

JGR Atmospheres

RESEARCH ARTICLE

10.1029/2024JD041112

Key Points:

- A volcanic ash retrieval was run on simulated radiances produced for a large range of ash cloud properties
- This constrained limitations on the detection and retrieval's performance at different ash column loadings, heights, and particle sizes
- Significant portions of ash mass present in dispersion model output were shown to be outside the satellite retrieval sensitivity range

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

C. Saint,
cameron.saint@metoffice.gov.uk

Citation:

Saint, C., Beckett, F. M., Dioguardi, F., Kristiansen, N., & Tubbs, R. N. (2024). Using simulated radiances to understand the limitations of satellite-retrieved volcanic ash data and the implications for volcanic ash cloud forecasting. *Journal of Geophysical Research: Atmospheres*, 129, e2024JD041112. <https://doi.org/10.1029/2024JD041112>

Received 14 MAR 2024

Accepted 31 OCT 2024

Author Contributions:

Conceptualization: C. Saint,

F. M. Beckett

Data curation: C. Saint, F. M. Beckett, F. Dioguardi

Formal analysis: C. Saint, F. M. Beckett

Investigation: C. Saint, F. M. Beckett

Methodology: C. Saint, F. M. Beckett

Project administration: C. Saint, F. M. Beckett

Software: C. Saint, F. M. Beckett




Visualization: C. Saint

© 2024 Crown copyright and British Geological Survey (C) UKRI. This article is published with the permission of the Controller of HMSO and the King's Printer for Scotland.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial License](https://creativecommons.org/licenses/by-nc/4.0/),

which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Using Simulated Radiances to Understand the Limitations of Satellite-Retrieved Volcanic Ash Data and the Implications for Volcanic Ash Cloud Forecasting

C. Saint¹ , F. M. Beckett¹, F. Dioguardi^{2,3} , N. Kristiansen¹, and R. N. Tubbs¹ 

¹Met Office, Exeter, UK, ²British Geological Survey, The Lyell Centre, Edinburgh, UK, ³Dipartimento di Scienze della Terra e Geoambientali, University of Bari "Aldo Moro", Bari, Italy

Abstract Volcanic Ash Advisory Centers (VAACs) have generated volcanic ash forecasts for the aviation industry since the mid-1990s. The excellent spatial and temporal coverage of satellite data makes them critical to the validation of ash dispersion model forecasts. This study investigates the limitations of satellite-retrieved volcanic ash data through the production of simulated radiances for a range of ash cloud properties encompassing the satellite retrieval's sensitivity. We run a detection and retrieval algorithm (Francis et al., 2012, <https://doi.org/10.1029/2011JD016788>) on these simulated ash clouds and assess the sensitivity and performance of the algorithms. Expected limitations are highlighted, including a lack of sensitivity to particles larger than $\sim 10 \mu\text{m}$ in radius and challenges in accurately retrieving heights in the stratosphere. However, other previously poorly defined limitations are also constrained, such as the reduction in sensitivity as ash column loading increases in optically thick ash clouds and increasingly underestimated column loading when column loadings are $> \sim 7 \text{ g m}^{-2}$. We consider the implications of the identified limitations when using satellite-retrieved ash column loadings to verify dispersion model output. We show that, accounting for the limitations of the satellite retrieval, a significant proportion of mass in the model output can lie outside the sensitivity range of the satellite detection and retrieval. This demonstrates the importance of understanding observations' limitations when comparing to model output. This knowledge should be used when verifying operational volcanic ash cloud forecasts.

Plain Language Summary Volcanic Ash Advisory Centers (VAACs) were created to forecast the location of airborne volcanic ash as it is a hazard to aircraft. These forecasts are produced by dispersion models, which combine estimates of the ash ejected by the volcano with meteorological information to forecast the location and amount of ash in the atmosphere. Observations, particularly satellite data, are required to validate the forecasts. However, satellite data also has uncertainty and this can vary significantly for different ash clouds. This paper investigates the uncertainty in quantitative satellite data, constraining previously poorly defined biases, such as limitations for larger ash particles and regions where large amounts of ash are present. These limitations are then considered when comparing quantitative satellite data to quantitative dispersion model output. This showed that significant portions of ash mass present in dispersion model output may be outside the quantitative satellite data's sensitivity range. This has important implications for the use of quantitative satellite volcanic ash data to validate dispersion model output.

1. Introduction

As volcanic ash represents a hazard to aviation there is significant interest from the aviation industry in volcanic ash forecasts. Since the mid-1990s, a network of nine volcanic ash advisory centers (VAACs) has generated forecasts, which advise the aviation industry of the presence of volcanic ash in the atmosphere (Mastin et al., 2022). The forecasts are produced based on the output of volcanic ash transport and dispersion models (VATDMs) and observations. The VATDMs take input from numerical weather prediction (NWP) models and information about the volcanic ash source to model the transport and dispersal of the ash (Bonadonna et al., 2012; Witham et al., 2007). At the London VAAC, the Numerical Atmospheric-dispersion Modelling Environment (NAME) is used (Beckett et al., 2020; Jones et al., 2007). Although the forecasts are currently only required to indicate the expected location of the ash cloud, there will soon be a new requirement, referred to as quantitative volcanic ash (QVA), for VAACs to produce quantitative forecasts of ash concentration and an indication of the probability of exceeding the prescribed concentration thresholds (ICAO, 2019). This has placed more emphasis

Writing – original draft: C. Saint,
F. M. Beckett
Writing – review & editing: C. Saint,
F. M. Beckett, F. Dioguardi,
N. Kristiansen, R. N. Tubbs

on accurately modeling the amount of ash in the atmosphere and on observations to provide accurate and reliable quantitative data for validation and forecast adjustment in addition to monitoring. This will also require a comprehensive understanding of the uncertainty associated with the eruption source terms, dispersion model parameters, and NWP data used with VATDMs, as well as a similar understanding of the uncertainty in any observations which may be used to constrain the dispersion model output. Note that throughout this paper “cloud” will be used to refer to ash downwind of the source and “plume” will be used to refer to ash directly above the source location.

Thermal infrared (TIR) satellite data can be used to identify and analyze volcanic ash clouds (e.g., Francis et al., 2012; Pavolonis et al., 2015a, 2015b; Prata et al., 2022) using the reverse absorption effect (Prata, 1989a, 1989b; Wen & Rose, 1994). This is where silicate particles become less absorbing as the wavelength of incident radiation increases from 10 to 12 μm , whereas water and ice particles show the opposite behavior. The primary instrument used at the London VAAC to observe volcanic ash is the Spinning Enhanced Visible Infra-Red Imager (SEVIRI) aboard Meteosat Second Generation (MSG), which has channels at 10.8 and 12.0 μm . This is because it provides frequent (15 minutely) coverage of the eastern North Atlantic and northern Europe region the London VAAC is responsible for. A brightness temperature difference (BTD) is generated by subtracting the 12.0 μm brightness temperature (BT) from the 10.8 μm BT. In the presence of ash, this BTD will usually be negative, while the opposite is true for meteorological clouds. The retrieval of volcanic ash properties has been demonstrated in a variety of schemes including those described by Prata and Grant (2001), Francis et al. (2012), and Prata et al. (2022). The method of retrieval in this paper is that described by Francis et al. (2012), which is a 1D-Var (one-dimensional variational analysis) method based on the Bayesian optimal estimation previously described by Rodgers (2000). The method searches for the optimal solution for the ash column loading, ash cloud height, and lognormal distribution of ash particle radius (specifically, the optimal effective radius is found for eight different assumed geometric standard deviations: from 1.25 to 3.0 in 0.25 intervals). This is based on the observed radiances seen in the 10.8, 12.0, and 13.4 μm channels and prior (background) probability distributions for the retrieved ash cloud properties. The search for the optimum solution involves the minimization of a cost function, which is comprised of the sum of the “background” and “observation” cost. The observation cost is a measure of the difference between the observed BTs and the simulated BTs given the presence of an ash cloud of the retrieved properties. The background cost is a measure of the difference between the a priori ash cloud properties (provided in Table 1) and retrieved properties. Lower costs are indicative of scenarios where the retrieval has been able to find a solution that closely fits the observed BTs (and does not differ significantly from the a priori values). Higher costs indicate poorer fits and less reliable retrieval results.

The method, like most satellite TIR retrieval schemes, relies on a semitransparent ash cloud to produce the BTDs that allow for detection of volcanic ash and then for accurate retrieval. Above $\sim 10 \text{ g m}^{-2}$ column loading, the cloud starts to become optically thick and the reliability and accuracy of satellite retrievals decrease. Similarly, the BTD method relies on the temperature contrast between the surface and the cloud (volcanic or meteorological). The BT observed at the satellite depends on the temperature at the altitude the radiation has been emitted from. Usually, the radiation received will be a combination of radiation from the surface and radiation from any absorbers, such as clouds, in the atmosphere. As such, lower BTs are observed when there is thicker cloud present, which will usually be at a lower temperature than the surface. How much lower the BTs are will depend on how much colder the cloud is than the surface (where higher clouds are usually colder), how thick the cloud is (optically thicker clouds will result in a greater proportion of the observed radiation coming from the colder cloud than the warmer surface) and the wavelength being observed (ash is more absorbing at 10.8 μm than 12.0 μm , and the opposite is true for water, which leads to the BTD behavior described above). As such, there could be an unknown bias in observations when ash outside the observing method's sensitivity range is present, such as with ash particles larger than 10 μm in radius, which do not produce the BTD behavior typically associated with ash. Once the properties of volcanic ash have been retrieved from the satellite data, they can be used for both dispersion model validation and adjustment. The latter can take many forms including data insertion (Prata et al., 2021; Wilkins et al., 2016), inversion (Pelley et al., 2021; Stohl et al., 2011), and data assimilation (Fu et al., 2017; Mingari et al., 2022; Pardini et al., 2020). There are strengths and weaknesses to each of these approaches, but all require a good understanding of the uncertainty and potential for bias in the satellite data.

This paper intends to quantify biases and uncertainties that may be present in the Francis et al. (2012) satellite volcanic ash retrieval and assess the impact any biases present may have on the operational use of satellite-retrieved ash data. This will be achieved through the production of simulated radiances based on the method

Table 1
A Priori Ash Cloud Properties and A Priori Uncertainties

Ash cloud property	A priori value	A priori uncertainty
Column loading (g m^{-2})	~3	20.0
Cloud top pressure (hPa)	600.0	750.0
Ash effective radius (μm)	3.5	10.0

Note. The a priori column loading is calculated for each of the size distribution standard deviations (1.25–3.0 in 0.25 intervals) by dividing an optical depth of 0.5 by the absorption coefficients of the SEVIRI 10.8 μm channel for ash with a lognormal size distribution defined by the a priori effective radius (3.5 μm) and the relevant standard deviation. This results in a priori column loading values in the range of 2.76–4.30 g m^{-2} .

of Millington et al. (2012). They show how simulated imagery could be used to make a real-time like-for-like comparison between satellite data and dispersion model data. This scheme is running operationally at the London VAAC and can be used in near-real time in response to an ongoing eruption. Simulated imagery has also been used previously to assess the performance of ash detection and retrieval algorithms (Corradini et al., 2014; Kylling et al., 2015; Piontek et al., 2021; Stevenson et al., 2015) and is useful because it avoids the challenges of verifying the performance of ash retrievals with real data. This is difficult due to the rarity of events and the lack of reliable and accurate independent observations to verify against. In this paper, the use of simulated radiances allows for an assessment of the performance of the volcanic ash retrieval across a range of ash cloud properties where the truth is known. We will show how awareness of biases present in the ash retrieval could impact decision making when comparing satellite-retrieved data with dispersion model output.

2. Background Information

To provide optimal quantitative and probabilistic ash cloud forecasts, it is imperative that we are able to constrain the uncertainties associated with the data being used. Previous studies have explored the sensitivity of model simulations to the driving data used, including the NWP and eruption source parameter (ESP) data, and considered the uncertainty on the model output. Crawford et al. (2022) showed that significant differences in predicted ash cloud area can occur when different meteorological ensemble members are used to drive the dispersion model. Devenish et al. (2012) considered the sensitivity of dispersion model output to the representation of several key physical processes in the model source parameters, including the eruption plume height and fraction of mass that remains within the distal ash cloud. They found that there was significant uncertainty in the latter and suggested that this could be due to various factors, including aggregation and emission of material with large grain sizes. Webster et al. (2012) found that their peak concentrations predicted by dispersion modeling were within one and a half orders of magnitude of the observed peak concentrations. Beckett et al. (2015) examined the sensitivity of the dispersion model output to the physical characteristics assigned to the ash; in particular, the particle size distribution (PSD) used to initialize the model simulations. They found that knowledge of the total grain size distribution (TGSD) is critical to reducing uncertainty in ash cloud dispersion forecasts. Dioguardi et al. (2020) highlight how predicted peak concentrations can vary by a factor of 35 depending on the method used to estimate the mass eruption rate (MER) and the observed plume height chosen.

The uncertainty on satellite-retrieved ash column loadings from TIR data was assessed by Gu et al. (2005) and Stohl et al. (2011) based on the work of Wen and Rose (1994). Gu et al. (2005) suggest a slightly smaller uncertainty of ~40%, with the assumption that the PSD and refractive index of the ash particles are known, compared to 40%–60% suggested by Stohl et al. (2011). Based on the sensitivity analyses of Wen and Rose (1994) and Gu et al. (2005), Prata and Prata (2012) suggest ash column loading uncertainties of 40%–50% associated with measurement error, modeling assumptions, and knowledge of the parameters required in the radiative transfer model. Western et al. (2015) investigated the sensitivity of retrieved column loadings to a variety of parameters including refractive index, surface temperature, ash cloud top temperature, ash bulk density, and uncertainties related to the spread of the ash particle size distribution, which they found to cause the largest uncertainty in the retrieved mass from the Eyjafjallajökull ash cloud. Corradini et al. (2008) completed a sensitivity analysis on the uncertainty in retrieved masses caused by uncertainty in surface temperature and emissivity, plume geometry (altitude and thickness), ash type, and atmospheric water vapor and found an uncertainty of 40% on total mass retrieval. Kylling et al. (2014) consider the effect of assuming spherical particles in the radiative transfer modeling and find an increase in the uncertainty of ash retrieval mass of 40%–50%, with the assumption of spherical particles resulting in an underestimation of ash mass. Prata et al. (2022) provide a detailed analysis of the sources of uncertainty in retrieved ash column loadings and note the usefulness of pixel-scale uncertainty estimates in all retrieved variables, which allow the data to be used in data assimilation schemes.

In addition to satellite data, there are also other types of observations that can be used to verify dispersion model forecasts. Bonadonna et al. (2015) provide a comprehensive assessment of the uncertainties associated with measuring the properties of tephra deposits and find significant uncertainty on the estimation of the MER and

eruption duration. Turnbull et al. (2012) consider the aircraft observations made during the 2010 Eyjafjallajökull eruption by Johnson et al. (2012) and Schumann et al. (2011) to suggest a factor of 2 uncertainty on aircraft observed ash concentrations. Marengo et al. (2011) consider the airborne lidar observations that were completed from the same aircraft as used in the Johnson et al. (2012) study and note the patchy and fractal nature of the observed ash clouds. Ansmann et al. (2011) describe ground-based lidar and photometer measurements of volcanic ash during the 2010 Eyjafjallajökull eruption and note the high uncertainty within the retrieval scheme when an unknown fraction of particles have radius $>15 \mu\text{m}$. Observations from the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) aboard the CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations) satellite are described by Winker et al. (2012) who found an uncertainty of around a factor of five in ash mass concentrations from the uncertainty in particle size distribution and composition.

Considering the vast amount of work on understanding uncertainties associated with volcanic ash modeling and observations completed since the 2010 eruption of Eyjafjallajökull, there have been surprisingly few detailed comparisons of ash estimates from different methods where the differences are analyzed in the context of uncertainty estimates for each of the methods. It has been seen that dispersion model output and ground deposit data will frequently indicate more ash mass than is observed in satellite, lidar, and aircraft data. Gudmundsson et al. (2012) estimated that at least 70 Tg of ash $<28 \mu\text{m}$ in diameter was released during the 2010 eruption of Eyjafjallajökull based on tephra deposit data. They noted this was approximately an order of magnitude more mass than Stohl et al. (2011) estimated had been released ($8.3 \pm 4.2 \text{ Tg}$) in particles in the size range of 2.8–28 μm during the same eruption based on inversion of satellite retrievals. It is worth noting that both Gudmundsson et al. (2012) and Stohl et al. (2011) referred to the mass of ash within a size range defined by diameter. We have chosen to remain consistent with the use of diameter throughout this paper when referring to dispersion modeling and ground deposit data, while using radius when referring to observation data including satellite, lidar, and aircraft data. Corradini et al. (2011) found the area and total ash mass of volcanic clouds computed by the FALL3D dispersion model were significantly greater than the MODIS (Moderate-resolution Imaging Spectroradiometer) retrieved data indicated for three example images from the October 2002 eruption of Mount Etna. Marzano et al. (2018) suggested that the differences between ground deposit and TIR satellite-retrieved masses are largely due to the lower sensitivity of TIR retrievals to larger particles. By considering microwave (MW) data from the Advanced Technology Microwave Sounder (ATMS) they were able to derive similar mass estimates from satellite data as the proximal ground deposit data showed for the 2015 Calbuco eruption. Freret-Lorgeril et al. (2021) consider estimates for total erupted mass (TEM) and MER and show promising agreement between ground deposit data, near-source observation estimates, and satellite estimates when MW and TIR data are combined for two events at Mount Etna. They note that each of the data sources are sensitive to different phases of the eruption and the importance of considering this when estimating MER from TEM or TEM from MER.

