Available online at www.sciencedirect.com
ScienceDirect

Journal of Volcanology and Geothermal Research 164 (2007) 95–111

Journal of volcanology and geothermal research

www.elsevier.com/locate/jvolgeores

Statistical analysis of bubble and crystal size distributions: Formulations and procedures

Alexander A. Proussevitch [a,*], Dork L. Sahagian [b], Evgeni P. Tsentalovich [c]

[a] Complex Systems Research Center, University of New Hampshire, Durham, NH 03824, United States
[b] Department of Earth and Environmental Sciences, and the Environmental Initiative, Lehigh University, Bethlehem, PA 18015, United States
[c] Massachusetts Institute of Technology, Bates Linear Accelerator Center, 21 Manning Rd., Middleton, MA 01949-2846, United States

Received 12 April 2006; received in revised form 7 February 2007; accepted 13 April 2007
Available online 10 May 2007

Abstract

Bubble and crystal size distributions have previously been described only by either exponential or power law functions. Within this limited framework, it has not been possible to characterize size distributions in a fully quantitative manner. We have developed an analytical and computational formulation with which to characterize and study crystal and bubble size distributions (BSD). This formulation demonstrates that all distributions known to date belong to the logarithmic family of statistical distributions. Four functions within the logarithmic family are best suited to natural bubbles and crystals (log normal, logistic, Weibull, and exponential). This characterization is supported by the fact that the power law function widely used for crystal and bubble size analysis is not a statistical distribution function, but rather represents an approximation of the upper regions (larger bubbles/crystals) of the logistic distribution, whose sizes are much larger than the mode.

The coefficients for each of the four logarithmic functions can be derived by 1) best fit exceedance function of the logarithmic distribution, and 2) best fit of the linear transformation of the distribution probability density. A close match of the coefficients derived by the above two methods can be used as an indicator of correct function fitting (choice of initial values). Function fitting by exceedance curves leads to the most accurate statistical results, but has certain strict limitations, including 1) a requirement to rescale the base distribution function; 2) a higher failure rate for function fitting than that for distribution density; 3) uncertainty in observational data error estimates; and 4) unsuitability for visual interpretation. The most productive approach to visualization and interpretation of size distributions is through linear transformation of logarithmic distributions on the basis of probability densities. This also makes it possible to 1) clearly discern bimodal distributions; 2) assess the range of observed objects relative to the full range of the indicated distribution; 3) determine number densities for each mode directly; and 4) integrate to obtain total volume fraction for comparison with available observations. The latter could, in some cases, provide more accurate results than many measurement methods.

Unambiguous definition of Bubble Number Density (BND) must be based on the number of bubbles per melt volume (not number of bubbles per bulk volume), so that like is done with crystals, it can be directly used as an indicator of basic

* Corresponding author.
E-mail addresses: alex.proussevitch@unh.edu (A.A. Proussevitch), dork.sahagian@lehigh.edu (D.L. Sahagian), evgeni@mit.edu (E.P. Tsentalovich).

0377-0273/$ - see front matter
doi:10.1016/j.jvolgeores.2007.04.007

vesiculation processes such that: a) nucleation leads to increase of BND, b) diffusive or decompressive bubble growth keeps BND constant, and c) coalescence decreases BND.

Keywords: bubble size distribution; crystal size distribution; statistical analysis; bubble number density; crystal number density; exceedance; magma chamber processes; eruption dynamics; vesicular lavas

1. Introduction

Bubble and crystal size distributions have been studied for many years (Sahagian, 1985; Marsh, 1988; Toramaru, 1989; Bisperink and Ronteltap, 1992), as they reveal processes that are not directly observable, such as flow in magma chambers and conduits, nucleation events, and the history of parcels of magma throughout the eruption process and in lava flows at the surface (Sahagian et al., 1989; Cashman et al., 1992; Proussevitch et al., 1993; Cashman et al., 1994; Vergniolle, 1996; Marsh, 1998). Previous studies of bubble and crystal size distributions have found that many can be characterized by exponential (Marsh, 1988) or power law (Gaonac'h et al., 1996a) functions. However, these are special cases of more general distributions. The data upon which these studies were based, as well as our own new data from vesicular lavas (Proussevitch et al., 2007-this issue; Sahagian et al., 2002) show that they all fall in the family of logarithmic distributions. In this paper, we develop a more universally applicable formulation and present a methodology for treating all such size distributions.

1.1. Background

Our present generalized analysis is built upon the shoulders of seminal studies that have been conducted in the past. Pioneering work that explored the physics of crystal nucleation and growth dynamics to derive an analytical formulation for crystal size distribution was conducted by Marsh in the late 1980's (Marsh, 1988). This work predicted an exponential distribution for single episode crystal nucleation combined with crystal growth. Coefficients for the distribution functions were directly linked to growth rates. This "classic" formulation has been subsequently applied to numerous studies of bubble size distributions (Sarda and Graham, 1990; Cashman and Mangan, 1994; Cashman et al., 1994; Blower et al., 2003).

In another study, Toramaru developed an analytical formulation for bubble nucleation and growth rates and applied it to a numerical model to predict bubble size distributions (Toramaru, 1989). He imposed eight different initial conditions (e.g. depth, decompression rate, initial dissolved water concentration, etc.) in the model to determine bubble size distribution. However, this mechanistic approach did not lead to any statistical interpretation of distribution functions (although it is evident upon inspection that they are logarithmic distributions). The theoretical results were subsequently applied to a number of vesicular lavas ranging in composition from basalt to rhyolite with the goal of reconstructing the physical conditions and processes within the magma body that led to the observed distributions (Toramaru, 1990).

In a later study, Cashman applied the formulation of Marsh (Marsh, 1988) in an attempt to characterize the bubble size distributions of Kilauean basalts (Cashman and Mangan, 1994). It was not possible, however, to characterize the full distribution because only large bubbles were available for analysis. It was not possible to include the smaller part of the distribution because the methodology of counting bubbles from photographs of their cross-sections could not resolve the small bubbles. As a result, the individual bubbles were larger than the mode of the actual distribution.

A very thorough analysis of power law function was subsequently conducted by Gaonac'h (Gaonac'h et al., 1996a,b; Lovejoy et al., 2004; Gaonac'h et al., 2005). This function can be effectively used to characterize the upper part of the bubble size spectrum (when smaller bubbles are neglected). While this is adequate for it application to a limited part of the distribution, the power law formulation is actually a special case of a log logistic distribution used by statisticians for other applications. We explore log logistic distributions and their application to bubble size distributions in Section 7, below.

More recently, Blower et al. explored evolution of exponential and power law functions to describe the bubble distributions of observed samples (Blower et al., 2001, 2003). They formulated a model for single and multiple nucleation events and subsequent bubble growth and found that the distribution resulting from multiple nucleation events that are common for silicic systems could be characterized by variations in the power function. This is in contrast to the interpretations of Gaonac'h who attributed the distribution to coalescence (Gaonac'h et al., 1996b). In basaltic melts where vesiculation is dominated by a single nucleation event followed by bubble growth, their model demonstrated a likelihood of exponential distributions. These models then were applied to interpretations of nucleation history of some natural samples (Blower et al., 2003).

Each of the studies described above consisted of a theoretical formulation, followed by application to natural samples, so in keeping with that "tradition", we have structured the present paper and its companion (Proussevitch et al., 2007-this issue) in a similar fashion. We use bubble (or crystal) volumes rather than radii throughout our formulation. This is both analytically convenient, as well as being more physically relevant. However, in the text we use diameter for ease of visualization.

1.2. Overview

In order to provide the reader with perspective and a preview of our methods and results, we briefly describe the structure of the paper here. In Section 2, we discuss the key parameter of bubble number density. In order to match its meaning to that of crystal number density so that it can be similarly used for first order characterization of vesiculation processes, it is necessary to define this in terms of number per melt volume, rather than per bulk volume.

In Section 3, we present the basic statistical formulation, pairing the theoretical equations with those to be used for building distribution histograms and other statistical curves. We highlight the importance of exceedance and distribution densities as those are very useful for the analysis. Size distribution and exceedance densities are complementary, and thus can be used together to explore the details of observed bubble or crystal populations. Exceedance does not require binning of data, and thus circumvents any biases created by the vagaries of choosing bin sizes and boundaries. The advantages and disadvantages of both distribution and exceedance perspectives are presented.

In Section 4 we demonstrate that bubble and crystal size distributions can be more generally characterized as logarithmic. We highlight 13 distribution functions within the logarithmic family (Cox and Oakes, 1984), and select four as most useful for application to natural bubble and crystal distributions.

The kernel of the formulation is discussed in Section 5, which covers unit conversion and transformation of logarithmic distributions to linear forms. In Section 6 we demonstrate how to apply linear transformations to the four selected logarithmic distributions, providing a summary of equations for functional analysis. Application of logarithmic distributions to previously studied cases is discussed in Section 7.

In Section 8 we discuss the use of distribution functions to calculate bulk vesicularity of natural samples. In some cases, this approach can be more accurate than physical measurement of the bulk vesicularity because a few observed large bubbles can skew the vesicularity significantly, and sample size is rarely sufficient to include statistically meaningful numbers of these large bubbles (or crystals). In the final section (9), we present a new technique for function best fit analysis, rarely attempted in volcanology (Sarda and Graham, 1990).

2. Spatial aspects of bubble size distributions

Bubble number density (BND) is a key aspect of vesicular volcanic rocks, as it can provide insights regarding vesiculation processes. However, it is important to define BND in terms that are relevant to these processes.

