AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2016JD024779

Initializing HYSPLIT with satellite observations of volcanic ash: A case study of the 2008 Kasatochi eruption

Alice M. Crawford[1,2], Barbara J. B. Stunder[2], Fong Ngan[1,2], and Michael J. Pavolonis[3]

[1]Cooperative Institute for Climate and Satellites, University of Maryland, College Park, Maryland, USA, [2]Air Resources Laboratory, NOAA, College Park, Maryland, USA, [3]Center for Satellite Applications and Research, NOAA, Madison, Wisconsin, USA

Key Points:
- An effective way of initializing HYSPLIT with satellite retrievals of ash cloud properties is validated
- Model output, MODIS retrievals of cloud top height, and CALIOP measurements show good agreement
- Measurements and model output all suggest that the ash cloud develops into a complex 3-D structure

Supporting Information:
- Supporting Information S1
- Animation S1
- Animation S2
- Animation S3
- Animation S4
- Animation S5
- Model Configuration S1
- Model Configuration S2
- Model Configuration S3
- Model Configuration S4
- Model Configuration S5
- Model Configuration S6
- Model Configuration S7
- Model Configuration S8

Correspondence to:
A. M. Crawford,
alice.crawford@noaa.gov

Citation:
Crawford, A. M., B. J. B. Stunder, F. Ngan, and M. J. Pavolonis (2016), Initializing HYSPLIT with satellite observations of volcanic ash: A case study of the 2008 Kasatochi eruption, *J. Geophys. Res. Atmos.*, *121*, 10,786–10,803, doi:10.1002/2016JD024779.

Received 11 JAN 2016
Accepted 26 AUG 2016
Accepted article online 1 SEP 2016
Published online 20 SEP 2016

Abstract The current work focuses on improving volcanic ash forecasts by integrating satellite observations of ash into the Lagrangian transport and dispersion model, HYSPLIT. The accuracy of HYSPLIT output is dependent on the accuracy of the initialization: the initial position, size distribution, and amount of ash as a function of time. Satellite observations from passive infrared, IR, sensors are used both to construct the initialization term and for verification. Space-based lidar observations are used for further verification. We compare model output produced using different initializations for the 2008 eruption of Kasatochi in the Aleutian Islands. Simple source terms, such as a uniform vertical line or cylindrical source above the vent, are compared to initializations derived from satellite measurements of position, mass loading, effective radius, and height of the downwind ash cloud. Using satellite measurements of column mass loading of ash to constrain the source term produces better long-term predictions than using an empirical equation relating mass eruption rate and plume height above the vent. Even though some quantities, such as the cloud thickness, must be estimated, initializations which release particles at the position of the observed ash cloud produce model output which is comparable to or better than the model output produced with source terms located above and around the vent. Space-based lidar data, passive IR retrievals of ash cloud top height, and model output agree well with each other, and all suggest that the Kasatochi ash cloud evolved into a complex three-dimensional structure.

1. Introduction

Transport and dispersion models are used extensively to help produce forecasts of the location of ash produced by a medium to large volcanic eruption. These forecasts are used by the aviation industry to help ensure safety while minimizing impacts to travelers and the industry. The inputs required by these models are data from a meteorological model and an initialization: the initial position, size, and amount of ash all as functions of time [*Heffter et al.*, 1993; *Leadbetter and Hort*, 2011; *Devenish et al.*, 2012; *Folch*, 2012].

Because many volcanoes are located in remote areas, forecasters rely heavily on space-based instruments aboard satellites to observe eruptions and the ash clouds they produce. Available observations can then be used to construct an initialization for input into a model [*Mastin et al.*, 2009]. Relatively recently, improved algorithms have been developed which use data from passive imaging sensors on satellites to identify the spectral signature of ash and then retrieve properties of the ash cloud, such as mass loading, top height, and effective radius [*Corradini et al.*, 2010; *Prata and Prata*, 2012; *Pavolonis et al.*, 2013, 2015a, 2015b].

Several investigations have focused on using satellite retrievals of ash cloud properties to produce a source term using an inversion algorithm which estimates emissions as a function of time and height above the vent [*Kristiansen et al.*, 2010; *Stohl et al.*, 2011; *Kristiansen et al.*, 2012; *Moxnes et al.*, 2014]. This approach is promising; however, it can take some time to produce a forecast. Since our main interest is near-real-time forecasting, we investigate simple initialization methods in which the model propagates ash forward from an observed ash cloud location which is downwind from the vent.

Vernier et al. [2013] and *Prata et al.* [2015] have looked at initializing trajectory models with information from space-based lidar. These studies looked at the 2011 eruption of Cordon-Caulle and the 2008 eruption of Chaiten, respectively. Both of these eruptions produced large, long-lived ash clouds. The accuracy and

resolution of meteorological model winds affect the accuracy of transport and dispersion model forecasts. How big an effect they have depends on the meteorological conditions and the length of the simulation. One advantage to propagating ash forward from an observed location downwind of the vent is that a shorter simulation is needed.

Analysts at the Tokyo Volcanic Ash Advisory Center (VAAC) have the capability to initialize their model from an observed location rather than the vent if they observe that the forecast cloud position or extent departs from the observed [*Elíasson and Yoshitani*, 2015]. They find that this helps reduce positional errors.

We investigate constructing a model initialization in which the initial positions of model particles are determined from passive IR retrievals of latitude, longitude location of ash, and ash cloud top height. Further, the size of the particles and the amount of mass are determined from the passive IR retrievals of effective radius and mass loading. The 2008 eruption of Kasatochi in the Aleutian Islands is an excellent test case for this type of initialization. The ash cloud produced was large and long-lived. Passive satellite sensors capable of taking high spatial resolution (1 km) measurements in the IR were able to observe the ash cloud at several times, mostly without interference from meteorological clouds. The first IR observation occurred near the end of the last explosive event [*Waythomas et al.*, 2010] and thus captured most of the emitted ash. A space-based lidar instrument also made several passes over the ash cloud providing some information on the vertical structure. Thus, this paper deals with how such an initialization may be constructed and how it performs under rather ideal conditions. How this type of initialization may be used in less ideal conditions is discussed in section 6 along with a discussion of *Wilkins et al.* [2014, 2016] who used similar downwind model initializations with a different model, passive IR sensor, retrieval algorithm, and eruption.

Satellite observations may also be used to verify model output and thus help drive model development. We employ commonly used statistical measures to compare model output with the passive IR satellite retrievals and draw conclusions about how the model initialization affects model performance.

We also compare space-based lidar time-height plots, model output of computational particle positions, and passive IR satellite retrievals of ash cloud top height. We use these comparisons to reinforce our conclusions about the efficacy of different initializations and to make some observations about the complex spatial structure of the ash cloud.

2. Satellite Observations

Satellite observations from two instruments are used.

1. Data from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument aboard the Aqua satellite using the algorithm described in *Pavolonis et al.* [2013, 2015a, 2015b] were utilized to create initializations for HYSPLIT and to verify HYSPLIT model output. The algorithm can be broken into two parts. The first part identifies pixels which contain ash. The second part calculates the mass loading, effective radius, and radiative cloud height for each pixel at which ash was detected. The algorithm has been applied to other instruments, such as imagers aboard geostationary satellites (see *Pavolonis et al.* [2015a], for a list of relevant sensors). Under ideal conditions, the lowest value of mass loading detectable by the passive IR satellite retrieval is approximately 0.1 g m^{-2} [*Prata and Prata*, 2012]. Hereafter we will refer to these data as a passive IR retrieval as the algorithm utilizes measurements of upwelling infrared radiation made by a satellite-based instrument. We use passive IR retrievals which occurred at five times as summarized in Table 1.
2. Data from the CALIOP (Cloud-Aerosol Lidar with Orthogonal Polarization) instrument aboard CALIPSO were utilized to verify HYSPLIT model output but were not used to create initializations. *Winker et al.* [2012] and *Vernier et al.* [2013] discuss various ways to distinguish volcanic ash from other features in the lidar data, such as dust or meteorological clouds. We located four overpasses which intersected the observed or modeled ash cloud. The times of intersection are summarized in Table 1. The location of features in time-height plots of the 532 nm backscatter were compared with the location of passive IR retrievals of ash and/or the location of HYSPLIT particles. Images of the attenuated color ratio (ratio of the 1064 nm to 532 nm backscatter) were then inspected [*Winker et al.*, 2009, 2010]. The color ratio is usually lower (near 0.5) for volcanic ash [*Vernier et al.*, 2013] (see supporting information for color ratio plots).