The requirement for quantitative forecasts (ICAO, 2019) makes it more critical than ever to understand the inconsistency of mass observed in tephra deposits with that observed by other methods including satellite data. To accurately forecast the concentration of ash in the atmosphere both the MER and impact of near-source removal processes, such as aggregation (e.g., Brown et al., 2012), gravitational instabilities (Carazzo & Jellinek, 2012; Manzella et al., 2015), hydrometeor formation (Durant & Rose, 2009; Durant et al., 2008), particle-particle interactions (Del Bello et al., 2017; Eychenne et al., 2015), and topography-induced perturbations to local wind fields which can enhance sedimentation on the leeside of mountains (Eychenne et al., 2017; Poulidis et al., 2017; Watt et al., 2015), must be constrained. Often the quoted uncertainties of observations used to constrain these variables are provided within the sensitivity range of that observation and the uncertainty outside this is not fully explored. In particular, TIR satellite retrievals are only sensitive to particles of $<\sim 10 \mu\text{m}$ in radius (Prata et al., 2022; Stevenson et al., 2015; Wen & Rose, 1994). Limitations in the ability to observe the full range of particles present is a concern for other observation methods with some methods for analyzing tephra deposits unable to identify smaller particles ($<\sim 20 \mu\text{m}$ diameter) (Stevenson et al., 2015) and challenges observing larger particles ($>\sim 50 \mu\text{m}$ diameter) with aircraft instrumentation (Turnbull et al., 2012). Kylling et al. (2015) demonstrate a further limitation of satellite-retrieved data by showing the negative impact that meteorological clouds can unsurprisingly have on the ability of satellites to detect and retrieve ash cloud properties.

Partially to correct for the discrepancies between TIR satellite data and dispersion model output, a scaling factor is applied to the MER used to initialize NAME in the operational setup for the London VAAC (Beckett et al., 2020; Webster et al., 2012), referred to as the distal fine ash fraction (DFAF). This computationally efficient approach is based on the assumption that the dispersion model output is overestimating the ash column loadings in distal ash

clouds because it does not represent near-source processes, which act to prematurely remove ash. In addition, only the smallest ash ($\leq 100 \mu\text{m}$ in diameter) is considered in the modeling, as it is assumed that larger particles will be quickly removed from the atmosphere given their sedimentation rate; Durant and Rose (2009) showed that particles larger than a millimeter have a residence time of less than 30 min. Dioguardi et al. (2018) suggested that particles coarser than $100 \mu\text{m}$ falling from an altitude of 20 km a.s.l. in a uniform 10 m s^{-1} cross flow would travel only few tens of km, excluding processes other than settling. Devenish et al. (2012) compared satellite-retrieved (MSG—SEVIRI) and modeled ash column loadings in the Eyjafjallajökull 2010 distal ash cloud and estimated the DFAF to be $\sim 5\%$. This agreed with observations of ash clouds from eruptions of Spurr, El Chichón, Láscaar, and Hudson (Rose et al., 2000) and was consistent with other studies including Dacre et al. (2011) who found that around 3% of the total mass remained in the distal cloud. A default DFAF of 5% is now applied in the operational setup at the London VAAC, although there is an option to vary this parameter, as was done during the response to the 2011 eruption of Grímsvötn. The ability to vary the DFAF is important as demonstrated by Gouhier et al. (2019) who noted from combined satellite and field data of sustained eruptions that the DFAF can vary by two orders of magnitude, which agreed with Mastin et al. (2009). Mastin et al. (2009) considered many studies where the TGSD, and therefore DFAF, had been determined and found the DFAFs ranged by nearly two orders of magnitude. They also considered the relationship between plume height and MER and found a fit that did not differ significantly from that found by Sparks et al. (1997), or the theoretical studies which indicated plume height should increase with the fourth root of MER and this is now often referred to as the Mastin relationship.

Although some studies indicate that a DFAF of 5% might be overestimating the proportion of ash reaching the distal cloud, Cashman and Rust (2019) present a different view. They suggest that very small ash ($< 20 \mu\text{m}$ diameter) is missing from the records of well-documented eruptions, including mapped deposits, eye-witness accounts, satellite-based observations (where eruptions pre-date satellites or for particles $< 1 \mu\text{m}$ diameter), and cryptotephra records. Further, considering the significant mismatch between the total mass estimated from deposits and satellite retrievals for a number of well-studied events, using 5% of the total mass erupted to represent the DFAF could be severely under-estimating the ash column loading in far-traveled volcanic clouds. There is also a growing body of evidence which shows that particles with diameter $> 100 \mu\text{m}$ can travel significant distances (e.g., Beckett et al., 2015; Cashman & Rust, 2019; Saxby et al., 2018; Stevenson et al., 2015; Watson et al., 2016). Watson et al. (2016) found particles from the 1875 eruption of Askja (Iceland) with lengths of up to $190 \mu\text{m}$ in Northern Poland $\sim 2,500 \text{ km}$ from the vent, and Saxby et al. (2019) observed particles from an eruption of Katla 12.1 ka BP (before present), also in Iceland, with a diameter of that of an equivalent sphere up to $191 \mu\text{m}$ in Norway, more than 1,200 km from the vent. Further, Saxby et al. (2018) demonstrated that when the physical characteristics of the ash are correctly represented (their shape and density) in the dispersion model, the observations of large ash ($\sim 100 \mu\text{m}$) in the UK following the eruption of Eyjafjallajökull in 2010 can be accounted for. Finally, it has now been shown that aggregation, rather than acting to preferentially remove ash from the atmosphere, can also raft particles to much further distances (Rossi et al., 2021). Constraining the DFAF is critical for comparisons of satellite and dispersion model data as shown by Crawford et al. (2022). They noted the variation in DFAF (from 1% to 50%), alongside other uncertainties in the method for estimating MER, as possible explanations for the order of magnitude difference they found in the effective (distal) MER estimated using the empirical Mastin relationship and a DFAF of 10%, to the effective (distal) MER estimated from satellite observations.

3. Methods

The three following subsections describe the methods used in this paper. These describe in detail: the production of volcanic ash simulated imagery; how the large data set produced was analyzed; and how the results from the experiments were used to recompare satellite data and dispersion model output. The detection method is identical to the current operational satellite volcanic ash detection method used at the London VAAC and described in detail by Saint (2023). In brief, the method uses BTM and beta ratio thresholds in combination with some spatial and geographical information in order to identify ash affected pixels. This is typically very effective for semi-transparent ash clouds, but, as it is reliant on a spectral signal indicative of ash, is often unable to identify optically thick ash clouds soon after an eruption or ash clouds with significant water content. The retrieval method is very similar to the method described by Francis et al. (2012) and in Section 1. The only difference was the reduction of the low-pressure hard limit from 50 to 1 hPa to ensure the retrieval could produce retrieved altitudes as high as the

highest altitudes used in the simulated data set (up to 25 km). The hard limits are present to ensure the retrieval does not produce unphysical values. Testing indicated the change had minimal influence on the results.

3.1. Production of Simulated Imagery

We use the method described by Millington et al. (2012) to produce simulated imagery, updated to use a more generic data set. Typically, simulated imagery is produced by feeding NWP data and modeled ash concentration data from a dispersion model such as NAME into a radiative transfer model such as RTTOV (Radiative Transfer for TOVS: <https://nwp-saf.eumetsat.int/site/software/rttov/>) to simulate radiances. These simulated radiances (produced by vertical 1D simulations with no horizontal variation for each simulated radiance) can then be combined into images as the dispersion model data will have a spatially coherent ash cloud.

Simulated imagery can be very useful for forecast evaluation, both qualitatively by comparing the real and simulated Dust RGB imagery (Figures 1a and 1c), and quantitatively by comparing the column loadings retrieved from the real image with those retrieved from the simulated image (Figures 1b and 1d). Dust RGB imagery is described in more detail by Francis et al. (2012). In an operational setting, it can be challenging to compare satellite imagery with dispersion model output. Regions where the forecast ash concentrations produced by the dispersion model do not align with the observations, and where there may be a forecast error that needs correcting, can be difficult to identify. Even when dispersion model output is presented as column loadings rather than concentrations, the limitations of the satellite-retrieved data can result in incorrect interpretation of the data. For example, the dispersion model could indicate ash present in the same horizontal location as the satellite imagery, but if the ash in the dispersion model output was at low altitude (<3 km) it may not be sufficient to produce the signal seen in the satellite imagery. Simulated imagery helps to avoid errors in interpretation like this by producing the expected satellite image given the dispersion model output and allowing a like-for-like comparison to the real satellite image.

Our method was different to the typical production of simulated imagery in two main ways. First, the properties of neighboring pixels are uncorrelated, so the position of each pixel in the final image has no implications for the parameters modeled. Second, we inserted meteorological data from the “Diverse profile data sets from the ECMWF 137-level short-range forecasts” (Eresmaa & McNally, 2012) and inserted ash in specified amounts and model levels rather than using NAME model output concentrations. Here, we will describe how the diverse profile data set was reduced and then how the profiles were used. We will then describe how the ash data were defined and the resulting range of ash cloud conditions that were simulated. Finally, we will describe the addition of noise to the BTs, which is representative of the noise seen in real SEVIRI imagery due to the instrument.

3.1.1. Reduction of the Diverse Profile Data Set

To produce volcanic ash simulated imagery it is necessary to have complete knowledge of the model environment in order to simulate not only the impact of the ash on the observed radiances, but also of any other atmospheric constituents that may impact the observed BTs. We take only the full pressure, humidity, and temperature profiles alongside the corresponding surface variables from the diverse profile data set, as the remaining data do not affect the simulated radiances in the four TIR channels used in the detection and retrieval. We also fixed the latitude and satellite zenith angle to 50° and 60°, respectively, as these also affect the simulated radiances. We chose not to investigate the impact of altering the satellite zenith angle as the impact of the corresponding increase in path length with satellite zenith angle is well understood (Gu et al., 2005). To simplify the analysis, meteorological cloud was not simulated in this investigation so the only cloud impacting the radiances simulated was ash cloud. RTTOV v13.0 was used in this paper, details for which can be found from the Science and Validation report for v13 (Saunders et al., 2020).

In addition to selecting which parameters from the data set it was necessary to use, we decided to filter the data set to remove a large number of the atmospheric profiles. The original data set has 25,000 profiles, including 5,000 profiles representing the variation in each of temperature, specific humidity, ozone mixing ratio, cloud condensates, and precipitation. For the 5,000 profiles representing the variation in temperature for example, the profiles for the other variables are taken from the same time and location as the respective temperature profile. Initially, only the 10,000 profiles which represented the variation in humidity and temperature were used. Further studies into the impact of variables related to the presence of meteorological cloud would be encouraged. The filtering then involved removing profiles where the meteorological cloud cover was greater than 10% or which

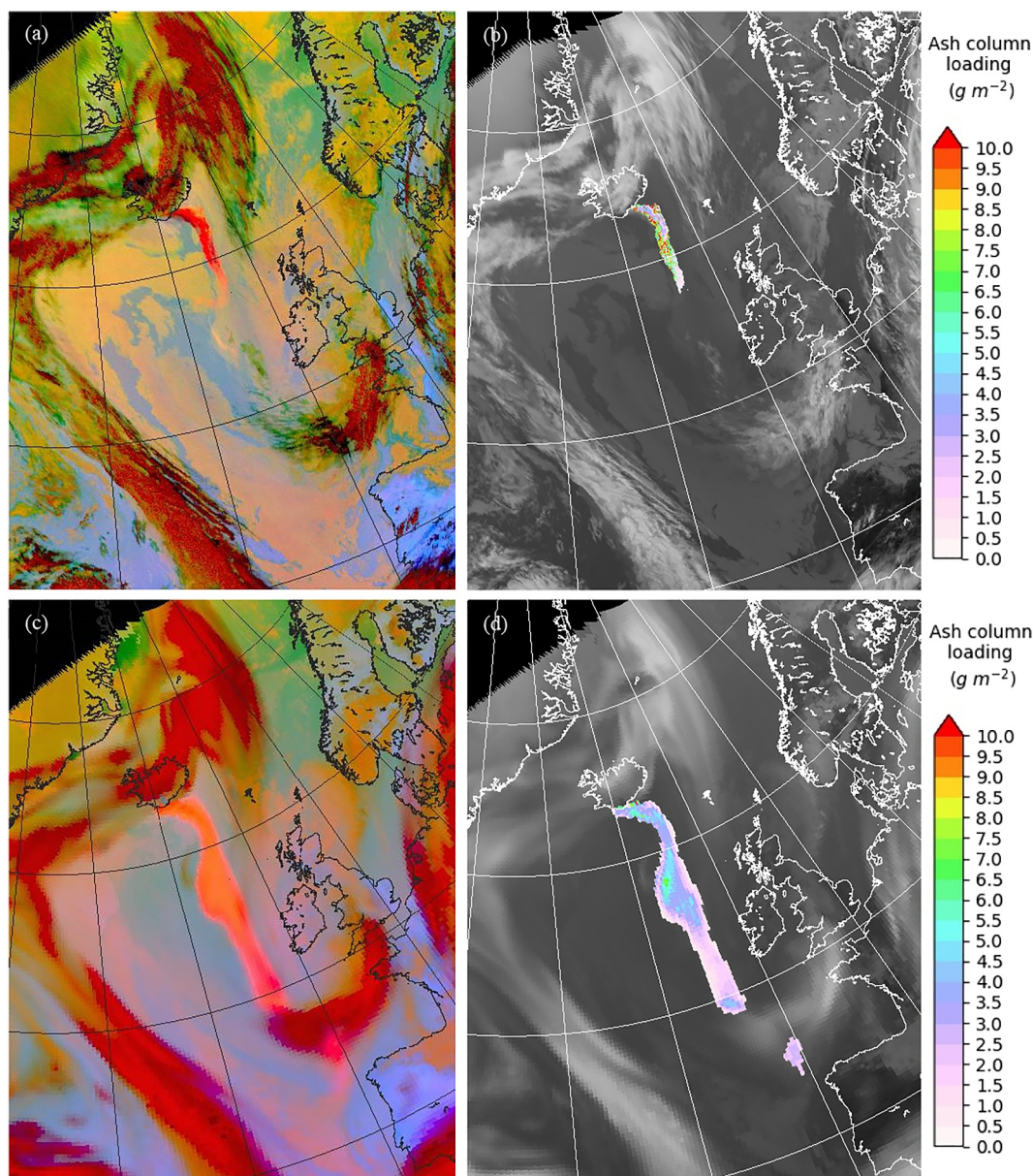


Figure 1. (a) Dust RGB, (b) satellite-retrieved column loadings, (c) Simulated Dust RGB, and (d) column loadings retrieved from the simulated image at 1200UTC on 6 May 2010. The simulated imagery was produced using the NAME data shown in Figure 9b and described in Section 3.3.

were not over ocean surfaces. By removing these profiles, we deliberately removed a large number of conditions from the model data set that are known to be challenging for satellite volcanic ash detection and retrieval methods. We did this to determine how the retrieval performs in what would be considered close to ideal conditions for volcanic ash detection and retrieval methods. This filtering resulted in only 539 of the original 10,000 profiles being used. The profiles retained were those over the ocean, containing less than 10% cloud and from the data sets representing the variation in temperature and humidity.

3.1.2. Selection and Insertion of Ash Into Simulated Imagery

In the operational production of volcanic ash simulated imagery at the London VAAC and in the previous work by Millington et al. (2012), realistic volcanic ash clouds are simulated based on NAME model output. We modeled 1D-columns of ash at an angle (satellite zenith angle) through the atmosphere by inserting ash into specific model

Table 2
Range of Ash Cloud Properties Simulated Radiances Were Produced for

Ash cloud property	Range of values
Column loading (g m^{-2})	0.1, 0.25, 0.5, 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0, 12.5, 15.0, 17.5, 20.0, 25.0, 30.0, 40.0, 50.0, 75.0, 100.0
Cloud top altitude (m)	987, 1,600, 2,453, 3,564, 4,896, 6,341, 7,797, 9,256, 10,714, 12,190, 13,727, 15,329, 16,999, 18,756, 20,695, 22,918, 25,489
Ash median radius (μm)	0.62, 1.9, 5.9, 9.3, 12.4, 15.6

levels with concentrations such that the column loading was set at the values indicated in Table 2. Note that RTTOV requires the number concentration of the aerosol being simulated. This was calculated from the mass density and the mean mass per particle, which was determined from the particle size distribution. The radiative properties of the aerosol being simulated were based on the particle size distribution (which was varied as described in the following paragraph) and required the recalculation of RTTOV coefficients. For both the calculation of the mean mass per particle and evaluation of the radiative properties, the particles were assumed to be spherical. As the ash particles are assumed to be spherical both when simulating the radiances and within the retrieval, this investigation does not further explore the behavior Kylling et al. (2014) observed that assuming spherical particles when retrieving volcanic ash particles will result in underestimated column loadings.

There is a tool in RTTOV v13.0 that allows for the creation of new coefficients files when provided with refractive indices, a particle size distribution and whether the aerosol species is hydrophilic. The general process for simulating radiances in the presence of aerosols is described by Matricardi (2005). The tool requires a refractive index data set and particle size distribution as inputs. The Mie properties, which describe the scattering of light, are then computed using the refractive index interpolated to the channel wavenumber for each particle size from a grid covering the desired particle size distribution. The final values can then be obtained by integrating the computed properties over the size distribution. Using this tool, we created new coefficients files for five further lognormal size distributions. Each of these had a geometric standard deviation of 1.85 and had median radii in terms of mass as listed in Table 2. In each case we used the same Pollack et al. (1973) andesite refractive indices data set used in the operational simulated imagery. Although interpolating the refractive index of ash (represented by Pollack et al. (1973) andesite) to the central wavelength of the instrument channel (in this case SEVIRI) is sufficient to account for the impact of ash reasonably accurately, this approach would not be accurate enough for the gaseous absorbers. Therefore, when producing the simulated radiances for the MSG-2 SEVIRI channels centered at 8.7, 10.8, 12.0, and 13.4 μm , RTTOV accounts for the spectral response functions of the instrument within the gas absorption parameterization.

