ShapeCalc – Documentation

This document has been adapted from the Supplementary Information given in Mangler et al. (2022), where ShapeCalc was first published:

Mangler, M.F., Humphreys, M.C.S., Wadsworth, F.B., Iveson, A.A., Higgins, M. (2022). Variation of plagioclase shape with size in intermediate magmas: a window into incipient plagioclase crystallisation. Contributions to Mineralogy and Petrology, 177, 64.
https://doi.org/10.1007/s00410-022-01922-9

ShapeCalc_v1.0 - Documentation

ShapeCalc provides 3D (*S:I:L*) shape estimates for 2D (*w-l*) sample data. Best estimates are found by comparing the sample *w/l* distribution with 2618 model *w/l* distributions covering shapes from 1:1:1 to 1:20:20. The approach taken is analogous to *CSDslice*, including the binning of the *w/l* distribution into 25 bins with a width of $\Delta w/l = 0.04$. However, *ShapeCalc* is distinguished by a seven key differences:

(1) The model *w/l* distributions for *ShapeCalc* are derived from *CSDCorrections* (v 1.61; M. Higgins, 2022), as they allow a more robust shape differentiation for $S \approx I$. Furthermore, using the same models for shape estimates and crystal size distribution analysis via *CSDCorrections* offers superior internal consistency for any CSD work.

(2) Each model distribution is based on 20,000 random sections of a given model shape (*CSDslice*: 10,000 random sections).

(3) The database for *ShapeCalc* includes shapes up to 1:20:20, allowing adequate representation of crystals with extreme aspect ratios, which may be expected in rapidly cooled samples (*CSDslice*: 1:1:1 - 1:10:10).

(4) The *ShapeCalc* data grid offers a higher resolution for 3D intermediate and long dimensions *I* and *L,* and it reflects the greater uncertainty on *L* with a wider spacing (Table 1; Figure 1). The total number of model shapes included in the *ShapeCalc* database is 2618 (*CSDslice*: 703).

(5) The best *S:I:L* estimate is found by maximising the cumulative goodness of fit (GOF) between sample and model *w/l* distributions (Fig. 2), calculated as

$$R_c{}^2 = 1 - \frac{\sum_{j=1}^{25}\sum_{i=1}^{j}(s_i - m_i)^2}{\sum_{j=1}^{25}\sum_{i=1}^{j}\left[s_i - \frac{(\sum_{i=1}^{n} s_i)}{25}\right]^2} \quad (1)$$

where s_i is the normalised frequency of crystals in bin i for the sample *w/l* distribution, and m_i is the normalised frequency of crystals in bin i for the model *w/l* distribution.

The *cumulative* assessment of the GOF using $R_c{}^2$ is distinct from $R^2_{CSDSlice}$, which assesses the GOF based on the *modes* of the distribution. However, in natural samples, variations in both mode and tail frequencies are to be expected compared to *w/l* distributions derived from a perfect model crystal. The $R_c{}^2$ GOF test parameter accounts for such natural variability, as it is less sensitive to the modal value and variations in individual bins but rather assesses the overall fit of the *w/l* distributions. An $R_c{}^2$ <0.975 indicates a larger discrepancy between sample and model *w/l* distributions than reasonably expected for a natural sample; such results are highlighted in red by *ShapeCalc*. In such cases, better results may be achieved by binning the *w-l* sample dataset.

(6) *ShapeCalc* lists and displays the 300 best-ranked *S:I:L* estimates, colour coded by $R_c{}^2$. This threshold was empirically chosen as the limit in variation between sample and model. If desired, the threshold value can be easily changed in the spreadsheet.

(7) *ShapeCalc* provides constraints on the uncertainties on the best shape estimate as 1σ standard deviations on *S/I* and *I/L* of the 300 best shape estimates. These uncertainties are deliberately conservative to reflect the variability of crystal shapes both between and within crystal populations, as well as generally poor statistical constraints on *I/L*. Typical standard deviations are ± 0.04-0.11 for *S/I,* and ± 0.12-0.26 for *I/L*.

