



RESEARCH ARTICLE

10.1002/2015JD023895

Using data insertion with the NAME model to simulate the 8 May 2010 Eyjafjallajökull volcanic ash cloud

Key Points:

- Eyjafjallajökull ash cloud is simulated using data insertion
- Captures regional-scale ash cloud structure
- Results evaluated using spatial metrics

Correspondence to:

K. L. Wilkins,
kate.wilkins@bristol.ac.uk

Citation:

Wilkins, K. L., I. M. Watson, N. I. Kristiansen, H. N. Webster, D. J. Thomson, H. F. Dacre, and A. J. Prata (2016), Using data insertion with the NAME model to simulate the 8 May 2010 Eyjafjallajökull volcanic ash cloud, *J. Geophys. Res. Atmos.*, 121, 306–323, doi:10.1002/2015JD023895.

Received 8 JUL 2015

Accepted 2 DEC 2015

Accepted article online 12 DEC 2015

Published online 7 JAN 2016

K. L. Wilkins¹, I. M. Watson¹, N. I. Kristiansen², H. N. Webster³, D. J. Thomson³, H. F. Dacre⁴, and A. J. Prata⁵

¹School of Earth Sciences, University of Bristol, Bristol, UK, ²Norwegian Institute for Air Research, Kjeller, Norway,

³Met Office, Exeter, UK, ⁴Department of Meteorology, University of Reading, Reading, UK, ⁵Nicarnica Aviation, Kjeller, Norway

Abstract A data insertion method, where a dispersion model is initialized from ash properties derived from a series of satellite observations, is used to model the 8 May 2010 Eyjafjallajökull volcanic ash cloud which extended from Iceland to northern Spain. We also briefly discuss the application of this method to the April 2010 phase of the Eyjafjallajökull eruption and the May 2011 Grímsvötn eruption. An advantage of this method is that very little knowledge about the eruption itself is required because some of the usual eruption source parameters are not used. The method may therefore be useful for remote volcanoes where good satellite observations of the erupted material are available, but little is known about the properties of the actual eruption. It does, however, have a number of limitations related to the quality and availability of the observations. We demonstrate that, using certain configurations, the data insertion method is able to capture the structure of a thin filament of ash extending over northern Spain that is not fully captured by other modeling methods. It also verifies well against the satellite observations according to the quantitative object-based quality metric, SAL—structure, amplitude, location, and the spatial coverage metric, Figure of Merit in Space.

1. Introduction

The eruption of Eyjafjallajökull in Iceland (19.60°W, 63.62°N) during April and May 2010 caused major disruption to air traffic as the ash cloud was transported over Europe and the North Atlantic, costing \$1.8 billion to the aviation industry and approximately \$5 billion to the global economy [International Air Transport Association, 2011]. Volcanic ash transport and dispersion models (VATDMs) are important tools which are used by Volcanic Ash Advisory Centers (VAACs) for forecasting the transport and dispersion of ash in order to provide safety advisories to civil aviation authorities. The Numerical Atmospheric-dispersion Modelling Environment (NAME) [Jones *et al.*, 2007] is a Lagrangian VATDM used operationally by the London VAAC, based at the UK Met Office. The eruption source parameters (ESPs; e.g., mass eruption rate, plume altitude and particle-size distribution) that form the source term required by models such as NAME can often be highly uncertain, which, in addition to structural uncertainty inherent in the model and its parameterizations, can lead to considerable uncertainty in model output [e.g., Devenish *et al.*, 2012a].

1.1. Uncertainties in Eruption Source Parameters

NAME does not include some of the complex physical processes that govern the behavior of the plume in proximity to the volcanic vent, e.g., particle aggregation. In standard practice an “effective” source is therefore calculated to represent ash which has already undergone near-source processes and is transported distally. Note that here the term “ash cloud” is used to describe distal ash which has been transported downwind, and “plume” is used to describe the column of ash erupting from the vent.

The difficulties in calculating effective source terms include estimating an effective particle-size distribution (PSD). PSDs calculated from ash deposits do not necessarily represent ash within distal ash clouds. Calculated PSDs can also be highly method dependent and are usually measured after an eruption, rather than during the eruption when VATDMs are run operationally [Bonadonna and Houghton, 2005]. An effective PSD that has been used operationally by the London VAAC is based on empirical measurements from Hobbs *et al.* [1991] and represents only the percentage of fine ash that survives near-source fallout (the fine ash fraction) [Webster *et al.*, 2012]. In addition to the PSD, particle density (g m^{-3}) determines the mass and hence sedimentation

©2015. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

velocity of particles, but this is also very difficult to determine in an operational situation. A default density of 2300 kg m^{-3} is assumed by the London VAAC for all eruptions unless other information is available.

Often, mass eruption rate (MER) is calculated using an empirical relationship between the height of the plume above the volcano summit and the MER of material [e. g., *Sparks et al.*, 1997; *Mastin et al.*, 2009]. One such relationship is given by:

$$Q_m = 140.876z_{\text{max}}^{4.15}, \quad (1)$$

where z_{max} is plume height above the vent in kilometer and Q_m is MER in kg s^{-1} [*Mastin et al.*, 2009]. The calculated MER and an estimation of the fine ash fraction can then be used to calculate the effective source strength required by NAME. Typically, the mass of ash in the distal cloud is the order of $\sim 5\%$ of the total erupted mass, although this percentage is usually not well known and can therefore cause large uncertainties in the modeled concentration of distal ash clouds [*Devenish et al.*, 2012a]. The empirical relationship in equation (1) does not take into account meteorological conditions at the time of the eruption or the effects of moisture at the source which influence the characteristics of an eruption plume [*Devenish et al.*, 2012a]. The relationship may also be biased toward larger events which are easier to study [*Woodhouse et al.*, 2013]. Eruptions in moist atmospheres, weak eruptions subject to high wind speeds, or those with a high proportion of large particles may therefore vary significantly from the calculated MER [*Sparks et al.*, 1997; *Woodhouse et al.*, 2013]. Modeling methods have been developed to deal with some of these uncertainties. Plume models such as PLUMERIA [*Mastin*, 2007] and PlumeRise [*Woodhouse et al.*, 2013] include some of the physical processes that govern the behavior of an erupting plume and can be used to make estimates of ESPs to use within VATDMs [e.g., *Kristiansen et al.*, 2012]. Inversion methods have also been used to estimate effective ESPs using satellite imagery for both volcanic ash and SO_2 [*Eckhardt et al.*, 2008; *Kristiansen et al.*, 2010; *Stohl et al.*, 2011; *Kristiansen et al.*, 2012; *Boichu et al.*, 2014; *Moxnes et al.*, 2014; *Kristiansen et al.*, 2015; *Pelley et al.*, 2015].

Plume rise height determines the level at which ash enters the atmosphere and hence which direction it will travel due to the wind direction at that height. Radar plume height measurements can be used for estimating the plume rise height, but they are subject to some uncertainty due to interpolation between discrete bands, the curvature of the Earth, and local topography [*Arason et al.*, 2011; *Folch et al.*, 2012]. Also, the height at which ash spreads laterally into the atmosphere may not be the maximum measured plume height. This could be due to strong winds causing a weak plume to bend over, moving the location of the maximum plume height to downwind of the vent or due to the vertical momentum of the eruptive column causing it to extend higher than the level of neutral buoyancy before relaxing back down to that level [*Webster et al.*, 2012]. The plume can then spread laterally and thin under gravity (a gravity current), until it approaches the ambient wind speed [*Sparks*, 1986; *Bursik et al.*, 1992]. The physics describing gravity currents is not currently included in most VATDMs (including NAME), but for small to medium-sized eruptions such as Eyjafjallajökull (2010), gravity currents may only dominate transport (over the effects of advection and turbulence) in the earliest phase [*Bursik et al.*, 1992; *Costa et al.*, 2013].

One basic effective emission profile in NAME is a column of ash extending from the vent to the maximum plume height with a uniform vertical distribution, but the emission profile can be prescribed in different ways to represent effective plumes affected by the near-source processes mentioned above. A number of studies have investigated the effect of the emission profile on the downwind ash layer thickness and on modeled ash concentration [e.g., *Dacre et al.*, 2011; *Devenish et al.*, 2012a, 2012b; *Grant et al.*, 2012; *Dacre et al.*, 2015]. It has been found that ash layer thickness far from the source can be dominated by subgrid scale diffusion, wind shear and vertical resolution of the model output, over the effects of assumed emission profile shape and PSD [*Devenish et al.*, 2012a, 2012b; *Dacre et al.*, 2015].

In calculating an effective source to account for the near-source processes that are not included in VATDMs, a number of uncertainties will be introduced. In this study we implement a data insertion modeling method to create a downwind “virtual” ash source that does not require the calculation of an effective eruption source. This does, however, require assumptions to be made about the downwind source, including ash layer depth, vertical distribution, and PSD. This method was introduced in *Wilkins et al.* [2014] for the 2010 Eyjafjallajökull eruption and showed promising initial results. Here we expand on that work to qualitatively and quantitatively compare the results of this method with three other previously studied simulations described in section 2.2, specifically focussing on the 8 May 2010 Eyjafjallajökull ash cloud.

Table 1. SEVIRI 1D-Var Retrievals of the Eyjafjallajökull Ash Cloud Used for Data Insertion

Retrieval	Time (UTC)
R1	1900 06 May 2010
R2	0700 07 May 2010
R3	1800 07 May 2010
R4	0000 08 May 2010
R5	0300 08 May 2010
R6	0600 08 May 2010

2. Methods

2.1. Data Insertion

In data insertion, the model state is set to observed values where they are available, and the model is used to propagate that information with time. Here NAME is initialized using the physical properties of downwind ash clouds estimated from satellite observations. Instead of using these data to inform us about the eruption

source, we create a downwind source term in NAME, where a source is created for each satellite pixel, and a set number of particles are released per source. Ash from all sources is released into the model atmosphere instantaneously. This method requires little prior knowledge of the eruption and hence allows us to run volcanic ash simulations without the usual NAME source term. In doing this we can ignore some of the uncertainties associated with estimating an effective source. However, this does not take account of observation uncertainty and therefore introduces uncertainties attributable to the derivation of the observation data, in mapping observations to formats required by the model, and in the assumptions about the downwind source mentioned above.

