

Supplementary information

S.1 Expert elicitation for Fuego PDC volumes

Background on elicitations work. Structured expert elicitation is a quantitative way of pooling expert knowledge to estimate the value of parameters, as well as reflecting the collective confidence in those estimates (Aspinall 2010). In short, it aims to take a snapshot of the current (approximate) view of a group of experts. It is a statistically robust method of establishing a ‘rational consensus’, that faithfully represents the uncertainty of the group. The most commonly used method of structured elicitation in volcanology (Cooke method Cooke (1991)) was developed in the 1990s, and has been applied widely in the aviation industry, nuclear industry, civil engineering, climate science and volcanology.

There are now many methods by which expert elicitation is conducted, each designed to achieve the most accurate result given the nature of the parameters being estimated, and the dynamics of the group of experts. For Fuego, we have conducted one expert elicitation so far with eleven experts using an equal weights method for the computing the results. The authors of this paper were all included. The question asked was: *For Fuego volcano, what is the maximum volume of a pyroclastic flow, in any given barranca, that could occur ?*. We ask for the best estimate, as well as upper and lower reasonable bounds for these PDC volumes. This has allowed the elicitation facilitators to obtain the private views of individual experts, capturing their opinion as well as their degree of certainty.

Elicitation software. We used the open-source, Python-based application, “ELICIPY 1.0” (Vitturi et al. 2024), which combines data collection, expert weighting (not used here), results processing, and results presentation into one package. ELICIPY 1.0 was developed by volcanologists following the body of work by Aspinall and others, and has been used operationally for volcanological expert elicitations at INGV (Italy), the IMO (Iceland) and MVO (Montserrat). There is flexibility to adjust important parameters, modify functions if necessary, and the process is automated, and fast. The software allows answers to be filled out using a secure webform, with data submitted to a secure GitHub repository. The automated submission and formatting of these responses reduces the risk of copying errors, and expedites the process significantly. Responses are downloaded and processed offline, and are automatically collated into a PowerPoint presentation containing the results and comparative plots. The pooled response, as well as its uncertainty, is calculated using three weighting schemes, the Classical, ERF and the Equal weighting scheme (the methods used here), and the results of these are presented side-by-side. Once the correct structural components are in place, a full expert elicitation (excluding questionnaire response times) can be performed in less than 10 minutes by a trained user. This allows results to be presented and discussed during the same meeting that they were elicited.

Statistical background of weighting and aggregation. The statistical techniques used to weight and aggregate data in structured elicitations is summarised

below but a fuller description, including their background can be found in (Aspinall et al. 2006; Flandoli et al. 2011; Vitturi et al. 2024; Rougier et al. 2013; Colson and Cooke 2017).

1844

Once elicitation responses are gathered, the responses must be aggregated to calculate the ‘decision maker’ (DM) response for each question. This is conducted by a Monte Carlo sampling approach. For each expert, a distribution is fitted to the reported percentiles for a particular question. For the Equal weights method used here, this is a “maximum entropy distribution” (Vitturi et al. 2024). The DM response is built by sampling this set of distributions N times (where N is a large number; 10000 by default). Each sample is taken from the distribution of one of E experts, and the probability of expert e being sampled on each iteration is equal to $w_{(e)}$. In an equal weighting scheme, $w_{(e)} = 1/E$. This generates a set of N realisations for the response, which is weighted implicitly by the sampling procedure. The mean, median, 5th and 95th percentiles are then calculated to generate a summary of the DM response. This process is repeated for each elicited question.

1857

Results. Results of the elicitation described above are summarized in Figs. S1 and S2, and in Table S1.