Historically, crystal number density (CND) was introduced before BND with an original purpose to characterize crystal nucleation (Brandeis et al., 1984). CND was defined as the number of crystals per unit volume of vesicle-free rock. Since molar volume of mineral components in crystals and melts are about the same, crystal growth does not change CND. Thus CND characterizes the number of crystals nucleated in the system regardless of their sizes. In order to keep a similar meaning for BND, it must be defined as *number of bubbles per melt volume* (not bulk volume, as traditionally used) because, unlike crystals, the molar volume of dissolved volatiles is much smaller than that of bubbles/gases (Table 1). It is important to define BND

Table 1
Differences in definition of crystal and bubble number density

"Traditional" CND	$CND = \frac{N_{\mathrm{crystals}}}{\upsilon_{\mathrm{bulk}}}$	$\upsilon_{\mathrm{bulk}} \approx$ const as crystals grow because $v_\mu^{\mathrm{crystals}} \approx v_\mu^{\mathrm{melt}}$ where v_μ is molar volume
CND in this work	$CND = \frac{N_{crystals}}{\upsilon_{melt+solids}}$	It is used for bubble free samples so that $\upsilon_{\mathrm{bulk}} = \upsilon_{\mathrm{melt+solids}}$
"Traditional" BND	$BND \neq \frac{N_{bubbles}}{\upsilon_{bulk}}$	$\upsilon_{\mathrm{bulk}} \neq$ const as bubbles grow because $v_\mu^{\mathrm{gas}} \neq v_\mu^{\mathrm{melt}}$
BND in this work	$BND = \frac{N_{bubbles}}{\upsilon_{melt+solids}}$	$\upsilon_{\mathrm{melt/solids}} \approx$ const as bubbles grow because $v_\mu^{\mathrm{gas}} \gg v_\mu^{\mathrm{melt/solids}}$ where v_μ is molar volume

In order for CND and BND to be used to gain insights on initial nucleation density, they must be independent of crystal/bubble growth history. Therefore, the denominators above must remain constants throughout crystal/bubble growth (or dissolution). In practical use, υ_{bulk} in CND is close to $\upsilon_{\mathrm{melt+solids}}$ in the absence of bubble growth.

in this way so that it remains invariant as bubbles grow (in the absence of additional nucleation or coalescence) due to decompression or diffusion of volatiles into the bubbles (Klug et al., 2002; Gurioli et al., 2005).

BND defined by melt volume is thus relevant to vesiculation processes such as nucleation (increases BND) or coalescence (decreases BND), regardless of individual bubble growth.

3. Statistical formulation

We now derive the basic formulation and couple it to discrete forms of the equations that refer to bin or individual bubble indices (subscripts i). The discrete forms are helpful guidance for a practical analysis of actual samples. Table 2 summarizes variable notations used.

Table 2
Notation list used in statistical formulations

Notation	Description	Units
BSD	Bubble Size Distribution	
BND	Bubble Number Density	
CND	Crystal Number Density	
$f(V)$**	Population density by bubble/crystal size in volume units*	m^{-3}
$\hat{f}(V)$	Distribution density by bubble/crystal size in volume units	m^{-3}
i (subscript)	Bin or individual object index (order number)	
n_V	Number of objects (bubbles) of arbitrary size V*	None
n'_{total}	Observed total number of objects (bubbles) in a population*	None
n_{total}	Total number of objects (bubbles) in a population*	None
N_{total}	Total number density (all bubbles) per unit volume of melt	m^{-6}
$n(V)$	Number density (bubbles with size V) per unit volume of melt	m^{-6}
$N(V'>V)$	Exceedance density	m^{-3}
$N(V'>V)$	Raw (observed) exceedance density (not adjusted)	m^{-3}
P	Probability, range of values is from 0 to 1	None
S	Exceedance or complimentary probability as $S=1-P$	None
S'	Raw exceedance (not adjusted)	None
V	Volume as a measure of bubble/crystal sizes	m^3
υ_m, υ_g	Melt+solids and gas/void volume in a sample*	m^3
ΔV	Bin width of bubble/crystal sizes in volume units	m^3
v_μ	Molar volume	
v, α	Coefficients of linear distributions (Table 4)	
k, μ, $\hat{\mu}$	Coefficients of logarithmic distributions (Table 4)	

* Indicates extensive (proportional to sample volume) quantities and functions as opposed to others (intensive) where applicable.
** Here and below in the table (V) or ($V'>V$) indicates a function of bubble/crystal size measured in volume units.

Population Density is the total number of bubbles in each size range in the sample. The first step in constructing a histogram of the data is binning, in which size ranges (bins) are defined (by bubble volume), and bubbles are placed within the appropriate bin. The analytical form of population density, $f(V)$ as a function of bubble sizes V, for continuous distributions is

$$f(V)=\frac{n_V}{\mathrm{d}V} \tag{1}$$

where n_V is number of bubbles of size V. From Eq. (1) it follows that a population density histogram can be built by dividing the number of individuals in each bin n_i by the range of bubble sizes ΔV in a bin, so that

$$f(V_i)=\frac{n_i}{\Delta V} \tag{2}$$

It is important to note that any reasonable alteration of bin width ΔV leaves the distribution density invariant for the histograms built by Eq. (2). Since the population density is an extensive function that depends on sample size, bubble "concentration", and size distribution, it can not be directly used in this analysis. However, its components can be readily addressed (below).

Distribution Density is population density divided by total number of bubbles in a sample. The area under the distribution density curve is always equal to 1. Sometimes referred to as probability density in the literature, it is one of the most fundamental statistical parameters that can be readily found from observed population density $f(V)$. An obvious benefit of obtaining distribution density from population density is that it can be tested and matched against known analytically defined distribution functions such as log normal, etc. By definition, the total area under distribution density curve is always equal to unity, so to get distribution density $\hat{f}(V)$ from $f(V)$ we have to normalize the latter by total number of objects (bubbles) in a population which is $n_{\mathrm{total}}=\int_0^\infty f(V)\mathrm{d}V$. Then distribution density follows from population density as

$$\hat{f}(V)=\frac{\mathrm{d}P}{\mathrm{d}V}=\frac{1}{n_{\mathrm{total}}}\frac{n_V}{\mathrm{d}V}=\frac{1}{n_{\mathrm{total}}}f(V) \tag{3}$$

where P is probability of observing a bubble between size 0 and size V ($P_{V\Rightarrow\infty}=1$), and V is a measure of bubble size. Based on Eq. (3) a histogram of distribution (probability) density can be constructed using

$$\hat{f}(V_i)=\frac{1}{n'_{\mathrm{total}}}\frac{n_i}{\Delta V} \tag{4}$$

where n'_{total} is observed bubble population count, and ΔV is a bin width in the histogram. Eq. (4) is valid only if the condition of $n'_{total} \approx n_{total}$ is met, which it rarely is true since the range of observed bubble sizes often does not reflect the actual range of bubble sizes due, for instance, to insufficient instrument resolution, or small bubble population (see discussions in Section 5 and illustration in Fig. 5). In fact, it is commonly difficult to know whether small and/or large bubbles are missed in observations. If any of these are missed, then the total number of bubbles in the population n'_{total} cannot be determined accurately, and so calculations based on Eq. (4) will be in error.

Cumulative bubble number density (N_{total}) or "concentration" is just simple normalization of total number of bubbles of all sizes to volume of melt in a sample. As we explained in Section 2, N_{total} should involve normalization per unit volume of melt+solids (υ_m). So it can be expressed as

$$N_{total} = \frac{1}{\upsilon_m}\int_0^\infty f(V)\mathrm{d}V = \frac{n_{total}}{\upsilon_m} \tag{5}$$

Therefore,

$$f(V) = n_{total}\,\hat{f}(V) = \upsilon_m N_{total}\,\hat{f}(V) \tag{6}$$

The difference between n_{total} and N_{total} with regard to distribution density is that the former is an extensive quantity that depends on sample size, while the latter is an intensive quantity normalized to melt volume which remains essentially invariant during nucleation and bubble growth.

Number density is "concentration" of bubbles of certain size, V, in a sample. Number density can be calculated by dividing population density by melt volume. This makes it an intensive function that does not depend on the sample size, nor on time, as bubbles grow due to decompression or diffusion and exsolution of dissolved gas from the melt. Number density function $n(V)$ is particularly informative because it combines information about statistical distribution of bubble sizes $\hat{f}(V)$ and their total density in the sample N_{total}. $n(V)$ relates to distribution and population densities as

$$n(V) = \frac{1}{\upsilon_m} f(V) = N_{total}\,\hat{f}(V) \tag{7}$$

where V is a measure of bubble sizes by volume. As follows from Eq. (7), it could be interpreted as local bubble number density, i.e. in comparison to cumulative number density N_{total} (compare Eqs. (5) and (7)). Consequently, for number density histograms it takes the form

$$n(V_i) = \frac{1}{\upsilon_m}\frac{n_i}{\Delta V} \tag{8}$$

where υ_m melt+solids volume in a sample, and ΔV is a bin size in the histogram. Number density $n(V_i)$ should not be confused with a parameter $N_{V\bar{i}}$ (particle population in a bin i per unit volume) that is commonly used in statistical stereology (Sahagian and Proussevitch, 1998) since the value of the latter depend on bin size ΔV while $N(V_i)$ does not.