Ash column loading (ACL) and ash cloud height estimates for a range of times during the May 2010 phase of the Eyjafjallajökull eruption (see Table 1) are retrieved from Meteosat Second Generation Spinning Enhanced Visible and InfraRed Imager (SEVIRI) infrared channels. The retrieval algorithm optimally estimates ash cloud properties with a one-dimensional variational (1D-Var) method using the 8.7, 10.8, 12.0, and 13.4 μm wavelength channels (for details see Francis *et al.* [2012]). Some example ACL retrievals are shown in Figure 1. The retrieval is run on the pixels in which ash has been detected, using four different refractive index data sets; 1.5% Haematite mineral dust [Balkanski *et al.*, 2007], volcanic dust [Volz, 1973], andesite and obsidian [Pollack *et al.*, 1973], and the algorithm provides solution costs for each. The Pollack *et al.* [1973] andesite data set is used here because all of the retrievals in Table 1 attained the lowest total solution costs using that data set. This result is in agreement with the Francis *et al.* [2012] study and with Millington *et al.* [2012], who found that andesite refractive indices gave the best results for simulation of SEVIRI imagery for this eruption.

As the retrieval data is two dimensional, the vertical structure of the ash and the thickness of the layer are estimated based on observations of the Eyjafjallajökull ash cloud from various times in May 2010 reported in a number of studies [Marenco *et al.*, 2011; Grant *et al.*, 2012; Johnson *et al.*, 2012]. Here three options for the ash layer thicknesses are assumed: 0.5, 1.0, or 2.0 km for the entire ash cloud. The mass is distribute vertically following a Gaussian distribution with the standard deviation matching the standard deviation of a uniform distribution over the nominal thickness. This distribution is used so that the peak of the distribution is around

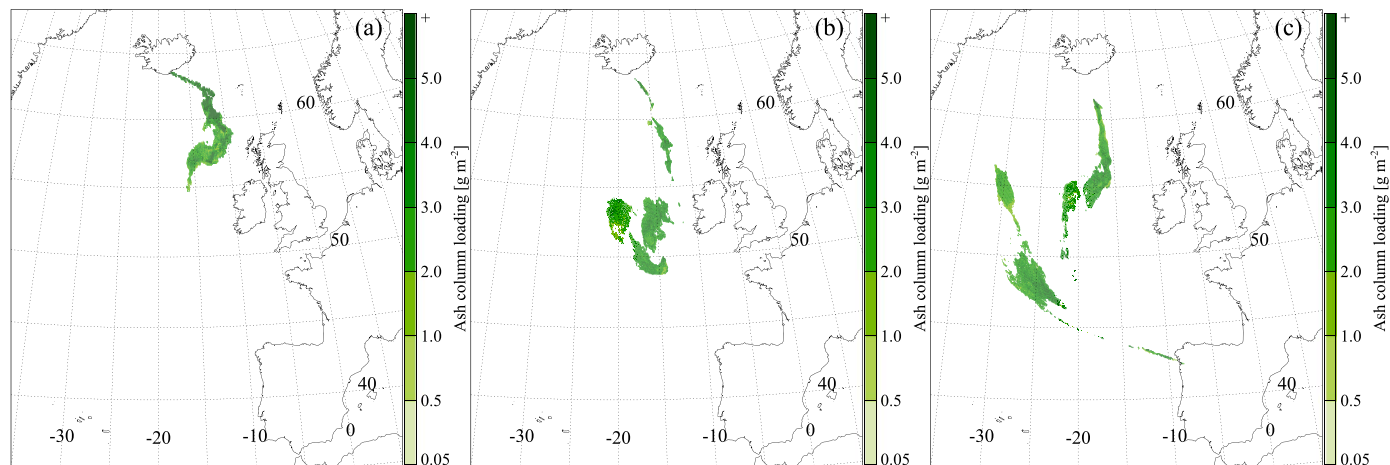


Figure 1. SEVIRI 1D-Var ash column loading retrievals of the Eyjafjallajökull ash cloud. (a) R1: 1900 UTC 6 May 2010, (b) R2: 0700 UTC 7 May 2010, and (c) R4: 0000 UTC 8 May 2010. Note that the lower part of the color scale is nonlinear for these figures and others.

the middle of the layer and for simplicity in running NAME (discussed further in section 3.4). The Gaussian distribution has a standard deviation of $dz/\sqrt{12}$, where dz is the layer thickness, and z , the altitude of the center of the layer, is given by $z = \text{retrieved ash cloud height} - dz/2$. It is noted that the Gaussian distribution will distribute small amounts of ash either side of this layer. Also, the retrieved ash cloud height is likely to represent the top of the ash cloud for optically thick ash layers but may be toward the center for optically thinner ones [Francis *et al.*, 2012], and this could lead to an underestimation of the ash cloud top height when inserted into NAME.

The SEVIRI satellite retrieval is most sensitive to particles 2–32 μm in diameter [Stohl *et al.*, 2011]. This range falls within, but is narrower than, the range covered by the PSD used operationally in NAME for volcanic ash [see Webster *et al.*, 2012, Figure 1]. Here we use a PSD derived from measurements taken on the United Kingdom's Facility for Airborne Atmospheric Measurements BAe-146 research aircraft on 14 May 2010 [see Dacre *et al.*, 2013, Table 1], which has a similar size range to the SEVIRI-sensitive range. A particle density of 2300 kg m^{-3} is used, and 1 hour average ACLs are calculated in NAME over each output $0.25 \times 0.25^\circ$ grid cell.

2.2. Forecasts

In this section the construction of our two main types of data insertion forecasts, which we refer to as layered model (LM) or single retrieval (SR) forecasts, are described. The SR forecasts are simulations initialized with data from one retrieval. They are named according to the retrieval in Table 1 used to initialize them, e.g., SR1 was initialized using retrieval R1.

For the LM forecasts, output from a number of SR simulations are combined into a composite simulation in order to capture ash erupted at different times and ash potentially obscured by cloud in certain retrievals. The maximum ACL value for each output grid cell from all SR simulations is used for the LM forecast, i.e., for a set of N SR simulations ending at the same time, the LM forecast ACL value for grid cell xy is given by

$$ACL_{xy}^{\text{forecast}} = \max(\mathbf{ACL}_{xy}), \quad (2)$$

where \mathbf{ACL}_{xy} is grid cell xy in the set of simulations SR1 to N ($ACL_{xy}^{\text{SR1}}, \dots, ACL_{xy}^{\text{SRN}}$).

We only use ash-flagged pixels to create the SR forecasts and do not include a meteorological cloud detection scheme to indicate the presence of cloud (i.e., where ash may be obscured). Therefore, taking the maximum value per grid cell from a number of simulations in the LM method should provide more conservative forecasts (for observed ash) than the SR method. Take the example case where ash is detected in an early retrieval and inserted into NAME, but undetected due to cloud cover in a later retrieval, and hence not inserted into NAME. In this case, assuming that transport errors are not prohibitively large, we would like to take the maximum value at the location at which the observed ash has been transported in the model, rather than preferentially choosing the value from the most recent simulation. A cloud detection mask would provide more information on the areas that are likely to be clear or to contain cloud, which could be used to indicate which areas are likely to give reliable ash (or no ash) retrievals. We aim to employ such a mask in future work. It is noted that using the maximum value method, ash mass will not be conserved in the layered forecasts.

For short-lived or instantaneous eruptions, or where there is minimal cloud cover, gains from using a large number of retrievals to create simulations may be negligible, and a small number of retrievals may be sufficient to capture much of the ash cloud mass and structure. For long-lived eruptions with some cloud cover, or short-lived eruptions with a lot of cloud cover, it may be appropriate to allow a larger number of SR simulations to contribute to the final LM forecast. It is noted that this method will be most effective at times with little cloud cover and ineffective when parts of the ash cloud are consistently undetected in the retrievals.

The LM forecasts are named according to the number of contributing SR simulations, starting with SR1, e.g., LM(3) is a combination of SR1, SR2, and SR3. For both SR and LM forecasts the number denoted in kilometer is the ash layer thickness assumed at initialization (e.g., SR1 1.0 km).

The final layered forecasts could be created in a number of ways, so alongside taking the maximum ACL, we will show three other alternative options. The first is a weighted average of the simulations, where ACL values for each xy grid cell are weighted according to the time since their insertion:

$$\mathbf{w} = \frac{(48 - \mathbf{t})}{\sum (48 - \mathbf{t})} \quad (3)$$

Table 2. Previously Studied Model Simulations Used Here for Comparison to the Data Insertion Forecasts

Forecast	Met Data	Details
NAME best guess	MetUM	A forecast using NAME initialized with the Met Office's best estimate of the ESPs using plume rise height observations for the duration of the eruption and MER calculated using the relationship in equation (1) [see Webster <i>et al.</i> , 2012].
NAME a posteriori	MetUM	A forecast using NAME initialized with ESPs derived using the inversion method of Kristiansen <i>et al.</i> [2012]. An a priori estimate of the ESPs was made and a posteriori ESPs calculated by applying an inversion algorithm to a series of model runs and satellite-retrieved ash properties.
FLEXPART a posteriori	ECMWF	Same as NAME a posteriori but with FLEXPART replacing NAME, both in applying the inversion method to estimate the source and in calculating the forecast, and with the European Centre for Medium-Range Weather Forecasts (ECMWF) meteorological data.

$$ACL_{xy}^{\text{forecast}} = \sum (\mathbf{w}(ACL_{xy})), \quad (4)$$

where \mathbf{w} is an N length vector of weights and \mathbf{t} is an N length vector of the duration in hours of simulations $SR1 : N$. The neighborhood is somewhat arbitrarily chosen to be 48 h, beyond which simulations would have no impact on the forecast (the longest running simulation used here is 38 h). It is noted that a different neighborhood value would provide different weights for each simulation. For the second alternative forecast option we take the mean of simulated values to give the forecast ACL value for xy . To create the final forecast option, the ACL value for grid cell xy from the most recent simulation replaces an ACL value from an older simulation, but only if it is greater than zero. These methods will be denoted the weighted average, mean, and recent value methods, respectively.

We compare the results of the data insertion forecasts with three other forecasts. The NAME “best guess” forecast is a NAME simulation initialized using a posteruption analysis of plume eruption height measurements and a fine ash fraction of 5% [see Webster *et al.*, 2012]. The NAME a posteriori and FLEXPART dispersion model (FLEXPART) a posteriori forecasts are simulations using NAME and FLEXPART [Stohl *et al.*, 1998], respectively, initialized using ESPs derived with the inversion method of Kristiansen *et al.* [2012]. Details of these methods are summarized in Table 2.

2.3. SAL

SAL is a quality measure comprising three components: structure (S), amplitude (A), and location (L) [Wernli *et al.*, 2008]. This diagnostic tool has previously been applied to NAME output by Dacre [2011] for air quality forecasts and has been adapted for use here. Perfect forecasts will have scores of 0 for all components of the assessment, with ± 2 being the worst possible score for both S and A and +2 being the worst score for L. Objects from the observed and modeled ACL fields must be identified for S and L evaluations. An object is a group of adjacent grid cells which have an ACL value above a given threshold. In Wernli *et al.* [2008] the threshold is taken as one fifteenth of the maximum value in the field. The SEVIRI ash detection limit is considered to be $\sim 0.2\text{--}0.3 \text{ g m}^{-2}$ or more conservatively $0.5\text{--}1.0 \text{ g m}^{-2}$ [Devenish *et al.*, 2012a; Francis *et al.*, 2012; Prata and Prata, 2012], so the threshold used here is chosen to be 0.5 g m^{-2} . Model ACL values below this threshold are excluded from all S, A, and L calculations. Here groups of at least two adjacent grid cells constitute an object so that the number of objects is not increased by spurious single ash-containing grid cells. Objects are given as O_n , $n = 1, \dots, M$, where M is the number of objects in the domain. The domain is -45 to 50°E , 30 to 80°N .