Table 1: *Step sizes for I and L used to produce model w/l distributions*

S	I Range	I Step	L Range	L Step
1	1 – 2	0.05	1 – 2	0.05
1	2 – 5	0.1	2 – 3	0.1
1			3 – 5.2	0.2
1	5 – 10	0.2	5.2 – 10	0.4
1	10 – 20	0.5	10 – 20	1

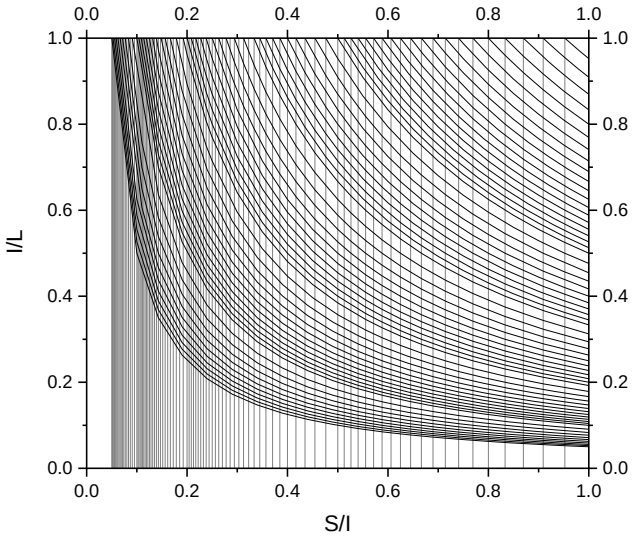

Fig. 1 *Grid of model shapes used by ShapeCalc. Every intersection between S/I and S/L tie lines represents a model shape in the ShapeCalc database*

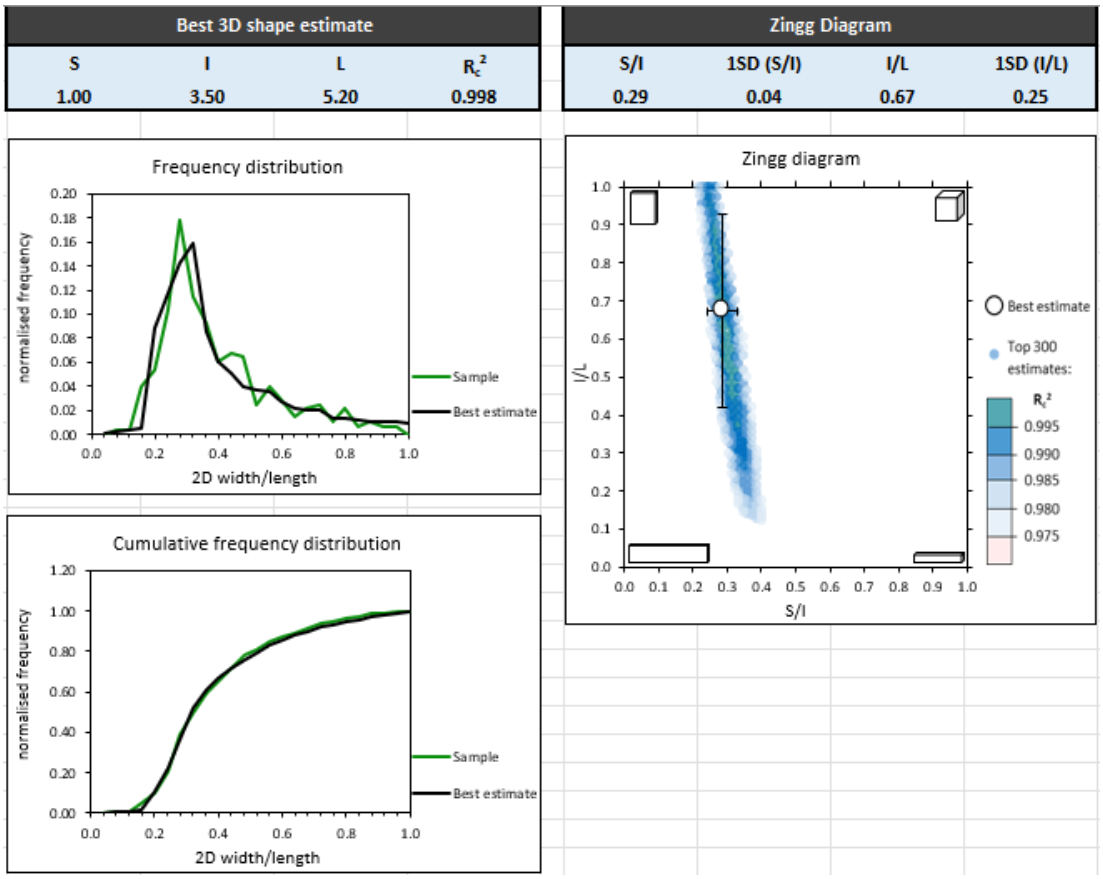

Fig. 2: *Screenshot showing ShapeCalc outputs: Best 3D shape estimate, R_c^2, 1SD uncertainties on S/I and I/L, distributions of sample and best estimate model data, and visualisation of 300 best fits on a Zingg diagram.*