Multimodal distributions are treated as a superposed mixture of monomodal distributions (7) so that

$$n(V) = \sum_j N_{j_{total}}\,\hat{f}_j(V) \tag{9}$$

where j is mode number and $N_{j_{total}}$ is cumulative number density for mode j so that $N_{total} = \sum_j N_{j_{total}}$ and $\hat{f}_j(V)$ is distribution density function for mode j. Most multimodal distributions are bimodal rendering Eq. (9) in the form

$$n(V) = N_{1_{total}}\,\hat{f}_1(V) + N_{2_{total}}\,\hat{f}_2(V) \tag{10}$$

If distribution densities of multimodal populations are of the same type (e.g. normal) and closely superposed, then it is possible to use moment recursion relations, leading to fewer parameters than in the mixture (Cobb et al., 1983), but this problem is beyond the scope of this paper.

Number density function is one of the most fundamental characteristics of observed bubble populations, and it relates to distribution density $\hat{f}(V)$ with a very simple scaling relation (factor of N_{total}) shown in Eq. (7). Distribution density, in turn, is necessary to match observations against known distribution functions in order to answer the question "what type of distribution characterizes the observed population?". This question will be addressed in the next section and answers will be tested in the following companion paper (Proussevitch et al., 2007-this issue).

Exceedance $S(V)$ is a statistical parameter also known as the survival function (Connor et al., 2003), and can interpreted as a complementary probability function $(1-P)$. Exceedance is the fraction of observed objects (bubbles) of size larger than V. We use it instead of probability since it has been used implicitly in recent studies of volcanic bubbles (Gaonac'h et al., 1996b; Blower et al., 2001). In these studies, parameter $N(V'>V)$ was referred to as *exceedance density* such that

$$N(V'>V) = N_{total}S(V) \tag{11}$$

As a complementary function to probability, exceedance is defined as

$$S(V) = \int_V^\infty \hat{f}(V)\mathrm{d}V \tag{12}$$

In both Eqs. (11) and (12) V is a measure of bubble sizes by volume. So exceedance and exceedance density graphs of observed bubble or crystal populations can be constructed using

$$S'(V_i) = \frac{n'_{\mathrm{total}} - i}{n'_{\mathrm{total}}} \tag{13}$$

$$N'(V'>V) = \frac{n'_{\mathrm{total}} - i}{\upsilon_m} \tag{14}$$

where i is the index (order number) for sorted bubbles with ascending sizes (volumes V), so that in accordance with Eq. (13) exceedance for the smallest bubble/crystal is 1 and for the largest one is zero, and similarly exceedance density for the smallest bubble is cumulative bubble number density N_{total} (compare Eqs. (5) and (14) with $i=0$, provided $n'_{\mathrm{total}} \approx n_{\mathrm{total}}$) and for the largest one is zero. From Eq. (14) it also follows that exceedance density $N(V'>V)$ is an intensive quantity.

Exceedance is widely used in statistical analysis primarily because it does not involve histogram binning, so the investigator does not need to choose bin size and binning space (a potentially subjective process). Exceedance curves are typically smooth and can be constructed over wide ranges of observed values (analytical problems introduced by zero objects in any bin are circumvented). They therefore have certain advantages for analytical fitting of distribution functions (with some restrictions indicated below). Calculation of exceedance is prone to the same problem as probability density with regard of n'_{total}. Because n'_{total} can never be known with perfect confidence, we denote exceedance and exceedance density as "raw" or "observed" S' and $N'(V'>V)$ in our notation by using an apostrophe. In contrast to the case of probability density, this does not introduce a fatal problem for exceedance because when small and large objects (bubbles) are not detected in the population we can still write a relation between "true" exceedance $S(V)$ and "observed" $S'(V)$ by simply rescaling such that

$$S' = \frac{S - (S)_{V=V_{\max}}}{(S)_{V=V_{\min}} - (S)_{V=V_{\max}}} \tag{15}$$

where S' is "observed" exceedance, and S is "true" exceedance to be found using a best fit function, and $V_{\min}$ and $V_{\max}$ are sizes (volumes) of the smallest and largest bubbles in the observed population. Eq. (15) rather than the seemingly definitive form (13) must be used for distribution function fitting (discussed below). The immediate benefit of this is that the range $(S)_{v=v_{\min}} \div (S)_{v=v_{\max}}$ is calculated mathematically by best fit function minimization rather than by subjective truncation.

3.1. Summary of basic statistical formulations

The formulations described above address the issue of initial data processing in preparation for determining an analytical distribution function that best fits the observed data. In the past, theoretical distributions functions have been expressed in terms of probability densities, because observational data have been most commonly pre-processed as distribution density. However, in the course of our analysis, we have found that distribution density histograms (Eq. (4)) should be avoided (since n'_{total} represents observed but not actual population) and that number densities (Eq. (8)) should be used instead. Exceedance curves are a useful means for representing observational data in any attempt to identify a best fit distribution function, because they are free of bias related to subjective choice of bin sizes for constructing histograms.

4. Logarithmic family of continuous distribution functions

The size distributions produced by previous studies of bubble populations in volcanic rocks can be characterized by various members of *the logarithmic family of distributions*. In general, logarithmic distributions can be distinguished by two primary features —

1. The range of volumes between small and large objects in the population covers several orders of magnitude (at least 6 orders of magnitude for the volcanic bubbles we have analyzed.
2. Distribution density varies also within several orders of magnitude between small and large bubbles (about 4 orders of magnitude for volcanic bubbles we have analyzed). This leads to an exponential exceedance curve.

Other examples of natural objects that belong to the logarithmic family of distributions are crystals in magmatic rocks, river basin sizes, islands, star brightness, city/town population, etc.

There are at least 13 analytical functions within the logarithmic family, (Cox and Oakes, 1984, page 17).

Table 3
List of distribution functions from the family of continuous logarithmic distributions that are most suitable for bubble and crystal populations

Distribution	Notes	Previously used for vesicles
Log Normal	Most common distribution	Some in (Sarda and Graham, 1990)
Log Logistic	Similar to Log Normal but with added versatility in asymmetry and skewness	Asymptotes only (Gaonac'h et al., 1996b)
Weibull	Highly asymmetric	No, but see Exponential
Exponential	Special case of Weibull distribution	Yes, based on (Marsh, 1988)

Most of these are very specialized and only two have been previously tested in application to volcanic bubbles. Some of the functions do not have analytical forms for both exceedance and distribution density and thus cannot be readily tested by fitting observed bubble populations. Others have more than 3 coefficients that cannot be physically interpreted, rendering them difficult to fit to observed data. On this basis, we have found that there are four that are best suited to application to vesicular basalts and comparable distributions such as crystals, stars, etc. These are indicated in Table 3.

An additional useful logarithmic function is the *Gamma distribution* —

$$\hat{f}(t) = \mu^{-1}\left(\frac{t}{\mu}\right)^{k-1}\frac{\exp\left(-\frac{t}{\mu}\right)}{\Gamma(k)} \tag{16}$$

where $k>0$ is a gamma function index, $k\mu$ is mean and $k^{-1/2}$ is variance. The gamma distribution is analytically expressed only in terms of distribution density, and cannot be explicitly integrated to provide exceedance or converted to a linear scale except for few special cases in which k is a small integer (1 or 2). Integration and differentiation operations must generally be done numerically. Due to these limitations we have not developed the gamma distribution in our explorations of bubble populations.

5. Transformation of logarithmic distributions to linear forms

While logarithmic distributions can be used to accurately characterize bubble (or crystal) populations, it is difficult to display them in forms that lend themselves to simple visualization and interpretation. The statistical and physical meaning of their coefficients and moments are not always obvious, and they cannot be readily associated with the physical processes that generate the distributions. Because the distribution density of logarithmic functions monotonically decreases with bubble size, it is impossible explicitly define multimodal distributions or display the modes, asymmetry features, etc. Consequently, logarithmic functions are not well suited to identifying the details, multiple modes, and other small features of populations.

This shortcoming of logarithmic distributions can be overcome by transforming observed distributions to logarithmic scale and treating them with a linear analog of the logarithmic distribution. This enables one to visualize all the features that are otherwise masked by logarithmic distributions. The basic approach to accomplish this is relatively straightforward-one merely converts all measurements to their logarithms. For the case of bubbles, all units become logarithm of volume. *However, it is important to note that transformation of the distribution density from log to linear cannot be done by simply substituting the dimensional variable (t) with its logarithm (log t).* For example, in the case of a distribution density function, since

$$\hat{f}(x) = \frac{\partial P(x)}{\partial x}\mathrm{d}x \tag{17}$$

substitution of $x=\log(t)$ leads to the following distribution density relations between logarithmic scale (in t) and linear scale (in x).

$$\hat{f}^{\log}(t) = \frac{1}{t}\hat{f}^{\mathrm{lin}}(log\ t) \tag{18}$$

and

$$\hat{f}^{\mathrm{lin}}(x) = \exp(x)\hat{f}^{\log}(\exp \mathrm{x}) \tag{19}$$

where superscripts “lin” and “log” refers to linear and corresponding logarithmic distribution. If we illustrate this transformation for a classical normal distribution in linear scale,

$$\hat{f}(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-v)^2}{2\sigma^2}\right] \tag{20}$$

then its logarithmic analog is a log normal distribution density that follows after substitution of Eq. (20) into Eq. (18)

$$\hat{f}(t) = \frac{k}{t\sqrt{2\pi}}\cdot\exp\left[-\frac{1}{2}k^2\ln^2\left(\frac{t}{\mu}\right)\right] \tag{21}$$

where coefficients between Eqs. (20) and (21) relate as $\mu=\exp v$, and $k=\frac{1}{\sigma}$.

