TITLE

Revisiting the statistical analysis of pyroclast density and porosity data

AUTHORS

Benjamin Bernard[1*], Ulrich Kueppers[2], Hugo 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 – Germany

* Corresponding author: bbernard@igepn.edu.ec

Journal: Solid Earth

Manuscript type: short communication

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 to statistical methods 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~~ the incorporation of this analysis into future investigations to check the objectivity of results achieved by different working groups and guarantee the meaningfulness of the interpretation.

Keywords: explosive eruptions, pyroclast textures, porosity, density, statistical analysis

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:

$$\rho = \frac{m}{V} \quad (1)$$

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 shapes. According to the Archimedes' principle, V~~$_p$~~ can be calculated using the volume of water displaced by the pyroclast V_w that can be directly measured or calculated using thethe ~~the~~ following equation:

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

~~W~~where the water density ρ_w depends on the ambient temperature and m_w corresponds to the mass of water ~~volume weight~~ displaced by the pyroclast.

If the density DRE (Dense Rock Equivalent ~~(DRE~~, ρ~~$_{dr}$~~$_{DRE}$) is known, either assumed using the rock composition or measured in laboratory (i.e. rock powder density using water or helium pycnometry), it can be used along with the pyroclast density to calculate the pyroclast porosity (φ~~$_p$~~):

$$\varphi = 1 - \frac{\rho}{\rho_{DRE}} \quad (3)$$

It is important to note that measuring the density and the porosity of irregularly shaped pyroclasts is not ~~a~~ straightforward ~~analysis~~. In particular, the parameter m_w is ~~hard~~difficult to constrain ~~precisely~~accurately ~~as it has to be achieved before or better without a significant portion of the pore space having been filled with water~~ due to water infiltration in the pyroclast. The impact on the measurement increases for samples with high porosity and permeability.~~.~~ 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 there are many different techniques to obtain density and porosity ~~other methods~~ such as water saturation, pycnometry (water or helium), photogrammetry, ~~caliper~~calliper techniques, and X-ray tomography ~~are also used to calculate density and porosity~~ (Hanes, 1962; Manger, 1966; Giachetti et al., 2011). The increasing use of regularly shaped ~~core~~ samples (cores) ~~with regular shape in the laboratory solves the problem of imbibition~~ allows for an easy way to derive average density but provides partial information on the bulk density and porosity of the starting pyroclasts~~s~~ due to 3D effects such as heterogeneous vesicle size and density distribution. The purpose of this paper is not to compare the different methods used to obtain the density/porosity data but to discuss how they should be treated statistically.

Another important aspect of density/porosity analysis is that pyroclastic deposits commonly present a large range of density values, so sample sets must comprise a ~~significant~~large number of clasts. Additionally, the results must be checked for a low amount of bias ~~during sample selection~~due to preferential sampling during ~~field work~~fieldwork. Then the density and porosity results are generally treated statistically using frequency analysis including average and distribution histograms. These analyses are often used to interpret~~ed 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~~that- are not additive unlike extensive properties such as mass or volume (White, 2012). In consequence, if it cannot be added, it should not be possible to average (sum divided by number of measurement) intensive properties. ~~For homogeneous material such as native elements (diamond, gold) it is not a problem because the~~Pyroclast d~~-d~~ensity is ~~will be the same independently of the scale~~size dependent even for samples with a homogeneous bubble distribution (increase in density for particles smaller than the average bubble size, e.g., Eychenne ~~et al~~and Le Pennec~~.~~, 201~~3~~2). This effect can be even stronger ~~. F~~for heterogeneous ~~matter such as~~ pyroclastic material that commonly shows bubble gradients. ~~T~~Therefore, ~~an~~the average density ρ_a can be estimated as the total mass of the pyroclasts m_t divided by their total volume V_t:

$$\rho_a = \frac{m_t}{V_t} \approx \frac{\sum_{i=1}^{n} m_i}{\sum_{i=1}^{n} V_i} \quad . \quad (4)$$

The non-additive property of density and porosity also ~~forbid~~limits the use of frequency histograms. For statistical analysis on the density/porosity distribution, the measurements must be weighted ~~cannot be summed or averaged. In fact intensive properties must be weighted in order to be treated statistically.~~adequately to be physically meaningful.

