The effect of wind and plume height reconstruction methods on the accuracy of simple plume models- a second look at the 2010 Eyjafjallajökull eruption

Tobias Dürig[1], Magnús T. Gudmundsson[1], Thorbjörg Ágústsdóttir[2], Thórdís Högnadóttir[1], Louise S. Schmidt[3]

1. Institute of Earth Sciences, University of Iceland, Sturlugata 7, 101 Reykjavík, Iceland
2. Iceland GeoSurvey (ÍSOR)
3. Department of Geosciences, University of Oslo, Sem Saelands Vei 1, 0371 Oslo, Norway

* Corresponding author, tobi@hi.is

Abstract

Real-time monitoring of volcanic ash plumes with the aim to estimate the mass eruption rate is crucial for predicting atmospheric ash concentration. Mass eruption rates are usually assessed by 0D and 1D plume models, which are fast and require only a few observational input parameters, often only the plume height. A model's output, however, depends also on the plume height data handling strategy (sampling rate, gap reconstruction methods and statistical treatment), especially in long-term eruptions with incomplete plume height records. To represent such an eruption, we used Eyjafjallajökull 2010 to test the sensitivity of six simple and two explicitly wind-affected plume models against 22 data handling strategies. Based on photogrammetric measurements, the wind deflection of the plume was determined and used to re-calibrate radar-derived height data. The resulting data was then subjected to different data handling strategies, before being used as input for the plume models. The model results were compared to the erupted mass measured on the ground, allowing us to assess the prediction accuracy of each combination of data handling strategy and model. Combinations that provide highest prediction accuracies vary, depending on data coverage, eruption intensity, and fragmentation mechanism. However, for this type of moderate-to-weak eruption (VEI 3 in terms of maximum intensity), the most important factor was found to be the prevailing wind speed. When wind speeds exceed 20 m/s, most combinations of

strategies and models provide predictions that underestimate the erupted mass by more than 40%. Under such conditions, the optimal choice of data handling strategy and plume model is of particular importance.

Introduction

Volcanic ash injected into the atmosphere during explosive eruptions can pose a serious threat for aviation and air-travel infrastructure (Kienle et al. 1980; Grindle and Burcham 2002). Next to direct observations of the ash cloud (e.g. satellite imagery and lidar retrievals) predictions of movement of volcanic ash clouds are based on atmospheric ash dispersion models (Dacre et al. 2011; Kristiansen et al. 2012; Dioguardi et al. 2016, 2020). Inaccurate predictions can on one hand lead to severe damage to and even loss of aircraft (Guffanti et al. 2010), or the other hand to airport closures framed or perceived as over-cautious (Harris et al. 2012; Harris 2015) and flight diversions or cancellations, which involve large preventable costs (e.g., Brannigan 2011; Macrae 2011; Ragona et al. 2011). This underlines the need for accurate, reliable and confident model forecasts, the lack of which is the current bottle neck in our forecasting capability. Key to achieving such forecasts is thus to be able to provide an ash dispersion model with accurate near-term eruption source parameters, i.e. physical quantities, which characterize the eruptive source. In this regard the most important eruption source parameter is the mass eruption rate (MER) (e.g., Degruyter and Bonadonna 2012; Mastin 2014; Bonadonna et al. 2016), that is the mass flux of the eruptive gas-ash mixture, expressed in kg/s (Wilson and Walker 1987).

MER cannot be directly measured. However, methods have been developed that infer MER using observable properties of the plume. These include using video analyses of ash plumes and ejecta (e.g., Wilson and Self 1980; Valade et al. 2014; Dürig et al. 2015b, a; Pioli and Harris 2019; Tournigand et al. 2019), emitted infra-sound waves (Johnson and Ripepe 2011; Ripepe et al. 2013), thermal infrared signatures (e.g., Harris 2013; Ripepe et al. 2013; Harris et al. 2013; Cerminara et al. 2015), electrostatic field (Büttner et al. 2000; Calvari et al. 2012), interpretation of microwave radar signals (Montopoli 2016; Marzano et al. 2020) or satellite-based estimates (e.g., Pouget et al. 2013; Pavolonis et al. 2018; Gouhier et al. 2019; Bear-Crozier et al. 2020). For real-time MER assessment, however many of these approaches are still in an experimental state or struggle with high uncertainties as they often depend on data that are hard to obtain *in situ*, e.g. the vent geometry (Dürig et al. 2015a). These methods will thus not be considered here.

The most straightforward and reliable observable in an explosive eruption is usually the height of the eruptive column *H*. A number of plume models linking *H* with the mass eruption rate at the vent have

been developed (for in-depth overview see Costa et al. 2016). These range from "simple" theoretical (Wilson and Walker 1987; Woods 1988) and empirical 0D models (Sparks et al. 1997; Mastin et al. 2009), through explicitly wind-affected steady 1D models (Bursik 2001; Degruyter and Bonadonna 2012; Devenish 2013; Woodhouse et al. 2013; Mastin 2014; de'Michieli Vitturi et al. 2015; Folch et al. 2016; Aubry et al. 2017) to elaborate time-dependent multi-phase models in 2D (Neri et al. 1998) or 3D space (Esposti Ongaro et al. 2007; Suzuki and Koyaguchi 2012; Cerminara et al. 2016).

At present, real-time MER assessments must rely on simple 0D and explicitly wind-affected 1D models, which have the advantage of providing fast results (e.g., Sparks et al. 1997; Bursik 2001; Woodhouse et al. 2013). A monitoring system that uses a suite of such models to assess the mass eruption rate in near-real time is the software REFIR (Real-time Eruption source parameters FutureVolc Information and Reconnaissance system) (Dürig et al. 2018; Dioguardi et al. 2020). Simple 0D models are limited by initial assumptions, accuracy of measured parameters (plume height and mass) and simplifications on which they are based. To assess the effect of these shortcomings, this study explores three aspects that might limit the models' accuracies:

1. the effect of plume height uncertainties resulting from a stepwise horizontal ground-based radar scanning technique on MER estimates.
2. the sensitivity of data handling strategies on model predictions when dealing with incomplete sets of plume height recordings.
3. the impact of wind on MER prediction by non-explicitly wind-affected plume models in comparison to the impact of using different data handling strategies.

The term "data handling strategies" refers to the statistical treatment of plume height data before it is used as a model input. Note that with "plume height" we refer to the maximum elevation of the plume above vent, unless stated otherwise. Since plume height *H* and mass eruption rate *Q* are related in a highly non-linear way, the time-averaged mass eruption rate $\overline{Q(t)}$ is in most cases *not* identical to *Q* resulting from using the time-averaged plume height $\overline{H(t)}$ as input. Thus:

$$\overline{Q(t)} \neq Q\left(\overline{H(t)}\right) \quad (1)$$

Although rarely acknowledged, this fact implies that not only the plume height *H* itself, but also the time interval over which it is measured, as well as the statistical strategy for its assessment, are expected to affect a model's outcome.

In our model sensitivity study, we focus on the 2010 Eyjafjallajökull (Iceland) eruption. This event featured a wind-affected plume (Gudmundsson et al. 2012), which was monitored by a stepwise horizontally scanning radar. The resulting plume height records were incomplete (Arason et al. 2011).

The 2010 Eyjafjallajökull eruption therefore represents an ideal test case for the exploration of the three aspects mentioned above.

The Eyjafjallajökull 2010 eruption

The 2010 eruption of Eyjafjallajökull had four phases of activity.

- Phase I: lasting from 14 April 2010 until noon of 18 April, was an initial explosive phase with phreatomagmatic activity (Gudmundsson et al. 2012; Dellino et al. 2012).
- Phase II: (second half of 18 April - 4 May) was a phase of low discharge that was effusive with relatively weak but sustained explosive activity (Gudmundsson et al. 2012).
- Phase III: (5 - 17 May) saw a significant increase of explosive activity, coinciding with a change in melt composition from benmoreite to trachyte (Gudmundsson et al. 2012). This phase was characterized by the emission of distinct explosive ash pulses (Dürig et al. 2015b, a) and is often referred to as "second explosive phase" (Gudmundsson et al. 2012; Dellino et al. 2012).
- Phase IV: (18 - 22 May) was characterized by a decline in eruption activity and plume height. Although the end of eruption is set on 22 May, minor isolated explosive events occurred on 4-8 June 2010 (Gudmundsson et al. 2012). No ash plumes were detected on radar after 10:20 UTC on 21 May, when the plume fell below the detection limit of 2.5 km (Arason et al. 2011).

The eruption's fallout was measured in considerable detail at about 400 locations in Iceland (Gudmundsson et al. 2012), allowing us to compare model predictions on the erupted mass with the actual amount of mass deposited as tephra. For the part of the tephra that fell into the ocean south and southeast of Iceland a considerable uncertainty exists, but the magnitude of this component can nevertheless be estimated using well-established thinning behaviour of tephra layers with distance (e.g., Thorarinsson 1954; Pyle 1989; Fierstein and Nathenson 1992). On 14-16 April, during phase I, westerly upper-tropospheric winds prevailed, changing to northerly winds by 17 April, directing the plume to the south (Gudmundsson et al. 2012). As a result, the erupted mass could be assessed separately as phases *Ia* and *Ib* (see Table 1).

The monitoring conditions of this eruption are classified as relatively challenging since it is a moderate-to-weak eruption that took place under adverse weather conditions (frequent clouds at low and medium altitudes and strong wind) (Arason et al. 2011; Gudmundsson et al. 2012; Björnsson et al. 2013), which meant direct observations of the volcanic plume were often not possible.

Data

The data used in this study are:

i) Plume height obtained by the Icelandic Meteorological Office (IMO) from the C-band radar in Keflavík (Arason et al. 2011).

ii) Plume height as measured from photographs taken from inspection aircraft (Figure 1).

iii) Plume height as measured on photographs taken from the ground from areas to the west of the volcano.

iv) Wind speed and direction at plume-relevant altitudes retrieved either from ERA5 or ICRA. ERA5 is a global reanalysis product by the European Centre for Medium-Range Weather Forecasts with a horizontal resolution of 30 km (Hersbach et al. 2020). ICRA is a local reanalysis product by the Icelandic Met Office over Iceland, using the non-hydrostatic numeric weather prediction model HARMONIE-AROME with a horizontal resolution of 2.5 km (Nawri et al. 2017; Schmidt et al. 2018). Both models have a temporal resolution of 1h. Since ICRA data were only available up to an altitude of 6 km a.s.l., we used composite data sets for the ICRA labelled wind speed profiles: ICRA reanalysis data below 6 km, and ERA5 reanalysis data above.

Radar data

Located at 155 km distance from the vent, the C-band radar at Keflavík airport, was the only weather radar operating in Iceland at the time of the eruption. The radar's sampling strategy was to scan at vertical angles of 0.5°, 0.9°, 1.3°, 2.4°, 3.5°, 4.5° and 6° (Arason et al. 2011). The width of the beam was 0.9° providing some overlap between scans. For Eyjafjallajökull, the vertical distance between the beams of the lowest scanning angles was about 1.1 km and about 3 km between 1.3° and 2.4°. For a target above the volcano, this leads to a stepping in the plume height records, with preferential sampling at about 2.5, 3.9, 5.0 and ~8 km height above sea level (for details see Arason et al., 2011).

Photos

The monitoring of vent activity and visual observations of plume behaviour took place on inspection flights (see Fig. 1; for complete list of surveillance flights, see Appendix. 6.3 in Thorkelsson et al. (2012)). For flights where the aircraft location was recorded with GPS, it was possible to determine the position of the camera at the time of a photo with an accuracy higher than 1 km (Gudmundsson et al. 2015). This is the maximum error in the distance between aircraft and the volcanic vents, or any other reasonably well-defined target.

In most cases the plume was bent over by the wind, and its bearing was usually obtained from the position of the aircraft when it was upwind and in line with the plume. Occasionally this was also done by flying along the plume and determining the location relative to known landmarks. Errors in distance to the top of the plume were in this way contained within 1 km in most cases, and less than 2 km in all cases. In the vast majority of cases this amounted to 5-10% of the distance.

The height of plume was either assessed directly by using the altimeter of the plane (Fig. 1d) or calculated using the Python software Pixelcalc (Magnússon 2012). The plume top was referenced by triangulation, using photos of the plume taken in short time intervals from different angles. Given a known camera-target distance Pixelcalc converts distances between two points on the photo to actual 3D distances at the location of the target by using the lens and camera specifications, which are derived from the photo's metadata. The software corrects for Earth's curvature at distances where it becomes significant. In most cases the plume height is obtained by measuring the distance on the photo between the plume top and the surface, taking into account plume deflection by wind and surface elevation under the plume top, to obtain height above sea level. In some cases, where data on cloud top height are available, height over clouds (i.e., the distance between cloud top and plume top) is determined. Such cloud heights are in some cases obtained by direct observation using aircraft altimeter or through independent cloud height measurement using Pixelcalc. The uncertainties of the heights and horizontal deflection distances vary from one photo to another, depending on the uncertainties in distances, pixel size and image geometry. Assuming maximum errors for these parameters, we found the uncertainties of plume height and deflection to be less than 0.5 km.

In addition to the airborne imagery, photographs taken on the ground were used. For these the location of the camera is accurately known which allowed us to calculate the distance to the plume top. Independent information of plume azimuth was used to correct for plume top deflection relative to vents.

The total number of photos taken on inspection flights are counted in thousands. Not all flights provided photos that allow reliable determination of plume height, e.g., due to cloud cover. Only the photos best suited for analysis were used, yielding in total 165 observations of plume heights, covering 17 of the 38 eruption days examined. From these photos, 148 were taken in the air, and 17 on the ground. In 85 cases these photos could also be used to assess the plume deflection (i.e., the horizontal shift of the plume top towards the vent location).

Methods

Models

For estimating the MER, we used six *simple* (non-explicitly wind-affected) plume models and two wind-affected models, which are for simplicity referred to by the name of their first authors:

- *Wilson* (Wilson and Walker, 1987) - a theoretical model based on the theory of buoyant plumes by Morton et al. (1956) which estimates the mass eruption rate *Q* by:

$$Q_{Wilson} = (H/c)^4 \quad (2)$$

where *H* denotes the plume height (in m) and *c* is a constant which is 236 m $(s/kg)^{1/4}$.

To the group of non-explicitly wind-affected numerical models tested in this study we include three derivatives of the Wilson model, which use Eq. (2), but with different constants *c* based on empirical data from plumes with basaltic or andesitic tephra. They are referred to as:

- *Wehrmann* (Wehrmann et al. 2006), which uses for *c* a value of 295 $m(s/kg)^{1/4}$.
- *Scollo* (Scollo et al. 2007), using 247 $m(s/kg)^{1/4}$ for *c*.
- *Andronico* (Andronico et al. 2008), for which 244 $m(s/kg)^{1/4}$ is selected for *c*.

Two empirical models were used, which are based on correlations between plume height during eruption and the quantity of tephra found in deposits:

- *Sparks* (Sparks et al., 1997) - an empirical model which approximates *Q* by:

$$Q_{Sparks} = \rho \cdot (H/c)^{3.86} \tag{3}$$

where ρ is the dense-rock equivalent (DRE) density of the tephra within the plume. Following Gudmundsson et al. (2012), a DRE density of 2600 kg/m^3 was used for all computations and the constant *c* is given by Sparks et al. (1997) as 1670 $m(s/kg)^{1/3.86}$.

- *Mastin* (Mastin et al., 2009) – an empirical model with ρ defined as in (3) which estimates the mass eruption rate by:

$$Q_{Mastin} = \rho \cdot (H/c)^{4.15} \tag{4}$$

where *c* is given by Mastin et al. (2009) as 2000 $m(s/kg)^{1/4.15}$.

Since these models are empirically constrained they reflect the influence of the wind on plume-height to some degree (Mastin 2014) and can therefore be seen as implicitly wind-affected.

