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

3 Antonio Capponi[1], Natalie J. Harvey[2], Helen F. Dacre[2], Keith Beven[1], Cameron Saint[3], Cathie A.
4 Wells[2], and Mike R. James[1]

5 [1] Lancaster Environment Centre, LEC Building, Lancaster, LA1 4AQ, UK
6 [2] Department of Meteorology, University of Reading, Earley Gate, Reading, RG6 6ET, UK
7 [3] Met Office, Fitzroy Road, Exeter EX1 3PB, UK

8 *Correspondence to*: Antonio Capponi (a.capponi2@lancaster.ac.uk)

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

28 ## 1 Introduction

29 Volcanic ash in the atmosphere poses a hazard to aircraft (Casadevall, 1994). It is therefore important to accurately
30 forecast the evolution of volcanic ash cloud in the atmosphere for the aviation industry. Forecasting the distribution
31 of volcanic ash in the atmosphere at a given time is typically performed using a volcanic ash transport and dispersion
32 model (VATDM). VATDMs solve dynamic equations to evolve the system state (volcanic ash cloud) forward in time.
33 However, such simulated volcanic ash distributions are subject to errors due to inaccurate parametrisations of physical
34 processes, errors in the driving meteorological fields and errors in the volcanic eruption source parameters.
35 Observations of volcanic ash distributions may be obtained from ground-based, aircraft or satellite-based instruments.

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

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

62   In most cases eruption source parameters (input parameters), physical processes (internal parameters) and the driving
63   meteorology are uncertain so an ensemble of VATDM simulations can be formed by perturbing the input,
64   meteorological and internal parameters. This results in a probability density function (pdf) of simulated volcanic ash
65   distributions. In this case, the data assimilation step involves conditioning the VATDM simulated pdf based on a
66   comparison with the observed volcanic ash cloud to create a filtered pdf at each time the data assimilation is performed.
67   In volcanic ash forecasting, ensemble source inversion (Harvey et al., 2020) and ensemble sequential filtering
68   methods, such as Ensemble Kalman Filters (EnKFs) (Fu et al., 2015; Pardini et al., 2020; Osores et al., 2020) and
69   Particle Filters (Wang et al., 2017; Zidikheri and Lucas, 2021), have been employed. EnKFs were developed for non-
70   linear systems and so are suitable for dispersion problems. However, they assume that the parameters to be estimated
71   have Gaussian prior pdfs, which may not be true. Conversely, no assumptions on the form of the prior pdf of simulator

72   states are needed for Particle Filtering techniques meaning that they are more sensitive to the tails of the prior

73   distribution, although this is at the cost of a lot more simulations.

74   Bayesian inference is used in particle filtering to derive constraints on the parameters of a simulation from

75   observations. In this framework, the posterior pdf of the simulation parameters given the observed data is derived

76   from a prior pdf and from the likelihood of the data given a choice of simulator parameters. Bayesian inference

77   therefore relies on the ability to compute a formal likelihood function. For volcanic eruption source parameters their

78   exact likelihood function is unknown or computationally intractable and so direct Bayesian analysis is therefore not

79   possible. A technique known as Approximate Bayesian Computation uses simulations to bypass the need to evaluate

80   a likelihood function. Approximate Bayesian Computation systematically explores the prior parameter space and

81   compares the simulated and observed data sets using a distance metric. By accepting simulations for which this

82   distance metric is smaller than a given threshold, the method provides an approximation to the Bayesian posterior pdf.

83   One Approximate Bayesian Computation method frequently used in hydrology forecasting is known as Generalised

84   Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 1992). The GLUE methodology is based on the

85   concept of equifinality, which acknowledges that there exist many combinations of simulation input and internal

86   parameters that provide equally good simulations of the observed system.

87   There are several steps in the GLUE methodology:

88   1.   Realistic ranges are defined for the simulator input and internal parameters. These are known as prior pdfs

89   since they are defined prior to the comparison with observational data. When there is a lack of strong prior

90   information about the parameter distributions and their interactions uniform pdfs are often used.

91   2.   Rejection criteria are defined to determine the accepted agreement between the simulators and the observed

92   system state. These can be based on subjectively chosen thresholds limits or as accepted minimum levels of

93   performance allowing for the expected uncertainties in the observational data.

94   3.   Input and internal parameter sets are sampled from the prior pdfs to generate an ensemble of simulation

95   predictions of the system for a given analysis time.

96   4.   Simulations that are not in agreement with the observed system, using the selected rejection criteria, are

97   discarded from the analysis. The subset of retained simulators with known parameter sets forms the posterior

98   pdfs for each input and internal parameter. Thus, the posterior pdfs are the conditional pdfs of each parameter

99   given the observations.

100   5.   Input and internal parameter values are then sampled from the posterior pdf to generate an ensemble

101   prediction of the system at a future time.

102   6.   Steps 4–5 can be repeated for subsequent analysis times and the joint posterior distributions compared to the

103   preceding analysis time.

104   The main aim of this paper is to contribute to the development of data assimilation methods to improve quantitative

105   ash dispersion forecasts. To this end we will determine whether satellite retrievals of volcanic ash column loading can

106   be used to filter an ensemble of volcanic ash simulations using the particle filtering methodology. We determine which

107   of the input and internal model parameters are most constrained by the satellite observations and quantify how the

108   assimilation of satellite data changes the uncertainty estimate of the ensemble. Finally, for communicating the volcanic

109   ash forecasts, we apply to the ensemble output the risk-matrix approach described in Prata et al. (2019) and applied

110   retrospectively to the 2011 Grimsvotn eruption by Harvey et al. (2020), where risk is defined as the likelihood of

111   exceeding ash concentrations considered a potential risk to aircraft. This approach will be demonstrated using

112   simulations and observations of the Raikoke volcanic eruption which was a short duration eruption lasting less than

113   24 hours occurring between the 21–22 June 2019.

2 Methods and data

114

115   In this study, the Numerical Atmospheric–dispersion Modelling Environment, NAME (Jones et al., 2007), was used

116   to simulate the dispersion of volcanic ash. It is the VATDM used by the London Volcanic Ash Advisory Centre

117   (LVAAC) for producing volcanic ash advisories following an eruption. Each simulated ash cloud was quantitively

118   evaluated using retrievals from Himawari–8.

2.1 Himawari–8

119

120   Himawari–8 is a geostationary satellite that came into operation in July 2015 (Bessho et al., 2016). It has 16 spectral

121   channels and provides observations of high temporal frequency (10 min) and spatial resolution (2 km for the infrared

122   bands). The high temporal and spatial resolution make these observations ideally suited to evaluate the transport of

123   volcanic ash following an eruption. The volcanic ash retrievals used in this study are based on the method by Francis

124   et al. (2012) with slight adaptations for the channels of the AHI instrument aboard the Himawari–8 satellite.

125   The Met Office volcanic ash retrieval algorithm has two steps. The first step detects which pixels contain volcanic ash

126   using the channels at 8.6 μm, 10.4 μm and 12.4 μm. The second step runs a one–dimensional variational (1D–Var)

127   analysis to determine an optimal estimate between the assumed background and the observed radiances in the channels

128   at 10.4 μm, 12.4 μm, and 13.3 μm for the column loading, ash cloud height and effective radius. The detection is

129   based on a combination of brightness temperature difference (BTD) tests and beta ratio tests (Pavolonis, 2010). The

130   beta ratio tests use a derived radiative parameter β, that is the effective absorption optical depth ratio of two channels

131   and are used to filter pixels marked as ash by the BTD tests. These tests have been improved by fine tuning of the

132   operational thresholds, to optimise coverage of the June 2019 Raikoke eruption. In addition, several geographical

133   filters have been added, to reduce false detections at high satellite zenith angle and over arid land surfaces, and further

134   false detections have been removed by checking the consistency of ash detection in neighbouring pixels.

135   The retrieval algorithm also provides a measurement of the error on each of the retrieved values. The retrieval relies

136   on the minimization of a cost function to determine the optimal estimate from the assumed background and the

137   observed radiances. How well defined the minimum of the cost function is provides an indication of the likely accuracy

138   of the retrieval, in that the more well defined the minimum of the cost function is, the more accurate the retrieval is

139   likely to be. By considering the inverse of the second derivative of the cost function with respect to each of the

140 variables considered in the retrieval, we can provide an estimate of the error for the retrieved ash plume pressure, ash

141 column loading and ash effective radius.

142 Where ash is detected, these pixels are flagged as ash and this algorithm determines the ash column loading. If a pixel

143 is free from both ash and meteorological cloud, then it is flagged as a clear sky pixel. Pixels that neither have detectable

144 ash nor are flagged as clear skies are unclassified. As in Harvey et al. (2020), further processing is performed to regrid

145 the retrieved column loadings on to a grid of 0.375° latitude by 0.5625° longitude (approximately 40 km x 40 km in

146 mid–latitudes) and averaged over 1 h. This is to match the resolution of the VATDM ash concentration output and to

147 reduce data volumes. If all classified pixels within a grid box are flagged as clear sky pixels, then the grid box is

148 deemed to be a clear sky observation. Otherwise, the grid box is deemed to be an ash grid observation with the column

149 loading in this grid box given by the mean of all the classified pixels (including clear skies).