2.3.1. Structure (S)

The structure component compares normalized object sizes and shapes. A scaled ash mass (scaled area-integrated ACL), V_n , is calculated for each modeled and observed object:

$$V_n = \sum_{(xy) \in O_n} R_{xy} / R_n^{\max} = R_n / R_n^{\max}, \quad (5)$$

where xy is the grid cell location within the modeled or observed field, R_{xy} is the area-integrated ash column loading (i.e., ash mass) in grid cell xy , $R_n = \sum_{(xy) \in O_n} R_{xy}$ is the ash mass in object O_n and R_n^{\max} is the maximum grid cell ash mass within the object O_n . Weighted means (V) of the scaled masses (V_n) for all objects are calculated separately for the modeled and observed fields. Each object has a weight proportional to its ash mass R_n :

$$V = \frac{\sum_{n=1}^M R_n V_n}{\sum_{n=1}^M R_n}. \quad (6)$$

S is calculated as the normalized difference in observed (obs) and modeled (mod) V values:

$$S = \frac{V_{\text{mod}} - V_{\text{obs}}}{0.5[V_{\text{mod}} + V_{\text{obs}}]}. \quad (7)$$

S scores range between -2 and $+2$. A negative S score indicates that predicted fields cover too small an area or are too peaked, which can be achieved when the maximum modeled ACLs are much higher than the observed or where the observed field is more widespread [Dacre, 2011]. A positive S score indicates a predicted field that is too spread out and/or flat, which can be achieved when there are many more objects in the observed field, giving a smaller average observed object size [Dacre, 2011].

2.3.2. Amplitude (A)

The amplitude component of SAL compares the normalized difference between observed and modeled ash mass averaged over the domain area. A scores range between -2 (-1 is a factor of 3 underprediction of the domain-averaged values) and $+2$ ($+1$ is a factor of 3 overprediction) [Wernli et al., 2008], with 0 indicating a perfect match in amplitude. A is given as follows:

$$A = \frac{\bar{R}_{\text{mod}} - \bar{R}_{\text{obs}}}{0.5[\bar{R}_{\text{mod}} + \bar{R}_{\text{obs}}]}, \quad (8)$$

where \bar{R}_{mod} and \bar{R}_{obs} are the modeled and observed ash masses averaged over all grid cells in the domain (-45 to 50°E , 30 to 80°N), respectively, i.e., $\bar{R} = \sum_{(xy) \in \text{Domain}} R_{xy} / \text{domain area}$. Zero ash mass is assumed where no ash is detected.

2.3.3. Location (L)

The location component compares the distribution of the modeled and observed masses. It is made up of two parts, L_1 and L_2 . L_1 is a normalized distance measure, comparing the centers of mass between modeled and observed fields:

$$L_1 = \frac{|C_{\text{mod}} - C_{\text{obs}}|}{D}, \quad (9)$$

where $|C_{\text{mod}} - C_{\text{obs}}|$ is the distance between the modeled and observed centers of mass. D is the maximum distance within the entire domain. L_1 ranges between 0 (centers of mass in the same location) and 1 (centers of mass are as far apart as they can be within the domain). For a perfect match $L_1 = 0$, however, a field containing two separate ash clouds could have the same center of mass as a field with one ash cloud. L_2 measures the difference in distribution of masses in the modeled and observed fields about the respective centers of mass. For both fields, the weighted average distance (H) between the center of mass (C) and the center of mass of each object within the field (C_n) is given as follows:

$$H = \frac{\sum_{n=1}^M R_n |C - C_n|}{\sum_{n=1}^M R_n}, \quad (10)$$

where $|C - C_n|$ is the distance between the center of mass of the field and the center of mass of the object n . If there is only one object in the domain, $H = 0$. L_2 is calculated based on the difference between modeled and observed H values:

$$L_2 = 2 \left[\frac{|H_{\text{mod}} - H_{\text{obs}}|}{D} \right]. \quad (11)$$

If both the observed and modeled fields have only one object, $L_2 = 0$. A factor of 2 is used to scale L_2 to the same range as L_1 . L is given by the sum of the two components, L_1 and L_2 , and has a range between 0 and 2:

$$L = L_1 + L_2. \quad (12)$$

We note that the magnitude of L is dependent on the size of the domain. However, we use the same domain for each calculation so that the scores can be considered relative to each other; therefore, the size of the domain will not affect our interpretation of the results. L does not take account of the objects rotation or rotation about the center of mass [Wernli *et al.*, 2008]. An absolute score ($|S| + |A| + L$) can range from 0 to 6, with values closest to 0 indicating the best agreement between the model and observation.

2.4. FMS

The Figure of Merit in Space (FMS) is used to measure the spatial coverage of modeled compared with observed fields [Galmarini *et al.*, 2010] and is defined as the area of intersection of modeled and observed fields / the area of union (equation (13)). The method of interpolation from observation to model grid will therefore affect the results [Mosca *et al.*, 1998], but again, we consider the results in relation to each other, so this will not affect our interpretation. Spatial ash coverage of both the modeled and interpolated observation fields is calculated for grid cells with ACL values exceeding 0.5 g m^{-2} . Good model output will have a high FMS value, but a low FMS could correspond to two similar shapes shifted in space [Mosca *et al.*, 1998]. Hence, this is taken to be a good complement to the SAL metric which does not measure spatial overlap. FMS is given as follows:

$$\text{FMS} = 100\% \frac{B_{\text{mod}} \cap B_{\text{obs}}}{B_{\text{mod}} \cup B_{\text{obs}}}, \quad (13)$$

where B_{mod} and B_{obs} are the modeled and observed ash areas, respectively.

3. Results and Discussion

3.1. Observed and Retrieved Ash

A thin filament of ash can be seen extending over northern Spain at 0900 UTC 8 May 2010 in the brightness temperature difference (BTD) image shown in Figure 2a. This filament is not well captured using the ash detection method in the 1D-Var retrieval (Figure 2b). However, turning off the final ash detection test which utilizes β ratios adds many more ash flagged pixels, and the structure is partially observed at that time (Figure 2c). β ratios are ratios of effective absorption optical depth between pairs of spectral channels [Pavolonis *et al.*, 2013]. They facilitate isolation of cloud microphysical properties from contributions from the surface and atmosphere through conversion of spectral radiance to effective absorption optical depth [Pavolonis, 2010]. The calculation of β ratios negates the requirement for large offline look-up tables in the retrieval of ash physical properties [Pavolonis *et al.*, 2013], and they are used in the 1D-Var retrieval as a means of removing flags from pixels that have previously been tentatively flagged as containing ash [Francis *et al.*, 2012]. This technique largely removes the need for algorithm tuning and, under some conditions, has higher skill scores than simple band differencing but, under some circumstances, overconstrains ash cloud extent. Indeed, the ash detection method used within the Prata and Prata [2012] retrieval, which does not utilize β ratios, detects a larger amount of ash over Spain (Figure 2d). For an arid land surface such as Spain, emissivity is low and quite variable, so the $8.7 \mu\text{m}$ signal to which the β ratio test is sensitive is also likely to be variable. This means that turning the test off could be reasonable in this case (P. Francis, personal communication, 2015). Note that the retrievals we use in creating the data insertion simulations have the β ratio test turned on.

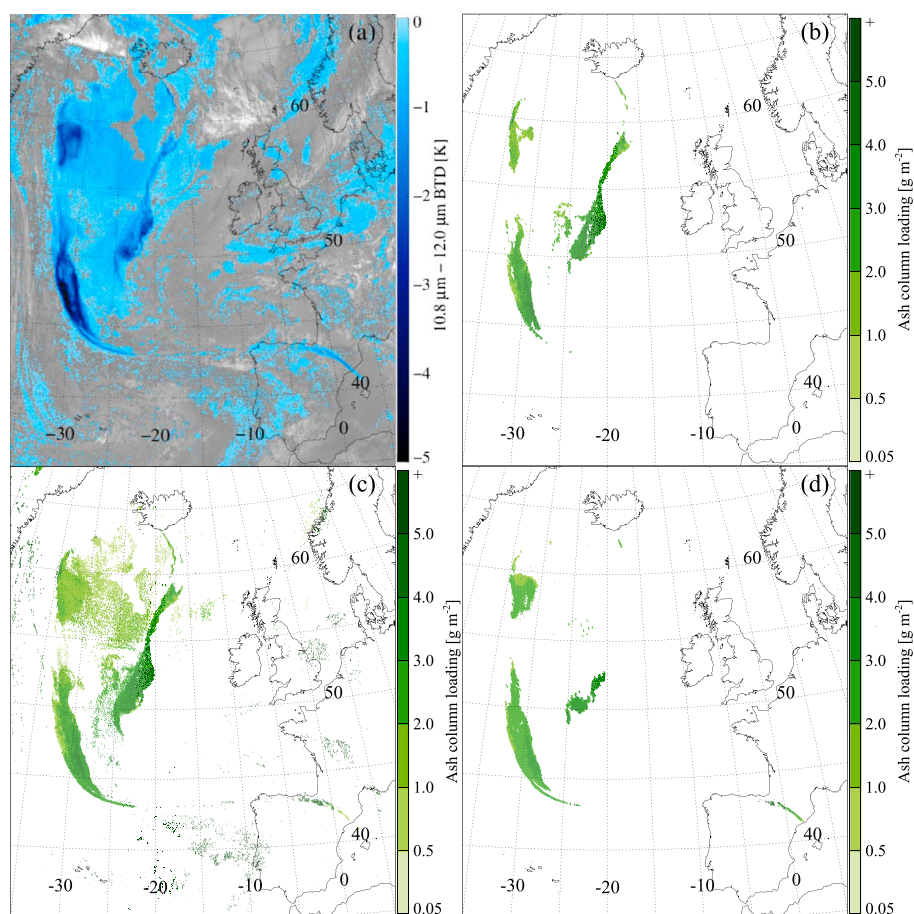


Figure 2. Observed Eyjafjallajökull ash cloud at 0900 UTC 8 May 2010. (a) Brightness temperature difference image with negative values shown in blue, (b) 1D-Var ash column loading retrieval [Francis *et al.*, 2012], (c) same as Figure 2b with the β ratio test removed from detection steps (see text), and (d) retrieved ash column loading using the look-up-table method of Prata and Prata [2012].