The purpose of this paper is to present a simple method to obtain weighted ~~averages and histograms~~statistics in order to ~~analysze~~analyse density and porosity data. We also propose ~~an *ipso facto*~~ a stability analysis that allows quantifying the quality of the sampling and the relevance of the results. ~~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.~~In order to standardize the description of grain-size distribution of sediments, Inman (1952) proposed a -set of graphical parameters based on statistical analysis. The new parameters such as graphical standard deviation and graphical skewness allowed ~~to put~~putting numbers on descriptive terms. Few years later Folk and Ward (1957) proposed revised parameters that better describe natural material in particular polymodal distributions. They also introduced the kurtosis that allows ~~to describe~~describing the shape of the mode. These parameters have been used ever since to characterize and distinguish volcanic deposits (Walker, 1971). We propose to adapt those equations to describe density and porosity distribution. This methodology is incorporated in an open source R script (http://www.r-project.org/). R is a high-functioning freeware with excellent statistical capacities that provide an optimal platform for such analysis. In order to promote this analysis we also provide ~~Those three steps are incorporated in an open source R script (http://www.r-project.org/) for easy use~~a similar MatLab numeric code. An Excel ~~spreadsheet~~spread sheet is also jointed but only with basic formulae as most of the ~~formulae~~protocol cannot be translate to a ~~spreadsheet~~spread sheet format. -Finally we illustrate and discuss this method using two 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 different outcrops of the 3640-3510 BC directed blast from Chachimbiro volcano, Ecuador (Appendix 1). 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 (Appendix 2). Each sample set contains 24-33 large pyroclasts (~~100-5000 g~~>64 mm) measured according to the methodology presented in Kueppers et al. (2005).

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 *n* must be weighted. Based on the Eq. (1) the density/porosity data can be weighted either by the volume or by the mass of the pyroclast as soon as the weighting parameter, here called the representativeness *R*, is defined as follows:

$$\rho_a = \sum_{i=1}^{n} \left(R_i \times \rho_i \right) \quad . \qquad (5)$$

Here we chose to present the weighting by volume but the same resolution can be used to weight by mass. The Eq. (1) can be reformulate as follows:

$$m_i = \rho_i \times V_i \quad . \qquad (6)$$

Then the Eq. (6) can be inserted in the Eq. (4):

$$\rho_a = \frac{m_t}{V_t} \approx \frac{\sum_{i=1}^{n} m_i}{V_t} = \frac{\sum_{i=1}^{n} \left(V_i \times \rho_i \right)}{V_t} = \sum_{i=1}^{n} \left(\frac{V_i \times \rho_i}{V_t} \right) \quad . \qquad (7)$$

Using the Eq. (5) and (7):

$$\sum_{i=1}^{n} \left(\frac{V_i \times \rho_i}{V_t} \right) = \sum_{i=1}^{n} \left(R_i \times \rho_i \right) \quad , \qquad (8)$$

On the basis of Eq. (1), it appears that the measurement must be weighted by the volume of the pyroclast . Therefore the reprthe representativeness by volume of theany pyroclast R_p, which is the part of the measurement in the whole sample setdefined as the volumetric portion of the pyroclast, is calculated as follows in the whole sample:

$$R_i = \frac{V_i}{V_t} \qquad (9)$$

Therefore if n = 1, R =1.4

$$R_i = \frac{V_i}{V} \qquad (4)$$

4 dotRhoVp) and porosity (dotPhiVp) as follows:

4 (5)

4 (6)

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

5 (7)

Therefore the weighted values do have a physical meaning whereas the frequency values don't.

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 $R_p$representativeness 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[1] (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 practicetheory, the bin size should be selected depending on the number of measurements and the density or porosity range, nevertheless for comparison purpose we chose a constant bin size (100 kg m^{-3} and 5% porosity) that can be changed in the numeric code. Cumulative plots (Fig. 1B and D) are easier to produce and have a unique representation as the data are used directly to produce the plot. 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 representativeness. The density and porosity cumulative plots should have the same shape but rotated 180°[2].

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 set size, here expressed as the number of measurements *n*, is primarily dependant on the dispersion of the data. Deposits with a large density range and a large standard deviation 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 (including all the measurements) and intermediate density averages (including part of the measurements). To avoid analytical skew, due to intentional or unintentional ordering of the samples during the measurements, the data must be ordered randomly several times. ~~Then t~~The ~~density~~intermediate average ρ_{aint} is calculated after each measurement and compared with the final average. ~~the~~An absolute error (AE) is calculated using as follows:

$$AE = \left|\frac{(\rho_a-\rho_{aint})}{\rho_a}\right|. \qquad (10)$$

~~with the final density average is determined.~~ Each run with random ordering leads to a different AE after a certain number of measurements. We chose to represent ~~T~~the 95th quantile (2 sigma) of the ~~absolute error~~AE ~~is then plotted~~ against the number of measurements (Fig. 2). We found that about 1,000 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 with low slopes associated to high quality sampling. The slope of the curve is also calculated below 1% of AE as an additional quality indicator but it ~~seems~~appears ~~not as~~less useful in practice.

