

# Refining an ensemble of volcanic ash forecasts using satellite retrievals: Raikoke 2019

Antonio Capponi<sup>1</sup>, Natalie Harvey<sup>2</sup>, Helen Dacre<sup>2</sup>, Keith Beven<sup>1</sup>, Cameron Saint<sup>3</sup>, Cathie Wells<sup>2</sup>, and Mike R. James<sup>1</sup>

<sup>1</sup> Lancaster Environment Centre, [Lancaster University](https://www.lancaster.ac.uk), Lancaster, LA1 4AQ, UK

<sup>2</sup> Department of Meteorology, University of Reading, Earley Gate, Reading, RG6 6ET, UK

<sup>3</sup> Met Office, Fitzroy Road, Exeter EX1 3PB, UK

Correspondence to: Antonio Capponi ([a.capponi2@lancaster.ac.uk](mailto:a.capponi2@lancaster.ac.uk))

**Abstract.** Volcanic ash advisories are produced by specialised forecasters who combine several sources of observational data and volcanic ash dispersion model outputs based on their subjective expertise. These advisories are used by the aviation industry to make decisions about where it is safe to fly. However, both observations and dispersion model simulations are subject to various sources of uncertainties that are not represented in operational forecasts. Quantification and communication of these uncertainties are fundamental for making more informed decisions. Here, we develop a data assimilation [method](#) which combines satellite retrievals and volcanic ash transport and dispersion model (VATDM) output, considering uncertainties in both data sources. The methodology is applied to a case study of the 2019 Raikoke eruption. To represent uncertainty in the VATDM output, 1000 simulations are performed by simultaneously perturbing the eruption source parameters, meteorology and internal model parameters (known as the prior ensemble). The ensemble members are filtered, based on their level of agreement with [the ash column loading, and their uncertainty, of the Himawari-8](#) satellite retrievals, to produce a [constrained](#) posterior ensemble. For the Raikoke eruption, filtering the ensemble skews the values of mass eruption rate towards the lower values within the wider parameters ranges initially used in the prior ensemble (mean reduces from 1 Tgh<sup>-1</sup> to 0.1 Tgh<sup>-1</sup>). Furthermore, including satellite observations from subsequent times increasingly constrains the posterior ensemble. These results suggest that the prior ensemble leads to an overestimate of both the magnitude and uncertainty in ash column loadings. Based on the prior ensemble, flight operations would have been severely disrupted over the Pacific Ocean. Using the constrained posterior ensemble, the regions where the risk is overestimated are reduced potentially resulting in fewer flight disruptions. The data assimilation methodology developed in this paper is easily generalisable to other short duration eruptions and to other VATDMs and retrievals of ash from other satellites.

## 1 Introduction

Volcanic ash in the atmosphere poses a hazard to aircraft (Casadevall, 1994). It is therefore important to accurately forecast the evolution of volcanic ash cloud in the atmosphere for the aviation industry. Forecasting the distribution of volcanic ash in the atmosphere is typically performed using a volcanic ash transport and dispersion model (VATDM). VATDMs solve [numerical representations of equations related to ash dispersal processes in the atmosphere](#), to evolve the system state (volcanic ash cloud) forward in time. However, such simulated volcanic ash distributions are subject to errors due to inaccurate parametrisations of physical processes, errors in the driving meteorological fields and errors in the volcanic eruption source parameters. Observations of volcanic ash

36 [concentrations, size distributions, and mass loadings](#) may be obtained from ground-based, aircraft or satellite-based  
37 instruments. These observations can be used to evaluate the accuracy of VATDM simulations (Harvey and Dacre  
38 2016, Dacre et al. 2016). Geostationary satellite measurements are of particular interest as they provide information  
39 at high temporal frequency and, thanks to the increasingly growing network of satellites, over large spatial extents.  
40 Ash retrievals from geostationary satellite data use an [inverse](#) model to transform observations of radiance into  
41 vertically integrated [volcanic ash column loadings](#). However, retrievals of volcanic ash column loading from satellite  
42 data are subject to [errors, including measurement, retrieval and forward model errors, and](#) interference with other  
43 atmospheric constituents (e.g., Krotkov et al. 1999, Francis et al. 2012). This [error](#) information is often disregarded  
44 and only the mean ash retrievals are used for verification purposes. Therefore, to improve estimates of the volcanic  
45 ash cloud in the atmosphere, VATDM simulations and observations can be combined to create an analysis. Combining  
46 satellite-based observations and VATDM simulations allows the modelled volcanic ash cloud to be continuously  
47 adjusted and thus improves the accuracy of volcanic ash forecasts (Fu et al. 2017).

48 The most straightforward combination of VATDM and satellite observations is data insertion, whereby satellite  
49 observations of volcanic ash column loading are used as initial conditions in a VATDM simulation (Wilkins et al.  
50 2015). More sophisticated combinations of VATDM simulations and satellite observations involve data assimilation  
51 techniques such as variational and sequential methods. In variational data assimilation a cost function is defined to  
52 quantify the difference between a VATDM simulation and a satellite observation of volcanic ash column loading,  
53 weighted by the VATDM and observation uncertainties. The cost function is typically minimised by adjusting one or  
54 more eruption source parameters (e.g., plume height, mass eruption rate, particle size distribution, ash density) to  
55 estimate their optimum value for simultaneously fitting the simulated column loadings to the satellite retrievals and  
56 for fitting to prior estimates of the eruption source parameters at a given time. In the volcanic ash literature this  
57 technique is often referred to as source inversion (Stohl et al. 2011, Kristiansen et al. 2012, Denlinger et al. 2012,  
58 Pelley et al. 2015). Variational data assimilation [typically](#) uses observations from a fixed time window, thus allowing  
59 time evolving eruption source parameters to be estimated, so is suitable for long duration volcanic eruptions which  
60 undergo several eruptive pulses. Alternatively, sequential data assimilation provides an estimation of the system state  
61 sequentially as it evolves forward in time using observations as they become available. Thus, sequential data  
62 assimilation is suitable for short duration single pulse volcanic eruptions (Chai et al. 2017, Zidikheri et al. 2017).

63 In most cases eruption source parameters (input parameters), physical processes (internal parameters) and the driving  
64 meteorology are uncertain so an ensemble of VATDM simulations can be formed by perturbing the input,  
65 meteorological and internal parameters. This [estimates](#) a probability density function (pdf) of simulated volcanic ash  
66 distributions. In this case, the data assimilation step involves conditioning the VATDM [prior](#) pdf based on a  
67 comparison with the observed volcanic ash cloud to create a [posterior](#) pdf at each time the data assimilation is  
68 performed. In volcanic ash forecasting, ensemble source inversion (Harvey et al. 2020) and ensemble sequential  
69 filtering methods, such as Ensemble Kalman Filters (EnKFs) (Fu et al. 2015, Pardini et al. 2020, Osoreo et al. 2020,  
70 [Mingari et al. 2022](#)) and Particle Filters (Wang et al. 2017, Zidikheri and Lucas 2021), have been employed. EnKFs  
71 were developed for non-linear systems and so are suitable for dispersion problems. However, they assume that the

72 parameters to be estimated have unbiased Gaussian prior pdfs, which may not be true. Conversely, no assumptions on  
73 the form of the prior pdf of simulator states are needed for Particle Filtering techniques, although this is at the cost of  
74 requiring more simulations.

75 Bayesian inference is used in particle filtering to constrain simulation parameters with observations. In this framework,  
76 the posterior pdf of the simulation parameters given the observed data is computed from a prior pdf and from the  
77 likelihood of the data given a choice of simulator parameters. Bayesian inference therefore relies on the ability to  
78 compute a formal likelihood function. For volcanic eruption source parameters their exact likelihood function is  
79 unknown or computationally intractable and so direct Bayesian analysis is therefore not possible. A technique known  
80 as Approximate Bayesian Computation uses simulations to bypass the need to evaluate a likelihood function.  
81 Approximate Bayesian Computation systematically explores the prior parameter space and compares the simulated  
82 and observed data sets using a distance metric. By accepting simulations for which this distance metric is smaller than  
83 a given threshold, the method provides an approximation to the Bayesian posterior pdf. One Approximate Bayesian  
84 Computation method frequently used in hydrology forecasting is known as Generalised Likelihood Uncertainty  
85 Estimation (GLUE) (Beven and Binley, 1992). The GLUE methodology is based on the concept of equifinality, which  
86 acknowledges that there exist many combinations of simulation input and internal parameters that provide equally  
87 good simulations of the observed system.

88 There are several steps in the GLUE methodology:

- 89 1. Realistic ranges are defined for the simulator input and internal parameters. These are known as prior pdfs  
90 since they are defined prior to the comparison with observational data. When there is a lack of strong prior  
91 information about the parameter distributions and their interactions uniform pdfs are often used.
- 92 2. Rejection criteria are defined to determine the accepted agreement between the simulators and the observed  
93 system state. These can be based on subjectively chosen thresholds limits or as accepted minimum levels of  
94 performance allowing for the expected uncertainties in the observational data.
- 95 3. Input and internal parameter sets are sampled from the prior pdfs to generate an ensemble of simulation  
96 predictions of the system for a given analysis time.
- 97 4. Simulations that are not in agreement with the observed system, using the selected rejection criteria, are  
98 discarded from the analysis. The subset of retained simulators with known parameter sets forms the posterior  
99 pdfs for each input and internal parameter. Thus, the posterior pdfs are the conditional pdfs of each parameter  
100 given the observations.
- 101 5. Input and internal parameter values are then sampled from the posterior pdf to generate an ensemble  
102 prediction of the system at a future time.
- 103 6. Steps 4-5 can be repeated for subsequent analysis times and the joint posterior distributions compared to the  
104 preceding analysis time.

105 The main aim of this paper is to contribute to the development of data assimilation methods to improve quantitative  
106 ash dispersion forecasts. To this end we will determine whether satellite retrievals of volcanic ash column loading can

107 be used to filter an ensemble of volcanic ash simulations using the particle filtering methodology. We determine which  
108 of the input and internal model parameters are most constrained by the satellite observations and quantify how the  
109 assimilation of satellite data changes the uncertainty estimate of the ensemble. Finally, for communicating the  
110 volcanic ash forecasts, we apply to the ensemble output the risk-matrix approach described in Prata et al. (2019) and  
111 applied retrospectively to the 2011 Grimsvotn eruption by Harvey et al. (2020), where risk is defined as the likelihood  
112 of exceeding ash concentrations considered a potential risk to aircraft. This approach will be demonstrated using  
113 simulations and observations of the Raikoke volcanic eruption which was a short duration eruption lasting less than  
114 24 hours occurring between the 21–22 June 2019.

## 115 **2 Methods and data**

116 In this study, the Numerical Atmospheric–dispersion Modelling Environment, NAME (Jones et al., 2007), was used  
117 to simulate the dispersion of volcanic ash. It is the VATDM used by the London Volcanic Ash Advisory Centre  
118 (LVAAC) for producing volcanic ash advisories following an eruption. Each simulated ash cloud was quantitatively  
119 evaluated using retrievals from Himawari–8.

### 120 **2.1 Himawari–8**

121 Himawari–8 is a geostationary satellite that came into operation in July 2015 (Bessho et al., 2016). It has 16 spectral  
122 channels and provides observations of high temporal frequency (10 min) and spatial resolution (2 km for the infrared  
123 bands). The high temporal and spatial resolution make these observations ideally suited to evaluate the transport of  
124 volcanic ash following an eruption. The [Met Office](#) volcanic ash retrievals used in this study are based on the method  
125 by Francis et al. (2012) with slight adaptations for the channels of the [Advanced Himawari Imager \(AHI\)](#) instrument  
126 aboard the Himawari–8 satellite.

127 The volcanic ash retrieval algorithm has two steps ([Francis et al. 2012](#)). The first step detects which pixels contain  
128 volcanic ash using the channels at 8.6  $\mu\text{m}$ , 10.4  $\mu\text{m}$  and 12.4  $\mu\text{m}$ . The second step runs a one–dimensional variational  
129 (1D–Var) analysis to determine an optimal estimate between the assumed background and the observed radiances in  
130 the channels at 10.4  $\mu\text{m}$ , 12.4  $\mu\text{m}$ , and 13.3  $\mu\text{m}$  for the column loading, ash cloud height and effective radius. The  
131 detection is based on a combination of brightness temperature difference (BTD) tests and beta ratio tests (Pavolonis,  
132 2010). The beta ratio tests use a derived radiative parameter  $\beta$ , that is the effective absorption optical depth ratio of  
133 two channels and are used to filter pixels marked as ash by the BTD tests. These tests have been improved by fine  
134 tuning of the operational thresholds, to optimise coverage of the June 2019 Raikoke eruption. In addition, several  
135 geographical filters have been added, to reduce false detections at high satellite zenith angle and over arid land  
136 surfaces, and further false detections have been removed by checking the consistency of ash detection in neighbouring  
137 pixels.

138 The retrieval algorithm also provides a measurement of the error on each of the retrieved values. The retrieval relies  
139 on the minimization of a cost function to determine the optimal estimate from the assumed background and the

140 observed radiances. How well defined the minimum of the cost function is provides an indication of the likely accuracy  
141 of the retrieval, in that the more well defined the minimum of the cost function is, the more accurate the retrieval is  
142 likely to be. By considering the inverse of the second derivative of the cost function with respect to each of the  
143 variables considered in the retrieval, we can provide an estimate of the error for the retrieved ash plume pressure, ash  
144 column loading and ash effective radius.

145 Where ash is detected, these pixels are flagged as ash and this algorithm determines the ash column loading. If a pixel  
146 is free from both ash and meteorological cloud, then it is flagged as a clear sky pixel. Pixels that neither have detectable  
147 ash nor are flagged as clear skies are unclassified. As in Harvey et al. (2020), further processing is performed to regrid  
148 the retrieved column loadings on to a grid of  $0.375^\circ$  latitude by  $0.5625^\circ$  longitude (approximately 40 km x 40 km in  
149 mid-latitudes) and averaged over 1 h. This is to match the resolution of the VATDM ash concentration output and to  
150 reduce data volumes. If all classified pixels within a grid box are flagged as clear sky pixels, then the grid box is  
151 deemed to be a clear sky observation. Otherwise, the grid box is deemed to be an ash grid observation with the column  
152 loading in this grid box given by the mean of all the classified pixels (including clear skies).