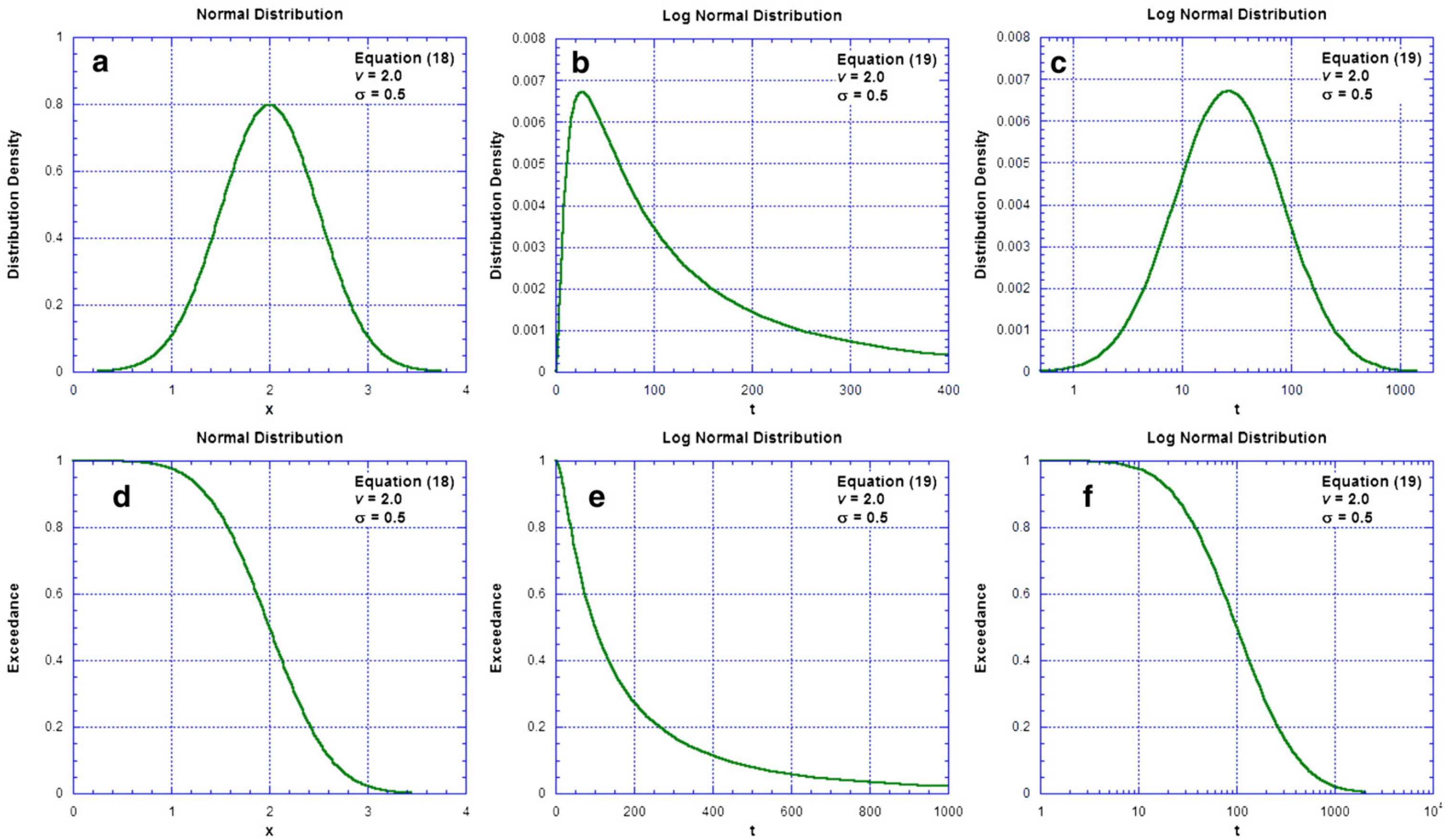

Fig. 1. Illustration for transformation of monomodal logarithmic distributions to linear forms with hypothetical distributions. a). Normal distribution with arbitrarily chosen position ($\nu=2$) and standard deviation ($\sigma=0.5$); b). Same distribution as log normal (units relate as $x=\log_{10} t$). Note that mode is shifted despite the fact that it is still exactly the same distribution (Eqs. (20) and (21)). This is because bin sizes and count events are different in both cases $\Delta x=\text{const}=x_2-x_1=x_3-x_2$ and $\Delta t=\text{const}=\log_{10} x_2-\log_{10} x_1 \neq \log_{10} x_3-\log_{10} x_2$. Mode of Eq. (21) is at $t_{\text{Mode}} = \mu \exp\left(-\frac{1}{k^2}\right)$ and for 10 base logarithms we have it at $t_{\text{Mode}}=10^{(\nu-\sigma^2 \ln 10)}$. For this particular case $t_{\text{Mode}}=26.6$ which corresponds to $x=\log_{10} 26.6=1.42$. As sigma increases the differences between normal and log normal modes becomes larger. c). Meaningless changing of t-axis scale of (b) to its logarithm makes a Gaussian curve (normal distribution). Its moment has no statistical or physical meaning (see discussion of (b)). However, with substitution $x'=\log_{10} t$ its function becomes a Gaussian curve of the form $\hat{f}(x') = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x'-\nu')(x'-\nu'')}{2\sigma^2}\right]$ where ν' and ν'' are real numbers. d). Exceedance for the same distribution as on (a). e). Exceedance for the same log normal distribution as on (b). f). Changing t-axis of (e) to log scale yields a match of (d). Since exceedance does not involve any derivatives of original probability (it is actually inverse probability) it is legitimate to change the t-axis to logarithmic scale.

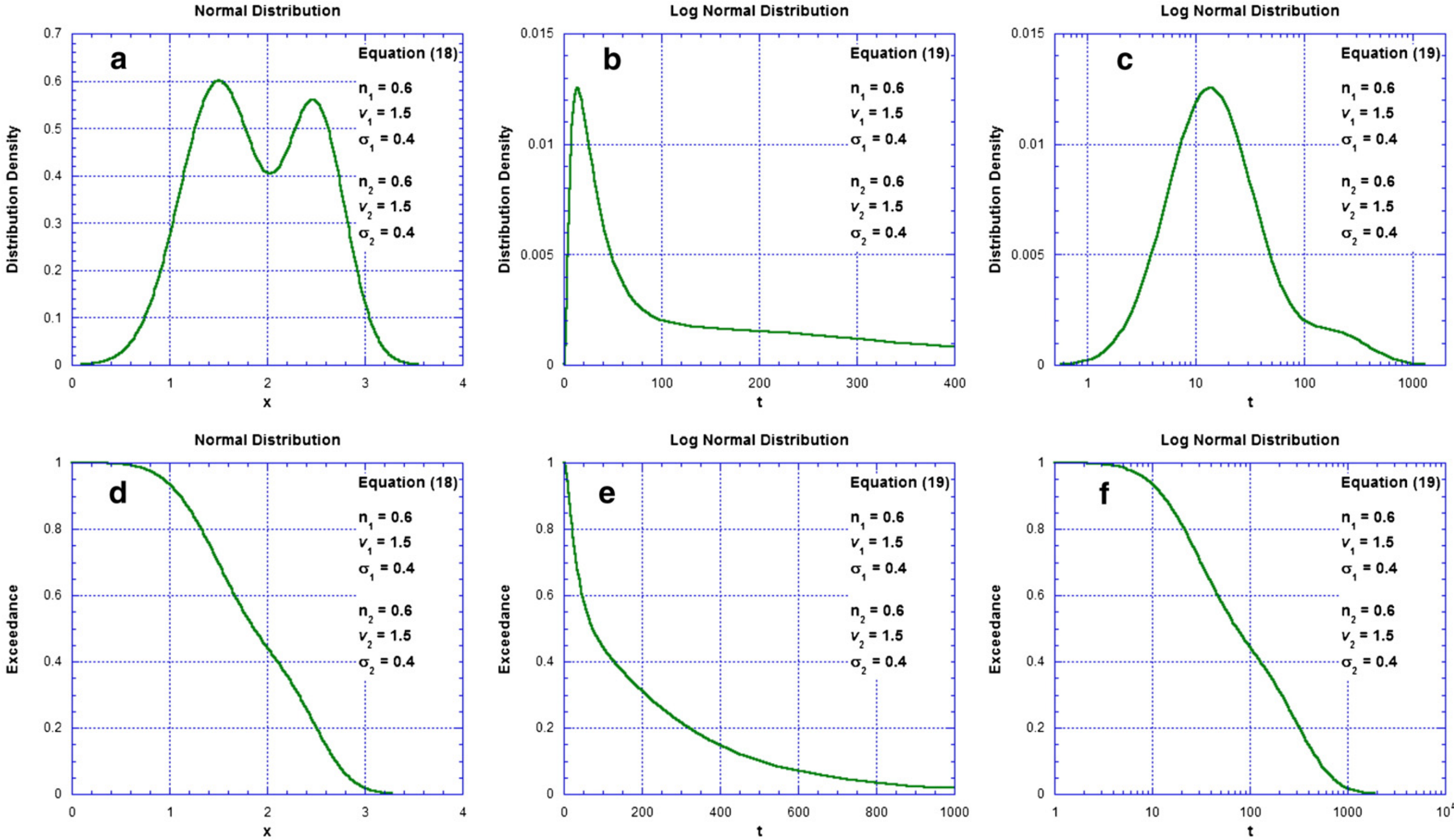

Fig. 2. Same graphs as on Fig. 1 but for bimodal hypothetical distribution. Note that small wiggles on the exceedance curve caused by the bimodal nature of the distributions are difficult to discern visually and could be easily masked by data scatter and thus overlooked in sample analyses.