Background

Crystal intersections are 2D representations of 3D objects, and 2D intersection sizes and shapes (approximated by w and l) are not a straightforward measure of 3D short, intermediate, and long crystal dimensions (S:I:L; e.g., Higgins 1994; 2000). Constraining true 3D shapes requires a robust statistic of 2D w-l measurements as well as knowledge of the size(s) and orientation of crystals (preferred vs. random). Higgins (1994) showed that for a population of randomly oriented, anisotropic 3D objects of same size and shape, the most likely intersection length l is close to the 3D intermediate dimension I, and the most likely intersection width w approximates the 3D short dimension S. Hence, the mode of a 2D w/l distribution yields the 3D S:I ratio of the true shape (Higgins 1994; 2000). Estimates of the long dimension L are more difficult to obtain and less robust, as sections along the elongate axis are rare. Higgins (1994; 2000) proposed an estimation of I/L based on the skewness of the w/l distribution.

Higgins' equations relating 2D w/l to 3D S:I:L are based on numerical models in which a rectangular block with a defined S:I:L is randomly sectioned 50,000 times to produce a representative w/l distribution for the known 3D shape (Higgins, 1994). A version of this numerical model using 20,000 random sections is included in the crystal size distribution program CSDCorrections (Higgins, 2000). In CSDCorrections, the user is required to input estimates for S:I:L, which are used to generate a model w/l distribution, which are the used to perform tailing corrections during the calculation of CSDs. ***ShapeCalc*'s database of 2618 model w/l distributions was produced using the same numerical model as CSDCorrections.**

A similar numerical modelling approach is used by the standalone program CSDslice (Morgan and Jerram 2006). Here, the w/l model distributions are based on 10,000 random intersections, which are obtained by rotating a model crystal into random orientations and cutting it perpendicular to the z-axis of a reference coordinate system. CSDslice automatically compares sample w/l distributions with its database of 703 model w/l distributions for 3D shapes from 1:1:1 to 1:10:10 and identifies the five best S:I:L estimates based on a goodness of fit test centred on the mode of the distribution. Best S:I:L estimates from CSDslice are commonly used as input for CSDCorrections.

Comparison of ShapeCalc and CSDslice

The respective model w/l distributions used by *ShapeCalc* and CSDslice are consistent for shapes with S/I << 1, where the modes of the w/l distributions reflect S/I (Fig. 3a, b). **However, significant differences in model w/l distributions are observed for shapes with S/I ≈ 1 (i.e.,**

equant and prismatic shapes; Fig. 3c). In particular, while *ShapeCalc* w/l distributions show a mode of w/l = 1 reflecting S≈I, CSDslice data lack a clear mode. This indicates that there may be an error in the algorithm used by CSDslice that is most evident for prismatic shapes. **The lack of a clear mode renders CSDslice model w/l distributions non-unique.** For example, CSDslice model w/l distributions for multiple different shapes can be combined to closely reproduce its 1:1:10 model distribution (>0.8) (Fig. 3d). This suggests that the model distributions underpinning CSDslice tend to return a shape with S ~ I (i.e., 1:1:L) for sample w/l distributions without a clear modal peak. Natural w/l distributions derived from igneous rocks often lack such clear modes due to the presence of multiple crystal populations as well as some variability in shape within a single population (e.g., Duchêne et al. 2008). We therefore speculate that many natural samples are likely to produce a 1:1:L outcome using CSDslice, irrespective of their true crystal shapes. Indeed, published plagioclase microlite shape estimates derived by CSDslice are dominated by shapes with S≈I (e.g., S:I = 1:1.1–1:1.5 in Preece et al. 2013, 2016; Bain et al. 2019; and Wallace et al. 2020). Furthermore, a core limitation of CSDslice is that its model database is limited to 3D shapes from 1:1:1 to 1:10:10, and natural crystal shapes with higher S:I or S:L can therefore not be reproduced. In our experience, this can generate artefacts suggesting a prevalence of 1:10:10 shapes, particularly for experimental samples, where shapes with S:L ≈ 1:20 have been reported (Muncill and Lasaga 1988; Hammer and Rutherford 2002). Finally, in light of the significant differences in model w/l distributions between CSDslice and *ShapeCalc* (Fig. 3), we suggest that it would be more internally consistent to use the same w/l model distributions for both 3D shape estimates and tailing corrections of CSDs, i.e., *ShapeCalc*.

