UNIVERSITÉ DE STRASBOURG

5 rue René Descartes
F-67084 Strasbourg Cedex
Tél : (33) 03 68 85 01 17
Fax : (33) 03 68 85 02 91

Ecole et Observatoire des Sciences de la Terre

Jamie Farquharson
farquharson@unistra.fr

Géophysique Experimentale
Institut de Physique de Globe de Strasbourg
Université de Strasbourg
5 rue René Descartes
67084 Strasbourg cedex
France

Review of "*Revisiting the statistical analysis of pyroclast density and porosity data*" by Benjamin Bernard, Ulrich Küppers, and Hugo Ortiz.

I have read the above manuscript with interest. The authors present an evaluation of frequency- and weight-based statistical analyses of pyroclastic deposits, specifically in terms of their density and/or porosity. To facilitate the standardisation of this type of analysis, the authors provide an open-access code (for the freeware statistical package *R*). The method presented has the potential to be of great use in field volcanology and related disciplines. However, I would argue that structural changes to the manuscript should be incorporated, in order to properly show where this work stands in terms of the pre-existing literature, and how it improves upon this. Further, the practical application of this analytical tool should be expounded on, specifically in terms of past and future interpretation of volcanic deposits. I recommend this manuscript is published with moderate revision.

Comments on the manuscript are laid out as general, specific and technical comments. Comments on the figures and figure captions, and the code (*stats.R*) are provided separately.

General comments:

· *R* is a high-functioning software environment with excellent statistical capabilities. Even better, it is freeware, and there are a wealth of support forums and online code repositories, which makes this an excellent platform for the open-source analysis tool presented herein. However, given that an aim of this paper is to promote a standardised analysis by different working groups, perhaps it would be worth providing the tool in other formats? A prior example of this in volcanology is the viscosity calculator of Giordano et al. (2008), who provide their model in both MatLab and Excel formats. Either way, it is worth stipulating in the introduction why *R* is a suitable platform. More detailed commenting on the provided code is recommended to account for potential users not familiar with this particular programming language.

· Section 2.5 - Graphical statistics - should not be a discussion section. Nothing new is presented, and the section concludes with the assertion that these analyses are inferior to the weighting method already established in the previous sections. These seven equations and the accompanying explanation should appear in the context of the introduction as a brief literature review. Thus, the previous work (established standards from the '50s and '60s) should be presented, and their relative strengths and limitations outlined. The authors can then describe how the presented work improves upon the established knowledge base.

UNIVERSITÉ DE STRASBOURG

5 rue René Descartes
F-67084 Strasbourg Cedex
Tél : (33) 03 68 85 01 17
Fax : (33) 03 68 85 02 91

Ecole et Observatoire des Sciences de la Terre

· While the analytical methods outlined in this communication are sound and statistically robust, the discussion is lacking in a few important areas:

- Firstly, the fact that traditional analyses can foster misinterpretation of density/ porosity data, as the authors contend, should comprise a more noteworthy part of the paper. Given this assertion, the logical next step is to assess the interpretations posited by previous authors. The comparative illustration between the two types of analysis (frequency vs. weighted: Figure 3) shows to some extent the difference between the two methods, but this is lacking in context. For example, what does a peak in porosity between 10 and 15% indicate (in terms of eruptive behaviour, dynamics, vesiculation events and so forth)? Does a peak between 20 and 25% (as identified with the new weighted method) indicate significantly different volcanic behaviour? The discussion does not necessarily have to be heavily process-based, but needs to show how and why this new method is beneficial.

- Secondly, the practical use and context of the weighted analysis needs to be described. Essentially, given a population of *n* [clasts, rocks, etc.] for which the volume, mass, density, porosity are known, this method dictates how many subsamples would be necessary in order to ensure that the subset is statistically representative of the population dataset (assuming random sampling of the population). Crucially, this analysis can only be performed on data already collected, *i.e.* it cannot inform how many samples should be collected in the first instance. As a consequence, there appear two main uses for this analysis:

1. Having performed an initial field campaign, the weighted method presented herein could inform the researchers on an ideal number of randomly sampled pyroclasts for a secondary study (e.g. laboratory experimentation; *in situ* testing of rock strength/ permeability/ chemical composition).

2. Establishing a standard for this kind of measurement campaign (porosity/ density measurements of pyroclasts) at different volcanoes and different sites at those volcanoes. The authors give such examples (Chachimbiro, Unzen). Future researchers making measurements at the same sites can thus assess the size required for their (randomly sampled) dataset, in order for their data to be comparable, statistically, to the original data and any future work. Caution is necessary though, as the required number relies on the porosity range and modality: if the mean porosity of pyroclasts changes over time at a given volcano, then the required sample number may also change.