3.2. Forecasts

3.2.1. Qualitative Assessment

The NAME best guess forecast (Figure 3a) places ash in a similar location to the NAME a posteriori forecast (Figure 3b) with increased ACL in general. The filament of ash extending over northern Spain is apparent in both of these simulations, but the best guess extends it slightly farther east. The overall shape of the FLEXPART a posteriori field (Figure 3c), which uses a different meteorology to the others, is similar to the other forecasts, but the filament of ash over northern Spain is not visible.

The LM data insertion forecast using retrievals R1–R5 (see Table 1), where a 1.0 km thick ash layer was assumed during insertion (hence, the forecast is denoted LM(5) 1.0 km), captures the ash cloud structure over Spain well, as seen in Figure 3d. (This forecast has effectively a 6 h lead time; the latest image to be inserted into NAME is valid from 6 h previous to the forecast time.) Qualitatively, the position of the ash agrees well with the BTD imagery, but the ACL values over Spain are generally lower than those in Figures 2c and 2d.

The ash filament over Spain in the LM forecast (Figure 3d) is more horizontally spread than the NAME best guess simulation and the observations. The LM forecast is a combination of a number of simulations, so it is likely to have more horizontal spread than a single simulation, particularly if ash is added in the wrong location or height in one or more of the simulations due to errors in the retrieval or layer thickness estimates. Errors in the meteorological fields at the ash insertion heights could also contribute to the increased spread. As mentioned above, the SEVIRI ash detection limit is considered to be ~ 0.2 – 1.0 g m⁻², so some ash may not have been detected around the periphery of the filament over Spain. Much of the modeled ash over Spain is below the lower bound of the conservative threshold (0.5 g m⁻²). The LM forecast has placed some ash over

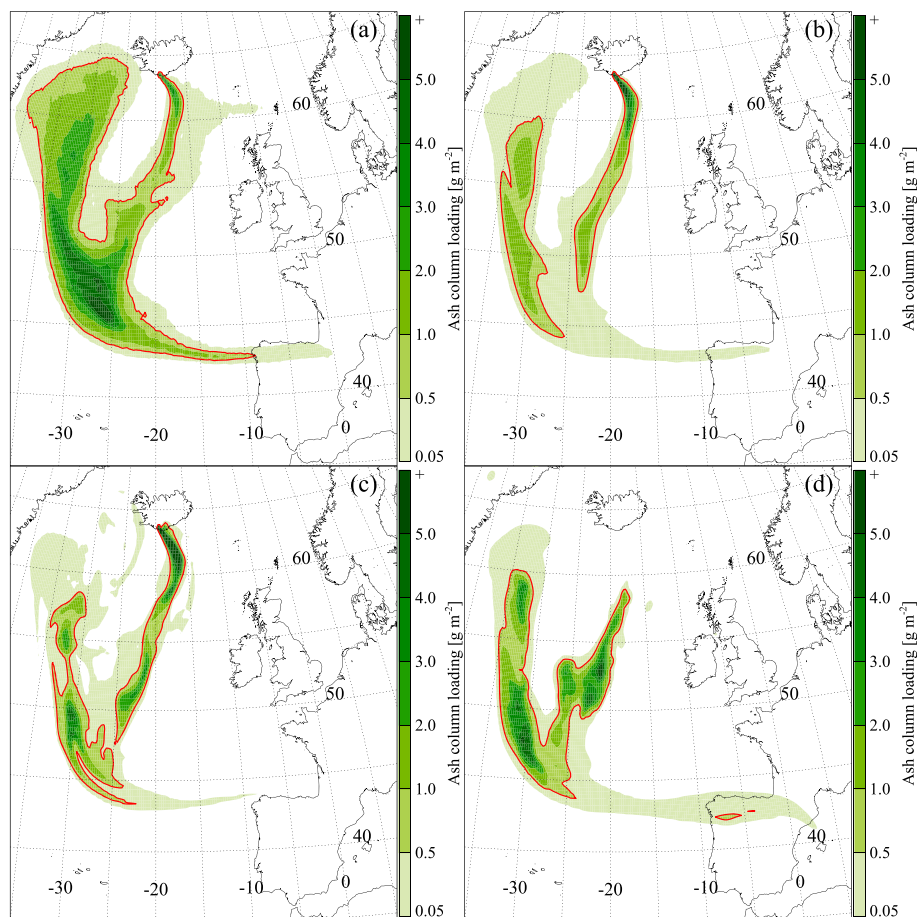


Figure 3. Ash column loading forecasts for 0900 UTC 8 May 2010 described in Table 2: (a) NAME best guess, (b) NAME a posteriori, (c) FLEXPART a posteriori, and (d) NAME data insertion forecast LM(5) 1.0 km; a combination of SR simulations initialized using retrievals R1–R5 (see Table 1) assuming a 1.0 km thick ash layer. Red lines show the outlines of objects identified for the SAL analysis.

Spain above that threshold, although this is shifted slightly to the south of the observed filament, while the other forecasts placed very little, if any, ash above that threshold over Spain.

Vertical cross sections of the modeled ash concentrations over northern Spain at 43°N between -10 and 0° E (with 200 m vertical resolution) are shown in Figure 4. Figure 4a shows that the NAME best guess ash layer is centered around ~ 5 km altitude and is 2–3 km thick. The NAME a posteriori ash layer is similar, centered at ~ 5 km altitude, and is 1–2 km thick (Figure 4b). The FLEXPART forecast (not shown) has a thin layer centered at ~ 2.5 km and another region of ash at ~ 9 km altitude. The concentrations for these layers are very low ($< 4 \mu\text{g m}^{-3}$). The LM(5) 1.0 km forecast layer is centered at ~ 5.5 km altitude and is ~ 2 –3 km thick, with an area of low concentration ash at -2 to 0° E, ~ 8 –11 km altitude (Figure 4c). The 1D-Var retrieved ash cloud height values (with the β ratio test turned off) at this latitude are shown as black asterisks. This shows that the three NAME ash layers and the retrieved ash cloud height are in reasonable agreement over this region.

The difference in ash position across Spain between the NAME and FLEXPART a posteriori forecasts may be due to the accuracy of the MetUM and ECMWF wind fields at the altitudes at which the ash is transported. There are differences in the derived ESPs which could also be causing this discrepancy. These differences are due to the inversion algorithm finding different optimal ESPs based on the different VATDMs and meteorologies used.

Figure 5 shows a series of SR forecasts with the observed ash clouds that were used to initialize those simulations outlined in red. It is apparent from Figure 5a that much of the ash cloud coverage over Spain in the LM forecast (Figure 3d), specifically at the eastern edge, has been contributed by SR1. The SR1 forecast does not, however, include ash in the region 47 to 60° N, -25 to -15° E, which is seen in the observations, because the

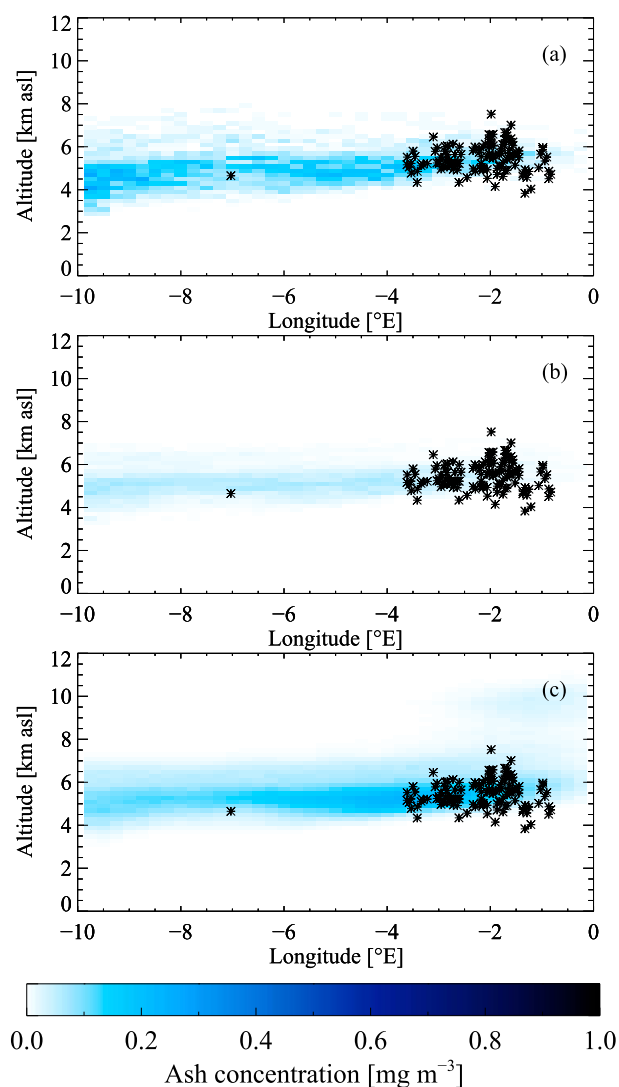


Figure 4. Cross sections of modeled ash concentrations at 0900 UTC 8 May 2010, 43°N over Spain. (a) NAME best guess forecast, (b) NAME a posteriori forecast, (c) NAME data insertion forecast LM(5) 1.0 km shown in Figure 3d. 1D-Var retrieved ash cloud heights (with the β ratio test turned off) from between 42.95 and 43.05°N are plotted with black asterisks.

ash had not yet been emitted at the time of the R1 retrieval. This shows that in this case, it has been necessary to use retrieval data from different times to create the LM forecast.

Figure 6 shows the three alternative data insertion forecast methods. Figure 6a shows the weighted average of the same five simulations used to create LM(5) 1.0 km, Figure 6b is the mean of the same five simulations, and in Figure 6c preference has been given to recent, nonzero values. The weighted average and mean fields have generally lower ACL values and are spread wider than the recent value forecast field. Figure 6c is the most similar to the observation for that time (Figure 2b), as would be expected. None of the fields are as spread as the maximum value method (Figure 3d) nor do they show ash $>0.05 \text{ g m}^{-2}$ (the lower contour level in the figure) as far east in the filament over northern Spain. As the ash in the tip of this filament has been shown to mainly originate from the SR1 simulation, creating the forecasts using the mean of the simulations dilutes the ash from SR1 in that region. In the recent value method, much of the ash in this region is below the range shown in the plot due to the influence of the most recent simulations, which have very low ACLs here.