## 153 2.2 VATDM

154 All the simulations were performed in parallel mode using NAME version 8.1 on the Joint Analysis System Meeting  
155 Infrastructure Needs (JASMIN) super data cluster (Lawrence et al., 2012). Each simulation ran for 10 to 60 minutes  
156 to complete a -96-h forecast. The total computational time necessary to run a complete 1000-member ensemble varied  
157 from a minimum of 7 hours up to 110 hours, depending on the available resources on JASMIN and queuing times. To  
158 simulate the dispersion and removal of volcanic ash, NAME includes parametrization of the effects of turbulence on  
159 the transport and dispersion, sedimentation, dry deposition and wet deposition. In the operational configuration used  
160 by the LVAAC (Beckett et al., 2020) aggregation of ash particles, near source plume rise and processes driven by the  
161 eruption dynamics (e.g., Woodhouse et al., 2013) are not explicitly modelled. The default particle size distribution  
162 used is based on data from Hobbs et al. (1991) and the shape of the particles are assumed to be spherical.

163 Ensembles of NAME simulations were created by varying nine parameters covering the meteorology, information  
164 about the eruption source and the parameterisation of turbulence in NAME (as in Harvey et al. 2018; Prata et al.,  
165 2019). Uniform distributions between the specified ranges were used as prior probability distributions to generate the  
166 initial ensemble (Table 1). Full details of how these parameters are sampled is given in Sect.4.1.

167 In each ensemble, all simulations share the same start and end time, 18:00 UTC on 21 June 2019 and 1200 UTC on  
168 25 June 2019 respectively, for a total run time of 96 hours. The eruption start time matches the simulation start time.  
169 Volcanic ash within the simulations is released along a vertical line, between the lower and upper plume heights  
170 (Witham et al., 2019). Ash column loadings ( $\text{g}/\text{m}^2$ ) and ash concentrations ( $\text{g}/\text{m}^3$ ) are output onto a global grid of  $800$   
171  $\times 600$  points, corresponding to a grid of  $0.45^\circ$  longitude and  $0.3^\circ$  latitude, giving a horizontal resolution of  
172 approximately 40 km in midlatitudes. Ash column loadings are instantaneous loading values outputted every 6h, and  
173 ash concentrations are output every 6 hours using a 6-hour time average at 22 Flight Levels (FL000–FL550) with a  
174 vertical resolution of 25 FL (“thin layers”). All the Flight Levels are then combined to form three “thick” layers

175 (FL000–200, FL200–350, FL350–550) by taking the maximum concentration values from the component thin layers  
 176 for the corresponding thick layer value (Witham et al., 2019).

177 **Table 1: Parameters sampled and their control and initial sampling ranges.**

| Parameter   | Symbol      | Control value                               | Initial sampling range  |
|---|-------------|---|---|
| Plume height (km) <sup>1</sup>  | $H$         | 12.45                                       | 9–17  |
| Mass eruption rate factor <sup>2</sup>  | $MER F$     | 1   | 0.33–3  |
| Ash density (kg/m <sup>3</sup> )  | $\rho$      | 2300  | 1350–2500   |
| Source duration (hr)  | $L$         | 12  | 9–15  |
| Distal fine ash fraction (%)  | $DFAF$      | 5   | 0.5–20  |
| Horizontal (vertical) Lagrangian timescale for free tropospheric turbulence (s)               | $\tau$      | 300 (100)                                   | 100–900 (33.33–300)   |
| Standard deviation of horizontal (vertical) velocity for free tropospheric turbulence (m/s)   | $\sigma$    | 0.25 (0.1)                                  | 0.0025–2.5 (0.001–1)  |
| Standard deviation ( $\sigma$ ) of horizontal velocity for unresolved mesoscale motions (m/s) | $m\sigma U$ | 0.8   | 0.27–1.74   |
| Meteorological fields   | $MET$       | Met Office Unified Model<br>global analysis | <a href="#">Met Office Global and Regional Ensemble Prediction System</a><br>members 0–17 |

178

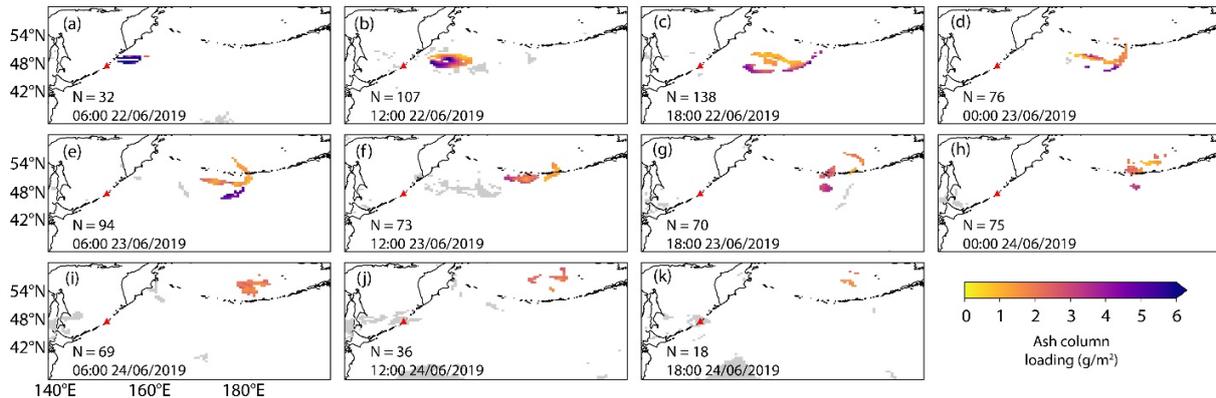
179 <sup>1</sup>Plume rise height above the summit

180 <sup>2</sup>scaling factor applied to the default mass eruption rate value calculated using the equation from Mastin et al. (2009); see Sect.  
 181 4.1.1 for details on how mass eruption rate is calculated.

### 182 3 Description of the 2019 Raikoke eruption

183 Raikoke is an uninhabited volcanic Island near the centre of the Kuril Island chain in the Sea of Okhotsk in the  
 184 northwest Pacific Ocean located at 48.2°N, 153.3°E. Its most recent explosive eruption, after 95 years of dormancy,  
 185 started at approximately 1800 UTC on 21 June 2019 and is estimated to have an initial eruptive plume height of 10–  
 186 13 km above sea level (asl) (Global Volcanism Project, 2019). The eruption lasted approximately 12 hours and ended  
 187 at approximately 0600 UTC on 22 June 2019. There is evidence from visible satellite imagery to suggest that there  
 188 was an umbrella cloud which was quickly advected eastwards towards a large extratropical cyclone which distorted  
 189 the dispersed ash cloud (Fig. 1).

190 The number of satellite grid boxes that are classified as containing ash at each time are the ones available to be used  
 191 to refine the prior pdfs. The largest number of boxes are available at 1800 UTC on 22 June. Before 0600 UTC on 22  
 192 June the number of grid boxes available is limited by the small time the ash has had to be transported. After 1800 UTC  
 193 on 24 June the number of grid boxes is limited due to the presence of meteorological cloud associated with an  
 194 extratropical cyclone situated to the east of Raikoke.



195 **Figure 1: Hourly mean ash column loadings from the Himawari-8 satellite at (a) 0600 UTC, (b) 1200 UTC, (c) 1800 UTC**  
 196 **on 22 June 2019, (d) 0000 UTC, (e) 0600 UTC, (f) 1200 UTC, (g) 1800 UTC on 2300 June 2019, (h) 0000 UTC, (i) 0600 UTC,**  
 197 **(j) 1200 UTC, (k) 1800 UTC on 24 June 2019. Grey shading indicates grid boxes that are classified as clear sky. The red**  
 198 **triangle indicates the location of Raikoke. N indicates the number of gridboxes classified as containing ash. The same**  
 199 **colorbar applies to all maps.**  
 200

#### 201 4 Particle filter construction

202 Drawing from the GLUE methodology described in Sect. 1, we developed a new particle filter for refining a series of  
 203 ensembles moving forward in time based on their level of agreement with Himawari-8 satellite observations. All  
 204 ensemble members are evaluated and filtered by comparing the NAME-simulated ash column loadings ( $\text{g}/\text{m}^2$ ) with  
 205 the satellite-detected ash column loadings ( $\text{g}/\text{m}^2$ ) for a given analysis time. Column loading retrievals from Himawari-  
 206 8 cover a time range from 1800 UTC on 21 June 2019 to 0000 UTC on 24 June 2019. However, due to the low number  
 207 of grid boxes containing ash before 1800 UTC on 22 June (Fig. 1), for the initial ensemble (*ENS01*) we chose as first  
 208 verification time,  $T_1$ , the observations at 0600 UTC on 22 June 2019. For this given time, 32 grid boxes containing  
 209 ash are available (Fig. 1). Each satellite hourly average represents an average of data available at two times for each  
 210 hour of the observation (e.g., for 0600 UTC on 22 June 2019, the satellite observations are an average of data from  
 211 0600 UTC and 0530UTC). Comparing NAME-simulated ash column loadings with observations that had been time-  
 212 averaged over 1, 3 or 6 hours showed little change with averaging duration. However, averaging over longer periods  
 213 led to a decrease of the number of grid boxes available for comparison for some of the observations. Therefore, we  
 214 retained 1-hour average observations for comparisons.

215 The particle filtering operation is designed such that, once all the members of an ensemble have been evaluated based  
 216 on user defined rejection criteria, only those within the limits of acceptability are retained and used to produce posterior  
 217 pdfs. A posterior ensemble is created by resampling the perturbed parameters from the posterior pdfs which are then  
 218 compared forward in time with a new set of observations,  $T_n$ . Therefore, each posterior ensemble represents a new  
 219 possible state of our simulated ash cloud at a future time based on the evaluation performance of the prior ensemble  
 220 at a previous time. The main steps involved in the methodology are summarised (Fig. 2):

- 221 1. Prior pdfs are created for the 9 perturbed parameters, including eruption source parameters, driving  
 222 meteorology, and NAME model internal parameters (Table 1; Fig. 2). We assign an initial range for each

- 223 parameter that is then sampled independently from a uniform distribution to create an ensemble of 1000  
224 members, *ENS01*. Perturbed parameters and their initial ranges are detailed in Sect. 4.1.
- 225 2. We define our rejection criteria, based on the threshold values for our verification metrics: Hit Rate (*HR*) and  
226 Mean Percentage Difference (*MPD*). Sect. 4.2 describes how we calculate *HR*, *MPD* and how we set the  
227 thresholds (Fig. 2).
  - 228 3. Each model prediction from *ENS01* is compared with the satellite observations at a given time, *T1*. All  
229 posterior ensembles are verified at a future time using observations every 6 hours (Fig. 2):

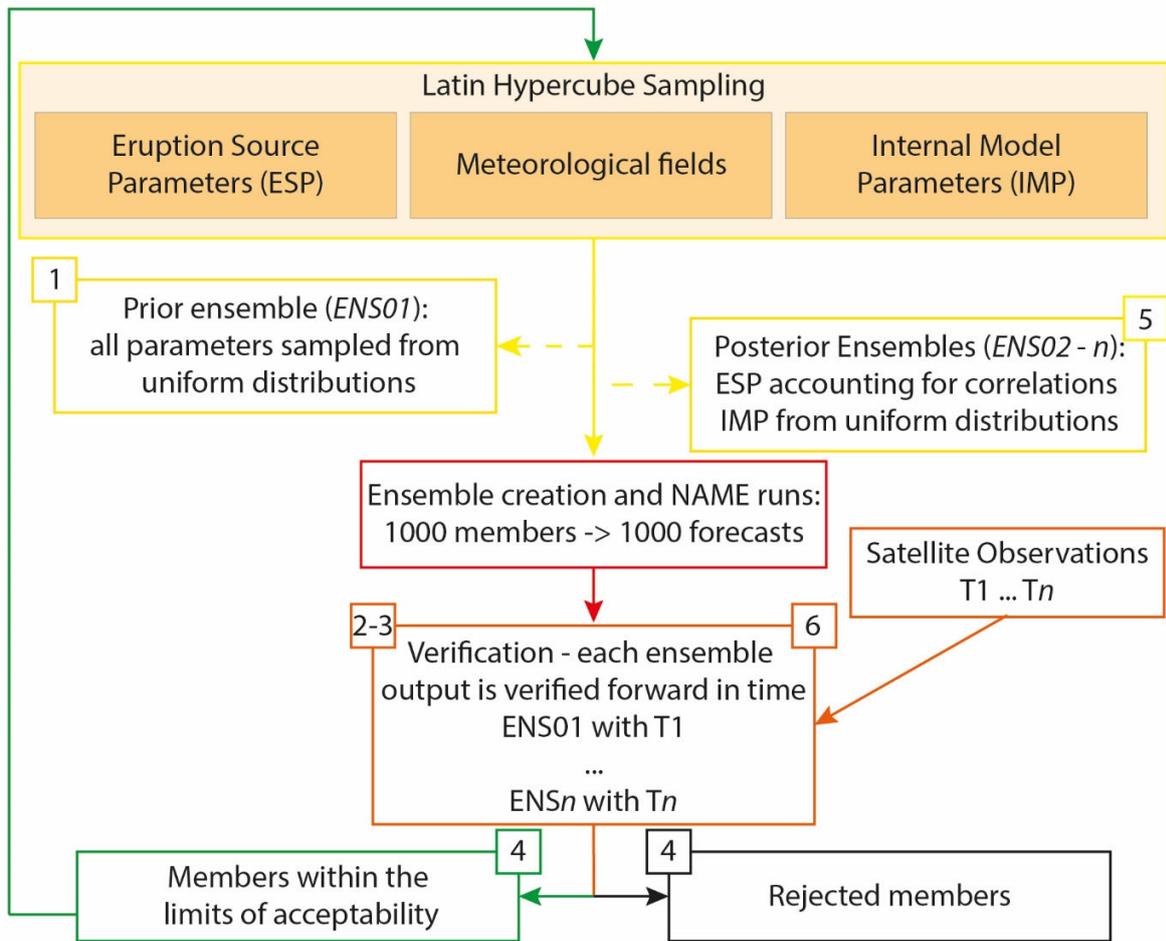
230  $ENS01 \rightarrow T1 = 22/06/19\ 0600\ UTC$

231  $ENS02 \rightarrow T2 = 22/06/19\ 1200\ UTC$

232 ...