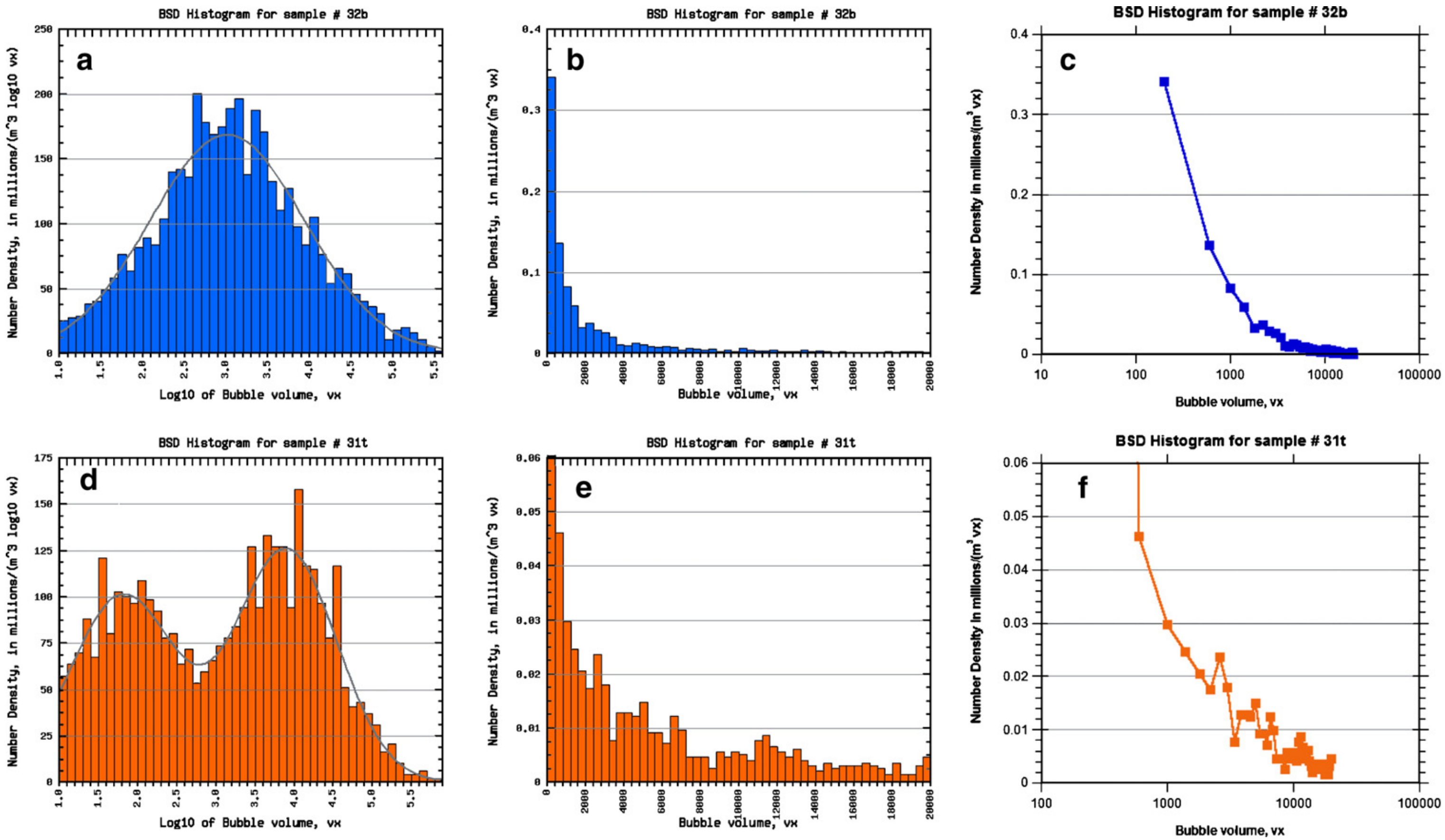

Fig. 3. Examples with actual physical data for graphs corresponding to Figs. 1 and 2. Here we illustrate BSD converted to linear distributions by changing unit scale (graphs in first column). Graphs in second column are logarithmic distributions with unchanged units. Graphs in third column are same as in second but using logarithmic scale in abscissa. Top row is normal unimodal distribution. Second row is normal bimodal distribution. Samples 32b and 31t are used for the illustration (Colorado Plateau basalts collection processed by high-resolution X-ray computed tomography and object recognition software (Proussevitch and Sahagian, 2001)) with voxel size of 47 microns.

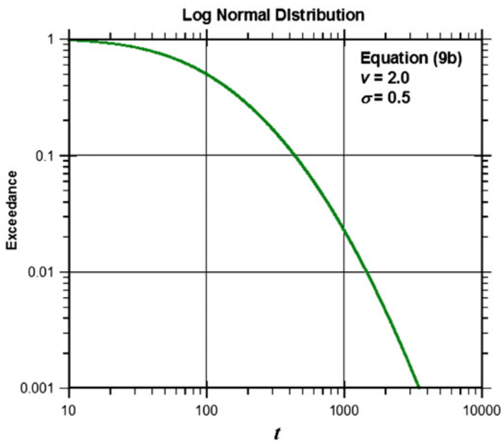

Fig. 4. Typical log–log representation of exceedance curve in bubble size analysis that completely obscures distribution features and shifts the focus onto the large size tail of the distribution. This commonly leads to systematic misinterpretation of the distribution.

The above transformation can be visualized simply for the case of a log normal bubble size (volume) distribution (21), but its features are not apparent or interpretable in terms of physical meaning for coefficients k and μ (Figs. 1 and 2). The distribution density (histogram) takes on the form of a monotonically decreasing exponential curve that quickly falls from high values at low volumes to small values as volume increases (Figs. 1b, 2b and corresponding Fig. 3b, e). If one takes the standard approach of data binning, then the small size bins quickly fill, while larger bins remain almost empty. This situation is not improved by examining exceedance curves. However, the picture comes into focus if one takes the logarithm of volume for every bubble and re-plots the results in the scale from negative to some positive $t(s)$, with uniform bin sizes in log units (Figs. 1a, 2a and corresponding Fig. 3a, d). This leads to a more interpretable bell shaped distribution density curve with mean (v) and standard deviation (σ) well characterized.

The example with bimodal distribution (Fig. 2) highlights the importance of correct treatment of log normal and other logarithmic distributions to enable physical interpretation and proper visualization (Fig. 3a).

In contrast to distribution density, exceedance is not prone to the errors inherent in direct substitution of units used to measure object (bubble) sizes (Figs. 1 and 2, second row), because exceedance is inverse probability and does not involve derivatives (as in Eq. (17)). It is important to note that in any case (before or after unit transformations) bimodal distributions are virtually impossible to recognize on exceedance curves (Figs. 1d and 2d). This is one of the primary reasons for using distribution density graphs for diagnostic purposes.

Despite the various limitations of exceedance curves, they do reveal one feature of bubble distributions particularly well. Upon inspection of an exceedance curve, it is possible to determine if the complete range of object's sizes is represented in the observed population. Because distribution density has a "bell shaped" form, exceedance in turn gets a sigmoidal form with the wings asymptotically approaching 1 and 0 on either end of the

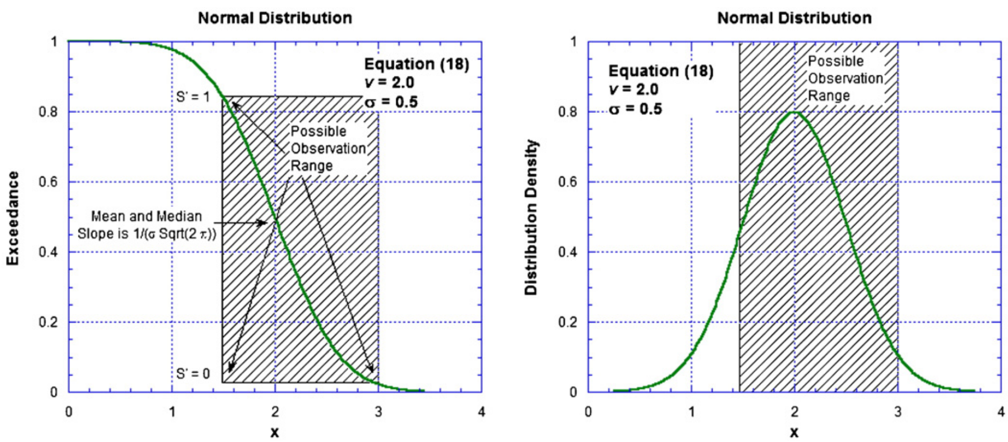

Fig. 5. Exceedance rescaling in applications to natural samples. In actual distributions, small objects in log normal distributions are rarely detected, and some large objects can also be missed. The shaded area shows the range of observations. In the observation area the smallest object gets observation exceedance $S'=1$ and the largest one gets $S'=0$, although those values actually correspond to $S=0.85$ and $S=0.05$ respectively. Rescaling S' using Eq. (15) can salvage exceedance function fitting, but lack of sufficient characteristic points on the curve can cause the fitting procedure to fail or yield large uncertainties. Distribution coefficients (v, σ) obtained from function fitting of both exceedance and density curves should generally match to ensure the results are valid.