To juxtapose these simple plume models with explicitly wind-affected models we included two examples of the latter category:

- *Woodhouse* (Woodhouse et al. 2013) – an empirical relationship of a numerical 1D model, which estimates the MER by:

$$Q_{Woodhouse} = \left(\frac{h}{0.318} \cdot \frac{1 + 4.266\widetilde{W}_s + 0.3527\widetilde{W}_s^{\,2}}{1 + 1.373\widetilde{W}_s} \right)^{3.953} \tag{5}$$

where *h* denotes the centreline plume height (in km) and $\widetilde{W}_s$ quantifies the strength of the wind shear from the ground to a reference height H_1, according to:

$$\widetilde{W}_s = 1.44 \frac{V_1}{\bar{N} H_1} \tag{6}$$

with V_1 being the wind speed at H_1 and $\overline{N}$ being the average buoyancy frequency. The latter parameter is determined by:

$$\overline{N}^2 = \frac{1}{h}\int_0^h N^2(z)dz = \frac{1}{h}\frac{g}{C_{a0}T_{a0}}\int_0^h \left(1 + \frac{C_{a0}}{g}\frac{dT_a}{dz}\right)dz \quad (7)$$

with *g*, *C*, and *T* being the gravitational acceleration, heat capacity and temperature. Parameter *z* denotes the vertical coordinate above the source, subscript *a* refers to the atmosphere, and *0* to the volcanic source vent height. Note that a plume's top height *H* coincides with the elevation of its centreline *h* only for strong eruptions with vertical ash columns (Mastin 2014), which introduces an error when using *H* as an input for a weak or moderate eruption. Since for Eyjafjallajökull 2010 the exact conversion from *H* to *h* is unknown, for simplicity, in this study we assumed that the centreline height *h* was always 10% lower than *H*.

- *Aubry* (Aubry et al. 2017): A scaling relationship which uses multiple linear regression to relate the logarithm of the MER to the logarithm of the plume height *H* (in km), the logarithm of the average buoyancy frequency $\overline{N}$ (eq. (7)), and a windscaling parameter $\mathcal{W}_0^*$:

$$\log\left(Q_{Aubry}\right) = C + C_H \log(H) + C_N \log(\overline{N}) + C_W \mathcal{W}_0^* \quad (8)$$

where the constants $C = 7.4$, $C_H = 3.4$, $C_N = 2.3$ and $C_W = 1.1$ are empirically determined. The wind scaling parameter $\mathcal{W}_0^*$ is described by:

$$\mathcal{W}_0^* = \frac{W}{U_0} = \frac{W}{1.85\sqrt{Rn_0T_0}} \quad (9)$$

where *W* is the horizontal wind speed averaged over the plume height, U_0 is the eruption exit velocity, R=461.5 J/K/kg is the gas constant of volcanic gas at vent, n_0 is the magma volatile content, and T_0 is the exit temperature. Here we assume that n_0 = 3 wt% (Woodhouse et al. 2013) and T_0 = 1323 K for Eyjafjallajökull(Keiding and Sigmarsson 2012).

For the goals of this study, we put our focus on the simple plume models. These have the advantage that they use plume height as the only independent variable. In their entirety, the six simple models represent the empirical variation of the predictions by the theoretical *Wilson* model. In contrast to simple models, the more sophisticated explicitly wind-affected models depend on additional variables and parameters, (e.g., atmospheric parameters, wind entrainment coefficient, volatile content, magmatic temperature, top vs centreline height) which are further sources of uncertainties that have to be taken into account when analysing the model results.

Plume height data handling strategies

The radar data set is based on sampling intervals of 5 minutes, but the data set is incomplete (Arason et al. 2011). Reasons for missing data were (Arason et al. 2011):

i) The plume was below the visible horizon from Keflavík, which limited the lowest observable plume to 2.5 km.
ii) The plume was obscured by precipitating clouds.
iii) The radar was malfunctioning.
iv) The plume was absent, and the data gap was "real".

Here we explore 22 strategies to construct a data set with an improved coverage of the complete eruption (see Table 2). In the example of Figure 2a, a data set was segmented into 3h and 6h static blocks, composed of 36 and 72 bundled five-minute slots, respectively, illustrating how much of the eruption was covered according to the blocks. For example, a data coverage of 50% means for a 6h block that 36 of its 72 entries are missing. In this study, we distinguish between data voids and gaps. *Voids* are defined as empty or missing 5-minute slots in a block that otherwise contains data. In contrast, a completely empty block of data is called a *gap*. Figure 2b and c summarize the 3h- and 6h gaps in the radar records of the Eyjafjallajökull 2010 eruption.

The tested strategies can be divided into five main categories (see Table 2). Aspects included in the tested strategies are: sampling method, void reconstruction, gap reconstruction, timebase, time steps and a factor introduced to correct for data reduction bias (Y-correction):

Static plume height reconstruction (*SH*)

Static plume height reconstruction (*SH*) strategies bundle data by splitting the data set into static time intervals (blocks) and using plume height averages. The vast majority of studies using plume heights for MER modelling apply *SH*-strategies. Studies on Eyjafjallajökull used the data by Arason et al. (2011) with blocks typically ranging from 1 h (Woodhouse et al. 2013) to 6 h (Kaminski et al. 2011; Bursik et al. 2012; Mastin 2014). Others report plume heights averaged over eruptive episodes that can include multiple eruptive phases (Aubry et al. 2021).

***Sampling method**:* SH strategies applied static sampling whereby the whole data set was simply split into fixed time intervals (static blocks) of a given uniform duration (see Fig. 2f).

***Void reconstruction**:* Voids, i.e., missing plume height values within a block, were filled with the mean value of the interval (see Fig. 2e).

***Gap reconstruction**:* When applied, the heights of the missing blocks were reconstructed. The missing blocks were replaced by linearly interpolating between the previous and following block (see Fig. 2f). Where only void reconstruction but no gap reconstruction was used, strategies are labelled

with *gap*. In contrast, static sampling strategies which applied both void and gap reconstruction are termed *interpolated (int).*

Timebase: describes the width (i.e., duration) of a block. *SH*-strategies were tested for four timebases, covering the range from highest to lowest resolution (see time steps).

Time steps: describes which time interval is used for data sampling. For static sampling strategies, the time step is identical to the timebase.

No data bundling is required when using five-minute time steps. In this case the integrated mass M_{5min} over the tested eruption period of duration D (in seconds) was calculated by:

$$M_{5min} = \sum_{j=1}^{f} Q_{j,model}(H_{5min,j}) \cdot \Delta t \quad (10)$$

with the time step Δt=300s, $f = D/\Delta t$, and $Q_{j,model}(H_{5min,j})$ being the mass eruption rate Q_j provided by a specific plume model with $H_{5min,j}$ as input parameter.

By increasing the time steps, the data coverage is increased (see Fig. 2) at the cost of temporal resolution. For example, when using static blocks with a time step of 3h, the integrated mass M_{3h} is given by:

$$M_{3h} = \sum_{j=1}^{f} Q_{j,model}(H_{3h,j}) \cdot \Delta t \quad (11)$$

where $H_{3h,j}$ is the 3 h - averaged plume height and the time step Δt=300s × 36 = 10,800s. Correspondingly, Δt is doubled when using a 6h time step, and a 6h plume height average is used as model input. As time step of lowest temporal resolution, we used the complete eruptive phase for a static block (see Table 1):

$$M_{phase} = Q_{model}(H_{phase}) \cdot \Delta t \quad (12)$$

with Δt being the phase's duration and H_{phase} being the phase-averaged plume height. Equation (12) was also applied for testing the complete eruption with $\Delta t = E$, where E is the whole recorded eruption period. The corresponding diagrams are labelled as *all* (for all phases). In total, we used four different time steps for *SH* strategies (*5min, 3h, 6h, phase*). For example, an *SH*-strategy with a timebase of 6 h and applied gap reconstruction is denoted *SH_6h_int*.

Y-correction*:* To adjust for the bias induced by data reduction, four of the tested *SH* strategies multiplied the resulting masses with a correction factor $1/Y$ (see below). Strategies with Y-correction use the additional letter "Y" as identifier. For example, if Y-correction was applied to the strategy *SH_6h_int*, it is denoted *SH_Y6h_int*.

Static Sampling of five-minute intervals (*S_5min_gap*)

Being a special case of the *SH* strategy, the *S_5min_gap* strategy uses equation (10) for the computation of the total erupted mass and does not apply gap reconstruction. Since it uses the highest possible temporal resolution, voids cease to exist, and every missing data is a gap, which is treated as a "real measurement" by the *S_5min_gap* strategy.

Static MER reconstruction (*SM*)

When using average plume heights for reconstructing voids and gaps, we would expect to produce a systematic arithmetic error in the computation of erupted mass by the equations (10)-(12), since, according to equation (1), the average plume heights do not yield representative MERs. The static MER reconstruction (*SM*) strategy follows therefore a different path. Instead of using average heights, void and gap reconstruction is applied by reconstructing the "missing" MERs for each time step.

***Sampling method**: SM* strategies apply static sampling.

***Void reconstruction**:* Voids were filled with the average MER value of the interval. The erupted mass *M* of a block *B* was therefore computed by:

$$M_{3h,B} = \left(\sum_{j=1}^{36} Q_{j,model}(H_{5min,j}) \cdot \Delta t\right) \cdot \frac{36}{(36-\lambda)} \quad (13)$$

$$M_{6h,B} = \left(\sum_{j=1}^{72} Q_{j,model}(H_{5min,j}) \cdot \Delta t\right) \cdot \frac{72}{(72-\lambda)} \quad (14)$$

where λ is the number of voids in block *B* and Δt = 300 s.

***Gap reconstruction**:* When applied, the MERs of the missing blocks were reconstructed. The missing intervals were replaced by linearly interpolating between the masses of the previous and following block.

Timebase: Two different timebases were tested (3 h and 6 h) with the *SM* strategies. In addition, we tested a strategy with a five minute timebase and MER-reconstructed gap (*SM_5min_int*).

Time steps: see timebase.

Dynamic plume height and MER reconstruction (*DHM*)

The DHM strategies combine plume height and MER reconstruction, and apply dynamic sampling, which is particularly useful for real-time monitoring purposes (Dürig et al. 2018; Dioguardi et al. 2020).

***Sampling method**:* Dynamic sampling, using blocks of a uniform duration with moving start and end times. The width of the block is defined by the timebase, while the temporal increment with which

the block is moved is described by the time step. For example, a dynamic sampling strategy using a 3h timebase means that at any given time the data of the last 3 hours are considered.

***Void reconstruction**:* Voids were filled with the interval-averaged plume heights, as described for *SH* strategies.

***Gap reconstruction**:* The missing blocks were replaced by linearly interpolating between the masses of the previous and following block, as described for *SM* strategies. For the presented DHM strategies, gap reconstruction was always applied.

Timebase: Two different timebases were tested (3 h and 6 h) with the *DHM* strategies.

Time steps: All *DHM* strategies strategies used five-minute time steps.

The REFIR strategies (*REFIR*)

The strategies discussed so far produce best MER estimates independently from plume height uncertainties. To consider these, the *DHM* strategies were modified by computing the MER for a dynamic block *B* with:

$$Q_{model,B} = (Q_{model}(H_B - s_B) + Q_{model}(H_B) + Q_{model}(H_B + s_B))\ /3 \qquad (15)$$

where H_B is the average height and s_B the plume height uncertainty of the block at time *t*.

These strategies were computed with the software REFIR (Dürig et al. 2018; Dioguardi et al. 2020), which uses equation (15) to obtain a best MER estimate. We note that while the software is designed to apply a combination of models linked by weight factors, here we applied it exclusively to individual models.

***Sampling method**:* Dynamic sampling, see DHM.

***Void reconstruction**:* see DHM.

***Gap reconstruction**:* see DHM.

Timebase: Four different timebases were tested (15 min, 30 min, 1 h and 3 h) with *REFIR* strategies.

Time steps: All *REFIR* strategies used five-minute time steps.

Erupted tephra mass

The erupted mass *M* was used as validation parameter to test the capability of the assessed models and data handling strategies in reproducing mass eruption rates. The individual model-derived results for the four main stages of the eruption were compared to the total masses obtained for each

of the analyzed phases (Ia, Ib, II, III, IV and all, see Table 1), based on the mapped fallout and exponential dispersal models for the offshore part (Gudmundsson et al. 2012). We denote the measured range of uncertainty *ME*.

Computation of *Y*-ratios

To study the impact of data reduction on prediction quality, a subset of 49 selected six-hour blocks was used, for which the radar data provides high data coverage (containing at least 60 of 72 data points measured at five-minute intervals). The voids were filled with 6h averages, as described for the *SH* strategies. For each of the 6h blocks with the best data coverage the predicted mass erupted M_i was computed, based on:

I. Five minute radar height data H_{5min}. The integrated mass M_{i_5min} was calculated by using eq. (8) with f = 72.

II. Mean, median and maximum heights (H_{3h_avg}, H_{3h_med}, H_{3h_max}) within two 3h intervals. The integrated mass M_{i_3h} over a 6h interval was quantified by:

$$M_{i\_3h} = \sum_{j=1}^{2} Q_{j,model}(H_{3h,j}) \cdot \Delta t \quad (16)$$

with Δt=10,800s (= 3h) and $H_{3h,j}$ being H_{3h_avg}, H_{3h_med}, H_{3h_max}..

III. Mean, median and maximum heights (H_{6h_avg}, H_{6h_med}, H_{6h_max}) within the 6h interval, using

$$M_{i\_6h} = Q_{model}(H_{6h}) \cdot \Delta t \quad (17)$$

with Δt=21600s (= 6h) and H_{6h} being H_{6h_avg}, H_{6h_med} or H_{6h_max}, depending on the tested procedure.

Varying plume model, time interval (3h, 6h) and statistical treatment (mean, median, maximum) results in 24 estimates for M_i for each of the *49* analyzed 6h blocks. These mass estimates were compared to the respective value M_{i_5min} based on the 5 minutes data sets, by generating the ratios *Y*:

$$Y_i = M_i / M_{i\_5min} \quad (18)$$

Strategies which apply Y-correction (see Table 2) multiply the estimated mass *M* with the according time interval-specific correction factor 1/*Y*.

Plume height uncertainties

We take two types of plume height uncertainties into account for the MER calculations: i) the observation uncertainty of the corrected radar plume heights, and ii) the uncertainties introduced by using the average plume height of a block to calculate the MER.

The uncertainty of the radar, ΔH, is assumed to be 0.5 km based on the maximum uncertainty of the photo heights and the uncertainty estimates given in Aubry et al. (2021). The uncertainty arising from the block averaging of plume heights is assumed to be equal to two standard deviations σ within

each block, corresponding to a 95% significance level. The total plume height uncertainty of a block s_B is then calculated as:

$$s_B = \sqrt{\Delta H^2 + \left(2 \cdot \sigma(H_{block})\right)^2} \quad (19)$$

For strategies where no height averaging is used, σ is equal to zero, and therefore only radar uncertainties are considered.

Erupted mass uncertainty

The results obtained depend critically on the error in erupted mass. The isopach map reported in Gudmundsson et al. (2012) is based on about 400 data points obtained during and immediately after the eruption. Post-depositional erosion and compaction that often limit the accuracy of tephra layer thickness and size (Biass et al. 2013; Engwell et al. 2013) are therefore considered to be minor for the Eyjafjallajökull data. Good spatial coverage in the near field where thicknesses are several meters on a relatively regular glacier surface result in relative errors in thickness of only a few percent. In the far field the relative errors rise, reaching up to 40% for thicknesses <0.2 cm. The resulting error reported for volume on land is 15% (Gudmundsson et al., 2012). For the offshore part the error was estimated as 40%, based on minimum and maximum extrapolations that are consistent with observations in the Faroe Islands and mainland Europe (for more detail see methods in Gudmundsson et al. 2012). The resulting best estimate of maximum total error was 25%. The contemporary measurements of density of the tephra yielded 1400+/-40 kg m^{-3} (Gudmundsson et al., 2012) a relative error of 3%. Treating these two errors as independent, the effect of the error in density is negligible, resulting in a relative error in mass of 25%.

Results

Plume heights and ERA5 wind speeds

In Figure 3, the radar-derived plume heights are compared with plume heights obtained by photos. The comparison of the 165 data pairs suggests that on average, photo-based plume heights show an offset from the radar data of 0.5 km (see Figures 3b and 3c). As a first step for data quality improvement, this systematic shift was therefore corrected by adding 0.5 km to all radar-derived plume heights (see Figure 3d).

Figure 4a shows the wind speed profiles over the vent during the Eyjafjallajökull 2010 eruption, based on the ERA5 reanalysis data. The highest wind speeds, with up to 64 m/s, occurred during phase I in regions above 6000 m at a time when the recorded plume heights reached similar altitudes

(for more detailed, phase-specific plots, see Online Resource 1). In Figure 4b, we show the differences in wind speeds between ICRA and ERA5.