Figures 1a, 1c, and 1e show passive IR satellite retrievals of top height for the three Aqua MODIS overpasses for which there was a corresponding CALIOP overpass through the ash cloud. Figures 1b, 1d, and 1f show plots of the CALIOP 532 nm backscatter as well as passive IR retrievals of ash cloud top height along the path of the

Table 1. Top Shows at Which Times the Satellite Retrievals From the MODIS Instrument Were Made (to the Nearest Tens of Minutes) and at Which Times the CALIOP Instrument Made an Overpass of the Ash Cloud[a]

Times at Which Observations Are Available			
Abbreviation	Date HH:MM UTC	MODIS	CALIOP Lidar
T1	8 Aug 2008 13:40	x	x
T2	9 Aug 2008 00:50	x	
T3	Aug 2008 12:50	x	
T4	10 Aug 2008 00:00	x	x
T5	10 Aug 2008 11:50	x	x
T6	10 Aug 2008 13:30		x
Model Initializations			
Abbreviation	Initial Position	Initial Mass	
Observed Initializations			
RT1	MODIS passive IR retrieval at T1	T1 mass retrieval	
RT2	MODIS passive IR retrieval at T2	T2 mass retrieval	
RT3	MODIS passive IR retrieval at T3	T3 mass retrieval	
Uniform Line Source			
LINE A	uniform line above vent	$MER_{fine} = 2.8 \times 10^4 kg\,s^{-1}$	
Cylindrical Sources			
CYL A	cylinder d = 100 km at vent[b]	$MER_{fine} = 2.8 \times 10^4 kg\,s^{-1}$	
CYL B	cylinder d = 100 km at vent	$MER_{fine} = 2.8 \times 10^3 kg\,s^{-1}$	
CYL C	cylinder d = 100 km at vent	$MER_{fine} = 2.8 \times 10^5 kg\,s^{-1}$	
CYL D	cylinder d = 100 km at vent	$MER_{fine} = 2.8 \times 10^6 kg\,s^{-1}$	

[a]Bottom summarizes the initializations used as input into the HYSPLIT model. The abbreviations given in the first column are used throughout the text.
[b]d is the diameter of the cylindrical source.

lidar track. Lidar features which were identified as possible ash are labeled with capital letters. Comparison of the two data sets (Figures 1b, 1d, and 1f) shows that there is good agreement between the passive IR satellite retrievals and the CALIOP data. Passive IR satellite height retrievals tended to be a few kilometers higher than the top aerosol level indicated by CALIOP, but within the expected error of the retrieval.

In Figure 1b, features A and B coincide with the location of the passive IR satellite retrieval of ash. Feature A also has a low color ratio. Feature C coincides with meteorological ice clouds shown in MODIS multispectral imagery and is connected to feature A. It may be a mixture of meteorological cloud and ash.

In Figure 1d, CALIOP indicates two thin levels of ash (features A and B), and the retrieved top height lies between these two levels. For layers of ash or a geometrically thick ash cloud, the retrieved top height may lie somewhere in the middle of the layers or ash cloud rather than at the top [*Pavolonis et al.*, 2013]. Feature C lies at a low altitude and may be part volcanic and part meteorological in nature. *Kristiansen et al.* [2010] identify this feature as volcanic in origin.

In Figure 1f, feature A has a low color ratio and partially coincides with the passive IR satellite retrieval of ash. The rest of the feature surrounding A may have some ash mixed with meteorological cloud.

3. Modeling

In this section, we describe the Lagrangian transport and dispersion model, HYSPLIT, the three meteorological data sets used as input into HYSPLIT, and the different ways to initialize the model.

3.1. HYSPLIT

HYSPLIT is a Lagrangian transport and dispersion model developed by the National Oceanic and Atmospheric Administration's Air Resources Laboratory (NOAA ARL). The model is used operationally at the Washington,

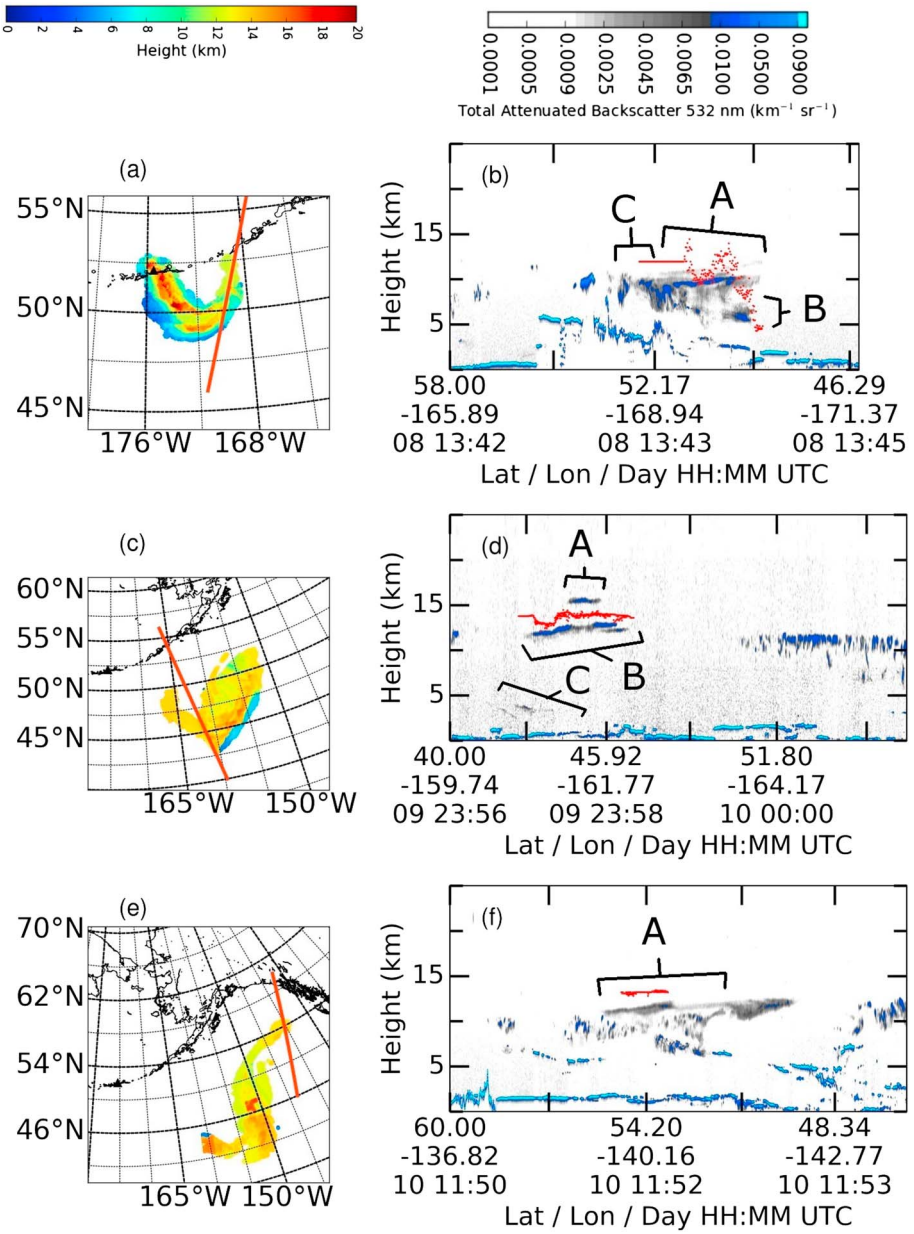

Figure 1. Satellite-retrieved location and top height of ash cloud (a) T1, (c)T4, and (e) T5. Top height is shown in the height color bar. In Figures 1a, 1c, and 1e the orange line shows the CALIOP track. (b, d, and f) The red points show the satellite-retrieved top heights at points that were within 0.1° of the CALIOP track. Other colors show the total attenuated 532 nm backscatter as measured by CALIOP and values are described by the color bar. Features in the 532 nm backscatter which were identified as possible ash are labeled with capital letters.

Anchorage, Darwin, and Wellington Volcanic Ash Advisory Centers (VAACs) for modeling the transport and dispersion of volcanic ash [*Dare*, 2015; *Stunder et al.*, 2007]. It is also used to provide forecasts of smoke from wild fires [*Rolph et al.*, 2009; *Stein et al.*, 2009], forecasts of windblown dust [*Draxler et al.*, 2010], and atmospheric dispersion products for chemical and nuclear accidents [*Draxler and Rolph*, 2012]. The model details and history are described in *Draxler and Hess* [1997, 1998] and *Stein et al.* [2015]. Here the model is used in particle (rather than puff) mode. The model is configured to calculate wet and dry deposition. Column mass loadings are calculated over a 0.05° latitude by 0.1° longitude grid. At the latitudes of interest, this corresponds to ~6 km × 6 km grid.

