Solid Earth Discuss., 7, 1077–1095, 2015
www.solid-earth-discuss.net/7/1077/2015/
doi:10.5194/sed-7-1077-2015

Solid Earth Discussions
Open Access

Revisiting the statistical analysis of pyroclast density and porosity data

B. Bernard[1], **U. Kueppers**[2], **and H. Ortiz**[1]

[1]Instituto Geofísico, Escuela Politécnica Nacional, Ladrón de Guevara E11-253, Aptdo. 2759, Quito, Ecuador
[2]Ludwig-Maximilians-Universität München (LMU), Department of Earth and Environmental Sciences, Munich, Germany

Received: 19 February 2015 – Accepted: 22 February 2015 – Published: 11 March 2015

Correspondence to: B. Bernard (bbernard@igepn.edu.ec)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Explosive volcanic eruptions are commonly characterized based on a thorough analysis of the generated deposits. Amongst other characteristics in physical volcanology, density and porosity of juvenile clasts are some of the most frequently used characteristics to constrain eruptive dynamics. In this study, we evaluate the sensitivity of density and porosity data and introduce a weighting parameter to correct issues raised by the use of frequency analysis. Results of textural investigation can be biased by clast selection. Using statistical tools as presented here, the meaningfulness of a conclusion can be checked for any dataset easily. This is necessary to define whether or not a sample has met the requirements for statistical relevance, i.e. whether a dataset is large enough to allow for reproducible results. Graphical statistics are used to describe density and porosity distributions, similar to those used for grain-size analysis. This approach helps with the interpretation of volcanic deposits. To illustrate this methodology we chose two large datasets: (1) directed blast deposits of the 3640–3510 BC eruption of Chachimbiro volcano (Ecuador) and (2) block-and-ash-flow deposits of the 1990–1995 eruption of Unzen volcano (Japan). We propose add the use of this analysis for future investigations to check the objectivity of results achieved by different working groups and guarantee the meaningfulness of the interpretation.

1 Introduction

Pyroclast density and porosity are commonly used to reconstruct eruptive dynamics and feed numerical models. The pyroclast density ρ_p is defined as:

$$\rho_p = \frac{m_p}{V_p} \quad (1)$$

The mass of a pyroclast m_p is easily measured using a precision balance. The measurement of its volume V_p is a much greater task as pyroclasts have irregular

SED
7, 1077–1095, 2015

shapes. According to the Archimedes' principle, V_p can be calculated using the following equation:

$$V_p = \frac{m_w}{\rho_w} \quad (2)$$

Where the water density ρ_w depends on the ambient temperature and m_w corresponds to the water volume weight displaced by the pyroclast.

If the Dense Rock Equivalent (DRE, ρ_{dr}) is known, either assumed using the rock composition or measured, it can be used along with the pyroclast density to calculate the pyroclast porosity (φ_p):

$$\varphi_p = 1 - \frac{\rho_p}{\rho_{dr}} \quad (3)$$

It is important to note that measuring the density and the porosity of pyroclasts is not a straightforward analysis. In particular, the parameter m_w is hard to constrain precisely as it has to be achieved before or better without a significant portion of the pore space having been filled with water. In any case, the properties of the pore network, such as the permeability or the pore tortuosity, have to be taken into account because they affect the m_w. Over the last decades several methods have been developed to minimize the effect of intruding water (Houghton and Wilson, 1989; Schiffman and Mayfield, 1998; Polacci et al., 2003; Kueppers et al., 2005). It is worth indicating that other methods such as water saturation, caliper techniques, and X-ray tomography are also used to calculate density and porosity (Hanes, 1962; Manger, 1966; Giachetti et al., 2011).

Another important aspect is that pyroclastic deposits commonly present a large range of density values, so sample sets must comprise a significant number of clasts. Additionally, the results must be checked for a low bias during sample selection. Then the density and porosity results are generally treated statistically using frequency analysis including average and distribution histograms. These analyses are often

SED
7, 1077–1095, 2015

interpreted as indicators of volcanic structures or explosivity (Kueppers et al., 2005; Belousov et al., 2007; Kueppers et al., 2009; Shea et al., 2010; Mueller et al., 2011). The main issue in this approach is that density and porosity are considered thermodynamically as intensive properties and cannot be summed or averaged. In fact intensive properties must be weighted in order to be treated statistically.