· A sentence which ends in a formula or equation should end with a period, even where the equation is displayed (as opposed to inline). Similarly, if a comma, semicolon etc., would be required were the equation inline, it should be included after the displayed equation.

Specific and technical comments:

· Line 26: Suggest changing to "We propose the incorporation of this analysis into future investigations…".

· Line 46: A measurement is not an analysis. Suggest rewording to avoid this misnomer.

UNIVERSITÉ DE STRASBOURG

5 rue René Descartes
F-67084 Strasbourg Cedex
Tél : (33) 03 68 85 01 17
Fax : (33) 03 68 85 02 91

Ecole et Observatoire des Sciences de la Terre

· Line 46 - 48: This is a slightly confusing sentence, partly due to the punctuation. Perhaps rephrase for clarity and add a little more detail, for example: "In particular, the parameter m_W is nontrivial to constrain accurately due to the process of imbibition, whereby water infiltrates the sample pore space. For low porosity samples this has a relatively small impact over the timescale of the measurement, with a correspondingly greater influence at progressively higher porosities.".

· Line 46 - 54 Further, it is of great importance to indicate that all of these issues are associated with making field measurements. Most published datasets (to date) comprise laboratory-derived values measured on core samples, where the problems of imbibition, irregular shape etc., are circumvented through different means.

· Line 55: Another important aspect of what? Suggest rewording.

· Line 56: Care should be taken with the use of the word "significant": here it seems as though you mean "a large amount", whereas later in the manuscript it has statistical relevance.

· Line 57: "low bias" could be a little obscure for a reader without a statistics background, and is also liable to be misinterpreted: a dataset that is biased towards low values is different to a model or analysis with a low amount of bias (i.e. high variance). Suggest rewording.

· Line 59: It is not the analyses that are interpreted; rather, the analyses allow inferences to be drawn from data.

· Line 61 - 64: These fundamental principles underpin much of this paper, as such they require suitable references.

· Line 66, 213: I'm not convinced *ipso facto* is being used in its proper sense here - usually it means "by that very fact", "in and of itself", or something equivalent.

· Line 86: Consider rewording so that "…a number of measurements *n* must be weighted…", and italicise *n* from hereon in where it appears in the main text (lines 104, 120).

· Line 86 - 88: Justify why the measurement must be weighted by V_P rather than m_P.

· Line 88: For clarity, I think you should emphasise that the representativeness R_P of any given pyroclast is its volume as a proportion of the volume of the entire population. Thus if $n = 1$, then $R_P = (V_P / V_{P)} \equiv 1$. Similarly, $\sum_{i=1}^{n} R_{Pj} \equiv 1$.

· Line 94, 95: Due to the above point, I think brackets should be incorporated into equations (5) and (6), in order to distinguish between the summation and the multiplication. If $\sum_{i=1}^{n} R_{Pj}$ is not distinguished, then $\dot{\rho}_{Vp} = 1 \times \rho_{Pj}$. To avoid this misinterpretation, I suggest presenting them such that:

UNIVERSITÉ DE STRASBOURG

5 rue René Descartes
F-67084 Strasbourg Cedex
Tél : (33) 03 68 85 01 17
Fax : (33) 03 68 85 02 91

Ecole et Observatoire des Sciences de la Terre

$$\dot{\rho}_{Vp} = \sum_{i=1}^{n} (R_{Pj} \times \rho_{Pj})$$

which more clearly shows that the weighted mean density is the sum of all of the clast densities multiplied by their respective representativeness values.

· Line 99: Avoid contractions in academic writing, *i.e.* "don't" should be written "do not".

· Line 111 - 112: You state that "[i]n practice, the bin size should be selected depending on the number of measurements and the density or porosity range", however it is not apparent that the code you present does this. Rather, it seems that the bin size is consistently in 5% porosity increments, regardless of the number or range of measurements.

· Line 113: Indicate why the representation is "unique", *i.e.* in that the shape of the curve is not affected by differing bin sizes, etc.

· Line 121: Missing an article. Change to "Deposits with a large density range and a large standard deviation…".

· Line 137: I think you are referring to Figures 1B and 1D.

· Line 143: Please indicate what the subscript (and those in the following equations) correspond to (*i.e.* the value of ρ or φ at that given vol % cumulative abundance?).

· Line 150: Certainly Robert Folk is going to agree with his own work. Suggest rewording or removing this sentence. See comment for lines 227 - 230.