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, are difficult to obtain. Based on the work achieved to characterize better ~~studies of~~ grain-~~-~~size distribution (Inman, 1952; Folk and Ward, 1957), we propose for the first time a similar approach to calculate the graphical statistics of density and porosity using the cumulative plots (Fig. 1~~C~~B and D). The main difference between graphical statistics for grain-size distribution or for density data is not the equations but the data itself. Grain-size data obtained through sieving are partial data as the grain-size distribution inside each size class (1 phi, ½ phi or ¼ phi) ~~we cannot~~ is unknown ~~the grain-size distribution~~. The density data, on the other hand, are continuous through the whole sample set. ~~An other difference is that grain-size data are weighted by mass whereas density data are weighted by volume.~~ For informational purpose~~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~~ the Graphical Median Md is a proxy of the average:

- ~~EMBED Microsoft Equation 3.08~~ $$Md_{\rho}=\rho_{50}$$ ~~(8)~~

$$\rho_{Md}=\rho_{50}\quad ;\qquad (10)$$

~~W~~where ρ_{50} corresponds to the value of ρ at 50% of cumulative abundance. Same notation is used for the following equations;~~:~~

- ~~The~~ the Graphical Standard Deviation σ describe the dispersion of the dataset:

- ~~EMBED Microsoft Equation 3.08~~ $$\sigma_{\rho}=\frac{\rho_{84}-\rho_{16}}{2}$$ ~~(9)~~

$$\rho_{\sigma}=\frac{\rho_{84}-\rho_{16}}{2}\quad ;\qquad (11)$$

- ~~The~~ the Graphical Skewness Sk~~G~~ characterize the asymmetry of the data distribution:

- ~~EMBED Microsoft Equation 3.08~~ $$SkG_{\rho}=\frac{\rho_{84}+\rho_{16}-2\rho_{50}}{2\left(\rho_{84}-\rho_{16}\right)}$$ ~~(10)~~

$$\rho_{Sk}=\frac{\rho_{84}+\rho_{16}-2\rho_{50}}{2\left(\rho_{84}-\rho_{16}\right)}\quad .\qquad (12)$$

2.5.2. Folk and Ward graphical statistics

Folk and Ward (1957) proposed different parameters ~~that are considered by some authors (Folk, 1966)~~supposed to be more representative of natural distributions, in particular for bimodal or polymodal distributions. The main difference with Inman parameters is the inclusion of a 1-sigma parameter for the mean and a 2-sigma parameter for standard deviation and skewness. In addition Folk and Ward (1957) included the Kurtosis, a statistical parameter that allows ~~to characterize~~characterizing the shape of the distribution peak:

- ~~The~~ the Graphical Mean Mz:

- ~~EMBED Microsoft Equation 3.09~~ $$Mz_{\rho}=\frac{\rho_{16}+\rho_{50}+\rho_{84}}{3}$$ ~~(11)~~

$$\rho_{Mz}=\frac{\rho_{16}+\rho_{50}+\rho_{84}}{3}\quad ;\qquad (13)$$

- The the Inclusive Standard Deviation σI:
- EMBED Microsoft Equation 3.09 $$\sigma I_{\rho} = \frac{\rho_{84} - \rho_{16}}{4} + \frac{\rho_{95} - \rho_{5}}{6.6}$$ (12)

$$\rho_{\sigma I} = \frac{\rho_{84} - \rho_{16}}{4} + \frac{\rho_{95} - \rho_{5}}{6.6}$$;. (14)

- The the Inclusive Skewness SkI:
- EMBED Microsoft Equation 3.09 $$SkI_{\rho} = \frac{\rho_{84} + \rho_{16} - 2\rho_{50}}{2(\rho_{84} - \rho_{16})} + \frac{\rho_{95} + \rho_{5} - 2\rho_{50}}{2(\rho_{95} - \rho_{5})}$$ (13)

$$\rho_{SkI} = \frac{\rho_{84} + \rho_{16} - 2\rho_{50}}{2(\rho_{84} - \rho_{16})} + \frac{\rho_{95} + \rho_{5} - 2\rho_{50}}{2(\rho_{95} - \rho_{5})}$$;. (15)