For each of these size distributions we tested a range of ash cloud heights (above ground level) and ash column loadings, as given in Table 2. Ash cloud heights refer to the highest altitude model level the ash was placed in. For each of these ash cloud heights the depth of the ash cloud was varied by placing ash in a varying number of levels below the cloud top. The number of model levels containing ash were 1, 5, 10, 15, and 20, resulting in ash cloud depths ranging from 97.16 to 8,490.07 m. The model levels referred to here are as defined by the ECMWF diverse profile set (geometric altitudes of each level are provided in Table S1 in Supporting Information S1), and these are used in both the production of the simulated radiances and the volcanic ash retrieval. Note that the model levels become geometrically thicker with increasing height and the depths tested are therefore deeper at higher altitudes. This range of data was selected with the intent of testing the volcanic ash detection and retrieval across the full range they are sensitive to and to identify the limits of the sensitivity to each variable.

3.1.3. Addition of Instrument Noise

To ensure the results verifying the detection and retrievals performance with simulated imagery were as applicable as possible to real imagery, we added noise to the simulated BTs. Observed radiances or BTs from a satellite will always have an error associated with the measurement. This error can be averaged over a period of time and separated into two parts: the instrument or channel noise and the instrument or channel bias (as both the noise and bias will be different for each channel on an instrument). The bias is the difference from some reference truth value, such as a more accurate observation, and the noise represents the random variation around this bias. Typically, bias correction is used to remove the bias from the observation. Zero-mean stochastic Gaussian noise

was added to our simulated BTs based on the noise values in the EUMETSAT “Typical Radiometric Noise, Calibration Bias and Stability for Meteosat-8, -9, -10, and -11 SEVIRI” document (https://www-cdn.eumetsat.int/files/2020-04/pdf_typ_radiomet_acc_msg-1-2.pdf). Here, we use zero-mean Gaussian noise, implicitly assuming the radiance bias correction to be optimal. Assuming no bias correction and using the channel bias data from the EUMETSAT “Typical Radiometric Noise, Calibration Bias and Stability for Meteosat-8, -9, -10, and -11 SEVIRI” document would have resulted in us assessing the impact of a bias that will be mostly accounted for in the real observations by the bias correction. The noise was added to all four channels used in the Francis et al. (2012) automated detection and retrieval algorithms, which are the 8.7 (detection only), 10.8, 12.0, and 13.4 μm (retrieval only) channels.

3.1.4. Production of Simulated “Imagery”

For each of the ash cloud properties described in Section 3.1.2, imagery was produced using the 539 model profiles, one profile for each NWP model grid box (re-gridded to a 17 km horizontal resolution). As each grid box covers the area of multiple image pixels (3 km horizontal resolution at the sub satellite point), there were 13,871 image pixels containing ash, each with one of the 539 model profile data sets and with independent stochastic noise added as described in Section 3.1.3. As the operational detection, retrieval, and simulated imagery have been designed to work with and produce images, producing pseudo-images by arranging the model profiles in neighboring grid boxes significantly simplified the processing required by allowing all of the model profiles to be tested simultaneously for each of the ash cloud conditions. Examples of this imagery can be seen in Figure 2, although it should be noted that this imagery is merely illustrative of the impact of changing model data and instrument noise on both the simulated radiances and retrieved parameters. Upon close inspection, particularly of Figure 2c, the 539 NWP model grid boxes can be identified by the similar retrieved data with some variation within these noticeable due to the different instrument noise applied to each pixel within.

3.2. Method of Analysis of Satellite Retrieval Data

The volume of information produced in this investigation was significant because ash clouds of five different depths were created with every possible combination of ash cloud height, ash column loading, and particle size distribution (Table 2). As a result, 11,730 individual runs with unique ash cloud properties were completed. Figure 2 shows the (a) Dust RGB along with (b) the automated volcanic ash confidence product described by Saint (2023) and the retrieved (c) height, (d) loading and (e) effective radius for one of these runs. The location of each of the pixels is identical when the radiances are simulated and the retrieval is run. Therefore, in total, the experiments produced $13,871 \times 11,730 = 162,706,830$ data points that could be analyzed. However, we were not focused on the impact of the instrument noise or the variation in meteorological conditions. As a result, the variation in retrieved properties for the 11,730 different ash cloud conditions was the primary feature of focus for this analysis.

Histograms were plotted for the retrieved variables from each image pixel for each of these unique ash cloud conditions (e.g., in Figure 3), and the skewed nature of some of these indicated that it may be useful to take both the mean and mode of the data. In many cases the data were likely to be skewed due to the mode being close to zero and the inability to have negative values. The variation in retrieved values seen across the histograms is primarily driven by differences in the NWP model profiles across the model cells, but some variation driven by the noise applied to the image pixels is also present. The mean, mode, standard deviation, and retrieval success of the pixel values in each “image” were used to summarize statistically each of the 11,730 images. The value described as the mode is found by taking the central point of the modal bin in the histograms, which have bin widths of 0.1 g m^{-2} , 150 m, and $0.06 \mu\text{m}$ for the ash column loading, ash cloud height, and ash effective radius. These bin spacings allow for a good estimate of the maximum of the probability density function to be found by being narrow while ensuring a sufficient number of data points within each bin. The retrieval success percentage indicates the percentage of pixels that the detection successfully marked as ash and that the retrieval then produced a solution with a low enough cost for. The maximum acceptable cost is set to 12, which was empirically found to be a reasonable upper limit for good quality retrievals, and only retrievals with a cost lower than this can be seen in Figure 3. It is important to note that this does not necessarily mean that the retrieved data are accurate. The mean, mode, standard deviation, and retrieval success data from each histogram were then further reduced by averaging across more variables, including all of the depths tested and all of the size distributions tested. The results of this are shown and discussed in Section 4.

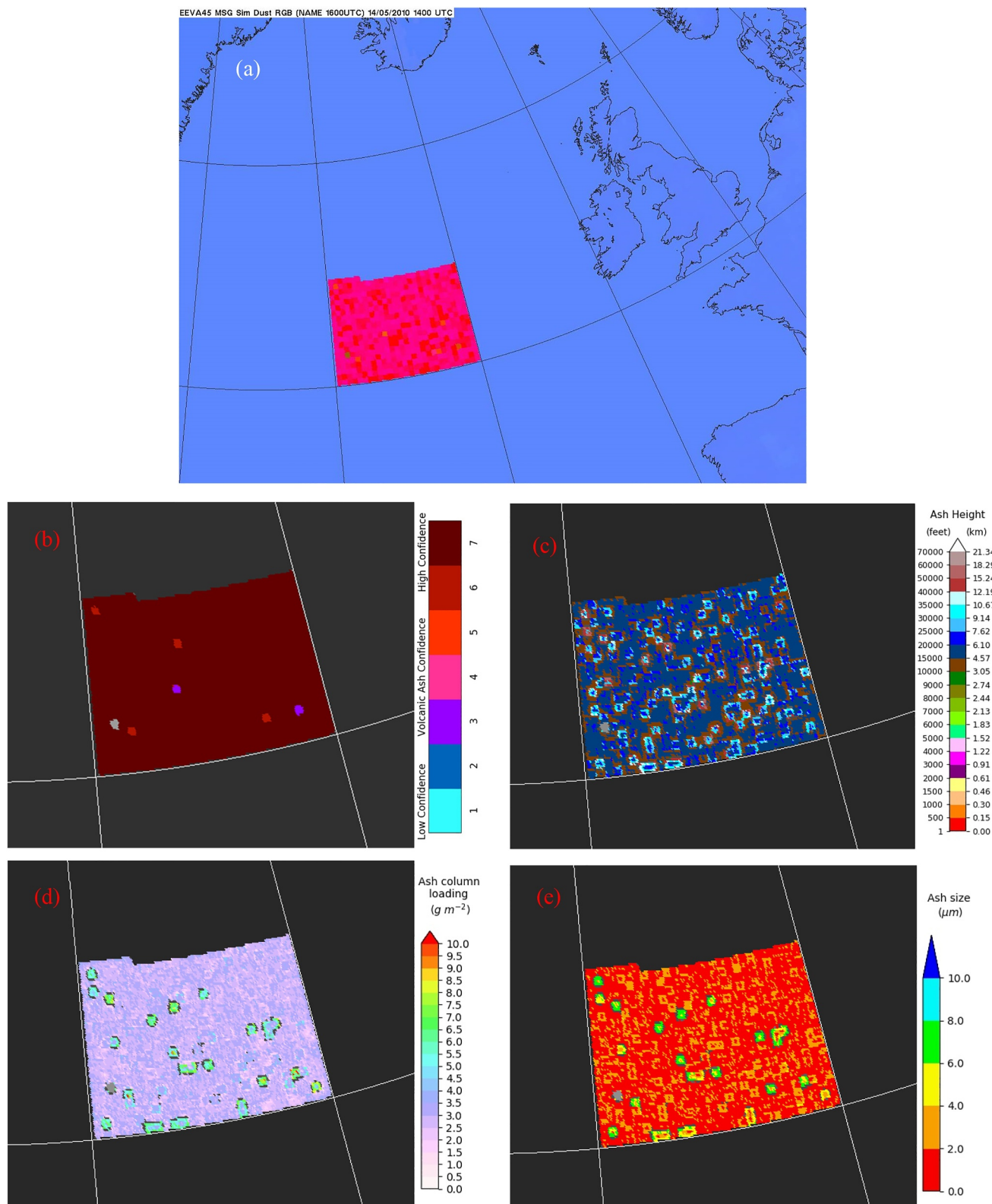


Figure 2. Example images of the simulated ash cloud (produced assuming a median radius of $1.9\ \mu m$, $3\ g\ m^{-2}$ column loading, 4,896 m cloud top height with one model level depth): (a) Dust RGB and, cropped to highlight the region containing ash, (b) volcanic ash confidence, retrieved (c) height, (d) loading and (e) effective radius. The ash clouds look spotty primarily due to the different model data (from the diverse profile data set) used in each model cell and the resulting impact on the detection and retrieval. Smaller variations within each model cell can be seen from the different instrument noise applied in each pixel.

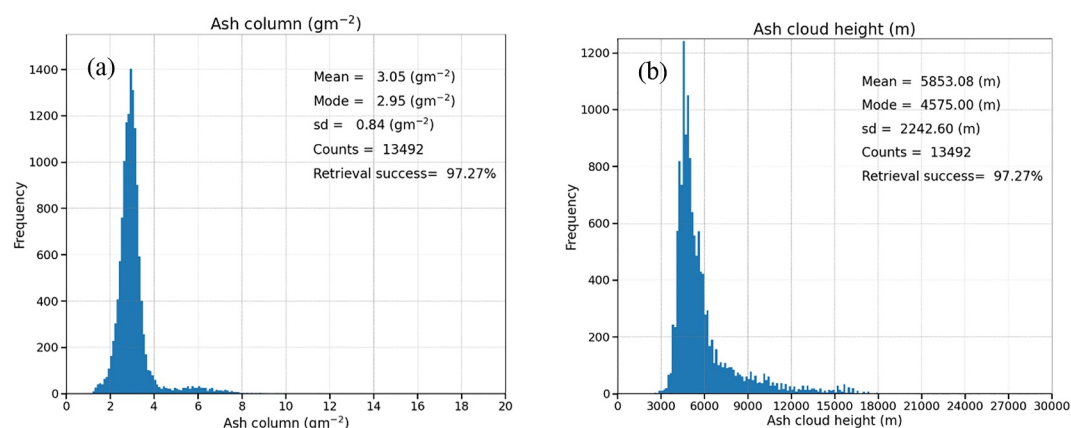


Figure 3. Examples of retrieved data plotted in histograms with relevant statistical data evaluated (including the mode which may differ from the mean significantly). Plots show retrieved (a) loading and (b) height for a simulated ash cloud with a median radius of $1.9 \mu\text{m}$, 3 g m^{-2} column loading, and $4,896 \text{ m}$ cloud top height with one model level depth.

3.3. Comparison to Dispersion Model Output

We assess the differences between NAME modeled and satellite-retrieved ash column loadings in the ash cloud from the eruption of Eyjafjallajökull volcano in Iceland (14 April to 23 May 2010). This is achieved by first comparing the total mass in NAME output with the total satellite-retrieved mass and then successively removing ash the satellite may not be sensitive to from the NAME output and repeating the comparison with the total satellite-retrieved mass. Model particles are advected by 4D wind fields; for this study, we use analysis data from the Global configuration of the Met Office's Unified Model (UM; Davies et al., 2005). Small-scale three-dimensional atmospheric turbulence and unresolved mesoscale motions are parameterized using random walk techniques (Webster et al., 2018), and parameterizations of sedimentation, dry deposition, and wet deposition are used to simulate the removal of volcanic ash from the atmosphere (Webster & Thomson, 2011).

NAME simulations use a set of source parameters which represent the changing eruption conditions through time. Model particles, each representing a mass of volcanic ash, are released as a vertical line source with a uniform distribution between the vent ($1,666 \text{ m asl}$) and the plume top height. We used the software tool REFIR (Realtime Eruption source parameters FutureVolc Information and Reconnaissance system; Dürig et al., 2018) to constrain the time series of the plume top height and its uncertainty due to the radar type and its distance from the target. We used plume height data from the C-band ground-based radar of Keflavík airport (Arason et al., 2011) averaged over 1 hr; the plume top height uncertainty was estimated to be $\pm 1.2 \text{ km}$ (Dioguardi et al., 2020). The temporally varying MER was then obtained using the six 0-D buoyant plume models implemented in REFIR (four non wind-affected: Wilson and Walker (1987), Sparks et al. (1997), Mastin et al. (2009), and Gudmundsson et al. (2012); two wind-affected: Degruyter and Bonadonna (2012), and Woodhouse et al. (2013)), to obtain the best estimate of MER and related uncertainty. We consider three different solutions: (Hmin, MERmin), (Havg, MERavg), (Hmax, MERmax) as outlined in Dioguardi et al. (2020). REFIR allows assigning different weights to the six models, and the fine tuning of wind entrainment coefficients (Dürig, Gudmundsson, et al., 2023; Dürig, Schmidt, et al., 2023). In this study, each model was equally weighted, and the wind entrainment coefficient for Degruyter and Bonadonna (2012) was set to the default value of 0.5. The TGSD of the 2010 Eyjafjallajökull ash (Bonadonna et al., 2011), with diameters ranging from $0.1 \mu\text{m}$ to 8 mm (Table S2 in Supporting Information S1), was applied to the total mass released, where $\sim 50\%$ of the mass is assigned to particles with diameters $< 100 \mu\text{m}$, and the DFAF scaling was not applied. Particles are assumed to be spherical with a density of $2,300 \text{ kg m}^{-3}$ (Beckett et al., 2015).

4. Results

The results are separated into two sections: first we will describe the results of the simulated imagery experiments and the resulting conclusions about the sensitivity of the volcanic ash retrieval to ash cloud properties. We will then discuss how this impacts the comparison of satellite-retrieved data, in particular satellite-retrieved ash column loadings, with NAME dispersion model output.

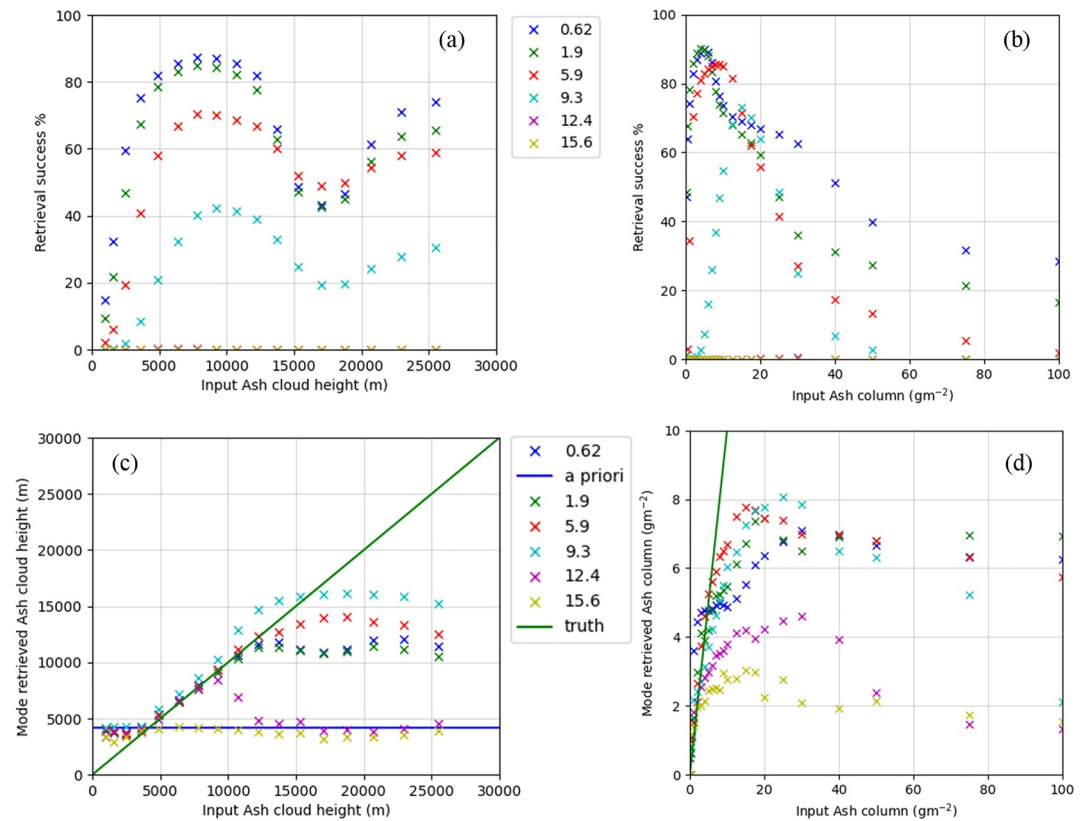


Figure 4. (a) Retrieval success and (c) average retrieved ash cloud height against input ash cloud top height and (b) retrieval success and (d) average retrieved column loading against input column loading for each size distribution labeled by median radius in μm . The retrieval success percentage indicates the percentage of pixels that the detection successfully marked as ash and the retrieval then retrieved a low enough cost solution to deem it successful. The average in (c) and (d) has been calculated by taking the mean of the modes for each set of ash cloud properties (height, depth, and column loading) weighted by retrieval success for given ash cloud properties. The optimum truth line is added to (c) and (d) in green and a line indicating the approximate altitude of the a priori height (~ 4.2 km as inferred from the a priori pressure for each model profile) used in the retrieval is shown in blue in (c).