· Line 164: The equations aren't simplified, as far as I can see. The *R* code serves to automate the analysis.

· Line 167: "as follows".

· Line 177, 178: In both cases the file path is also required, I believe? Further, the end of lines 178 and 181 should be punctuated. See the general comment about the use of *R*, as well as the specific comments on the *stats.R* code.

· Lines 182 - 185: See comments on the code, below.

· Lines 196 - 200: This is an important point, and should be emphasised. Frequency analysis has been used a lot (*e.g.* Kueppers et al., 2005; Belousov et al., 2007; Shea et al., 2010; Mueller et al., 2011; Barker et al., 2012). These few sentences may seem fairly innocuous at first glance, but actually underpin the importance of the type of analysis presented in the manuscript. The fact that misinterpretations can arise in traditional treatment of pyroclast data seems to me one of the critical reasons for "[r]evisiting the statistical analysis of pyroclast density and porosity data". See the general comments.

· Line 215: "analyse" should be "analysis".

UNIVERSITÉ DE STRASBOURG

5 rue René Descartes
F-67084 Strasbourg Cedex
Tél : (33) 03 68 85 01 17
Fax : (33) 03 68 85 02 91

Ecole et Observatoire des Sciences de la Terre

· Line 217: This should also appear in the introduction.

· Line 225: Suggest introducing a comma to improve the flow of the sentence, such that it reads "The porosity distribution for Unzen deposits…for Chachibiro deposits, mostly associated with a larger tail of data and wider porosity modes." Further, some indication of what this means in terms of the pyroclastic deposits would be useful.

· Line 227: Is there sufficient evidence to assert this?

· Lines 227 - 230: Grammatically these sentences are confusing. Suggest changing to "As indicated by Folk (1966), the Folk and Ward parameters generally represent natural distributions (in particular bimodal or polymodal distributions) better than do the Inman parameters. Consequently, the bimodal distribution of most subsamples of the Chachimbiro dataset explains why they appear to be better described by the former than the latter."

· Line 236: Suggest changing "proper" to "improved", "more robust", or something similar.

· Line 237: Indicate the type of datasets (*i.e.* field-derived pyroclast porosity/ density data).

· Line 240: "Better" than what? Indicate that you are referring to traditional characterisation of porosity and/or density data.

· Line 243: Suggest changing to "Finally we propose the use of graphical statistics to present density/porosity data. The differences observed between the two datasets indicate 15 that such representations can be useful to distinguish pyroclastic deposits."

Figures and figure captions:

· Line 260 and Figure 2: Vertical lines, corresponding to the number of samples required at 5% and 1% would be a useful graphical representation of how the required sample number is derived (for intermediate and high stability).

· Line 262: "analyse" should be "analyses".

· Line 264, 266 and Figure 3: Arrow colour is unnecessary (see Wong, 2010; 2011). Solid black arrows would be easier to see in these cases. Further, what are the arrows showing in Figure 3A? The comparison of the two methods (frequency against weighted) seems to indicate that for Unzen pyroclasts there is very little difference observed between either method. While the choice of analysis appears to have a larger influence on the Chachimbiro dataset, the arrows between different subsamples seems slightly misleading. Perhaps the data would be better described in terms of their spread around $x = y$?

· Line 265: "fluctuation" is an inaccurate term in this context. Suggesting changing to "difference" or "differential".

· Line 267: "analysis" should be "analyses".

UNIVERSITÉ DE STRASBOURG

5 rue René Descartes
F-67084 Strasbourg Cedex
Tél : (33) 03 68 85 01 17
Fax : (33) 03 68 85 02 91

Ecole et Observatoire des Sciences de la Terre

· Line 272: On what basis are you contending that the Folk and Ward parameters better distinguish the data? This is by no means clear.

Comments on the *stats.R* code.

· Whilst sparse or informal commenting is fine for coding intended for internal or personal use, given the nature of this paper (the aim of which is to promote the easy and widespread use of this specific code), it is important that the commenting is thorough, widely understandable, and appropriately checked for errors in the English.

· Line 4: "The csv file" is not needed

· Lines 5, 6: Spelling error: should be "columns" and "column", respectively.

· Lines 7, 141: Spelling error: should be "random", rather than "ramdom".

· Line 11: Should be "spans", not "spams".

· Line 26: Should be "samples".

· Line 28: Spelling error: change "desviations" to "deviations".

· Line 30: Spelling error: change "Frequencial" to "Frequential".

· Lines 34, 36: Spelling error: change "weigthed" to "weighted". Also in line 36 you refer to "Vesicularity", which is discussed everywhere else as "porosity".