233  $ENS\ n \rightarrow T_n = T_{n-1} + 6h$

- 234 4. Simulations that, based on the rejection criteria, are not in agreement with the satellite observations are  
235 discarded. The retained simulations are used to form the posterior pdfs for each input and internal model  
236 parameter (Sect. 4.3; Fig. 2).
- 237 5. Posterior pdfs are generated from the parameter sets of the retained simulations, also considering possible  
238 interaction among them by including effects of covariation between eruption source parameters. Parameters  
239 are resampled from these posterior pdfs, creating a new 1000-member posterior ensemble (Fig. 2). The newly  
240 created posterior ensemble represents a possible state of our system at a future time (Sect. 4.3). Each member  
241 of the new ensemble produces a new 96-h forecast, from 1800 UTC on 21 June 2019 to 1200 UTC on 25  
242 June 2019.
- 243 6. Each model prediction from the posterior ensemble is compared forward in time with a new set of satellite  
244 observations (Table 2; Fig. 2).
- 245 7. The steps 4–6 are repeated until all the available satellite observations are covered.



246

247 **Figure 2: Overall workflow of the particle filtering methodology.**

248 **4.1 Ensemble creation**

249 Each ensemble of NAME simulations is created by perturbing nine parameters, including eruption source parameters,  
 250 driving meteorology, and internal NAME parameters. To generate the ensemble efficiently, we use latin hypercube  
 251 sampling (LHS), that covers our entire parametric space maintaining orthogonality among the different perturbed  
 252 parameters (Prata et al., 2019). For *ENS01*, each parameter in the LHS is chosen from prior pdfs which ranges are  
 253 defined by a set of minima and maxima (Table 1) and sampled from a uniform distribution assuming that all values  
 254 within these ranges are equally likely. For each posterior ensemble the ESPs in the updated LHS designs are chosen  
 255 from the posterior pdfs, while NAME internal parameters and members of [the Met Office Global and Regional](#)  
 256 [Ensemble Prediction System \(MOGREPS-G\)](#) forecasts continue to be sampled from uniform distributions (Sect.4.3).

257 **4.1.1 Eruption Source parameters (ESPs)**

258 To represent an initial uncertainty associated with ESPs for *ENS01*, we define a minimum and a maximum possible  
259 value for each perturbed parameter. Then, a parameter value is sampled from a uniform distribution across this range.

260 We selected a total of 5 ESPs to perturb from those that have been shown to have most effect on the simulated ash  
261 cloud (e.g., Dacre et al., 2013, Harvey et al., 2018, Prata et al., 2019): plume height ( $H$ ), Distal Fine Ash Fraction  
262 ( $DFAF$ ), Mass Eruption Rate factor ( $MERF$ ), ash density ( $\rho$ ) and eruption duration ( $L$ ).  $H$  is used to calculate the Mass  
263 Eruption Rate ( $MER$ ) for each member using the empirical relationship from Mastin et al. (2009). For each ensemble  
264 member,  $MER$  is scaled by the  $DFAF$ , representing the percentage of ash transported at long distances, and multiplied  
265 for the  $MERF$ , to account for uncertainties associated with the Mastin et al. (2009) relationship. Hence,  $MER$  is not  
266 perturbed explicitly.

### 267 ***Plume height***

268 Plume height ( $H$ ) constrains the lower and upper limits of the ash particles' release height, therefore significantly  
269 impacts both the vertical and horizontal structure of the simulated ash cloud. Although the Raikoke eruption was  
270 characterized by the formation of an umbrella cloud, in the NAME ash release is defined as a vertical line along which  
271 the ash is uniformly distributed, with the lower and upper bounds representing the volcano summit height (551 m asl)  
272 and the reported plume height respectively. We based our initial  $H$  range on information available at that time. The  
273 Kamchatkan Volcanic Eruption Response Team (KVERT) and the Tokyo and Anchorage VAAC reported a large ash  
274 plume extending from 10 to 13 km (asl) within the first few hours of eruption (Crafford et al., 2019), while data from  
275 CALIPSO satellite indicate that the plume may have reached altitudes up to 17 km (asl; Hedelt et al., 2019). For our  
276 initial ensemble, we selected  $H$  ranges between 9–17 km (asl) and for the control run we set  $H = 13$  km (asl; Table 1).

### 277 ***Mass Eruption Rate (MER)***

278 Mass Eruption Rate,  $MER$ , is estimated from the plume height using the empirically derived relationship from Mastin  
279 et al. (2009):

$$280 \quad MER = 50.7 \times 10^7 H^{2/0.241} \quad (1)$$

281 where  $H$  represents the plume height in km.  $MER$  is expressed in  $g \text{ hr}^{-1}$ . By calculating the  $MER$  from  $H$ , any  
282 uncertainty associated with  $H$  propagates in the resulting  $MER$  calculation. Furthermore, there is also an uncertainty  
283 associated with the nature of the relationship, being an empirical one and based on a relatively small number of  
284 eruptions of variable magnitude (Mastin et al., 2009). In order to take this into consideration, for all the ensemble  
285 members, the  $MER$  is perturbed for a factor between 0.33 and 3, while we use 1 for the control run (Mass Eruption  
286 Rate Factor,  $MERF$ ; Harvey et al., 2018, Prata et al. 2019; Table 1).

### 287 ***Distal Final Ash Fraction (DFAF)***

288 The  $MER$  calculated with the Mastin et al. (2009) relationship (Eq. 1) estimates the total mass released during an  
289 eruption. However, the particle size distribution (PSD) of the erupted particles includes both larger particles ( $>1$  mm)

290 that are usually removed from the column in the first phases of an eruption and an additional finer fraction that may  
291 leave the column due to aggregation processes. Particles larger than 100  $\mu\text{m}$  are removed rapidly, without travelling  
292 long distances, and as result, only a fraction of fine particles  $<100 \mu\text{m}$  is transported at long distances. Details of the  
293 true PSD are often unknown. Here, the default LVAAC PSD is used in each simulation (Table 1 in Witham et al,  
294 2019), and aggregation processes are not modelled in the NAME simulations for this study. To account for this, the  
295 model assumes that most of the ash falls out close to the volcano, with only a small percentage of it reaching the distal  
296 plume. The NAME default value for this percentage (distal fine ash fraction, DFAF, Dacre et al. 2011), used here for  
297 the control run, is 5%; however, the real value is uncertain and varies with each eruption (Witham et al., 2019).  
298 Consequently, the uncertainty associated with DFAF can be very high (Grant et al. 2012). Recent studies challenged  
299 the 5% assumption by reaching contrasting conclusions: either 5% is too high for most of the eruptions (Gouhier et  
300 al. 2019). Or it is too low, severely underestimating mass loadings (Cashman and Rust, 2020). For our prior ensemble,  
301 the range used is 0.5–20 % (Table 1).

### 302 *Ash density*

303 The default LVAAC value for particle density is 2300  $\text{kg/m}^3$  (Witham et al., 2019) and particle shape is assumed to  
304 be spherical in the NAME simulations. At the time of writing, no specific ash density or shape information is available  
305 for the 2019 Raikoke eruption. Density was selected as parameter to perturb as it may help in representing uncertainty  
306 attributed to ash aggregation and particle shape (e.g., Harvey et al., 2018). The range used is 1350–2500  $\text{kg m}^{-3}$ , while  
307 we use the default 2300  $\text{kg/m}^3$  for the control run (Table 1).

### 308 *Source duration*

309 The overall duration of the intense phase of the Raikoke eruption is relatively well constrained, with KVERT reporting  
310 a strong explosive eruption beginning about 1805 UTC on 21 June and a weaker explosive event reported at 0540  
311 UTC on 22 June. However, ash emission continued, possibly until around 0800 UTC on 22 June, when KVERT  
312 reported a gas–steam plume with some ash content (Crafford et al., 2019). As uncertainty in the duration of ash  
313 emission may lead to uncertainty in both the location and timing of the modelled ash cloud (e.g., Prata et al., 2019),  
314 we considered a duration range of 9–15 hours for *ENS01* and 12 hours for the control run (Table 1). In the simulations,  
315 eruption source parameters are assumed constant throughout the release duration, although this is unlikely to be true  
316 for the Raikoke eruption.

### 317 **4.1.2 Driving Meteorology**

318 In this study, NAME was driven by the operational forecasts from the MOGREPS-[G](#). The global forecasts have 17  
319 ensemble members plus a control member. The horizontal resolution is approximately 20 km in the mid-latitudes and  
320 there are 70 vertical levels with the lid at approximately 80 km. Each forecast is run out for 7 days and they are  
321 initialised 4 times per day at 00, 06, 12 and 18 UTC (Bowler et al. 2008). At the time of the Raikoke eruption,  
322 MOGREPS-[G](#) used an on-line inflation factor calculation to calibrate the spread of the ensemble in space and time

323 and a stochastic physics scheme to account for model uncertainty (Flowerdew and Bowler, 2011; Flowerdew and  
324 Bowler, 2013). The MOGREPS–G forecasts used in this study were initialized at 1200 UTC 21 June 2019.

#### 325 **4.1.3 NAME internal model parameters**

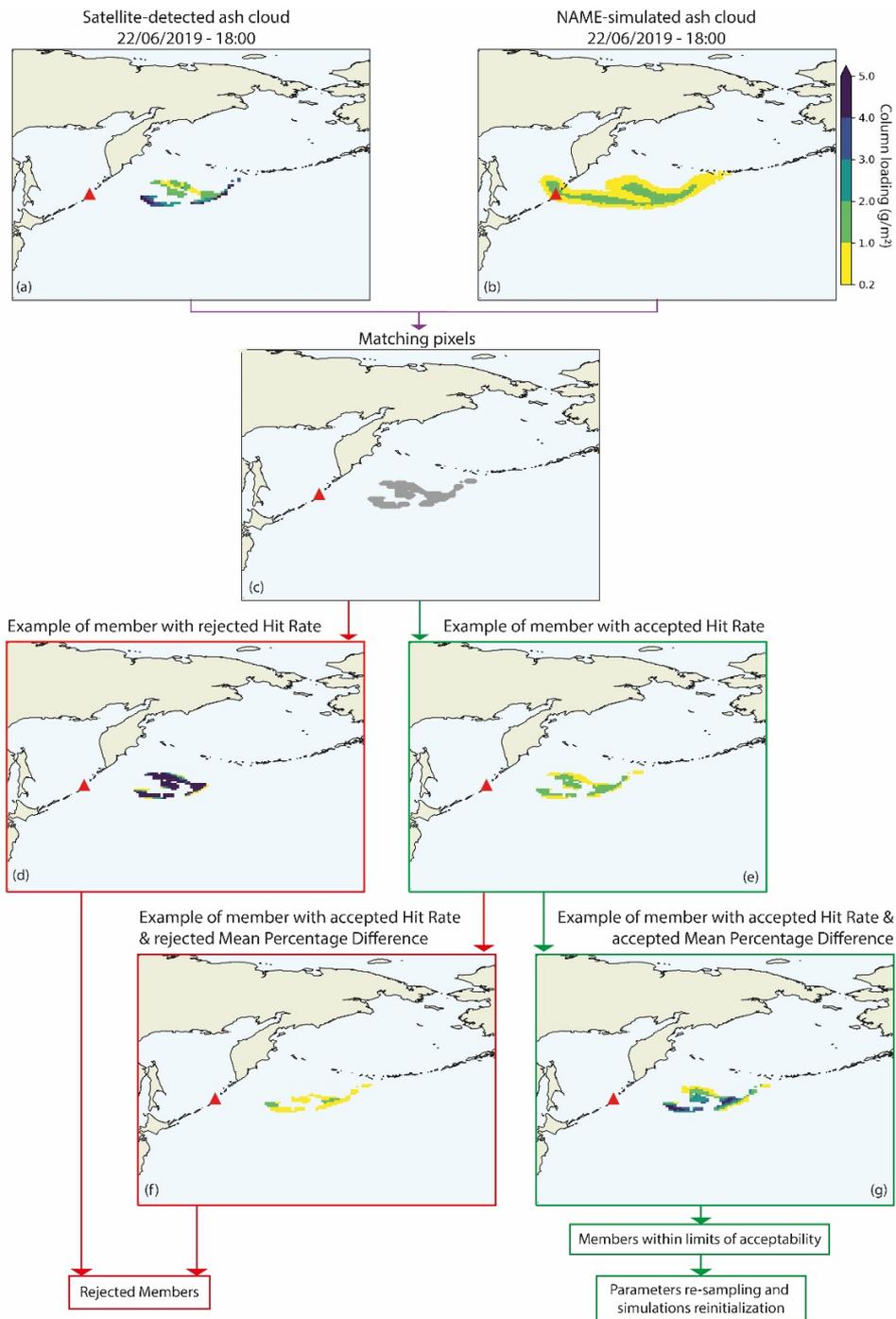
326 Previous studies have demonstrated how the NAME internal model parameters used for representing the free  
327 tropospheric turbulence can significantly impact the model output as they affect the vertical thickness of the simulated  
328 cloud and the overall motion of particles (Dacre et al., 2015; Harvey et al., 2018; Prata et al., 2019). To represent  
329 uncertainty in free tropospheric turbulence, we perturb the standard deviation ( $\sigma$ ) and Lagrangian timescales ( $\tau$ ) of the  
330 horizontal and vertical velocity components in NAME, by sampling them from a uniform distribution using the same  
331 ranges specified in Harvey et al. (2018) and Prata et al. (2019), with the horizontal component of  $\sigma$  sampled on a  
332 logarithmic scale. The horizontal and vertical components of these parameters are varied in proportion to each other.  
333 Similarly,  $\sigma$  of the horizontal velocity for unresolved mesoscale motions is also varied using the same range as in  
334 Harvey et al. (2018) and Prata et al. (2019) and sampled from a uniform distribution. For the control run, we use the  
335 default NAME values (Table 1). For both the control run and the ensembles, these values are fixed in time.

#### 336 **4.2 Particle filter verification metrics**

337 Each simulation is either discarded or retained based on its level of agreement with the satellite retrievals. We restrict  
338 the comparison only to the area covered by both NAME simulated ash cloud and by detectable ash in the satellite  
339 observations. The comparison is performed based on two verification metrics: Hit Rate (*HR*) and Mean Percentage  
340 Difference (*MPD*).