150 **2.2 VATDM**

151 All the simulations were performed using NAME version 8.1 on the Joint Analysis System Meeting Infrastructure

152 Needs (JASMIN) super data cluster (Lawrence et al., 2012). To simulate the dispersion and removal of volcanic ash,

153 NAME includes parametrization of the effects of turbulence on the transport and dispersion, sedimentation, dry

154 deposition and wet deposition. In the operational configuration used by the LVAAC (Beckett et al., 2020) aggregation

155 of ash particles, near source plume rise and processes driven by the eruption dynamics (e.g., Woodhouse et al., 2013)

156 are not explicitly modelled. The default particle size distribution used is based on data from Hobbs et al. (1991) and

157 the shape of the particles are assumed to be spherical.

158 Ensembles of NAME simulations were created by varying nine parameters covering the meteorology, information

159 about the eruption source and the parameterisation of turbulence in NAME (as in Harvey et al., 2018; Prata et al.,

160 2019). Uniform distributions between the specified ranges were used as prior probability distributions to generate the

161 initial ensemble (Table 1). Full details of how these parameters are sampled is given in Sect.4.1.

162 All simulations share the same start and end time, 18:00 UTC on 21 June 2019 and 1200 UTC on 25 June 2019

163 respectively, for a total run time of 96 hours. The eruption start time matches the simulation start time. Volcanic ash

164 within the simulations is released along a vertical line, between the lower and upper plume heights (Witham et al.,

165 2019). Ash column loadings ($g\,m^{-2}$) and ash concentrations ($g\,m^{-3}$) are output onto a global grid of 800×600 points,

166 corresponding to a grid of 0.45° longitude and 0.3° latitude, giving a horizontal resolution of approximately 40 km in

167 midlatitudes every 6 hours using a 6–hour time average. Ash concentrations are output at 22 Flight Levels (FL000–

168 FL550) with a vertical resolution of 25 FL ("thin layers") that are then combined to form three "thick" layers (FL000–

169 200, FL200–350, FL350–550) by taking the maximum concentration values from the component thin layers for the

170 corresponding thick layer value (Witham et al., 2019).

171

172

Parameter	Symbol	Control value	Initial sampling range
Plume height (km)[1]	H	12.45	9–17
Mass eruption rate factor[2]	$MER\,F$	1	0.33–3
Ash density (kg m^{-3})	ρ	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)	τ	300 (100)	100–900 (33.33–300)
Standard deviation of horizontal (vertical) velocity for free tropospheric turbulence (m s^{-1})	σ	0.25 (0.1)	0.0025–2.5 (0.001–1)
Standard deviation (σ) of horizontal velocity for unresolved mesoscale motions (m s^{-1})	$m\sigma U$	0.8	0.27–1.74
Meteorological fields	MET	Met Office Unified Model global analysis	MOGREPS–G members 0–17

173   **Table 1: Parameters sampled and their control and initial sampling ranges.**
174   [1]Plume rise height above the summit
175   [2]scaling factor applied to the default mass eruption rate value calculated using the equation from Mastin et al. (2009); see Sect.
176   4.1.1 for details on how mass eruption rate is calculated.

177   **3 Description of the 2019 Raikoke eruption**

178   Raikoke is an uninhabited volcanic Island near the centre of the Kuril Island chain in the Sea of Okhotsk in the
179   northwest Pacific Ocean located at 48.2°N, 153.3°E. Its most recent explosive eruption, after 95 years of dormancy,
180   started at approximately 1800 UTC on 21 June 2019 and is estimated to have an initial eruptive plume height of 10–
181   13 km above sea level (asl) (Global Volcanism Project, 2019). The eruption lasted approximately 12 hours and ended
182   at approximately 0600 UTC on 22 June 2019. There is evidence from visible satellite imagery to suggest that there
183   was an umbrella cloud which was quickly advected eastwards towards a large extratropical cyclone which distorted
184   the dispersed ash cloud (Fig. 1).

185   The number of satellite grid boxes that are classified as containing ash at each time are the ones available to be used
186   to refine the prior pdfs. The largest number of boxes are available at 1800 UTC on 22 June. Before 0600 UTC on 22
187   June the number of grid boxes available is limited by the small time the ash has had to be transported. After 1800 UTC
188   on 24 June the number of grid boxes is limited due to the presence of meteorological cloud associated with an
189   extratropical cyclone situated to the east of Raikoke (Fig. 1).

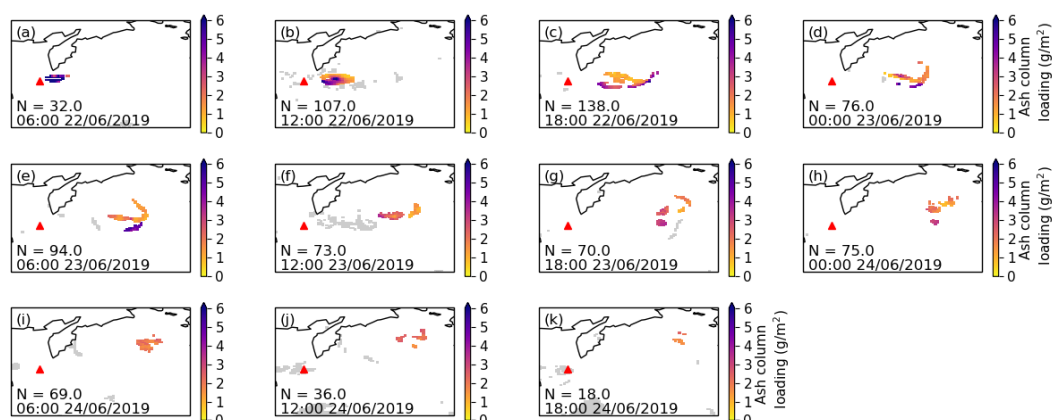



Figure 1: Hourly mean ash column loadings from the Himawari satellite at (a) 0600 UTC, (b) 1200 UTC, (c) 1800 UTC on 22 June 2019, (d) 0000 UTC, (e) 0600 UTC, (f) 1200 UTC, (g) 1800 UTC on 2300 June 2019, (h) 0000 UTC, (i) 0600 UTC, (j) 1200 UTC, (k) 1800 UTC on 24 June 2019. Grey shading indicates grid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

4 Particle filter construction

Drawing from the GLUE methodology described in Sect. 1, we developed a new particle filter for refining a series of ensembles moving forward in time based on their level of agreement with Himawari satellite observations. All ensemble members are evaluated and filtered by comparing the NAME–simulated ash column loadings (g m^{-2}) with the satellite–detected ash column loadings (g m^{-2}) for a given analysis time. Column loading retrievals from Himawari cover a time range from 1800 UTC on 21 June 2019 to 0000 UTC on 24 June 2019. However, due to the low number of grid boxes containing ash before 1800 UTC on 22 June (Fig. 1), for the initial ensemble (*ENS01*) we chose as first verification time, *T1*, the observations at 0600 UTC on 22 June 2019. For this given time, 32 grid boxes containing ash are available (Fig. 1).

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

1. Prior pdfs are created for the 9 perturbed parameters, including eruption source parameters, driving meteorology, and NAME model internal parameters (Table 1; Fig. 2). We assign an initial range for each parameter that is then sampled independently from a uniform distribution to create an ensemble of 1000 members, *ENS01*. Perturbed parameters and their initial ranges are detailed in Sect. 4.1.

214 2. We define our rejection criteria, based on the threshold values for our verification metrics: Hit Rate (HR) and
215   Mean Percentage Difference (MPD). Sect. 4.2 describes how we calculate HR, MPD and how we set the
216   thresholds (Fig. 2).

217 3. Each model prediction from *ENS01* is compared with the satellite observations at a given time, *T1*. All
218   posterior ensembles are verified at a future time using observations every 6 hours (Fig. 2):

$$\text{ENS01} \rightarrow \text{T1} = 22/06/19\ 0600\ \text{UTC}$$

$$\text{ENS02} \rightarrow \text{T2} = 22/06/19\ 1200\ \text{UTC}$$

$$\ldots$$

$$\text{ENS } n \rightarrow \text{T}_n = \text{T}_{n-1} + 6\text{h}$$

223 4. Simulations that, based on the rejection criteria, are not in agreement with the satellite observations are
224   discarded. The retained simulations are used to form the posterior pdfs for each input and internal model
225   parameter (Sect. 4.3; Fig. 2).