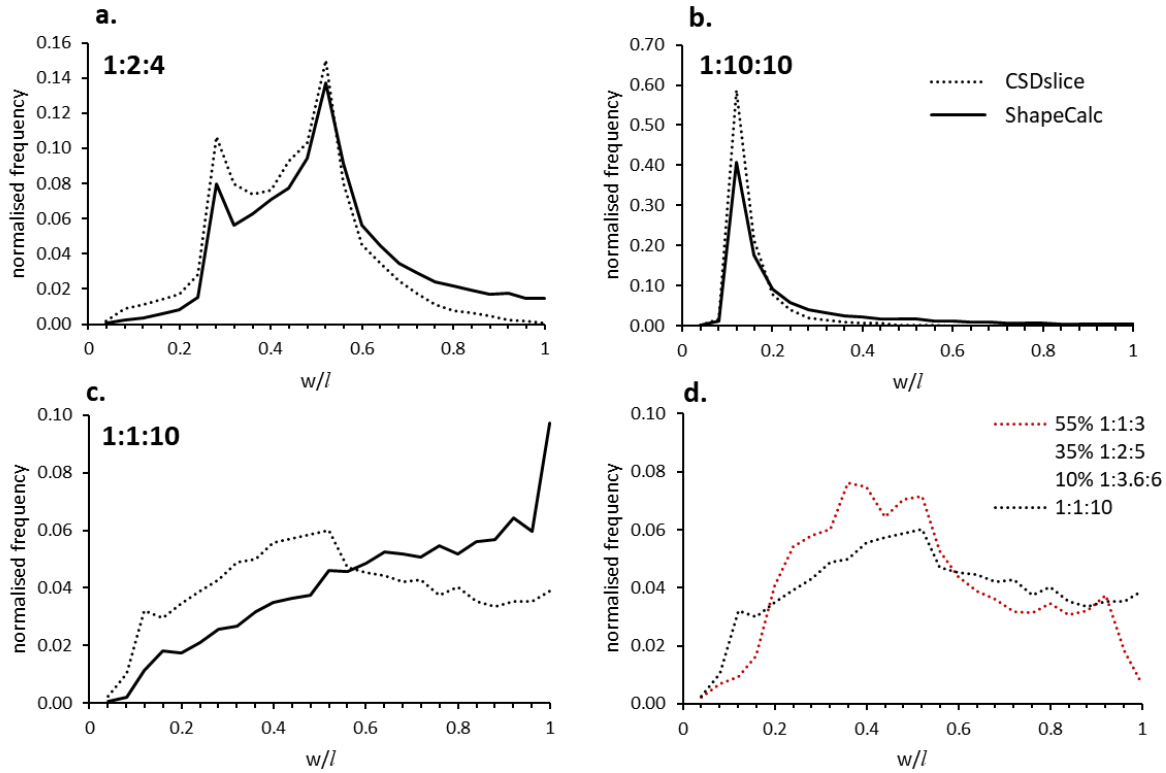

Fig. 3: *(a–c) Comparison of model 2D w/l distributions used by CSDslice (Morgan and Jerram 2006; dotted lines) and ShapeCalc (solid lines). Each distribution is obtained by 10,000 (CSDslice) or 20,000 (ShapeCalc) random sections of a model crystal shape. (a) 1:2:4, (b) 1:10:10, (c) 1:1:10. Note the large discrepancies for 1:1:10 model shapes. (d) Mixing of three CSDslice model populations shows good agreement with its 1:1:10 model distribution (>0.8), illustrating how may yield erroneous 3D shape estimates if multiple crystal populations with varying shapes are present in a sample.*

Evaluation of measuring technique

Image processing packages such as ImageJ calculate the aspect ratio of a polygon by measuring the width and length of a **best-fit ellipse**, i.e., an ellipse of equal area, orientation and centroid as the original shape. Combined with thresholding algorithms, which can identify the desired shapes, *w-l* data can be rapidly extracted from an image. However, as crystals in natural samples are commonly intergrown, automatic thresholding is often not feasible, and even manual outlining of individual crystals can be challenging. **Direct line measurements** of crystal length and width are more time consuming, but they offer greater conceptual robustness, as it is not necessary to know the complete 2D crystal outline in order to determine *l* and *w* accurately. To test whether the method of acquiring *w* and *l* affects shape results, we measured plagioclase

crystals in six BSE images of Mount St. Helens sample SH131A-2 (c.f. Mangler et al., 2022) using both crystal outlines and direct line measurements. The *w-l* distributions ($n \approx 600$) for both methods are in overall good agreement, the only slight difference being the greater relative abundance of $w/l > 0.9$ (Fig. 4a).