The purpose of this paper is to present a simple method to obtain weighted averages and histograms in order to analyse density and porosity data. We also propose an ipso facto stability analysis that allows quantifying the quality of the sampling. Then we introduce graphical statistical parameters similar to those used for the analysis of grain-size distribution (Inman, 1952; Folk and Ward, 1957) that can help the interpretation of density and porosity datasets. Those three steps are incorporated in an open source R script (http://www.r-project.org/) for easy use. Finally we illustrate and discuss this method using large datasets from different pyroclastic deposits.

2 Methodology

2.1 Density and porosity datasets

We chose two large datasets from different pyroclastic deposits in order to assess the validity of our approach. The Chachimbiro dataset (Bernard et al., 2014) is made of 32 sample sets from the 3640–3510 BC directed blast from Chachimbiro volcano, Ecuador (Supplement A). Each sample set contains between 15 and 103 clasts of the 16–32 mm fraction measured using the methodology of Houghton and Wilson (1989). The Unzen dataset (Kueppers et al., 2005) is made of 31 sample sets from block-and-ash-flow deposits from the 1990–1995 eruption of Unzen volcano, Japan (Supplement B). Each sample set contains 24–33 large pyroclasts (100–5000 g) measured according to the methodology presented in Kueppers et al. (2005).

SED
7, 1077–1095, 2015

2.2 Weighting measurements

In order to perform a thorough statistical analysis of density and porosity data, each clast measurement in a sample set with a number of "n" measurements must be weighted. On the basis of Eq. (1), it appears that the measurement must be weighted by the volume of the pyroclast V_p. Therefore the representativeness of the pyroclast R_p, which is the part of the measurement in the whole sample set, is calculated as follows:

$$R_p = \frac{V_p}{\sum_{i=1}^{n} V_{pi}} \quad (4)$$

It is then possible to calculate the weighted average density ($\dot{\rho}_{Vp}$) and porosity ($\dot{\varphi}_{Vp}$) as follows:

$$\dot{\rho}_{Vp} = \sum_{j=1}^{n} R_{pj} \cdot \rho_{pj} \quad (5)$$

$$\dot{\varphi}_{Vp} = \sum_{j=1}^{n} R_{pj} \cdot \varphi_{pj} \quad (6)$$

In order to check if the weighting equation is correct, it is possible to solve the Eq. (5) using Eqs. (1) and (4):

$$\dot{\rho}_{Vp} = \sum_{j=1}^{n} \frac{V_{pj}}{\sum_{i=1}^{n} V_{pi}} \cdot \frac{m_{pj}}{V_{pj}} = \frac{\sum_{j=1}^{n} m_{pj}}{\sum_{i=1}^{n} V_{pi}} \quad (7)$$

Therefore the weighted values do have a physical meaning whereas the frequency values do not.

SED
7, 1077–1095, 2015

2.3 Abundance histograms and cumulative plots

Abundance histograms and cumulative plots are typical graphical representations of density and porosity data (Fig. 1). The representativeness can be used to create weighted graphs. For the abundance histogram, in each interval we sum the R_p of the measurements instead of counting the number of measurements and dividing it by n. It is important to note that density and porosity histograms can have different shapes due to the selected bin size (Fig. 1a and c). Several studies have used mixed histograms, with the main axis for density and a secondary axis for porosity (Houghton and Wilson, 1989; Formenti and Druitt, 2003; Belousov et al., 2007; Shea et al., 2010; Komorowski et al., 2013). There is no consensus for the histogram representation; nonetheless most studies used bin sizes between 50 to 100 $kg\,m^{-3}$ for the density (Cashman and McConnell, 2005; Kueppers et al., 2005; Bernard et al., 2014). In practice, the bin size should be selected depending on the number of measurements and the density or porosity range. Cumulative plots (Fig. 1b and d) are easier to produce and have a unique representation. The data are sorted by increasing density or porosity and these values are then plotted against the cumulative abundance that is the sum of R_p. The density and porosity cumulative plots should have the same shape but rotated 180°.