The data insertion forecasts do not include newly erupted ash south of Iceland, probably for two reasons. First, there is no constant ash injection from the volcano vent included in the data insertion method; therefore, newly erupted ash will only contribute if new images are used. Radar and webcam observations of the

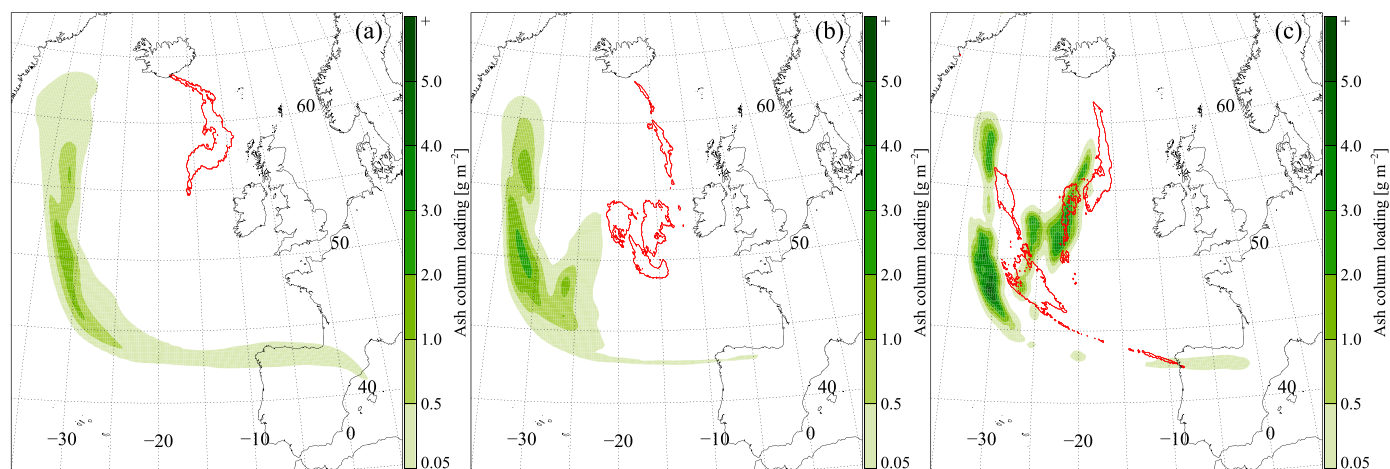


Figure 5. Eyjafjallajökull ash column loading SR data insertion forecasts for 0900 UTC 8 May (filled contours) using retrievals from different times (from Figure 1, outlined in red) assuming a 1.0 km thick ash layer. (a) SR1 forecast initialized using R1: 1900 UTC 6 May 2010, (b) SR2 forecast using R2: 0700 UTC 7 May 2010, and (c) SR4 forecast using R4: 0000 UTC 8 May 2010.

eruption plume reported in *Arason et al.* [2011] show that the plume height was ~4–5 km from 6 to 8 May, indicating that the eruption rate was fairly continuous over that time. Second, the newly erupted ash is not fully visible in a number of the retrievals. This may be due to a high optical thickness of the newly erupted ash cloud making it opaque and hence undetectable using the BTM method; the difference between the two channels tends toward zero for increasingly opaque clouds [*Prata and Grant, 2001*]. Also, it is possible that water is present in the plume, that there is meteorological cloud obscuring the ash signal, or that there is a high concentration of large particles making the detection difficult [*Rose et al., 1995; Stevenson et al., 2015*].

3.2.2. Quantitative Assessment
3.2.2.1. SAL

SAL is performed on the model output for 0900 UTC 8 May 2010. The “observed” field compared here comprises the maximum values over the previous 1 h from each pixel retrieved using 1D-Var. The aim of this is to create a slightly smoother observation field because the retrievals are generally patchier than modeled fields. This period coincides with the 1 h averaging time of the NAME forecasts. Observed and modeled fields must be compared on the same grid; therefore, the “smoothed” observed field is mapped onto the model grids using nearest neighbor interpolation.

Absolute SAL scores are shown in Figure 7a. The four model runs with lowest (best) absolute SAL scores are SR5 1.0 km, SR5 0.5 km, SR6 0.5 km, and SR6 1.0 km (see section 2.2 for a description of the notation). Hence,

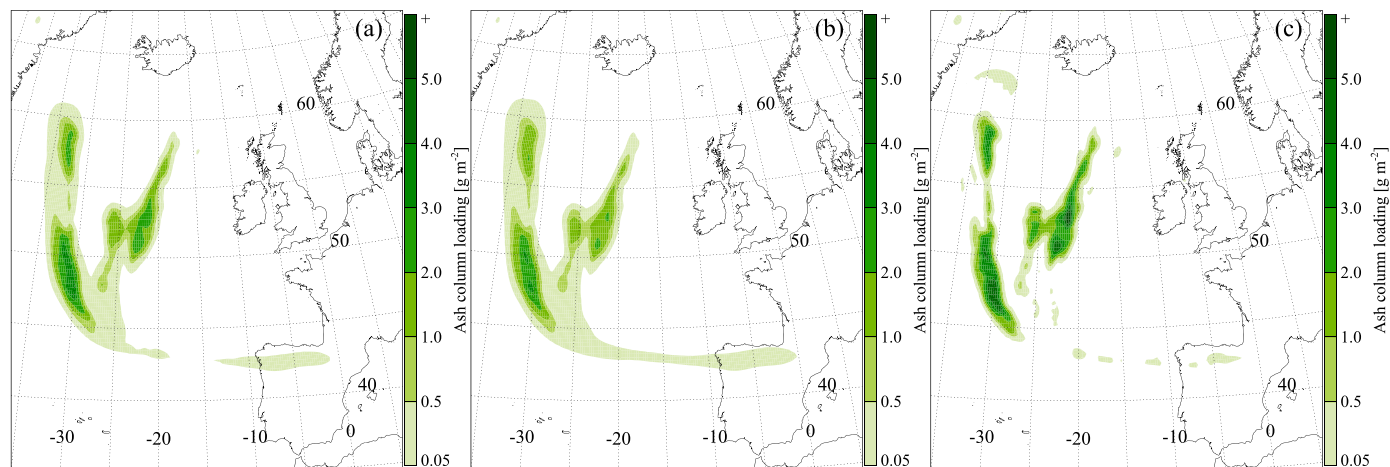


Figure 6. Ash column loading forecasts for 0900 UTC 8 May 2010 created by combining simulations SR1–SR5 using the (a) weighted average, (b) mean, and (c) recent value methods.

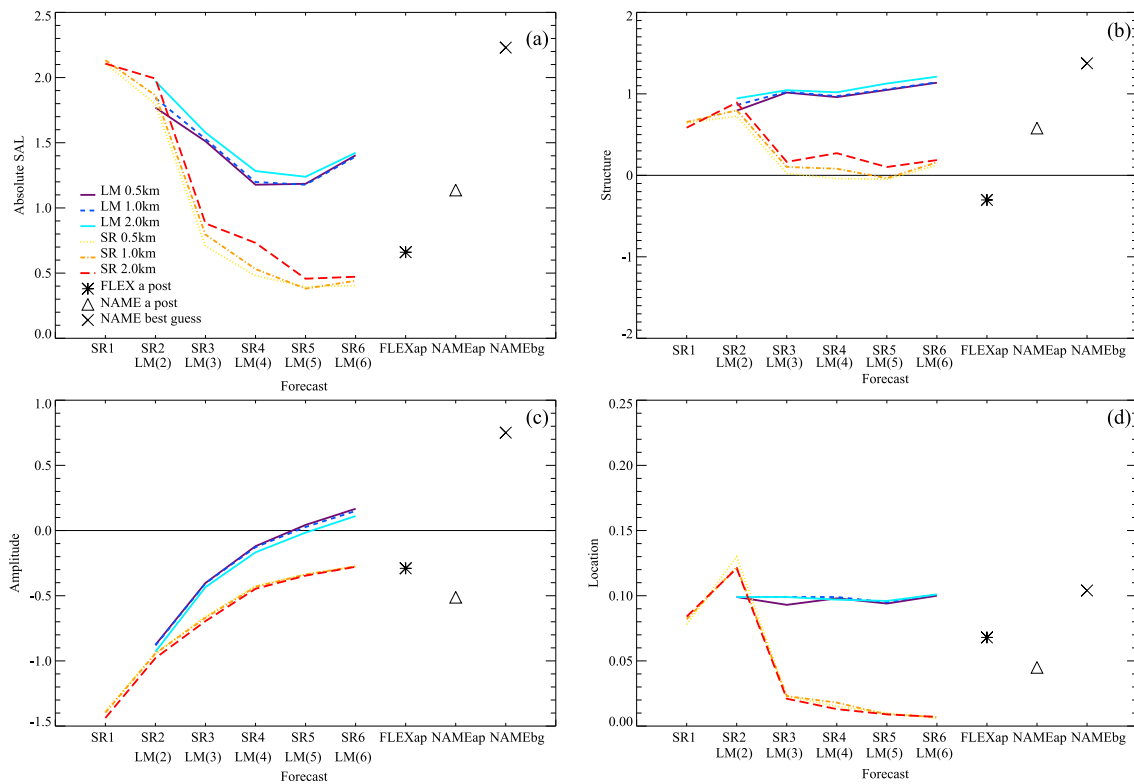


Figure 7. SAL scores for the smoothed observed field versus modeled fields at 0900 UTC 8 May 2010. FLEXap, NAMEap, and NAMEBg are the FLEXPART and NAME a posteriori forecasts and NAME best guess forecast, respectively. (a) Absolute SAL scores, (b) structure, (c) amplitude and (d) location.

according to this measure, the SR simulations with short lead times are the best performing of the forecasts studied here. The highest (worst) overall SAL scores were achieved by SR1 0.5 km, SR1 2.0 km, SR1 1.0 km, and NAME best guess. Hence, using later observations to initialize the model in the SR forecasts generally improves the SAL score, i.e., forecasts get worse with longer forecast times (as would be expected). Figure 7a shows that assuming a thinner layer depth for the entire ash cloud during insertion gives a slightly better score for the SR forecasts, indicating that 0.5 km may be a reasonable assumption for the layer thickness at those times. However, this increase in skill is very small.

Forecasts with the best S scores (Figure 7b) are SR5 0.5 km, SR4 0.5 km, SR5 1.0 km, and SR3 0.5 km. Negative S values are achieved by the three former of these forecasts and the FLEXPART a posteriori forecast. This indicates that those forecasts have predicted fields that cover too small an area and/or are too peaked. It can be seen in Figure 3c that the FLEXPART a posteriori field is peaked immediately south of Iceland. The NAME best guess and NAME a posteriori fields in Figures 3a and 3b, respectively, are also peaked at that location, but these and all other forecasts have positive S values, indicating they are too spread and/or flat. The spread of the fields in Figures 3a and 3b compensate for the peaked values, giving a positive S score. It can also be seen that the NAME data insertion LM(5) 1.0 km forecast (Figure 3d) is more spread and has larger objects than the observed field, resulting in a positive S score. All of the model fields contain fewer objects than the observation field (not shown).