Evaluation of plume height data handling strategies and models

Mean values and standard deviations of the resulting *Y*-ratios based on the 49 selected 6h blocks are presented in Table 3. Three statistical treatments were tested for model input, using average, median and maximum plume height, respectively. The larger the deviation of *Y* from 1, the less reliable the tested method of approximation (assuming M_{i_5min} to be the true value). The *Y*-ratios closest to 1 are found using median plume heights. However, using the median also results in significantly larger standard deviations compared to average plume heights. When using 3h- or 6h-blocks, all models must be corrected with 1/*Y* to be comparable to the summation of masses with 5min time step. For *Y*-corrections we focused on averaged heights, for which *Y* is almost constant for all models. *Y* ranges between 0.89 and 0.91 for 6h-blocks and 0.91-0.92 for 3h-blocks. Therefore, *Y*-correction strategies for 6h-blocks (denoted *Y6h*, see Table 2) were tested by using average heights and a model-independent correction factor of 1/0.9 = 1.1. Accordingly, a correction factor of 1/0.915 = 1.09 was used for *Y3h* strategies.

Figure 5 demonstrates how the temporal evolution of the erupted mass varies, depending on the applied plume height strategy and plume model. In both cases, the whole eruption was used as the reference period, with the measured range of the erupted mass (reported in Gudmundsson et al. 2012) indicated by the red error bar. The curves of Figure 5a show the modelled erupted mass using strategy *SH_3h_gap*. This strategy uses 3h-averaged plume heights in combination with applied void-reconstruction but without gap reconstruction (see Table 2). While for this strategy both wind-affected models provide estimates that lie within the measured range, this applies only for one out of the six simple plume models tested (*Wilson*). Figures 5b-d show the outcome of *Wilson* when applying different strategies. Comparison of model-estimates to the measured mass for individual eruptive phases is illustrated in Figure 6. When using strategy *SH_3h_gap* for phase I, most of the models underestimate the erupted mass, while for phases II to IV *Wilson, Andronico, Scollo* and *Sparks* provide estimates that coincide with the measured range. The same applies to *Woodhouse* for the phases I, II and III, while this wind-affected model provides too large estimates for phase IV. The other wind-affected model tested with *SH_3h_gap*, *Aubry*, provides estimated masses that are above the measured range suggested by Gudmundsson et al. (2012), but provides estimates that are on the lower margin for phase I. This combination results in a coinciding estimate, when considering the whole eruptive period (see Fig. 5a). Systematically on the lower end of the estimated range of modelled masses are *Wehrmann* and *Mastin*, suggesting, that *SH_3h_gap* is not the optimal choice for these models. This overview is an example for just one of the 22 tested strategies with reference

to the whole eruption. The phase-specific results for *Wilson* model outcomes according to the strategies in Fig. 5b can be found in Online Resource 2.

Figure 7 compares the measured masses of all reference periods with the estimates of four models using six different strategies with a timebase of 3 hours. The ranges of measured mass for each eruption phase are marked in green. For example, for phase III the *Mastin* model (see Fig. 7b) achieves best fitting estimates when applying the *REFIR_3h* strategy. Figure 7 demonstrates the impact that the choice of a strategy has on the modelled MER prediction. Phase II is the phase with the most abundant gaps in the radar records (see Fig. 2). Thus, it is not surprising that this is also the phase where the selected strategy has the biggest influence on the predicted MER. For example, for *Mastin* the prediction with sampling strategy *REFIR_3h* is 246 % larger than with *SH_3h_gap*.

As a general finding, applying sampling strategies with gap reconstruction (*int*) always results in larger masses than applying no gap reconstruction methods (*gap*). The *SM_3h_int* strategies provide systematically larger results than *DHM_3h* or *SH_3h_int* strategies, due to the non-linear relation between H and Q. For the same reason, the mass eruption rates $Q_{model}(H_B + s_B)$ and $Q_{model}(H_B - s_B)$ in equation (15) are not symmetrically distributed towards $Q_{model}(H_B)$. Consequently, the *REFIR_3h* strategy provides larger values as best MER estimate compared to other strategies.

When using the whole eruption as the reference period, the effect of the selected timebase for the *REFIR* strategy on the model outcome (i.e., the erupted mass estimates) is at maximum 6.8% and therefore small in comparison to the modelled mass uncertainties (see Online Resource 3). The largest effect of the selected timebase on the outcome is found for phase IV, where the model results vary with 12.0% at maximum (see Online Resource 4). Presented results for *REFIR* methods in the following are limited to the two endmembers (*REFIR_3h* and *REFIR_15min*).

For the static sampling strategies *SH* and *SM*, the timebase can have some considerable impact on the model outcome, especially for phases with abundant gaps. For example, for phase II, the mass predictions increase by 39 to 42% (depending on the model), if strategy *SH_6h_gap* is applied instead of *SH_3h_gap*.

Discongruence indices and congruency maps

A result table containing the modelled masses for all phases, data handling strategies and models can be found in Online Resource 5. In order to systematically assess how well the model-estimates coincide with the measured data, we defined discongruence indices (*DI*s) in Table 4, based on the error *ME* of measured mass data (Table 1). A model prediction that lies within the range of the measured mass is characterized by a discongruence index of 0. The more the model output deviates from the ground truth, the larger the *DI*. To better visualize the results, we produced congruency

maps, where each index is assigned a colour. Examples for such congruency maps are shown in Figure 8. There, the dark blue fields mark methods and strategies that lead to estimates that coincide with the range of mass from ground-based measurements. Combinations of strategies and plume models, which provide estimates that are just outside of the range of the measured values are marked in light blue. Predictions that considerably over- or underestimate the erupted mass (*DI* of 5) are displayed in orange. The resulting congruency map for the complete phase is presented as Online Resource 6, and all phase-specific results are provided in Online Resource 7.

Based on the resulting *DI*s, it was possible to construct tables with their averaged values that allow us to analyse the prediction qualities of models and data handling strategies for individual eruptive phases, as well as for the complete eruption (Tables 5 and 6).

Phase-specific comparison of plume height data handling strategies

Table 5 lists for each phase the strategies' success, represented by discongruence indices averaged over all tested plume models (separated between simple and explicitly wind-affected models). When using the entire eruption (*all*) as the reference period, the overall average *DI* for all strategies is 0.7. According to the results, the most successful strategy for the simple models is *REFIR_3h* with a *DI* of 0.3. For wind-affected models a number of strategies resulted in predicted mass values congruent to the measured mass, including *SH_6h_gap, SH_6h_int, SM_3h_gap, DHM_6h and DHM_3h*.

According to the model-specific *DI*s averaged over all strategies (Table 6), the models by *Wilson* and *Sparks* were found to have the highest prediction success rate, with DIs of 0.05 and 0.015, respectively, when considering the entire eruption. When studying individual phases, *Andronico* and *Sparks* turn out to be most successful, with an average *DI* of 0.05 for phase III.

Below we summarize the results for each of the phases:

Phase Ia: the overall average *DI* is 2.16, significantly larger than for any other phase. According to Table 6, the most successful strategy for both simple and explicitly wind-affected models is *REFIR_15min* with *DIs* of 1.7 and 0.5, respectively.

Phase Ib: here strategies provide higher success rates. Optimal strategies for simple models are the *REFIR* strategies, as well as *SM_6h_gap* and *SM_6h_int,* featuring *DI*s of 0.5. For wind-affected models, however, other strategies (e.g., *SH_6h_int* or *DHM_6h*) provide better results, while *REFIR* strategies result in overestimates of predicted masses.

Phase II: due to relatively weak activity, phase II has the lowest data coverage per block (see Fig. 2d). Many strategies and models tend to overestimate the erupted mass of this phase. This applies especially to strategies where gaps were filled by interpolated data (*DHM*, *REFIR* and static sampling

methods with gap reconstruction). For simple models, the best results with a *DI* of 0.3 are found for *SM_6h_gap*, *SH_6h_gap* and *SH_phase*, which are strategies with large timebases and no gap reconstruction. For wind-affected models, the best results were achieved with *SM_3h_gap*, *SH_3h_gap* and *SH_Y3h_gap.*

Phase III: featured increased explosive eruption at relatively low wind speeds. There, strategies showed the highest success rates of all phases with an average *DI* of 0.45 (0.5 for simple models, 0.2 for explicitly wind-affected models). The most successful strategies for simple models were the *REFIR* strategies with a *DI* of 0.3. For *Woodhouse* and *Aubry* most of the tested strategies result in predictions that coincide with the measured mass. Exceptions are the *REFIR* strategies (*DI* = 1.0) , and strategies that apply Y-correction (*DI* = 0.5).

Phase IV: For the weaker final eruptive phase, the optimal strategies for simple and explicitly wind-affected models were found to be *REFIR_15min* (*DI* = 0.2) and *S_5min_gap* (*DI* = 0), respectively.

Effect of using ERA5 vs ICRA weather data

So far, the results presented for the explicitly wind-affected models (*Woodhouse*, *Aubry*) are based on ERA5 wind speed data. How much do the results of these models (i.e., estimated erupted mass) change, when the ICRA data of (considerably) higher spatial resolution are used? In most cases (except for phase IV), the ICRA data leads to slightly larger mass estimates, but these changes are almost insignificant, considering the uncertainties (see Online Resource 8 for *Woodhouse*). For example, when applying the *REFIR 3h* strategy to *Woodhouse*, the differences between model-results based on ERA5 versus ICRA wind speeds is only approximately 2% for the whole eruption (and 1.3% of the uncertainty). The absence of sizeable changes can be explained by the fact that both sets of weather data feature similar wind speeds over most of the eruption period (see Fig. 4b). The only major difference (of more than 10 m/s) occurred during phase II, on 29 April below 2000 m a.s.l. For this phase, the modelled mass differed by 4.7%, which is small considering the model uncertainties. We thus conclude that for reanalysis of the Eyjafjallajökull 2010 eruption, using weather data of higher spatial resolution only effects the overall outcome in a minor way. However, this result should not be generalized. A reanalysis of other eruptions, especially short-termed ones, will probably benefit from a higher spatial resolution of weather data.

Wind effect on plume deflection and plume model predictions

The horizontal deflection of the plume top plotted over the contemporary wind speed at plume top level, based on ICRA reanalysis weather data (Figure 9a) shows an approximately linear correlation between these parameters. This is also indicated by a Pearson correlation coefficient (*R*) of ~0.68. For Figure 9a, wind speeds at variable elevation were used. To find out for which fixed altitude the wind speeds correlate best with deflection, the according correlation coefficients were plotted, using wind

speeds of both ERA5 and ICRA data (Fig. 9b). Next to the raw data (*ERA5 1h* and *ICRA 1h*), wind speeds averaged over 2h and 3h were also used. The best correlation at fixed elevation is observed for 2h-averaged ERA5 wind speeds at 2900 m a.s.l. (R = 0.65). At that altitude ICRA wind speeds correlate slightly less with deflection (R = 0.63), but below 2500 m a.s.l. it is the ICRA data that shows a better correlation with wind speeds compared to ERA5.

From the models tested, not only *Woodhouse* and *Aubry* incorporate a wind-effect. There is an implicit dependency on wind for the other (simple) models as well, as revealed by Figure 10. There, the ratio of modelled vs measured mass was plotted over phase-averaged wind speeds. For the modelled mass, three non-explicitly wind-affected models were investigated in combination with four strategies: *Wilson* (*Wil*, as representative for the simple numerical 0D models), *Sparks* (*Spa*) and *Mastin* (*Mas*). Here we selected strategies that do not apply gap reconstruction, to avoid the bias caused by the fact that there is a large variance in data gaps between the different phases. Two general trends (indicated by trendlines) can be observed:

1) At low wind speeds the different strategies and models lead to predictions with a large span. This range decreases for larger wind speeds.
2) With higher wind speeds all shown strategies and non-explicitly wind-affected models tend to increasingly underestimate the actual erupted mass. While for a wind speed of 14.5 m/s $M_{modelled}$ / $M_{measured}$ averages at around 103%, this ratio drops for 16.2, 19.2 and 36.4 m/s to 79%, 67% and 46%, respectively. Thus, for wind speeds over 20 m/s the model predictions underestimate the erupted mass on average by more than 40%.

Both trends are also confirmed in Figure 8, which compares the congruency maps of high-wind eruptive phase Ia (Fig. 8a) with the phase of lowest wind speeds (phase II, Fig. 8b), thus contrasting the two end members of Fig. 10. Compared to phase Ia, phase II was of much longer duration, and it was characterized by abundant data gaps, which should result in large differences between gap-reconstructing strategies and those which ignore them. Yet, for phase II, 54 combinations of strategies and models provide predictions that lie within the uncertainties of the measurements *ME* (tiles marked with dark blue color), as opposed to only 14 combinations found for phase Ia. While for most combinations, the explicitly wind-affected models *Woodhouse* and *Aubry* appear to be robust against this wind effect, for the simple models only very few combinations give predictions close to the measurements.

Discussion

The comparison between radar and photographic data suggests that the former systematically underestimated the plume height by 0.5 km. This finding is in contrast to the assessment of webcam

images analyzed by Arason et al. (2011), which suggested that the radar over-estimated the plume heights. These webcams were, however, located to the north and west-northwest of the volcano, while the plume was deflected away from the cameras, towards the south to southeast. Using the distance to the vents instead of the plume top would therefore lead to an underestimation in webcam-derived plume heights which might explain the differences with our results (see Fig. 3d). For example, a plume top of 5 km height that is shifted towards the south with a deflection *d* of 6 km would be perceived as being only 3 km high on images produced by the webcam at Þórólfsfell, located 9 km to the north from the vents. The difference between photos and radar measurements is mainly a consequence of the radar's horizontally stepping scanning procedure and the eruptive and atmospheric circumstances during the 2010 eruption. This implies that such an offset could change in a future eruption at Eyjafjallajökull, even if the same radar sensor is applied.

The data sets on which the empirical models are based include observations for various situations regarding wind speed, but it is known that scenarios of weak eruptions under strong wind-conditions are underrepresented (Mastin 2014). Furthermore, under high wind speeds plume height is reduced, due to the facts that the plume is bent-over and that there is more efficient turbulent entrainment of air into the column (Bursik 2001). For the phases of weak activity and relatively low wind speeds, however, the empirical models by *Mastin* (phase II with 0.35) and *Sparks* (phase IV with 0.10) still have shown to be among the ones with highest prediction success.

As a general finding, the explicitly wind-affected models tend to have larger ranges of uncertainty and over rather than underestimate the measured mass (see, e.g., Fig. 11). Results from the simple models fall at the other end of the modelled range, with *Wehrmann* being the most extreme example, which tend to systematically underestimate the mass eruption rates, especially under windy conditions. This discrepancy between estimates from simple and explicitly wind-affected models under windy conditions is consistent with previous findings (Bursik 2001; Mastin 2014).

When analyzing the entire eruptive period by comparing the strategy-averaged *DIs* for simple plume models (see Table 5), the *REFIR_15min* approach turned out to be the optimal strategy. In contrast, using *REFIR* strategies with explicitly wind-affected models lead often to overestimated mass predictions. These models are affected by larger uncertainties than the simple plume models (see, e.g., Fig 11), which leads to larger MER predictions, due to the definition of the best estimate by equation (15).

Of particular interest is also the comparison between phase Ib and III, which were similarly well covered (9.3 vs 9.2 plume height measurements per hour), had only slightly dissimilar prevailing wind speeds (19.2 vs 16.2 m/s), but differed in eruptive styles: phase Ib was phreatomagmatic, whereas

phase III was magmatic. Despite different outcomes for most of the other strategies, the *REFIR* strategies appear to be the most robust for both types of eruptive styles. This robustness is noteworthy, considering that phreatomagmatism is based on thermohydraulic fuel-coolant interaction processes (Wohletz 1986; Büttner and Zimanowski 1998; Dürig et al. 2020), which produce more fine-grained ash particles, release larger kinetic energy and generate steam - all of which are factors that are expected to affect the plume rise behaviour (Koyaguchi and Woods 1996; Sparks et al. 1997).