1860

1861

1862

1863

1864

1865

1866

1867

1868

1869

1870

1871

1872

1873

1874

1875

1876

1877

1878

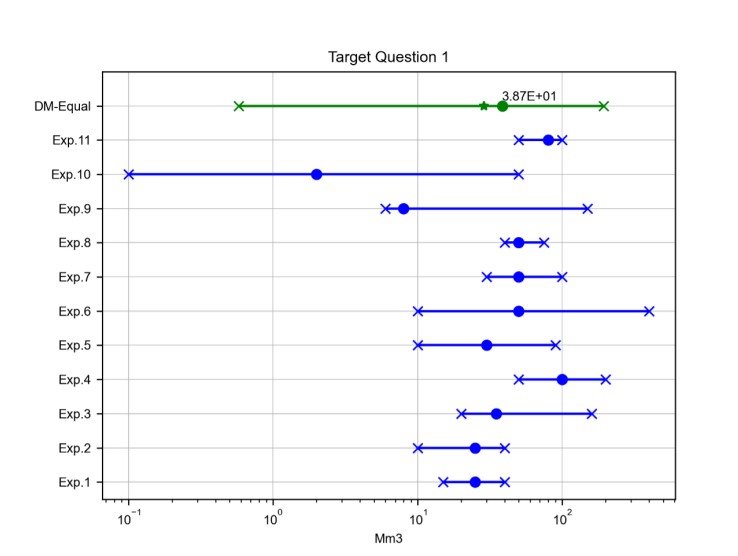


Fig. S1 Show the responses from the eleven experts for the target question, *For Fuego volcano, what is the maximum volume of a pyroclastic flow, in any given barranca, that could occur?*, as well as the computed results for the equal weights decision-maker (DM).

1882

1883

1884

1885

1886

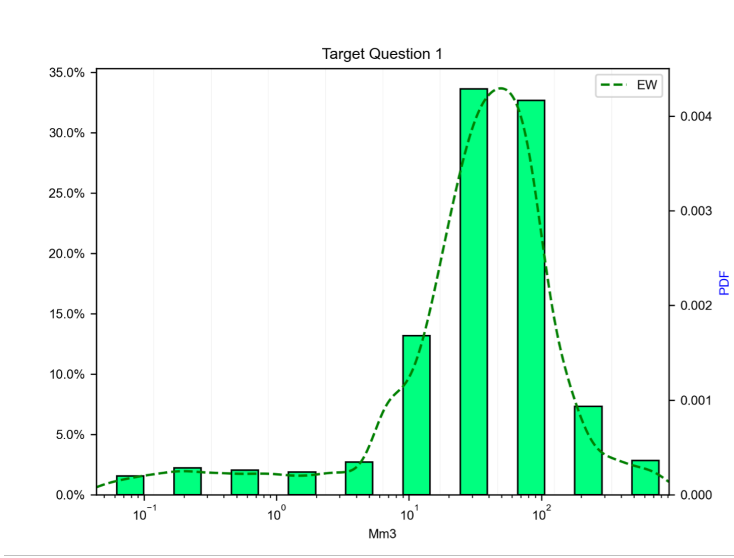


Fig. S2 Shows the distribution from which the percentiles are calculated and thus what provides the final answer. The way this is calculated for a given target question is: Each expert provides three percentiles. Each experts' response is fitted with a probability distribution. In the Equal weights case, this is a maximum entropy distribution. 10,000 samples are taken to generate the histogram. Each sample is (1) taken according to a multinomial distribution defined by the experts' weights. Eg. in an equal weight setup, two experts would be samples 5000 times each to make the total. (2) The distribution for each author is sampled according to the distribution defined in step 1. Collectively, these samples therefore represent a weighted distribution of 10k actualisations given the collective opinion of experts. The histogram is fitted non-parametrically by gaussian kernel density estimation (the dotted line). The 5^{th} , 50^{th} , 95^{th} percentiles and the mean of the 10,000 samples are taken as the overall answer, and presented in the Table S1.

Q_{05}	Q_{50}	Q_{95}	Mean
0.58 Mm ³	38.7 Mm ³	194 Mm ³	28.7 Mm ³

Table S1 The 5^{th} , 50^{th} , 95^{th} percentiles and the mean for maximum estimated PDC volumes that could occur in any given barranca.