##### 341 **4.2.1 Step 1: Identifying matching pixels**

342 Most satellite retrievals are unable to detect column loadings less than 0.2 g/m<sup>2</sup> (Prata and Prata, 2012). Consequently,  
343 before comparing the Himawari-8 data and NAME output (Figs 3a and 3b), we apply a minimum threshold of 0.2  
344 g/m<sup>2</sup> to the NAME-simulated ash column loading to align with the minimum detection limit of the satellite  
345 observations. The Himawari-8 observations and their error are both regrided over the NAME horizontal grid, to  
346 facilitate inter comparisons. The regridding process is performed using IRIS (Met Office, 2021). There was no  
347 noticeable difference between regridding either each ensemble member based on the satellite grid, or each observation  
348 and its error based on the NAME grid, using the different regridding schemes available. However, regridding each  
349 ensemble requires long computational times. Therefore, we decided to regrid the observation and its error based on  
350 NAME as target grid, using an area-weighted regridding scheme (Met Office, 2021). Finally, we identify the grid  
351 boxes in which both the NAME output and the satellite retrievals detect ash as ‘matching pixels’ (Fig. 3c). For each  
352 ensemble member, we then calculate the Hit Rate and Mean Percentage Difference based on all matching pixels.



353

354 **Figure 3: Ash column loading at 1800 UTC on 22 June 2019, (a) as detected by the satellite, (b) NAME simulation of one**  
 355 **ensemble member in NAME, and (c) after pixel–matching (grey pixels represent a match between satellite and simulation).**  
 356 **Examples of ensemble members rejected depending on values of Hit Rate and Mean Percentage Difference (red boxes)**  
 357 **are shown in panels (d) and (f). Examples of accepted ensemble members depending on values of Hit Rate and Mean Percentage**  
 358 **Difference (green boxes) are shown in panels (e) and (g). All panels except (c) use the colour scale shown in (b). The red**  
 359 **triangle in each panel shows the location of Raikoke.**

#### 360 4.2.2 Step 2: Calculating Hit Rate

361 The Hit Rate, *HR*, is a widely used categorical metric applied to many meteorological phenomena for forecast  
362 verification, representing the proportion of observed events that are successfully forecast by a simulation. The *HR* can  
363 be used to discriminate “yes events” and “no events”, often by specifying a threshold to separate “yes” and “no”  
364 (Joliffe and Stephenson, 2012). Similarly, *HR* has also been used to provide information on how ash forecast model  
365 outputs compare to observations in terms of binary *ash yes/ash no* events (e.g., Stefanescu et al, 2014; Marti and  
366 Folch, 2018). In such cases, a grid box would represent a hit if both simulation and observation detected ash above a  
367 threshold of  $>0.2 \text{ g/m}^2$ .

368 Here, we calculate the Hit Rate as the percentage of matching pixels for which the simulated value lies within one  
369 standard deviation of the corresponding mean satellite–detected ash column loading. The standard deviation is  
370 provided by the retrieval algorithm as a measurement of the error for the retrieved ash values in each grid box  
371 (Sect.2.1). Hence, we both directly compare the ash column loadings between simulation and observations for each  
372 individual matching pixel, and we complement this by considering the error associated with the satellite observations.

373 Once we have identified the total number of hits in each member, the *HR* can be calculated:

$$374 \quad HR = \frac{HITS}{Matching\ Pixels} \times 100 \quad (2)$$

375 At this point, members that have an *HR* below a specific threshold are discarded (Sect. 4.2.4; Fig. 3d) and the  
376 remaining members (Fig. 3e) are then verified further by testing against the observations using Mean Percentage  
377 Difference.

#### 378 4.2.3 Step 3: Calculating the Mean Percentage Difference

379 The final step in the evaluation process determines, for the members retained following the *HR* verification, how much  
380 the NAME–simulated ash column loading values differ from the satellite–detected ones on a grid box basis. The  
381 percentage difference magnitude (*PD*) for each matching pixel is the absolute value of the difference between the  
382 simulated and observed column loadings, divided by the mean of the two values and expressed in terms of percentage.

383 Then, the mean of the *PDs* over all grid boxes (*MPD*) for each ensemble member is calculated. Ensemble members  
384 with *MPD* below a threshold (Sect. 4.2.4; Fig. 3g) are retained and used to form our posterior. Those above the  
385 threshold are rejected (Fig. 3f).

#### 386 4.2.4 HR and MPD Thresholds

387 Both *HR* and *MPD* are sensitive to the total number of satellite grid boxes containing detectable ash. Therefore, when  
388 there are only a few satellite grid boxes available, fixed *HR* and *MPD* thresholds may retain too few ensemble members  
389 to form posterior pdfs. As our verification method is based on limits of acceptability, to avoid a situation where either

390 all simulations are rejected or too few members are retained for resampling, we implemented dynamic thresholding  
 391 for both *HR* and *MPD*. The acceptability thresholds for *HR* and *MPD* are adjusted dynamically during the verification  
 392 to ensure that a minimum of 50 ensemble members are retained (i.e., 5% of total number of ensemble members). This  
 393 dynamic adjustment is carried out by initially setting the *HR* threshold to 95% and the *MPD* threshold to the minimum  
 394 *MPD* value. Ensemble members are evaluated against these thresholds, and the thresholds adjusted (by increasing the  
 395 *MDP* value up to the median value, and by decreasing the *HR* value, in 5% increments) until at least 50 members lie  
 396 within the limits of acceptability.

397 This dynamic thresholding method, therefore, guarantees that the members within limits of acceptability are always  
 398 retained using the “best” threshold available for both *HR* and *MPD*, for a given time (Table 2). Only for *T1*, when 32  
 399 grid boxes containing detectable ash are available, were both thresholds varied substantially and fewer than 50  
 400 members were retained. Thereafter, a *HR* of 95% was maintained at each verification cycle. *MPD* values had to be  
 401 varied more to retain at least 50 members but were always within the range 20–50 % of the minimum *MPD*.

402 **Table 2: number of members within limits of acceptability (WLoA), and *HR* and *MPD* thresholds for each ensemble at each**  
 403 **verification time**

| Ensemble | Verification Time   | Num. of grid boxes | Posterior members WLoA <sup>1</sup> | Prior members WLoA <sup>2</sup> | H.R. threshold (%) | M.P.D. threshold (%) |
|----------|---------------------|--------------------|-------------------------------------|---------------------------------|--------------------|----------------------|
| 01       | T1: 22/06/19 06:00  | 32                 | 22                                  | 22                              | 64                 | 152                  |
| 02       | T2: 22/06/19 12:00  | 107                | 53                                  | 0                               | 95                 | 60                   |
| 03       | T3: 22/06/19 18:00  | 138                | 90                                  | 7                               | 95                 | 65                   |
| 04       | T4: 23/06/19 00:00  | 65                 | 176                                 | 30                              | 95                 | 68                   |
| 05       | T5: 23/06/19 06:00  | 71                 | 108                                 | 33                              | 95                 | 78                   |
| 06       | T6: 23/06/19 12:00  | 55                 | 66                                  | 20                              | 95                 | 56                   |
| 07       | T7: 23/06/19 18:00  | 41                 | 60                                  | 58                              | 95                 | 71                   |
| 08       | T8: 24/06/19 00:00  | 70                 | 59                                  | 181                             | 95                 | 99                   |
| 09       | T9: 24/06/19 06:00  | 66                 | 52                                  | 35                              | 95                 | 54                   |
| 10       | T10: 24/06/19 12:00 | 31                 | 62                                  | 53                              | 95                 | 57                   |
| 11       | T11: 24/06/19 18:00 | 18                 | 118                                 | 13                              | 95                 | 29                   |

404 <sup>1</sup>total number of retained ensemble members for a given time. Parameters for Ensemble 01 (prior ensemble) were  
 405 sampled from uniform distributions. Parameters for each subsequent ensemble (posterior ensemble) were sampled  
 406 from posterior pdfs obtained from the retained members of the previous ensemble, verified at the previous verification

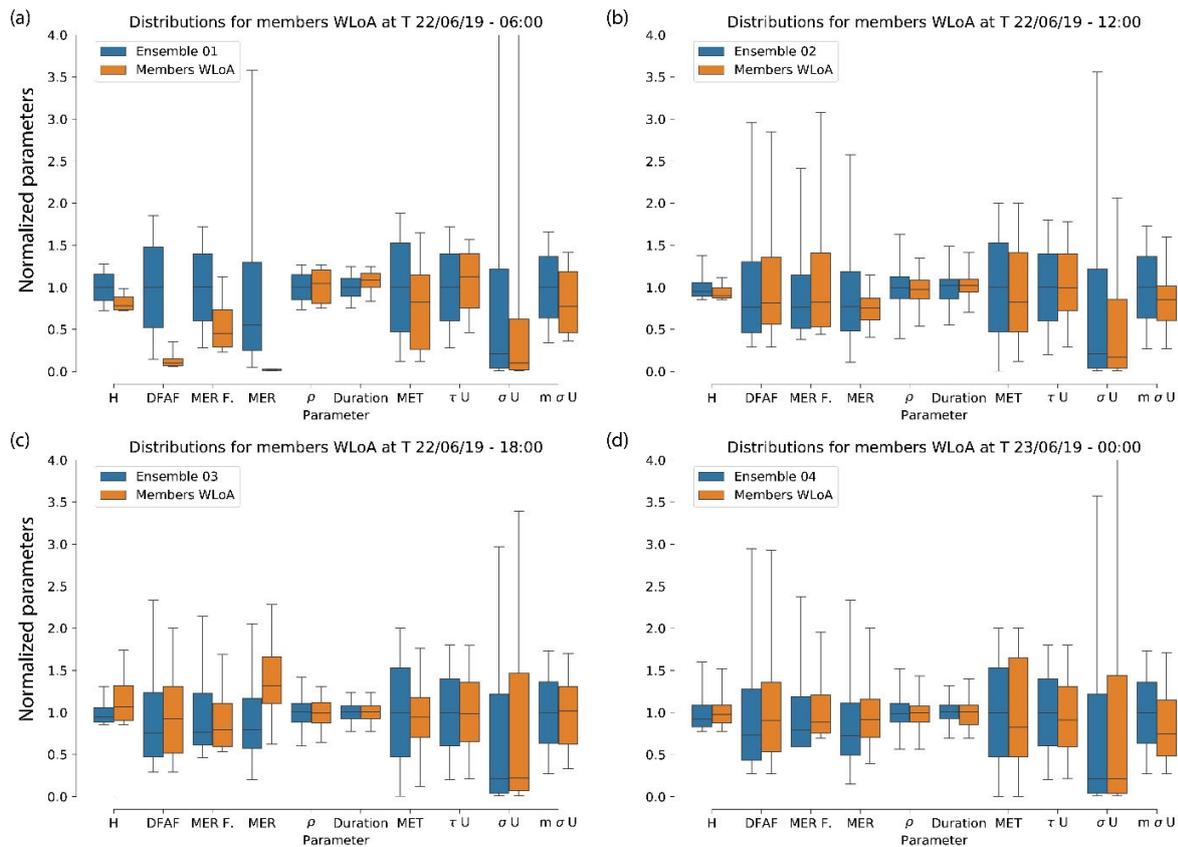
407 time. Threshold values for *HR* (higher is better) and *MPD* (lower is better) used at each verification time are shown  
408 in the last two columns.

409 <sup>2</sup> total number of ensemble members that would be retained from the prior ensemble, *ENS01*, if the verification was  
410 run at each verification time, using the same *HR* and *MPD* thresholds for which each posterior ensemble has been  
411 evaluated.

### 412 4.3 Posterior resampling

413 The comparison of the prior and posterior pdfs for each parameter allows us to identify the range of parameters that  
414 represent a good approximation of those generating the observed volcanic ash cloud at the verification time. Fig. 4  
415 shows this comparison for *ENS01*, *ENS02*, [ENS03](#) and *ENS04*; each ensemble was verified using observations at *T1*,  
416 *T2*, [T3](#) and *T4* respectively (Table 2). For *ENS01* (Fig. 4a), most of the ESP pdfs from the retained simulations are  
417 highly skewed. In particular, this can be seen for *H*, *DFAF* and *MERF*. *H* is used to calculate the mass eruption rate  
418 (shown in Fig. 4 but not explicitly perturbed), that is then perturbed further by *DFAF* and *MERF*. Therefore, these  
419 ESPs significantly influence both the vertical and horizontal structure of the modelled ash cloud and ash  
420 concentrations. However, *ENS02* (Fig. [4b](#)), [ENS03](#) (Fig. [4c](#)) and *ENS04* (Fig. 4d), which use the particle filter  
421 described above, show how as more ensembles are run and evaluated forward in time with new observations, the  
422 parameter ranges of each posterior ensemble gradually reduce. Additionally, the differences between posterior pdfs  
423 and the prior pdfs decrease and the ESPs become increasingly constrained.

424 Although this evolution is evident for many of the ESPs, the input model parameters ( $\tau U$ ,  $\sigma U$  and  $m\sigma U$ ) do not show a  
425 similar behaviour among the different ensembles. [This seems to hold for all the 11 ensembles \(Fig. 4 and A1 in](#)  
426 [Appendix A\), suggesting](#) that the Raikoke simulations are not sensitive to these internal parameters.