An important choice when selecting the plume height data handling strategy is the decision of whether gap reconstruction is applied or not. In a real-time monitoring scenario (as well as in re-analyses), where radar is the only source for plume heights, it is difficult to distinguish "real" gaps from data missing for other reasons. For example, it was suggested that the data gap between 15 and 16 April 2010 (within phase Ia) was real, and a consequence of low plume heights (Mastin 2014). According to our results, however, the optimal strategies for that phase are those that apply gap reconstruction (e.g., *REFIR_15min*), indicating that the gap was not reflecting the status of the actual plume. This is corroborated by photographs taken on that day, that prove the existence of a plume in that period. In contrast, our findings for phase II indicate that the gaps during that phase were real. In such scenarios of very weak eruptive activities, strategies with large timebases (such as *SH_phase*) or applied gap reconstruction (such as *SM_6h_int*) result in overestimated mass predictions, especially for explicitly wind-affected models. For modelling long-lasting eruptions with diverse episodes, comparable to Eyjafjallajökull 2010, we would therefore recommend a mixture of strategies, tailored to phase and model-type.

Similar to applying gap reconstruction, using a too large timebase result in real data gaps being missed. Using a too short timebase, on the other hand, might lead to errors when monitoring an evolving plume that has not reached stable buoyancy yet. Since this would violate the steady state assumption of the models, the model predictions would not be reliable. We consider the latter effect, however, as insignificant, when modelling a long-lasting eruption like Eyjafjallajökull 2010.

A key finding of this study is the dominating influence of the wind speed on the prediction quality of plume height data for mass eruption rates, at least for moderate eruptions with bent-over plumes. This is highlighted by the results of the phase-specific comparison of plume height data handling strategies (Table 5) and illustrated in Figures 8 and 10. Phase Ia is characterized by the highest wind speeds, and despite having a good data coverage (8.1 measurements per hour, see Table 1) the strategies are on average considerably less successful than for phases Ib and III. All our results indicate that for a moderate-to-weak eruption like Eyjafjallajökull 2010, it is increasingly challenging for wind speeds exceeding ~20 m/s to find combinations of plume models and data handling

strategies with which the mass eruption rate can be accurately predicted in real-time. This inference highlights how important it is to monitor such eruptions in real-time and use the 'optimal' combination of plume height data handling strategy and plume models.

The strategies presented were tested by focusing only on the best MER estimates, without considering the ranges of uncertainty (with the *REFIR* strategies being notable exceptions). These can become substantial and exceed the span of the measured mass by several orders of magnitude (Figure 11, for uncertainty ranges in modelled MERs, see also online resource 5). The uncertainties for $M_{modelled}$, as defined by equation (19), depend on model, timebase, strategy and the relative data coverage of the studied period. For example, when considering the entire eruption period, the uncertainties for *SH_3h_int* increase by 9 - 12% (depending on the model) compared to *SH_3h_gap* (15 - 17% for *SM_3h_int* vs *SH_3h_gap*). For phase II alone, the uncertainty ranges between the two strategies increase by up to 41% (up to 43% for for *SM_3h_int* vs *SH_3h_gap*).

Selecting a short timebase has a significant effect on the modelled mass uncertainties. For example, compared to the 3h-timebase, the *REFIR* uncertainties for a timebase of 15 minutes is reduced by up to 24% when considering the whole eruption period. For phase IV, this decrease reaches 32.5%. The largest uncertainties are found for the largest possible timebase, with strategy *SH_phase* (see Fig. 11b).

It is important to note that our error estimates did not consider model-inherent uncertainties (an example for the uncertainties of the *Mastin* model is shown in Fig. 11a). For some strategies, the model uncertainties could be larger than the ranges defined by equation (19). A model-strategy combination with a large *DI* does not necessarily provide a "wrong" prediction, since its uncertainty interval could (at least partly) overlap with the measured range. On the other hand, combinations that result in predictions affected by a large uncertainty range are unfavorable for real-time monitoring purposes. Our approach is optimized for finding the model-strategy pairs which provide the best estimates that are closest to the measurement, but we note that this is not the only method to assess the quality of strategies and models, and more advanced comparison methods might also consider the range and overlap of modelled mass uncertainties.

The radar data by Arason et al. (2011) has been used by a plethora of studies on the 2010 ash plume of Eyjafjallajökull (Kaminski et al. 2011; Bursik et al. 2012; Gudmundsson et al. 2012; Degruyter and Bonadonna 2012; Devenish 2013; Ripepe et al. 2013; Woodhouse et al. 2013; Mastin 2014; Dürig et al. 2015b; Dioguardi et al. 2020). Our suggested height correction still lies within the reported error margins, but when used as input for plume models it will increase the MER predictions. Due to our definition of uncertainties by equation (19), for strategies that do not apply height reconstruction,

699 the plume height uncertainties are only dependent on the radar height uncertainty itself. In Figure
700 11b, the effect of the height correction on the modelled mass is represented by the distance
701 between the lower error bar and the best estimate for these strategies (e.g., all *SM*-strategies).

702 In their study on Eyjafjallajökull, Woodhouse et al. (2013) applied an *SH* strategy with a timebase of
703 1h, but using the maximum plume height within each block instead of plume height averages. We
704 note that the 10% plume height reduction applied to convert *H* into *h* for *Woodhouse* is in the same
705 range as the height correction applied to the radar data. We therefore used comparable plume
706 heights as input as Woodhouse et al. (2013). Using the maximum plume heights per block instead of
707 averages resulted in MER predictions for phase Ia that exceeded our estimates by an order of
708 magnitude. We suggest that, when using *Woodhouse* to re-model the Eyjafjallajökull 2010 eruption,
709 using data handling strategies like *SH_3h_gap* with block-averaged plume heights might be the more
710 suitable strategy.

711 The *SH* strategies are the most commonly used for plume modelling (Kaminski et al. 2011; Bursik et
712 al. 2012; Degruyter and Bonadonna 2012; Woodhouse et al. 2013; Mastin 2014; Dürig et al. 2015b),
713 and our findings indicate that these strategies (especially *SH_3h_gap*) are often a good choice when
714 using explicitly wind-affected models, but they are far less successful with simple models. From the
715 dynamic sampling strategies, which are typically used for real-time monitoring, *DHM* strategies prove
716 to provide best results for explicitly wind-affected models. When used with simple models, however,
717 the optimal choice is the *REFIR* strategy.

718

719 Conclusions

720 Focusing on the different eruptive phases of the Eyjafjallajökull 2010 eruption, we used aerial and
721 ground-based photographs of the plume together with ground-based measurements of the erupted
722 mass to study and compare the predictive qualities of eight plume models in combination with 22
723 data handling strategies. Although the latter has a significant influence on the model outcome, their
724 influence has, to our knowledge, not yet been studied to this extent before.

725 The best re-analysis results for simple (non-explicitly wind-affected) plume models are in most cases
726 obtained by using strategies implemented in the software REFIR (Real-time Eruption source
727 parameters FutureVolc Information and Reconnaissance system), which apply dynamic sampling in
728 the form of a moving average and consider the plume height uncertainties. The only exception is
729 found for phase II, where the eruptive activity was weak, and the plume height was under the
730 detection limit of the radar for a considerable amount of time, causing data gaps in the records.

There, interpolating strategies generate ghost-data points, which lead to an overestimation of the erupted mass. Under such conditions, strategies without gap reconstruction and with long timebases (such as *SH_phase* or *SM_6h_gap*) are more reliable. For explicitly wind-affected models, the best strategies were found to be static sampling strategies with short timebases and without gap reconstruction (such as *SH_3h_gap*).

Our findings suggest that no single model has always the best answers, and different models and data handling strategies work best under different plume and data conditions. Which model and strategy to choose for optimal prediction results depends on data coverage, eruption intensity and, according to our findings from Eyjafjallajökull, to a lesser extent on fragmentation mechanism (e.g., magmatic vs phreatomagmatic). The dominant factor on the prediction quality, however, was found to be the wind speed. The higher the wind speed, the more the non-explicitly wind-affected models tend to underestimate the MER. Conversely, the explicitly wind-affected *Woodhouse* model overestimates it, when applying the *REFIR* strategies. Namely for wind speeds over 20 m/s, only a few combinations of data handling strategies and plume models provide accurate predictions when monitoring an Eyjafjallajökull 2010 – type of eruption.

A possible solution for obtaining accurate estimates in real time for such long-term eruptions of variable and moderate intensity under varying wind speeds would therefore be to not rely on a single combination of data handling strategy and plume model, but to be able to use a range of such combinations, thus providing the possibility to adapt to the scenario monitored.

Tables

Table 1:

Modelled eruptive phases and ground truth. Modelled phases of the 2010 Eyjafjallajökull eruption are listed, together with the estimated ranges of erupted airborne tephra mass, as reported in Gudmundsson et al. (2012). The errors of erupted mass are denoted *ME*. In our study these values were used as ground truth for testing the model predictions with different plume height data handling strategies. Additional columns indicate duration, number of recorded radar measurements, data points (i.e., plume height measurements) per hour and time-averaged wind speed for each phase. Computation of wind speeds were based on ERA5 reanalysis data for the (radar-detected) plume heights.

phase	from (dd/mm)	until (dd/mm/yy)	duration (min)	erupted mass ± *ME* ($\times 10^{10}$ kg)	number of radar measurements	data points per hour	wind speed m/s
Ia	14/04	16/04/10	3785	9.80 ± 2.10	511	8.1	36.4
Ib	17/04	18/04/10	2165	3.50 ± 0.98	335	9.3	19.2
I (total)	14/04	18/04/10	5945	13.30 ± 3.08	846	8.5	30.1
II	18/04	4/05/10	23765	4.20 ± 1.40	1005	2.5	11.8
III	5/05	17/05/10	18725	18.90 ± 4.90	2858	9.2	16.2
IV	18/05	21/05/10	4950	1.40 ± 0.42	430	5.2	14.5
all	**14/04**	**21/05/10**	**53385**	**37.8 ± 9.8**	**5139**	**5.8**	**15.6**

763 Table 2:

764 Tested data handling strategies. The columns list the aspects that were varied: sampling method,
765 void reconstruction, gap reconstruction, timebase (i.e., duration of a block), time stepping and Y-
766 correction.

Strategy	sampling	void reconstructed variable	gap reconstructed variable	timebase	time step	Y-correction
S_5min_gap	static	- n. a. -	- none -	5 min	5 min	- none -
SH_phase	static	height	- n. a. -	phase	phase	- none -
SH_6h_gap	static	height	- none -	6 h	6 h	- none -
SH_Y6h_gap	static	height	- none -	6 h	6 h	Yes
SH_3h_gap	static	height	- none -	3 h	3 h	- none -
SH_Y3h_gap	static	height	- none -	3 h	3 h	Yes
SH_6h_int	static	height	height	6 h	6 h	- none -
SH_Y6h_int	static	height	height	6 h	6 h	yes
SH_3h_int	static	height	height	3 h	3 h	- none -
SH_Y3h_int	static	height	height	3 h	3 h	yes
SH_5min_int	static	- n. a. -	height	5 min	5 min	- none -
SM_6h_gap	static	MER	- none -	6 h	6 h	- none -
SM_3h_gap	static	MER	- none -	3 h	3 h	- none -
SM_6h_int	static	MER	MER	6 h	6 h	- none -
SM_3h_int	static	MER	MER	3 h	3 h	- none -
SM_5min_int	static	- n. a. -	MER	5 min	5 min	- none -
DHM_3h	dynamic	height	MER	3 h	5 min	- none -
DHM_1h	dynamic	height	MER	1 h	5 min	- none -
REFIR_3h	dynamic	height	MER	3 h	5 min	- none -
REFIR_1h	dynamic	height	MER	1 h	5 min	- none -
REFIR_30min	dynamic	height	MER	30 min	5 min	- none -
REFIR_15min	dynamic	height	MER	15 min	5 min	- none -

767

768 Table 3:

769 Results for Y-values, based on 49 static 6h-blocks with a data coverage of 60/72 or higher. Numerical
770 0D models stand for *Wilson, Wehrmann, Scollo* and *Andronico* models. For *Woodhouse*, ERA5
771 weather data was used. Y-correction strategies multiplied the predicted mass with 1/Y.

		6h		3h	
		mean	stdev	mean	stdev
numerical 0D models	average	**0.90**	0.10	**0.91**	0.09
	median	**0.95**	0.25	**0.95**	0.19
	maximum	**2.36**	2.15	**2.06**	1.45
Sparks	average	**0.91**	0.09	**0.92**	0.08
	median	**0.96**	0.24	**0.95**	0.19
	maximum	**2.29**	2.00	**2.01**	1.36
Mastin	average	**0.89**	0.11	**0.91**	0.10
	median	**0.95**	0.26	**0.95**	0.20
	maximum	**2.44**	2.31	**2.12**	1.55
Woodhouse	average	**0.91**	0.09	**0.92**	0.09
	median	**0.97**	0.26	**0.95**	0.19
	maximum	**2.37**	2.13	**2.08**	1.46

772

773

774 Table 4:

775 List of discongruence indices (*DI*s). A *DI* of 0 refers to a model-estimate, which lies within the range of
776 measured mass according to Gudmundsson et al. (2012). This range is illustrated by red error bars in
777 Fig. 5 and Fig. 6 and specified by the uncertainty *ME*. *A DI* of 0 is therefore congruent with the
778 "ground truth". The larger the discongruence index, the more the model-output deviates from the
779 ground truth.

Discongruence Index (*DI*)	minimum deviation from measured mass	maximum deviation from measured mass
0	0	*ME*
1	*ME*	1.5 *ME*
2	1. 5 *ME*	2 *ME*
3	2 *ME*	2.5 *ME*
4	2.5 *ME*	3 *ME*
5	3 *ME*	> 3 *ME*

780

781

Table 5:

Average discongruence indices listed by data handling strategy and reference period. The first value is the resulting *DI* for the six simple models tested (*Wilson*, *Andronico*, *Scollo*, *Wehrmann*, *Sparks*, *Mastin*), the second column presents the averages for the two wind-affected models (*Woodhouse*, *Aubry*). The smaller the value, the more often fits the modelled result with the measured mass. Values representing the most successful strategies for a phase are printed in bold.