4.1. Simulated Imagery Results

Figure 4a shows the retrieval success averaged across all depths and loadings for each size distribution for all points with a given ash cloud height. Similarly, Figure 4b shows the retrieval success averaged across all depths and heights for each size distribution for all points with a given ash column loading. Averaging in this way allows the impact of ash cloud height (a) and ash column loading (b) on retrieval success to be assessed individually, alongside how these dependencies vary with the size of the ash. It does not mean the retrieval success is independent of the variables being averaged over. It is also worth noting that the exact percentages on this graph are dependent on the data set being tested and as such it is more valuable to consider the differences in retrieval success as each variable changes on a given graph rather than the specific values. For example, the retrieval success at 9,000 m for the size distribution with a mass median of 1.9 μm (green crosses) is around 85%. The reason this is not closer to 100% is not necessarily a lack of sensitivity to the detection and retrieval of ash at this height, but an indication that there are a number of high and low column loadings that the detection and retrieval are not sensitive to at all heights. It is therefore more useful to say that the detection and retrieval are more sensitive to ash at 9,000 m than 2,000 m for example.

In Figures 4a and 4b, we can see that the retrieval success is close to zero for all heights and loadings for the two largest size distributions (those with median radius in terms of mass of 12.4 and 15.6 μm). As a result, some of the crosses for each are not visible where they overlap. For the 4 distributions with the smallest particle sizes, Figure 4a shows that the retrieval success, and therefore the sensitivity of the detection and retrieval, peaks around 4–12 km with a secondary peak at around 25 km in the stratosphere. This is expected as ash clouds below 4 km

will have temperatures closer to that of the surface so the BTD will not always be significant enough to allow for detection. The dip around the tropopause and lower stratosphere may be caused by optically thick ash clouds at these heights resulting in very cold BTs which are difficult for the retrieval to fit.

If the additional noise added to each pixel results in a colder BT than is present in the model profile, it may not be possible for the retrieval to reproduce this with any ash cloud properties. Only optically thick clouds with higher column loadings lead to the coldest simulated BTs that are difficult for the retrieval to fit. Thinner ash clouds around the tropopause result in warmer BTs that are easier to fit as there is more contribution from the surface than for the optically thicker clouds. In addition, the background cost of retrievals near the tropopause is larger due to the increased distance from the a priori height (~ 4.2 km), which is in the mid to lower troposphere. It is worth noting that the a priori height referred to here is not used within the retrieval methodology but is simply the height that the a priori pressure (600 hPa) approximately refers to. The exact value of the a priori height is dependent on the pressure profile and will therefore vary slightly for each of the pressure profiles used from the diverse profiles data set described in Section 3.1. The total cost is made up of the sum of the observation cost (the difference between the observed radiances and simulated radiances for the retrieved parameters) and the background cost (the difference between the retrieved parameters and the a priori parameters). The a priori ash cloud height is at ~ 4.2 km so when the retrieved height is significantly above this the background cost will increase. This means the maximum acceptable cost in the retrieval is more likely to be exceeded. As the temperature begins to increase again in the upper stratosphere, the retrieval tends to misplace the ash lower in the upper troposphere and this reduces the cost by bringing the retrieved height closer to the a priori than it would be at the tropopause. This increases the chance of the retrieval being successful.

In Figure 4b, we can see that the retrieval is most sensitive to column loadings between 0.5 and 20 g m^{-2} . This is largely as expected as the retrieval is reliant on a semitransparent ash cloud, and therefore clouds that are too optically thick or thin are challenging to detect and retrieve properties for. It is worth noting that the column loadings that the detection and retrieval are most sensitive to are strongly dependent on the size distribution. It can be seen in Figure 4b that the peak of sensitivity moves to higher column loadings as the size of the particles present increases. In fact, calculating the absorption coefficients for each of the size distributions shows that the column loadings where the retrieval success peaks for each distribution corresponds to an optical depth around 1. This is an interesting result, as it shows the retrieval is most strongly dependent on the height and optical depth of the ash. Much of the impact of changing size distribution or column loading is through how it impacts the optical depth.

However, it is worth noting that there is reduced sensitivity as the number of large particles increases in addition to the sensitivity peak moving to higher column loadings. This is because of the reduced difference in extinction between 10.8 and $12.0 \mu\text{m}$ as the number of larger particles increases. When the particle size is similar to or larger than the wavelength of the incident light Mie scattering occurs, preferentially scattering light forward and reducing the difference in extinction between 10.8 and $12.0 \mu\text{m}$. In contrast, the retrieval shows good sensitivity to very high column loadings, even greater than 50 g m^{-2} , for the smallest size distribution (mass median radius of $0.62 \mu\text{m}$). It is also interesting to note that the way the peak in sensitivity moves to higher column loadings as the number of larger particles increases is actually favorable for detecting ash clouds soon after an eruption, when both column loadings and the number of large particles present are higher.

Figure 4c shows the average retrieved ash cloud height against the true ash cloud height averaged over all depths and column loadings for each size distribution. Note that the average in (c) and (d) has been calculated by taking the mean of the modes for each set of ash cloud properties (height, depth, and column loading) weighted by retrieval success for given ash cloud properties. The retrieval is fairly accurate for the four smallest size distributions between 3 and 12 km but is unable to consistently discriminate heights above and below this. This is partially as expected as the impact of the stratosphere warming with height and the a priori height being in the troposphere is that the retrieval prefers a solution at or below the tropopause. As such, one might expect retrieved heights to reduce as the true height increases in the stratosphere as the retrieval places the ash at a location in the troposphere where the temperature is the same. Further investigation revealed that this did indeed occur for individual column loadings but is not visible in Figure 4c due to averaging over successful retrievals with retrieval success varying with both column loading and ash cloud height. The dip in retrieval success around the tropopause (between 12 and 20 km) in Figure 4a is primarily caused by reduced retrieval success of higher column loadings at these heights. As higher column loadings generally result in higher retrieved heights, as optically thick

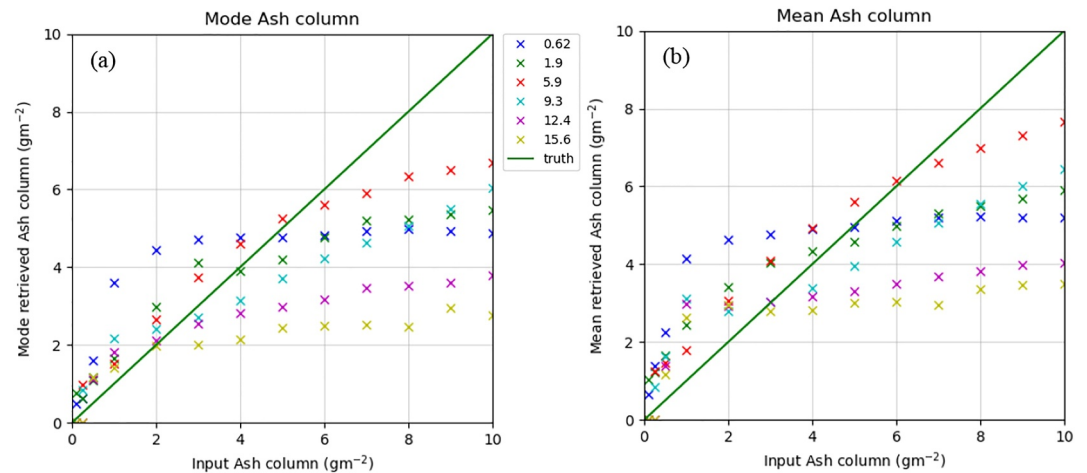


Figure 5. (a) Average (calculated as in Figure 4d) and (b) mean retrieved column loading against input column loading between 0 and 10 g m^{-2} for each size distribution labeled by median radius in μm . The green line shows where the optimal truth would lie.

clouds have less surface contribution and therefore retrieving heights significantly below the cloud height is unlikely, the average retrieved height in Figure 4c is effectively lowered around the tropopause by the dependence on the retrieval success. The region of sensitivity appears greatly reduced for the two largest size distributions (mass median radius equal to 12.4 and 15.6 μm), but this is based on a very limited number of successful retrievals as can be seen in Figures 4a and 4b. It is unclear why the retrieval appears to show reduced sensitivity to the true height below 3 km, but it is possible that low-level temperature inversions and the corresponding selection of profiles that allow successful detections may contribute to this, although a more detailed investigation of this is beyond the scope of the current paper.

Figure 4d shows the average retrieved ash column loading against the true ash column loading averaged over all depths and heights for each size distribution. The sensitivity to ash column loadings is restricted to column loadings less than around 10 g m^{-2} . As none of the average retrieved column loading values are greater than this, we have clipped the y-axis at 10 g m^{-2} . Above this true column loading, retrieved column loadings are increasingly underestimated. Figure 5 shows the (a) average (as calculated in Figure 4d) and (b) mean retrieved ash column loading against the true ash column loading averaged over all depths and heights for each size distribution for column loadings less than 10 g m^{-2} . This shows that for the largest five distributions the retrieval has good accuracy for column loadings below $2\text{--}6 \text{ g m}^{-2}$, with the distribution with a mass median of 5.9 μm showing good accuracy over the largest range of column loadings. For the smallest size distribution (blue crosses representing a mass median radius of 0.62 μm), there is little retrieval accuracy above $\sim 5 \text{ g m}^{-2}$ as with the other distributions. However, the retrieved column loading is also overestimated below this value for this distribution. The skewed nature of the distributions means that this overestimation appears more significant in the mean plot in Figure 5b. Although this overestimation is less significant for the other distributions, it is still present. This may be due to the a priori column loading being larger than most of these values ($\sim 3 \text{ g m}^{-2}$) and the retrieval therefore favoring higher column loadings, but investigating this further is beyond the scope of this study.

Figure 6a shows boxplots of the retrieval success against the true ash effective radius. The retrieval success reduces fairly steadily as the ash effective radius increases until it is very close to zero at and above $10 \mu\text{m}$. This is in line with expectations due to the reducing spectral dependence of extinction as the size of the particles increases and also in line with previous results (Prata et al., 2022; Stevenson et al., 2015; Wen & Rose, 1994). The large spread in the boxplots demonstrates that for some ash cloud properties (ash column loading, ash cloud height, and ash cloud depth) with effective radii $<10 \mu\text{m}$, the retrieval success can be very large, whereas for ash clouds with different properties it can be very low. For the size distributions with effective radii $<2 \mu\text{m}$, the median retrieval success is very high, which indicates that for most of the ash cloud properties tested the retrieval success is very high for these smaller particles. Similarly, for the size distribution with an effective radii close to $8 \mu\text{m}$ the median retrieval success is very low, indicating that for most of the ash cloud properties tested the retrieval success is very low for an ash cloud with this particle size distribution. Figure 6b shows the retrieval slightly overestimates ash

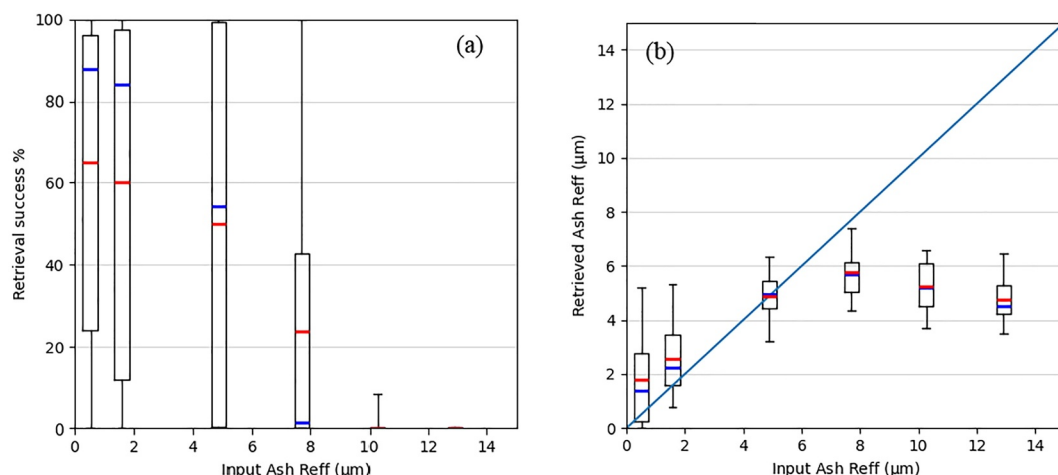


Figure 6. (a) Boxplot showing the maximum and minimum (whiskers), 25th and 75th percentile (box limits), mean (red line), and median (blue line) retrieval success for each of the unique ash cloud properties (ash column loading, ash cloud height, and ash cloud depth) against input effective radius, and (b) boxplot showing the 5th and 95th percentile (whiskers), 25th and 75th percentile (box limits), mean (red line) and median (blue line) retrieved effective radius for all simulations against input effective radius. In (b), the optimal truth line is plotted in light blue.

effective radius for small radii ($< \sim 2 \mu\text{m}$) and underestimates ash effective radius for large radii ($> \sim 7 \mu\text{m}$). Again, this is likely due to the a priori state vector. In many cases, similar radiances to the truth can be achieved by moving each of the retrieved variables slightly from their a priori values. This may be lower cost than moving just one of the retrieved variables far from its a priori value and therefore favored by the retrieval. However, for the two largest size distributions the retrieval success is very low (Figure 6a) so care should be taken when drawing conclusions from the retrieved data for the distributions with effective radii $> 10 \mu\text{m}$.

Figure 7a shows boxplots of the retrieval success against ash cloud height for all the simulations. We can see from the mean in Figure 7a that the retrieval sensitivity is good above about 2.5 km and peaks in the mid and upper troposphere (between 5 and 13 km). The retrieval sensitivity can also be seen to increase higher in the stratosphere ($> 17 \text{ km}$). Figure 7b shows boxplots of the retrieval success against ash column loading for all the simulations, and the mean demonstrates that the retrieval has good sensitivity between 0.25 g m^{-2} and 40 g m^{-2} . Note in this case “good sensitivity” is defined as the mean being greater than 20% in Figures 7a and 7b, but as noted previously, the dependence of the percentages on the chosen data set means the threshold for good sensitivity is quite subjective. The large spread in the box and whiskers indicates the importance of considering all of the ash properties to understand whether the satellite detection and retrieval are likely to be sensitive to a given ash cloud. For example, for all but the lowest ash cloud heights, the retrieval success can vary from 0% to 100%, indicating that the other ash cloud properties would need to be considered to determine whether the satellite detection and retrieval would likely be successful. This is a large part of the reason why precisely defining single thresholds for the sensitivity of satellite-retrieved ash data to different ash cloud properties is not possible. Figure 7c confirms that the retrieval can accurately retrieve heights between 4 and 12 km but is unable to consistently discriminate heights above and below this. The boxplot whiskers also show that the retrieval does not retrieve heights in the stratosphere with significant frequency. Figure 7d demonstrates the increasing underestimation in retrieved column loadings as the true column loading increases. Plotting on a logarithmic scale also demonstrates the overestimation in retrieved column loadings for low column loadings. The boxplot whiskers show that the retrieval can produce values as low as 0.4 g m^{-2} and as high as 30 g m^{-2} .

Figure 8 demonstrates the impact changing the depth of the simulated ash cloud has on the retrieval. Figure 8a shows boxplots of the retrieval success against the ash cloud depth. Each boxplot includes data for all ash cloud properties simulated with ash present in the number of model levels given by the value in the x-axis. As model levels become deeper with increasing altitude, this was considered the best way to separate the impact of increasing depth from the impact of increasing altitude. There is only a very weak trend of increasing sensitivity of the retrieval as the depth of the ash cloud increases. This is true for the mean, median, and spread seen in the boxplots, indicating that the retrieval success is much more strongly dependent on the other ash cloud properties.