2.4 Stability analysis

One of the main questions when performing a density and porosity analysis on pyroclastic deposits is: how many measurements are required to have a statistically representative sample set? The sample size, here expressed as the number of measurements n, is primarily dependant on the dispersion of the data. Deposits with large density range and large SD require a larger number of measurements. In order to assess the quality of the sampling we propose a stability analysis based on the comparison between the final density average and intermediate density averages. To avoid analytical skew, due to intentional or unintentional ordering of the samples during

SED
7, 1077–1095, 2015

the measurements, the data must be ordered randomly several times. Then the density average is calculated after each measurement and the absolute error with the final density average is determined. The 95th quantile (2 sigma) of the absolute error is then plotted against the number of measurements (Fig. 2). We found that about 1000 repetitive runs on one sample set are required to achieve identical results. Finally, the slope of the curve is calculated below a 5 % threshold of the absolute error to avoid the large error associated to a very small number of measurements. This slope is a direct indicator of the quality of the sampling.

2.5 Graphical statistics

As the frequency analysis is not suitable for density and porosity data, some interesting statistical parameters, such as the standard deviation (SD), are difficult to obtain. Based on studies of grain size (Inman, 1952; Folk and Ward, 1957), we propose a similar approach to calculate the graphical statistics of density and porosity using the cumulative plots (Fig. 1c and d). Here we present the equations for the density, which are identical to the equations for the porosity.

2.5.1 Inman graphical statistics

Inman (1952) defined three parameters:

- The graphical median Md:

$$\mathrm{Md}_\rho = \rho_{50} \tag{8}$$

- The graphical standard deviation (SD) σ:

$$\sigma_\rho = \frac{\rho_{84} - \rho_{16}}{2} \tag{9}$$

SED
7, 1077–1095, 2015

– The graphical skewness SkG:

$$\mathrm{SkG}_{\rho} = \frac{\rho_{84} + \rho_{16} - 2\rho_{50}}{2(\rho_{84} - \rho_{16})} \tag{10}$$

2.5.2 Folk and Ward graphical statistics

Folk and Ward (1957) proposed different parameters that are considered by some authors (Folk, 1966) more representative of natural distributions, in particular for bimodal or polymodal distributions:

– The graphical mean Mz:

$$\mathrm{Mz}_{\rho} = \frac{\rho_{16} + \rho_{50} + \rho_{84}}{3} \tag{11}$$

– The inclusive SD σI:

$$\sigma\mathrm{I}_{\rho} = \frac{\rho_{84} - \rho_{16}}{4} + \frac{\rho_{95} - \rho_{5}}{6.6} \tag{12}$$

– The inclusive skewness SkI:

$$\mathrm{SkI}_{\rho} = \frac{\rho_{84} + \rho_{16} - 2\rho_{50}}{2(\rho_{84} - \rho_{16})} + \frac{\rho_{95} + \rho_{5} - 2\rho_{50}}{2(\rho_{95} - \rho_{5})} \tag{13}$$

– The graphical kurtosis KG:

$$\mathrm{KG}_{\rho} = \frac{\rho_{95} - \rho_{5}}{2.44(\rho_{75} - \rho_{25})} \tag{14}$$

It is important to note that the values of graphical median and mean should be relatively close to the weighted average. Nevertheless, as the weighted average is physically the most accurate value, we propose to use it for graphical representation.

SED
7, 1077–1095, 2015

2.6 R code

An open access R code has been created to simplify the calculations presented above. Additionally it facilitates the automatic creation of abundance histograms, cumulative plots, and stability curves. The input file must be in the format csv (field separated by comma) and structured as follow:

1. first column: pyroclast mass (in kg or g);
2. second column: pyroclast volume (in m^3 or cm^3);
3. third column: pyroclast density (in kg m^{-3} or g cm^{-3});
4. fourth column: pyroclast porosity (in decimal from 0 to 1).

The columns should have a header. All the values must have the decimal point separator for the R code to run properly. The name of the file should correspond to the name of the sample set to avoid confusion when compiling large datasets. The R code is provided in the Supplement C and to run the code only two commands are required in R:

1. load the code: `source("stats.R");`
2. run the code: `results<-stats("Input file name.csv")`

   For large datasets it is possible to create a list of csv files and treat them with a loop:

3. create the list: `l<-list.files(path=".",pattern="csv")`
4. run the code for the list:
   `for (i in 1:length(l)){a<-stats(l[i],plot=FALSE)}`

The R code generates a text file with the statistical results and the figures in pdf format. Compiling the Chachimbiro (33 sample sets, 1492 clasts) and Unzen (32 sample sets,

SED
7, 1077–1095, 2015

922 clasts) datasets with the R code with 1000 runs for the stability analysis of each sample set take respectively 36 and 22 s on a 4 Gb ram computer (~ 42 clasts s^{-1} in both cases).