S.2 TITAN2D friction-volume design details

Ogburn *et al.* (2016, Ogburn *et al.* 2016) devises a hierarchical linear model (HLM) of the relationship (on a log-log scale) between the observed coefficient of friction, H/L , and the observed PDC volume (m³). Here we implement this model for Fuego, leveraging data from Colima, Merapi, Soufrière Hills, Unzen, and Semeru. The hierarchical linear model assumes that each volcano has its own regression slope, but that those slopes are samples from from a Normal distribution with variance τ^2 ($\tau^2 = 0$ implies that the various volcanoes' regression slopes are identical while $\tau^2 \rightarrow \infty$ implies that each volcano's regression is completely separate). Figure S3(a) shows the modeled Fuego regression curve, $l_{HLM}(\log_{10} V)$ along with credible intervals and prediction

1933 intervals.

1934

1935

1936

1937

1938

1939

1940

1941

1942

1943

1944

1945

1946

1947

1948

1949

1950

1951

1952

1953

1954

1955

1956

1957

1958

1959

1960

1961

1962

1963

1964

1965

1966

1967

1968

1969

1970

1971

1972

1973

1974

1975

1976

1977

1978

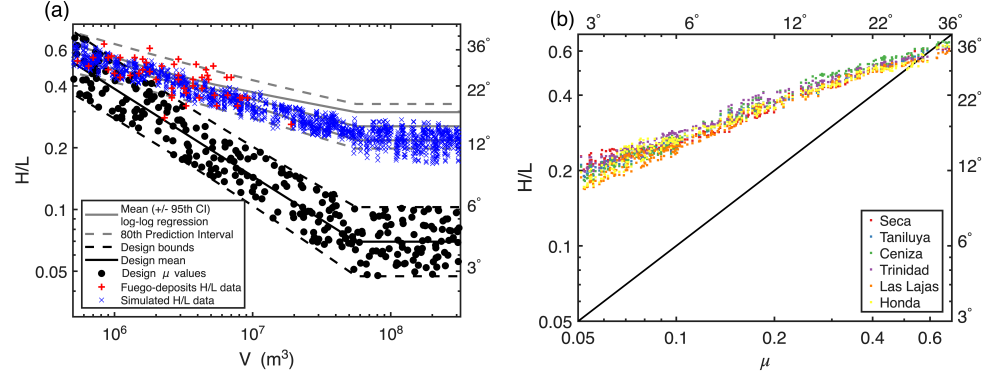


Fig. S3 (a) Log-log plots of μ vs volume, the simulation design; simulated H/L vs volume for all barrancas; and estimated Fuego deposit H/L versus volume. (b) Simulated H/L for all barrancas vs μ for each corresponding TITAN2D run, and a one-to-one reference line for visualization. (Additionally, for reference, angles corresponding to the inverse tangent of H/L are labeled on the right side axis of both figures along with the top for (b).)