Forecasts with the best A scores (Figure 7c) are LM(5) 2.0 km, LM(5) 1.0 km (Figure 3d), LM(5) 0.5 km, and LM(6) 2.0 km. LM forecasts show a general improvement in the A score compared with the SR forecasts but have a worse overall SAL score due to the large S component, which is probably contributed by earlier retrievals. To test whether the poor S score is contributed by earlier retrievals, simulations SR1, SR2, and SR3 are removed from the LM(5) 1.0 km forecast (Table 3). This gives improved S scores for the three forecasts tested, LM(SR2–SR5) gives an improved A score, and LM(SR3–SR5) and LM(SR4–SR5) give improved L scores. All absolute SAL scores are improved by excluding the simulations initialized using the earliest retrievals from the forecast. However, excluding these also removes some of the ash over northern Spain (not shown).

Table 3. SAL Scores for LM Forecasts (Created Using the Maximum Method), Combining Simulations Initialized From Different Retrievals Are Shown in the Top Three Rows^a

Forecast	S	A	L	SAL
LM(SR2–SR5)	1.048	0.015	0.105	1.168
LM(SR3–SR5)	0.149	-0.112	0.012	0.273
LM(SR4–SR5)	0.075	-0.202	0.012	0.289
LM(5)	1.056	0.027	0.095	1.178
Weighted average	0.120	-0.619	0.016	0.755
Mean	0.137	-0.779	0.021	0.937
Recent	-0.035	-0.337	0.009	0.381

^aSee Table 1 for retrieval times. LM(5) (Figure 3d) is included for comparison. The bottom three rows show SAL scores for the weighted average, mean, and recent value forecasts combining simulations SR1–SR5. In all simulations a 1.0 km thick ash layer is assumed. See text for details.

For the LM forecasts in Figure 7c the best A score is achieved using a combination of SR1–SR5. The NAME best guess A score indicates an overprediction of ash mass. This could be due to the satellite sensor failing to detect ash where the model has included it or too much ash having been released in the model. The combination of positive S scores and negative A scores for many of the other forecasts could indicate that there is too much diffusion in NAME, ash having been inserted at the wrong height, or transport errors leading to spread out clouds with large objects compared with the observation. NAME output represents an ensemble average over unresolved motions, and this is likely to be wider than an instantaneous ash plume. The overspreading leads to large areas of ACL below the lower threshold used here (i.e., ash outside the red outlines in Figure 3). That ash is not included in the SAL calculations, reducing the modeled ash mass available for comparison, and making the A score more negative. For the LM forecasts, the high S scores could be due to the same reasons mentioned above or due to the number of simulations contributing to the forecast leading to a wider ash cloud, as previously discussed.

The best L scores (Figure 7d) are achieved by SR runs utilizing retrievals from 3–6 h previous to the forecast time: SR6 0.5 km, SR6 2.0 km, SR6 1.0 km, SR5 2.0 km, and SR5 1.0 km. This is likely to be because observed data have been propagated by the model for a short period, and the ash detection is consistent over that time. The worst L scores are achieved by NAME best guess, SR2 2.0 km, SR2 1.0 km, and SR2 0.5 km. This may indicate that R2 was a poor retrieval or that the assumptions made during insertion into NAME were unsuitable for that time.

The SAL scores for the weighted average, mean, and recent value data insertion forecast methods are shown in Table 3. This shows that these three alternative methods score better overall than the LM(5) 1.0 km forecast (created using the maximum value method), due to smaller S and L scores, but do not improve on the A score. This is likely to be because the fields for these methods do not contain as much ash around the peripheries of the ash clouds as the LM(5) 1.0 km forecast, reducing the area of ash exceeding the 0.5 g m⁻² ACL threshold for which the S component is calculated. The recent value method has the best SAL scores of the three methods, which are the same as those for SR5 1.0 km.

3.2.2.2. FMS

Using the same model and observation fields as for SAL above, the Figure of Merit in Space is also calculated. The result is shown in Figure 8. This indicates that the SR simulations tend to have the best degree of spatial overlap with the observations, generally improving as the time since model initialization decreases. However, SR2 simulations show worse skill than all other data insertion simulations, again indicating that SR2 may be a poor simulation. The LM forecasts perform similarly to the a posteriori forecasts. The reason for the LM forecasts generally scoring worse than the SR forecasts is attributable to wider forecast fields resulting from combining a number of simulations. The NAME best guess forecast has the lowest score, as it has the widest spatial ash coverage (and the least input of observed data).

One reason for the general trend in increasing simulation skill with decreasing time between the last retrieval used and the forecast time is that the most recent inserted retrieval has been propagated by the model for a shorter period. That retrieval and the one used for comparison should have a high correlation, i.e., similar parts of the ash cloud are likely to be observed/observed, and hence, simulations using the later retrievals

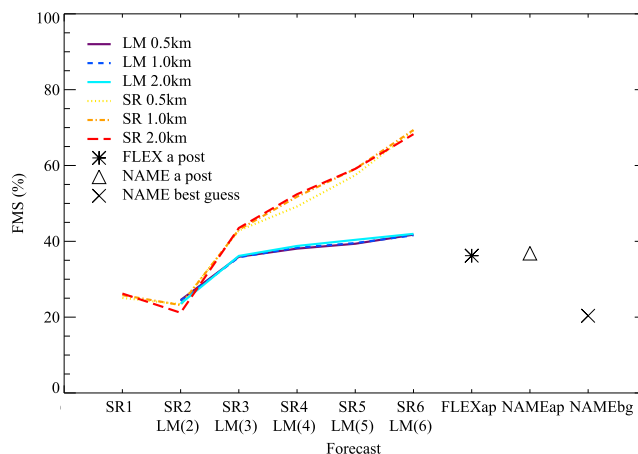


Figure 8. Figure of Merit in Space scores for the smoothed observed field versus modeled fields at 0900 UTC 8 May 2010. The nomenclature is the same as Figure 7.

should score better. Another possible reason is that, in this case, the ash cloud may be better observed in the later retrievals. A good earlier retrieval (observing most of the ash cloud and with very little cloud to obscure the observation) may give a better forecast than a later, poor retrieval. This may be the reason for the SR2 simulations generally scoring worse than the SR1 simulations.

The FMS scores for the weighted average, mean, and recent value forecast methods using SR1–SR5 are 54%, 52%, and 59% (same as SR5 1.0 km), respectively. These are higher scores than those achieved by the LM(5) 1.0 km forecast. Again, this is likely to be for the same reason as for the increase in skill for the S component of SAL; the fields for these methods do not contain as much ash as the LM(5) 1.0 km forecast, reducing the area of ash exceeding the 0.5 g m^{-2} ACL threshold for which the FMS score is calculated.

3.3. Applicability

For near-instantaneous eruptions, one or two good retrievals (and meteorological data) may be sufficient to simulate ash transport at long-range. For long-lived eruptions, such as Eyjafjallajökull, a good quality recent retrieval may be sufficient to predict ash transport in the short term, but to predict at a number of times over a longer range, a number of retrievals are required to maintain the addition of ash mass. For example, Figure 5a shows that a single retrieval from 38 h previous to the forecast time gives model output in which a large amount of ash is missing due to the continued eruption. Frequent updates would also be useful for the short emission to constrain advection errors.

To further demonstrate the applicability of the data insertion method, the shorter eruption of Grímsvötn (Iceland, May 2011) has been modeled, and the ACL results are shown in Figure 9a, with the corresponding SEVIRI BT image shown in Figure 9b. In Figures 9c and 9d, part of the April 2010 phase of the Eyjafjallajökull eruption is shown. In both cases, four SR simulations assuming 1.0 km thick ash clouds have been combined to create LM forecasts (using the maximum value method) at 1200 UTC 24 May 2011 and 1200 UTC 16 April 2010, respectively. Both forecasts are shown 24 h after the first retrieval was inserted and 6 h after the last. In a study of the effects of clouds on detection of ash in simulated SEVIRI imagery using the BT image method assuming a detection threshold of 0.2 g m^{-2} ACL, *Kylling et al.* [2015] found that on average clouds reduced the detection of ash affected pixels by $\sim 12\%$ in the period 14–21 April 2010 (i.e. $\sim 25\%$ were detected in cloudless scenes and $\sim 13\%$ in cloudy scenes) and up to 40% in some images for the Eyjafjallajökull eruption. The effects of clouds on images from 15 April 2010 to 6–8 May 2010 were found to be small or negligible but were larger on 16 April 2010. For the eruption of Grímsvötn in May 2011, fewer ash contaminated pixels were detected and the average reduction in detection of those pixels due to clouds was $\sim 6\%$ (i.e. from 10.0% of pixels detected in cloudless scenes to 3.6% in those with clouds).

In the early phase of the Grímsvötn eruption in May 2011, SO_2 traveled northward [*Kylling et al.*, 2015], and a small amount of ash was detected north of Iceland in SEVIRI imagery by *Cooke et al.* [2014]. During the later phase, differential settling and different emission altitudes of ash and SO_2 particles caused SO_2 to travel toward the north and most of the ash southeastward [*Moxnes et al.*, 2014]. As retrievals of the early phase are not included in this study, ash only traveled in a southeasterly direction in the data insertion simulations used

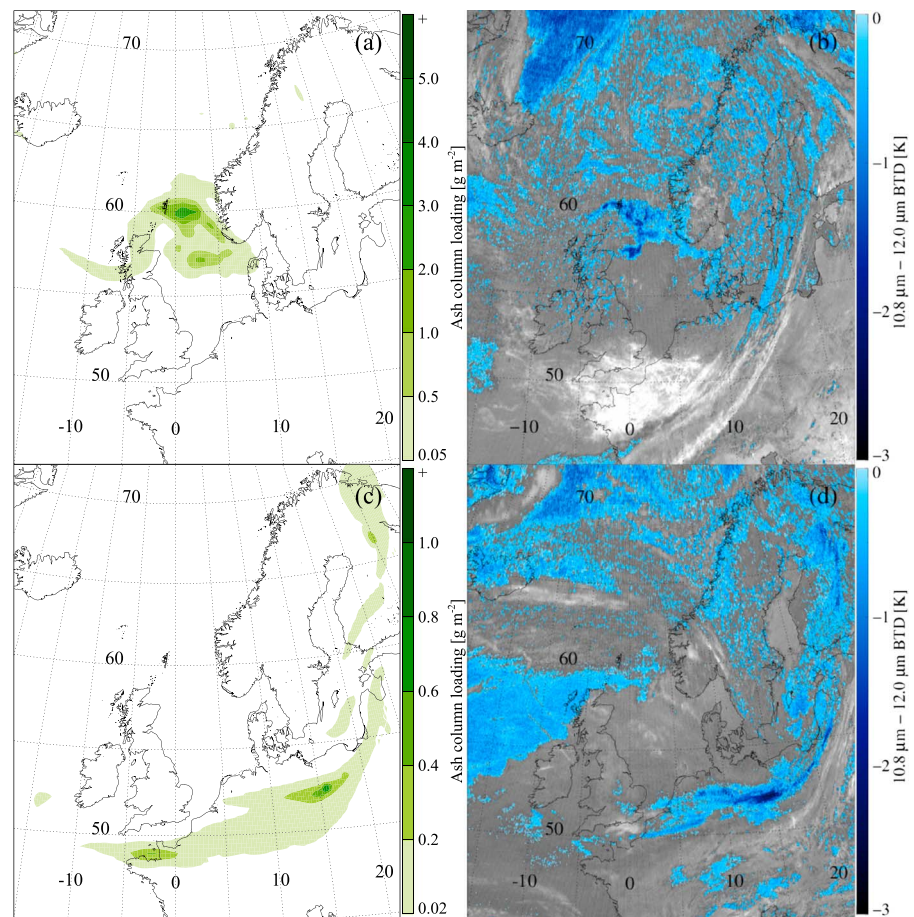


Figure 9. (a) Grímsvötn ash cloud at 1200 UTC 24 May 2011 modeled using the layered model (LM) data insertion method, (b) the corresponding BTM SEVIRI image for that time, and (c and d) the same as Figures 9a and 9b, respectively, but for the Eyjafjallajökull ash cloud at 1200 UTC 16 April 2010.