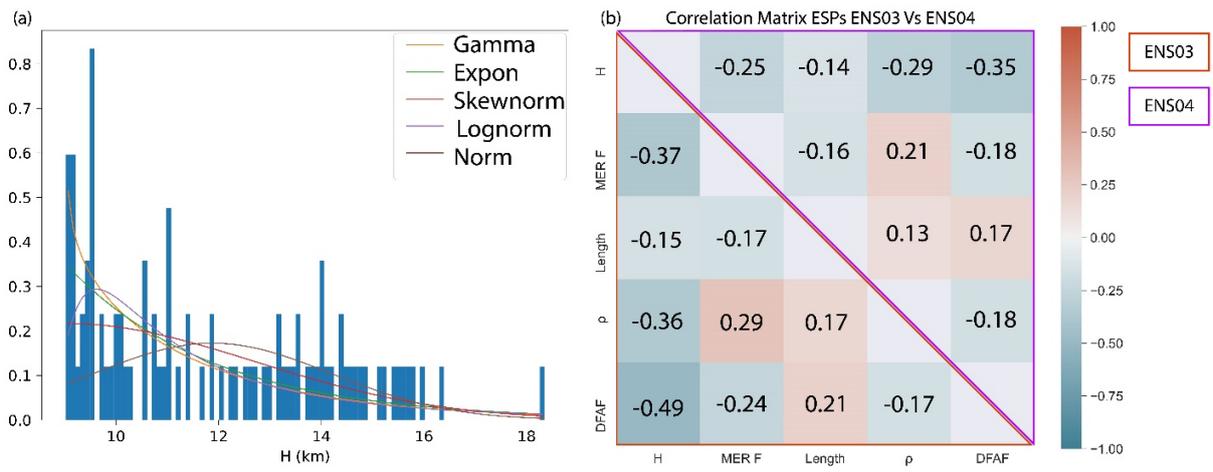
427  
 428 **Figure 4: Evolution of parameters distributions for (a) members Within Limits of Acceptability (WLoA) for (a) Ensemble**  
 429 **01, (b) Ensemble 02, (c) Ensemble 03, and (d) Ensemble 04.  $\rho$ ,  $\tau U$ ,  $\sigma U$  and  $m\sigma U$  represent density, horizontal Lagrangian**  
 430 **timescale for free tropospheric turbulence, standard deviation of horizontal velocity for free tropospheric turbulence, and**  
 431 **Standard deviation ( $\sigma$ ) of horizontal velocity for unresolved mesoscale motions. In (a), (b), (c) and (d), Mass Eruption Rate**  
 432 **(MER) is plotted to show its variation, although it is not explicitly perturbed in any of the ensembles. Each parameter in**  
 433 **the boxplots is normalized by dividing each individual value from the ensemble members by the mean of that entire**  
 434 **parameter range from the selected ensemble. The evolution of parameters distributions for ENS05-ENS11 are shown in the**  
 435 **Appendix A (Fig. A1).**

436 This constraining behaviour is the result of the refinement of each posterior ensemble, achieved by repeatedly refitting  
 437 the parameters sets of the retained simulations at each verification time, and resampling posterior ensemble parameters  
 438 from the newly fitted posterior pdfs. For the prior ensemble, the LHS is performed independently for each parameter.  
 439 However, eruption source parameters are likely to be correlated, especially the plume height, distal fine ash fraction  
 440 and mass eruption rate factor, which are used to estimate and perturb the mass eruption rate. Thus, we want to maintain  
 441 dependency among the ESPs during the resampling process. The main steps of this process are the following:

- 442 1. The input parameter ranges from the retained members are fitted with a gamma distribution (which  
 443 represented the overall best-suited distribution, Fig. 5a), and used to generate a correlation matrix for the  
 444 ESPs (Fig. 5b).

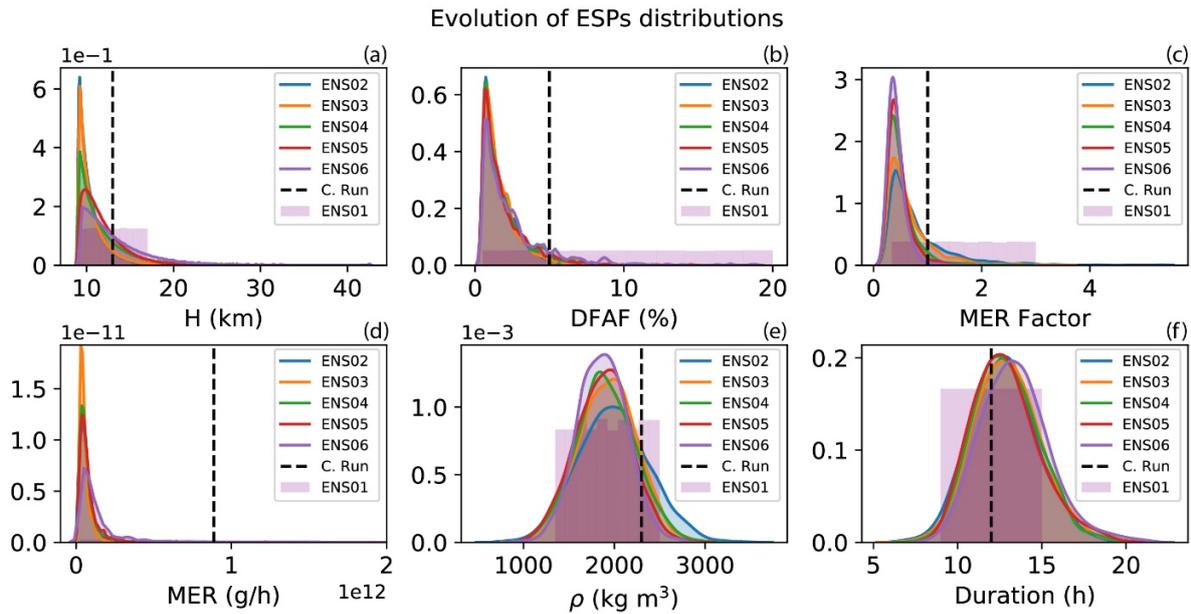
2. A new LH design is created (1000 samples for 5 parameters), and each sample is mapped to values applying inverse cumulative distribution function (cdf) of a normal variable  $N(0,1)$ . The generated values are independent and follow a normal distribution.
3. Using a Cholesky decomposition of the correlation matrix, a correlation structure is enforced to the new LHS design, adding dependency to the normal values of our parameters.
4. Finally, by applying a normal cdf to the normal variables, we transform them into uniform random variables, and then map the uniform distribution to the specified distribution function (gamma), applying the inversion cdf of our gamma distributions.

Each ESP is sampled from the newly generated posterior pdfs in the updated LHS design, therefore maintaining the dependency among the parameters (Fig. 5b). In contrast, as we did not observe the same constraining behaviour for the model input parameters and driving meteorology (Fig. 4 and A1), those parameters are treated independently and their posterior pdfs are sampled again from uniform pdfs, using the same ranges as for the prior ensemble (Table 1).



**Figure 5: (a) Example of distribution identification for fitting plume height (H) from members WLoA of ENS03 and (b) comparison of correlation matrices between ENS03 ESPs in the members WLoA (lower matrix, red triangle) and ESPs posterior pdfs for ENS04 (upper matrix, purple triangle).**

For the majority of the ESPs in *ENS02*, this resampling strategy modifies the ESPs distributions away from the initial uniform distribution of the prior ensemble to a distribution that is highly skewed towards the lower end of the initial range (Fig. 6). Then, as the posterior pdfs are refined by moving forward in time with new observations, the pdfs for some of the ESPs, such as particle density and duration, seem to approximate normal distributions (Fig. 5c).



467

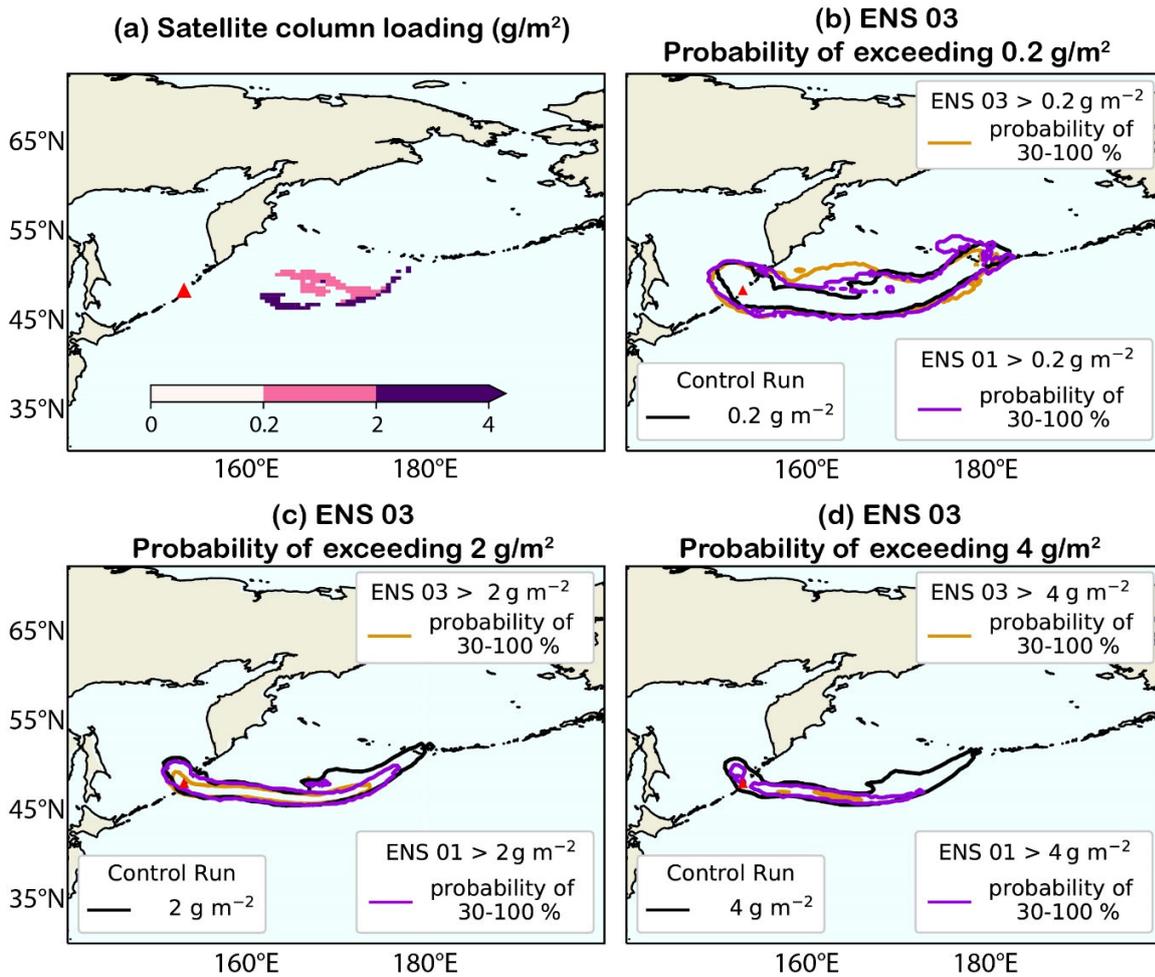
468 **Figure 6:** Evolution of ESPs distributions for ENS01 (purple shading), ENS02 (blue shading), ENS03 (orange shading),  
 469 ENS04 (green shading), ENS05 (red shading) and ENS06 (dark purple shading). Each panel shows: (a) plume height ( $H$ ),  
 470 (b) distal fine ash fraction ( $DFAF$ ), (c) mass eruption rate factor ( $MERF$ ), (d) mass eruption rate ( $MER$ ), (e) ash density ( $\rho$ ),  
 471 and (f) eruption duration. For ENS01, each parameter is sampled from a uniform distribution. For ENS02–06, each  
 472 parameter is sampled from posterior pdfs following a fit of the members WLoA and considering the interaction among the  
 473 parameters.  $MER$  is plotted to show its evolution, although it is not explicitly perturbed in any of the ensembles. Dashed  
 474 black lines in each plot represent the parameter value used for the control run. **The evolution of ESPs distributions for**  
 475 **ENS07-ENS11 are shown in the Appendix B (Fig. B1).**

476 **5 Discussion**

477 The particle filtering data assimilation technique described in Sect.4 demonstrates how a series of ensembles of  
 478 volcanic ash simulations can be successfully constrained based on the level of agreement between the simulation  
 479 output and satellite retrievals. Each ensemble is verified forward in time with new retrievals. Compared to the  
 480 parameter ranges used for the prior ensemble (Table 1), the range of eruption source parameters used to produce  
 481 simulated ash clouds that represent a good approximation to the observed volcanic ash cloud reduces as the posterior  
 482 ensembles become more constrained by the satellite retrievals.

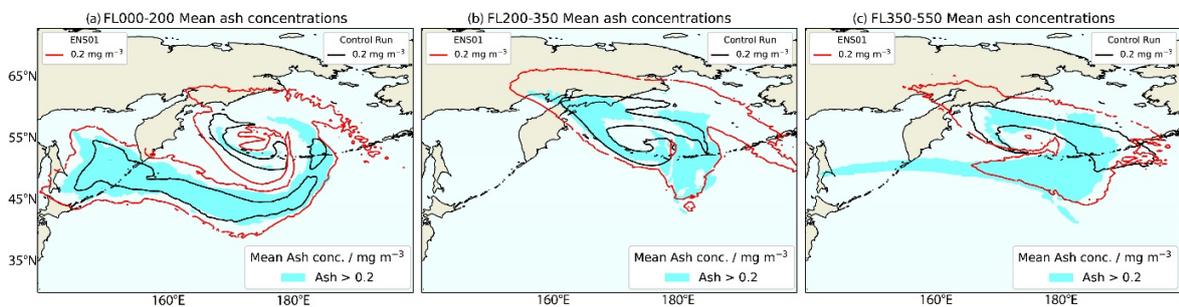
483 The effects of the refinement on the posterior ensemble can be observed in probability of exceedance maps, given  
 484 here for three different thresholds of ash column loadings,  $0.2 \text{ g/m}^2$  (Fig. 7b),  $2 \text{ g/m}^2$  (Fig. 7c) and  $4 \text{ g/m}^2$  (Fig. 7d).  
 485 The satellite retrieval at 1800 UTC 22/06/2019 detects two distinct regions where ash loadings exceed  $0.2 \text{ g/m}^2$  and  $2$   
 486  $\text{g/m}^2$  (Fig. 7a). The posterior ensemble (ENS03 – brown contour) agrees with the control run (black contour) and the  
 487 prior ensemble (purple contour) on regions where loadings  $> 0.2 \text{ g/m}^2$  are likely (30–60 %) and very likely (60–100  
 488 %; not plotted in Fig. 7). The simulated ash cloud regions are more extensive than the area of satellite-detected ash.  
 489 However, this overestimation is greater for both the prior ensemble (ENS01 – purple contour) and the control run

490 (black contour) than for the posterior ensemble (ENS03 – brown contour) and is largest for column loadings  $> 2 \text{ g/m}^2$   
 491 (Fig. 7c) and  $4 \text{ g/m}^2$  (Fig. 7d). The refined ensemble shows a much-reduced region with a 30–100 % probability of  
 492 exceeding these loadings, showing better agreement with the observations especially when considering loadings  $> 2$   
 493  $\text{g/m}^2$  detected by the satellite (Fig. 6a7a). Thus, by accounting for uncertainties, a wider region where those loadings  
 494 are less likely has been shown instead, for all the considered thresholds (up to 30%, dashed blue contour, not plotted  
 495 in Fig. 67 but in Appendix C, Fig. C1).