For TITAN2D with Mohr-Coulomb rheology, Ogburn (2014, Ogburn 2014) finds the assumption that TITAN2D exercised at $\mu = l_{HLM}(\log_{10} V)$ adequately reproduces mapped PDC run out capturing the relationship that larger volume flows have a lower effective friction. Figure S3(b) indicates that TITAN2D with Voellmy-Salm rheology run at Fuego severely violates this one-to-one relationship. As such, we instead seek a relationship between basal friction and volume that 1) produces simulated H/L s for volumes $\leq 10\text{Mm}^3$ that are consistent with the observed H/L at Fuego and 2) produces simulated H/L s that are roughly consistent with the HLM model of H/L for volumes larger than 10Mm^3 . Initially we explored a one-to-one relationship with a constant shift regardless of volume. That is, we took $\log_{10} \mu = l_{HLM}(\log_{10} V) - a$. Implementing this relationship resulted in simulated H/L s that were too low (compared to observed H/L s) for small volumes and too large (compared to the HLM model of H/L) for large volumes. Motivated by this trial-and-error, we next explore shifting and titling the relationship between the HLM regression line and the μ center-line of our design. Specifically, we consider $\log_{10} \mu = l_{HLM}(\log_{10} V) - (0.3 + 0.25(\log_{10} V - 6.7))$. This relationship is the center-line of our $\mu - V$ design as plotted in Fig. S3(a). We cut off this trend when the lower bound of the design curve reaches a minimum basal friction of $\mu = 0.05$ which at a volume of $V = 56\text{Mm}^3$ resulting in a hockey-stick design that is similar to the one proposed in Rutarindwa et al. (2019). The simulated H/L s for this design are also plotted in Fig. S3(b) for all barrancas. Note that this design does now achieve our objectives of 1) producing simulated H/L s for volumes $\leq 10\text{Mm}^3$ that are consistent with the observed H/L s at Fuego and 2) producing simulated H/L s that are roughly consistent with the HLM model of H/L s for volumes larger than 10Mm^3 .

S.3 Additional figures

Figure S4 is the equivalent to Fig. 4 for the remaining barrancas. Likewise, Fig. S5 is the equivalent to Fig. 5 for the remaining barrancas. Fig. S6 is a complementary figure to Fig. 6 to visualize the two components in the aerial ratio calculations.