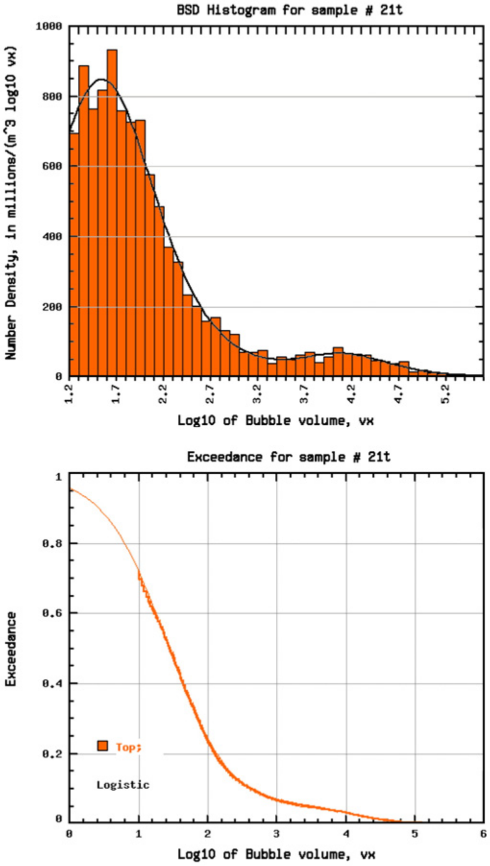

Fig. 6. Example to illustrate partial range of observed bubble sizes (Sample # 21t, collection of Colorado Plateau basalts processed by high-resolution X-ray computed tomography and object recognition software (Proussevitch and Sahagian, 2001). Voxel size is 47 μm.). This is a bimodal log logistic distribution shown as normalized distribution density histogram on the upper graph and exceedance on the lower graph. The units on both graphs has been rescaled to base 10 logarithms, and linear distribution equations (Table 4) have been used to find the best fit distribution curves. On the density histogram you can clearly see two modes. While the larger mode is represented throughout its full range, the smaller mode is cut off so that about 30% of the bubbles are not observed and presented in the collected data (due to resolution limits). This is reflected in the exceedance curve (Eq. (15)) as missing data for small bubbles (thin line). Also note that the second mode is not evident on the exceedance graph and could be easily missed.

distribution (see Figs. 1 and 2, second row). If this shape is not observed on a plot then it means that the observed sizes range is incomplete (as on Figs. 1e and 2e). Thus, although it is tempting to use log–log scale for bubbles' exceedance or exceedance density plots (Fig. 4) such graphs do not provide any meaningful information upon inspection.

Another important but largely overlooked issue in statistical analysis of bubble and crystal distributions is the verification of object size range in observed populations to determine completeness (Fig. 5). Most commonly, small bubbles or crystals are missing from the observed population due to measurement limitations, and large bubbles or crystals are missing due to finite size of the sample. This undermines the utility of exceedance unless the population is properly rescaled (Eq. (15)). A problem remains in that cutoff thresholds on both ends of the population are unknown. These need to be estimated by function best fit analysis (Section 9, below). Distribution density plots are not so adversely affected by the observational cutoffs (see Fig. 5) and thus are generally more useful for population visualization, assessment, and preliminary analysis.

An example of a population observed in actual rocks is shown on Fig. 6. The distribution density graph (based on unit substitution) brings to light both the bimodal distribution and the fact that the small size mode is cutoff by observational limitations. Appropriate rescaling of the exceedance function yields good function fit that in turn matches the function used to fit the density histogram. In this way, the two curves can be used to enhance each others' utility in the interpretation of population statistics.

6. Analytical forms of selected logarithmic distributions

Now that we have established the principles of transformation from log to linear formulations, we can discuss the details of the distribution equations indicated in Table 3. Table 4 contains the logarithmic and linear forms of the four essential functions from the logarithmic family of distributions. Conversion of the function coefficients is given for both natural and base 10 logarithms. (Base 10 logarithms are actually used in our studies).

We have produced logarithmic forms from linear normal distributions using relations (18–19). We have also used both forms of the logistic distribution (Cox and Oakes, 1984). Finally, we have produced linear forms of logarithmic representations for Weibull and exponential distributions.

7. Distribution functions used to analyze bubbles in volcanic rocks in previous studies

Some functions indicated in Table 4 have been used in a number of studies in the past to explore bubble size distributions. The studies have generally concluded that one or the other distribution is appropriate for the various observed populations, with the tacit implication that each

Table 4
Logarithmic family of distribution functions used to describe bubble size distributions

Logarithmic distribution		Linear distribution	Coefficient conversions	
Probability density	Exceedance	Probability density	Natural log scale	Base 10 log scale
Normal				
$\hat{f}(t)=\frac{k}{t\sqrt{2\pi}}\cdot\exp\left[-\frac{1}{2}k^2\ln^2\left(\frac{t}{\mu}\right)\right]$	$S(t)=\frac{1}{2}\mathrm{erfc}\left[\frac{k\ \ln\left(\frac{t}{\mu}\right)}{\sqrt{2}}\right]$	$\hat{f}(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-\nu)^2}{2\sigma^2}\right]$	$\mu=\exp\nu$ $k=\frac{1}{\sigma}$	$\mu=10^{\nu}$ $k=\frac{1}{L_{10}\sigma}$
Logistic				
$\hat{f}(t)=\frac{\frac{k}{\mu}\left(\frac{t}{\mu}\right)^{k-1}}{\left[1+\left(\frac{t}{\mu}\right)^k\right]^2}$	$S(t)=\frac{1}{1+\left(\frac{t}{\mu}\right)^k}$	$\hat{f}(x)=\frac{\sigma^{-1}\exp\left(\frac{x-\nu}{\sigma}\right)}{\left[1+\exp\left(\frac{x-\nu}{\sigma}\right)\right]^2}$	$\mu=\exp\nu$ $k=\frac{1}{\sigma}$	$\mu=10^{\nu}$ $k=\frac{1}{L_{10}\sigma}$
Weibull				
$\hat{f}(t)=\left(\frac{k}{\hat{\mu}}\right)\left(\frac{t}{\hat{\mu}}\right)^{k-1}\exp\left[-\left(\frac{t}{\hat{\mu}}\right)^k\right]$	$S(t)=\exp\left[-\left(\frac{t}{\hat{\mu}}\right)^k\right]$	$\hat{f}(x)=\sigma^{-1}\exp\left[\frac{x-\nu}{\sigma}-\exp\left(\frac{x-\nu}{\sigma}\right)\right]$	$\hat{\mu}=\exp\nu$ $k=\frac{1}{\sigma}$	$\hat{\mu}=10^{\nu}$ $k=\frac{1}{L_{10}\sigma}$
Exponential (It is special case of Weibull at $\sigma=1$)				
$\hat{f}(t)=\frac{1}{\hat{\mu}}\exp\left(-\frac{t}{\hat{\mu}}\right)$	$S(t)=\exp\left(-\frac{t}{\hat{\mu}}\right)$	$\hat{f}(x)=\exp[(x-\nu)-\exp(x-\nu)]$	$\hat{\mu}=\exp\nu$	Use Weibull at $\sigma=1$

Notations: $\hat{f}$ — probability density; S — exceedance; μ — mean; $\hat{\mu}$ — median; σ — standard deviation; N_0 — nucleation density; $L_{10}=\mathrm{Ln}(10)$;
Note # 1: Distributions are given in order of their significance.
Note # 2: Population and probability densities relate as $f(t)=N_0\hat{f}(t)$ and $\hat{f}(x)=N_0\hat{f}(x)$. In both cases N_0 is the same value.
Note # 3: BND$=N_0/Vm$.

interpretation was in conflict with the others. Here we present a more general formulation of the problem in which each of these is part of a family of distributions that applies to all of the observations. A few examples are given below:

7.1. Exponential distribution (Marsh, 1988)

This is the only logarithmic distribution function that has been used to analyze and interpret bubble size distributions in the past (Cashman and Mangan, 1994). Its logarithmic form has been used (first column in Table 4), but there have been no previous attempts to interpret the free coefficient that precedes the exponent (such as the one related to Bubble Number Density (BND) (see notes # 2–3 for Table 4 or Eq. (6))). Comparison of the exponential distribution with the Weibull distribution (Table 4) clearly indicates that the former is a special case of more general Weibull distribution. The special case occurs when sigma becomes equal to 1, reducing number of match coefficients (parameters) to just two.

7.2. Log logistic distribution

Although this formulation has never been explicitly used in bubble size analysis, its asymptotic case for large bubbles has been investigated (Gaonac'h et al., 1996a,b). The log logistic nature of the distribution was not recognized by previous investigators, but Fig. 6 of (Gaonac'h et al., 1996a) is an ideal case of the distribution throughout the whole range of bubble sizes. Instead, they found that for large bubbles (set $t\gg\mu$ in the log logistic distribution in Table 4), their data fit a power relation

$$N(V)\propto V^{-B-1} \tag{22}$$

$$S(V)\propto V^{-B} \tag{23}$$

where B is an empirical exponent.