3.2. Meteorological Data

For this study, three meteorological data sets were used as input into HYSPLIT.

1. The U.S. National Centers for Environmental Prediction (NCEP) Global Data Assimilation System (GDAS) data set with 1° latitude longitude resolution and output every 3 h. The output is every 25 hPa from 1000 to 900 hPa and 50 hPa from 900 to 50 hPa and then output at 20 hPa.
2. The European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim global atmospheric reanalysis [*Dee et al.*, 2011]. The data set used here has approximately 0.75° latitude longitude resolution and analyses every 6 h. The output is every 25 hPa from 1000 to 750 hPa, every 50 hPa from 750 to 250 hPa, and every 25 hPa from 250 to 100 hPa and then levels at 70, 50, 30, 20, 10, 7, 5, 3, 2, and 1 hPa.
3. Advanced Research dynamic core of the Weather Research Forecasting model (WRF). [*Skamarock et al.*, 2008], version 3.5.1 with 27 km horizontal grid spacing (approximately 0.25°) and output every hour. The vertical grid consisted of 52 layers with the model top at 10 hPa. More information can be found in the supporting information.

The tropopause was present at around 11 km. Ash from the eruption was quickly swept into a passing low-pressure system and transported east, over ocean, toward the west coast of Canada.

3.3. Model Initialization

Three types of model initializations were considered. The uniform line source and cylindrical source release particles above and around the vent, while the observed initialization releases particles at locations where ash is observed by satellite. These sources are described in detail below and summarized in Table 1. Some of the HYSPLIT input files are provided in the supporting information.

3.3.1. Uniform Line Source

For this type of source, ash is initialized in a uniform vertical line from the top of the vent, 300 m above sea level, to the observed plume height. Ash is released over the following three periods from the vent top to the given height: 7 August 2008 22:00 to 23:00 UTC to 14 km, 8 August 2008 02:00 to 03:00 UTC to 14 km, and 8 August 2008 04:00 to 12:00 UTC to 18 km. These time periods and plume heights approximately correspond to the three eruption events described in *Waythomas et al.* [2010]. Four particle sizes were used. Each particle size is represented by the same number of computational particles but a different amount of mass (10 μm radius: 67% of mass, 3 μm radius: 25.4% of mass, 1 μm radius: 6.8% of mass, and 0.3 μm radius: 0.8% of mass). These particle sizes are used as the default when HYSPLIT is run operationally by the Washington VAAC.

3.3.2. Cylindrical Source

The cylindrical source is the same as the uniform line source except particles are released in a cylindrical volume, 100 km in diameter, centered along the previous line source. According to *Waythomas et al.* [2010], the continuous plume width was observed in GOES thermal imagery to be about 75 km on 8 August 2008 from about 06:00 to 10:00 UTC. For large eruptions, the initial spreading of the buoyant plume can transport ash particles hundreds of kilometers from the vent regardless of the direction of the model winds [*Costa et al.*, 2013; *Mastin et al.*, 2014; *Johnson et al.*, 2015]. A simple way to take this into account is to initialize ash in an observed area around, rather than a line above, the vent.

For the line and cylindrical source, the HYSPLIT model was configured to emit one mass unit per hour, and mass loading outputs are in unit mass m^{-2}. Since satellite mass loading is given in g m^{-2}, we then convert the unit mass to grams which is equivalent to choosing a mass eruption rate of fine ash, MER_{fine}. We specify fine ash since mass eruption rate, MER, generally refers to the total mass erupted. This includes many larger particles (greater than a few tens of microns) which fall out fairly close to the source and are not of concern for this application. The mass eruption rates considered are shown in Table 1. The cylindrical source is denoted as CYL throughout the text and letters A, B, C, and D are used to denote the MER_{fine} as described in Table 1.

Deciding on an MER_{fine} and a threshold of detectable ash can significantly impact the modeled area of detectable ash. Although the term detectable ash is somewhat vague, we will use it here to mean that the model output should provide information on what the satellite is expected to detect. This process is subject to both uncertainties in the mass of the ash cloud and in the minimum mass loading which can be detected by the satellite at each point in space and time. We pick a range of mass eruption rates and explore how sensitive the results are to this choice. One common way of determining a MER_{fine} when a measurement of the plume height is available is to use an empirical equation relating MER and plume height and then estimate the fraction of fine ash. Using the relationship in *Mastin et al.* [2009], a plume height of 14–18 km, magma density of 2500 kg m^{-3} and 0.10 fraction of fine ash, the MER_{fine} ranges from 8.0×10^5 to 2.3×10^6 kg s^{-1} which is close to D in Table 1.

There is a large uncertainty in the mass eruption rate calculated with the empirical relationship. Figure 1 of *Mastin et al.* [2009] shows that for plumes with heights between about 15 and 20 km, MER's between about 1×10^6 and 1×10^8kgs^{-1} have been measured. The mass fraction of ash less than 63 μm has been found to vary between about 0.01 and 0.5 [*Mastin et al.*, 2009]. Using those ranges, we could expect to see a MER_{fine} of anywhere between 1×10^4 and 5×10^7 kg s^{-1}. The model produced mass loading values over about 5 orders of magnitude, and the values of MER_{fine} used here were chosen so a threshold of 0.1 g m^{-2} would span this range.

In this work, the rate of mass emission is kept constant with time. The MER, as well as the mass fraction of fine ash, is expected to vary with time for any eruption. However, since it is difficult to determine how it should vary, this simplifying assumption is often used.

3.3.3. Observed Initialization

The observed initializations use the passive IR satellite retrieval of latitude, longitude, top height, effective radius, and mass loading as input into the model. The satellite retrieval was regridded onto a regular grid which matched the HYSPLIT grid for calculating mass loading (0.05° latitude by 0.1° longitude). The total mass to be released from each grid box was calculated by multiplying the retrieved mass loading by the area of the grid box. Seven particle sizes were used, and these were determined for each grid box by creating a histogram of the retrieved effective radius with seven equally spaced bins. The particle size histograms for each observed initialization are shown in the supporting information. Particle radii ranged between 1.4 and 11.5 μm.

Since no information about the thickness of the ash cloud is retrieved, several options could be considered for the initial altitude range of the initialization. Here we consider a simple scheme in which particles were initialized from the retrieved top height to a point 2.5 km below the top height.

We pick this scheme since the CALIOP observations indicated that the retrievals mostly overestimated the height slightly, and the ash layers were not thick. We looked at different estimations of thickness for a few of the cases. This is discussed in Appendix A.

Figure 2 (column 1) shows passive IR retrievals of mass loading at times T1 through T5. Retrievals at T1, T2, and T3 were used to create model initializations RT1, RT2, and RT3 (see Table 1 and Figure 2). The model initializations are used to create HYSPLIT outputs at subsequent times.

Contour plots of mass loading for model outputs are shown in Figure 2 (columns 2–5). Figure 2 illustrates the relationship between the initializations and outputs. In Figure 2 (row 1, column1), the passive IR retrieval used to create the RT1 initialization is shown. Plotted to the right are the HYSPLIT outputs at T1 using the cylindrical source and the line source, respectively. The line and cylindrical sources are the only initializations which produce output which can be compared to the satellite retrieval at T1.

In Figure 2 (row 2, column 1), the passive IR retrieval used to create the RT2 initialization is shown. Plotted to the right are model outputs at T2 created with the cylindrical source and RT1 initialization. (Output from the line source is shown only for T1.)

In Figure 2 (row 3, column 1), the passive IR retrieval used to create the RT3 initialization is shown. Plotted to the right are outputs at T3 created with the cylindrical source, RT1 and RT2 initializations.

In Figure 2 (row 4, column 1), the passive IR retrieval at T4 is shown. Plotted to the right are model outputs at T4 from the cylindrical source, RT1, RT2, and RT3 initializations. In Figure 2 (row 5), the same plots are shown valid at T5. Passive IR retrievals at T4 and T5 were not used to create a model initialization.