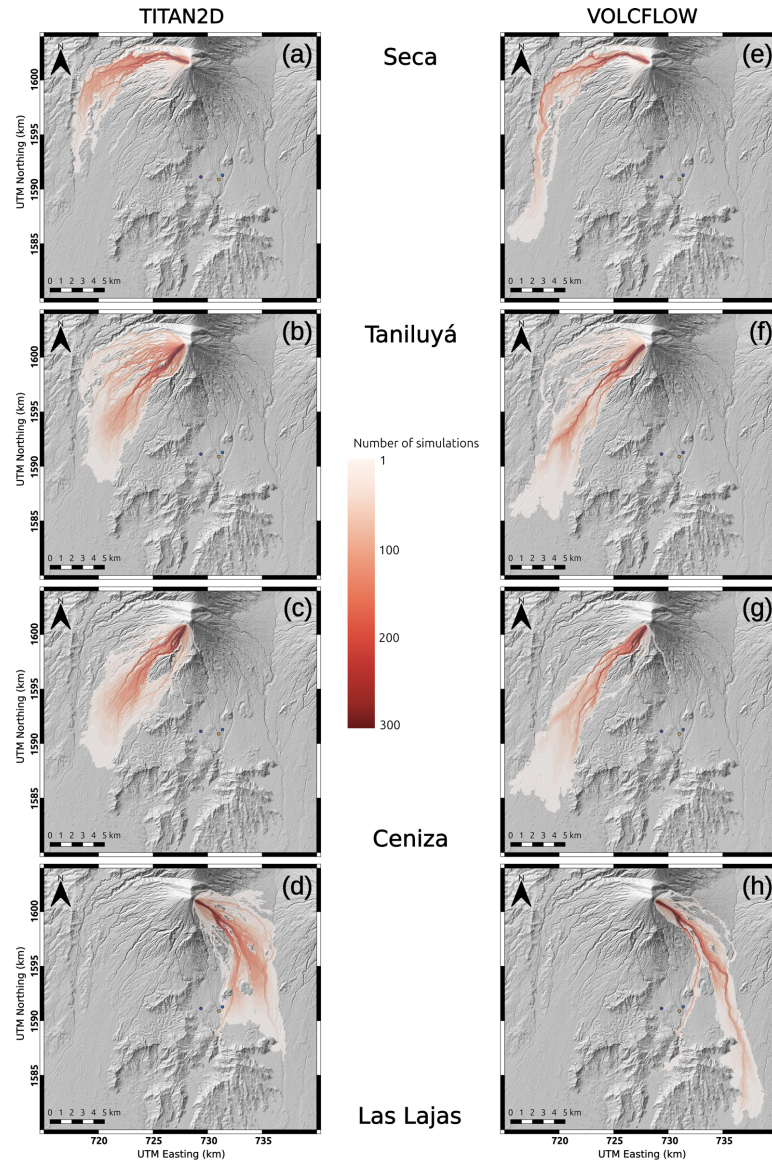


Fig. S4 Maps showing the cumulative areal extent of the design simulations run for the two PDC models explored (TITAN2D, a-d; and Volcflow, e-h) and the other four barrancas at Fuego Volcano: Seca (a, e), Taniluyá (b, f), Ceniza (c, g), and Las Lajas (d, h). NB. Points of interest A, B and C from Figure 1 are displayed in the same colors, but not labeled, to help visualization of PDC footprints.