226 5. Posterior pdfs are generated from the parameter sets of the retained simulations, also considering possible
227   interaction among them by including effects of covariation between eruption source parameters. Parameters
228   are resampled from these posterior pdfs for the posterior ensemble (Fig. 2). The newly created posterior
229   ensemble represents a possible state of our system at a future time (Sect. 4.3).

230 6. Each model prediction from the posterior ensemble is compared forward in time with a new set of satellite
231   observations. The ensemble output is also compared with all the previous times, to determine sensitivity for
232   the resampled parameters of the posterior pdf to the observations used (Fig. 2).

233 7. The steps 4–6 are repeated until all the available satellite observations are covered.

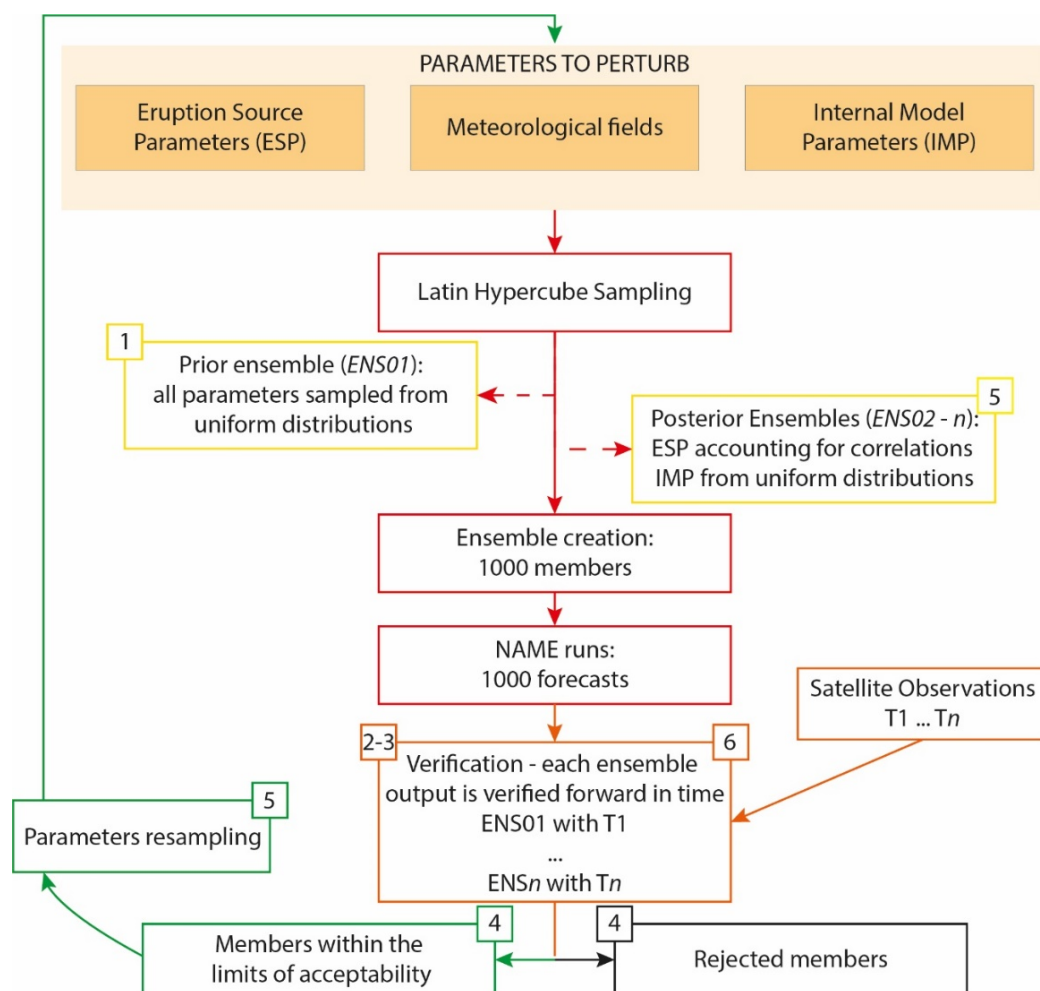



Figure 2: Overall workflow of the developed methodology.

4.1 Ensemble creation

Each ensemble of NAME simulations is created by perturbing nine parameters, including eruption source parameters, driving meteorology, and internal NAME parameters. To generate the ensemble efficiently, we use Latin Hypercube Sampling (LHS), that covers our entire parametric space maintaining orthogonality among the different perturbed parameters (Prata et al., 2019). For *ENS01*, each parameter in the LHS is chosen from prior pdfs which ranges are defined by a set of minima and maxima (Table 1) and sampled from a uniform distribution assuming that all values within these ranges are equally likely. For each posterior ensemble the ESPs in the updated LHS designs are chosen from the posterior pdfs, while NAME internal parameters and members of MOGREPS forecasts continue to be sampled from uniform distributions (Sect.4.3).

4.1.1 Eruption Source parameters (ESPs)

246   To represent an initial uncertainty associated with ESPs for *ENS01*, we define a minimum and a maximum possible

247   value for each perturbed parameter. Then, a parameter value is sampled from a uniform distribution across this range.

248   We selected a total of 5 ESPs to perturb from those that have been shown to have most effect on the simulated ash

249   cloud (e.g., Dacre et al., 2013, Harvey et al., 2018, Prata et al., 2019): plume height (*H*), Distal Fine Ash Fraction

250   (*DFAF*), Mass Eruption Rate factor (*MERF*), ash density (ρ) and eruption duration (*L*). *H* is used to calculate the Mass

251   Eruption Rate (*MER*) for each member using the empirical relationship from Mastin et al. (2009). For each ensemble

252   member, *MER* is scaled by the *DFAF*, representing the percentage of ash transported at long distances, and multiplied

253   for the *MERF*, to account for uncertainties associated with the Mastin et al. (2009) relationship. Hence, *MER* is not

254   perturbed explicitly.

Plume height

256   Plume height (*H*) constrains the lower and upper limits of the ash particles' release height, therefore significantly

257   impacts both the vertical and horizontal structure of the simulated ash cloud. Although the Raikoke eruption was

258   characterized by the formation of an umbrella cloud, in the NAME ash release is defined as a vertical line along which

259   the ash is uniformly distributed, with the lower and upper bounds representing the volcano summit height (551 m asl)

260   and the reported plume height respectively. We based our initial *H* range on information available at that time. The

261   Kamchatkan Volcanic Eruption Response Team (KVERT) and the Tokyo and Anchorage VAAC reported a large ash

262   plume extending from 10 to 13 km (asl) within the first few hours of eruption (Crafford et al., 2019), while data from

263   CALIPSO satellite indicate that the plume may have reached altitudes up to 17 km (asl; Hedelt et al., 2019). For our

264   initial ensemble, we selected *H* ranges between 9–17 km (asl) and for the control run we set *H* = 13 km (asl; Table 1).

Mass Eruption Rate (MER)

266   Mass eruption rate, *MER*, is estimated from the plume height using the empirically derived relationship from Mastin

267   et al. (2009):

$$MER = 50.7 \times 10^7 H^{2/0.241} g hr^{-1} \tag{1}$$

269   where *H* represents the plume height in km. By calculating the *MER* from *H*, any uncertainty associated with *H*

270   propagates in the resulting *MER* calculation. Furthermore, there is also an uncertainty associated with the nature of

271   the relationship, being an empirical one and based on a relatively small number of eruptions of variable magnitude

272   (Mastin et al., 2009). In order to take this into consideration, for all the ensemble members, the *MER* is perturbed for

273   a factor between 0.33 and 3, while we use 1 for the control run (Mass Eruption Rate Factor, *MERF*; Harvey et al.,

274   2018; Prata et al. 2019; Table 1).

Distal Final Ash Fraction (DFAF)

276 The *MER* calculated with the Mastin et al. (2009) relationship (Eq. 1) estimates the total mass released during an
277 eruption. However, the particle size distribution (PSD) of the erupted particles includes both larger particles (>1 mm)
278 that are usually removed from the column in the first phases of an eruption and an additional finer fraction that may
279 leave the column due to aggregation processes. Particles larger than 100 μm are removed rapidly, without travelling
280 long distances, and as result, only a fraction of fine particles <100 μm is transported at long distances. Details of the
281 true PSD are often unknown. Here, the default LVAAC PSD is used in each simulation (Table 1 in Witham et al,
282 2019), and aggregation processes are not modelled in the NAME simulations for this study. To account for this, the
283 model assumes that most of the ash falls out close to the volcano, with only a small percentage of it reaching the distal
284 plume. The NAME default value for this percentage (distal fine ash fraction, DFAF, Dacre et al., 2011), used here for
285 the control run, is 5 %; however, the real value is uncertain and varies with each eruption (Witham et al., 2019).
286 Consequently, the uncertainty associated with DFAF can be very high (Grant et al., 2012). Recent studies challenged
287 the 5 % assumption by reaching contrasting conclusions: either 5 % is too high for most of the eruptions (Gouhier et
288 al., 2019). Or it is too low, severely underestimating mass loadings (Cashman and Rust, 2020). For our prior ensemble,
289 the range used is 0.5–20 % (Table 1).