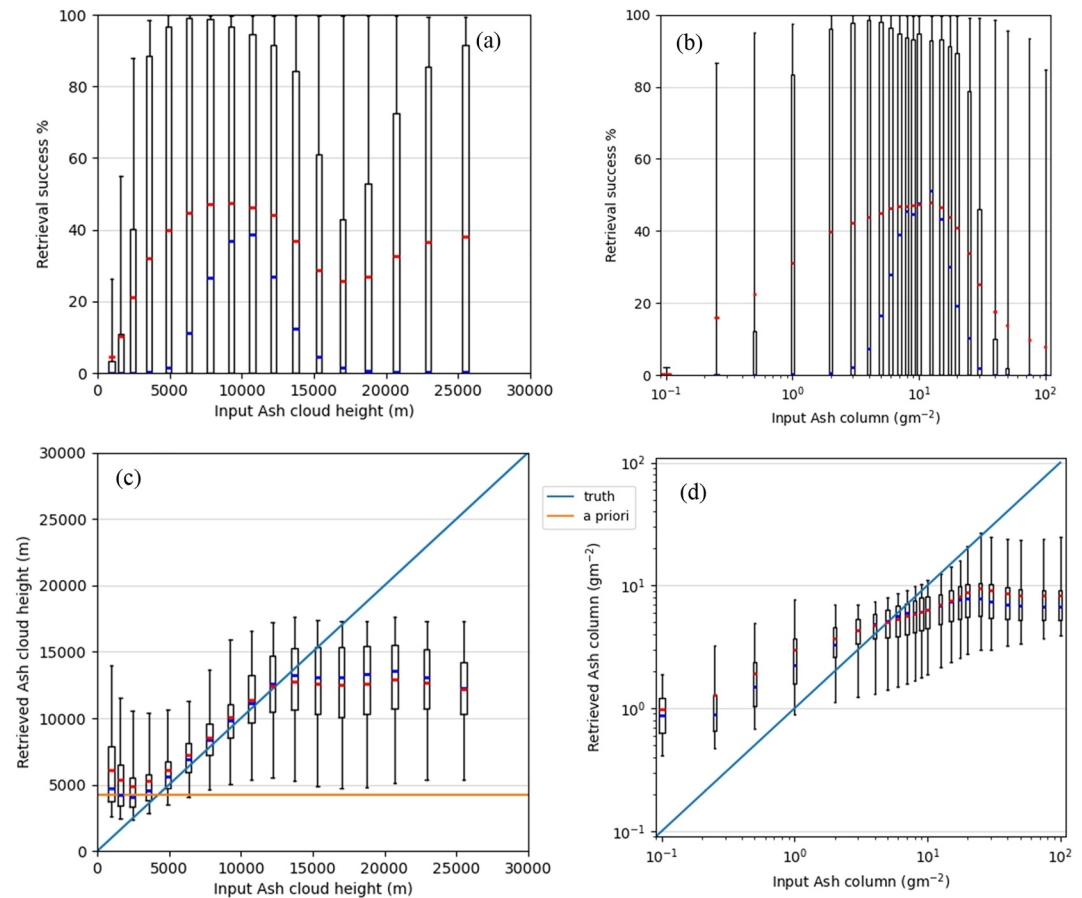


Figure 7. Boxplots showing the 5th and 95th percentile (whiskers), 25th and 75th percentile (box limits), mean (red line) and median (blue line) retrieval success for each of the unique ash cloud properties against (a) input ash cloud height and against (b) input ash column loading. In (a), the boxplot shows the data from the simulations of all of the column loadings, size distributions, and depths (described in Section 3.1.2), and in (b), the boxplot shows the data from the simulations of all of the heights, size distributions, and depths. Boxplots showing the 5th and 95th percentile (whiskers), 25th and 75th percentile (box limits), mean (red line) and median (blue line) retrieved ash (c) cloud height and (d) column loading for all simulations. The optimum truth line is added to (c) and (d) in light blue and a line indicating the approximate altitude of the a priori height used in the retrieval is shown in orange in (c).

Figure 8b shows the retrieved ash cloud height against true ash cloud height for each of the depths tested at 7 g m^{-2} for our default size distribution (mass median radius of $1.9 \mu\text{m}$). For each of the depths the retrieval is fairly accurate between 4 and 16 km, with retrieved heights overestimated below this and increasingly underestimated above this. Throughout the mid and upper troposphere, the deeper the ash cloud is the lower the retrieved height. This is expected as the deeper clouds will have more ash at lower heights and therefore warmer temperatures, so the overall BT of the cloud will be higher and the retrieval will find a height from lower within the cloud. Whereas in the stratosphere, this effect is largely flipped with thinner ash clouds being retrieved as lower altitude. Again, this is expected as increasing the depth in the stratosphere places more ash lower where it is cooler and therefore reduces the overall BT making it effectively appear to be higher in the troposphere to the retrieval. Figure 8b also shows the expected reduction in retrieved heights with increasing input heights in the stratosphere that was not visible in Figures 4 and 7c due to the averaging and impact of retrieval success varying with column loading. This behavior is also present for other column loadings and size distributions tested but is illustrated most clearly by the properties used in Figure 8b.

4.2. Comparison to Dispersion Model Output

We now compare NAME modeled and MSG-SEVIRI satellite-retrieved ash column loadings of the ash cloud from the eruption of Eyjafjallajökull (Iceland) in 2010. We consider the implications of applying the identified

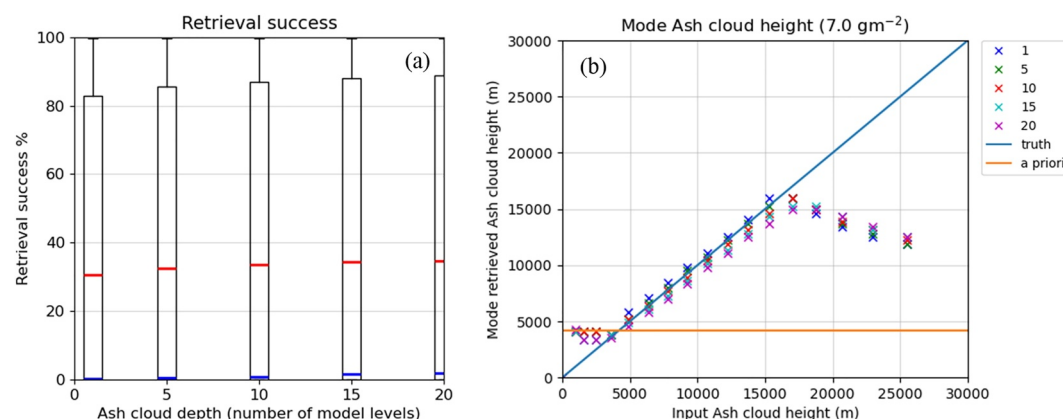


Figure 8. (a) Boxplot showing the 5th and 95th percentile (whiskers), 25th and 75th percentile (box limits), mean (red line) and median (blue line) retrieval success for each of the unique ash cloud properties (ash column loading, ash cloud height, and particle size distribution) against depth and (b) an example plot of the mode retrieved height against input ash cloud top height for each ash cloud depth (for a column loading of 7 g m^{-2} and for the size distribution with a median radius of $1.9 \mu\text{m}$). The legend indicates the relative depth of the ash cloud by showing the number of model levels containing ash for each point. The optimum truth line is added to (b) in blue and a line indicating the approximate altitude of the a priori height used in the retrieval is shown in orange.

limitations of the satellite retrieval to the modeled output for the comparison. To assess the sensitivity, the limitations are applied in a binary method, where it is assumed that ash that the satellite detection and retrieval are less sensitive to will never be successfully detected and retrieved, and that ash that the detection and retrieval are more sensitive to will always be successfully detected and retrieved. We use our judgment to analyze our results from Section 4.1 and identify the specific values of particle size, column loading, and height where we would consider the satellite retrieval to be insensitive in the comparisons shown in Figures 9–11. The NAME output is therefore altered by the removal of the ash the satellite detection and retrieval are less sensitive to.

Figure 9 compares the satellite-retrieved ash column loadings to NAME modeled ash column loadings on 6 May 2010 at 12:00 UTC. Here, the satellite data has been included primarily to provide spatial context to the NAME extents and the color scale has been chosen to illustrate the changes in NAME extent as the restrictions are applied. In Figure 9b, the NAME model output is generated using the ESPs outlined in Section 3.3 and the temporally varying plume top height and MER from REFIR (HAvG, MERAvG). The total mass in the atmosphere is predicted to be $2.27 \times 10^9 \text{ kg}$, which is 7.8 times greater than the total mass retrieved by the satellite ($2.90 \times 10^8 \text{ kg}$). Note that, for Figures 9–11, the total mass in the satellite-retrieved data only includes mass at latitude $>30^\circ$ and $-50^\circ < \text{longitude} < 10^\circ$ to not include mass from any pixels far from the ash cloud which have been falsely marked as containing ash by the detection algorithm.

When the identified limitations associated with the satellite data are applied to the NAME output, the total modeled mass is reduced and more closely matches the satellite retrieval. However, the intention of this comparison is not to achieve a better match between the satellite-retrieved and NAME model output total masses. It is to allow us to analyze how significant the impact of the limitations of the satellite data on our interpretation of the event is. Although it is interesting to note that for this case a better match to the satellite-retrieved total mass is found once all of the restrictions have been applied. The spatial extent of the ash cloud is also reduced in better agreement with the satellite data. In Figure 9c, we only consider NAME modeled ash column loadings associated with particles with diameter $<16 \mu\text{m}$, and this reduces the total modeled mass in the atmosphere to $1.64 \times 10^9 \text{ kg}$, 72% of the original mass. The spatial extent is largely unchanged due to this restriction, although the 10 g m^{-2} contour is reduced in area, indicating that most of the larger modeled particles remaining in the atmosphere in (b) were in the near-source region as one might expect. In Figure 9d, we only include the mass from particles with diameter of $<16 \mu\text{m}$ and at altitudes of $>3 \text{ km}$, and this further reduces the total modeled mass to $4.01 \times 10^8 \text{ kg}$, 18% of the original mass. For this case, this restriction was the most significant for both reducing the total mass and spatial extent of the NAME data. This indicates that most of the mass in the NAME output is below 3 km, particularly further from the source where particles would have had the most time to descend. Finally, in Figure 9e, we apply the preceding restrictions to the NAME data and in addition only consider ash column

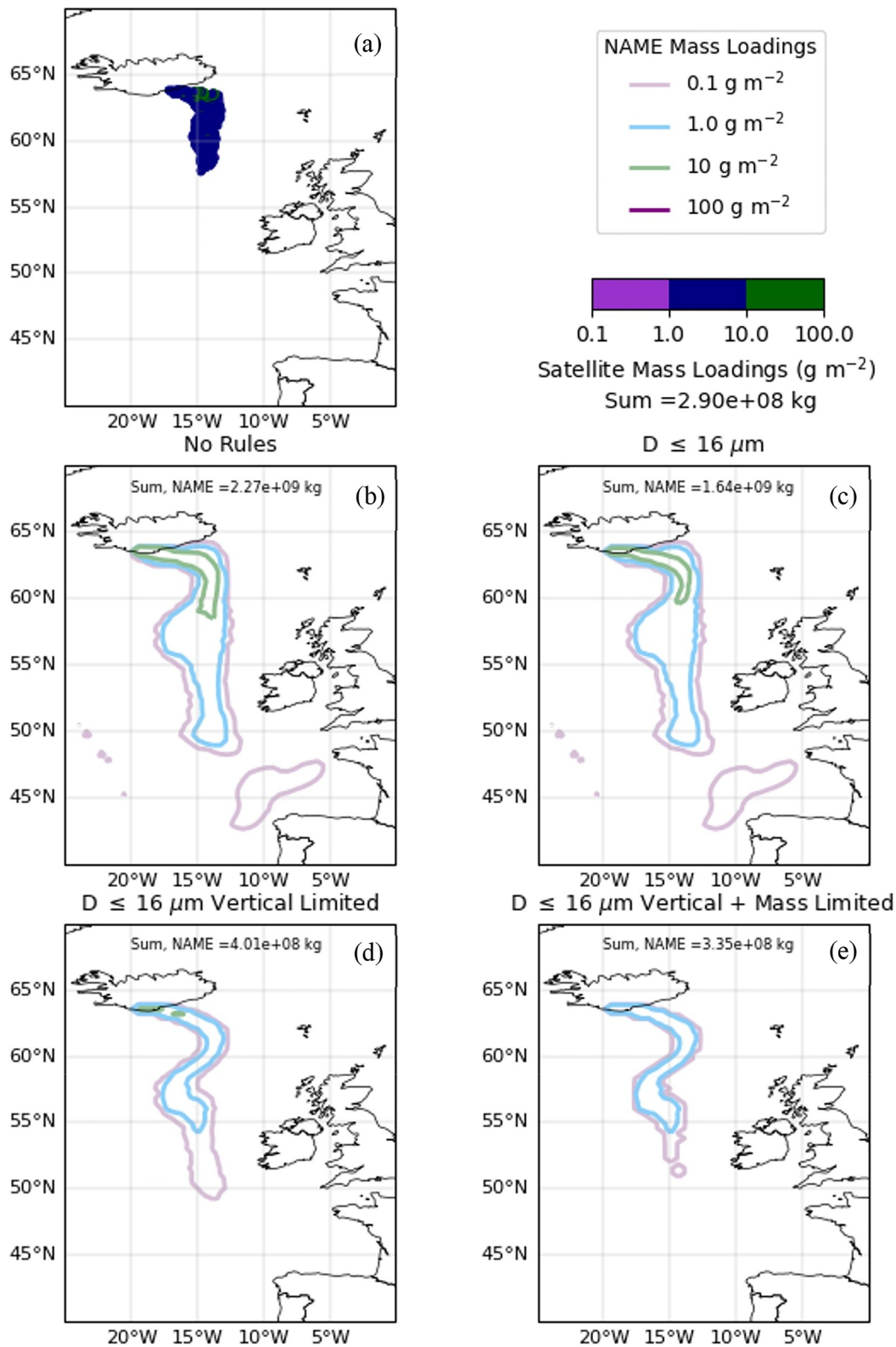


Figure 9.

loadings which are $>0.5 \text{ g m}^{-2}$, and when modeled ash column loadings are $>7 \text{ g m}^{-2}$, the value is limited to 7 g m^{-2} to reflect the fact that the satellite retrievals are saturated above these loadings. This reduces the total modeled mass to $3.35 \times 10^8 \text{ kg}$, 15% of the original mass. In this specific case, this restriction is not very impactful on either the total mass or the extent of the NAME data as, following the application of the preceding restrictions, there is not a significant number of column loadings above 7 g m^{-2} or below 0.5 g m^{-2} . Figure S1 in Supporting Information S1 shows how the distribution of the modeled mass compares to the satellite retrieval, as the identified limitations of the satellite retrievals are applied the distribution of mass is confined to a narrower range of column loadings in better agreement with the satellite-retrieved distribution. After applying all the restrictions, the total mass in the NAME output more closely matches the satellite output with 87% of the NAME mass now represented by the satellite retrieval.

Figure 10 shows a time series of total mass in the NAME output with no restrictions with each of the restrictions applied as in Figures 9c–9e, and the total mass in the satellite-retrieved data at 12:00 UTC from 6 May 2010 to 14 May 2010. On each day, the total mass in the satellite-retrieved data is less than the total mass in the NAME output with no restrictions applied. For all days except 6 May 2010, the total mass in the satellite-retrieved data is higher than the total mass in the NAME data with all of the restrictions applied. For this phase of the 2010 Eyjafjallajökull eruption, applying the vertical limiting rule, which removes mass below 3 km altitude, is the most impactful in reducing the total mass in the NAME output. Each stage of applying the restrictions removes significant mass from the NAME output.

It is important to recognize the uncertainties associated with the source terms used by the NAME simulations, in particular the plume height and MER, which can significantly impact the predicted downwind ash column loadings in the atmosphere. Figure 11 shows the NAME predicted ash column loadings when (Hmax, MERmax) and (Hmin, MERmin) output from REFIR are used alongside the satellite data (a) as in Figure 9. Using (Hmax, MERmax) the predicted total mass in the atmosphere at 12:00 UTC on 6 May is $3.51 \times 10^9 \text{ kg}$, which is 12.1 times greater than the satellite-retrieved mass (Figure 11b). This is reduced to $4.69 \times 10^8 \text{ kg}$ when the identified satellite data limitations are applied, and as such 62% of the modeled mass is now represented by the satellite-retrieved mass (Figure 11c). When (Hmin, MERmin) are used, the total modeled mass in the atmosphere is reduced and is 3.0 times more than the satellite retrieval, and when all the limitations of the satellite retrieval are applied, the modeled total mass is just $1.26 \times 10^8 \text{ kg}$ and is underestimating the mass in the atmosphere by a factor of 2.3 compared to the observations.

5. Discussion

We have shown that simulated imagery can be used to assess the performance and sensitivity of the volcanic ash detection and retrieval. Although there are differences between simulated and real imagery, the benefit of knowing the truth in the verification data set vastly outweighs this limitation when using simulated volcanic ash imagery. Despite this, investigations into the sensitivity of the results to the method of simulating the radiances, including the radiative transfer model used and assumptions about the ash composition, would be valuable. Millington et al. (2012) discuss the sensitivity of the simulated BTDs to the composition assumed and show notable differences in Dust RGB imagery when the refractive indices for volcanic dust (Volz, 1973), andesite ash (Pollack et al., 1973), obsidian (rhyolite) ash (Pollack et al., 1973), and desert dust with 0.9% hematite (Balkanski et al., 2007) are assumed. Although we would expect significant quantitative differences in our results were we to repeat the investigation with a different assumed composition in the production of the simulated radiances, in particular with a different composition than that assumed within the retrieval, we would expect qualitatively similar results.

The results shown in this paper are strictly only applicable to the Francis et al. (2012) retrieval scheme applied to MSG-SEVIRI data. However, given the intrinsic limitations of the spectral data, other similar retrieval schemes using TIR wavelengths are likely to show results that are at least qualitatively similar. Nonetheless, running

Figure 9. (a) Satellite-retrieved and (b, c, d, and e) NAME modeled ash column loadings for 12:00 UTC on 6 May 2010. NAME simulations are initialized using the temporally varying average plume top height and MER generated from REFIR. Figure (b) shows the raw NAME output, the subsequent figures show the NAME output corrected to align with the identified satellite limitations: (c) only particles with diameter $\leq 16 \mu\text{m}$ are included, (d) in addition only the particles residing at altitudes $>3 \text{ km}$ are included, and (e) the model output is further limited to only include ash column loadings $>0.5 \text{ g m}^{-2}$ and where ash column loadings are $>7 \text{ g m}^{-2}$ they are limited to 7 g m^{-2} .