496  
 497 **Figure 7:** (a) Satellite-detected column loading at T3 (22/06/2019 1800 UTC) shown in filled contours. Probability of  
 498 exceedance maps for ENS03 at T3 for column loading thresholds of (b)  $0.2 \text{ g m}^{-2}$ , (c)  $2 \text{ g m}^{-2}$  (d)  $4 \text{ g m}^{-2}$  shown with dashed  
 499 **brown** contour (30 % probability region – the region within the **brown** line goes from 30 % up to 100 %). The black contour  
 500 in panels (b), (c) and (d) show the relevant column loading threshold for the control run. The **purple** contours show the 30  
 501 % probability region for ENS01 of exceeding the relevant column loading threshold (the region within the **purple** line goes  
 502 from 30 % up to 100 %). For generating the **purple** contours, members from the prior ensemble were retained by verifying  
 503 ENS01 with observations at T3 (a) and using the same *HR* and *MPD* thresholds for which ENS03 has been evaluated (Table  
 504 2) to ensure a fair comparison between the prior and posterior ensembles. See Fig. C1 in Appendix C for the same  
 505 **probability maps including the 0-30% probability regions for ENS03.**

506 By considering mean ash concentrations ( $\text{mg}/\text{m}^3$ ) at T08 for *ENS08*, thus at 0000 UTC on 24 June 2019, for the three  
 507 “thick” Flight Layers, FL000–200 (Fig. 8a), FL200–350 (Fig. 8b) and FL350–550 (Fig. 8c), the control run seems to  
 508 underestimate the areas with concentrations exceeding  $0.2 \text{ mg}/\text{m}^3$  compared to the subset of retained members of both  
 509 *ENS01* and *ENS08*. Contrary, *ENS01* forecasts concentrations exceeding  $0.2 \text{ mg}/\text{m}^3$  over larger regions compared to  
 510 the posterior ensemble. The overestimation increases considerably for FL200–350 but also for FL350–550, even with  
 511 the posterior ensemble forecasting an additional plume tail extending to the east of Raikoke (Fig. 8c). Indeed, the areas  
 512 forecasted by *ENS08* with concentrations  $> 0.2 \text{ mg}/\text{m}^3$  for both FL200–350 (Fig. 8b) and FL350–550 (Fig. 8c) are,  
 513 respectively, around 60% and 30% less extended than the ones forecasted by *ENS01*. In general, for all three flight  
 514 levels, the area that a posterior ensemble would forecast with high ash concentrations is drastically reduced compared  
 515 to the prior ensemble.



516  
 517 **Figure 8:** Mean ash concentration values  $>0.2 \text{ mg}/\text{m}^3$  for *ENS08* members (cyan region) for (a) FL000–200, (b) FL200–350  
 518 and (c) FL350–550 at 0000 UTC 24/06/2019. Each panel shows also the associated  $0.2 \text{ mg}/\text{m}^3$  ash concentration contour for  
 519 both the control run (black contour) and *ENS01* (red contour).

## 520 5.1 Application to aviation operations

521 The use of probability of either exceedance maps (Fig. 7) or concentration maps (Fig. 78) condense the information  
 522 given by the ensemble of VATDM simulations. However, multiple charts are still needed to cover all the relevant  
 523 information, such as different flight levels, times and ash concentration thresholds. To reduce information overload  
 524 from these numerous charts, which could impede fast decision-making during emergency response, they can be  
 525 condensed further into a single chart using a risk matrix (Prata et al., 2019). Here we apply this risk-based approach  
 526 to international flight routes in the vicinity of Raikoke using both the prior and posterior ensembles outlined in Sect.4.

### 527 5.1.1 Flight routes and dosage risk

528 To simulate potential aircraft encounters with volcanic ash, flight routes with minimised travel time were generated  
 529 by solving a time-optimal control problem as described in Wells et al. (2021). Trans-Pacific flights were generated  
 530 assuming a constant true airspeed of  $240 \text{ m/s}$  ( $\sim 864 \text{ km/hr}$ ) at a cruise altitude of FL380 (or 200hPa) and using  
 531 horizontal wind speeds extracted from the National Center for Atmospheric Research re-analysis data (Kalnay et al.,  
 532 1996). Eastbound and westbound time-optimal routes from Sapporo (CTS) to Honolulu (HNL), Los Angeles (LAX)  
 533 to Seoul (ICN) and San Francisco (SFO) to Shanghai (PVG) international airports were calculated for each day of the  
 534 dispersion model output (i.e., at 24 hr intervals).

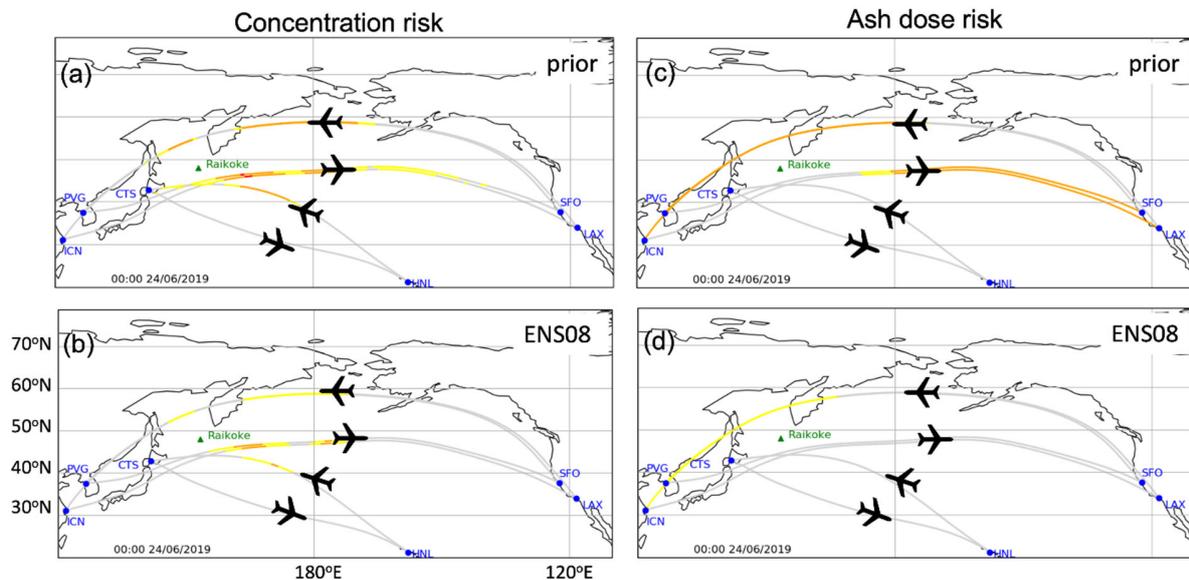
535 As in Prata et al. (2019), the along-flight ash dose,  $D$ , was defined as the ash concentration multiplied by the duration  
536 in that concentration (duration of exposure), integrated along an aircraft's flight path at cruise altitude (assumed to be  
537 FL350–550). This definition means that dose always increases monotonically along the route. All dose calculations  
538 assume that the modelled ash concentration fields at a given time step are fixed (i.e., do not change with time) as the  
539 aircraft flies from the origin to destination at its true airspeed (Prata et al. 2019).

### 540 **5.1.2 Risk based approach**

541 The first step in determining the risk is to calculate the fraction of ensemble members that have concentrations above  
542 specified impact thresholds for each of the three flight levels. In line with the current International Civil Aviation  
543 Organisation (ICAO) guidance, the impact concentration thresholds used are  $0.2\text{--}2\text{ mg m}^{-3}$ ,  $2\text{--}4\text{ mg m}^{-3}$  and greater  
544 than  $4\text{ mg m}^{-3}$ , for low, medium and high impact respectively. The risk of encountering ash is then determined by  
545 combining the likelihood ranges (less likely, 0–10 %; likely, 10–90 %; very likely, 90–100 %) and the impact. The  
546 risk of flying in a specific location and at a flight level is then assigned to be low, medium or high. The overall risk  
547 presented is the maximum risk over the three flight levels. In Prata et al. (2019), each risk level has a set of actions  
548 that may be implemented by the decision maker. These range from checking updated ash forecasts to considering  
549 alternative routes and scheduling extra maintenance.

550 The risk can be visualised as a 2D map or projected on to flight tracks of interest (Fig. 9). Considering risk based on  
551 ash concentrations at 0000 UTC on 24 June 2019, there are large portions of the flight tracks that encounter low and  
552 mid-level risk, with a small region of high risk to the east of Raikoke when using the prior ensemble (Fig. 9a). Thus,  
553 based on the prior ensemble output, flight operations could be expected to be severely disrupted at this time. However,  
554 determining risk from the posterior ensemble (ENS08) removes the region of highest risk and, overall, the amount of  
555 flight track potentially impacted and therefore requiring action from the flight operator is greatly reduced (Fig. 9b).

556 To account for the overall exposure of the aircraft to ash, the risk approach can also be applied to dose along a flight  
557 track (Fig. 9). To do this the ash dose impact thresholds used are  $4.4\text{--}14.4\text{ g s m}^{-3}$ ,  $14.4\text{--}28.8\text{ g s m}^{-3}$  and greater than  
558  $28.8\text{ g s m}^{-3}$  (Clarkson and Simpson, 2017; Prata et al., 2019). The likelihood ranges used are the same as those used  
559 for the concentration approach. In this scenario the risk is only determined at cruising altitude, which is assumed to  
560 be at FL350–550 (not the maximum over all flight levels). For the prior ensemble, the flight tracks to and from the  
561 West Coast of North America encounter mid-level risk and could potentially require specific actions by the airlines  
562 (e.g., more fuel and engine checks). Using this metric, flights between Honolulu and Sapporo do not reach doses that  
563 reach the lowest level of risk (Fig. 9c). For the posterior ensemble (ENS08), only the route from SFO to PVG and  
564 ICN reach sufficiently high doses to be highlighted by the risk approach. The other routes have very few ensemble  
565 members where doses are above  $4.4\text{ g s m}^{-3}$  and therefore are not highlighted by the risk approach (Fig. 9d). This could  
566 greatly reduce the need for an operator to implement any mitigation strategies.



567  
 568 **Figure 9:** Concentration risk along time optimal flight routes at 0000 UTC on 24 June 2019 for (a) prior ensemble (b)  
 569 *ENS08*. Ash dose risk along the same time optimal flight routes at 0000 UTC on 24 June 2019 for (c) prior ensemble (d)  
 570 *ENS08*. Yellow shading indicates the lowest level of risk, orange shading indicates mid-level risk and red indicates the  
 571 highest level of risk. Note that dose risk only considers risk at FL350–550, whereas concentration risk considers at all levels.

## 572 6 Conclusions

573 This study presents a new particle filtering data assimilation [system](#) that combines VATDM simulations with satellite  
 574 retrievals including their uncertainty estimates, to improve forecasts of volcanic ash cloud location and concentration.  
 575 A prior ensemble is created by simultaneously varying nine parameters representing the meteorology, eruption source  
 576 and internal parameters. Members from the prior ensemble are retained or discarded based on their level of agreement  
 577 with the satellite retrievals. The retained simulations are then used to create a posterior ensemble. Each posterior  
 578 ensemble is verified and filtered using satellite data at subsequent verification times.

579 The ESPs ranges in the constrained posterior ensembles are both smaller and skewed towards lower values than those  
 580 used in the control run and prior ensemble, [with the exception of ash density and eruption duration](#). Therefore, a single  
 581 ensemble designed with unconstrained parameters ranges (i.e., the prior ensemble) seems insufficient for estimating  
 582 ESPs ranges that may approximate more accurately the observed volcanic ash cloud. This is not the case for internal  
 583 model parameters which remain unconstrained by the data assimilation.

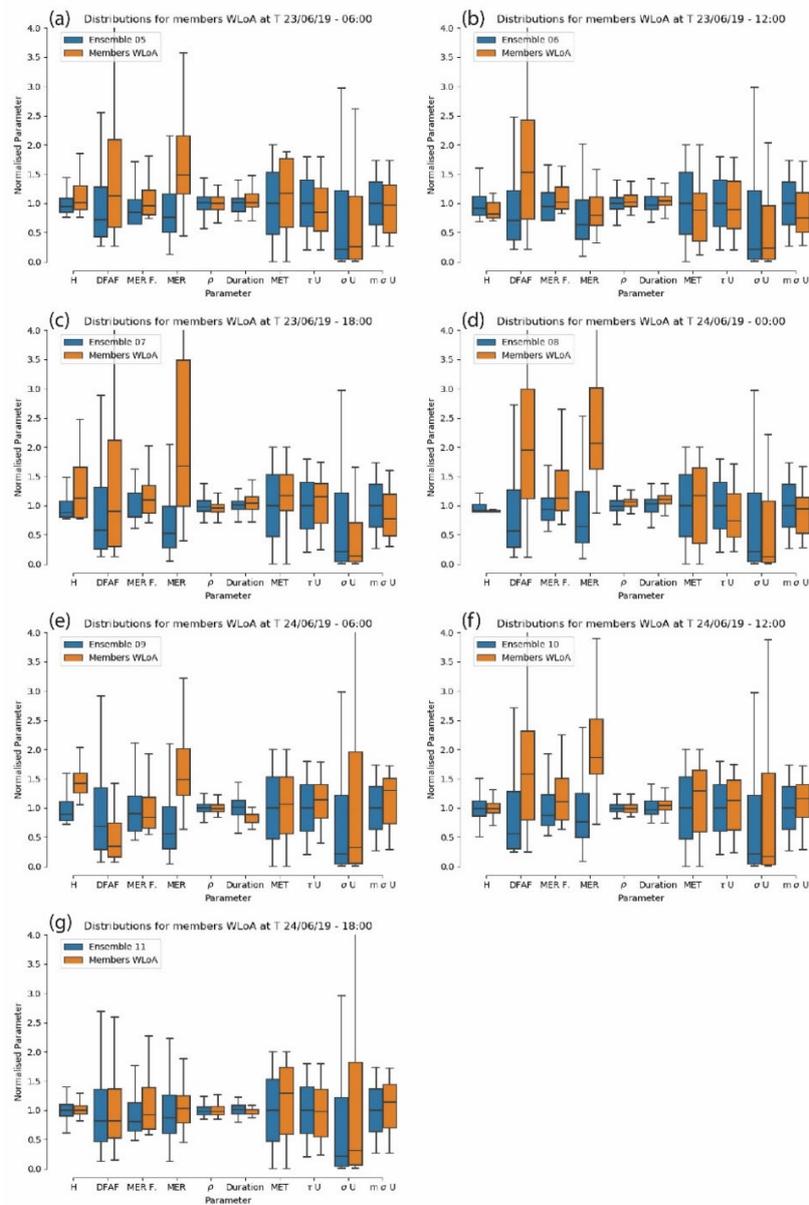
584 Communicating the risk of volcanic ash to aviation using risk maps and risk trajectories shows that the prior ensemble  
 585 forecasts mid-level and highest risk for both ash concentration and dose thresholds for much of a set of representative  
 586 flight tracks. Based on this information flight operations could be severely disrupted by the eruption. However, using  
 587 the constrained posterior ensemble, the region of highest risk is removed, and the mid-level risk is reduced. Thus,

588 using the refined posterior ensembles potentially reduces the need for the operator to implement any mitigation  
589 strategies and hence reduces disruption to airline operations.