Row	**Ia**		**Ib**		**II**		**III**		**IV**		**all**		***average***	
S_5min_gap	3.0	**0.5**	1.5	**0**	2.8	1.0	0.8	**0**	2.3	**0**	1.7	**0**	*2.0*	***0.3***
SH_phase	3.0	**0.5**	2.0	**0**	**0.3**	2.0	1.3	**0**	0.5	1.0	1.0	1.0	*1.4*	*0.8*
SH_6h_gap	2.8	**0.5**	1.5	**0**	**0.3**	1.5	0.5	**0**	0.5	1.0	1.3	**0**	*1.2*	*0.5*
SH_Y6h_gap	2.5	**0.5**	1.3	0.5	0.5	1.5	0.5	0.5	0.5	1.0	0.7	1.0	*1.0*	*0.8*
SH_3h_gap	3.5	**0.5**	1.7	**0**	0.5	**0.5**	0.5	**0**	0.5	1.0	1.3	**0**	*1.3*	***0.3***
SH_Y3h_gap	2.8	1.0	1.3	**0**	0.5	**0.5**	0.5	**0**	0.5	1.0	1.2	**0**	*1.1*	*0.4*
SH_6h_int	2.8	**0.5**	1.5	**0**	0.7	2.0	0.5	**0**	0.5	1.0	1.0	**0**	*1.2*	*0.6*
SH_Y6h_int	2.5	**0.5**	1.3	0.5	0.7	2.5	0.5	0.5	0.3	1.5	0.5	1.0	*1.0*	*1.1*
SH_3h_int	3.0	1.0	1.7	**0**	0.8	2.0	0.5	**0**	0.5	1.0	0.8	**0**	*1.2*	*0.7*
SH_Y3h_int	2.8	**0.5**	1.3	**0**	0.7	2.0	0.5	0.5	0.3	1.5	0.5	1.0	*1.0*	*0.9*
SH_5min_int	2.7	1.0	1.3	**0**	0.7	2.0	0.5	**0**	0.5	1.0	0.5	1.0	*1.0*	*0.8*
SM_6h_gap	2.2	**0.5**	**0.5**	0.5	**0.3**	1.5	0.5	**0**	0.5	1.0	0.5	0.5	*0.8*	*0.7*
SM_3h_gap	2.8	1.0	0.7	0.5	0.5	**0.5**	0.5	**0**	0.5	1.0	0.8	**0**	*1.0*	*0.5*
SM_6h_int	2.2	**0.5**	**0.5**	0.5	0.7	2.5	0.5	**0**	0.3	1.5	0.5	1.0	*0.8*	*1.0*
SM_3h_int	2.5	**0.5**	0.7	0.5	0.8	2.0	0.5	**0**	0.3	2.0	0.5	1.0	*0.9*	*1.0*
SM_5min_int	2.5	**0.5**	1.3	**0**	0.7	2.0	0.5	**0**	0.5	1.0	0.5	1.0	*1.0*	*0.8*
DHM_6h	3.0	1.0	1.7	**0**	0.7	2.0	0.5	**0**	0.5	1.0	0.8	**0**	*1.2*	*0.7*
DHM_3h	3.0	1.0	1.7	**0**	0.7	2.0	0.5	**0**	0.5	1.0	0.7	**0**	*1.2*	*0.7*
REFIR_3h	**1.7**	1.0	**0.5**	1.0	1.0	3.0	**0.3**	1.0	0.5	2.0	**0.3**	1.0	***0.7***	*1.5*
REFIR_15min	**1.7**	**0.5**	**0.5**	1.0	0.8	2.5	**0.3**	1.0	**0.2**	2.0	0.5	1.0	***0.7***	*1.3*
average	*2.7*	*0.7*	*1.2*	*0.3*	*0.7*	*1.8*	*0.5*	*0.2*	*0.5*	*1.2*	*0.8*	*0.5*	*1.1*	*0.8*

Table 6:

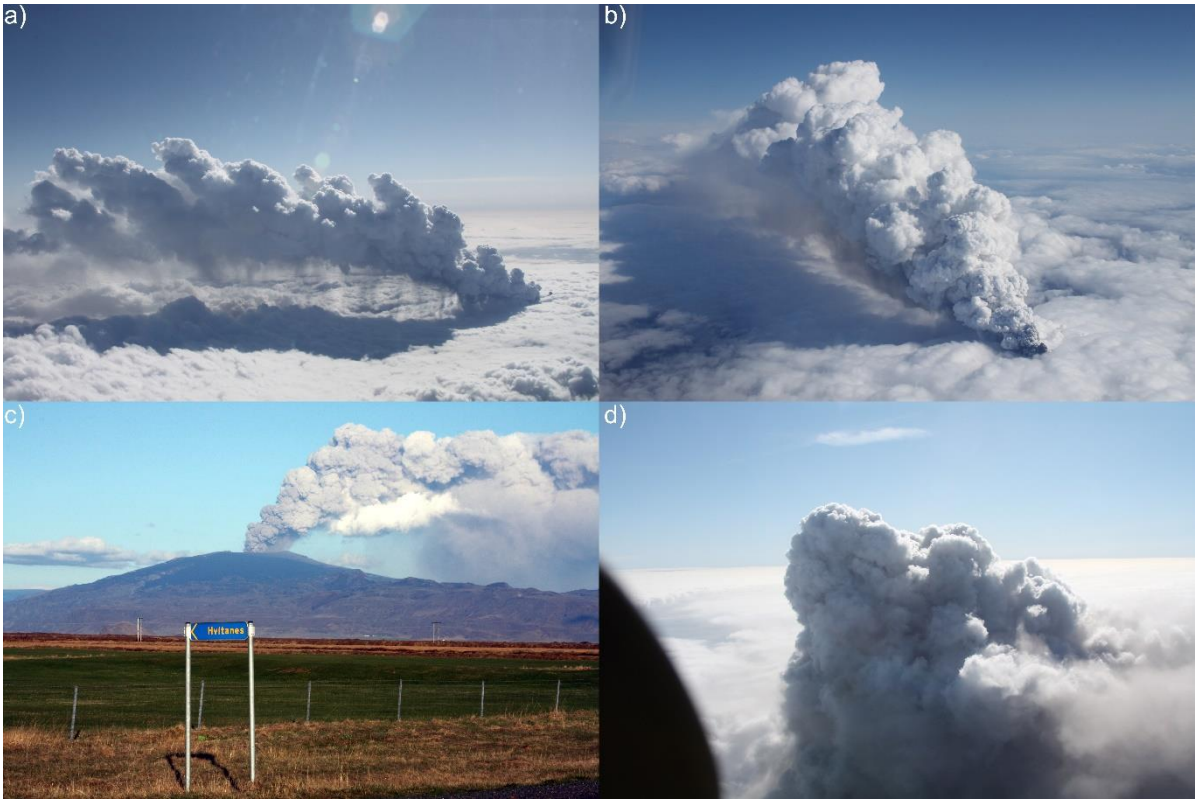
Average discongruence indices over all strategies, listed by reference period and plume model.

Row	Wilson	Wehrmann	Scollo	Andronico	Sparks	Mastin	Woodhouse	Aubry	**average**
Ia	1.55	4.90	2.20	1.85	1.85	3.55	0.95	0.40	**2.16**
Ib	0.40	3.10	0.70	0.70	0.70	1.75	0.10	0.40	**0.98**
II	1.05	0.80	0.70	0.70	0.80	0.35	1.10	2.45	**0.99**
III	0.10	2.00	0.10	0.05	0.05	0.95	0.10	0.25	**0.45**
IV	0.10	1.85	0.10	0.10	0.10	1.00	1.10	1.25	**0.70**
all	0.05	2.55	0.45	0.20	0.15	1.30	0.55	0.50	**0.72**
average	**0.54**	**2.53**	**0.71**	**0.60**	**0.61**	**1.48**	**0.65**	**0.88**	

Figures

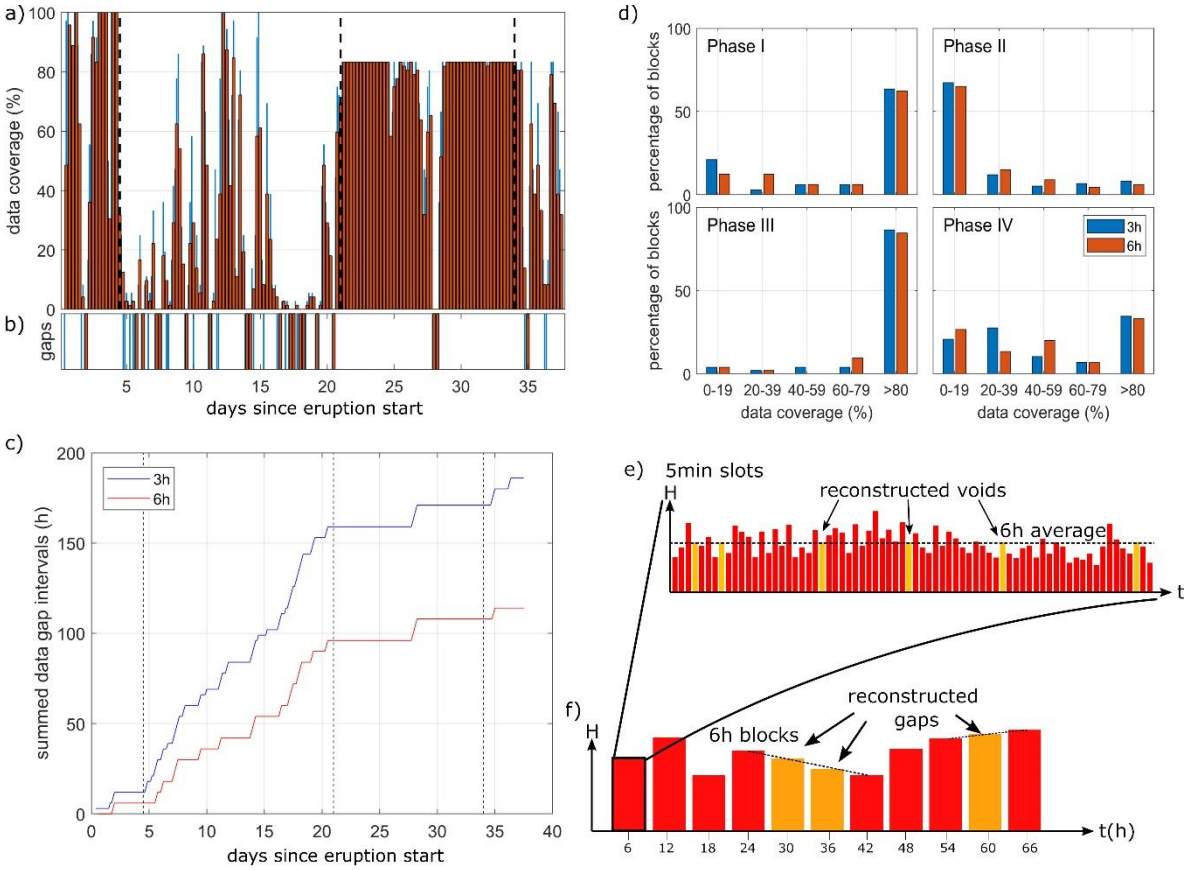
Figure 1

Examples of photos that were used to validate the plume height records from radar. a) Aerial picture of the Eyjafjallajökull plume, taken on 14 April 2010 at 10:20 UTC. b) This photo was taken 5 minutes later from a different angle. Since the GPS coordinates of the plane were logged, the horizontal location of the plume top could be constrained by triangulation. At that time, the plume top reached an altitude of (6.0 ± 0.3) km above vent. c) Example of a photo taken on the ground, from a distance of 29 km from the vent. This picture dates from 10 May (19:50 UTC), when the plume was deflected by 1.7 km towards 193° (SSW), with its top reaching an elevation of (2.2± 0.2) km above vent. d) This photo was taken on the last day of the eruption (21 May) at 18:00 UTC. In this case the elevation of the plume was directly assessed by flying at the same elevation as the plume top and using the on-board altimeter. The aircraft flew at 12,000ft a.s.l., which translates into a plume height of (2.3 ± 0.1) km. Photo credits: Th. Högnadóttir (a, b); M.T. Gudmundsson (c, d).



809 Figure 2

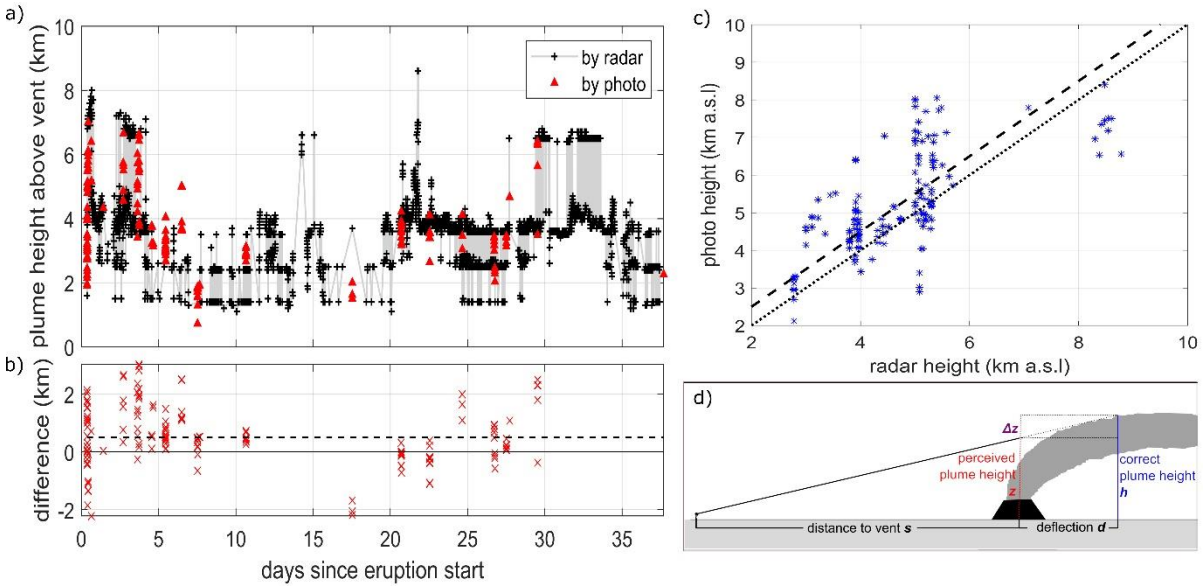
810 The radar record of the 2010 Eyjafjallajökull eruption. Data from Arason et al. (2011). a) Data
811 coverage of 3h- (blue) and 6h -(red) blocks. For example, a 6h-block with 50% data coverage
812 represents a set of 72 5-minute slots, of which 36 contain data. b) Timing of data gaps (i.e. blocks
813 that are completely empty). c) Increasing the time step (i.e., block width), leads to a reduction of
814 total overall gap time. d) Histograms show that data coverage of the blocks differed between the
815 eruptive phases. e) Voids in the plume height records are reconstructed by filling them with the
816 block's mean value (orange bars). f) Reconstructed gaps (orange bars) are obtained by interpolation.



817

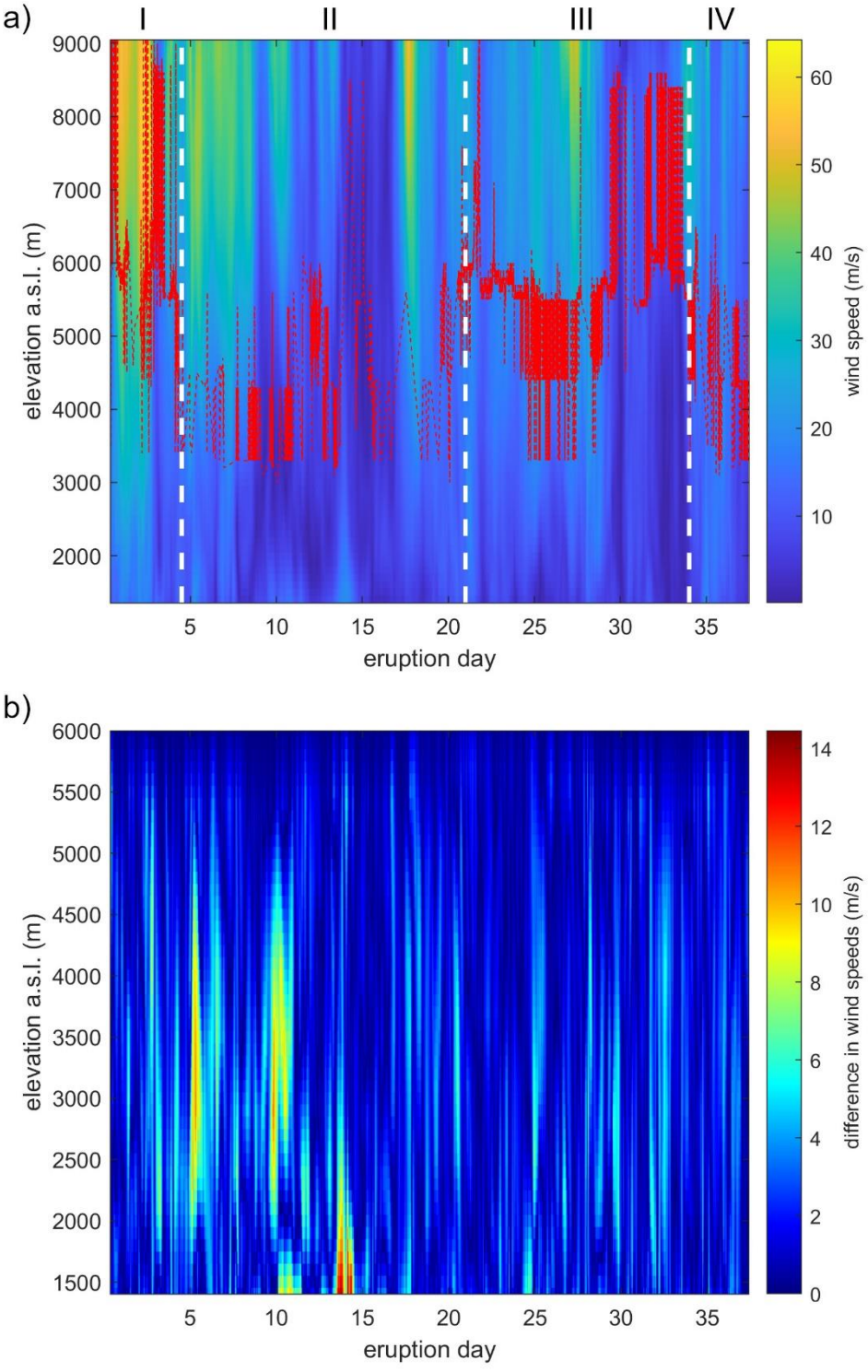
Figure 3

Plume height validation by means of photographic plume analysis. a) Radar-based plume height records of the Eyjafjallajökull 2010 eruption were compared with elevation data retrieved from photos. b) Photo plume heights are systematically shifted upward by 0.5 km relative to radar plume heights. c) This means that, statistically, the radar signal underestimated the actual plume height by 0.5 km. d) Due to windy conditions, the ash plume showed at times a considerable deflection *d*, which would lead to an underestimation of the plume if the distance to the plume top is assumed to equal the distance to the vent *s*.