4. Spatial Extent and Mass Loading

In ash forecasting applications, the model's ability to determine the horizontal spatial extent of the detectable ash cloud is very important to the aviation community for safety purposes. The horizontal spatial extent of the ash is also the easiest quantity to determine from satellite data. In section 4.1 we describe verification statistics used to evaluate how well the model predicts the horizontal spatial extent of the ash cloud. Then, in section 4.2 we explore the following questions: Which observed initialization predicts the horizontal spatial extent of the ash cloud the best? How do uncertainties in the passive IR retrievals affect the forecasts produced using the observed initializations? Does using a cylindrical source produce a more accurate forecast than simply using a line source? How sensitive is the forecast produced using the line and cylindrical sources to the choice of MER_{fine}? How do forecasts produced using an observed initialization compare to those produced using a

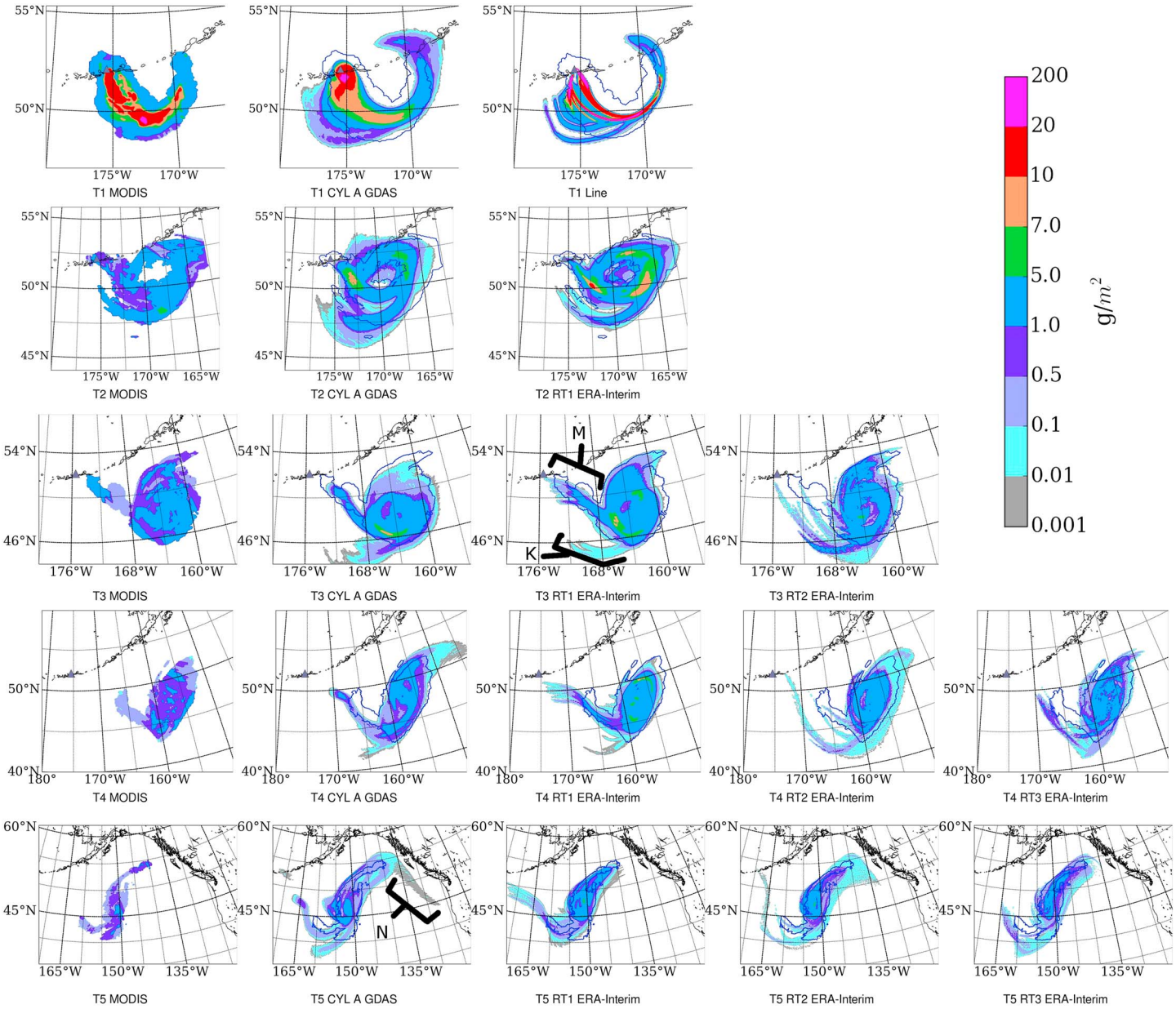

Figure 2. Contour plots of mass loading. (column 1) The satellite retrievals at (rows 1–5) T1-T5. (columns 2–5) HYSPLIT mass loading output with initializations and meteorology as labeled. The thick blue lines outline the footprint of the satellite retrieval shown in Figure 2 (column 1). The black triangle shows the location of Kasatochi. The color bar in the upper right corner applies to all plots. The feature labeled M in the T3 RT1 ECMG plot is an arm of high-altitude ash stretching northwest from the main ash cloud toward the volcano. The feature labeled K is an arm of low-altitude ash stretching southward from the main ash cloud. The altitude of these arms can be seen in Figure 7. Although these features are only annotated in one plot, they can be observed in the other plots at times T3, T4, and T5. The feature labeled N (T5 CYL A GDAS in Figure 2, row 5, column 2) is a long thin arm of ash which stretches along the coast. For the CYL A source shown, the concentrations in this arm are below the 0.1 g m^{-2} threshold. However, for the CYL C and D sources, concentrations in this arm are above this threshold.

cylindrical source? Are the verification metrics used here able to adequately distinguish which forecasts are better? How does the meteorological data set used affect the forecasts?

4.1. Method

The critical success index, CSI, is a statistic which measures the spatial overlap of two data sets [*Webley et al.*, 2009; *Wilks*, 2011].

$$\mathrm{CSI} = \frac{n_{\mathrm{hit}}}{n_{\mathrm{hit}} + n_{\mathrm{miss}} + n_{\mathrm{fa}}} \tag{1}$$

n_{hit} is the number of grid points at which both the satellite detected ash and the HYSPLIT mass loading output was ≥ 0.1 g m^{-2}. n_{miss} is the number of grid points at which the satellite detected ash and the HYSPLIT mass loading output was < 0.1 g m^{-2}. n_{fa} is the number of grid points at which the satellite did not detect ash and the HYSPLIT mass loading output was ≥ 0.1 g m^{-2}. For the model output, a 1 h average mass loading which encompasses the satellite retrieval time was used.

At T4 and T5, HYSPLIT produced mass loading output above the threshold in areas which were outside the view of the satellite. Only grid points which were within the range of the satellite were used to calculate the CSI. We also do not consider model mass which is below 2 km mean sea level, msl, in height since the satellite retrieval is less sensitive to low-level ash. Throughout the text, model mass loading refers to model mass above 2 km.

Two related statistics are the false alarm ratio (FAR) and the probability of detection (POD).

$$\mathrm{FAR} = \frac{n_{fa}}{n_{fa} + n_{hit}} \tag{2}$$

$$\mathrm{POD} = \frac{n_{hit}}{n_{miss} + n_{hit}} \tag{3}$$

While the CSI gives an overall measure of performance, the POD and FAR provide more information.

Distributions of the mass loading are also compared by calculating the Kolmogorov Smirnov parameter (KSP) which is the largest difference between the two cumulative distribution functions, CDFs. The value of the KSP is affected by the chosen threshold of 0.1 g m^{-2} since all points with model mass below this amount do not contribute to the modeled CDF used to calculate the KSP.

In some cases, a lower KSP could be achieved by using a higher or lower value for the threshold. However, without more knowledge about the detection limit of the satellite retrieval as a function of time and space, we must use a single estimated value. Values of KSP above 0.6 generally mean that there is little overlap between the mass loading distributions of the two data sets. In this case, it was not necessary to set a threshold for the upper detection limit of the satellite, but in some cases it might be. The satellite retrieval may not work in areas where the ash cloud is too optically thick [*Pavolonis et al.*, 2015a].

4.2. Results

First, we compare results for the observed initializations, RT1, RT2, and RT3. Then we compare results for the LINE and CYL sources and discuss the importance of choosing a mass eruption rate. Finally, we compare output produced using observed initializations to output produced using the LINE and CYL sources and discuss how choice of meteorological data set affected the forecasts.

All model simulations produced output at two to five times which could be compared to passive IR satellite retrievals. We look at trends in the verification statistics to evaluate the efficacy of different model initializations. Using multiple observations for verification helps assure that conclusions are not sensitive to uncertainties in the observations.

4.2.1. Comparison of Observed Initializations

CSI, POD, FAR, and KSP are plotted for the RT1, RT2, and RT3 initializations in Figures 3a–3d. For the observed initializations, the POD stays within a fairly narrow range, and it is changes in the FAR that have the most effect on the CSI. For instance, a large increase in FAR for the RT1 GDAS source at T4 causes a large decrease in CSI when compared to output at earlier times. This suggests that it is important to identify modeled areas which are not ash and/or areas which might be ash which the satellite did not detect (possibly due to meteorological cloud).