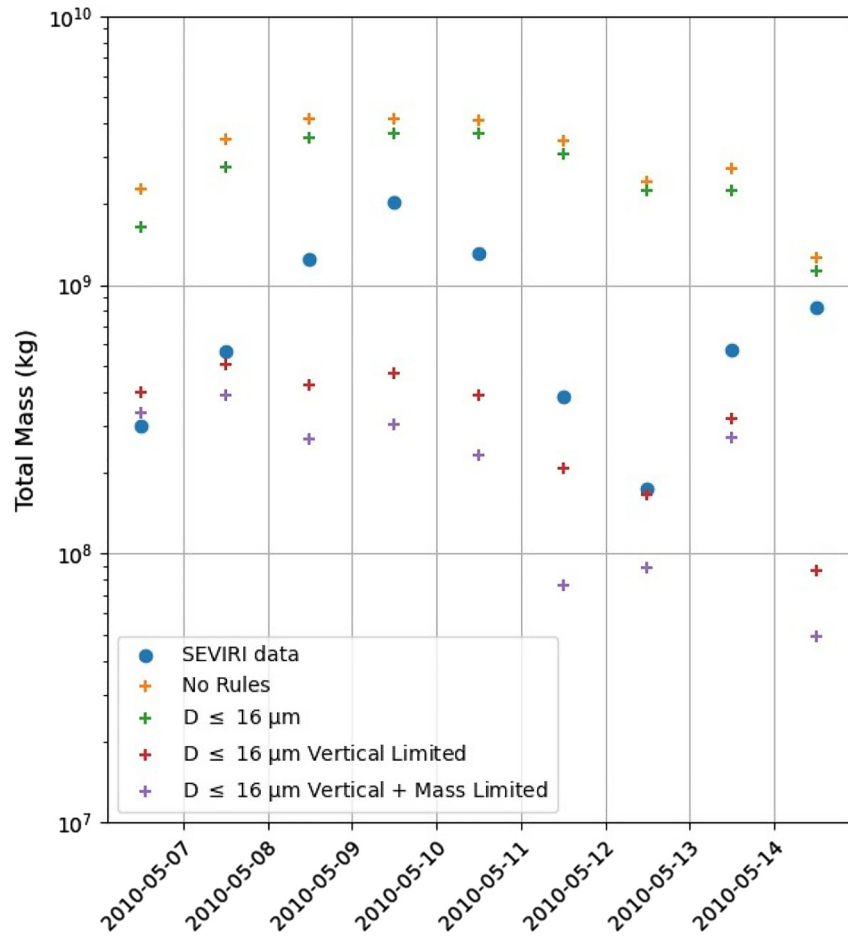


Figure 10. Time series comparing the total mass in the satellite-retrieved data with the total mass in the NAME output with no restrictions applied, and with each of the restrictions as applied in Figures 9c–9e at 12:00 UTC from 6 May 2010 to 14 May 2010.

different detection and retrieval schemes on the simulated data set would be encouraged, as well as testing improvements to this scheme with the data set. Using the same retrieval scheme, Stevenson et al. (2015) found a similar leveling off in retrieved radius above 5 μm. Piontek et al. (2021) found qualitatively similar results to our retrieved loading results with retrieved optical depth overestimated at low optical depths (column loadings) and retrieved optical depth underestimated at high optical depths. Additionally, investigations into the impact of meteorological cloud and varied land surfaces would be encouraged. Similar investigations to understand the performance of retrieval methods on different instruments (e.g., MODIS, VIIRS (Visible/Infrared Imager Radiometer Suite), ABI (Advanced Baseline Imager), and AHI (Advanced Himawari Imager)) would also be useful. Differences in performance stemming from differences in the instruments spectral response functions are difficult to predict, although sensitivity to lower column loadings would be expected to increase for instruments with higher spatial resolution. Using simulated imagery, we have been able to both confirm and constrain a number of previously known limitations for the detection and retrieval of volcanic ash at TIR wavelengths. In particular, we can see from Figures 6a, 7a, and 7b that we would not expect the satellite retrieval to be sensitive to ash below ~3 km, below ~0.25 g m⁻², and above ~40 g m⁻² and for particles larger than ~10 μm in radius.

When using the results from Section 4.1 to compare dispersion model output and satellite-retrieved data, we used a binary approach to apply the satellite retrieval limitations to the NAME model output. Although a more sophisticated approach, which considers the probability of ash being detected, would be preferred, this approach indicates how much of the mass that the satellite retrieval may be insensitive to is present in the NAME output. Similar approaches, where selected portions of the model mass were removed from comparisons with satellite data, have been used in the past including Crawford et al. (2016) who did not consider modeled mass below 2 km

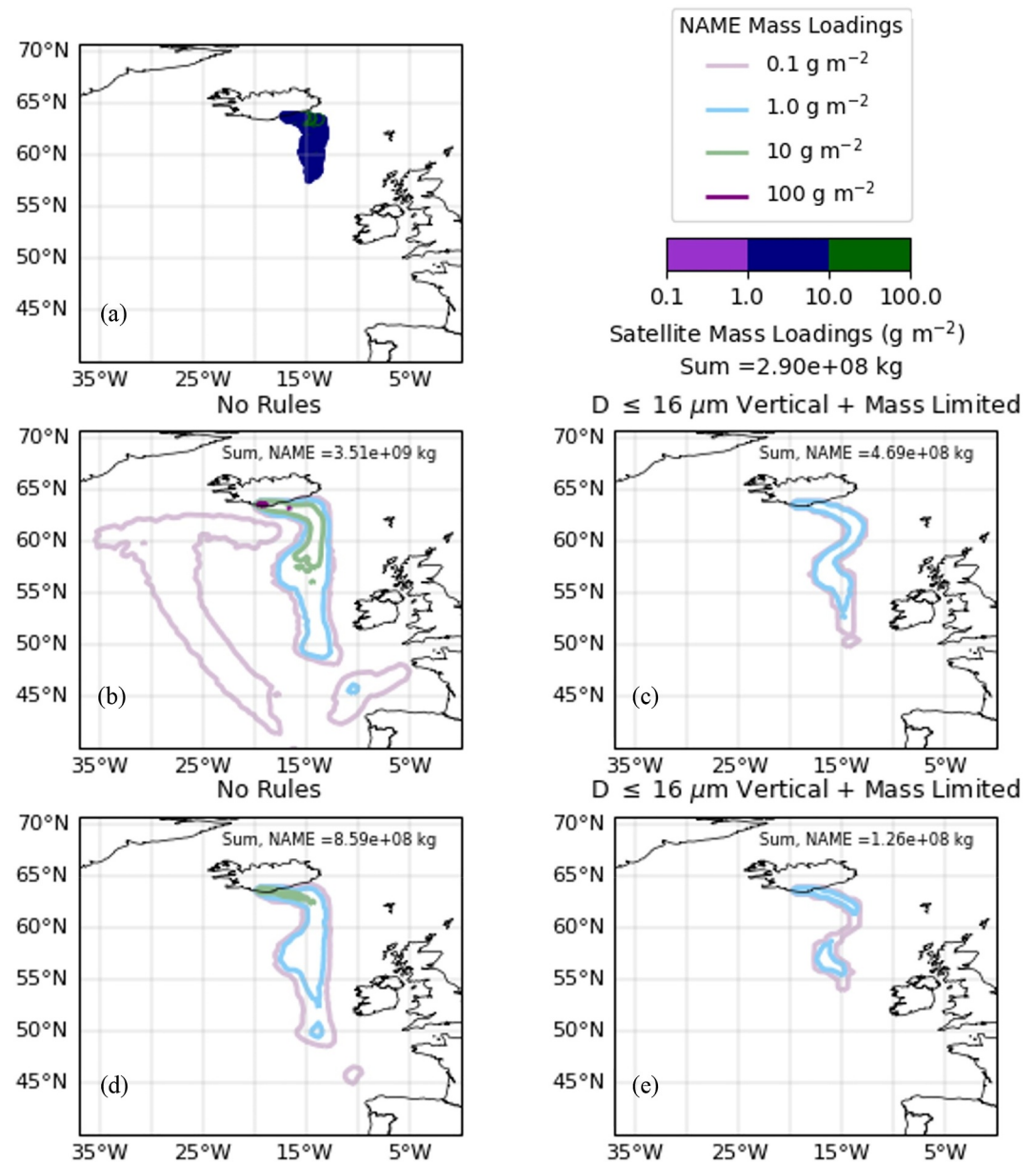


Figure 11. (a) Satellite-retrieved ash column loadings and the NAME modeled ash column loadings for 12:00 UTC on 6 May 2010 generated using (HMax, MERMax) (b) with no retrieval limitations and (c) with all identified satellite retrieval limitations applied to the model data, and using (HMin, MERMin) (d) with no retrieval limitations and (e) with all identified satellite retrieval limitations applied to the model data.

and only considered modeled loadings $\geq 0.1 \text{ g m}^{-2}$ to be consistent with the satellite detection identifying ash affected pixels. Crawford et al. (2022) note the importance of correcting any biases in the observations before comparing these to the modeled column loadings.

Our findings suggest that a significant proportion of the total mass in the dispersion model output can lie outside of the satellite retrieval detection limits; in the example shown in Figure 9, the total satellite-retrieved mass is similar to the total mass in the dispersion model output with the identified satellite retrieval limitations applied. However, this does not mean that once the satellite retrieval limitations are considered there is generally agreement between the satellite-retrieved total mass and modeled total mass. The proportion of mass in the dispersion model that is outside the satellite retrieval's sensitivity range varies significantly for different cases and even for the same eruption at different times. Although in this case and time $1.9 \times 10^9 \text{ kg}$, or 85% of the original

mass, is shown to be outside the identified sensitivity range of the satellite retrieval, this number can vary from very small values (<10%) to very large values (>95%). For example, in Figure 10, it can be seen that significantly more mass is removed from the NAME output when the restrictions are applied on 11 May compared to 7 May. More significant variation in the amount of mass removed would likely be seen if other eruptions were considered. Well-dispersed ash clouds in the upper troposphere containing mostly moderate column loadings and smaller particles would likely not be affected significantly by applying the restrictions and better agreement between the satellite-retrieved data and the output of VATDMs should be expected. This agreement may then decrease as the ash cloud further disperses and column loadings below the sensitivity range of the satellite detection algorithm become more prevalent.

The relative significance of the application of each of the restrictions is also likely to be highly dependent on the specific eruption being considered and the age of the ash cloud. Ash clouds which have recently been produced by an eruption with a high plume height are likely to be most impacted by the application of the particle size restriction and maximum column loading rescaling, whereas ash clouds that have been dispersing and descending for a long time are likely to be most impacted by the application of the altitude restriction and the removal of low column loadings. It is also worth noting that the sequential application of the restrictions will have reduced the impact of applying the vertical limiting and especially the mass loading restrictions. This is because the application of the preceding restrictions will have removed mass that would otherwise be removed by these restrictions. For example, in Figure 9 the area of column loadings $>10 \text{ g m}^{-2}$ is reduced by applying the restriction on diameter and the impact of later applying the upper column loading restriction is therefore reduced.

We have shown that applying the identified sensitivities of the satellite retrievals (to particle size, altitude, and ash column loadings in the atmosphere) to dispersion model output can strongly influence what is inferred from comparing the two data sources. It is therefore important that the limitations of both the satellite and model data are considered for forecast verification and when using satellite data for calibration of MERs, including in inversion methods, data assimilation, and wider data fusion techniques. It is also important to correctly interpret what our results show. We have demonstrated that significant proportions of the mass present in the dispersion model output can be outside the satellite retrieval's sensitivity range, and therefore the quantitative satellite data does not provide evidence that this ash is not present in the atmosphere. However, this does not by extension mean that the ash outside the retrieval's sensitivity range is present. It is possible that this ash is not present and errors in the ESPs, meteorological data, and the representation of near-source processes could contribute to an over-estimation of ash mass in the dispersion model output. In some cases these errors are clearly present, such as the example shown in Figure S2 in Supporting Information S1, where the dispersion model does not appear to have captured the location of the entire ash cloud accurately when compared to the satellite-retrieved data. Understanding how much ash outside the retrieval's sensitivity range is present in the atmosphere is essential for future quantitative ash forecasts, as we have shown this could represent more ash than is within the satellite retrieval's sensitivity range. Further work comparing observations with different limitations will likely be crucial to improving this understanding.

Discrepancies between dispersion model output and satellite retrievals are often attributed to the lack of representation of near source processes in the modeling, such as aggregation. Our findings suggest it is possible limitations in the satellite-retrieved data could account for some of these discrepancies, and that near source processes removing large quantities of ash may not be required to achieve consistency in the data from different sources. This also supports the findings from the tephra community studying deposits, including cryptotephra, which suggest that large amounts of ash can be transported long distances from the source (e.g., Cashman & Rust, 2019), and new evidence which shows that aggregation can, in some cases, extend the lifetime of the ash in the atmosphere (Rossi et al., 2021).

In contrast, there is evidence that much of the mass outside the sensitivity range of the satellite retrieval does not remain in the distal ash cloud. For example, Prata and Prata (2012) used aircraft, ground-based lidar, and air quality data to validate satellite-retrieved ash column loadings and found good qualitative agreement as did Newman et al. (2012) when comparing IASI (Infrared Atmospheric Sounding Interferometer) retrievals and aircraft observations. It is worth noting that the validation methods have their own sensitivities and limitations. For example, aircraft observations are often limited to less dense parts of ash clouds for safety reasons (Johnson et al., 2012).

Limitations in satellite-retrieved volcanic ash data are also dependent on the detection and retrieval scheme used. The VOLCAT (VOLcanic Cloud Analysis Toolkit) system achieves significant improvements in the detection of optically thick ash clouds by considering the rate of growth of the umbrella cloud (Pavolonis et al., 2018) and improvements in the detection of optically thin ash by identifying cloud objects to increase the likelihood of ash being identified if neighboring pixels contain ash (Pavolonis et al., 2015a). Some biases shown here for the Francis et al. (2012) retrieval scheme may also be less significant for other retrieval schemes. For example, schemes that use alternative a priori probability distributions, such as lognormal probability as a function of the retrieved variables, may well be able to more consistently retrieve higher column loadings than the Francis et al. (2012) method.

Additionally, even in scenarios where the ash cloud may be out of the satellite detection and retrieval's sensitivity range (e.g., if the dispersion model output indicates this), the satellite may still provide useful information. This is because many of the ash clouds that the detection and retrieval would not be sensitive to will still produce a noticeable signal in the satellite imagery. In particular, scenarios where there is a lot of ash often result in optically thick clouds which the automated detection is unable to discern from meteorological cloud. However, a skilled observer will often be able to identify these as volcanic clouds from an animation of the Dust RGB where the cloud can easily be tracked from the volcano. As such, even where the satellite detection or retrieval have failed, it will likely be possible to assess whether this is because there is too much or too large ash, or whether it is because significant ash simply is not present. Similarly, where the biases and limitations of the detection and retrieval method are well-known, information can be inferred from whether the detection and retrieval were successful. For example, where the Francis et al. (2012) ash detection and retrieval have successfully identified ash, the likelihood of extremely high column loadings being present is reduced as the likelihood of successful detection and retrieval decreases as the column loadings present increase (above $\sim 10 \text{ g m}^{-2}$).

Furthermore, the limited available microwave satellite data indicate that a significant proportion of the mass released by a volcano may be on particles that are too large to reach the distal ash cloud in significant amounts. Marzano et al. (2018) found a fine ash fraction of $\sim 1\%$ by comparing estimates of total mass in TIR satellite retrievals with ground deposit total mass estimates for the April 2015 Calbuco eruption. In the same study, they found that total mass estimates from MW observations near the source were comparable with the total mass estimated from ground deposits. MW observations from 14 hr later indicated that little of the ash observed immediately after the eruption that the MW data were sensitive to was still present. The rapid reduction in mass in the MW observations is consistent with the sensitivity of the instruments, which are only able to observe more significant volcanic eruptions and the subsequent ash cloud, and are only sensitive to the larger particles in the erupted cloud. The precise lower limit of size that the MW observations are sensitive to is not well-defined with values ranging from 10 to 50 μm in radius quoted in previous studies (Marzano et al., 2013, 2018; Romeo et al., 2023). As such, it is not clear how much mass may be present that neither the TIR nor MW observations are sensitive to. However, given the good agreement between total mass present in the MW observations and ground deposit data, it is likely most of the ash is within the two observations' combined sensitivity range. Further evidence for this is provided by Corradini et al. (2016) where they find the retrieved size distribution from ground-based MW data gives comparable results to the retrieved size distribution from SEVIRI observations around 10 μm radius.

The emergence of new evidence, including the results of this investigation, has potential implications for the DFAF applied at the London VAAC. It is important to consider the sensitivity and uncertainty of any observations used for setting default initial values for the DFAF and when modifying the DFAF in near-real time eruption response. Given the DFAF is known to vary significantly between eruptions, it would be extremely valuable if the processes dictating the proportion of mass ejected from the volcano that reaches the distal cloud were better understood, monitored in real time, and modeled, such that more realistic and dynamic values for DFAF could be determined in near real time. This would constrain a significant uncertainty that is currently poorly defined in the forecast ash concentrations.