290 ***Ash density***

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

296 ***Source duration***

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

305 **4.1.2 Driving Meteorology**

306 In this study, NAME was driven by the operational forecasts from the Met Office Global and Regional Ensemble
307 Prediction System (MOGREPS). The global forecasts have 17 ensemble members plus a control member. The
308 horizontal resolution is approximately 20 km in the mid–latitudes and there are 70 vertical levels with the lid at

309   approximately 80 km. Each forecast is run out for 7 days and they are initialised 4 times per day at 00, 06, 12 and 18

310   UTC (Bowler et al., 2008). At the time of the Raikoke eruption, MOGREPS–G used an on-line inflation factor

311   calculation to calibrate the spread of the ensemble in space and time and a stochastic physics scheme to account for

312   model uncertainty (Flowerdew and Bowler, 2011; Flowerdew and Bowler, 2013). The MOGREPS–G forecasts used

313   in this study were initialized at 1200 UTC 21 June 2019.

4.1.3 NAME internal model parameters

315   Previous studies have demonstrated how the NAME internal model parameters used for representing the free

316   tropospheric turbulence can significantly impact the model output as they affect the vertical thickness of the simulated

317   cloud and the overall motion of particles (Dacre et al., 2015; Harvey et al., 2018; Prata et al., 2019). To represent

318   uncertainty in free tropospheric turbulence, we perturb the standard deviation (σ) and Lagrangian timescales (τ) of the

319   horizontal and vertical velocity components in NAME, by sampling them from a uniform distribution using the same

320   ranges specified in Harvey et al. (2018) and Prata et al. (2019), with the horizontal component of σ sampled on a

321   logarithmic scale. The horizontal and vertical components of these parameters are varied in proportion to each other.

322   Similarly, σ of the horizontal velocity for unresolved mesoscale motions is also varied using the same range as in

323   Harvey et al. (2018) and Prata et al. (2019) and sampled from a uniform distribution. For the control run, we use the

324   default NAME values (Table 1). For both the control run and the ensembles, these values are fixed in time.

4.2 Particle filter verification metrics

326   Each simulation is either discarded or retained based on its level of agreement with the satellite retrievals. We restrict

327   the comparison only to the area covered by both NAME simulated ash cloud and by detectable ash in the satellite

328   observations. The comparison is performed based on two verification metrics: Hit Rate (HR) and Mean Percentage

329   Difference (MPD).

4.2.1 Step 1: Identifying matching pixels

331   Most satellite retrievals are unable to detect column loadings less than 0.2 g m^{-2} (Prata and Prata, 2012). Consequently,

332   before comparing the Himawari data and NAME output (Fig. 3a and 3b), we apply a minimum threshold of 0.2 g m^{-2}

333   to the NAME–simulated ash column loading to align with the minimum detection limit of the satellite observations.

334   The Himawari observations are then regridded over the NAME horizontal grid, to facilitate inter comparisons. Finally,

335   we identify the grid boxes in which both the NAME output and the satellite retrievals detect ash as 'matching pixels'

336   (Fig. 3c). For each ensemble member, we then calculate the Hit Rate and Mean Percentage Difference based on all

337   matching pixels.

338

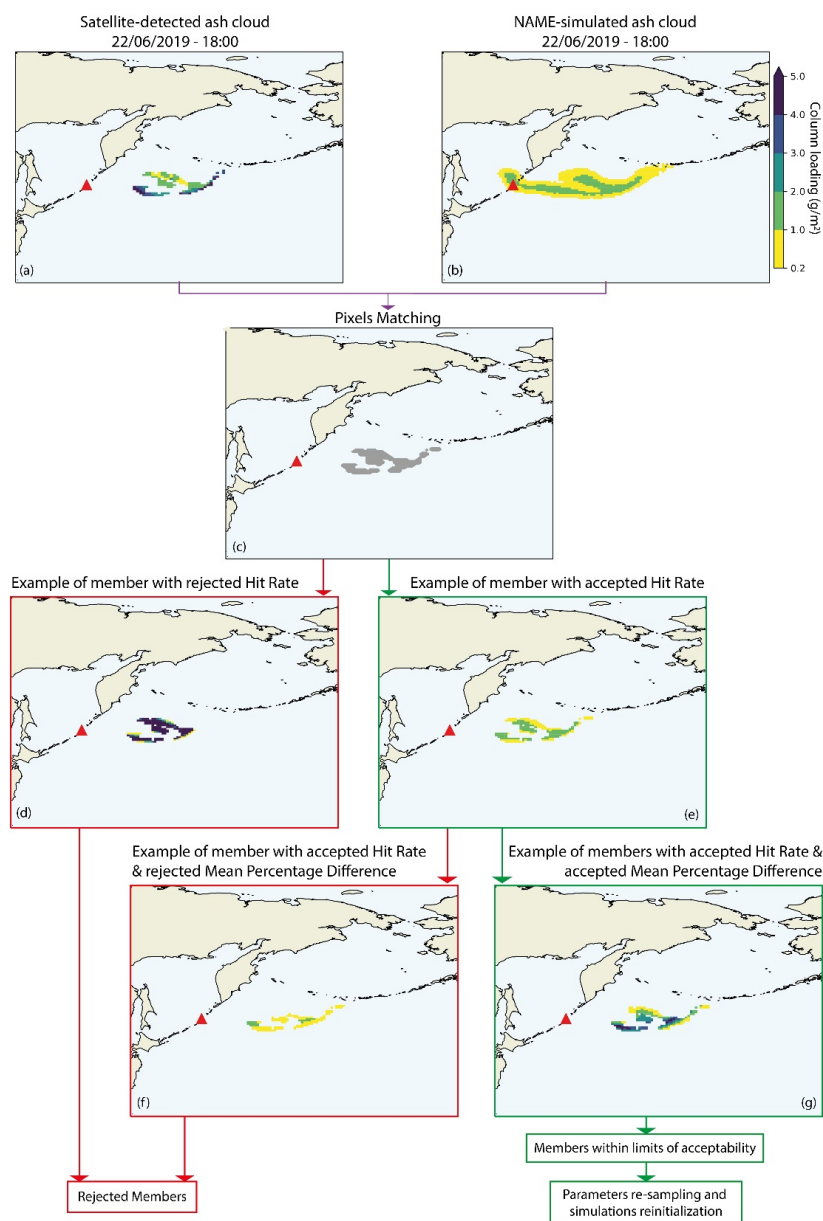



Figure 3: Ash column loading at 1800 UTC on 22 June 2019, (a) as detected by the satellite, (b) NAME simulation of one ensemble member in NAME, and (c) after pixel–matching (grey pixels represent a match between satellite and simulation). Examples of ensemble members rejected depending on values of Hit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

4.2.2 Step 2: Calculating Hit Rate

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

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

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

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

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

4.2.3 Step 3: Calculating the Mean Percentage Difference

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

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

4.2.4 HR and MPD Thresholds

Both *HR* and *MPD* are sensitive to the total number of satellite grid boxes containing detectable ash. Therefore, when there are only a few satellite grid boxes available, fixed *HR* and *MPD* thresholds may retain too few ensemble members to form posterior pdfs. To avoid this, the acceptability thresholds for *HR* and *MPD* are adjusted dynamically during the verification to ensure that a minimum of 50 ensemble members are retained (i.e., 5 % of total number of ensemble

375 members). This dynamic adjustment is carried out by initially setting the *HR* threshold to 95% and the *MPD* threshold

376 to the minimum *MPD* value. Ensemble members are evaluated against these thresholds, and the thresholds adjusted

377 (by increasing the *MDP* value up to the median value, and by decreasing the *HR* value, in 5 % increments) until at

378 least 50 members lie within the limits of acceptability.

379 This dynamic thresholding method, therefore, guarantees that the members within limits of acceptability are always

380 retained using the "best" threshold available for both *HR* and *MPD*, for a given time (Table 2). Only for *T1,* when 32

381 grid boxes containing detectable ash are available, were both thresholds varied substantially and fewer than 50

382 members were retained. Thereafter, a *HR* of 95% was maintained at each verification cycle. *MPD* values had to be

383 varied more to retain at least 50 members but were always within the range 20–50 % of the minimum *MPD*.

Ensemble	Verification Time	Num. of grid boxes	Posterior members WLoA[1]	Prior members WLoA[2]	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

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

386 [1]total number of retained ensemble members for a given time. Parameters for Ensemble 01 (prior ensemble) were
387 sampled from uniform distributions. Parameters for each subsequent ensemble (posterior ensemble) were sampled
388 from posterior pdfs obtained from the retained members of the previous ensemble, verified at the previous verification
389 time. Threshold values for *HR* (higher is better) and *MPD* (lower is better) used at each verification time are shown
390 in the last two columns.