3 Contribution of the renewed methodology

3.1 Frequency vs. weighted analysis

The absolute difference between frequency and weighted density/porosity averages for Chachimbiro and Unzen datasets is up to 4 and 2 % respectively (Fig. 3a, Supplement D). This difference is not as important as the relative difference between individual sample sets per volcano. To highlight this we chose two sample sets from the Chachimbiro, 021-B and 089-A. These samples have almost the exact same frequency density average (1961 and 1960 kg m^{-3}) but a distinct weighted density averages (2039 and 1892 kg m^{-3}). In contrast, two other sample sets from Chachimbiro (018-C and 095-A) show similar weighted density averages (2246 and 2242 kg m^{-3}) but distinct frequency density averages (2284 and 2154 kg m^{-3}). Abundance histograms can also be biased by the use of frequency analysis. We observed significant modification of the histogram shape such as fluctuation of the density/porosity modes (Fig. 3b), variation of the mode fraction, or change of the general density/porosity distribution (unimodal or plurimodal). Therefore, the use of frequency analysis alone can lead to misinterpretations.

3.2 Sample size

The stability analysis (c.f. Sect. 2.4) can be used to assess the quality of the sampling and also to estimate the minimum number of measurements required to obtain meaningful results. When comparing the slope of the stability curve below the 5 % threshold and the number of measurements from the Chachimbiro dataset, it appears that sample sets with more than 40 clasts have a high stability (Fig. 4, Supplement D).

SED
7, 1077–1095, 2015

Below 40 measurements there is scattering in the results (from high to low stability) probably associated to the difference of SD. The Unzen dataset exhibits a much smaller spread with a high stability for most of the sample sets. This difference indicates that natural heterogeneity of pyroclasts and eruption, transport and deposition dynamics require a deposit-adapted sampling strategy. Houghton and Wilson (1989) propose a minimum of 30 clasts per sample set. Our analysis shows that the minimum number of measured clasts per sample set must be established according to the characteristics of the deposit itself and therefore based on an ipso facto approach. Moreover, the stability analysis might be used to select only high stability samples for further analyse (Fig. 5).

3.3 Distinguishing deposits

Graphical statistics for grain-size analysis have been commonly used to identify the nature of volcanic deposits (Walker, 1971). The same might be applied for density analysis. Figure 5 highlights the differences between the Chachimbiro and Unzen datasets. For values of similar density/porosity averages the Chachimbiro dataset shows almost systematically a higher SD than the Unzen dataset (Supplement D). The two datasets also display a small degree of overlap when looking at skewness and kurtosis parameters. The Unzen deposits have principally a symmetric porosity distribution (SkG and SkI around 0) while the Chachimbiro deposits have a clear asymmetric distribution (SkG and SkI mostly positive and up to 0.4). The porosity distribution for Unzen deposits is typically mesokurtic (KG ~ 1) while it is generally highly leptokurtic (KG > 1) for Chachimbiro deposits mostly associated to a larger tail of data and wider porosity modes. It appears that the Folk and Ward parameters allow for a better distinction than the Inman parameters. This is probably due to the bimodal distribution of most sample sets from the Chachimbiro dataset and agree with Folk (1966) conclusions made for grain-size analysis. It is possible that the distinction made thanks to the graphical parameters is related to the origin of the deposits (directed blast

SED
7, 1077–1095, 2015

vs. block-and-ash-flow) but more data from different deposits are required to support this hypothesis.

4 Conclusions

This study presents a new methodology to treat density and porosity measurements from pyroclastic deposits. It presents weighting equations that allow a proper statistical analysis. The evaluation of Chachimbiro and Unzen datasets indicate that frequency analysis alone can lead to misinterpretations and that weighted analysis should be used to avoid analytical bias. The stability analysis provides a tool to assess the quality of the sampling while the graphical parameters allow for a better characterization of the deposits. The results obtained show that for small numbers of measurements the Chachimbiro dataset is less stable than the Unzen one. This can be interpreted as being due to either the sampling method or due to the deposit density/porosity distribution. Finally we propose to use graphical statistics to represent the density/porosity data. The differences observed between the two datasets indicate that such representations can be useful to distinguish pyroclastic deposits.

