% of our data are ≥ 8 wt% and this 1.6% exists in a range of samples, from a range of sites, we think this is appropriate for our purpose here.

I am also concerned about the removal of individual data values that fall below the single analysis detection limits, despite this being commonly applied. It is much better to include the actual data values and also report what the detection limits are. This way, information is not discarded, and each end user can determine what is fit for their own purpose.

Fundamentally, analytical techniques like EPMA and ICP-MS are based on signal counting and essentially follow Poisson statistics. Consequently, replicate analyses on a homogeneous material follow a probability distribution. At low enough concentrations, a portion of that distribution will begin to fall below the single analysis detection limit at some selected level of confidence (often chosen as 99% or 3 sigma, but 2 sigma and 1 sigma are valid confidence levels as well). At this point, those specific analyses that fall below the detection limit are no less valid as members of the population than those analyses which fall above the limit. It is therefore completely inappropriate to compute summary statistics like mean, median, mode, or standard deviation when the low end of that population has been removed. Removing those below detection limit values and then computing the mean and standard deviation of those that remain biases the mean to too high a value and reduces the standard deviation, misrepresenting the true precision.

We greatly appreciate this detail from Assoc. Prof Kuehn and will include more details about detection limits within our text.

Another community-recommended best practice is to have all tephra analyses and glass standard analyses linked on an analytical session (run) basis. This is because results on common reference glasses are perhaps the best demonstration of analytical quality and because instrument performance can vary from session to session. In SM Table 4, trace element standard analyses are clearly labeled with the individual runs. Please add equivalent information to the other tables for both standards and samples.

We have split our data into analytical sessions to allow further clarity to the data reduction and standardisation practises.

Where multiple isotopes are measured for the same element, could these be used in combination rather than reporting concentrations from each with assumed isotopic ratios? (Or alternately use the one less prone to interferences, or an average.)

We will provide a single isotope value for the duplicate analyses of trace elements on multiple isotopes. This method is run as a standard in our lab to test for interferences, but the data do not necessarily need to be presented here.

SM Table 2 puts some EPMA and LA-ICP-MS analyses together on the same line. Presumably this means that both were done on the same shards. If so, state this explicitly.

Yes, this is our process. We have added a statement about this in the text.

SM Table 3 Major Standards omits results on VG-A99 despite the paper text indicating that this was run as a secondary (and primary) standard. Please add the data for VGA99 to the table.

Amended

This same table lists some individual analyses with surprisingly low Na₂O concentrations for ATHO-G. What is going on with these? Could these represent cases where the electron beam analyzed the same location twice? These problematic results should be set aside and labeled as such but not entirely removed from the table.

Yes, we believe the low Na₂O values are indeed where the points have not been plotted far enough away and the area has effectively been run twice. We will remove these points from the statistical assessment of the data.

SM Table 3 also includes some plots which apparently show results on reference glasses over time. Please add explanation for these.

Amended

Also, the plots for ATHO-G show some discontinuities and trends that suggest problems with the standardization and/or drift. These results on ATHO-G are worrying. Please explain what is going on here. Do analyses of VG-A99 show the same patterns?

We believe this to be an artefact of re-standardisation after leaving the machine idle for two days. This appears to have only impacted Channel 3 (SiO₂ and Al₂O₃ values), and no other parameters were changed. But, regardless, the data have been standardised separately for these two run sessions.

The text mentions that offline data reduction was used to correct for variability in the VG-568 primary calibration. Are the ATHO-G results in this table the raw or corrected values?

They are corrected values. A note has been made in the table to clarify this.

SM Table 4 – Offset is a percentage. Label it as such for clarity. Relabel 2*Std as 2*StDev for clarity. It may be more useful to compare the offset from reference in ppm to the StDev in ppm as a ratio rather than reporting an offset in % of the reference ppm (or do both). This way, if a result is e.g. 3 standard deviations from reference, this would be more readily apparent.

Amended

Line 192 – The proximal-distal differences mentioned raise the question of how it would be known that these are the same tephra?

Agreed, hence the need to publish these data to allow rigorous investigation for future studies.

Lines 224- – What was the kV? What is total time for an entire analysis, including spectrometer movement and other instrument overhead?

Amended

Lines 232-34 – The text mentions monitoring for drift. Was there drift? How was the drift correction implemented – linear interpolation?

No drift was identified

Line 235 – By “applied to all the data” does this mean just all data lines, or all lines and all elements? Where an element is at low concentration on the reference glass (e.g. MnO, MgO in VG-568), the precision may not be sufficient to apply a reliable correction to other analyses. Additionally, such a correction will make little difference where VG568 is both the primary calibration standard and the offline correction standard for the same element.