391 [2] total number of ensemble members that would be retained from the prior ensemble, *ENS01*, if the verification was
392 run at each verification time, using the same *HR* and *MPD* thresholds for which each posterior ensemble has been
393 evaluated.

394 **4.3 Posterior resampling**

395 The comparison of the prior and posterior pdfs for each parameter allows us to identify the range of parameters that

396 represent a good approximation of those generating the observed volcanic ash cloud at the verification time. Fig. 4

397 shows this comparison for *ENS01*, *ENS02* and *ENS04*; each ensemble was verified using observations at *T1*, *T2* and

398 *T4* respectively (Table 2). For *ENS01* (Fig. 4a), most of the posterior ESP pdfs from the retained simulations are highly

399 skewed. In particular, this can be seen for *H*, *DFAF* and *MERF. H* is used to calculate the mass eruption rate (shown

400   in Fig. 4 but not explicitly perturbed), that is then perturbed further by *DFAF* and *MERF.* Therefore, these ESPs
401   significantly influence both the vertical and horizontal structure of the modelled ash cloud and ash concentrations.
402   However, *ENS02* (Fig. 4c) and *ENS04* (Fig. 4d), which use the particle filter described above, show how as more
403   ensembles are run and evaluated forward in time with new observations, the parameter ranges of each posterior
404   ensemble gradually reduce. Additionally, the differences between posterior pdfs and the prior pdfs decrease and the
405   ESPs become increasingly constrained.

406   Although this evolution is evident for many of the ESPs, the input model parameters (τU, σU and $m\sigma U$) do not show a
407   similar behaviour among the different ensembles. This suggests that the Raikoke simulations are not sensitive to these
408   internal parameters.

409

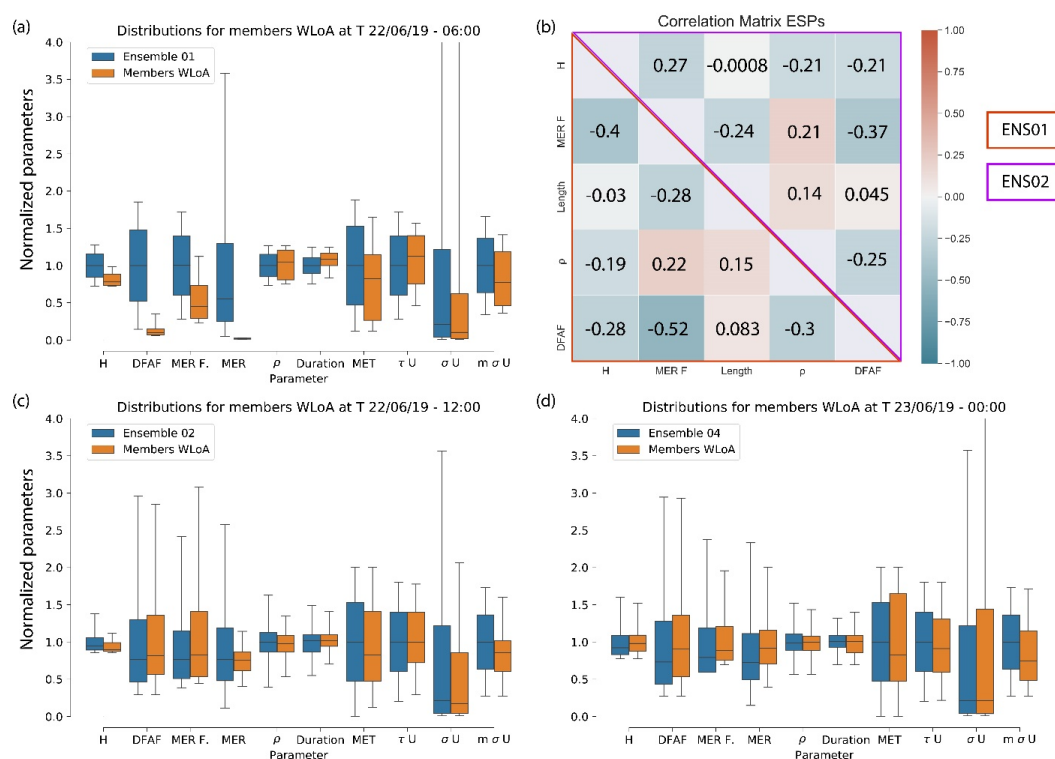



**Figure 4: Evolution of parameters distributions for (a) members Within Limits of Acceptability (WLoA) for Ensemble 01,
(b) comparison of correlation matrixes between ENS01 ESPs in the members WLoA (lower matrix, red triangle) and ESPs
posterior pdfs for ENS02 (upper matrix, purple triangle), and (c) distributions for members WLoA for Ensemble 02 and
(d) Ensemble 04. ρ, τU, σU and $m\sigma U$ represent density, horizontal Lagrangian timescale for free tropospheric turbulence,
standard deviation of horizontal velocity for free tropospheric turbulence, and Standard deviation of horizontal velocity
for unresolved mesoscale motions. In (a), (c) and (d), Mass eruption rate (*MER*) is plotted to show its variation, although it
is not explicitly perturbed in any of the ensembles. Each parameter in the box plots is normalized by dividing each individual
value from the ensemble members by the mean of that entire parameter range from the selected ensemble.**

418   This constraining behaviour is the result of the refinement of each posterior ensemble, achieved by repeatedly refitting

419   the parameters sets of the retained simulations at each verification time, and resampling posterior ensemble parameters

420   from the newly fitted posterior pdfs. For the prior ensemble, the LHS is performed independently for each parameter.

421   However, eruption source parameters are likely to be correlated, especially the plume height, distal fine ash fraction

422   and mass eruption rate factor, which are used to estimate and perturb the mass eruption rate. Thus, before the

423   resampling process, the input parameter ranges from the retained members are used to generate both posterior pdfs

424   reflecting the distribution of each ESP and a correlation matrix for the ESPs (Fig. 4b). Then, using a Cholesky

425   decomposition of the correlation matrix, a correlation structure is enforced to a new LHS design (Fig. 4b). Finally,

426   each ESP is sampled from the newly generated posterior pdfs in the updated LHS design, maintaining the dependency

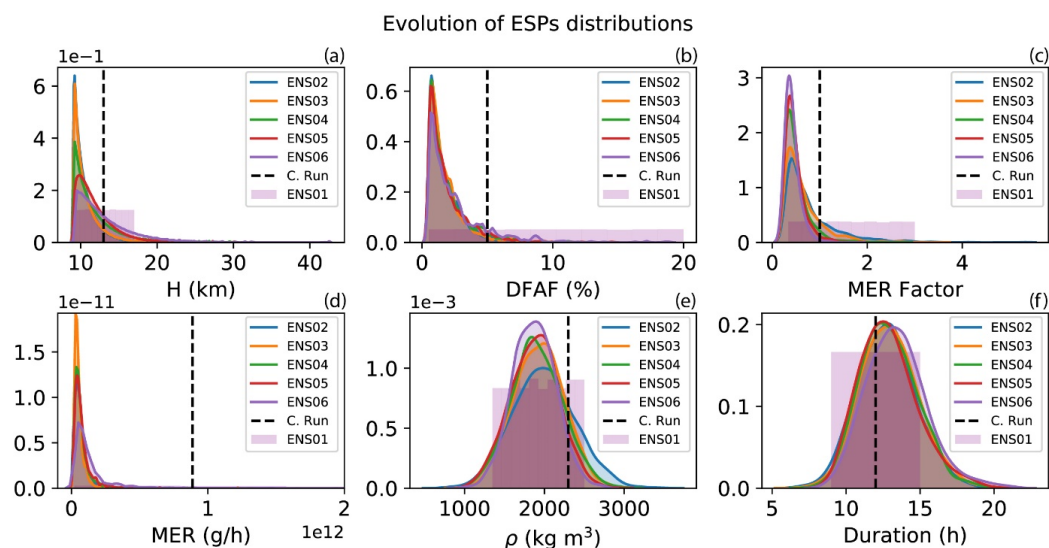
427   among the parameters. In contrast, model input parameters and driving meteorology are treated independently and

428   their posterior pdfs are sampled again from uniform pdfs, using the same ranges as for the prior ensemble (Table 1).

429   For the majority of the ESPs in *ENS02*, this resampling strategy modifies the ESPs distributions away from the initial

430   uniform distribution of the prior ensemble to a distribution that is highly skewed towards the lower end of the initial

431   range (Fig. 5). Then, as the posterior pdfs are refined by moving forward in time with new observations, the pdfs for

432   several of the ESPs tend toward normal distributions (Fig. 5).