The RT1 initialization tended to produce higher CSI than the RT2 or RT3 initialization. By examining Figure 2 we see that RT1 better captures the placement of the arm of high-altitude ash which stretches northwest from the main ash cloud toward the volcano at T3, T4, and T5 (feature M, row 3, column 3).

Figure 4 shows the ash cloud top heights predicted with the RT1 at T2, the passive IR height retrieval at T2, and a couple cross sections comparing the vertical position of the HYSPLIT particles with the top heights retrieved at T2. At T2, RT1 places the northwest arm of ash higher than the T2 passive IR satellite retrieval does. The higher placement of the arm by RT1 better predicts how this arm will look at T3 through T5. The height retrievals at T2 in this area may be underestimated, possibly because of the presence of multiple layers of ash as observed by CALIOP at T4 (shown in Figure 1d).

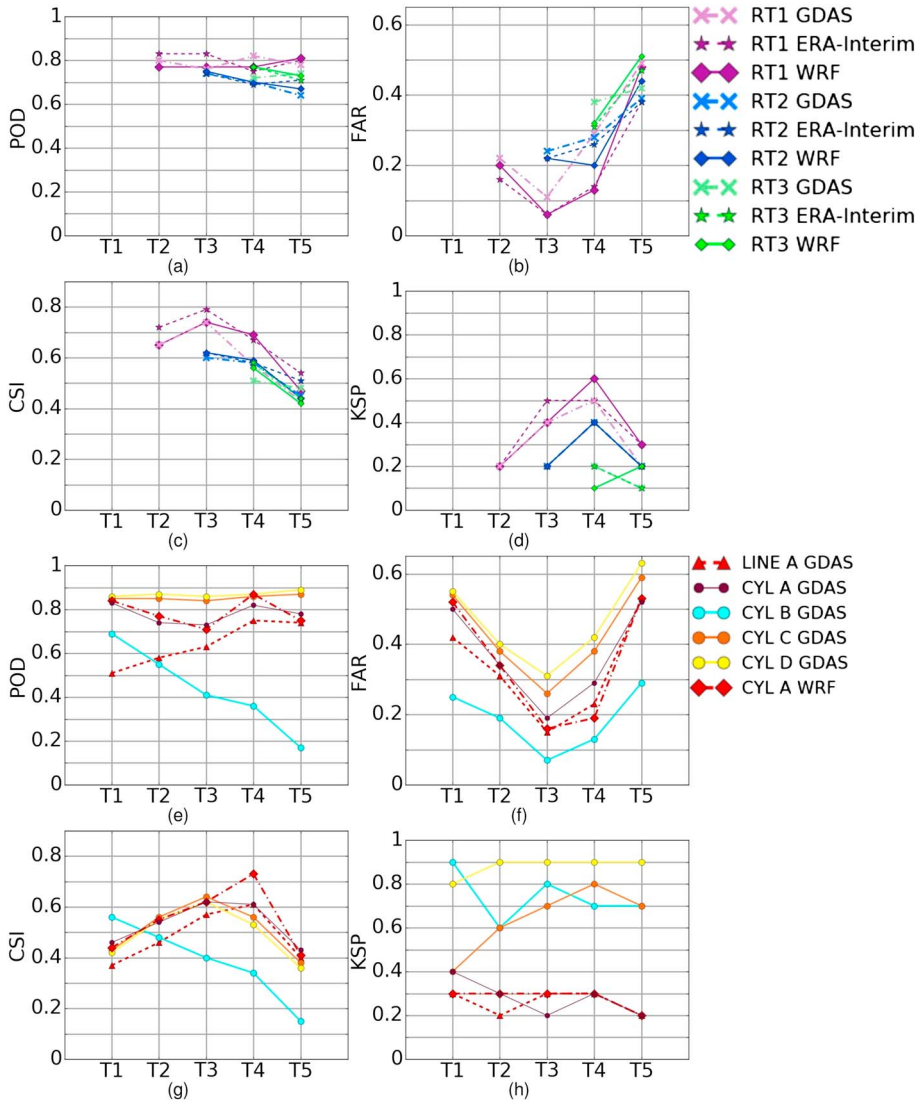

Figure 3. Statistics comparing HYSPLIT model output with satellite retrievals of ash as described in section 4.1. (a) POD, (b) FAR, (c) CSI, and (d) KSP produced with observed initializations. (e) POD, (f) FAR, (g) CSI, and (h) KSP produced with CYL and LINE sources.

We also note that all outputs show a low-altitude diffuse arm of ash stretching southward from the main ash cloud which is not shown in the satellite retrieval (Figure 2, feature K, row 3, column 3). This arm has a smaller mass loading when the model is initialized with RT1 which leads to a lower FAR since more of it is below the threshold. Feature K is faintly identifiable as volcanic ash in multispectral MODIS imagery, but it went undetected by the automated retrieval algorithm. It is possible that either mass loading of this low-level feature was below the detection threshold of 0.1 g m^{-2} or the detection threshold was higher in this area.

To explore the sensitivity of the modeled detectable area to the threshold of detectable ash we look at Figure 5a which shows the CDF for the five satellite retrievals (solid lines) along with CDF for the four outputs produced with the RT1 ERA-Interim initialization. (Note the CDFs in Figure 5 were not used to calculate KSP as the modeled area below threshold is removed before calculating KSP.) As expected, over time the maximum mass loading value decreases, and lower mass loading values make up a larger percentage of the ash cloud. This is due to dispersion of the ash cloud as well as mass moving below 2 km msl or being deposited. Very few passive IR satellite retrievals under 0.1 g m^{-2} are made which supports our decision to use this value as the threshold of detectable ash but does not rule out that the detection limit may have been higher at some places and times. For the data shown in Figure 5a, 18% to 34% of the modeled area is below the threshold.

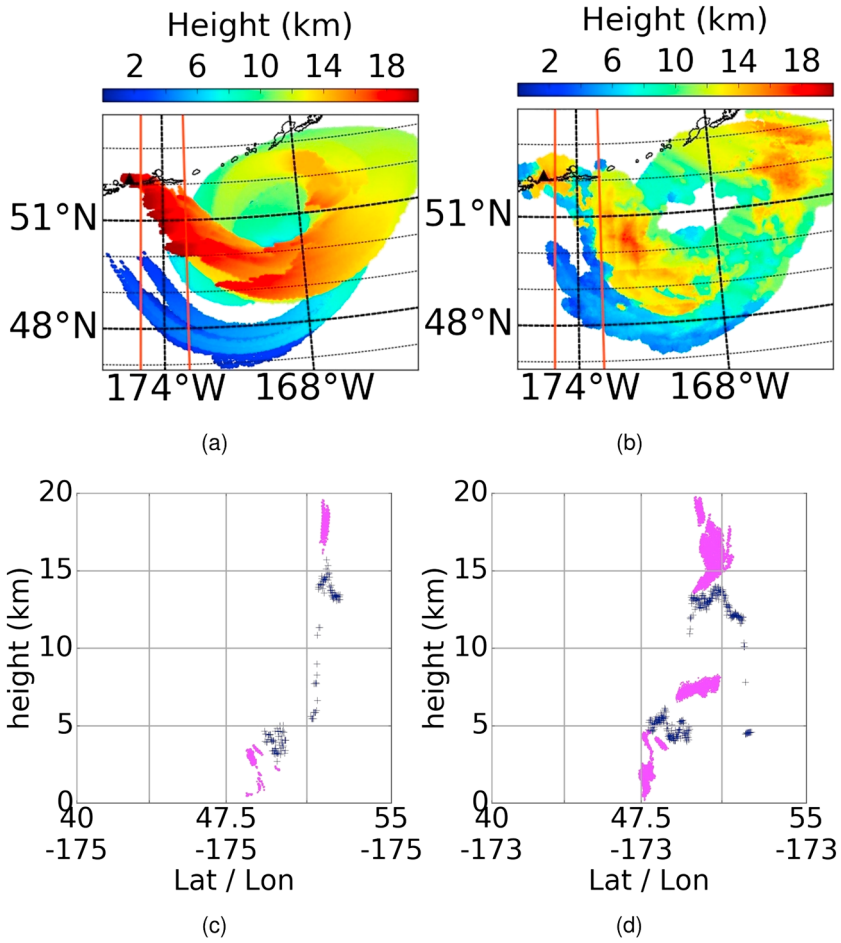

Figure 4. (a) Top heights of modeled ash at T2 produced using RT1 (ERA-Interim). (b) Passive IR satellite retrieval of top height of the ash cloud at T2. The orange lines in Figures 4a and 4b show lines of longitude 175° and 173°. (c, d) Plots of latitude versus altitude along lines of longitude 175°W and 173°W. The magenta points show the position of HYSPLIT particles that lie within 0.1° of the line of longitude. The blue crosses show the passive IR retrieved top heights at T2 that lie within 0.1° of the line of longitude.