Despite the unanswered questions, we can make recommendations for best practice in using satellite data during operational response in future based on the results of this investigation. The current operational process (at the London VAAC) is to use the satellite data, including the Dust RGB, to evaluate the NAME model forecast. Simulated imagery is a useful tool to aid this process, with comparison of the real and simulated Dust RGBs allowing for a good comparison of the location of ash in the observations and dispersion model output. It can also

be useful for quantitative comparison of mass loadings as the loadings retrieved from the simulated image can be compared to the loadings retrieved from the real image. This reduces the impact of any biases in the satellite-retrieved data, because the biases present in the loadings retrieved from the real image are likely to be similar to the biases in the loadings retrieved from the simulated image. The QVA requirement to produce quantitative data with associated uncertainties means that the process for utilizing satellite-retrieved volcanic ash data is likely to become increasingly automated, both at the London VAAC and other centers. As such, the focus for producers of satellite-retrieved data is likely to be how it can be used in automatic verification and assimilation methods. The data from this study could be used to create a product which reflects what column loadings are consistent with the satellite-retrieved data. Probability distributions of the likely true column loading present in the atmosphere given a retrieved column loading could be produced and used within verification and inversion schemes. This would automatically account for the identified limitations of the satellite-retrieved data.

At the London VAAC, a source inversion model called InTEM (Inversion Technique for Emissions Modeling; Pelley et al., 2021) can be run operationally. Inversion schemes, such as the InTEM tool, are used to provide revised time-varying MER and vertical distribution of ash at the vent, which can correct both the location of the modeled ash cloud as well as the modeled ash column loadings to better match the satellite observations. InTEM infers the optimum ESPs based on the amount of ash retrieved from the observations and a prior estimate of the ESPs. Currently within InTEM, the satellite's sensitivity range is compensated for simply through limiting the PSD used to between 1 and 30 μm in diameter. It is therefore implicitly assumed that the satellite is sensitive to all fine ash in the atmosphere. As such, it may be better to use a product that has accounted for the identified limitations of the satellite data, such as the one described above, within the inversion. Alternative approaches, such as producing simulated radiances with the dispersion model data and running the detection and retrieval algorithm on these to compare to the retrieved loadings from the real imagery, could also be considered. Accounting for the satellite limitations in InTEM is likely to lead to increased reliance on (i.e., give more weight to) the a priori estimate (in effect a first guess model forecast) for modeling the part of the ash cloud which is outside of the satellite sensitivity ranges.

6. Conclusions

We have demonstrated the value of simulated satellite imagery in assessing the performance of a satellite ash detection and retrieval algorithm with a known truth. We have shown that TIR satellite retrievals using the method of Francis et al. (2012) are most sensitive to ash loadings of between 0.25 and 40 g m^{-2} , at altitudes higher than 3 km above ground level, and for particles smaller than 10 μm in radius. We also showed that, for this retrieval method, retrieved column loadings become increasingly underestimated when more than $\sim 7 \text{ g m}^{-2}$ of ash is present. The results of this investigation have important implications for the use of satellite data in verifying dispersion model forecasts in an operational environment. In many scenarios, there will be significant uncertainty in the amount of ash present in the atmosphere that may have gone undetected or underestimated by satellite detection and retrieval algorithms. Although comparisons to other observations indicate good agreement with satellite data, limitations in the other observations may also lead to systematic underestimations. As such, further work comparing different forms of observations and dispersion model output with consideration of their respective limitations would be valuable in constraining the possibility of significant systematic errors from satellite and other observations. In particular, further studies with both MW and TIR data combined, and studies defining the sensitivity of MW observations to different particle sizes and ash column loadings, would be very useful for understanding how much ash typically remains in the distal ash cloud.

Data Availability Statement

All retrieved ash data, data used in the NAME mass comparisons, and data used by REFIR are available at Zenodo (Saint et al., 2024). The simulated radiances were produced using a version of SPS (Satellite Processing System) that is stored in internal Met Office repositories alongside the python plotting code used to produce Figures 1–8. Unfortunately, there is no existing license to make this software openly available, but interested parties are encouraged to contact the author who will assist in sharing as much relevant code as possible. The software to simulate the radiances included the use of RTTOV v13.0 (Saunders et al., 2020). The NWPSAF profile data (Eresmaa and McNally, 2012) used to set the meteorological data in the simulations and retrievals are available at <https://nwp-saf.eumetsat.int/site/software/atmospheric-profile-data/>. NAME Version 8.1 was used in the

dispersion model simulations. The UK Met Office NAME model and UM output to drive NAME are available via license from the UK Met Office (© Crown Copyright, Met Office) and interested parties can contact the Atmospheric Dispersion team at: atmospheric.dispersion@metoffice.gov.uk. Figures 9–11, Figures S1 and S2 in Supporting Information S1 were produced with code stored in an internal Met Office repository. REFIR v20.1 was used and is available at Zenodo (Dürrig & Dioguardi, 2020).

Acknowledgments

We would like to thank Luke Western and Andy Prata for inspiring discussions on uncertainty in satellite-retrieved volcanic ash data, and Pete Francis, Dave Thomson, Matthew Hort, and Samantha Engwell for their feedback on the manuscript. This work is published with permission of the Executive Director of British Geological Survey (UKRI). The work discussed here has been possible due to funding from the UK Civil Aviation Authority. We would also like to thank Peter Colarco and two anonymous reviewers for their useful comments which improved the paper.

References

- Ansmann, A., Tesche, M., Seifert, P., Groß, S., Freudenthaler, V., Apituley, A., et al. (2011). Ash and fine-mode particle mass profiles from EARLINET-AERONET observations over central Europe after the eruptions of the Eyjafjallajökull volcano in 2010. *Journal of Geophysical Research*, *116*, D00U02. <https://doi.org/10.1029/2010JD015567>
- Arason, P., Petersen, G. N., & Björnsson, H. (2011). Observations of the altitude of the volcanic plume during the eruption of Eyjafjallajökull, April–May 2010. *Earth System Science Data*, *3*(1), 9–17. <https://doi.org/10.5194/essd-3-9-2011>
- Balkanski, Y., Schulz, M., Claquin, T., & Guibert, S. (2007). Reevaluation of mineral aerosol radiative forcings suggests a better agreement with satellite and AERONET data. *Atmospheric Chemistry and Physics*, *7*(1), 81–95. <https://doi.org/10.5194/acp-7-81-2007>
- Beckett, F. M., Witham, C. S., Hort, M. C., Stevenson, J. A., Bonadonna, C., & Millington, S. C. (2015). Sensitivity of dispersion model forecasts of volcanic ash clouds to the physical characteristics of the particles. *Journal of Geophysical Research: Atmospheres*, *120*(22), 11636–11652. <https://doi.org/10.1002/2015JD023609>
- Beckett, F. M., Witham, C. S., Leadbetter, S. J., Crocker, R., Webster, H. N., Hort, M. C., et al. (2020). Atmospheric dispersion modelling at the London VAAC: A review of developments since the 2010 Eyjafjallajökull volcano ash cloud. *Atmosphere*, *11*(4), 352. <https://doi.org/10.3390/atmos11040352>
- Bonadonna, C., Biass, S., & Costa, A. (2015). Physical characterization of explosive volcanic eruptions based on tephra deposits: Propagation of uncertainties and sensitivity analysis. *Journal of Volcanology and Geothermal Research*, *296*, 80–100. <https://doi.org/10.1016/j.jvolgeores.2015.03.009>
- Bonadonna, C., Folch, A., Loughlin, S., & Puempel, H. (2012). Future developments in modelling and monitoring of volcanic ash clouds: Outcomes from the first IAVCEI-WMO workshop on ash dispersal forecast and civil aviation. *Bulletin of Volcanology*, *74*, 1–10. <https://doi.org/10.1007/s00445-011-0508-6>
- Bonadonna, C., Genco, R., Gouhier, M., Pistolesi, M., Cioni, R., Alfano, F., et al. (2011). Tephra sedimentation during the 2010 Eyjafjallajökull eruption (Iceland) from deposit, radar, and satellite observations. *Journal of Geophysical Research*, *116*(B12), B12202. <https://doi.org/10.1029/2011JB008462>
- Brown, R., Bonadonna, C., & Durant, A. (2012). A review of volcanic ash aggregation. *Physics and Chemistry of the Earth, Parts A/B/C*, *45–46*, 65–78. <https://doi.org/10.1016/j.pce.2011.11.001>
- Carazzo, G., & Jellinek, A. (2012). A new view of the dynamics, stability and longevity of volcanic clouds. *Earth and Planetary Science Letters*, *325–326*, 39–51. <https://doi.org/10.1016/j.epsl.2012.01.025>
- Cashman, K. V., & Rust, A. (2019). Far-travelled ash in past and future eruptions: Combining tephrochronology with volcanic studies. *Journal of Quaternary Science*, *35*(1–2), 1–12. <https://doi.org/10.1002/jqs.3159>
- Corradini, S., Merucci, L., & Folch, A. (2011). Volcanic ash cloud properties: Comparison between MODIS satellite retrievals and FALL3D transport model. *IEEE Geoscience and Remote Sensing Letters*, *8*(2), 248–252. <https://doi.org/10.1109/LGRS.2010.2064156>
- Corradini, S., Montopoli, M., Guerrieri, L., Ricci, M., Scollo, S., Merucci, L., et al. (2016). A multi-sensor approach for volcanic ash cloud retrieval and eruption characterization: The 23 November 2013 Etna lava fountain. *Remote Sensing*, *8*(1), 58. <https://doi.org/10.3390/rs8010058>
- Corradini, S., Pugnaghi, S., Piscini, A., Guerrieri, L., Merucci, L., Picchiani, M., & Chini, M. (2014). Volcanic ash and SO₂ retrievals using synthetic MODIS TIR data: Comparison between inversion procedures and sensitivity analysis. *Annals of Geophysics*, *57*(2), 1–6. <https://doi.org/10.4401/ag-6616>
- Corradini, S., Spinetti, C., Carboni, E., Tirelli, C., Buongiorno, M. F., Pugnaghi, S., & Gangale, G. (2008). Mt. Etna tropospheric ash retrieval and sensitivity analysis using Moderate Resolution Imaging Spectroradiometer measurements. *Journal of Applied Remote Sensing*, *2*(1), 023550. <https://doi.org/10.1117/1.3046674>
- Crawford, A., Chai, T., Wang, B., Ring, A., Stunder, B., Loughner, C. P., et al. (2022). Evaluation and bias correction of probabilistic volcanic ash forecasts. *Atmospheric Chemistry and Physics*, *22*(21), 13967–13996. <https://doi.org/10.5194/acp-22-13967-2022>
- Crawford, A. M., Stunder, B. J. B., Ngan, F., & Pavolonis, M. J. (2016). Initializing HYSPLIT with satellite observations of volcanic ash: A case study of the 2008 Kasatochi eruption. *Journal of Geophysical Research: Atmospheres*, *121*(18), 10786–10803. <https://doi.org/10.1002/2016jd024779>
- Dacre, H. F., Grant, A. L. M., Hogan, R. J., Belcher, S. E., Thomson, D. J., Devenish, B. J., et al. (2011). Evaluating the structure and magnitude of the ash plume during the initial phase of the Eyjafjallajökull eruption using lidar observations and NAME simulations. *Journal of Geophysical Research*, *116*, D00U03. <https://doi.org/10.1029/2011JD015608>
- Davies, T., Cullen, M., Malcolm, A., Mawson, M., Staniforth, A., White, A., & Wood, N. (2005). A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Quarterly Journal of the Royal Meteorological Society*, *131*(608), 1759–1782. <https://doi.org/10.1256/qj.04.101>
- Degruyter, W., & Bonadonna, C. (2012). Improving on mass flow rate estimates of volcanic eruptions. *Geophysical Research Letters*, *39*(16), L16308. <https://doi.org/10.1029/2012GL052566>
- Del Bello, E., Taddeucci, J., de' Michieli Vitturi, M., Scarlato, P., Andronico, D., Scollo, S., et al. (2017). Effect of particle volume fraction on the settling velocity of volcanic ash particles: Insights from joint experimental and numerical simulations. *Scientific Reports*, *7*(1), 39620. <https://doi.org/10.1038/srep39620>
- Devenish, B. J., Francis, P. N., Johnson, B. T., Sparks, R. S. J., & Thomson, D. J. (2012). Sensitivity analysis of dispersion modeling of volcanic ash from Eyjafjallajökull in May 2010. *Journal of Geophysical Research*, *117*(D20), D00U21. <https://doi.org/10.1029/2011JD016782>
- Dioguardi, F., Beckett, F., Dürrig, T., & Stevenson, J. A. (2020). The impact of eruption source parameter uncertainties on ash dispersion forecasts during explosive volcanic eruptions. *Journal of Geophysical Research: Atmospheres*, *125*(17), e2020JD032717. <https://doi.org/10.1029/2020JD032717>
- Dioguardi, F., Mele, D., & Dellino, P. (2018). A new one-equation model of fluid drag for irregularly shaped particles valid over a wide range of Reynolds number. *Journal of Geophysical Research: Solid Earth*, *123*(1), 144–156. <https://doi.org/10.1002/2017JB014926>