434   **Figure 5: Evolution of ESPs distributions for ENS01 (purple shading), ENS02 (blue shading), ENS03 (orange shading),**

435   **ENS04 (green shading), ENS05 (red shading) and ENS06 (dark purple shading). Each panel shows: (a) plume height (*H*),**

436   **(b) distal fine ash fraction (*DFAF*), (c) mass eruption rate factor (*MERF*), (d) mass eruption rate (*MER*), (e) ash density (*ρ*),**

437   **and (f) eruption duration. For ENS01, each parameter is sampled from a uniform distribution. For ENS02–06, each**

438   **parameter is sampled from posterior pdfs following a fit of the members WLoA and considering the interaction among the**

439   **parameters. Mass eruption rate (*MER*) is plotted to show its evolution, although it is not explicitly perturbed in any of the**

440   **ensembles. Dashed black lines in each plot represent the parameter value used for the control run.**

441   **5 Discussion**

442   The particle filtering data assimilation technique described in Sect.4 demonstrates how a series of ensembles of

443   volcanic ash simulations can be successfully constrained based on the level of agreement between the simulation

444   output and satellite retrievals. Each ensemble is verified forward in time with new retrievals. Compared to the

445   parameter ranges used for the prior ensemble (Table 1), the range of eruption source parameters used to produce

446   simulated ash clouds that represent a good approximation to the observed volcanic ash cloud reduces as the posterior

447   ensembles become more constrained by the satellite retrievals.

448   The effects of the refinement on the posterior ensemble can be observed in probability of exceedance maps, given

449   here for three different thresholds of ash column loadings, 0.2 g m^{-2} (Fig. 6b), 2 g m^{-2} (Fig. 6c) and 4 g m^{-2} (Fig. 6d).

450   The satellite retrieval at 1800 UTC 22/06/2019 detects two distinct regions where ash loadings exceed 0.2 g m^{-2} and 2

451   g m^{-2} (Fig. 6a). The posterior ensemble (*ENS03* – brown contour) agrees with the control run (black contour) and the

452   prior ensemble (*ENS01* – purple contour) on regions where loadings > 0.2 g m^{-2} are likely (30–60 %) and very likely

453   (60–100 %; not plotted in Fig. 6). The simulated ash cloud regions are more extensive than the area of satellite–

454   detected ash. However, this overestimation is greater for both the prior ensemble (purple contour) and the control run

455   (black contour) than for the posterior ensemble (brown contour) (Fig. 6b) and is largest for column loadings > 2 g m^{-2}

456   (Fig. 6c) and 4 g m^{-2} (Fig. 6d). The refined ensemble shows a much–reduced region with a 30–100 % probability of

457   exceeding these loadings, showing better agreement with the observations especially when considering loadings > 2

458   g m^{-2} detected by the satellite (Fig. 6a). Thus, by accounting for uncertainties, a wider region where those loadings are

459   less likely (up to 30%, dashed blue contour) has been shown instead, for all the considered thresholds (Fig. 6).

460

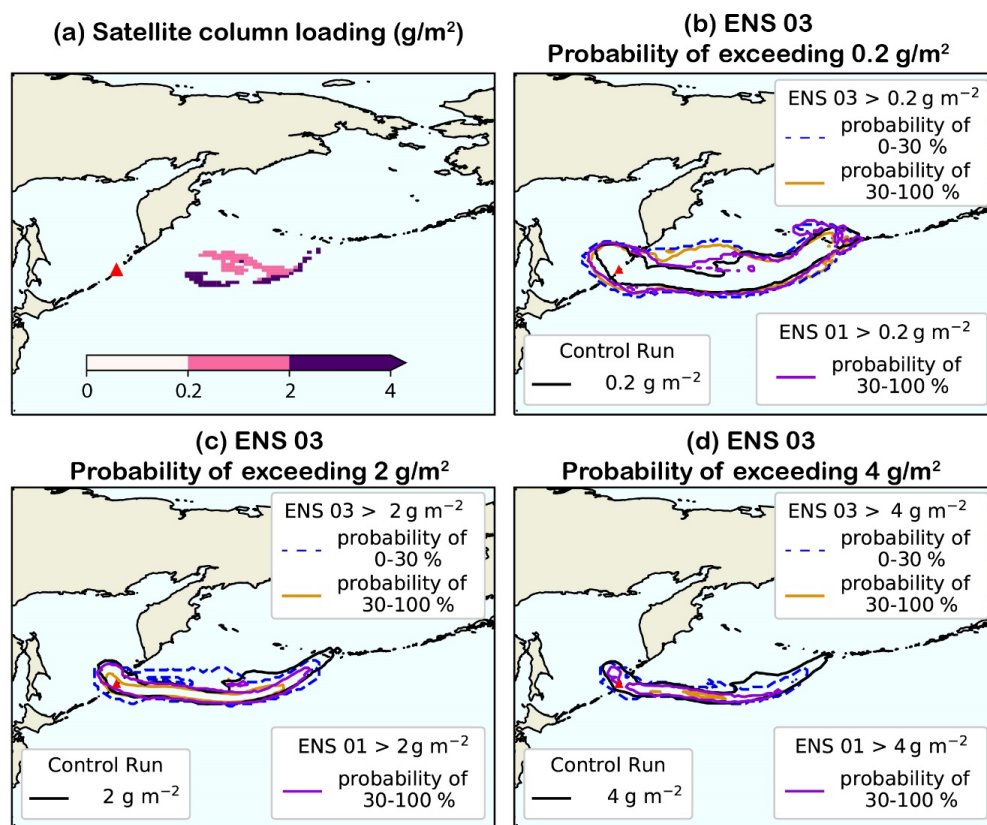



Figure 6: (a) Satellite–detected column loading at *T3* **(22/06/2019 1800 UTC) shown in filled contours, and probability of**

exceedance maps for *ENS03* at *T3* for column loading thresholds of (b) 0.2 g m⁻², (c) 2 g m⁻² (d) 4 g m⁻² shown with dashed

blue contours (the region within the blue line is a probability region up to 30 %) and brown contour (30 % probability

region – the region within the brown line goes from 30 % up to 100 %). The black contour in panels (b), (c) and (d) show

the relevant column loading threshold for the control run. The purple contours show the 30 % probability region for *ENS01*

of exceeding the relevant column loading threshold (the region within the purple dashed line goes from 30 % up to 100 %).

For generating the purple contours, members from the prior ensemble were retained by verifying *ENS01* with observations

at *T3* (a) and using the same *HR* and *MPD* thresholds for which *ENS03* has been evaluated (Table 2) to ensure a fair

comparison between the prior and posterior ensembles.

By considering mean ash concentrations (mg m⁻³) at *T10* for *ENS08*, thus at 0000 UTC on 24 June 2019, for the three

"thick" Flight Layers, FL000–200 (Fig. 7a), FL200–350 (Fig. 7b) and FL350–550 (Fig. 7c), the control run seems to

underestimate the areas with concentrations exceeding 0.2 mg m⁻³ compared to the subset of retained members of both

ENS01 and *ENS08*. Contrary, *ENS01* forecasts concentrations exceeding 0.2 mg m⁻³ over larger regions compared to

the posterior ensemble. The overestimation increases considerably for FL200–350 but also for FL350–550, even with

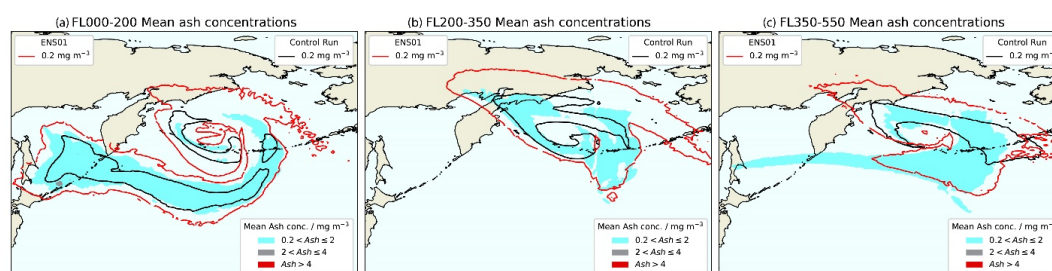
the posterior ensemble forecasting an additional plume tail extending to the east of Raikoke (Fig. 7c). Indeed, the areas

forecasted by *ENS08* with concentrations > 0.2 mg m⁻³ for both FL200–350 (Fig. 7b) and FL350–550 (Fig. 7c) are,

477   respectively, around 60 % and 30 % less extended than the ones forecasted by *ENS01*. In general, for all three flight

478   levels, the area that a posterior ensemble would forecast with high ash concentrations is drastically reduced compared

479   to the prior ensemble.