The slope of the modeled CDFs around the threshold is fairly steep, meaning that relatively small changes in threshold can change the area of the detectable plume by a significant amount.

The modeled mass loading values cover a wider range than the measured mass loading values, and the maximum modeled mass loading is greater than the retrieved mass loading. This can be seen in the CDF and in the contour plots in Figure 2 where streaks of higher mass loading ($> 5 gm^{-2}$) for the RT1 source (rows 2–5, column 3) are observed. These areas of higher mass loading, which are modeled but not observed, may be due to the presence of multiple layers of ash in some cases (the satellite retrieval is most relevant to the highest ash cloud layer). Thus, metrics which compare modeled and observed mass loading, such as the KSP, should be used only as a coarse measure of whether the modeled mass loading is in the same range as the observed mass loading.

4.2.2. Comparison of Sources Initiated at the Vent

The CSI, POD, FAR, and KSP for the CYL and LINE source are plotted in Figures 3e–3h. For simplicity the results for the line source in Figure 3 are shown only for conversion factor A. For the CYL A, C, and D and LINE source, CSI was lowest at T1 and T5. At early times, the model may have trouble reproducing the position of the ash cloud since the position of the ash may be heavily influenced by the growth of the plume itself. At the intermediate times, T2, T3, and T4, the ash is swept into a synoptic-scale low-pressure system and is being passively transported by the wind field which is captured by the meteorological model.

Using the cylindrical source term to partially account for plume dynamics did increase model performance. Model output produced with the CYL A, C, and D sources has a higher CSI and POD than that produced using

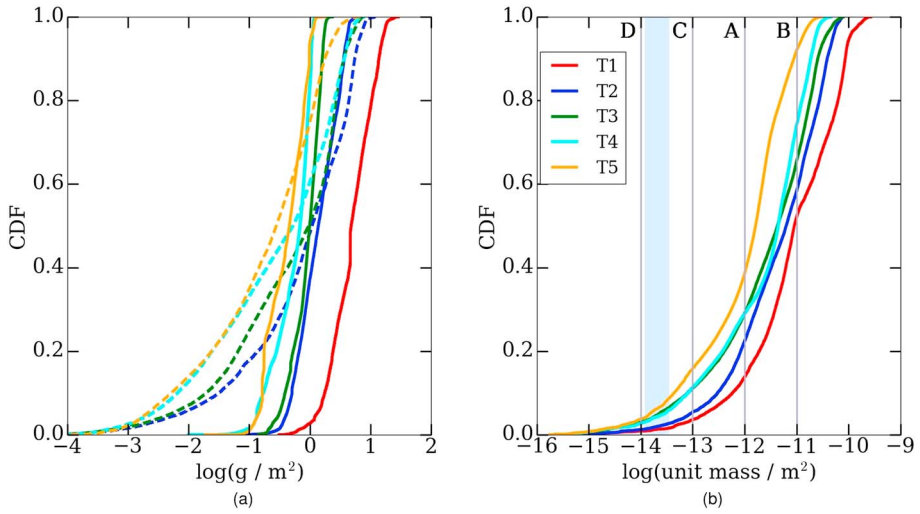

Figure 5. (a) CDF of the mass loading. Solid lines show CDF for satellite retrievals at T1–T5. Dotted lines show CDFs for HYSPLIT output produced using RT1 (ERA-Interim) at times T2–T5. (b) CDF of the mass loading of the model output produced using the cylindrical source (GDAS) in unit mass m^{-2}. The grey vertical lines indicate threshold values assuming different mass eruption rates and a value of 0.1 g m^{-2} for detectable ash. The different mass eruption rates are shown by the grey vertical lines; A: $MER_{fine} = 2.8 \times 10^4$ kg s^{-1}, B: $MER_{fine} = 2.8 \times 10^3$ kg s^{-1}, C: $MER_{fine} = 2.8 \times 10^5$ kg s^{-1}, and D: $MER_{fine} = 2.8 \times 10^6$ kg s^{-1}. The blue shaded area shows values that would be chosen for a 14–18 km high plume using the empirical relationship as described in section 3.3.

the LINE source at T1 through T3. The LINE source produces a plume which is too narrow at these times (Figure 2 (row 1) illustrates this for T1).

Initializing the model with the observations (RT1, RT2, and RT3) has the advantage that the amount of mass in the model initialization is determined by the satellite measurement, while a mass eruption rate must be estimated for the CYL and LINE sources. Here we look at how choice of MER_{fine} and threshold of detectable ash affects the modeled area of detectable ash. This discussion focuses on the cylindrical source term but is expected to be applicable to a uniform line source as well.

Figure 5b shows the CDFs of mass loading of the model output produced using the cylindrical source in unit mass m^{-2}. This plot quickly illustrates how choosing a MER_{fine} and threshold of detectable ash affects the predicted lifetime of the ash cloud. If a MER_{fine} of 2.8×10^3 kg s^{-1} is used (CYL B source), then by T5 only a small percentage of the modeled ash cloud is detectable. However, if a mass eruption rate of 2.8×10^4 kg s^{-1} is chosen (CYL A source), then over half of the ash cloud at T5 is detectable. The plot also shows how the choice of MER_{fine} affects the modeled area of the detectable plume at a single time. At all time periods, adjusting MER_{fine} from B to A results in a larger change to the modeled detectable area than adjusting it from A to C, which in turn results in a larger change in the detectable area than adjusting it from C to D. Setting MER_{fine} greater than D would have very little effect on the detectable area for the time periods investigated here. It would affect later forecasts, as the plume dispersed. Note that the lines D, C, A, and B on the plot could also be drawn by keeping MER_{fine} constant and changing the threshold of detectable ash.

Examining Figure 2 (column 2), which shows contour plots of the ash mass loading at the five time periods, confirms that choice of MER_{fine} is critical, particularly for forecasting the area of older, larger, more diffuse ash clouds. Although only output produced using the CYL A source is shown in Figure 2 (column 2), output for CYL B, C, and D sources would be identical except the values for the contour lines would change. For the CYL B source, all contours below 1 g m^{-2} are below the threshold value. For the CYL C source, contours above 0.01 g m^{-2} are all above threshold. While for the D source, contours above 0.001 g m^{-2} are above threshold. Thus, at the fifth time period, the CYL B source shows an ash cloud which has almost dissipated, while the CYL D source shows a large ash cloud with a long tail stretching along the coast (feature N, Figure 2, row 5, column 2). The area produced using the CYL A source most closely matches the area identified as ash by the passive IR satellite retrieval.

Although the CSI for the CYL A source is only slightly better than the CYL C and CYL D source at T5, an examination of the mass loading contour plot may lead to a subjective judgment that its output is much

better since it does not show the long thin arm of ash extending along the coast (feature N in Figure 2, row 5, column 2). A different type of statistic may be needed to take into account that false alarms which lie farther outside the observed area are more important.

The CYL A source is the only cylindrical source for which the KSP$\leq$0.4 for all five time periods (Figure 3h). This indicates that it is best at reproducing the observed mass loading. The CYL C source has the same KSP as the CYL A source at T1. However, while the CYL A source slightly underpredicts mass loading values at T1, the CYL C source slightly over predicts them. This over prediction increases with time, leading to an increase of KSP values ($\geq$0.6) for CYL C at T2 through T5. Some plots of the CDFs used to calculate KSP can be found in the supporting information.

CSI values from the CYL A source are comparable to those from the CYL C and D sources at T1 through T3 and slightly better at T4 and T5. Which forecast is preferable would depend on the application. The CYL D source consistently produced the highest POD and FAR. So if a conservative forecast is needed, then using a high MER_{fine} with a cylindrical source term above the vent may be desirable. CYL A decreases FAR significantly while maintaining a fairly high POD. It also provides a better quantitative estimate of mass loading.

The CYL B source has lower POD and FAR than the other cylindrical sources at all times. The low FAR leads to a high CSI compared to the other sources at T1. However, the POD decreases markedly over time leading to a large decrease in CSI as well. Overestimating MER_{fine} may be preferable to underestimating it.

4.2.3. Comparison of Sources Initiated at the Vent and Observed Initializations

Outputs produced using the observed initializations often, but not always, have a higher CSI than the outputs from sources initiated at the vent (see Figures 3c and 3g).