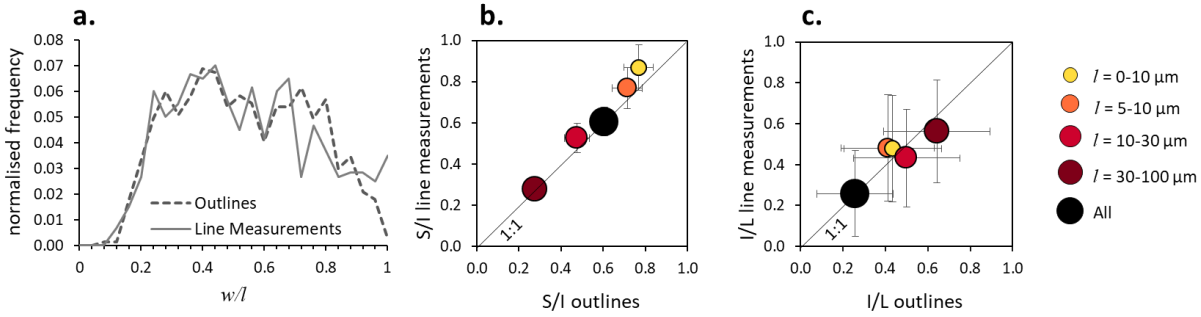

Fig. 4: *Shape results for ≈600 plagioclase crystals counted using crystal outlines or direct line measurements to obtain length and width data. (a) normalised frequency plot of 2D data. (b) and (c) ShapeCalc 3D best estimates for line measurement data plotted against those for outline data, (b) S/I and (c) I/L*

Similarly, processing of the 2D datasets using *ShapeCalc* yields consistent *S:I:L* best estimates across length fractions (Fig. 4b & c). Ratios of *S/I* and *I/L* show a difference between methods of ≤0.1, which constrain uncertainties on *S/I* and *I/L* derived from measurement errors to ± 0.05. These uncertainties are equal to or lower than those related to most *ShapeCalc* model results, suggesting that **both methods of constraining the aspect ratios of 2D crystal intersections are robust and reproducible.**

Evaluation of statistical effects

Morgan & Jerram (2006) suggested that a minimum of 250 sections is necessary for accurate *S:I:L* estimates in natural samples, and that more sections are required for shapes with *S≈I* due to the cut-section effect. We tested the effect of statistics on 2D *w-l* data and 3D *S:I:L* estimates by comparing results of two individual counts of plagioclase in BSE images of SH131A-2 (Fig. 5). The first dataset is the one obtained for the test of the measuring technique ($n = 599$), with the number of crystal sections per size bin varying from 126 to 265. The second dataset represents a larger area in SH131A-2 and therefore contains about four times as many *w/l* datapoints ($n = 2518$), with 354-1450 crystals per size bin.

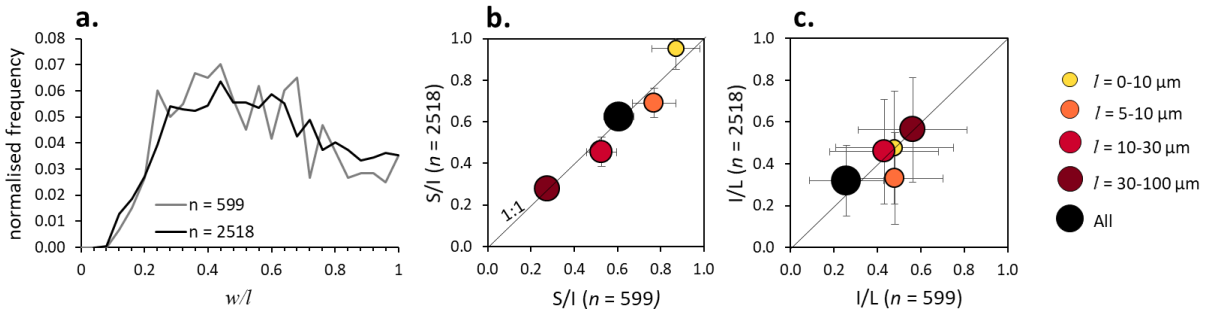

Fig. 5 *Shape results for different counting statistics of MSH sample SH131A-2. In a first run, widths and lengths of 599 crystals were measured; in a second run of the same sample, 2518 crystals were measured. (a) normalised frequency plot of 2D data. (b) and (c) ShapeCalc 3D best estimates for the two runs plotted against each other, (b) S/I and (c) I/L*