828 Figure 4

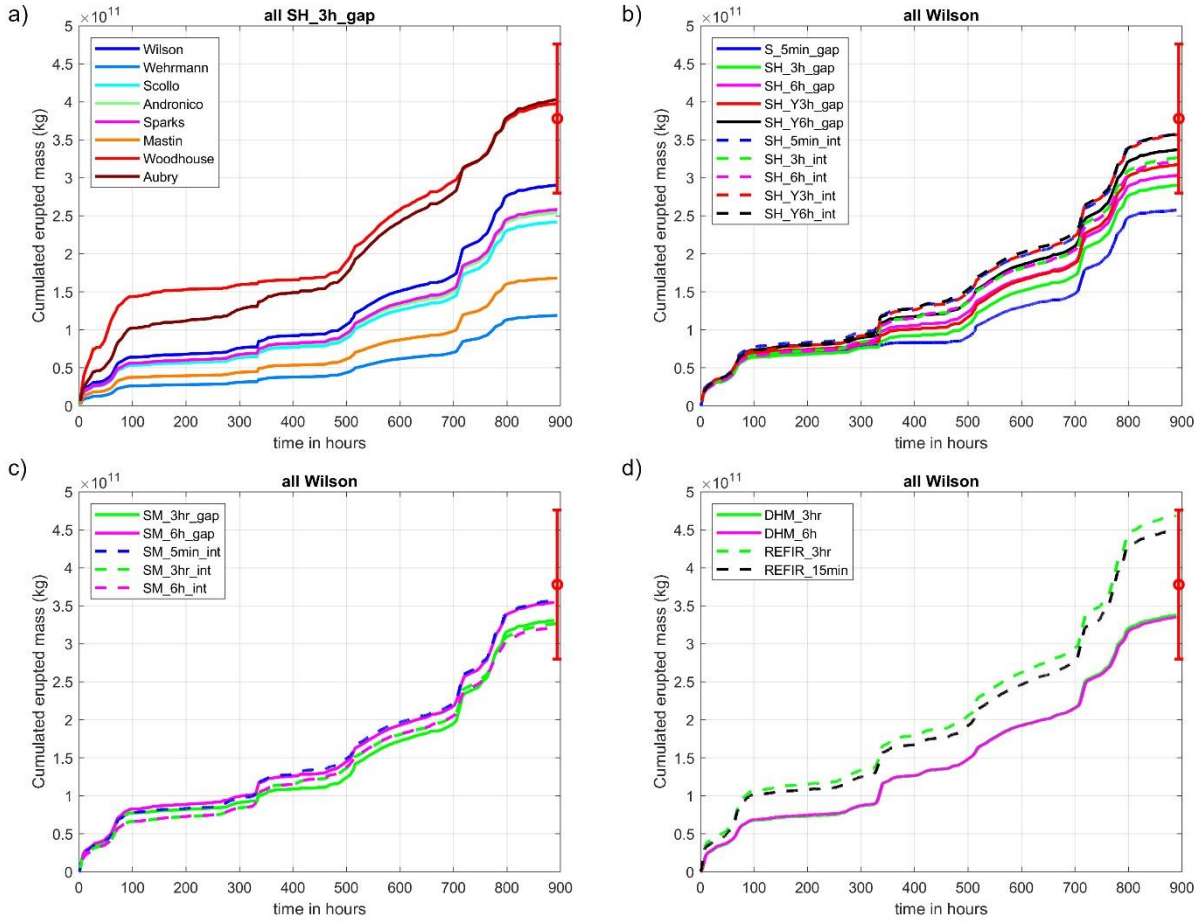
829 Wind speeds above Eyjafjallajökull during the 2010 eruption. a) Plume height (red curve) is plotted
830 together with altitude-specific wind speeds above the vent, based on ERA5 reanalysis data. Altitude
831 refers to sea level. Vertical dashed white lines mark margins of eruptive phases. b) Overall, the
832 absolute differences in wind speeds between ICRA and ERA5 were relatively small, with only sporadic
833 disparities, mainly at ground level on day 14.



834

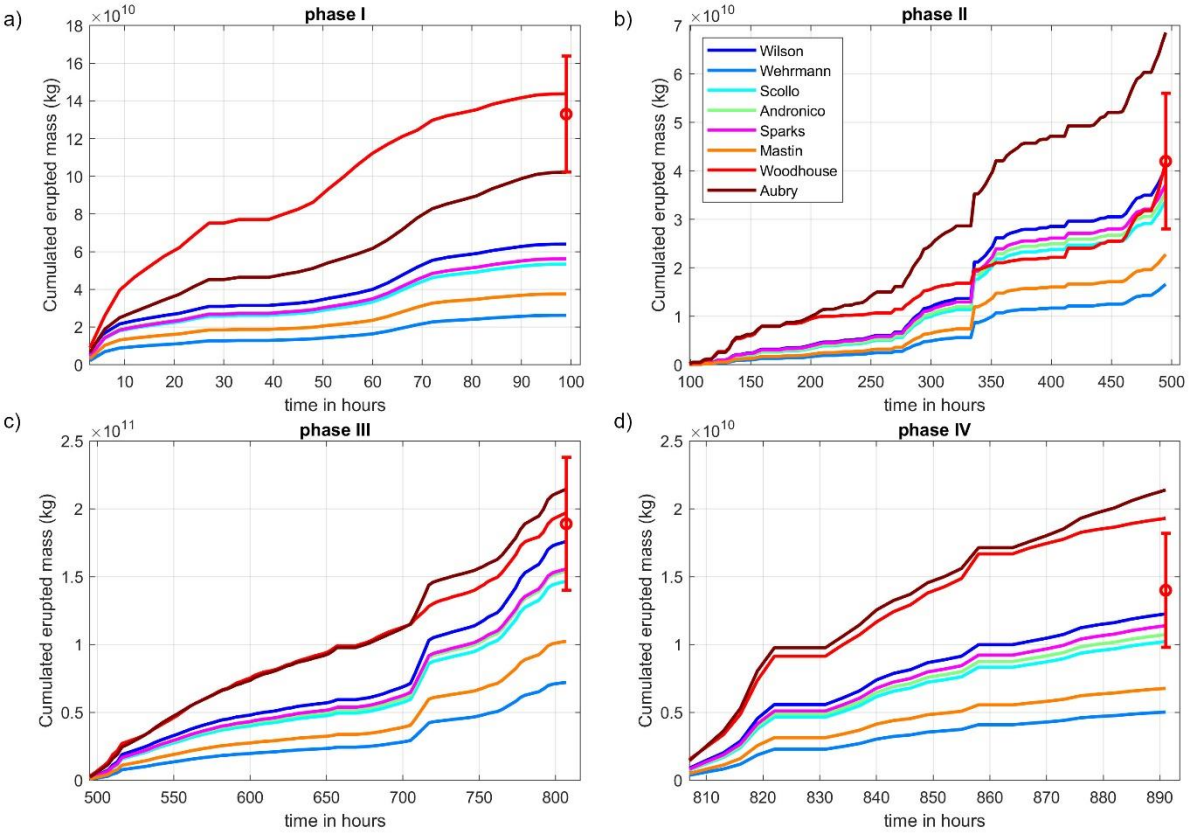
Figure 5

Modelled erupted mass. Erupted mass is computed by multiplying the modelled mass eruption rates by the block duration and summing the results over the reference period (here: the entire eruption). The range of measured erupted mass as reported by Gudmundsson et al. (2012) (see also Table 1) is displayed as red error bar with width *ME*. a) For each of the 8 tested models, 6h-averaged plume heights were used as input with void but no gap reconstruction, thus applying strategy *SH_3h_gap* (see Table 2). Wind speeds from ERA5 were used for *Woodhouse* and *Aubry*. b) Results for *S_5min_gap* and *SH* plume height strategies applied to the *Wilson* model. c) Results for *Wilson* with *SM* strategies. c) Predicted mass by *Wilson* with *DHM* and two *REFIR* strategies.



847 Figure 6

848 Modelled curves of erupted mass for each eruptive phase, applying strategy *SH_3h_gap* for eight
849 models. The range of measured erupted mass (ground truth) is indicated with a red error bar.
850 Reference periods are: a) phase I, b) phase II, c) phase III, d) phase IV.



851

852

Figure 7

Phase-specific comparisons of modelled versus measured mass. Results for six strategies (marked by different symbols) are presented for the whole eruption (*all*) and individual phases. The plots show the model-specific outcomes for: a) *Wilson*, b) *Mastin*, c) *Woodhouse* d) *Aubry* (with using wind speeds from ERA5 for both wind-affected models). The green lines indicate the range of erupted mass, interpolated according to ground measurements. A symbol lying within this corridor indicates that the respective strategy and plume model provides a 'successful' prediction under the tested eruptive conditions.

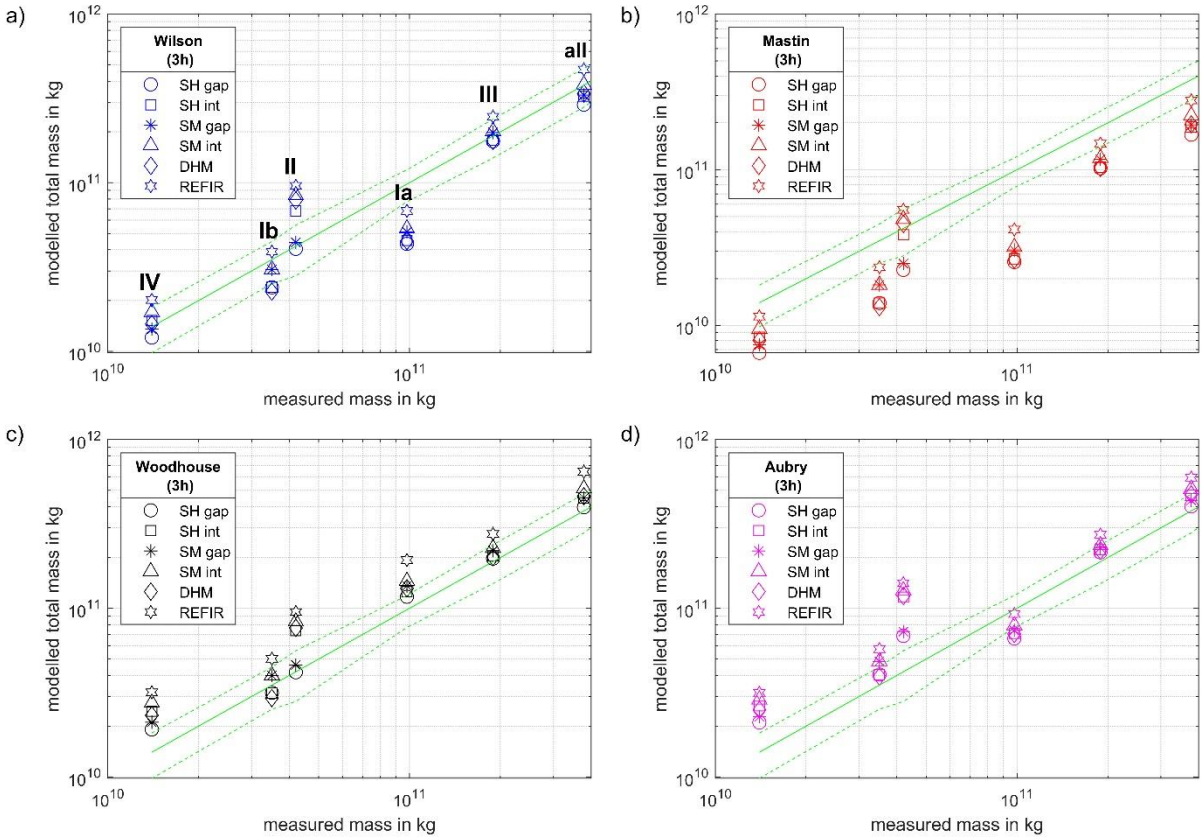
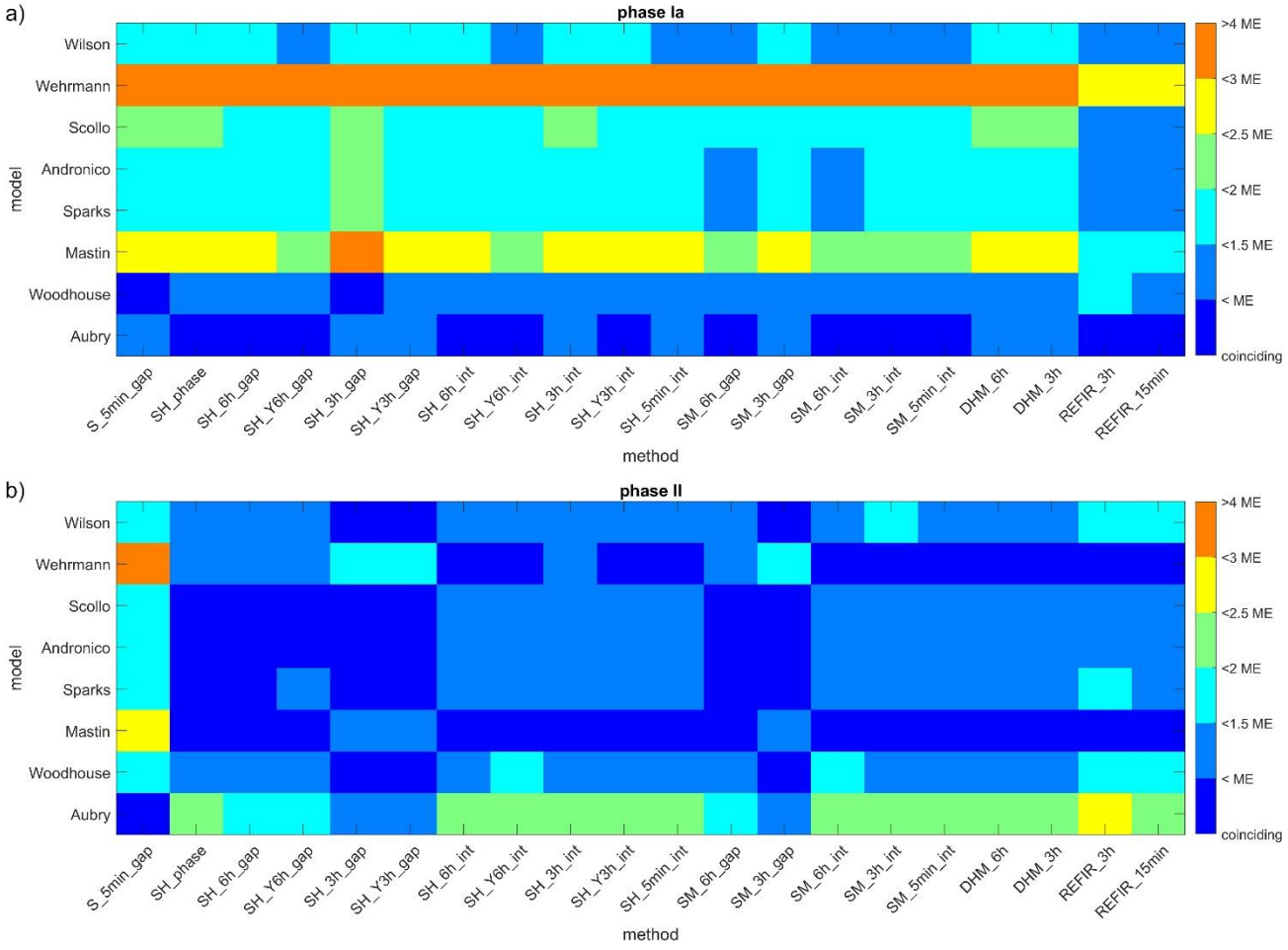


Figure 8

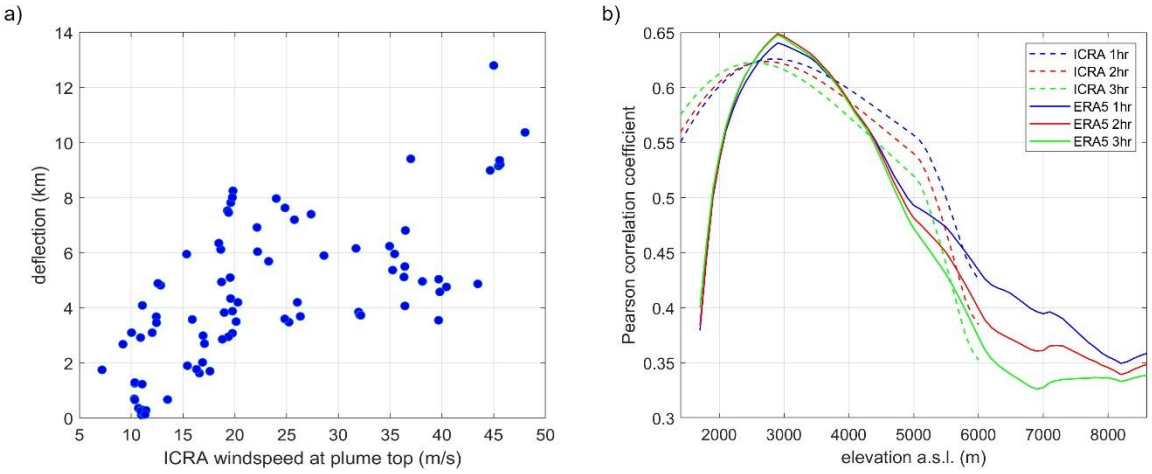
Congruency maps for combinations of strategies and models. ERA5 wind speeds were used for *Woodhouse* and *Aubry*. Each cell's colour indicates how well the estimate of a strategy and model pair fits with the measured data. Dark blue tiles indicate predictions that coincide with measurements. a) Congruency map of all strategies and models for phase Ia, which was characterized by the highest wind speeds. b) Congruency map for phase II, the phase with the most gaps in the radar records.