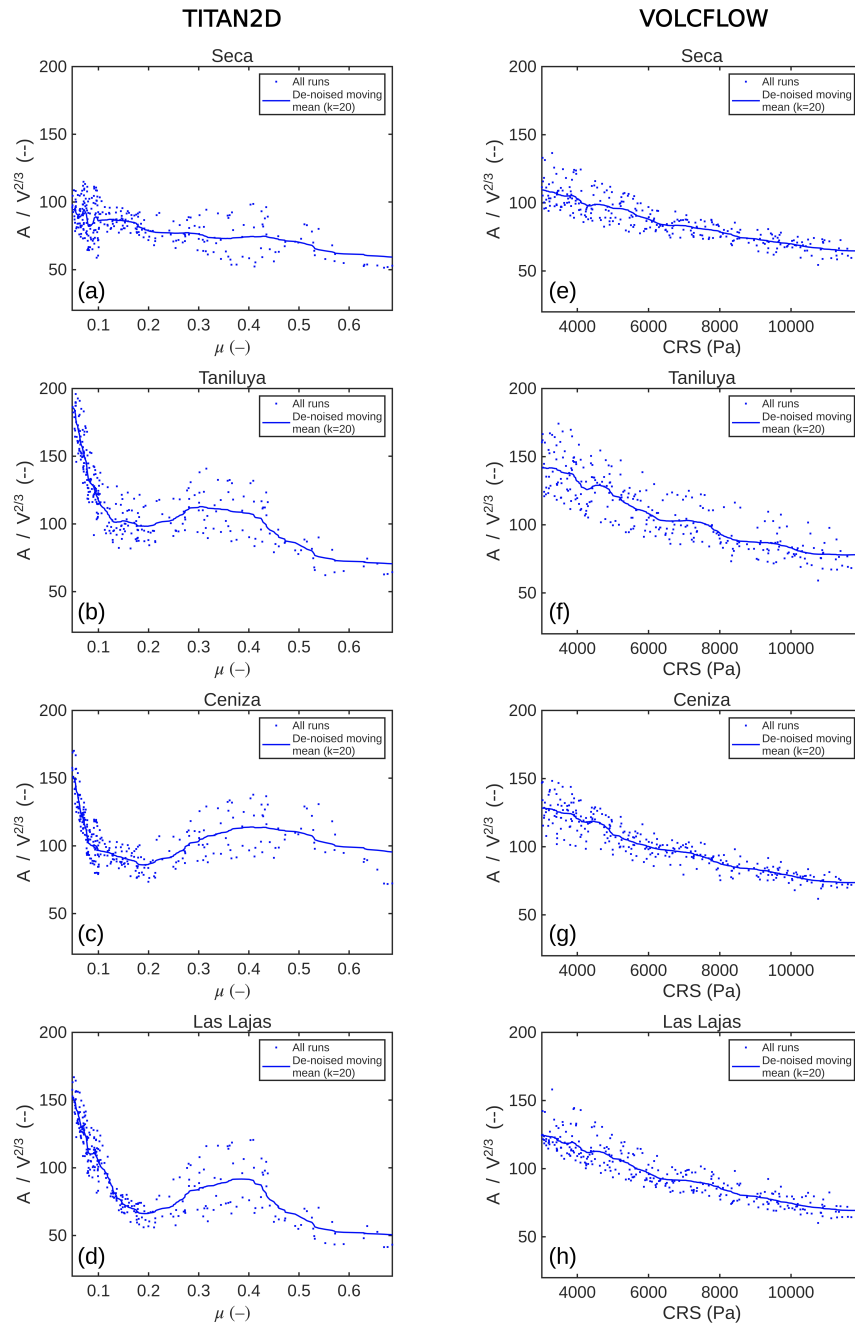


Fig. S5 Additional examples of PDC friction model parameters (TITAN2D: Coulomb-type friction parameter, μ , a-d; Volcflow: Constant Retarding Stress, CRS , e-h) versus PDC mobility metrics computed from the outputs of the design simulations at barrancas Seca (a, e), Taniluyá (b, f), Ceniza (c, g), and Las Lajas (d, h).

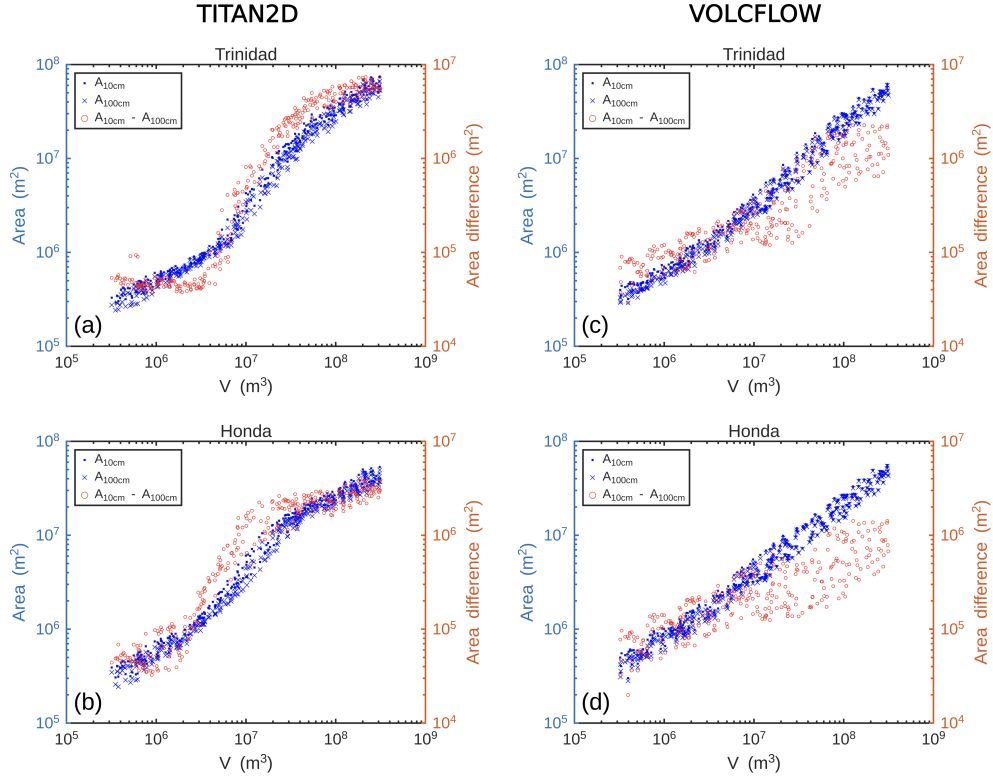


Fig. S6 Additional analyses of the PDC areas of inundation at barrancas Trinidad (a, c) and Honda (b, d), using the TITAN2D (a-b) and Volcflow (c-d) models. A_{10cm} , A_{100cm} denote, respectively, the areas contained within the 10 cm and 100 cm isolines of flow depth, calculated from the model outputs.