- The the Graphical Kurtosis KG:
- EMBED Microsoft Equation 3.08 $$KG_{\rho} = \frac{\rho_{95} - \rho_{5}}{2.44(\rho_{75} - \rho_{25})}$$ (14)

$$\rho_{K} = \frac{\rho_{95} - \rho_{5}}{2.44(\rho_{75} - \rho_{25})}$$ .. (16)

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. Standard deviation, skewness and kurtosis are important parameters that have never been used yet to characterize density and porosity distributions but they are useful.

2.6. R code

An open access R code has been created to simplifyautomate 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 follows:

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 supplementary material (Appendix 3) and to run the code only ~~two~~three commands are required in R:

1) set the Working Directory where the R code and the input file are located: setwd("~/");

~~1~~2) load the code: *source("stats.R");*

~~2~~3) 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~~4) create the list: *l<-list.files(path=".",pattern="csv");*

~~4~~5) 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, 922 clasts) datasets with the R code with 1000 runs for the stability analysis of each sample set take respectively 36 and 22 seconds on a 4 Gb ram computer (~42 clasts/s in both cases). A translation of the R code in MatLab format is also provided in the Appendix 3 as well as a basic spreadsheet including the formulae required to obtain weighted average.~~-~~

3. Contribution of the renewed methodology

3.1. Frequency versus 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, Appendix 4) that is close to the analytical error (<5%) ~~(Fig. 3A, Appendix 4)~~. 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). For both of our study cases, the number of measurements and the number of samples per deposit is large enough for the effect of one method compared to the other to be minimum (few percent of deviation). Even though, laboratory experiments have shown that porosity is one of the main parameters that controls fragmentation during ~~decompression~~explosive eruptions under the presence of bubbles with gas overpressure (Alidibirov and Dingwell, 1996; Spieler et al., 2004). ~~t~~Therefore a change of only few percent of porosity might induce a large error on the calculation of pre-eruptive conditions such as overpressure and fragmentation depth.~~Therefore, the use of frequency analysis alone can lead to misinterpretations.~~ It is difficult to assess the effect of the statistical method based on literature as most of the publications only provide the final density and porosity datasets and not the raw data (mass and volume).

3.2. Sample size

The stability analysis (c.f. 2.3) 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, Appendix 4). Below 40 measurements there is scattering in the results (from high to low stability) probably associated to ~~the~~ differences ~~of~~ in the standard deviation. 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~~. When more raw data are available on different deposits, ~~the~~stability analysis results ~~from this approach~~ could ~~also~~ be ~~use~~used to suggest a minimum number of measurements for future investigations. Moreover, the stability analysis might be used to select only high stability, ergo more representative~~,~~ samples for further analys~~e~~is such as laboratory experimentation or permeability measurements (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 standard deviation than the Unzen dataset (Appendix 4). 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~~with a larger tail of data and wider porosity modes. This might be interpreted as an expression of the outgassing processes in both contexts. The dome collapses, associated to Unzen deposits, probably affected the ~~outer~~upper part of the lava dome that has been ~~homogeneously~~fairly outgassed while the directed blast, associated to Chachimbiro deposit, removed most of the dome in one event, ~~including the highly outgassed carapace~~but also magma from the plumbing system ~~and the internal magma~~ with ~~still a~~ higher volatile content. There is no major difference between the Inman (1952) and ~~It appears that~~ the Folk and Ward (1957) parameters for the Unzen dataset while the Chachimbiro dataset behave differently. In particular the Inclusive Skewness (Fig. 5D) allows for a better distinction ~~than the Inman parameters~~between the Unzen and Chachimbiro datasets. As indicated by Folk (1966), the Folk and Ward parameters generally represent polymodal distribution better than do the Inman parameters. Consequently, the bimodal distribution of most samples from the Chachimbiro deposit explains why ~~they are better described by the former~~the former better describes them than the latter.~~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 vs block-and-ash-flow) but more data from different deposits are required to support this hypothesis.