The Supplement related to this article is available online at doi:10.5194/sed-7-1077-2015-supplement.

Author contributions. B. Bernard developed the methodology with contribution from all co-authors and prepared the Chachimbiro dataset. U. Kueppers prepared the Unzen dataset. H. Ortiz developed the R code. B. Bernard processed the data and prepared the manuscript with contributions from all co-authors.

Acknowledgements. The authors thank J. Anzieta and S. Hidalgo for useful discussions on the methodology and E. Gaunt for English proof reading. U. Kueppers acknowledges funding from DFG grant KU2689/2-1 that allowed for personal discussion during field work in Ecuador. This

SED
7, 1077–1095, 2015

research as been completed in the framework of the Laboratoire Mixte International “Séismes et Volcans dans les Andes du Nord” which link the IGEPN and three research entities in France including the Laboratoire Magmas et Volcans (Blaise Pascal University, Clermont-Ferrand).

References

Belousov, A., Voight, B., Belousova, M.: 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, B. Volcanol., 69, 701–740, 2007.

Bernard, B., Hidalgo, S., Robin, C., Beate, B., Quijozaca, J.: The 3640–3510 BC rhyodacite eruption of Chachimbiro compound volcano, Ecuador, a violent directed blast produced by a satellite dome, B. Volcanol., 76, 1–20, doi:10.1007/s00445-014-0849-z, 2014.

Cashman, K. and McConnell, S.: Multiple levels of magma storage during the 1980 summer eruptions of Mount St. Helens, WA, B. Volcanol., 68, 57–75, 2005.

Folk, R.: A review of grain-size parameters, Sedimentology, 6, 73–93, 1966.

Folk, R. L. and Ward, W. C.: Brazos river bar: a study of the significance of grainsize parameters, J. Sediment. Petrol., 27, 3–26, 1957.

Formenti, Y. and Druitt, T. H.: Vesicle connectivity in pyroclasts and implications for the fluidisation of fountain-collapse pyroclastic flows, Montserrat (West Indies), Earth Planet. Sc. Lett., 214, 561–574, doi:10.1016/S0012-821X(03)00386-8, 2003.

Giachetti, T., Burgisser, A., Arbaret, L., Druitt, T. H., Kelfoun, K.: Quantitative textural analysis of Vulcanian pyroclasts (Montserrat) using multi-scale X-ray computed microtomography: comparison with results from 2-D image analysis, B. Volcanol., 73, 1295–1309, doi:10.1007/s00445-011-0472-1, 2011.

Hanes, F.: Determination of porosity, specific gravity, absorption and permeability, and details of sample preparation for various other rock studies, Appendix XI in Jet Piercing Research Project, Investigation Report IB 62-27, Dept. Mines and Tech. Surveys, Ottawa, 332–358, 1962.

Houghton, B. F. and Wilson, C. J. N.: A vesicularity index for pyroclastic deposits, B. Volcanol., 51, 451–462, 1989.

Inman, D. L.: Measures for describing the size distribution of sediments, J. Sediment. Petrol., 22, 125–145, 1952.

SED
7, 1077–1095, 2015

Komorowski, J.-C., Jenkins, S., Baxter, P. J., Picquout, A., Lavigne, F., Charbonnier, S., Gertisser, R., Preece, K., Cholik, N., Budi-Santoso, A., Surono: Paroxysmal dome explosion during the Merapi 2010 eruption: processes and facies relationships of associated high-energy pyroclastic density currents, J. Volcanol. Geotherm. Res., 261, 260–294, doi:10.1016/j.jvolgeores.2013.01.007, 2013.

Kueppers, U., Scheu, B., Spieler, O., Dingwell, D. B.: Field-based density measurements as tool to identify preeruption dome structure: set-up and first results from Unzen volcano, Japan J. Volcanol. Geotherm. Res., 141, 65–75, 2005.