- Durant, A., & Rose, W. (2009). Sedimentological constraints on hydrometeor-enhanced particle deposition: 1992 Eruptions of Crater Peak, Alaska. *Journal of Volcanology and Geothermal Research*, 186(1–2), 40–59. <https://doi.org/10.1016/j.jvolgeores.2009.02.004>
- Durant, A., Shaw, R., Rose, W., Mi, Y., & Ernst, G. (2008). Ice nucleation and overseeding of ice in volcanic clouds. *Journal of Geophysical Research*, 113(D9). <https://doi.org/10.1029/2007JD009064>
- Dürig, T., & Dioguardi, F. (2020). British Geological Survey/REFIR: v20.1 [Software]. Zenodo. <https://doi.org/10.5281/zenodo.13285643>
- Dürig, T., Gudmundsson, M. T., Dioguardi, F., & Schmidt, L. S. (2023). Quantifying the effect of wind on volcanic plumes: Implications for plume modeling. *Journal of Geophysical Research: Atmospheres*, 128(2), e2022JD037781. <https://doi.org/10.1029/2022JD037781>
- Dürig, T., Gudmundsson, M. T., Dioguardi, F., Woodhouse, M., Björnsson, H., Barsotti, S., et al. (2018). REFIR- A multi-parameter system for near real-time estimates of plume-height and mass eruption rate during explosive eruptions. *Journal of Volcanology and Geothermal Research*, 360, 61–83. <https://doi.org/10.1016/j.jvolgeores.2018.07.003>
- Dürig, T., Schmidt, L. S., & Dioguardi, F. (2023). Optimizing mass eruption rate estimates by combining simple plume models. *Frontiers in Earth Science*, 11. <https://doi.org/10.3389/feart.2023.1250686>
- Eresmaa, R., & McNally, A. (2012). Diverse profile datasets from the ECMWF 137-level short-range Forecasts. *NWP SAF Report No. NWP SAF-EC-TR-17*, 12. https://www.nwpsaf.eu/publications/tech_reports/nwpsaf-ec-tr-017.pdf
- Eychenne, J., Cashman, K. V., Rust, A. C., & Durant, A. (2015). Impact of the lateral blast on the spatial pattern and grain size characteristics of the 5 May 1980 Mount St. Helens fallout deposit. *Journal of Geophysical Research: Solid Earth*, 120(9), 6018–6038. <https://doi.org/10.1002/2015JB012116>
- Eychenne, J., Rust, A. C., Cashman, K. V., & Wobrock, W. (2017). Distal enhanced sedimentation from volcanic plumes: Insights from the secondary mass maxima in the 1992 Mount Spurr fallout deposits. *Journal of Geophysical Research: Solid Earth*, 122(10), 7679–7697. <https://doi.org/10.1002/2017JB014412>
- Francis, P. N., Cooke, M. C., & Saunders, R. W. (2012). Retrieval of physical properties of volcanic ash using Meteosat: A case study from the 2010 Eyjafjallajökull eruption. *Journal of Geophysical Research*, 117(D20), D00U09. <https://doi.org/10.1029/2011JD016788>
- Freret-Lorgeril, V., Bonadonna, C., Corradini, S., Donnadieu, F., Guerrieri, L., Lacanna, G., et al. (2021). Examples of multi-sensor determination of eruptive source parameters of explosive events at Mount Etna. *Remote Sensing*, 13(11), 2097. <https://doi.org/10.3390/rs13112097>
- Fu, G., Prata, F., Lin, H. X., Heemink, A., Segers, A., & Lu, S. (2017). Data assimilation for volcanic ash plumes using a satellite observational operator: A case study on the 2010 Eyjafjallajökull volcanic eruption. *Atmospheric Chemistry and Physics*, 17(2), 1187–1205. <https://doi.org/10.5194/acp-17-1187-2017>
- Gouhier, M., Eychenne, J., Azzaoui, N., Guillin, A., Deslandes, M., Poret, M., et al. (2019). Low efficiency of large volcanic eruptions in transporting very fine ash into the atmosphere. *Scientific Reports*, 9(1), 1449. <https://doi.org/10.1038/s41598-019-38595-7>
- Gu, Y., Rose, W. I., Schneider, D. J., Bluth, G. J. S., & Watson, I. M. (2005). Advantageous GOES IR results for ash mapping at high latitudes: Cleveland eruptions 2001. *Geophysical Research Letters*, 32(2), L02305. <https://doi.org/10.1029/2004GL021651>
- Gudmundsson, M., Thordarson, T., Höskuldsson, Á., Larsen, G., Björnsson, H., Prata, F. J., et al. (2012). Ash generation and distribution from the April-May 2010 eruption of Eyjafjallajökull, Iceland. *Scientific Reports*, 2(1), 572. <https://doi.org/10.1038/srep00572>
- ICAO. (2019). Roadmap for international airways volcano watch (IAVW) in support of international air navigation. Retrieved from <https://www.icao.int/airnavigation/METP/MOGVAREferenceDocuments/IAVWRoadmap.pdf> 29 November 2021.
- Johnson, B., Turnbull, K., Brown, P., Burgess, R., Dorsey, J., Baran, A. J., et al. (2012). In situ observations of volcanic ash clouds from the FAAM aircraft during the eruption of Eyjafjallajökull in 2010. *Journal of Geophysical Research*, 117(D20), D00U24. <https://doi.org/10.1029/2011JD016760>
- Jones, A., Thomson, D., Hort, M., & Devenish, B. (2007). The U.K. Met Office's next generation atmospheric dispersion model, NAME III, in air pollution modeling and its application XV11. In C. Borrego & A.-L. Norman (Eds.), *Proceedings of the D00U08 WEBSTER ET AL.: PREDICTING EYJAFJALLAJÖKULL ASH LEVELS 27th NATO/CCMS international technical meeting on air pollution modeling and its application* (pp. 580–589). Springer. https://doi.org/10.1007/978-0-387-68854-1_62
- Kylling, A., Kahnert, M., Lindqvist, H., & Nousiainen, T. (2014). Volcanic ash infrared signature: Porous non-spherical ash particle shapes compared to homogeneous spherical ash particles. *Atmospheric Measurement Techniques*, 7(4), 919–929. <https://doi.org/10.5194/amt-7-919-2014>
- Kylling, A., Kristiansen, N., Stohl, A., Buras-Schnell, R., Emde, C., & Gasteiger, J. (2015). A model sensitivity study of the impact of clouds on satellite detection and retrieval of volcanic ash. *Atmospheric Measurement Techniques*, 8(5), 1935–1949. <https://doi.org/10.5194/amt-8-1935-2015>
- Manzella, I., Bonadonna, C., Phillips, J. C., & Monnard, H. (2015). The role of gravitational instabilities in deposition of volcanic ash. *Geology*, 43(3), 211–214. <https://doi.org/10.1130/G36252.1>
- Marenco, F., Johnson, B., Turnbull, K., Newman, S., Haywood, J., Webster, H., & Ricketts, H. (2011). Airborne lidar observations of the 2010 Eyjafjallajökull volcanic ash plume. *Journal of Geophysical Research*, 116, D00U05. <https://doi.org/10.1029/2011JD016396>
- Marzano, F. S., Corradini, S., Mereu, L., Kylling, A., Montopoli, M., Cimini, D., et al. (2018). Multi-satellite multi-sensor observations of a sub-Plinian volcanic eruption: The 2015 Calbuco explosive event in Chile. *IEEE Transactions on Geoscience and Remote Sensing*, 56(5), 2597–2612. <https://doi.org/10.1109/TGRS.2017.2769003>
- Marzano, F. S., Lamantea, M., Montopoli, M., Herzog, M., Graf, H., & Cimini, D. (2013). Microwave remote sensing of the 2011 Plinian eruption of the Grímsvötn Icelandic volcano. *Remote Sensing of Environment*, 129, 168–184. <https://doi.org/10.1016/j.rse.2012.11.005>
- Mastin, L., Pavolonis, M., Engwell, S., Clarkson, R., Witham, C., Brock, G., et al. (2022). Progress in protecting air travel from volcanic ash clouds. *Bulletin of Volcanology*, 84(1), 9. <https://doi.org/10.1007/s00445-021-01511-x>
- Mastin, L. G., Guffanti, M., Servranckx, R., Webley, P., Barsotti, S., Dean, K., et al. (2009). A multidisciplinary effort to assign realistic source parameters to models of volcanic ash-cloud transport and dispersion during eruptions. *Journal of Volcanology and Geothermal Research*, 186(1–2), 10–21. <https://doi.org/10.1016/j.jvolgeores.2009.01.008>
- Matricardi, M. (2005). The inclusion of aerosols and clouds in RTIASI, the ECMWF fast radiative transfer model for the infrared atmospheric sounding interferometer. *ECMWF Tech. Memo.*, 474.
- Millington, S. C., Saunders, R. W., Francis, P. N., & Webster, H. N. (2012). Simulated volcanic ash imagery: A method to compare NAME ash concentration forecasts with SEVIRI imagery for the Eyjafjallajökull eruption in 2010. *Journal of Geophysical Research*, 117(D20), D00U17. <https://doi.org/10.1029/2011JD016770>
- Mingari, L., Folch, A., Prata, A. T., Pardini, F., Macedonio, G., & Costa, A. (2022). Data assimilation of volcanic aerosol observations using FALL3D+PDAF. *Atmospheric Chemistry and Physics*, 22(3), 1773–1792. <https://doi.org/10.5194/acp-22-1773-2022>
- Newman, S. M., Clarisse, L., Hurtmans, D., Marenco, F., Johnson, B., Turnbull, K., et al. (2012). A case study of observations of volcanic ash from the Eyjafjallajökull eruption: 2. Airborne and satellite radiative measurements. *Journal of Geophysical Research*, 117(D20), D00U13. <https://doi.org/10.1029/2011JD016780>

- Pardini, F., Corradini, S., Costa, A., Esposti Ongaro, T., Merucci, L., Neri, A., et al. (2020). Ensemble-based data assimilation of volcanic ash clouds from satellite observations: Application to the 24 December 2018 Mt. Etna explosive eruption. *Atmosphere*, *11*(4), 359. <https://doi.org/10.3390/atmos11040359>
- Pavlonis, M. J., Sieglaff, J., & Cintineo, J. (2015a). Spectrally enhanced cloud objects - a generalized framework for automated detection of volcanic ash and dust clouds using passive satellite measurements: 2. Cloud object analysis and global application. *Journal of Geophysical Research: Atmospheres*, *120*(15), 7842–7870. <https://doi.org/10.1002/2014JD022969>
- Pavlonis, M. J., Sieglaff, J., & Cintineo, J. (2015b). Spectrally enhanced cloud objects-A generalized framework for automated detection of volcanic ash and dust clouds using passive satellite measurements: 1. Multispectral analysis. *Journal of Geophysical Research: Atmospheres*, *120*(15), 7813–7841. <https://doi.org/10.1002/2014JD022968>
- Pavlonis, M. J., Sieglaff, J., & Cintineo, J. (2018). Automated detection of explosive volcanic eruptions using satellite-derived cloud vertical growth rates. *Earth and Space Science*, *5*(12), 903–928. <https://doi.org/10.1029/2018EA000410>
- Pelley, R. E., Thomson, D. J., Webster, H. N., Cooke, M. C., Manning, A. J., Witham, C. S., & Hort, M. C. (2021). A near-real-time method for estimating volcanic ash emissions using satellite retrievals. *Atmosphere*, *12*, 1573. <https://doi.org/10.3390/atmos12121573>
- Piontek, D., Bugliaro, L., Kar, J., Schumann, U., Marengo, F., Plu, M., & Voigt, C. (2021). The new volcanic ash satellite retrieval VACOS using MSG/SEVIRI and artificial neural networks: 2. Validation. *Remote Sensing*, *13*(16), 3128. <https://doi.org/10.3390/rs13163128>
- Pollack, J. B., Toon, O. B., & Khare, B. N. (1973). Optical properties of some terrestrial rocks and glasses. *Icarus*, *19*(3), 372–389. [https://doi.org/10.1016/0019-1035\(73\)90115-2](https://doi.org/10.1016/0019-1035(73)90115-2)
- Poulidis, A., Takemi, T., Iguchi, M., & Renfrew, I. (2017). Orographic effects on the transport and deposition of volcanic ash: A case study of Mt. Sakurajima, Japan. *Journal of Geophysical Research: Atmospheres*, *122*(17), 9332–9350. <https://doi.org/10.1002/2017JD026595>
- Prata, A. J. (1989a). Infrared radiative transfer calculations for volcanic ash clouds. *Geophysical Research Letters*, *16*(11), 1293–1296. <https://doi.org/10.1029/GL016i011p01293>
- Prata, A. J. (1989b). Observations of volcanic ash clouds in the 10–12 μm window using AVHRR/2 data. *International Journal of Remote Sensing*, *10*(4–5), 751–761. <https://doi.org/10.1080/01431168908903916>
- Prata, A. J., & Grant, I. F. (2001). Retrieval of microphysical and morphological properties of volcanic ash plumes from satellite data: Application to Mt Ruapehu, New Zealand. *Quarterly Journal of the Royal Meteorological Society*, *127*(576), 2153–2179. <https://doi.org/10.1002/qj.49712757615>
- Prata, A. J., & Prata, A. T. (2012). Eyjafjallajökull volcanic ash concentrations determined using Spin Enhanced Visible and Infrared Imager measurements. *Journal of Geophysical Research*, *117*(D20), D00U23. <https://doi.org/10.1029/2011JD016800>
- Prata, A. T., Grainger, R. G., Taylor, I. A., Povey, A. C., Proud, S. R., & Poulsen, C. A. (2022). Uncertainty-bounded estimates of ash cloud properties using the ORAC algorithm: Application to the 2019 Raikoke eruption. *Atmospheric Measurement Techniques*, *15*(20), 5985–6010. <https://doi.org/10.5194/amt-15-5985-2022>
- Prata, A. T., Mingari, L., Folch, A., Macedonio, G., & Costa, A. (2021). FALL3D-8.0: A computational model for atmospheric transport and deposition of particles, aerosols and radionuclides – Part 2: Model validation. *Geoscientific Model Development*, *14*(1), 409–436. <https://doi.org/10.5194/gmd-14-409-2021>
- Rodgers, C. D. (2000). *Inverse methods for atmospheres: Theory and practice*. River edge, World Scientific.
- Romeo, F., Mereu, L., Scollo, S., Papa, M., Corradini, S., Merucci, L., & Marzano, F. S. (2023). Volcanic cloud detection and retrieval using satellite multisensor observations. *Remote Sensing*, *15*(4), 888. <https://doi.org/10.3390/rs15040888>
- Rose, W. I., Bluth, G. J. S., & Ernst, G. G. J. (2000). Integrating retrievals of volcanic cloud characteristics from satellite remote sensors: A summary. *Philosophical Transactions of the Royal Society A*, *358*(1770), 1585–1606. <https://doi.org/10.1098/rsta.2000.0605>
- Rossi, E., Bagheri, G., Beckett, F., & Bonadonna, C. (2021). The fate of volcanic ash aggregates: Premature or delayed sedimentation? *Nature Communications*, *12*(1), 1303. <https://doi.org/10.1038/s41467-021-21568-8>
- Saint, C. (2023). *Improved satellite detection of volcanic ash* (Vol. 656). Forecasting Research Technical Report; Met Office: Exeter, UK, 2023. Retrieved from https://digital.nmla.metoffice.gov.uk/IO_12dd8ba8-aca0-4fc6-baa1-5127edb46916/
- Saint, C., Beckett, F., & Dioguardi, F. (2024). Data for “Using simulated radiances to understand the limitations of satellite-retrieved volcanic ash data and the implications for volcanic ash cloud forecasting” [Dataset]. *Zenodo*. <https://doi.org/10.5281/zenodo.10562538>
- Saunders, R., Hocking, J., Turner, E., Havemann, S., Geer, A., Lupu, C., et al. (2020). RTTOV-13 science and validation report. *EUMETSAT NWP SAF. Technical report, NWPSAF-MO-TV-46*. https://nwp-saf.eumetsat.int/site/download/documentation/rtm/docs_rttov13/rttov13_svr.pdf
- Saxby, J., Beckett, F., Cashman, K., Rust, A., & Tennant, E. (2018). The impact of particle shape on fall velocity: Implications for volcanic ash dispersion modelling. *Journal of Volcanology and Geothermal Research*, *362*, 32–48. <https://doi.org/10.1016/j.jvolgeores.2018.08.006>
- Saxby, J., Rust, A., Cashman, K., & Beckett, F. (2019). The importance of grain size and shape in controlling the dispersion of the Vedde cryptotephra. *Journal of Quaternary Science*, *35*(1–2), 175–185. <https://doi.org/10.1002/jqs.3152>
- Schumann, U., Weinzierl, B., Reitebuch, O., Schlager, H., Minikin, A., Forster, C., et al. (2011). Airborne observations of the Eyjafjalla volcano ash cloud over Europe during air space closure in April and May 2010. *Atmospheric Chemistry and Physics*, *11*(5), 2245–2279. <https://doi.org/10.5194/acp-11-2245-2011>
- Sparks, R. S. J., Bursik, M. I., Carey, S. N., Gilbert, J. S., Glaze, L., Sigurdsson, H., & Woods, A. W. (1997). *Volcanic plumes*. John Wiley & Sons, Inc.
- Stevenson, J. A., Millington, S. C., Beckett, F. M., Swindles, G. T., & Thordarson, T. (2015). Big grains go far: Understanding the discrepancy between tephrochronology and satellite infrared measurements of volcanic ash. *Atmospheric Measurement Techniques*, *8*(5), 2069–2091. <https://doi.org/10.5194/amt-8-2069-2015>
- Stohl, A., Prata, A. J., Eckhardt, S., Clarisse, L., Durant, A., Henne, S., et al. (2011). Determination of time- and height-resolved volcanic ash emissions and their use for quantitative ash dispersion modeling: The 2010 Eyjafjallajökull eruption. *Atmospheric Chemistry and Physics*, *11*(9), 4333–4351. <https://doi.org/10.5194/acp-11-4333-2011>
- Turnbull, K., Johnson, B., Marengo, F., Haywood, J., Minikin, A., Weinzierl, B., et al. (2012). A case study of observations of volcanic ash from the Eyjafjallajökull eruption: 1. In situ airborne observations. *Journal of Geophysical Research*, *117*(D20), D00U12. <https://doi.org/10.1029/2011JD016688>
- Volz, F. E. (1973). Infrared optical constants of ammonium sulfate, Sahara dust, volcanic pumice, and flyash. *Applied Optics*, *12*(3), 564–568. <https://doi.org/10.1364/AO.12.000564>
- Watson, E. J., Swindles, G. T., Stevenson, J., Savov, I., & Lawson, I. (2016). The transport of Icelandic volcanic ash: Insights from northern European cryptotephra records. *Journal of Geophysical Research: Solid Earth*, *121*(10), 7177–7192. <https://doi.org/10.1002/2016JB013350>
- Watt, S., Gilbert, J., Folch, A., Phillips, J., & Cai, X. (2015). An example of enhanced tephra deposition driven by topographically induced atmospheric turbulence. *Bulletin of Volcanology*, *77*(5), 35. <https://doi.org/10.1007/s00445-015-0927-x>

- Webster, H. N., & Thomson, D. J. (2011). Dry deposition modelling in a Lagrangian dispersion model. *International Journal of Environment and Pollution*, 47(1/2/3/4), 1–9. <https://doi.org/10.1504/IJEP.2011.047322>
- Webster, H. N., Thomson, D. J., Johnson, B. J., Heard, I. P. C., Turnbull, K., Marengo, F., et al. (2012). Operational prediction of ash concentrations in the distal volcanic cloud from the 2010 Eyjafjallajökull eruption. *Journal of Geophysical Research*, 117(D20), D00U08. <https://doi.org/10.1029/2011JD016790>
- Webster, H. N., Whitehead, T., & Thomson, D. J. (2018). Small-scale three-dimensional atmospheric turbulence and unresolved mesoscale motions are parameterized using random-walk techniques. *Journal of Applied Meteorology and Climatology*, 57(3), 645–657. <https://doi.org/10.1175/JAMC-D-17-0075.1>
- Wen, S., & Rose, W. I. (1994). Retrieval of sizes and total masses of particles in volcanic clouds using AVHRR bands 4 and 5. *Journal of Geophysical Research*, 99(D3), 5421–5431. <https://doi.org/10.1029/93JD03340>
- Western, L. M., Watson, M. I., & Francis, P. N. (2015). Uncertainty in two-channel infrared remote sensing retrievals of a well-characterised volcanic ash cloud. *Bulletin of Volcanology*, 77(8), 67. <https://doi.org/10.1007/s00445-015-0950-y>
- Wilkins, K. L., Watson, I. M., Kristiansen, N. I., Webster, H. N., Thomson, D. J., Dacre, H. F., & Prata, A. J. (2016). Using data insertion with the NAME model to simulate the 8 May 2010 Eyjafjallajökull volcanic ash cloud. *Journal of Geophysical Research: Atmospheres*, 121(1), 306–323. <https://doi.org/10.1002/2015JD023895>
- Wilson, L., & Walker, G. P. L. (1987). Explosive volcanic eruptions - VI. Ejecta dispersal in Plinian eruptions: The control of eruption conditions and atmospheric properties. *Geophysical Journal International*, 89(2), 657–679. <https://doi.org/10.1111/j.1365-246X.1987.tb05186.x>
- Winker, D. M., Liu, Z., Omar, A., Tackett, J., & Fairlie, D. (2012). CALIOP observations of the transport of ash from the Eyjafjallajökull volcano in April 2010. *Journal of Geophysical Research*, 117(D20), D00U15. <https://doi.org/10.1029/2011JD016499>
- Witham, C. S., Hort, M. C., Potts, R., Servranckx, R., Husson, P., & Bonnardot, F. (2007). Comparison of VAAC atmospheric dispersion models using the 1 November 2004 Grimsvotn eruption. *Meteorological Applications*, 14(1), 27–38. <https://doi.org/10.1002/met.3>
- Woodhouse, M. J., Hogg, A. J., Phillips, J. C., & Sparks, R. S. J. (2013). Interaction between volcanic plumes and wind during the 2010 Eyjafjallajökull eruption, Iceland. *Journal of Geophysical Research: Solid Earth*, 118(1), 92–109. <https://doi.org/10.1029/2012JB009592>