The *w-l* distributions for the respective complete datasets are broadly consistent (Fig. S4a), and *S:I:L* estimates show good agreement (Fig. S4b & c), including the bins with the smallest *n* (*l* = 30-100 µm). The largest differences between the two datasets are observed for the size fractions with shapes approaching *S/I* = 1, i.e., the *l* = 0-10 µm and 5-10 µm fractions ($\Delta S/I \leq 0.08$, $\Delta I/L \leq 0.15$). It should be noted that even the largest discrepancies related to counting statistics are equal to or smaller than uncertainties related to the modelling performed with *ShapeCalc*. We conclude that when using *ShapeCalc,* a **minimum number of 200 *w/l* data points is recommended in order to estimate a 3D shape for an unknown sample**; **for tabular crystal shapes (*S*<<*I*), robust results can be obtained with ~100 2D crystal intersections.**

References

Bain AA, Calder ES, Cortés JA, Cortés GP, Loughlin SC (2019) Textural and geochemical constraints on andesitic plug emplacement prior to the 2004–2010 vulcanian explosions at Galeras volcano. Colombia Bull Volcanol 81(1):1–25. https://doi.org/10.1007/s00445-018-1260-y

Duchêne S, Pupier E, De Veslud CLC, Toplis MJ (2008) A 3D reconstruction of plagioclase crystals in a synthetic basalt. Am Mineral 93(5–6):893–901. https://doi.org/10.2138/am.2008.2679

Hammer JE, Rutherford MJ (2002) An experimental study of the kinetics of decompression-induced crystallization in silicic melt. J Geophys Res Solid Earth. https://doi.org/10.1029/2001jb000281

Higgins MD (1994) Numerical modeling of crystal shapes in thin sections: estimation of crystal habit and true size. Am Mineral 79(1–2):113–119

Higgins MD (2000) Measurement of crystal size distributions. Am Mineral 85(9):1105–1116. https://doi.org/10.2138/am-2000-8-901

Mangler, M.F., Humphreys, M.C.S., Wadsworth, F.B., Iveson, A.A., Higgins, M. (2022). Variation of plagioclase shape with size in intermediate magmas: a window into incipient plagioclase crystallisation. Contributions to Mineralogy and Petrology, 177, 64. https://doi.org/10.1007/s00410-022-01922-9

Morgan DJ, Jerram DA (2006) On estimating crystal shape for crystal size distribution analysis. J Volcanol Geotherm Res 154(1–2):1–7. https://doi.org/10.1016/j.jvolgeores.2005.09.016

Muncill GE, Lasaga AC (1988) Crystal-growth kinetics of plagioclase in igneous systems; isothermal H 2 O-saturated experiments and extension of a growth model to complex silicate melts. Am Mineral 73(9–10):982–992

Preece K, Barclay J, Gertisser R, Herd RA (2013) Textural and micro-petrological variations in the eruptive products of the 2006 dome-forming eruption of Merapi volcano, Indonesia: implications for sub-surface processes. J Volcanol Geotherm Res 261:98–120. https://doi.org/10.1016/j.jvolgeores.2013.02.006

Preece K, Gertisser R, Barclay J, Charbonnier SJ, Komorowski J-C, Herd RA (2016) Transitions between explosive and effusive phases during the cataclysmic 2010 eruption of Merapi volcano, Java Indonesia. Bull Volcanol 78(8):1–16. https://doi.org/10.1007/s00445-016-1046-z

Wallace PA, Lamb OD, De Angelis S, Kendrick JE, Hornby AJ, Díaz-Moreno A, González PJ, von Aulock FW, Lamur A, Utley JE (2020) Integrated constraints on explosive eruption intensification at Santiaguito dome complex, Guatemala. Earth Planet Sci Lett. https://doi.org/10.1016/j.epsl.2020.116139