to create the forecast shown in Figure 9a. Qualitatively, the position of the observed ash cloud, which shows as a dark blue region over the North Sea between Scotland and Norway in Figure 9b, agrees well with the forecast in Figure 9a.

For the April 2010 Eyjafjallajökull case in Figure 9c (note a different color scale is used), it can be seen that the transport pattern of the ash is similar to the pattern of negative BTM in Figure 9d. However, as meteorological cloud may have significantly reduced the ash detection on 16 April, it is likely that additional information is required to capture the majority of ash in this case.

3.4. Advantages and Limitations

While simpler than other data assimilation methods, data insertion is more complex than methods based solely on eruption height, and it introduces different issues. The data insertion method is strongly dependent on good satellite observations. Compared to methods such as inversion modeling, which accounts for errors in the observations and the model forecast, this method is largely data centered and does not consider errors in the satellite retrievals. Ash will be missed from the simulations if it has erupted since the last retrieval, if it is consistently obscured by cloud, or if it is near to the volcano, where it may be undetected either because it is too optically thick or has a high water content (see Figure 3d). The issue of missing ash close to the source could be overcome by including ash initialized from the volcano location in NAME. This is beyond the scope of this paper but will be addressed in future work.

While missing ash is a major limitation, especially close to the source where the concentration could be very high, the data insertion method only relies on images of downwind ash and does not make any assumptions about the source (e.g., eruption plume height, MER and fine ash fraction) nor does it require modeling of

near-source processes (e.g., particle aggregation, transport by gravity currents, and others). This means that it could be applicable to volcanoes in remote areas which are not well observed via ground-based monitoring. It is reasonable to assume that this method could also be applied to atmospheric SO₂ transport as long as good satellite retrievals can be performed, but this has not been attempted here.

Another limitation of the data insertion method is that it requires estimation of the layer thickness, vertical distribution of ash, and PSD, which may be difficult to obtain in an operational setting. Where these parameters are unknown, it is suggested that a set of default values could be used. However, the use of a set of default values could introduce large errors in the forecast position. Also, these are properties that can generally only be measured for a point location or along a trajectory (e.g., ground based lidar, research flights and others) and are not available for the entire ash cloud. A small number of observations (if available) must, therefore, be applied to a larger region of the ash cloud. The data insertion method does, though, have the advantage of enabling the modeler to update the forecast with new retrievals or different observations as they become available. Updating the model in this way could allow inclusion of observed features even if they are not resolved by the model, and if a good retrieval close to the forecast time is available, model transport errors may be relatively small.

In the NAME configuration used here, vertical distributions other than uniform or Gaussian would require the separation of sources into a number of layers in order to represent the vertical ash profile in the NAME input file. This would involve some simplifications of the profile and increase the number of sources released by the number of vertical layers used. Although beyond the scope of this paper, a next step would be to use Cloud-Aerosol Lidar with Orthogonal Polarization data to better characterize the vertical distribution of the ash layer and determine its effect on the simulated ash cloud.

The manner in which simulations are used to create the forecast will have a significant effect on the result. As discussed above, the LM maximum value method is conservative and is used to overcome obscured ash in one or more retrievals. The LM method does not remove false alarms which could be present due to errors in the observation, the translation from observation to model input format, the dispersion model itself, or the meteorological data used to drive the model. The weighted average method places more trust in newer simulations initialized from later retrievals, which are likely to contain fewer cumulative transport errors than earlier ones, but also does not allow distinction between good and bad retrievals. By taking the mean, we give all simulations equal weight, but a simulation in which ash is obscured by cloud would reduce the column load in the forecast. The recent value method gives the same results as simply taking the SR5 simulation here, although this may not always be the case.

4. Conclusions

We discuss the application of a data insertion method, where the NAME dispersion model is initialized from satellite retrievals of downwind ash clouds, using the 8 May 2010 ash cloud from the eruption of Eyjafjallajökull as the main case study. The model results are compared against other established methods and satellite imagery. We also briefly discuss the application of this method to the April 2010 phase of the Eyjafjallajökull eruption and the May 2011 Grímsvötn eruption.

For the 8 May 2010 Eyjafjallajökull case, the data insertion forecasts perform as well, according to the SAL and FMS metrics, as the NAME best guess forecast, which is initialized from the volcanic source alone and includes measured eruption plume height data in its ESPs.

Layered model (LM) data insertion forecasts score worse overall than the single retrieval (SR) forecasts using SAL and FMS. Poorer S, L, and FMS scores reflect the broader, more extensive ash clouds that are generally provided by the LM forecasts compared with the observations. The LM forecasts thus provide a more conservative estimate designed to ensure most observable ash is captured. The weighted average, mean, and recent value forecast methods give better absolute SAL and FMS scores than the LM method, for the same reasons as above. However, the LM method achieves significantly better A scores and produces a more similar pattern of ash over northern Spain on 8 May 2010 to that observed in SEVIRI BTM imagery.

For satellite images in which a lot of ash is obscured, the data insertion method alone is likely to be insufficient, and more information such as a cloud mask or a combination of data insertion and initialization from the volcano source may be necessary.

Finally, it is noted that while observations are key in evaluating model output, it is important to keep in mind what we would like to achieve by simulating the transport of volcanic ash. Would we prefer a forecast that scores well against observations or a conservative one which scores worse, but which includes ash that may not be captured by observations? Either way, the limitations of quantitative metrics and those of the observations themselves should be acknowledged in any forecast evaluation.

Acknowledgments

Kate Wilkins is supported by the Natural Environment Research Council [Consortium on Risk in the Environment: Diagnostics, Integration, Benchmarking, Learning and Elicitation (CREDIBLE) and Cooperative Awards in Science and Technology (CASE, Met Office); grant reference NE/J017450/1]. We acknowledge use of the NAME atmospheric dispersion model, associated NWP meteorological data sets and 1D-Var code made available by the Met Office, and also EUMETSAT for the provision of SEVIRI data via the Earth Observation Portal (available upon registration). Data used to generate figures, graphs, plots, and tables are freely available via contacting the lead author: kate.wilkins@bristol.ac.uk. Thank you to Pete Francis for the help with implementing the 1D-Var method used here, to Heini Wernli for use of the SAL code, and to Shona Mackie, Natalie Harvey, and Luke Western for helpful comments on a draft of this paper. Thank you also to the reviewers for helpful comments.