4. Conclusion

This study presents a new methodology to treat density and porosity measurements from pyroclastic deposits. It presents weighting equations that allow a more robust ~~proper~~ statistical analysis. The evaluation of Chachimbiro and Unzen density/porosity 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 than the classical approach using only averages and histograms. The results obtained show that for small numbers of measurements the Chachimbiro ~~dataset~~ sample sets ~~is~~ are less stable than the Unzen ones. This can be interpreted as being due to either the sampling method or due to the deposit density/porosity distribution. Finally we propose ~~to~~the use of graphical statistics to re~~re~~present ~~the~~ density/porosity data. The differences observed between the two datasets indicate that such representations can be useful to distinguish pyroclastic deposits.

5. Author contribution

BB developed the methodology with contribution from all co-authors and prepared the Chachimbiro dataset. UK prepared the Unzen dataset. HO developed the R code and its translation to MatLab format. BB processed the data and prepared the manuscript with contributions from all co-authors.

6. Acknowledgement

The authors thank J Anzieta and S Hidalgo for useful discussions on the methodology and E. Gaunt for English proof reading. UK acknowledges funding from DFG grant KU2689/2-1 that allowed for personal discussion during ~~field work~~fieldwork in Ecuador. This 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). The authors thank Jamie Farquharson and Thomas Giachetti for their reviews that helped improving the original manuscript.

7. References

Alidibirov M, Dingwell DB (1996) High temperature fragmentation of magma by rapid decompression. Nature 380, 146-149.

Belousov A, Voight B, 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. Bull Volcanol 69:701–740.

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

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

Eychenne J, Le Pennec J-L (2012) Sigmoidal Particle Density Distribution in a Subplinian Scoria Fall Deposit. Bull Volcanol 74(10): 2243-49. doi:10.1007/s00445-012-0671-4.

~~Eychenne~~

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

Folk RL, Ward WC (1957) Brazos river bar: a study of the significance of grainsize parameters. J Sediment Petrol 27:3–26.

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

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

Hanes F (1962) Determination of porosity, specific gravity, absorption and permeability, and details of sample preparation for various other rock studies. 332–358.

Houghton BF, Wilson CJN (1989) A vesicularity index for pyroclastic deposits. Bull Volcanol 51:451–462.

Inman DL (1952) Measures for describing the size distribution of sediments. J Sediment Petrol 22:125–145.

Komorowski J-C, Jenkins S, Baxter PJ, Picquout A, Lavigne F, Charbonnier S, Gertisser R, Preece K, Cholik N, Budi-Santoso A, Surono (2013) 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

Kueppers U, Scheu B, Spieler O, Dingwell DB (2005) 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.

Kueppers U, Varley NR, Alatorre-Ibargüengoitia MA, Lavallée Y, Becker S, Berninger N, Goldstein F, Hanson JB, Kolzenburg S, Dingwell DB (2009) 2500 pyroclast puzzle: probing eruptive scenarios at

Volcán de Colima, Mexico. EOS Trans. AGU. San Francisco, p V23C–2072

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

Mueller S, Scheu B, Kueppers U, Spieler O, Richard D, Dingwell DB (2011) 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

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

Schiffman P, Mayfield JD (1998) Measuring the Density of Porous Volcanic Rocks in the Field Using a Saran Coating. J Geosci Educ 46:460–464.

Shea T, Gurioli L, Larsen JF, Houghton BF, Hammer JE, Cashman KV (2010) 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

Spieler O, Kennedy B, Kueppers U, Dingwell DB, Scheu B, Taddeucci J (2004) The fragmentation threshold of pyroclastic rocks. Earth Planet Sci Lett 226:139–148. doi: 10.1016/j.epsl.2004.07.016

Walker GPL (1971) Grain-size characteristics of pyroclastic deposits. J Geol 79:696–714.

White MW (2012) ~~ADD HERE~~Geochemistry. Wiley-Blackwell eds., 668 p. ISBN-13: 978-0470656686

Figure captions

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

Figure 2. Stability curves obtained after 1,000 runs for two samples from Chachimbiro and Unzen datasets. Note the constant slope below the 5% threshold.

Figure 3. Comparison between frequency and weighted analyses. A: weighted vs frequency density average for Chachimbiro and Unzen datasets, note the large relative differences highlighted by the ~~red~~~~black~~red arrows (see paragraph 3.1 for explanation); B: Porosity abundance histogram for one sample from the Chachimbiro dataset, note the large ~~fluctuation~~difference (10%) of the main porosity mode between the two statistical methods represented by the ~~red~~~~black~~red arrow.

Figure 4. Results of the stability analys~~i~~es 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.

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 two datasets ~~are better~~ show lower superposition with the Folk and Ward parameters than with the Inman parameters, in particular when using the Skewness (5D).