590 This methodology is easily generalisable to other VATDMs and could be used to run a comparison with other models.  
591 Different remote sensing datasets could be used to assess its sensitivity to the observations used. Running multiple  
592 1000–members ensembles requires either a high computational power or can be subject to variable queuing times on  
593 computer clusters such as JASMIN. Future work could include code optimization, to make runtime and ensemble size  
594 more efficient, potentially allowing an operational application. With a more manageable ensemble size, it could also  
595 be possible to introduce temporal variation of perturbed parameters, such as plume height, within each simulation.  
596 Furthermore, the evaluation method is based on limits of acceptability; a future improvement would be to define the  
597 posterior by weighting each simulation output according to some measure of its fit to the observations, in a way that  
598 takes proper account of the epistemic uncertainties in the satellite retrievals or any other available information.

599

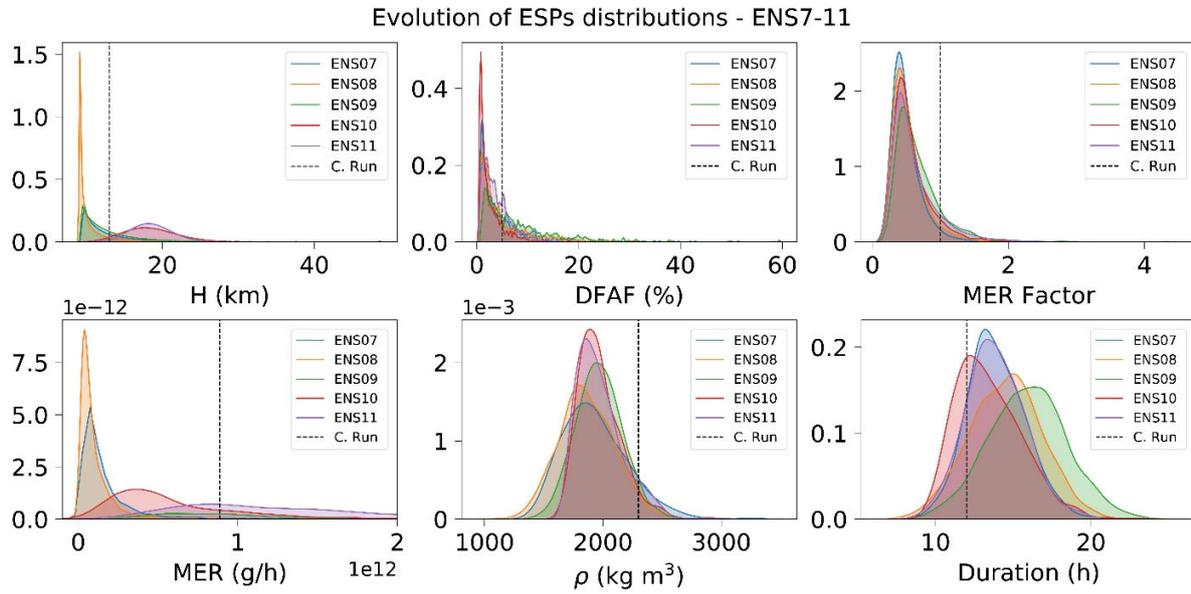
600



602  
 603 **Fig. A1: Evolution of parameters distributions for members Within Limits of Acceptability (WLoA) for (a)**  
 604 **Ensemble 05, (b) Ensemble 06, (c) Ensemble 07, (d) Ensemble 08, (e) Ensemble 09, (f) Ensemble 10, and (g)**  
 605 **Ensemble 11.  $\rho$ ,  $\tau U$ ,  $\sigma U$  and  $m \sigma U$  represent density, horizontal Lagrangian timescale for free tropospheric**  
 606 **turbulence, standard deviation of horizontal velocity for free tropospheric turbulence, and Standard deviation**  
 607 **( $\sigma$ ) of horizontal velocity for unresolved mesoscale motions. In all panels, Mass eruption rate (MER) is plotted**  
 608 **to show its variation, although it is not explicitly perturbed in any of the ensembles. Each parameter in the**  
 609 **boxplots is normalized by dividing each individual value from the ensemble members by the mean of that entire**  
 610 **parameter range from the selected ensemble.**

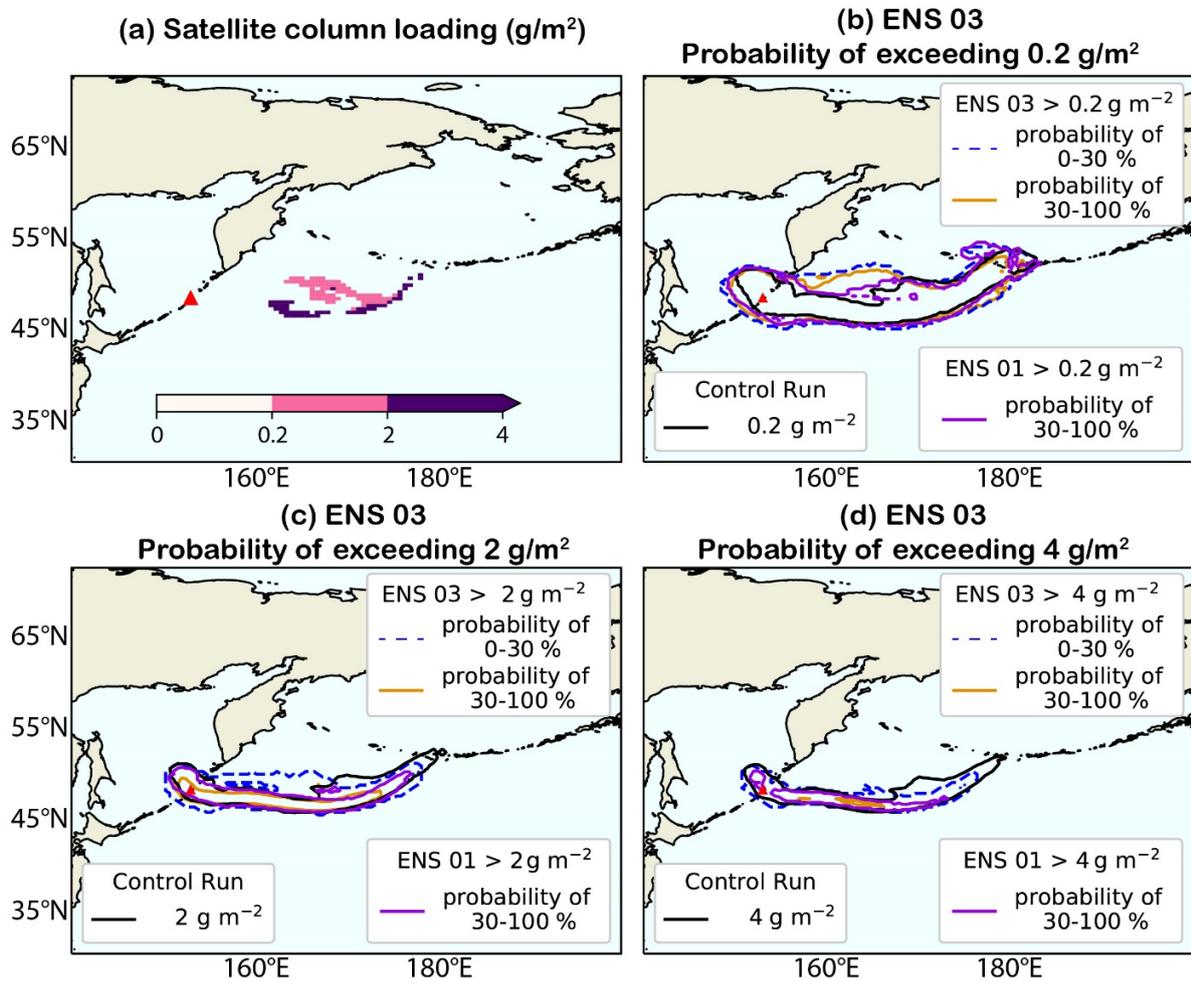
611 **Appendix B**

612



613

614 **Fig. B1: Evolution of ESPs distributions for ENS07 (blue shading), ENS08 (orange shading), ENS09 (green shading), ENS10 (red shading), and ENS11 (purple shading). Each panel shows: (a) plume height (H), (b) distal**  
615 **fine ash fraction (DFAF), (c) mass eruption rate factor (MERF), (d) mass eruption rate (MER), (e) ash density ( $\rho$ ), and (f) eruption duration. For ENS07–11, each parameter is sampled from posterior pdfs (see main text**  
616 **for a detailed description of the resampling method). Mass eruption rate (MER) is show for all the ensembles, but it is not explicitly perturbed. Dashed black lines in each plot represent the parameter value used for the**  
617 **control run (see Table 1 in the main manuscript).**  
618  
619  
620  
621



623  
 624 **Fig. C1: Same probability maps as in Fig. 7 in the main manuscript, with the addition of the 0-30% probability region for**  
 625 **ENS03 to exceed mass loading of (b)  $0.2 \text{ g m}^{-2}$ , (c)  $2 \text{ g m}^{-2}$ , and (d)  $4 \text{ g m}^{-2}$  (dashed blue contours; the region within the blue**  
 626 **line is a probability region up to 30 %).**

627 **Data availability**

628 NAME simulation output for the prior ensemble (ENS01) and posterior ensembles 03 (ENS03) and 08 (ENS08) are  
 629 available in the Lancaster University Research Directory at <https://doi.org/10.17635/lancaster/researchdata/491>. Raw  
 630 data for all the posterior ensembles can be provided by the authors upon request.

631 **Authors contributions**

632 Conceptualization, A.C, H.F.D., N.J.H., K.B., M.R.J.; methodology, A.C.; software, A.C. and C.W.; validation, A.C.;  
 633 investigation, A.C. and N.J.H.; resources, C.S.; writing – original draft preparation, A.C., N.J.H., H.F.D.; writing—

634 review and editing, A.C., N.J.H., H.F.D., K.B., C.S., M.R.J.; funding acquisition, M.R.J., H.F.D., K.B.; [A.C., H.F.D.,](#)  
635 [N.J.H., K.B., C.S., M.R.J](#) have read and agreed to the version of the manuscript.

#### 636 **Competing interests**

637 The authors declare that they have no conflict of interest.

#### 638 **Acknowledgments**

639 This research was funded by the Natural Environment Research Council (NERC) under the project *Radar-supported*  
640 *Next-Generation Forecasting of Volcanic Ash Hazard (R4AsH)*, NE/S005218/1. We thank Andrew Prata (University  
641 of Oxford) for providing a copy of the Python code for designing an ensemble of NAME simulations and guidance  
642 for its modification, and Ben Evans (Met Office) for providing the MOGREPS-G forecasts. We also thank Helen  
643 Webster (Met Office) for discussions and her constructive comments on the manuscript before submission. [We thank](#)  
644 [two anonymous reviewers for their positive comments, which helped improve the manuscript.](#)

#### 645 **References**

646 Beckett, F. M., Witham, C. S., Leadbetter, S. J., Crocker, R., Webster, H. N., Hort, M. C., Jones, A. R., Devenish, B.  
647 J., and Thomson, D. J.: Atmospheric Dispersion Modelling at the London VAAC: A Review of Developments since  
648 the 2010 Eyjafjallajökull Volcano Ash Cloud, *Atmosphere*, 11, 352, 2020

649 Bessho, K., Date, K., Hayashi, M., Ikeda, A., Imai, T., Inoue, H., Kumagai, Y., Miyakawa, T., Murata, H., Ohno, T.,  
650 Okuyama, A., Oyama, R., Sasaki, Y., Shimazu, Y., Shimoji, K., Sumida, Y., Suzuki, M., Taniguchi, H., Tsuchiyama,  
651 H., Uesawa, D., Yokota, H., and Yoshida, R.: An Introduction to Himawari-8/9; Japan's New-Generation  
652 Geostationary Meteorological Satellites, *Journal of the Meteorological Society of Japan. Ser. II*, 94, 151–183,  
653 <https://doi.org/10.2151/jmsj.2016-009>, 2016.

654 Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction,  
655 *Hydrological Processes*, 6, 279–298, <https://doi.org/https://doi.org/10.1002/hyp.3360060305>, 1992.

656 Bowler, N. E., Arribas, A., Mylne, K. R., Robertson, K. B., and Beare, S. E.: The MOGREPS short-range ensemble  
657 prediction system, *Quarterly Journal of the Royal Meteorological Society*, 134, 703–722,  
658 <https://doi.org/https://doi.org/10.1002/qj.234>, 2008.