References

- Arason, P., G. N. Petersen, and H. Björnsson (2011), Observations of the altitude of the volcanic plume during the eruption of Eyjafjallajökull, April–May 2010, *Earth Syst. Sci. Data*, 3(1), 9–17, doi:10.5194/essd-3-9-2011.
- Balkanski, Y., M. Schulz, T. Claquin, and S. Guibert (2007), Reevaluation of Mineral aerosol radiative forcings suggests a better agreement with satellite and AERONET data, *Atmos. Chem. Phys.*, 7, 81–95.
- Boichu, M., L. Clarisse, D. Khvorostyanov, and C. Clerbaux (2014), Improving volcanic sulfur dioxide cloud dispersal forecasts by progressive assimilation of satellite observations, *Geophys. Res. Lett.*, 41, 2637–2643, doi:10.1002/2014GL059496.
- Bonadonna, C., and B. F. Houghton (2005), Total grain-size distribution and volume of tephra-fall deposits, *Bull. Volcanol.*, 67, 441–456, doi:10.1007/s00445-004-0386-2.
- Bursik, M. I., S. N. Carey, and R. S. J. Sparks (1992), A gravity current mode for the May 18, 1980 Mount St. Helens plume, *Geophys. Res. Lett.*, 19(16), 1663–1666.
- Cooke, M. C., P. N. Francis, S. Millington, R. Saunders, and C. Witham (2014), Detection of the Grímsvötn 2011 volcanic eruption plumes using infrared satellite measurements, *Atmos. Sci. Lett.*, 15, 321–327, doi:10.1002/asl2.506.
- Costa, A., A. Folch, and G. Macedonio (2013), Density-driven transport in the umbrella region of volcanic clouds: Implications for tephra dispersion models, *Geophys. Res. Lett.*, 40, 4823–4827, doi:10.1002/grl.50942.
- Dacre, H. F. (2011), A new method for evaluating regional air quality forecasts, *Atmos. Environ.*, 45(4), 993–1002, doi:10.1016/j.atmosenv.2010.10.048.
- Dacre, H. F., et al. (2011), Evaluating the structure and magnitude of the ash plume during the initial phase of the 2010 Eyjafjallajökull eruption using lidar observations and NAME simulations, *J. Geophys. Res.*, 116, D00U03, doi:10.1029/2011JD015608.
- Dacre, H. F., A. L. M. Grant, and B. T. Johnson (2013), Aircraft observations and model simulations of concentration and particle size distribution in the Eyjafjallajökull volcanic ash cloud, *Atmos. Chem. Phys.*, 13(3), 1277–1291, doi:10.5194/acp-13-1277-2013.
- Dacre, H. F., A. L. M. Grant, N. J. Harvey, D. J. Thomson, H. N. Webster, and F. Marengo (2015), Volcanic ash layer depth: Processes and mechanisms, *Geophys. Res. Lett.*, 42, 637–645, doi:10.1002/2014GL062454.
- Devenish, B. J., P. N. Francis, B. T. Johnson, R. S. J. Sparks, and D. J. Thomson (2012a), Sensitivity analysis of dispersion modeling of volcanic ash from Eyjafjallajökull in May 2010, *J. Geophys. Res.*, 117, D00U21, doi:10.1029/2011JD016782.
- Devenish, B. J., D. J. Thomson, F. Marengo, S. J. Leadbetter, H. Ricketts, and H. F. Dacre (2012b), A study of the arrival over the United Kingdom in April 2010 of the Eyjafjallajökull ash cloud using ground-based lidar and numerical simulations, *Atmos. Environ.*, 48, 152–164, doi:10.1016/j.atmosenv.2011.06.033.
- Eckhardt, S., A. J. Prata, P. Seibert, K. Stebel, and A. Stohl (2008), Estimation of the vertical profile of sulfur dioxide injection into the atmosphere by a volcanic eruption using satellite column measurements and inverse transport modeling, *Atmos. Chem. Phys.*, 8, 3881–3897.
- Folch, A., A. Costa, and S. Basart (2012), Validation of the FALL3D ash dispersion model using observations of the 2010 Eyjafjallajökull volcanic ash clouds, *Atmos. Environ.*, 48, 165–183, doi:10.1016/j.atmosenv.2011.06.072.
- Francis, P. N., M. C. Cooke, and R. W. Saunders (2012), Retrieval of physical properties of volcanic ash using Meteosat: A case study from the 2010 Eyjafjallajökull eruption, *J. Geophys. Res.*, 117, D00U09, doi:10.1029/2011JD016788.
- Galmarini, S., F. Bonnardot, A. Jones, S. Potempski, L. Robertson, and M. Martet (2010), Multi-model vs. EPS-based ensemble atmospheric dispersion simulations: A quantitative assessment on the ETEX-1 tracer experiment case, *Atmos. Environ.*, 44(29), 3558–3567, doi:10.1016/j.atmosenv.2010.06.003.
- Grant, A. L. M., H. F. Dacre, D. J. Thomson, and F. Marengo (2012), Horizontal and vertical structure of the Eyjafjallajökull ash cloud over the UK: A comparison of airborne lidar observations and simulations, *Atmos. Chem. Phys.*, 12(21), 10,145–10,159, doi:10.5194/acp-12-10145-2012.
- Hobbs, P. V., L. F. Radke, J. H. Lyons, R. J. Ferek, D. J. Coffman, and T. J. Casadevall (1991), Airborne measurements of particle and gas emissions from the 1990 volcanic eruptions of Mount Redoubt, *J. Geophys. Res.*, 96(D10), 735–752.
- International Air Transport Association (2011), Volcanic Ash. Ashes to Ashes—What has been learned from recent volcanic eruptions and where does the industry go from here? [Available at <http://airlines.iata.org/analysis/volcanic-ash>, accessed: 22/12/2015.]
- Johnson, B. T., et al. (2012), In situ observations of volcanic ash clouds from the FAAM aircraft during the eruption of Eyjafjallajökull in 2010, *J. Geophys. Res.*, 117, D00U24, doi:10.1029/2011JD016760.
- Jones, A. R., D. J. Thomson, M. C. Hort, and B. J. Devenish (2007), The U.K. Met Office's Next-Generation Atmospheric Dispersion Model, NAME III, in *Air Pollution Modeling and Its Application XVII*, vol. 6, edited by C. Borrego and A.-L. Norman, pp. 580–589, Springer, New York.
- Kristiansen, N. I., et al. (2010), Remote sensing and inverse transport modeling of the Kasatochi eruption sulfur dioxide cloud, *J. Geophys. Res.*, 115, D00L16, doi:10.1029/2009JD013286.
- Kristiansen, N. I., et al. (2012), Performance assessment of a volcanic ash transport model mini-ensemble used for inverse modeling of the 2010 Eyjafjallajökull eruption, *J. Geophys. Res.*, 117, D00U11, doi:10.1029/2011JD016844.
- Kristiansen, N. I., A. J. Prata, A. Stohl, and S. A. Carn (2015), Stratospheric volcanic ash emissions from the 13 February 2014 Kelut eruption, *Geophys. Res. Lett.*, 42, 588–596, doi:10.1002/2014GL062307.
- Kylling, A., N. Kristiansen, A. Stohl, R. Buras-Schnell, C. Emde, and J. Gasteiger (2015), A model sensitivity study of the impact of clouds on satellite detection and retrieval of volcanic ash, *Atmos. Meas. Tech.*, 8(5), 1935–1949, doi:10.5194/amt-8-1935-2015.
- Marengo, F., B. T. Johnson, K. Turnbull, S. M. Newman, J. M. Haywood, H. N. Webster, and H. Ricketts (2011), Airborne lidar observations of the 2010 Eyjafjallajökull volcanic ash plume, *J. Geophys. Res.*, 116, D00U05, doi:10.1029/2011JD016396.
- Mastin, L. G. (2007), A user-friendly one-dimensional model for wet volcanic plumes, *Geochem. Geophys. Geosyst.*, 8, Q03014, doi:10.1029/2006GC001455.
- Mastin, L. G., et al. (2009), A multidisciplinary effort to assign realistic source parameters to models of volcanic ash-cloud transport and dispersion during eruptions, *J. Volcanol. Geotherm. Res.*, 186(1–2), 10–21, doi:10.1016/j.jvolgeores.2009.01.008.

- Millington, S. C., R. W. Saunders, P. N. Francis, and H. N. Webster (2012), Simulated volcanic ash imagery: A method to compare NAME ash concentration forecasts with SEVIRI imagery for the Eyjafjallajökull eruption in 2010, *J. Geophys. Res.*, *117*, D00U17, doi:10.1029/2011JD016770.
- Mosca, S., G. Graziani, W. Klug, R. Bellasio, and R. Bianconi (1998), A statistical methodology for the evaluation of long-range dispersion models: An application to the ETEX exercise, *Atmos. Environ.*, *32*(24), 4307–4324.
- Moxnes, E. D., N. I. Kristiansen, A. Stohl, L. Clarisse, A. Durant, K. Weber, and A. Vogel (2014), Separation of ash and sulfur dioxide during the 2011 Grimsvötn eruption, *J. Geophys. Res. Atmos.*, *119*, 7477–7501, doi:10.1002/2013JD021272.
- Pavlonis, M. J. (2010), Advances in extracting cloud composition information from spaceborne infrared radiances—A robust alternative to brightness temperatures. Part I: Theory, *J. Appl. Meteorol. Climatol.*, *49*, 1992–2012, doi:10.1175/2010JAMC2433.1.
- Pavlonis, M. J., A. K. Heidinger, and J. Sieglaff (2013), Automated retrievals of volcanic ash and dust cloud properties from upwelling infrared measurements, *J. Geophys. Res. Atmos.*, *118*, 1436–1458, doi:10.1002/jgrd.50173.
- Pelley, R. E., M. C. Cooke, A. J. Manning, D. J. Thomson, C. S. Witham, and M. C. Hort (2015), Initial Implementation of an Inversion Technique for Estimating Volcanic Ash Source Parameters in Near Real time using Satellite Retrievals Forecasting Research Technical Report 604, Met Office, Exeter.
- Pollack, J. B., O. B. Toon, and B. N. Khare (1973), Optical properties of some rocks and glasses, *Icarus*, *19*, 372–389, doi:10.1016/0019-1035(73)90115-2.
- Prata, A. J., and I. F. Grant (2001), Retrieval of microphysical and morphological properties of volcanic ash plumes from satellite data: Application to Mt. Ruapehu, New Zealand, *Q. J. R. Meteorol. Soc.*, *127*, 2153–2179.
- Prata, A. J., and A. T. Prata (2012), Eyjafjallajökull volcanic ash concentrations determined using Spin Enhanced Visible and Infrared Imager measurements, *J. Geophys. Res.*, *117*, D00U23, doi:10.1029/2011JD016800.
- Rose, W. I., D. J. Delene, D. J. Schneider, G. J. S. Bluth, A. J. Krueger, I. E. Sprod, C. McKee, H. L. Davies, and G. G. J. Ernst (1995), Ice in the 1994 Rabaul eruption cloud: Implications for volcano hazard and atmospheric effects, *Nature*, *375*, 477–479.
- Sparks, R. S. J. (1986), The dimensions and dynamics of volcanic eruption columns, *Bull. Volcanol.*, *48*, 3–15.
- Sparks, R. S. J., M. I. Bursik, S. N. Carey, J. S. Gilbert, L. S. Glaze, H. Sigurdsson, and A. W. Woods (1997), *Volcanic Plumes*, 574 pp., John Wiley, Chichester, U. K.
- Stevenson, J. A., S. C. Millington, F. M. Beckett, G. T. Swindles, and T. Thordarson (2015), Big grains go far: reconciling tephrochronology with atmospheric measurements of volcanic ash, *Atmos. Meas. Tech. Discuss.*, *8*(1), 65–120, doi:10.5194/amtd-8-65-2015.
- Stohl, A., M. Hittenberger, and G. Wotawa (1998), Validation of the lagrangian particle dispersion model FLEXPART against large-scale tracer experiment data, *Atmos. Environ.*, *32*(24), 4245–4264, doi:10.1016/S1352-2310(98)00184-8.
- Stohl, A., A. J. Prata, S. Eckhardt, L. Clarisse, A. Durant, S. Henne, N. I. Kristiansen, and A. Minikin (2011), Determination of time- and height-resolved volcanic ash emissions and their use for quantitative ash dispersion modeling: The 2010 Eyjafjallajökull eruption, *Atmos. Chem. Phys.*, *11*, 4333–4351, doi:10.5194/acp-11-4333-2011.
- Volz, F. E. (1973), Infrared optical constants of ammonium sulfate, sahara dust, volcanic pumice, and flyash, *Appl. Opt.*, *12*(3), 564–568.
- Webster, H. N., et al. (2012), Operational prediction of ash concentrations in the distal volcanic cloud from the 2010 Eyjafjallajökull eruption, *J. Geophys. Res.*, *117*, D00U08, doi:10.1029/2011JD016790.
- Wernli, H., M. Paulat, M. Hagen, and C. Frei (2008), SAL—a novel quality measure for the verification of quantitative precipitation forecasts, *Mon. Weather Rev.*, *136*(11), 4470–4487, doi:10.1175/2008MWR2415.1.
- Wilkins, K. L., S. Mackie, I. M. Watson, H. N. Webster, D. J. Thomson, and H. F. Dacre (2014), Data insertion in volcanic ash cloud forecasting, *Ann. Geophys.*, *57*(2), 1–6, doi:10.4401/ag-6624.
- Woodhouse, M. J., A. J. Hogg, J. C. Phillips, and R. S. J. Sparks (2013), Interaction between volcanic plumes and wind during the 2010 Eyjafjallajökull eruption, Iceland, *J. Geophys. Res. Solid Earth*, *118*, 92–109, doi:10.1029/2012JB009592.