Kueppers, U., Varley, N. R., Alatorre-Ibargüengoitia, M. A., Lavallée, Y., Becker, S., Berninger, N., Goldstein, F., Hanson, J. B., Kolzenburg, S., Dingwell, D. B.: 2500 pyroclast puzzle: probing eruptive scenarios at Volcán de Colima, Mexico, EOS Trans., AGU, San Francisco, p. V23C–2072, 2009.

Manger, G.: Method dependent values of bulk, grain and pore volume as related to observed porosity, US Geol. Surv. Bull., 1203, p. 20, 1966.

Mueller, S., Scheu, B., Kueppers, U., Spieler, O., Richard, D., Dingwell, D. B.: The porosity of pyroclasts as an indicator of volcanic explosivity, J. Volcanol. Geotherm. Res., 203, 168–174, doi:10.1016/j.jvolgeores.2011.04.006, 2011.

Polacci, M., Pioli, L., Rosi, M.: The Plinian phase of the Campanian Ignimbrite eruption (Phlegrean Fields, Italy): evidence from density measurements and textural characterization of pumice, B. Volcanol., 65, 418–432, 2003.

Schiffman, P. and Mayfield, J. D.: Measuring the density of porous volcanic rocks in the field using a saran coating, J. Geosci. Educ., 46, 460–464, 1998.

Shea, T., Gurioli, L., Larsen, J. F., Houghton, B. F., Hammer, J. E., Cashman, K. V.: Linking experimental and natural vesicle textures in Vesuvius 79 AD white pumice, J. Volcanol. Geotherm. Res., 192, 69–84, doi:10.1016/j.jvolgeores.2010.02.013, 2010.

Walker, G. P. L.: Grain-size characteristics of pyroclastic deposits, J. Geol., 79, 696–714, 1971.

SED
7, 1077–1095, 2015

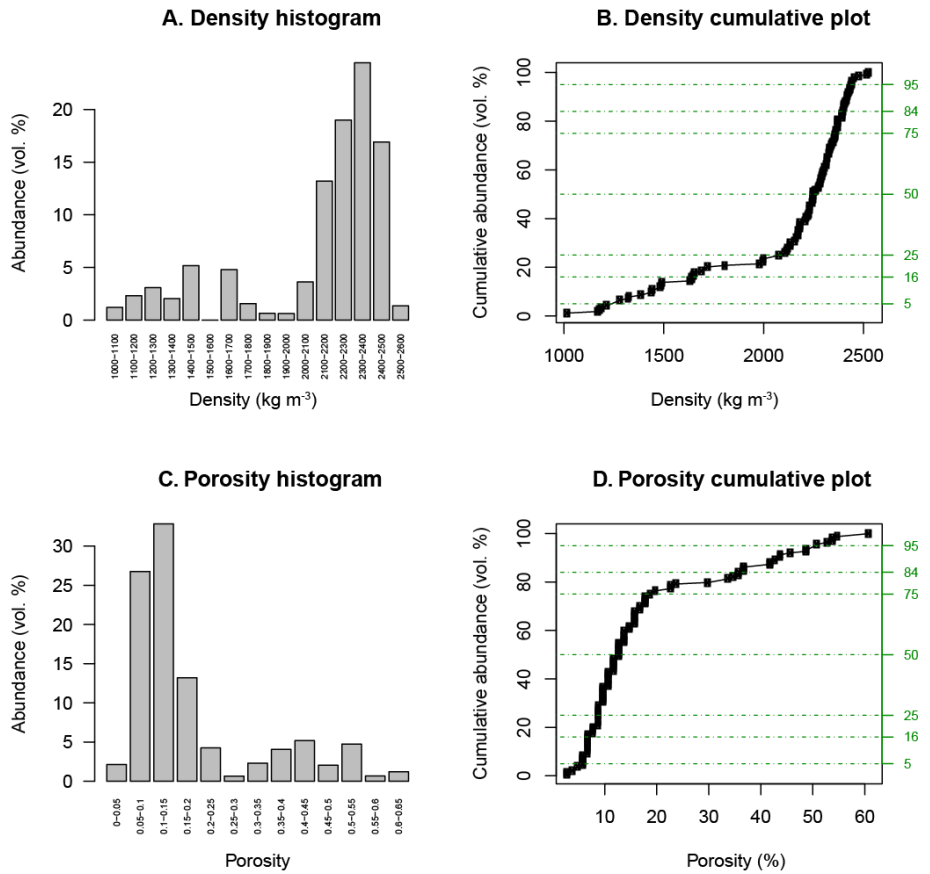

Figure 1. Abundance histograms and cumulative plots for pyroclast density and porosity data. Sample CHA-201-A (n = 103) from Chachimbiro directed blast deposit.