659 Casadevall, T. J.: The 1989–1990 eruption of Redoubt Volcano, Alaska: impacts on aircraft operations, *Journal of*  
660 *volcanology and geothermal research*, 62, 301–316, 1994.

661 Chai, T., Crawford, A., Stunder, B., Pavolonis, M. J., Draxler, R., and Stein, A.: Improving volcanic ash predictions  
662 with the HYSPLIT dispersion model by assimilating MODIS satellite retrievals, *Atmospheric Chemistry and Physics*,  
663 17, 2865–2879, <https://doi.org/10.5194/acp-2517-2865-2017>, 2017.

664 Clarkson, R. J., Majewicz, E. J., and Mack, P.: A re-evaluation of the 2010 quantitative understanding of the effects  
665 volcanic ash has on gas turbine engines, *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of*  
666 *Aerospace Engineering*, 230, 2274–2291, 2016.

667 Clarkson, R.J. and Simpson, H.: Maximising airspace use during volcanic eruptions: matching engine durability  
668 against ash cloud occurrence. Science and Technology Organization (STO) - Meeting Proceedings Paper MP-AVT-  
669 272-17, 15-17 May 2017, Vilnius, Lithuania, pp. 1–20, 2017.

670 Dacre, H.F., Grant, A.L., Hogan, R.J., Belcher, S.E., Thomson, D.J., Devenish, B.J., Marengo, F., Hort, M.C.,  
671 Haywood, J.M., Ansmann, A. and Mattis, I., 2011. Evaluating the structure and magnitude of the ash plume during  
672 the initial phase of the 2010 Eyjafjallajökull eruption using lidar observations and NAME simulations. *Journal of*  
673 *Geophysical Research: Atmospheres*, 116(D20).

674 Dacre, H. F., Harvey, N. J., Webley, P. W., and Morton, D.: How accurate are volcanic ash simulations of the 2010  
675 Eyjafjallajökull eruption?, *Journal of Geophysical Research: Atmospheres*, 121, 3534–3547,  
676 <https://doi.org/https://doi.org/10.1002/2015JD024265>, 2016.

677 Denlinger, R. P., Pavolonis, M., and Sieglaff, J.: A robust method to forecast volcanic ash clouds, *Journal of*  
678 *Geophysical Research*, 117, <https://doi.org/10.1029/2012JD017732>, 2012.

679 Flowerdew, J. and Bowler, N. E.: Improving the use of observations to calibrate ensemble spread, *Quarterly Journal*  
680 *of the Royal Meteorological Society*, 137, 467–482, <https://doi.org/https://doi.org/10.1002/qj.744>, 2011.

681 Flowerdew, J. and Bowler, N. E.: On-line calibration of the vertical distribution of ensemble spread, *Quarterly Journal*  
682 *of the Royal Meteorological Society*, 139, 1863–1874, <https://doi.org/https://doi.org/10.1002/qj.2072>,  
683 <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2072>, 2013.

684 Francis, P. N., Cooke, M. C., and Saunders, R. W.: Retrieval of physical properties of volcanic ash using Meteosat: A  
685 case study from the 2010 Eyjafjallajökull eruption, *Journal of Geophysical Research*, 117,  
686 <https://doi.org/10.1029/2011JD016788>, 2012.

687 Fu, G., Lin, H., Heemink, A., Segers, A., Lu, S., and Palsson, T.: Assimilating aircraft-based measurements to improve  
688 forecast accuracy of volcanic ash transport, *Atmospheric Environment*, 115, 170–184,  
689 <https://doi.org/https://doi.org/10.1016/j.atmosenv.2015.05.061>, 2015.

690 Fu, G., Prata, F., Lin, H. X., Heemink, A., Segers, A., and Lu, S.: Data assimilation for volcanic ash plumes using a  
691 satellite observational operator: a case study on the 2010 Eyjafjallajökull volcanic eruption, *Atmospheric Chemistry*  
692 *and Physics*, 17, 1187–1205, <https://doi.org/10.5194/acp-17-1187-2017>,  
693 <https://acp.copernicus.org/articles/17/1187/2017/>, 2017.

694 Global Volcanism Program: Report on Raikoke (Russia) (Crafford, A.E., and Venzke, E., eds., Bulletin of the Global  
695 Volcanism Network,44:8. Smithsonian Institution., <https://doi.org/10.5479/si.GVP.BGVN201908-290250>, 2019.

696 Grant, A.L., Dacre, H.F., Thomson, D.J. and Marengo, F., 2012. Horizontal and vertical structure of the  
697 Eyjafjallajökull ash cloud over the UK: a comparison of airborne lidar observations and simulations. *Atmospheric  
698 Chemistry and Physics*, 12(21), pp.10145-10159.

699 Harvey, N. J. and Dacre, H. F.: Spatial evaluation of volcanic ash forecasts using satellite observations, *Atmospheric  
700 Chemistry and Physics*, 16, 861–872, <https://doi.org/10.5194/acp-16-861-2016>,  
701 <https://acp.copernicus.org/articles/16/861/2016/>, 2016.

702 Harvey, N. J., Huntley, N., Dacre, H. F., Goldstein, M., Thomson, D., and Webster, H.: Multi-level emulation of a  
703 volcanic ash transport and dispersion model to quantify sensitivity to uncertain parameters., *Natural hazards and earth  
704 system sciences.*, 18, 41–63, 2018.

705 Harvey, N. J., Dacre, H. F., Webster, H. N., Taylor, I. A., Khanal, S., Grainger, R. G., and Cooke, M. C.: The Impact  
706 of Ensemble Meteorology on Inverse Modeling Estimates of Volcano Emissions and Ash Dispersion Forecasts:  
707 Grímsvötn 2011, *Atmosphere*, 11, 1022,55<https://doi.org/10.3390/atmos11101022>,  
708 <http://dx.doi.org/10.3390/atmos11101022>, 2020.

709 Hedelt, P., Efremenko, D. S., Loyola, D. G., Spurr, R., and Clarisse, L.: Sulfur dioxide layer height retrieval from  
710 Sentinel-5 Pre-cursor/TROPOMI using FP\_ILM, *Atmospheric Measurement Techniques*, 12, 5503–5517,  
711 <https://doi.org/10.5194/amt-12-5503-2019>,<https://amt.copernicus.org/articles/12/5503/2019/>, 2019.

712 Hobbs, P. V., Radke, L. F., Lyons, J. H., Ferek, R. J., Coffman, D. J., and Casadevall, T. J.: Airborne measurements  
713 of particle and gas emissions from the 1990 volcanic eruptions of Mount Redoubt, *Journal of Geophysical Research*,  
714 96, 18 735–18 752, 1991.

715 Jolliffe, I. T. and Stephenson, D. B.: *Forecast Verification A Practitioner’s Guide in Atmosphere Science*, Wiley-  
716 Blackwell, Chichester, 2012.

717 Jones, A., Thomson, D., Hort, M., and Devenish, B.: The UK Met Office’s next-generation atmospheric dispersion  
718 model, NAME III, in: *Air Pollution Modeling and its Application XVII*, pp. 580–589, Springer, 2007.

719 Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen,  
720 J., et al.: The NCEP/NCAR 40-year reanalysis project, *Bulletin of the American meteorological Society*, 77, 437–472,  
721 1996.

722 Kristiansen, N. I., Stohl, A., Prata, A. J., Bukowiecki, N., Dacre, H., Eckhardt, S., Henne, S., Hort, M. C., Johnson, B.  
723 T., Marengo, F.,Neininger, B., Reitebuch, O., Seibert, P., Thomson, D. J., Webster, H. N., and Weinzierl, B.:  
724 Performance assessment of a volcanic ash transport model mini-ensemble used for inverse modeling of the 2010  
725 Eyjafjallajökull eruption, *Journal of Geophysical Research*, 117,2012.

726 Krotkov, N. A., Flittner, D., Krueger, A., Kostinski, A., Riley, C., Rose, W., and Torres, O.: Effect of particle non-  
727 sphericity on satellite monitoring of drifting volcanic ash clouds, *Journal of Quantitative Spectroscopy and Radiative*  
728 *Transfer*, 63, 613–630, 1999.

729 Lawrence, B. N., Bennett, V., Churchill, J., Juckes, M., Kershaw, P., Oliver, P., Pritchard, M., and Stephens, A.: The  
730 JASMIN super-data-cluster, arXiv preprint arXiv:1204.3553, 2012.

731 Marti, A. and Folch, A.: Volcanic ash modeling with the NMMB-MONARCH-ASH model: quantification of offline  
732 modeling errors, *Atmospheric Chemistry and Physics*, 18, 4019–4038, 2018.

733 [Met Office: Iris: A Python package for analysing and visualising meteorological and oceanographic data sets, v3.1,](http://scitools.org.uk/)  
734 [2021, http://scitools.org.uk/.](http://scitools.org.uk/)

735 [Mingari, L., Folch, A., Prata, A. T., Pardini, F., Macedonio, G., and Costa, A.: Data assimilation of volcanic aerosol](https://doi.org/10.5194/acp-22-1773-2022)  
736 [observations using FALL3D+PDAF, \*Atmospheric Chemistry and Physics\*, 22, 1773–1792,](https://doi.org/10.5194/acp-22-1773-2022)  
737 [https://doi.org/10.5194/acp-22-1773-2022, 2022.](https://doi.org/10.5194/acp-22-1773-2022)

738 Osoreo, S., Ruiz, J., Folch, A., and Collini, E.: Volcanic ash forecast using ensemble-based data assimilation: an  
739 ensemble transform Kalman filter coupled with the FALL3D-7.2 model (ETKF–FALL3D version 1.0), *Geoscientific*  
740 *Model Development*, 13, 1–22, 2020.

741 Pardini, F., Corradini, S., Costa, A., Esposti Ongaro, T., Merucci, L., Neri, A., Stelitano, D., et al.: Ensemble-Based  
742 Data Assimilation of Volcanic Ash Clouds from Satellite Observations: Application to the 24 December 2018 Mt.  
743 Etna Explosive Eruption, *Atmosphere*, 11,80359, 2020.

744 Pavolonis, M. J. (2010): Advances in extracting cloud composition information from spaceborne infrared radiances:  
745 A robust alternative to brightness temperatures. Part I: Theory, *J. Appl. Meteorol. Climatol.*, 49, 1992–2012,  
746 doi:10.1175/2010JAMC2433.1.

747 Pelley, R. E., Cooke, M. C., Manning, A. J., Thomson, D. J., Witham, C. S., and Hort, M. C.: Initial implementation  
748 of an inversion techniques for estimating volcanic ash source parameters in near real time using satellite retrievals,  
749 *Forecasting Research Technical Report No. 644*, [https://www.metoffice.gov.uk/research/library-and-](https://www.metoffice.gov.uk/research/library-and-archive/publications/science/weather-science-technical-reports)  
750 [archive/publications/science/weather-science-technical-reports](https://www.metoffice.gov.uk/research/library-and-archive/publications/science/weather-science-technical-reports), 2015.

751 Prata, A. and Prata, A.: Eyjafjallajökull volcanic ash concentrations determined using Spin Enhanced Visible and  
752 Infrared Imager measurements, *Journal of Geophysical Research: Atmospheres* (1984–2012), 117, 2012.

753 Stefanescu, E., Patra, A., Bursik, M., Madankan, R., Pouget, S., Jones, M., Singla, P., Singh, T., Pitman, E., Pavolonis,  
754 M., et al.: Temporal, probabilistic mapping of ash clouds using wind field stochastic variability and uncertain eruption  
755 source parameters: Example of the 14 April 2010 Eyjafjallajökull eruption, *Journal of Advances in Modeling Earth*  
756 *Systems*, 2014.

757 Stohl, A., Prata, A., Eckhardt, S., Clarisse, L., Durant, A., Henne, S., Kristiansen, N., Minikin, A., Schumann, U.,  
758 Seibert, P., et al.: Determination of time-and height-resolved volcanic ash emissions and their use for quantitative ash  
759 dispersion modeling: the 2010 Eyjafjallajökull eruption, *Atmos. Chem. Phys.*, 11, 4333–4351, 2011.

760 Wang, R., Chen, B., Qiu, S., Zhu, Z., and Qiu, X.: Data assimilation in air contaminant dispersion using a particle  
761 filter and expectation-maximization algorithm, *Atmosphere*, 8, 170, 2017.

762 Wells, C. A., Williams, P. D., Nichols, N. K., Kalise, D., and Poll, I.: Reducing transatlantic flight emissions by fuel-  
763 optimised routing, *Environmental Research Letters*, 16, 025 002, 2021.

764 Wilkins, K., Mackie, S., Watson, M., Webster, H., Thomson, D., and Dacre, H.: Data insertion in volcanic ash cloud  
765 forecasting, *Annals of Geophysics*, 57, 2015.

766 Witham, C., Hort, M., Thomson, D., Leadbetter, S., Devenish, B., Webster, H., Beckett, F., and Kristiansen, N.: The  
767 current volcanic ash modelling setup at the London VAAC,  
768 [https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/services/transport/aviation/vaac/london\\_v](https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/services/transport/aviation/vaac/london_vaac_current_modelling_setup.pdf)  
769 [aac\\_current\\_modelling\\_setup.pdf](https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/services/transport/aviation/vaac/london_vaac_current_modelling_setup.pdf), accessed: 2021-06-23, 2019.

770 Woodhouse, M. J., Hogg, A. J., Phillips, J. C., and Sparks, R. S. J.: Interaction between volcanic plumes and wind  
771 during the 2010 Eyjafjallajökull eruption, Iceland, *Journal of Geophysical Research*, 118, 92–109,  
772 <https://doi.org/10.1029/2012JB009592>, <http://dx.doi.org/10.1029/2012JB009592>, 2013.

773 Zidikheri, M. J., Lucas, C., and Potts, R. J.: Toward quantitative forecasts of volcanic ash dispersal: Using satellite  
774 retrievals for optimal estimation of source terms, *Journal of Geophysical Research: Atmospheres*, 122, 8187–8206,  
775 2017.

776 Zidikheri, M. J. and Lucas, C.: A computationally efficient ensemble filtering scheme for quantitative volcanic ash  
777 forecasts. *Journal of Geophysical Research: Atmospheres*, 126, e2020JD033094, [10.1029/2020JD033094](https://doi.org/10.1029/2020JD033094), 2021.