At T2 and T3, the RT1 initialization produces a higher CSI than the LINE or CYL sources. However, the RT2 initialization at T3, produces a CSI which is comparable to or slightly lower than the CYL A, C, and D sources but still higher than the LINE A or CYL B source.

At T4, the CYL A WRF source produces the highest CSI (0.73), with the RT1 initializations with WRF and ERA-Interim slightly lower (0.69 and 0.67). CYL B (0.34) and RT3 GDAS (0.51) initializations produce the lowest CSIs at T4. The rest all lie between 0.56 and 0.61.

At T5, the observed initializations produce CSI values which range from 0.41 to 0.54, while the LINE and CYL sources produce CSI values which range from 0.15 to 0.43.

Again, we see that if a conservative forecast is needed, a cylindrical source term with a large mass eruption rate may be appropriate. However, the RT1 initialization generally produces forecasts with significantly lower FAR and a POD which remains above 0.75.

The resolution of the meteorological data set does not have a large effect in this case. This is not surprising as the ash is mostly being transported over ocean by a synoptic scale low-pressure system. Although changing the meteorological data set used did change the details of the forecast, no clear trend could be identified in which one meteorological data set consistently produced better forecasts.

5. Vertical Distribution of Ash

We make some qualitative assessments of the vertical distribution of ash and the three-dimensional structure of the ash cloud by using data from CALIOP and the passive IR retrieved top heights. We also provide further support for our assessment that the CYL source term is an improvement over the LINE source, the RT1 initialization produces more accurate forecasts than the RT2 or RT3 initialization, and the presence of multiple layers of ash may affect the value of the passive IR retrieval of mass loading and height. For these comparisons, the instantaneous positions of HYSPLIT particles at the hour closest to the time at which the satellite instrument made the measurements are used. Since we are simply looking at the position of the HYSPLIT particles, the plots shown in this section are not dependent on choosing a MER_{fine} or threshold of detectable ash.

A comprehensive analysis of the three-dimensional structure of the ash cloud is beyond the scope of this paper. Future work might include using the technique described in *Winker et al.* [2012] to obtain an estimate of mass concentration from the CALIOP data. This could then be compared to HYSPLIT concentration output. It may also be interesting to follow the approach of *Peng and Peterson* [2012] and investigate how features detected by CALIOP and the position of HYSPLIT computational particles correspond to attracting structures in the flow.

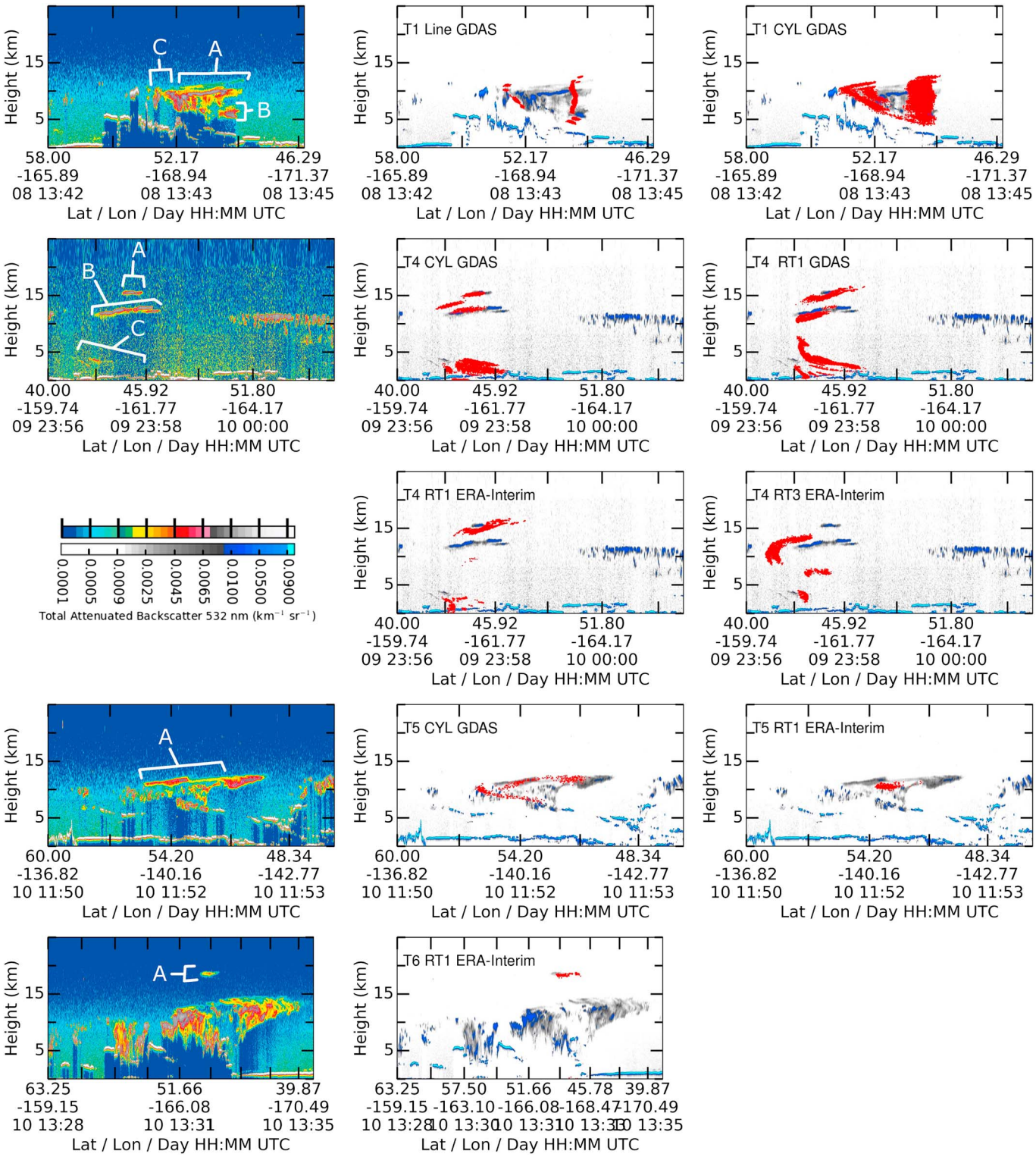

Figure 6. (column 1) The 532 nm backscatter from CALIOP using colors shown in the top bar. Features which may be ash are labeled with capital letters. (columns 2 and 3) The 532 nm backscatter from CALIOP shown in shades of grey and blue (values shown in bottom color bar). Position of HYSPLIT particles within 0.1° of the track of the lidar are shown in red. The time period, initialization, and meteorological data set are shown in the top left corner of each plot.