· Lines 39, 50: The correct spelling is "Inman", rather than "Imman".

· Line 86: Spelling error: change "cointaining" to "containing".

· Line 88: Should read "defined".

· Line 90: Spelling errors: "absolute" and "volume", respectively.

· Lines 107, 137: "volumen" should be "volume" in both cases.

· Line 141: Spelling error: "Ploting" should be "Plotting".

· Lines 115, 124, 126: For clarity, I would suggest putting *pe* in quotes "pe".

· Lines 179, 180, 182: As above, I would suggest putting *st* in quotes "st".

· Line 141: As above, for *pq*.

UNIVERSITÉ DE STRASBOURG

5 rue René Descartes
F-67084 Strasbourg Cedex
Tél : (33) 03 68 85 01 17
Fax : (33) 03 68 85 02 91

Ecole et Observatoire des Sciences de la Terre

· The full references for Inman (1952), Folk and Ward (1957), should be included in the comments, as should be the citation for this paper. The format of online code repositories etc. could result in the code being separate from the associated article, so it is important that it can still be properly referenced (a digital watermark could be implemented to this end).

· It would facilitate interpretation if there was a legend defining the parameters listed in lines 173 and 174 (*i.e.* the list *st*). This already exists in the accompanying Excel file, however as mentioned previously, one should be able to use and understand the code separate from the article.

· It would useful to comment around the calculations for increased transparency. For example (lines 41 - 43):

```
#Inman graphical statistics:
gmedianrho<-grho[5] #Calculates the graphical median density.

gsigmarho<-(grho[7]-grho[3])/2 #Calculates the graphical standard
deviation of density.

gskrho<-(grho[7]+grho[3]-2*grho[5])/(2*(grho[7]-grho[3]))#Calculates
the graphical skewness for density.
```

· Further, "troubleshooting" metadata would be a useful incorporation, to outline solutions to common potential problems with the code (e.g. comma vs. semicolon delimited data, g cm^{-3} vs. kg m^{-3} etc.). Failing this, a correspondence email address would be sufficient.

Literature cited in this review:

Barker, S. J., Rotella, M. D., Wilson, C. J., Wright, I. C., and Wysoczanski, R. J. (2012). Contrasting pyroclast density spectra from subaerial and submarine silicic eruptions in the Kermadec arc: implications for eruption processes and dredge sampling. *Bulletin of volcanology*, *74*(6), 1425-1443. doi: :10.1007/s00445-012-0604-2

Belousov, A., Voight, B., and Belousova, M. (2007). Directed blasts and blast-generated pyroclastic density currents: a comparison of the Bezymianny 1956, Mount St Helens 1980, and Soufrière Hills, Montserrat 1997 eruptions and deposits. *Bulletin of Volcanology*, *69*(7), 701-740. doi: 10.1007/s00445-006-0109-y

Giordano, D., Russell, J. K., and Dingwell, D. B. (2008). Viscosity of magmatic liquids: a model. *Earth and Planetary Science Letters*, *271*(1), 123-134. doi:10.1016/j.epsl.2008.03.038

Kueppers, U., Scheu, B., Spieler, O., and Dingwell, D. B. (2005). Field-based density measurements as tool to identify preeruption dome structure: set-up and first results from Unzen volcano, Japan. *Journal of volcanology and geothermal research*, *141*(1), 65-75. doi:10.1016/j.jvolgeores.2004.09.005.

UNIVERSITÉ DE STRASBOURG

5 rue René Descartes

F-67084 Strasbourg Cedex

Tél : (33) 03 68 85 01 17

Fax : (33) 03 68 85 02 91

Ecole et Observatoire des Sciences de la Terre

Mueller, S., Scheu, B., Kueppers, U., Spieler, O., Richard, D., and Dingwell, D. B. (2011). The porosity of pyroclasts as an indicator of volcanic explosivity. *Journal of Volcanology and Geothermal Research*, *203*(3), 168-174. doi:10.1016/j.jvolgeores.2011.04.006

Shea, T., Houghton, B. F., Gurioli, L., Cashman, K. V., Hammer, J. E., and Hobden, B. J. (2010). Textural studies of vesicles in volcanic rocks: an integrated methodology. *Journal of Volcanology and Geothermal Research,190*(3), 271-289. doi:10.1016/j.jvolgeores.2009.12.003

Wong, B. (2010). Points of view: Color coding. *nature methods*, 7(8), 573-573. doi:10.1038/nmeth0810-573

Wong, B. (2011). Points of view: Color blindness. *nature methods*, *8*(6), 441-441. doi:10.1038/nmeth.1618