We apply the correction to all the elements – we have noted this comment.

Line 237-238 – For older tephra glasses, most of this water can be secondary hydration (water absorbed from the environment post deposition) rather than magmatic. This is acknowledged later in equation 3.

Noted.

Lines 245-255 – Equations 1 and 2 essentially outline a standards-based (reference material based) normalization using a single standard. This could also be done using a consensus of the three EPMA reference glasses (VG-568, VG-A99, and ATHO-G) for even better consistency. Of course, any bad analyses would have to be set aside prior to computing consensus corrections.

Noted.

Line 315 – DFA with cross-validation is another way to test this by looking for high rates of misclassification.

Noted.

Line 321 – “of” instead of “or”

Amended

Lines 347-49 – This is perhaps OK as a rough criterion, but precision also scales with concentration due to signal intensity and counting statistics. At lower concentrations, standard deviations will be smaller in wt%, but the relative standard deviation (as a percentage of the analyzed concentration) will be greater. The reverse is true when going to higher concentrations such that a 4x increase in signal intensity corresponds to a relative standard deviation that is reduced by a factor of 1/2 (i.e. scaling with sqrt of total x-rays recorded). So, the standard deviation on a secondary standard only provides an estimate of what should be expected for a homogeneous sample when the concentrations of the same element are similar. Consequently the stated +/- 0.23wt% homogeneity cutoff for Fe would not be appropriate for a sample with a much higher or lower concentration of Fe than the standard.

We agree and we will edit the text to be more explicit.

Line 364 – Crystallization of biotite is not the only way to affect K₂O concentrations. Does the Shane et al (2008) reference provide evidence for biotite fractionation?

Shane et al. (2008), Shane et al. (2003 – reference added to the text) and Nairn et al. (2004 – reference added to the text), described two different biotite populations each of which could be used to model fractional crystallisation trends from a zoned magma chamber. We have added extra details into the text to clarify this.

Figure 4 – The TiO₂ and MgO appear “quantized” on the plots, probably due to rounding everything to the nearest 0.01 wt% prior to plotting. To avoid artifacts like this, it is often better to carry one extra decimal place.

Amended

Line 415 – Perhaps replace “can be used to maintain” with “may exhibit”

Amended

Line 453 – Not coincidentally, some of John Westgate’s early work in the 1960 and 1970s (e.g. Westgate & Evans 1978) also used the same three elements.

Reference added.

Line 454 – Perhaps replace “presence” with “inclusion”

Amended

Figure 10 – Was Ba not included in the histograms for some reason? Some of the text and later plots point to Ba as a useful discriminator.

Ba is a very useful discriminator, and this attribute can be seen clearly by the PCA graph. We have amended the histograms to reflect this.

Lines 504-505 – Remove “ratios”

Removed

Line 606 – Perhaps replace “acts to effectively reduce the variability of” with “exhibits little variability in”

Amended

Line 607 – Perhaps replace “can be used to maintain” with “exemplifies”

Amended

Line 610 – Replace “causes the” with “exhibits”

Amended

Figure 13 (and supplemental tables) – Where multiple populations are clearly evident as in this figure, it would be useful to clearly identify such populations in the data tables, e.g. as Pop 1, Pop 2. Means and standard deviations should then be computed and reported separately for each major population. Data repositories also need to support archival of such details.

A really good suggestion, thank you - amended

Lines 739-757 – Yes, discovering more tephra layers in marine and lacustrine sections than are known from proximal deposits at potential source volcanoes appears to be common in many volcanic regions. (Multiple examples have been reported for e.g. the Cascades arc too such as at Summer Lake, Oregon.) Therefore, developing more complete records of volcanic events (and also understanding their spatial distribution, timing, and eruptive/dispersal processes) requires the integration of proximal and distal tephra records, ideally into accessible databases.

Agreed, we are so pleased you recognise the value in this research, thank you.

Tables 1-5 – To facilitate data re-use, provide in the supplemental file spreadsheet versions of all of these tables in addition to the versions embedded within the manuscript.

The full excel file with all the Tables and SM Tables will be provided.

Table 1 – Relabel easting and northing as longitude and latitude. Also convert to decimal degrees for simpler presentation and easier reuse. Abbott et al. (2020) noted above also contains recommendations regarding reporting of sample details.

Amended

Table 2 – This table lists acquisition time, but it does not specify which is peak and which is background or whether one or two backgrounds were measured.

Amended