SED
7, 1077–1095, 2015

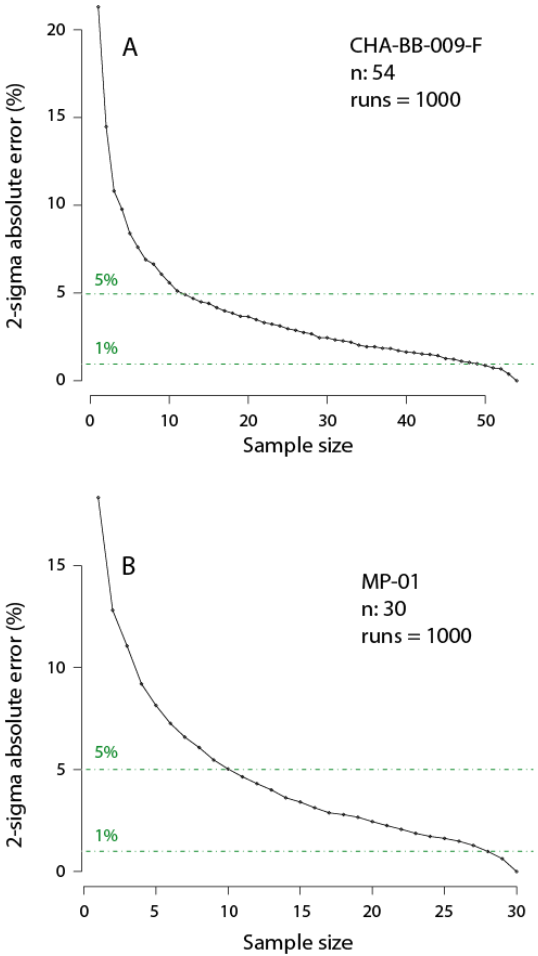

Figure 2. Stability curves obtained after 1000 runs for two samples from Chachimbiro and Unzen datasets. Note the constant slope below the 5 % threshold.

SED
7, 1077–1095, 2015

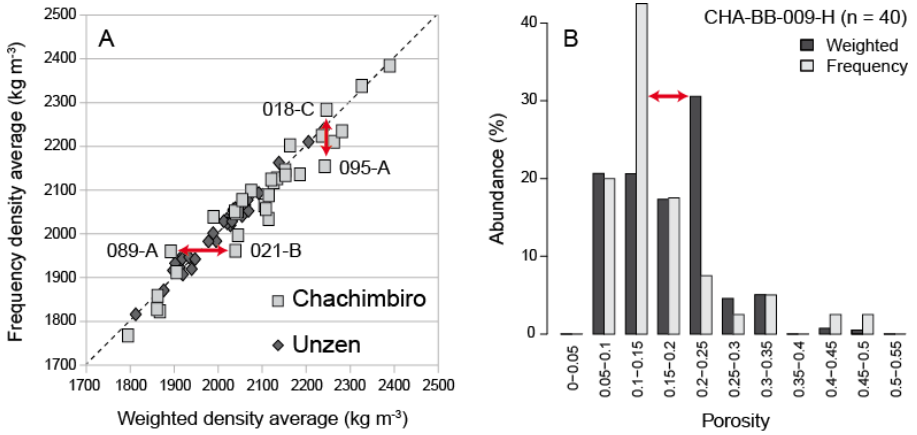

Figure 3. Comparison between frequency and weighted analyse. **(a)** Weighted vs. frequency density average for Chachimbiro and Unzen datasets, note the large relative differences highlighted by the red arrows; **(b)** porosity abundance histogram for one sample from the Chachimbiro dataset, note the large fluctuation (10 %) of the main porosity mode between the two statistical methods represented by the red arrow.