Naturally, a power function of this form can not formally be a distribution function because, for example, integration of Eq. (22) yields infinity (integration of probability density should be equal to one). Nevertheless, if one takes $t\gg\mu$ for the log logistic distribution (Table 4), then it transforms to power law (Eqs. (22) and (23)). Gaonac'h et al. suggested that $B\approx 0$ for bubble populations produced by a diffusive regime of growth, and $B\approx 1$ for a coalescence regime of bubble growth. Later it was suggested that high B could be caused by multiple nucleation events and bubble growth combined (Blower et al., 2001).

In sum, the power law functions (Eqs. (22) and (23)) represent truncation of log logistic distributions for the large scale bubbles far beyond the distribution mode. An obvious benefit of using power law functions is that they

Table 5
Largest viable objects (bubbles) of logarithmic distributions

Distribution	Descent rate within exponent	Reference distribution [a]	Radius of largest bubble [b]	Sample size [c]	Notes
Normal	x^2	ν=0.5 mm σ=0.05 mm ϕ=16.4%	4.26 mm	37.1 mm	*Fast descent rate*. Largest bubbles have size comparable to modal sizes.
Logistic	x	ν=0.5 mm σ=0.04 mm ϕ=48.4%	>1 m	>1 m	*Very slow descent rate*. Largest bubbles are extremely large causing total vesicularity to be much higher than that observed in samples of practical size.
Weibull	exp(x)	ν=0.5 mm σ=0.05 mm ϕ=7.09%	1.52 mm	13.2 mm	*Very fast descent rate*. Largest bubbles have almost same size as modal bubbles.
Exponential	exp(x)	ν=0.5 mm σ=0.06 mm ϕ=11.0%	2.40 mm	19.4 mm	*Very fast descent rate*. Same as Weibull.

[a] Reference distributions are chosen to be typical for Colorado Plateau basalts (Proussevitch et al., 2007-this issue). ν is modal radius, σ is standard deviation, and BND of 10^8 is the same for all distributions.
[b] The largest bubble size is taken as radius of bubble for upper limit to integrate (24) that yields 99% of total vesicularity.
[c] Sample size allows only one bubble of the largest size to be present as found from the probability density function of the reference distribution.

yield finite porosities for observed populations while the log logistic function tend to allow infinite size of largest bubbles and thus exaggerated porosities (Section 8, and Table 5). On the other hand, unlike the power law functions, the log logistic distribution functions enable us to analyze bubbles around and below the distribution mode (max probability density) which is a fundamental characterization of the distribution and thus observed population of bubbles or crystals.

7.3. Log normal distribution

This distribution has been used to interpret bubble size distributions in deep water mid-ocean ridge basalts (Sarda and Graham, 1990), but the density histogram included only 7 data points (bins), making any function fit problematic. The possibility of log normal distributions for bubbles in basaltic lava bombs has been suggested (Shin et al., 2005), but no systematic study of this distribution has been performed to date.

While the log-normal and exponential functions have been applied to natural samples, the Weibull and full, log-logistic distributions have been overlooked.

8. Bulk vesicularity of specific size distributions

If a distribution function is known, it is possible to calculate sample bulk vesicularity. This constitutes an important application of the statistical analysis of bubble distributions because it is often difficult to measure void fraction in typical rock samples with assurance that the full size spectrum is captured (particularly for very large or very small extremes).

Total volume of bubbles υ_g in a sample can be calculated directly from number density $N(V)$ or population density $f(V)$ as

$$\upsilon_g = \upsilon_m \int_0^\infty V \cdot N(V) \mathrm{d}V = \int_0^\infty V \cdot f(V) \mathrm{d}V \qquad (24)$$

where V is a measure of bubble sizes by volume. Vesicularity ϕ of a sample then comes from the volume of bubbles υ_g and melt υ_m as follows:

$$\phi = \frac{\upsilon_g}{\upsilon_g + \upsilon_m} \qquad (25)$$

Unfortunately only exponential distributions lend themselves to volume integration

$$\upsilon_g = \mathrm{BND} \cdot \hat{\mu} \qquad (26)$$

while other cases (normal, logistic and Weibull) must be integrated numerically using Romberg schemes for improper integrals (Press et al., 1992).

In order to physically measure bulk vesicularity, it is necessary to know the rate at which the frequency of large bubbles decreases with increasing size. This determines the minimum necessary size of rock sample as well as the size of the largest possible vesicle in the sample. This issue is summarized in Table 5.

The distribution descent rate is critical in defining the minimum sample size necessary to include a representative distribution of the largest vesicles for determining bulk vesicularity. As suggested in Table 5, relatively small samples are suited to study vesicularity of basalts if they

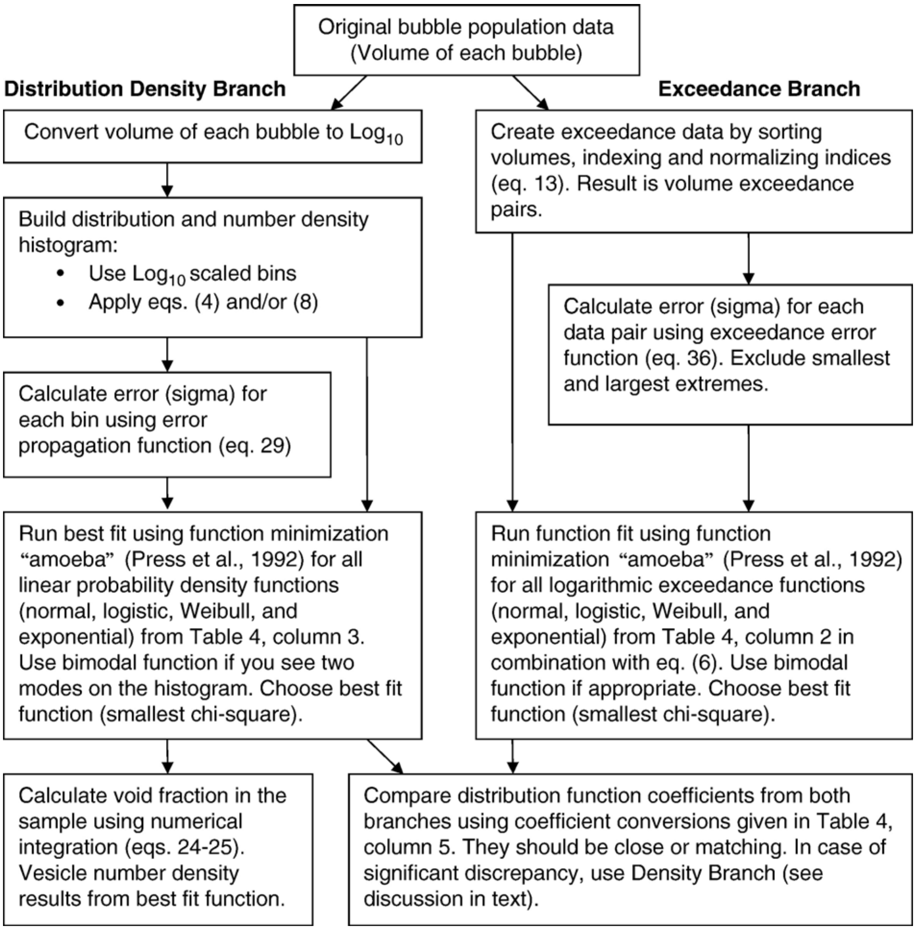

Fig. 7. Flowchart of bubble size distribution analysis.

contain Weibull or exponential distributions. Samples with normal distribution need to be larger, and those with logistic distributions must be larger still because they contain a few extremely large bubbles that would control the total vesicularity, but that could be missed altogether in a small sample. As logistic distributions relate to power law functions (Eqs. (22) and (23)), this issue has been discussed in detail in previous studies (Gaonac'h et al., 1996b) in which some elegant ways to use power law functions to estimate finite porosities have been suggested.

9. Function best fit analysis

In order to determine the appropriate distribution with which to characterize a population, it is necessary to find the function that best fits the observed data. There are a number of function fitting routines available in commercial graphic software. One of the most widely used methods is “Maximum Likelihood Estimation” (MLE) (Johnson and Kotz, 1970). However, these are inappropriate for this application because they are invariably based on minimizing the distance between the function and observed values, whereas the statistically robust approach is to minimize this distance in terms of observation error (sigma) for each data point. This requires a full error analysis for the data a priori. We present an approach to calculating this for both exceedance and probability or number density.

Our approach is based on chi-square function minimization such that

$$\chi^2 \equiv \sum_{i=1}^{N} \left(\frac{y_i - y(x_i; a_1 \dots a_M)}{\sigma_i} \right)^2 \tag{27}$$

$$\hat{\chi}^2 = \frac{\chi^2}{N - M} = \frac{\chi^2}{\nu} \tag{28}$$

where N is a number of observation points (x_i,y_i) with known errors σ_i (standard deviations) at each point, $y(x)$ is the function to fit that has its own coefficients $a_1,\dots,$

a_M, and v is number of degrees of freedom. The purpose of best fit analysis is to identify a combination of given function coefficients $a_1,\ldots, a_M$ such that χ^2 is minimized. We have developed an algorithm for this task that is based of function minimization routine "amoeba" from Numerical Recipes (Press et al., 1992).