872 Figure 9

873 Wind-induced deflection of plume top. a) Photo-derived deflection of plume top is plotted versus
874 wind speeds at maximum plume height, according to ICRA reanalysis data. The data plots with a
875 Pearson correlation coefficient of 0.68. b) Correlation between deflection of the plume top and wind
876 speeds at different altitude levels (elevation above sea level).

877

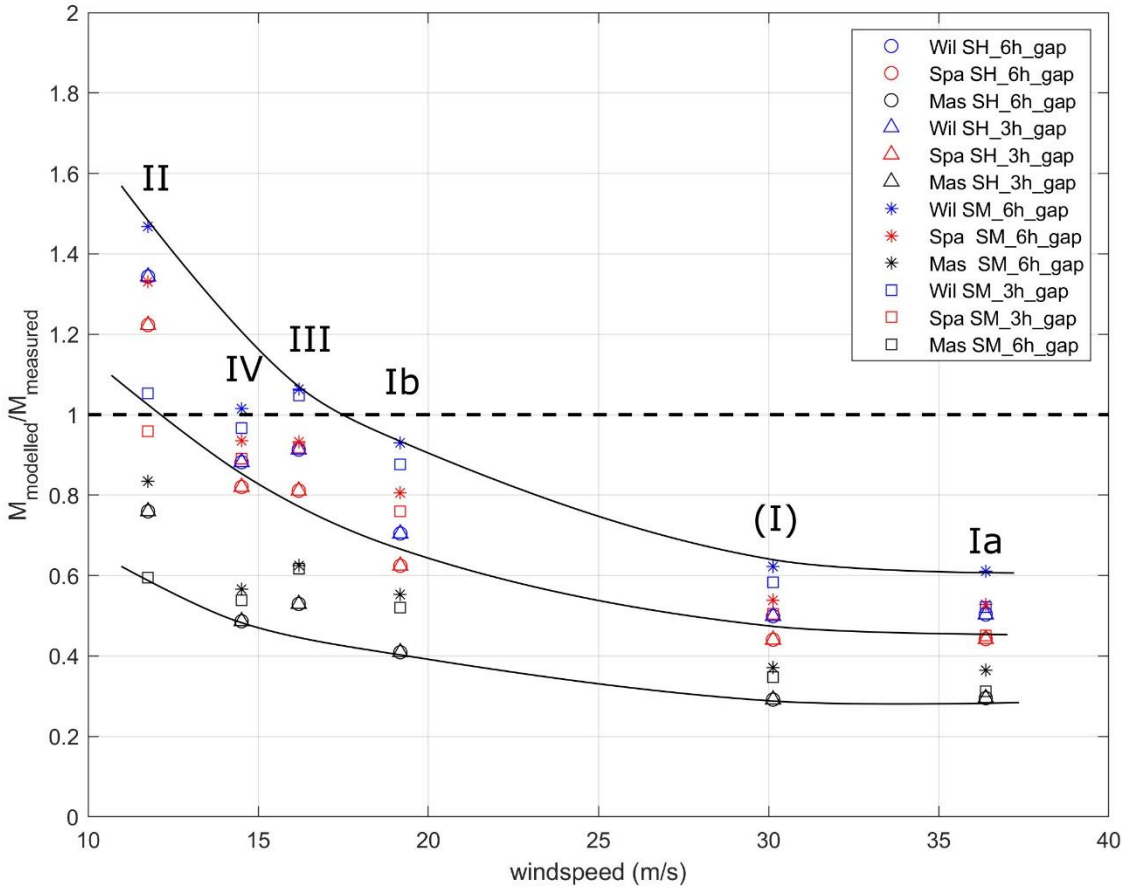


878

879

880 Figure 10

881 Wind-effect on simple (i.e., non-explicitly wind-affected) model predictions. Using four data handling
882 strategies for three models (*Wil: Wilson, Spa: Sparks, Mas: Mastin*), the ratio of modelled vs
883 measured mass is plotted over phase-averaged ERA5 wind speeds. The black curves serve for
884 visualization of what appears to be a general trend: the stronger the wind, the more the simple
885 plume models tend to underestimate the mass eruption rate, regardless of the applied strategy.



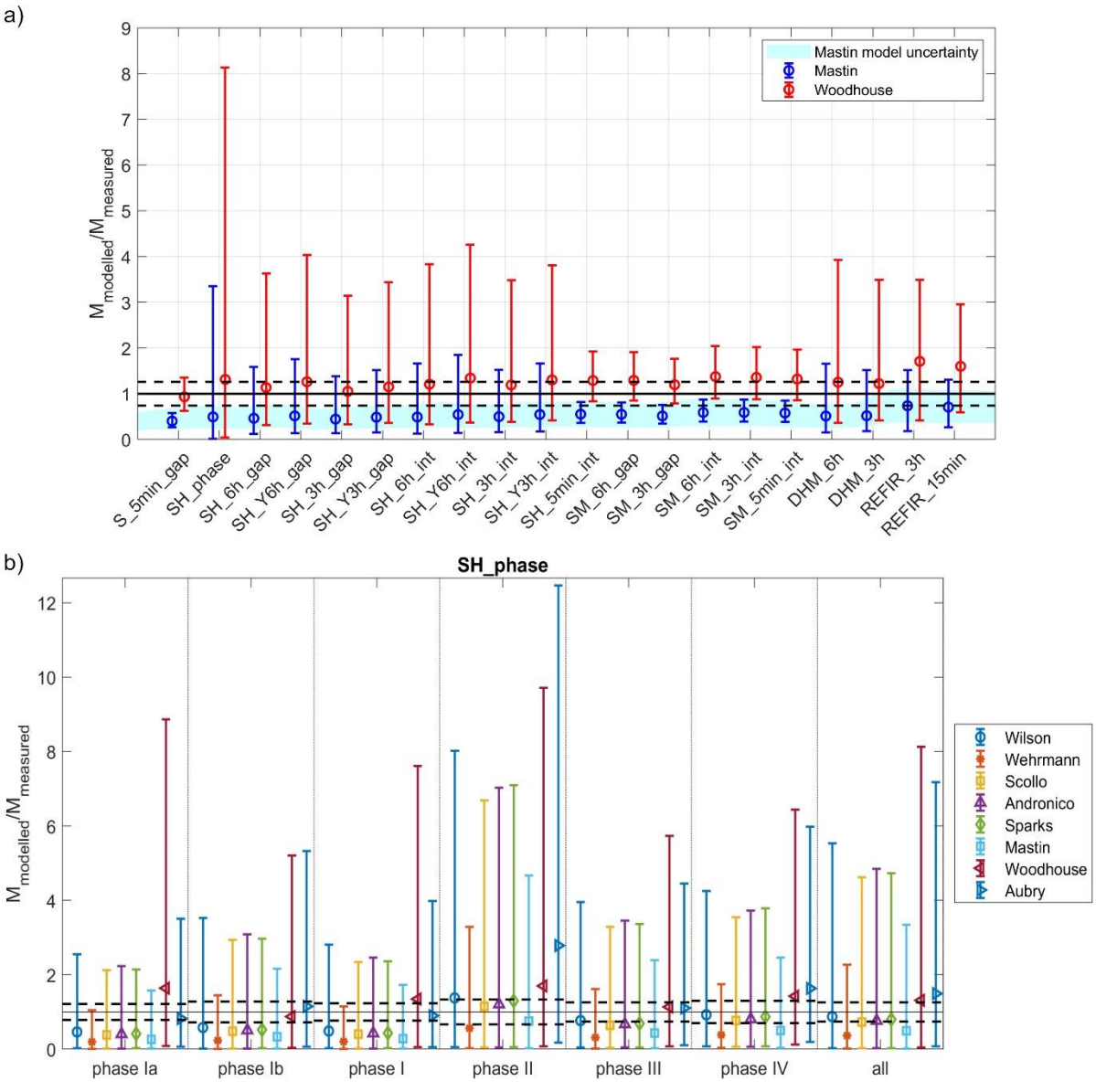
886

887

888 Figure 11

889 Ranges of uncertainty for modelled masses. a) Strategy-specific uncertainties for *Mastin* and
890 *Woodhouse*, when modelling the complete eruption period (*all*). For most strategies, the effect of
891 plume height errors according to equation (19) is larger than the model uncertainties (blue-shaded
892 background for *Mastin*). b) Ranges of predicted mass uncertainties for strategy *SH_phase* for all
893 studied eruptive phases and the complete eruption (*all*). Although depending on model and phase,
894 for this strategy the error bars are always considerably larger than the range of measured mass, due
895 to the large timebase.

896



897

Acknowledgements

We would like to thank Larry Mastin and two anonymous referees as well as handling editor Julia Eychenne and executive editor Andrew Harris for their constructive comments which helped to improve our manuscript. Rögnvaldur L. Magnússon did the geo-referencing of the photos, which were mostly taken on board the aircraft of the Icelandic Coast Guard or those of the private operator Eagle Air. Bolli Palmason from Icelandic Met Office is gratefully acknowledged for providing ICRA data. The geo-referencing and photo analysis was conducted under the EU Framework 7 FutureVolc project (2012-2016). This work contributes to project MAXI-Plume, supported by the Icelandic Research Fund (Rannís), grant Nr. 206527-051.

Online Resources

Online Resource 1 (.tif):

Corrected radar heights and ERA-5 wind speeds for different eruptive phases. a) phase I, b) phase II, c) phase III, d) phase IV.

Online Resource 2 (.tif):

Phase-specific results for five plume height strategies applied to the Wilson model. The range of measured erupted mass ("ground truth") is indicated with a red error bar. a) phase I, b) phase II, c) phase III, d) phase IV.

Online Resource 3 (.tif):

Results for all models, using the REFIR strategies (*REFIR_15min*, *REFIR_30min*, *REFIR_1h* and *REFIR_3h*), and the whole eruption (i.e, all phases) as reference period. *Woodhouse* and *Aubry* were used with wind speeds from ERA5. Dashed horizontal lines indicate the range of measured erupted mass.

Online Resource 4 (.tif):

Phase-specific results for all models, using the REFIR strategies (*REFIR_15min*, *REFIR_30min*, *REFIR_1h* and *REFIR_3h*). Dashed horizontal lines indicate the range of measured erupted mass. a) phase I, b) phase II, c) phase III, d) phase IV.

Online Resource 5 (.xlsx):

Modelled masses (in kg). Sheets are named after the plume height strategy applied. Columns show eruptive phases. Rows show model results. Next to the best estimates, the range of uncertainty is specified by providing the minimum (min) and maximum values (max).

Online Resource 6 (.jpg):

Congruency map for complete eruption (all).

Online Resource 7 (.jpg):

Phase-specific congruency maps for all strategies tested.

Online Resource 8 (.tif):

Phase-specific MER predictions for *Woodhouse* with *REFIR_3h*, using wind speeds of ERA 5 and ICRA. In addition to 1h-timed wind speeds, 3h-averaged wind speeds were tested.

Declarations

Funding

TD was supported by the IRF (Rannís) PostDoctoral project grant 206527-051

Conflicts of interest/Competing interests

There are no competing interests

Availability of data and material

All data will be provided upon request to the lead author

Code availability

Not applicable

References

Andronico D, Scollo S, Caruso S, Cristaldi A (2008) The 2002–03 Etna explosive activity: Tephra dispersal and features of the deposits. J Geophys Res 113:B04209. https://doi.org/10.1029/2007JB005126

Arason P, Petersen GN, Bjornsson H (2011) Observations of the altitude of the volcanic plume during the eruption of Eyjafjallajökull, April–May 2010. Earth Syst Sci Data 3:9–17. https://doi.org/10.5194/essd-3-9-2011

Aubry TJ, Engwell S, Bonadonna C, et al (2021) The Independent Volcanic Eruption Source Parameter Archive (IVESPA, version 1.0): A new observational database to support explosive eruptive column model validation and development. J Volcanol Geotherm Res 417:107295. https://doi.org/10.1016/j.jvolgeores.2021.107295

Aubry TJ, Jellinek AM, Carazzo G, et al (2017) A new analytical scaling for turbulent wind-bent plumes: Comparison of scaling laws with analog experiments and a new database of eruptive conditions for predicting the height of volcanic plumes. J Volcanol Geotherm Res 343:233–251. https://doi.org/10.1016/j.jvolgeores.2017.07.006

Bear-Crozier A, Pouget S, Bursik M, et al (2020) Automated detection and measurement of volcanic cloud growth: towards a robust estimate of mass flux,

966 mass loading and eruption duration. Nat Hazards 101:1–38.
967 https://doi.org/10.1007/s11069-019-03847-2

968 Biass S, Bagheri G, Aeberhard W, Bonadonna C (2013) TError: towards a better
969 quantification of the uncertainty propagated during the characterization of tephra
970 deposits. Stat Volcanol 1:1–27. https://doi.org/10.5038/2163-338X.1.2

971 Björnsson H, Magnusson S, Arason P, Petersen GN (2013) Velocities in the plume of
972 the 2010 Eyjafjallajökull eruption. J Geophys Res Atmos 118:11,611-698,711.
973 https://doi.org/10.1002/jgrd.50876

974 Bonadonna C, Cioni R, Costa A, et al (2016) MeMoVolc report on classification and
975 dynamics of volcanic explosive eruptions. Bull Volcanol 78:84.
976 https://doi.org/10.1007/s00445-016-1071-y

977 Brannigan V (2011) Paradigms Lost: Emergency Safety Regulation under Scientific
978 and Technical Uncertainty. In: Alemanno A (ed) Governing Disasters: The
979 Challenges of Emergency Risk Regulation. Edward Elgar Publishing,
980 Cheltenham, pp 101–114

981 Bursik M (2001) Effect of wind on the rise height of volcanic plumes. Geophys Res
982 Lett 28:3621–3624. https://doi.org/10.1029/2001GL013393

983 Bursik M, Jones M, Carn S, et al (2012) Estimation and propagation of volcanic
984 source parameter uncertainty in an ash transport and dispersal model:
985 application to the Eyjafjallajokull plume of 14–16 April 2010. Bull Volcanol
986 74:2321–2338. https://doi.org/10.1007/s00445-012-0665-2

987 Büttner R, Zimanowski B (1998) Physics of thermohydraulic explosions. Phys Rev E
988 57:5726–5729. https://doi.org/10.1103/PhysRevE.57.5726

989 Büttner R, Zimanowski B, Röder H (2000) Short-time electrical effects during volcanic
990 eruption: Experiments and field measurements. J Geophys Res Solid Earth
991 105:2819–2827. https://doi.org/10.1029/1999JB900370

992 Calvari S, Büttner R, Cristaldi A, et al (2012) The 7 September 2008 Vulcanian
993 explosion at Stromboli volcano: Multiparametric characterization of the event and
994 quantification of the ejecta. J Geophys Res Solid Earth 117:B05201.