480

481   **Figure 7: Mean ash concentration values (mg m⁻³) for *ENS08* members for (a) FL000–200, (b) FL200–350 and (c) FL350–**

482   **550 at 0000 UTC 24/06/2019. Each panel shows also the associated 0.2 mg m⁻³ ash concentration contour for both the control**

483   **run (black contour) and *ENS01* (red contour).**

484   **5.1 Application to aviation operations**

485   The use of either probability exceedance maps (Fig. 6) or concentration maps (Fig. 7) condense the information given

486   by the ensemble of VATDM simulations. However, multiple charts are still needed to cover all the relevant

487   information, such as different flight levels, times and ash concentration thresholds. To reduce information overload

488   from these numerous charts, which could impede fast decision–making during emergency response, they can be

489   condensed further into a single chart using a risk matrix (Prata et al., 2019). Here we apply this risk–based approach

490   to international flight routes in the vicinity of Raikoke using both the prior and posterior ensembles outlined in Sect.4.

491   **5.1.1 Flight routes and dosage risk**

492   To simulate potential aircraft encounters with volcanic ash, flight routes with minimised travel time were generated

493   by solving a time–optimal control problem as described in Wells et al. (2021). Trans–Pacific flights were generated

494   assuming a constant true airspeed of 240 m s⁻¹ (~864 km hr⁻¹) at a cruise altitude of FL380 (or 200 hPa) and using

495   horizontal wind speeds extracted from the National Center for Atmospheric Research re–analysis data (Kalnay et al.,

496   1996). Eastbound and westbound time–optimal routes from Sapporo (CTS) to Honolulu (HNL), Los Angeles (LAX)

497   to Seoul (ICN) and San Francisco (SFO) to Shanghai (PVG) international airports were calculated for each day of the

498   dispersion model output (i.e., at 24 hr intervals).

499   As in Prata et al. (2019), the along–flight ash dose, *D*, was defined as the ash concentration multiplied by the duration

500   in that concentration (duration of exposure), integrated along an aircraft's flight path at cruise altitude (assumed to be

501   FL350–550). This definition means that dose always increases monotonically along the route. All dose calculations

502   assume that the modelled ash concentration fields at a given time step are fixed (i.e., do not change with time) as the

503   aircraft flies from the origin to destination at its true airspeed (Prata et al., 2019).

504   **5.1.2 Risk based approach**

505   The first step in determining the risk is to calculate the fraction of ensemble members that have concentrations above

506   specified impact thresholds for each of the three flight levels. In line with the current International Civil Aviation

507   Organisation (ICAO) guidance, the impact concentration thresholds used are 0.2–2 mg m^{-3}, 2–4 mg m^{-3} and greater

508   than 4 mg m^{-3}, for low, medium and high impact respectively. The risk of encountering ash is then determined by

509   combining the likelihood ranges (less likely, 0–10 %; likely, 10–90 %; very likely, 90–100 %) and the impact. The

510   risk of flying in a specific location and at a flight level is then assigned to be low, medium or high. The overall risk

511   presented is the maximum risk over the three flight levels. In Prata et al. (2019), each risk level has a set of actions

512   that may be implemented by the decision maker. These range from checking updated ash forecasts to considering

513   alternative routes and scheduling extra maintenance.

514   The risk can be visualised as a 2D map or projected on to flight tracks of interest (Fig. 8). Considering risk based on

515   ash concentrations at 0000 UTC on 24 June 2019, there are large portions of the flight tracks that encounter low and

516   mid–level risk, with a small region of high risk to the east of Raikoke when using the prior ensemble (Fig. 8a). Thus,

517   based on the prior ensemble output, flight operations could be expected to be severely disrupted at this time. However,

518   determining risk from the posterior ensemble (*ENS08*) removes the region of highest risk and, overall, the amount of

519   flight track potentially impacted and therefore requiring action from the flight operator is greatly reduced (Fig. 8b).

520   To account for the overall exposure of the aircraft to ash, the risk approach can also be applied to dose along a flight

521   track (Fig. 8). To do this the ash dose impact thresholds used are 4.4–14.4 g s m^{-3}, 14.4–28.8 g s m^{-3} and greater than

522   28.8 g s m^{-3} (Clarkson and Simpson, 2017; Prata et al., 2019). The likelihood ranges used are the same as those used

523   for the concentration approach. In this scenario the risk is only determined at cruising altitude, which is assumed to

524   be at FL350–550 (not the maximum over all flight levels). For the prior ensemble, the flight tracks to and from the

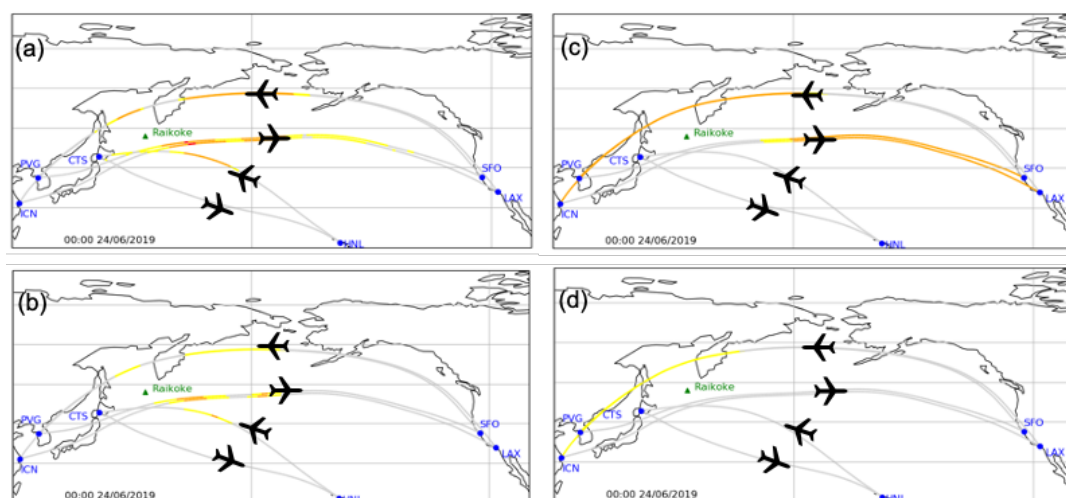
525   West Coast of North America encounter mid–level risk and could potentially require specific actions by the airlines

526   (e.g., more fuel and engine checks). Using this metric, flights between Honolulu and Sapporo do not reach doses that

527   reach the lowest level of risk (Fig. 8c). For the posterior ensemble (*ENS08*), only the route from SFO to PVG and ICN

528   reach sufficiently high doses to be highlighted by the risk approach. The other routes have very few ensemble members

529   where doses are above 4.4 g s m^{-3} and therefore are not highlighted by the risk approach (Fig. 8d). This could greatly

530   reduce the need for an operator to implement any mitigation strategies.

532   **Figure 8: Concentration risk along time optimal flight routes at 0000 UTC on 24 June 2019 for (a) prior ensemble (b)**
533   ***ENS08*. Ash dose risk along the same time optimal flight routes at 0000 UTC on 24 June 2019 for (c) prior ensemble (d)**
534   ***ENS08*. Yellow shading indicates the lowest level of risk, orange shading indicates mid–level risk and red indicates the**
535   **highest level of risk. Note that dose risk only considers risk at FL350–550, whereas concentration risk considers at all levels.**

536   **6 Conclusions**

537   This study presents a new particle filtering data assimilation methodology that combines VATDM simulations with
538   satellite retrievals including their uncertainty estimates, to improve forecasts of volcanic ash cloud location and
539   concentration. A prior ensemble is created by simultaneously varying nine parameters representing the meteorology,
540   eruption source and internal parameters. Members from the prior ensemble are retained or discarded based on their
541   level of agreement with the satellite retrievals. The retained simulations are then used to create a posterior ensemble.
542   Each posterior ensemble is verified and filtered using satellite data at subsequent verification times.

543   The ESPs ranges in the constrained posterior ensembles are both smaller and skewed towards lower values than those
544   used in the control run and prior ensemble. Therefore, a single ensemble designed with unconstrained parameters
545   ranges (i.e., the prior ensemble) seems insufficient for estimating ESPs ranges that may approximate more accurately
546   the observed volcanic ash cloud. This is not the case for internal model parameters which remain unconstrained by
547   the data assimilation.

548   Communicating the risk of volcanic ash to aviation using risk maps and risk trajectories shows that the prior ensemble
549   forecasts mid–level and highest risk for both ash concentration and dose thresholds for much of a set of representative
550   flight tracks. Based on this information flight operations could be severely disrupted by the eruption. However, using
551   the constrained posterior ensemble, the region of highest risk is removed, and the mid–level risk is reduced. Thus,
552   using the refined posterior ensembles potentially reduces the need for the operator to implement any mitigation
553   strategies and hence reduces disruption to airline operations.