As noted above, this method requires knowledge of errors at each observation point, and the best estimate for the error is square root of counts in each bin (given a statistically meaningful number of counts of at least $\sigma_i^{\text{count}} = \sqrt{n_i}$). For fewer counts, the direct number of counts can be used rater than the square root, and if there are no counts (n_i=0) it is set to 1. Using the error propagation equation, the error for each observation point is

$$\sigma_i = \left|\frac{\partial f(n)}{\partial n}\right|\sigma_i^{\text{count}} = \left|\frac{\partial f(n)}{\partial n}\right|\sqrt{n_i} \tag{29}$$

Thus, for a function to fit normalized distribution density (Eq. (8)) it is

$$\sigma_i = \frac{1}{V_m \Delta V}\sqrt{n_i} \tag{30}$$

Error in the exceedance function fit should be taken differently. If we write the exceedance function as

$$S = \frac{n_1}{n_1 + n_2} = \frac{n_1}{n_0} \tag{31}$$

where $n_1 = n_{V>V'}$, $n_2 = n_{V<V'}$, and n_0 is total number of bubbles. Then

$$\sigma_s^2 = n_1\left(\frac{\partial S}{\partial n_1}\right)^2 + n_2\left(\frac{\partial S}{\partial n_2}\right)^2 + 2\,\text{cov}(n_1, n_2)\frac{\partial S}{\partial n_1}\frac{\partial S}{\partial n_2} \tag{32}$$

Since

$$\frac{\partial S}{\partial n_1} = \frac{n_2}{n_0^2} \tag{33}$$

$$\frac{\partial S}{\partial n_2} = -\frac{n_1}{n_0^2} \tag{34}$$

$$\text{cov}(n_1, n_2) = -\sqrt{n_1}\sqrt{n_2} \tag{35}$$

we have

$$\sigma_s^2 = \frac{n_1 n_2(\sqrt{n_1} + \sqrt{n_2})^2}{n_0^4} = \frac{S(1-S)(\sqrt{S} + \sqrt{1-S})^2}{n_0} \tag{36}$$

We should note that Eq. (36) is applicable to the "true" range exceedance curve S and not for the "observed" S' (see Eq. (15)) that is normally available. Since one cannot know a priori how much of the bubble size range is truncated in observations we must exclude the smallest and largest few points from the calculations where the sigma becomes very small. If these are not excluded, then erroneously high chi-square values could result in the case of typical observations that do not actually sample all vesicles at the extreme ends of the distribution.

10. Conclusion

A fully robust statistical analysis of bubble (or crystal) size distributions can provide useful information in studies of vesicular (or crystalline) rocks. We have found that virtually all geologically realistic distributions fall within the logarithmic family of size distributions.

The general procedure of sample analysis for logarithmic distributions is summarized in the flowchart (Fig. 7). Analysis can be based on linear probability density and/or logarithmic exceedance (see two branches on the flowchart). Reliable results cannot be generated directly from observational data for cases with logarithmic probability densities because of problems inherent in data binning. Rather, individual volumes must be converted to a logarithmic scale as a first step. In contrast, an advantage of exceedance analysis is that it does not involve data binning, and so results can be generated without the necessity of arbitrarily selecting bin sizes.

Integration of distribution density function is a valuable supplement to the analysis since it allows estimation of bulk void fraction of bubbles as well as void fraction for each of the modes (in case of polymodal distributions). In many cases it yields more accurate results than direct physical measurements.

Acknowledgements

This material is based upon work supported by NSF under awards EAR-0207818 and EAR-0509859. We are also greatful to Prof. E. Linder from the Dept of Mathematics and Statistics at UNH for reviewing the paper during its preparation.

References

Bisperink, C.G.J., Ronteltap, A.D., 1992. Bubble-size distributions in foams. Adv. Colloid Interface Sci. 38, 13–32.

Blower, J.D., Keating, J.P., Mader, H.M., Phillips, J.C., 2001. Inferring volcanic degassing processes from vesicle size distributions. Geophys. Res. Lett. 28 (2), 347–350.

Blower, J.D., Keating, J.P., Mader, H.M., Phillips, J.C., 2003. The evolution of bubble size distributions in volcanic eruptions. J. Volcanol. Geotherm. Res. 120 (1–2), 1–23.

Brandeis, G., Jaupart, C., Allegre, C., 1984. Nucleation, crystal growth and thermal regime of cooling magmas. J. Geophys. Res. 89, 10161–10177.

Cashman, K.V., Mangan, M.T., 1994. Physical aspects of magma degassing, II, constraints on vesiculation processes from textural studies of eruptive products. In: Carroll, M.R., Holloway, J.R. (Eds.), Volatiles in Magmas. Reviews in Mineralogy. Mineral. Soc. of Am, Washington, D. C., pp. 446–478.

Cashman, K., Mangan, M., Newman, S., 1992. Surface modifications of vesicle size distributions in Kilauea basalt — implications for interpretation of vesicle distribution in lava flows. Eos 73, 629.

Cashman, K.V., Mangan, M.T., Newman, S., 1994. Surface degassing and modifications to vesicle size distributions in Kilauea basalt. J. Volcanol. Geotherm. Res. 61, 45–68.

Cobb, L., Koppstein, P., Chen, N.H., 1983. Estimation and moment recursion relations for multimodal distributions of the exponential family. J. Am. Stat. Assoc. 78 (381), 124–130.

Connor, C.B., Sparks, R.S.J., Mason, R.M., Bonadonna, C., Young, S.R., 2003. Exploring links between physical and probabilistic models of volcanic eruptions: The Soufriere Hills Volcano, Montserrat. Geophys. Res. Lett. 30 (13), 1701. doi:10.1029/2003GL017384.

Cox, D.R., Oakes, D., 1984. Analysis of survival data. Chapman and Hall, London. 201 pp.

Gaonac'h, H., Lovejoy, S., Schertzer, D., 2005. Scaling vesicle distributions and volcanic eruptions. Bull. Volcanol. 67 (4), 350–357.

Gaonac'h, H., Lovejoy, S., Stix, J., Scherzter, D., 1996a. A scaling growth model for bubbles in basaltic lava flows. Earth Planet. Sci. Lett. 139, 395–409.

Gaonac'h, H., Stix, J., Lovejoy, S., 1996b. Scaling effects on vesicle shape, size and heterogeneity of lavas from Mount Etna. J. Volcanol. Geotherm. Res. 74 (1–2), 131–153.

Gurioli, L., Houghton, B.F., Cashman, K.V., Cioni, R., 2005. Complex changes in eruption dynamics during the 79 AD eruption of Vesuvius. Bull. Volcanol. 67 (2), 144–159.

Johnson, N.L., Kotz, S., 1970. Continuous univariate distributions, 1. Hougton Mifflin, Boston. 300 pp.

Klug, C., Cashman, K.V., Bacon, C.R., 2002. Structure and physical characteristics of pumice from the climactic eruption of Mt. Mazama (Crater Lake), Oregon. Bull. Volcanol. 64, 486–501.

Lovejoy, S., Gaonac'h, H., Schertzer, D., 2004. Bubble distributions and dynamics: The expansion–coalescence equation. J. Geophys. Res. 109 (B11), B11203.

Marsh, B.D., 1988. Crystal size distributions (CSD) in rocks and the kinetics and dynamics of crystallization: 1. Theory Contr. Min. Pet. 99, 277–291.

Marsh, B.D., 1998. On the interpretation of crystal size distributions in magmatic systems. J. Petrol. 39 (4), 553–599.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numerical Recipes in C. The Art of Scientific Computing. University Press, Cambridge. 998 pp.

Proussevitch, A.A., Sahagian, D.L., 2001. Recognition and separation of discrete objects within complex 3D voxelized structures. Comput. Geosci. 27, 441–454.

Proussevitch, A., Sahagian, D., Kutolin, V., 1993. Stability of foams in silicate melts. J. Volcanol. Geotherm. Res. 59, 161–178.

Proussevitch, A.A., Sahagian, D.L., Carlson, W., 2007-this issue. Statistical analysis of bubble and crystal size distributions: Application to Colorado Plateau Basalts. J. Volcanol. Geotherm. Res. doi:10.1016/j.jvolgeores.2007.04.007.

Sahagian, D.L., 1985. Bubble migration and coalescence during the solidification of basaltic lava flows. J. Geol. 93, 205–211.

Sahagian, D.L., Proussevitch, A.A., 1998. 3D particle size distributions from 2D observations: stereology for natural applications. J. Volcanol. Geotherm. Res. 84, 173–196.

Sahagian, D.L., Anderson, A.T., Ward, B., 1989. Bubble coalescence in basalt flows: comparison of a numerical model with natural examples. J. Volcanol. Geotherm. Res. 52, 49–56.

Sahagian, D.L., Proussevitch, A.A., Carlson, W., 2002. Timing of Colorado Plateau uplift: initial constraints from vesicular basalt-derived paleoelevations. Geology 30 (9), 807–810.

Sarda, P., Graham, D., 1990. Mid-ocean ridge popping rocks: implication for degassing at ridge crests. Earth Planet. Sci. Lett. 97, 268–289.

Shin, H., Lindquist, W.B., Sahagian, D.L., Song, S.R., 2005. Analysis of the vesicular structure of basalts. Comput. Geosci. 31 (4), 473–487.

Toramaru, A., 1989. Vesiculation process and bubble size distributions in ascending magmas with constant velocities. J. Geophys. Res. 94 (B12), 17,523–17,542.

Toramaru, A., 1990. Measurement of bubble size distributions in vesiculated rocks with implications for quantitative estimation of eruption processes. J. Volcanol. Geotherm. Res. 43, 71–90.

Vergniolle, S., 1996. Bubble size distribution in magma chambers and dynamics of basaltic eruptions. Earth Planet. Sci. Lett. 140 (1–4), 269–279.