995 https://doi.org/10.1029/2011JB009048

996 Cerminara M, Esposti Ongaro T, Berselli LC (2016) ASHEE-1.0: a compressible,
997 equilibrium–Eulerian model for volcanic ash plumes. Geosci Model
998 Dev 9:697–730. https://doi.org/10.5194/gmd-9-697-2016

999 Cerminara M, Esposti Ongaro T, Valade S, Harris AJL (2015) Volcanic plume vent
1000 conditions retrieved from infrared images: A forward and inverse modeling
1001 approach. J Volcanol Geotherm Res 300:129–147.
1002 https://doi.org/10.1016/j.jvolgeores.2014.12.015

1003 Costa A, Suzuki YJ, Cerminara M, et al (2016) Results of the eruptive column model
1004 inter-comparison study. J Volcanol Geotherm Res 326:2–25.
1005 https://doi.org/10.1016/j.jvolgeores.2016.01.017

1006 Dacre HF, Grant ALM, Hogan RJ, et al (2011) Evaluating the structure and
1007 magnitude of the ash plume during the initial phase of the 2010 Eyjafjallajökull
1008 eruption using lidar observations and NAME simulations. J Geophys Res
1009 116:D00U03. https://doi.org/10.1029/2011JD015608

1010 de'Michieli Vitturi M, Neri A, Barsotti S (2015) PLUME-MoM 1.0: A new integral
1011 model of volcanic plumes based on the method of moments. Geosci Model Dev
1012 8:2447–2463

1013 Degruyter W, Bonadonna C (2012) Improving on mass flow rate estimates of volcanic
1014 eruptions. Geophys Res Lett 39:1–6. https://doi.org/10.1029/2012GL052566

1015 Dellino P, Gudmundsson MT, Larsen G, et al (2012) Ash from the Eyjafjallajökull
1016 eruption (Iceland): Fragmentation processes and aerodynamic behavior. J
1017 Geophys Res Solid Earth 117:B00C04. https://doi.org/10.1029/2011JB008726

1018 Devenish BJ (2013) Using simple plume models to refine the source mass flux of
1019 volcanic eruptions according to atmospheric conditions. J Volcanol Geotherm
1020 Res 256:118–127. https://doi.org/10.1016/j.jvolgeores.2013.02.015

1021 Dioguardi F, Beckett F, Dürig T, Stevenson JA (2020) The Impact of Eruption Source
1022 Parameter Uncertainties on Ash Dispersion Forecasts During Explosive Volcanic
1023 Eruptions. J Geophys Res Atmos 125:e2020JD032717.

https://doi.org/10.1029/2020JD032717

Dioguardi F, Dürig T, Engwell SL, et al (2016) Investigating Source Conditions and Controlling Parameters of Explosive Eruptions: Some Experimental-Observational- Modelling Case Studies. In: Németh K (ed) Updates in Volcanology - From Volcano Modelling to Volcano Geology. InTech, Rijeka

Dürig T, Gudmundsson MT, Dellino P (2015a) Reconstruction of the geometry of volcanic vents by trajectory tracking of fast ejecta - the case of the Eyjafjallajökull 2010 eruption (Iceland). Earth, Planets Sp 67:64. https://doi.org/10.1186/s40623-015-0243-x

Dürig T, Gudmundsson MT, Dioguardi F, et al (2018) REFIR- A multi-parameter system for near real-time estimates of plume-height and mass eruption rate during explosive eruptions. J Volcanol Geotherm Res 360:61–83. https://doi.org/10.1016/j.jvolgeores.2018.07.003

Dürig T, Gudmundsson MT, Karmann S, et al (2015b) Mass eruption rates in pulsating eruptions estimated from video analysis of the gas thrust-buoyancy transition—a case study of the 2010 eruption of Eyjafjallajökull, Iceland. Earth, Planets Sp 67:180. https://doi.org/10.1186/s40623-015-0351-7

Dürig T, White JDL, Murch AP, et al (2020) Deep-sea eruptions boosted by induced fuel–coolant explosions. Nat Geosci 13:498–503. https://doi.org/10.1038/s41561-020-0603-4

Engwell S, Sparks R, Aspinall W (2013) Quantifying uncertainties in the measurement of tephra fall thickness. J Appl Volcanol 2:5. https://doi.org/10.1186/2191-5040-2-5

Esposti Ongaro T, Cavazzoni C, Erbacci G, et al (2007) A parallel multiphase flow code for the 3D simulation of explosive volcanic eruptions. Parallel Comput 33:541–560. https://doi.org/10.1016/j.parco.2007.04.003

Fierstein J, Nathenson M (1992) Another look at the calculation of fallout tephra volumes. Bull Volcanol 54:156–167. https://doi.org/10.1007/BF00278005

Folch A, Costa A, Macedonio G (2016) FPLUME-1.0: An integral volcanic plume

model accounting for ash aggregation. Geosci Model Dev 9:431–450. https://doi.org/10.5194/gmd-9-431-2016

Gouhier M, Eychenne J, Azzaoui N, et al (2019) Low efficiency of large volcanic eruptions in transporting very fine ash into the atmosphere. Sci Rep 9:1449. https://doi.org/10.1038/s41598-019-38595-7

Grindle TJ, Burcham FW (2002) Even minor volcanic ash encounters can cause major damage to aircraft. ICAO J 57:12–14

Gudmundsson MT, Högnadóttir T, Dürig T, et al (2015) Field laboratory, aircraft observations and radar - Futurevolc Report D7.2. Reykjavík

Gudmundsson MT, Thordarson T, Hoskuldsson A, et al (2012) Ash generation and distribution from the April-May 2010 eruption of Eyjafjallajökull, Iceland. Sci Rep 2:1–12. https://doi.org/10.1038/srep00572

Guffanti M, Casadevall TJ, Budding KE (2010) Encounters of aircraft with volcanic ash clouds: A compilation of known incidents, 1953-2009

Harris A (2013) Thermal Remote Sensing of Active Volcanoes: A User's Manual. Cambridge University Press, New York

Harris AJL (2015) Forecast communication through the newspaper Part 2: perceptions of uncertainty. Bull Volcanol 77:30. https://doi.org/10.1007/s00445-015-0902-6

Harris AJL, Delle Donne D, Dehn J, et al (2013) Volcanic plume and bomb field masses from thermal infrared camera imagery. Earth Planet Sci Lett 365:77–85. https://doi.org/10.1016/j.epsl.2013.01.004

Harris AJL, Gurioli L, Hughes EE, Lagreulet S (2012) Impact of the Eyjafjallajökull ash cloud: A newspaper perspective. J Geophys Res Solid Earth 117:. https://doi.org/10.1029/2011JB008735

Hersbach H, Bell B, Berrisford P, et al (2020) The ERA5 global reanalysis. Q J R Meteorol Soc 146:1999–2049. https://doi.org/10.1002/qj.3803

Johnson JB, Ripepe M (2011) Volcano infrasound: A review. J Volcanol Geotherm

1081 Res 206:61–69. https://doi.org/10.1016/j.jvolgeores.2011.06.006

1082 Kaminski E, Tait S, Ferrucci F, et al (2011) Estimation of ash injection in the
1083 atmosphere by basaltic volcanic plumes: The case of the Eyjafjallajökull 2010
1084 eruption. J Geophys Res 116:. https://doi.org/10.1029/2011JB008297

1085 Keiding JK, Sigmarsson O (2012) Geothermobarometry of the 2010 Eyjafjallajökull
1086 eruption: New constraints on Icelandic magma plumbing systems. J Geophys
1087 Res Solid Earth. https://doi.org/10.1029/2011JB008829

1088 Kienle J, Kyle PR, Self S, et al (1980) Ukinrek Maars, Alaska, I. April 1977 eruption
1089 sequence, petrology and tectonic setting. J Volcanol Geotherm Res 7:11–37.
1090 https://doi.org/10.1016/0377-0273(80)90018-9

1091 Koyaguchi T, Woods AW (1996) On the formation of eruption columns following
1092 explosive mixing of magma and surface-water. J Geophys Res B Solid Earth.
1093 https://doi.org/10.1029/95jb01687

1094 Kristiansen NI, Stohl A, Prata AJ, et al (2012) Performance assessment of a volcanic
1095 ash transport model mini-ensemble used for inverse modeling of the 2010
1096 Eyjafjallajökull eruption. J Geophys Res Atmos 117:D00U11.
1097 https://doi.org/10.1029/2011JD016844

1098 Macrae D (2011) Which Risk and Who Decides When There Are So Many Players?
1099 In: Alemanno A (ed) Governing Disasters: The Challenges of Emergency Risk
1100 Regulation. Edward Elgar Publishing, Cheltenham, pp 13–24

1101 Magnússon RL (2012) Pixelcalc: Forrit til mælinga á stærð gosmakka út frá
1102 stafrænum myndum (Pixelcalc: Software to measure height of eruption plumes
1103 from digital photos). University of Iceland

1104 Marzano FS, Mereu L, Scollo S, et al (2020) Tephra Mass Eruption Rate From
1105 Ground-Based X-Band and L-Band Microwave Radars During the November 23,
1106 2013, Etna Paroxysm. IEEE Trans Geosci Remote Sens 58:3314–3327.
1107 https://doi.org/10.1109/TGRS.2019.2953167

1108 Mastin LG (2014) Testing the accuracy of a 1-D volcanic plume model in estimating
1109 mass eruption rate. J Geophys Res Atmos 119:2474–2495.

https://doi.org/10.1002/2013JD020604

Mastin LG, Guffanti M, Servranckx R, et al (2009) A multidisciplinary effort to assign realistic source parameters to models of volcanic ash-cloud transport and dispersion during eruptions. J Volcanol Geotherm Res 186:10–21. https://doi.org/10.1016/j.jvolgeores.2009.01.008

Montopoli M (2016) Velocity profiles inside volcanic clouds from three-dimensional scanning microwave dual-polarization Doppler radars. J Geophys Res Atmos 121:7881–7900. https://doi.org/10.1002/2015JD023464

Morton BR, Taylor G, Turner JS (1956) Turbulent gravitational convection from maintained and instantaneous sources. Proc R Soc London Ser A Math Phys Sci 234:1–23. https://doi.org/10.1098/rspa.1956.0011

Nawri N, Pálmason B, Petersen NG, et al (2017) The ICRA atmospheric reanalysis project for Iceland, Report VÍ 2017-005, Icelandic Met Office, pp 37. Reykjavík

Neri A, Papale P, Macedonio G, et al (1998) The role of magma composition and water content in explosive eruptions: 2. Pyroclastic dispersion dynamics. J Volcanol Geotherm Res 87:95–115. https://doi.org/10.1016/S0377-0273(98)00102-4

Pavolonis MJ, Sieglaff J, Cintineo J (2018) Automated Detection of Explosive Volcanic Eruptions Using Satellite-Derived Cloud Vertical Growth Rates. Earth Sp Sci. https://doi.org/10.1029/2018EA000410

Pioli L, Harris AJL (2019) Real-time geophysical monitoring of particle size distribution during volcanic explosions at stromboli volcano (Italy). Front Earth Sci 7:1–13. https://doi.org/10.3389/feart.2019.00052

Pouget S, Bursik M, Webley P, et al (2013) Estimation of eruption source parameters from umbrella cloud or downwind plume growth rate. J Volcanol Geotherm Res 258:100–112. https://doi.org/10.1016/j.jvolgeores.2013.04.002

Pyle DM (1989) The thickness, volume and grainsize of tephra fall deposits. Bull Volcanol 51:1–15. https://doi.org/10.1007/BF01086757

1138 Ragona M, Hannstein F, Mazzocchi M (2011) The financial impact of the volcanic
1139 ash crisis on the European airline industry. In: Alemanno A (ed) Governing
1140 Disasters: The Challenges of Emergency Risk Regulation, Edward Elg.
1141 Cheltenham, pp 27–50

1142 Ripepe M, Bonadonna C, Folch A, et al (2013) Ash-plume dynamics and eruption
1143 source parameters by infrasound and thermal imagery: The 2010 Eyjafjallajökull
1144 eruption. Earth Planet Sci Lett 366:112–121.
1145 https://doi.org/10.1016/j.epsl.2013.02.005

1146 Schmidt L, Langen P, Aðalgeirsdóttir G, et al (2018) Sensitivity of Glacier Runoff to
1147 Winter Snow Thickness Investigated for Vatnajökull Ice Cap, Iceland, Using
1148 Numerical Models and Observations. Atmosphere (Basel) 9:450.
1149 https://doi.org/10.3390/atmos9110450

1150 Scollo S, Del Carlo P, Coltelli M (2007) Tephra fallout of 2001 Etna flank eruption:
1151 Analysis of the deposit and plume dispersion. J Volcanol Geotherm Res
1152 160:147–164. https://doi.org/10.1016/j.jvolgeores.2006.09.007

1153 Sparks RSJ, Bursik MI, Carey SN, et al (1997) Volcanic plumes. John Wiley & Sons,
1154 Chichester

1155 Suzuki YJ, Koyaguchi T (2012) 3-D numerical simulations of eruption column
1156 collapse: Effects of vent size on pressure-balanced jet/plumes. J Volcanol
1157 Geotherm Res 221–222:1–13. https://doi.org/10.1016/j.jvolgeores.2012.01.013

1158 Thorarinsson S (1954) The tephra-fall from Hekla on March 29th 1947. In: Einarsson
1159 T, Kjartansson G, Thorarinsson S (eds) The Eruption of Hekla 1947–1948.
1160 Societas Scientarium Islandica, Reykjavík, pp 1–68

1161 Thorkelsson B, Karlsdóttir S, Gylfason ÁG, et al (2012) The 2010 Eyjafjallajökull
1162 Eruption -Iceland Report to ICAO-June, pp 206. Reykjavík

1163 Tournigand P-Y, Fernández JJP, Taddeucci J, et al (2019) Time evolution of
1164 transient volcanic plumes: Insights from fractal analysis. J Volcanol Geotherm
1165 Res 371:59–71. https://doi.org/10.1016/j.jvolgeores.2018.12.007

1166 Valade SA, Harris AJL, Cerminara M (2014) Plume Ascent Tracker: Interactive

1167 Matlab software for analysis of ascending plumes in image data. Comput
1168 Geosci. https://doi.org/10.1016/j.cageo.2013.12.015

1169 Wehrmann H, Bonadonna C, Freundt A, et al (2006) Fontana Tephra: A basaltic
1170 Plinian eruption in Nicaragua. In: Rose WI, Bluth GJS, Carr MJ, et al. (eds)
1171 Volcanic Hazards in Central America. Geological Society of America, Boulder,
1172 Colorado, pp 209–223

1173 Wilson L, Self S (1980) Volcanic explosion clouds: Density, temperature, and particle
1174 content estimates from cloud motion. J Geophys Res 85:2567.
1175 https://doi.org/10.1029/JB085iB05p02567

1176 Wilson L, Walker GPL (1987) Explosive volcanic eruptions - VI. Ejecta dispersal in
1177 plinian eruptions: the control of eruption conditions and atmospheric properties.
1178 Geophys J Int 89:657–679. https://doi.org/10.1111/j.1365-246X.1987.tb05186.x

1179 Wohletz KH (1986) Explosive magma-water interactions: Thermodynamics, explosion
1180 mechanisms, and field studies. Bull Volcanol 48:245–264.
1181 https://doi.org/10.1007/BF01081754

1182 Woodhouse MJ, Hogg AJ, Phillips JC, Sparks RSJ (2013) Interaction between
1183 volcanic plumes and wind during the 2010 Eyjafjallajökull eruption, Iceland. J
1184 Geophys Res Solid Earth 118:92–109. https://doi.org/10.1029/2012JB009592

1185 Woods AW (1988) The fluid dynamics and thermodynamics of eruption columns. Bull
1186 Volcanol 50:169–193. https://doi.org/10.1007/BF01079681

1187