SED
7, 1077–1095, 2015

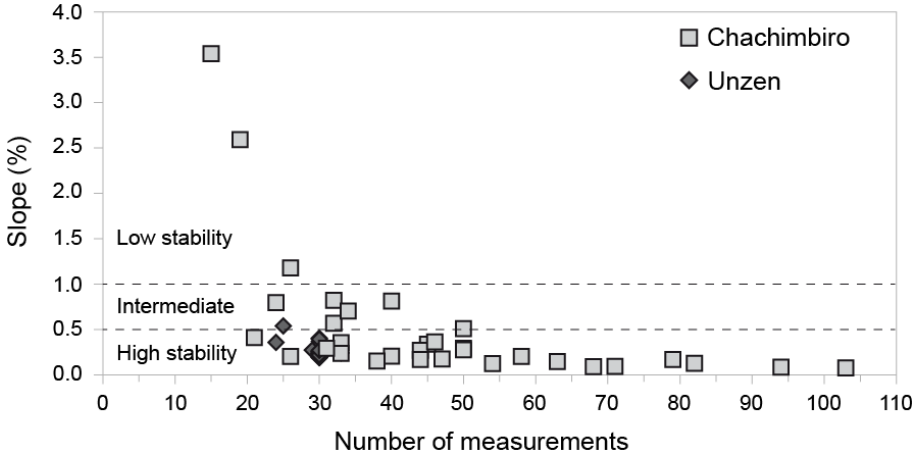

Figure 4. Results of the stability analysis for the Chachimbiro and Unzen datasets. Note that there is a large scattering for Chachimbiro dataset below 40 measurements while the Unzen dataset has much less dispersed values.

SED
7, 1077–1095, 2015

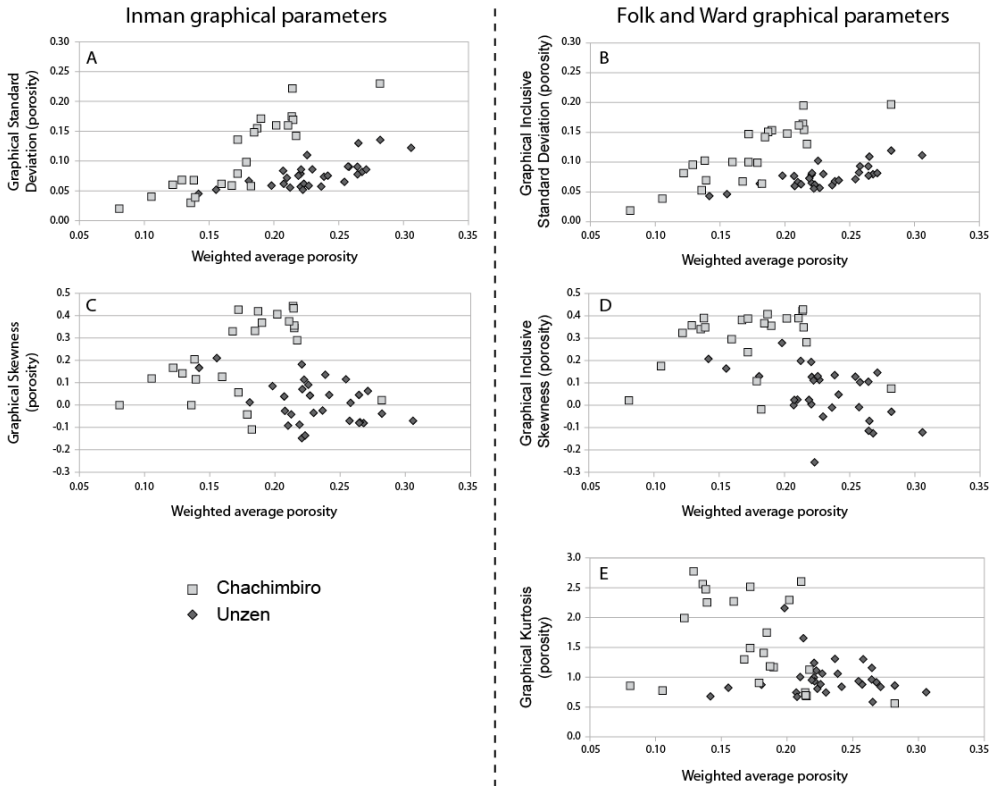

Figure 5. Graphical parameters for the Chachimbiro and Unzen datasets. Only high stability (slope < 0.5 %) sample sets are used in this figure. Note that the datasets are better distinguished using the Folk and Ward parameters.

SED
7, 1077–1095, 2015