554   This methodology is easily generalisable to other VATDMs and could be used to run a comparison with other models.
555   Different remote sensing datasets could be used to assess its sensitivity to the observations used. Running multiple
556   1000–members ensembles requires either a high computational power or can be subject to variable queuing times on
557   computer clusters such as JASMIN. Future work could include code optimization, to make runtime and ensemble size
558   more efficient, potentially allowing an operational application. Furthermore, the evaluation method is based on limits
559   of acceptability; a future improvement could be to define a formal likelihood measure, and weight the behaviour of
560   each simulation output with the observations for creating the posterior.

561   **Data availability**

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

565   **Authors contributions**

566   Conceptualization, H.F.D., N.J.H., A.C., K.B.; methodology, A.C.; software, A.C. and C.A.W.; investigation, A.C.
567   and N.J.H.; resources, C.S.; writing – original draft preparation, A.C., N.J.H., H.F.D.; writing—review and editing,
568   A.C., N.J.H., H.F.D., K.B., C.S., M.R.J.; funding acquisition, M.R.J., H.F.D. and K.B. All authors have read and
569   agreed to the version of the manuscript.

570   **Competing interests**

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

572   **Acknowledgments**

573   This research was funded by the Natural Environment Research Council (NERC) under the project *Radar–supported
574   Next–Generation Forecasting of Volcanic Ash Hazard (R4AsH),* NE/S005218/1. We thank Andrew Prata (University
575   of Oxford) for providing a copy of the Python code for designing an ensemble of NAME simulations and guidance
576   for its modification, and Ben Evans (Met Office) for providing the MOGREPS–G forecasts. We also thank Helen
577   Webster (Met Office) for discussions and her constructive comments on the manuscript before submission.

578   **References**

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

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

587   Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction,
588   Hydrological Processes, 6, 279–298, doi.org/10.1002/hyp.3360060305, 1992.

589   Bowler, N. E., Arribas, A., Mylne, K. R., Robertson, K. B., and Beare, S. E.: The MOGREPS short-range ensemble
590   prediction system, Quarterly Journal of the Royal Meteorological Society, 134, 703–722, doi.org/10.1002/qj.234,
591   2008.

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

594   Cashman, K.V. and Rust, A.C.: Far-travelled ash in past and future eruptions: combining tephrochronology with
595   volcanic studies. Journal of Quaternary Science, 35(1-2), pp.11-22, 2020.

596   Chai, T., Crawford, A., Stunder, B., Pavolonis, M. J., Draxler, R., and Stein, A.: Improving volcanic ash predictions
597   with the HYSPLIT dispersion model by assimilating MODIS satellite retrievals, Atmospheric Chemistry and Physics,
598   17, 2865–2879, doi.org/10.5194/acp-2517-2865-2017, 2017.

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

602   Clarkson, R.J. and Simpson, H.: Maximising airspace use during volcanic eruptions: matching engine durability
603   against ash cloud occurrence. Science and Technology Organization (STO) - Meeting Proceedings Paper MP-AVT-
604   272-17, 15-17 May 2017, Vilnius, Lithuania, pp. 1–20, 2017.

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

609   Dacre et al., 2013 ??

610   Dacre, H. F., Grant, A. L. M., Harvey, N. J., Thomson, D. J., Webster, H. N. and Marenco F.: Volcanic ash layer
611   depth: Processes and mechanisms, Geophys. Res. Lett., 42, 637–645, doi:10.1002/2014GL062454, 2015

...

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

615 Denlinger, R. P., Pavolonis, M., and Sieglaff, J.: A robust method to forecast volcanic ash clouds, Journal of
616 Geophysical Research, 117, doi.org/10.1029/2012JD017732, 2012.

617 Flowerdew, J. and Bowler, N. E.: Improving the use of observations to calibrate ensemble spread, Quarterly Journal
618 of the Royal Meteoro-35logical Society, 137, 467–482, doi.org/10.1002/qj.744, 2011.

619 Flowerdew, J. and Bowler, N. E.: On-line calibration of the vertical distribution of ensemble spread, Quarterly Journal
620 of the Royal Meteorological Society, 139, 1863–1874, doi.org/10.1002/qj.2072, 2013.

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

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

627 Fu, G., Prata, F., Lin, H. X., Heemink, A., Segers, A., and Lu, S.: Data assimilation for volcanic ash plumes using a
628 satellite observational operator: a case study on the 2010 Eyjafjallajökull volcanic eruption, Atmospheric Chemistry
629 and Physics, 17, 1187–1205, doi.org/10.5194/acp-17-1187-2017, 2017.

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

632 Gouhier, M., Eychenne, J., Azzaoui, N., Guillin, A., Deslandes, M., Poret, M., Costa, A. and Husson, P.: Low
633 efficiency of large volcanic eruptions in transporting very fine ash into the atmosphere. Sci Rep 9, 1449
634 doi.org/10.1038/s41598-019-38595-7, 2019.

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

638 Harvey, N. J. and Dacre, H. F.: Spatial evaluation of volcanic ash forecasts using satellite observations, Atmospheric
639 Chemistry and Physics, 16, 861–872, doi.org/10.5194/acp-16-861-2016, 2016.

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

643 Harvey, N. J., Dacre, H. F., Webster, H. N., Taylor, I. A., Khanal, S., Grainger, R. G., and Cooke, M. C.: The Impact
644 of Ensemble Meteorology on Inverse Modeling Estimates of Volcano Emissions and Ash Dispersion Forecasts:
645 Grímsvötn 2011, Atmosphere, 11, 1022,55, doi.org/10.3390/atmos11101022, 2020.

646 Hedelt, P., Efremenko, D. S., Loyola, D. G., Spurr, R., and Clarisse, L.: Sulfur dioxide layer height retrieval from
647 Sentinel-5 Pre-cursor/TROPOMI using FP_ILM, Atmospheric Measurement Techniques, 12, 5503–5517,
648 doi.org/10.5194/amt-12-5503-2019, 2019.

649 Hobbs, P. V., Radke, L. F., Lyons, J. H., Ferek, R. J., Coffman, D. J., and Casadevall, T. J.: Airborne measurements
650 of particle and gas emissions from the 1990 volcanic eruptions of Mount Redoubt, Journal of Geophysical Research,
651 96, 18 735–18 752, 1991.

652 Jolliffe, I. T. and Stephenson, D. B.: Forecast Verification A Practitioner's Guide in Atmosphere Science, Wiley-
653 Blackwell, Chichester, 2012.

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

656 Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen,
657 J., et al.: The NCEP/NCAR 40-year reanalysis project, Bulletin of the American meteorological Society, 77, 437–472,
658 1996.

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

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

666 Lawrence, B. N., Bennett, V., Churchill, J., Juckes, M., Kershaw, P., Oliver, P., Pritchard, M., and Stephens, A.: The
667 JASMIN super-data-cluster, arXiv:1204.3553v1, 2012.

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

670 Mastin, L.G., Guffanti, M., Servranckx, R., Webley, P., Barsotti, S., Dean, K., Durant, A., Ewert, J.W., Neri, A., Rose,
671 W.I., Schneider, D., Siebert, L., Stunder, B., Swanson, G., Tupper, A., Volentik, A. and Waythomas, C.F.: A
672 multidisciplinary effort to assign realistic source parameters to models of volcanic ash-cloud transport and dispersion
673 during eruptions. Journal of Volcanology and Geothermal Research, 186, 10–21, 2009.

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

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

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

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

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

689   Prata, A.T., Dacre H.F., Irvine, E.A., Mathieu, E., Shine, K.P. and Clarkson, R.J.: Calculating and communicating
690   ensemble-based volcanic ash dosage and concentration risk for aviation. Meteorol Appl.; 26:253–266.
691   doi.org/10.1002/met.1759, 2019.

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

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

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

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

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

705 Witham, C., Hort, M., Thomson, D., Leadbetter, S., Devenish, B., Webster, H., Beckett, F., and Kristiansen, N.: The
706 current volcanic ash mod-elling setup at the London VAAC,
707 https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/services/transport/aviation/vaac/london_v
708 aac_current_modelling_setup.pdf, accessed: 2021-06-23, 2019.

709 Woodhouse, M. J., Hogg, A. J., Phillips, J. C., and Sparks, R. S. J.: Interaction between volcanic plumes and wind
710 during the 2010 Eyjafjallajökull eruption, Iceland, Journal of Geophysical Research, 118, 92–109,
711 doi.org/10.1029/2012JB009592, 2013.

712 Zidikheri, M. J., Lucas, C., and Potts, R. J.: Toward quantitative forecasts of volcanic ash dispersal: Using satellite
713 retrievals for optimal estimation of source terms, Journal of Geophysical Research: Atmospheres, 122, 8187–8206,
714 2017.

715 Zidikheri, M. J. and Lucas, C.: A computationally efficient ensemble filtering scheme for quantitative volcanic ash
716 forecasts. Journal of Geophysical Research: Atmospheres, 126, e2020JD033094, 10.1029/2020JD033094, 2020.