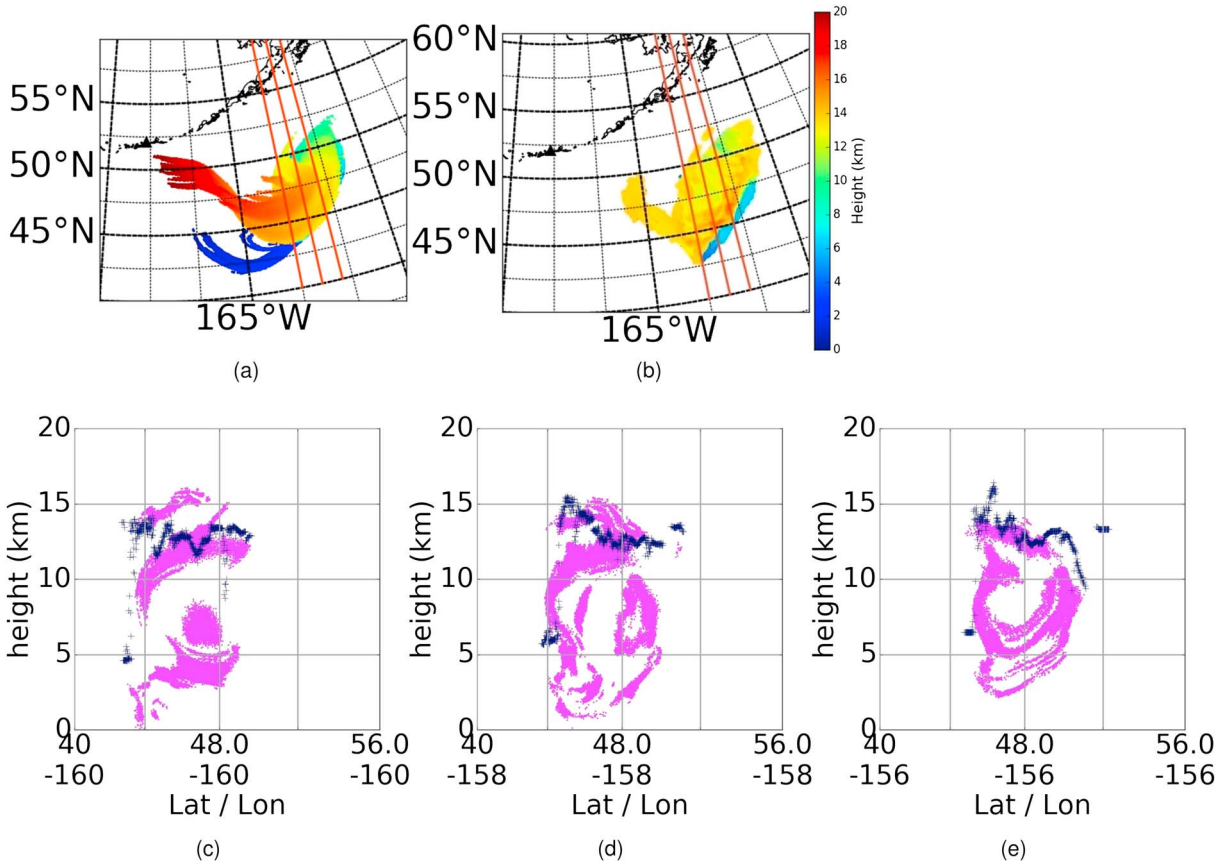

Figure 7. (a) Top height of model output at T4 produced using RT1 initialization and ERA-Interim. (b) Passive IR satellite retrieval of top height at T4. For Figures 7a and 7b the orange lines show lines of longitude along which a vertical profile is plotted in Figures 7c–7e. (c–e) Magenta circles show position of HYSPLIT particles within 0.1° of the line of longitude. Blue crosses show position of passive IR satellite retrieval within 0.1° of line of longitude. 160°W (Figure 7c), 158°W (Figure 7d), and 156°W (Figure 7e).

5.1. CALIOP Lidar and Model Comparison

Figure 6 compares the vertical profile of the HYSPLIT output to the CALIOP measurements using various model initializations and meteorology.

At T1, the plume produced by the line source is too narrow (Figure 6, row 1, column 2). The cylindrical source (Figure 6, row 1, column 3) reproduces the wedge shape (thin on the north side and thick on the south side) of the CALIOP feature, but the plume appears to be shifted slightly too far to the east.

At T4 (Figure 6, rows 2 and 3), three CALIOP features are identified as possible ash. Features A and B are almost certainly ash, while feature C may be ash. Most of the HYSPLIT forecasts, excepting those initiated with RT3, show ash near all three of these features. The lidar track at T4 cuts across the base of the arm of high-altitude ash which stretches northwest of the main ash cloud. As discussed in section 4.2.1, the RT2 and RT3 initializations had trouble placing the position of this arm correctly.

At T5 (Figure 6, row 4), all the initializations produce output which predicts ash at the location of feature A which is the portion of the CALIOP feature most likely to contain ash. RT1 places ash mainly at feature A, while the other initializations also place ash at other locations in the CALIOP feature. Interestingly, at this time none of the initializations produce output which places ash in an area where no CALIOP feature is seen.

A CALIOP overpass at T6 shows a thin feature at around 18 km which is most likely volcanic (feature A in Figure 6, row5, column 1). The only forecast which reproduces this feature is the forecast initiated with RT1. The height and thickness of the feature are forecast correctly, but the position is shifted to the south. A few of the other forecasts (initiated with the cylindrical source and RT2) show a similar feature, a high tail of ash stretching from the main ash cloud back toward the volcano, but it does not stretch far enough to the west to intersect the CALIOP overpass.

5.2. Comparison of Model Heights and Satellite-Retrieved Heights

In Figure 7 we compare the position of HYSPLIT particles to the satellite retrieval of top height at T4 by plotting latitude versus altitude along lines of longitude. Here we pick representative slices through the center of the ash cloud. Although there are large differences between the position of the highest HYSPLIT particle and the retrieved top height at many positions, the position of the HYSPLIT particles tends to follow the trend of the top height retrieval. Agreement between the retrieved top height and model output tends to be less at the edges of the ash cloud. The supporting information contains animations which show a sequence of these plots at 0.2 ° longitude intervals beginning on the west side of the ash cloud and traveling through to the east side of the ash cloud. The shape of the ash cloud as defined by the position of the computational particles is quite complicated, and it would be difficult to assign a single thickness or top height to the entire ash cloud. The CALIOP retrieval and the retrieved top heights from the MODIS instrument support this picture of the ash cloud. The patchy looking nature of the satellite retrievals may reflect the complex nature of the ash cloud. For instance, the abrupt jump in height shown on the satellite retrieval at T4 on the southeast side of the ash cloud agrees with the modeled profile of the ash particles as shown in Figure 7.

This supports the conjecture that some differences between modeled mass loading and retrieved mass loading could be due to the presence of multiple layers of ash.

6. Discussion and Conclusions

A vertical cylindrical source around the vent may produce better forecasts than a vertical line source above the vent for large eruptions, particularly if an estimate of the width of the umbrella cloud is available. This is a simple way to account for plume dynamics which can transport ash far from the vent.

Using satellite retrievals of mass loading to determine the mass of fine ash in the ash cloud can decrease uncertainty in the long-term forecast of a transport and dispersion model. In an operational setting, the empirical equation relating MER to plume height and an estimate of the fraction of fine ash may first be used to determine MER_{fine} if no other information is available. If this estimation produces a modeled ash cloud with an area either larger or smaller than that observed by an experienced analyst, the mass eruption rate may be adjusted up or down so the area better matches. However, this case illustrates that adjusting the MER so the output area matches the observed area at one time does not necessarily lead to a better forecast at later times. Mass loading contours may be very close together at early times, and thus, area matching may not be a very sensitive way to differentiate between choices.

If the mass is overestimated, then forecasts at later times are apt to overestimate the area which will be covered with detectable ash. If the mass is underestimated, then forecasts at later times will be apt to underestimate this area. Consequently, mass loading information from a passive IR satellite retrieval or other source should be incorporated as soon as it becomes available. Although the passive IR satellite retrieval used here often fails at the onset of the eruption since the cloud is too opaque, an estimate of MER from cloud growth may eventually be available for use [*Pouget et al.*, 2013]. Additionally, even a measurement of mass loading made for only part of the ash cloud (as may happen if meteorological cloud is covering part of the ash cloud) may be used to help set a MER for a source term starting at the vent.

The HYSPLIT transport and dispersion model was able to predict the height and thickness of ash clouds reasonably well and was able to resolve multiple thin ash layers. The HYSPLIT forecast of vertical ash position generally agreed well with CALIOP data. HYSPLIT vertical ash position and retrieved top heights generally showed the same pattern, particularly in the middle of the ash cloud. Differences were most pronounced at the edges. The passive IR retrieved top heights, CALIOP measurements and model output all support the idea that ash clouds can develop a complex three-dimensional structure which would make assigning a single top and bottom height to an ash cloud challenging.

Even though the thickness of the ash cloud needs to be estimated, initializing the model with the observed ash top height, mass loading, and effective radius usually performed as well or better than initializing the model with a source located at the vent (as measured by CSI). For this case, using the earliest satellite observation for initialization produced comparable or better forecasts than using later observations. This may be because the ash cloud had a simpler, less layered structure at the earliest time. Retrieved top heights may also be more accurate when the ash cloud is more optically thick. This suggests that using an early passive IR retrieval when the ash cloud is more optically thick and has a simpler structure may be more desirable than using the latest

retrieval to initialize the model. However, when using forecast meteorology, in which the model winds will have larger errors, it may be more important to use the most recent satellite observation. This may also be the case for more complex wind fields or when the ash cloud approaches a bifurcation point in the wind field.

The observed initialization is simple to construct if an appropriate passive IR retrieval is available. It requires very little computational resources and no information about the previous history of the ash. This is important since the history of the ash can be complex. The assumption that ash was passively transported by the wind field from some height above the vent may not be a good one in many eruptions. Ash may be transported hundreds of kilometers from the vent by the inertial spreading of the plume [*Costa et al.*, 2013; *Pouget et al.*, 2013; *Johnson et al.*, 2015]. In some eruptions, resuspended ash from locations not particularly close to the vent may make a significant contribution to the detected ash cloud [*Thorsteinsson et al.*, 2009]. Inversion algorithms can be a powerful tool, but they do require some bounds on the initial locations and time of release. Usually, they assume ash originated above the vent and then calculate the vertical distribution of ash above the vent as a function of time. Increasing the number of possible source locations and times generally increases the computational resources required.

When initializing the model with a satellite retrieval, as performed here, the main uncertainties arise from the need to estimate the ash cloud thickness and uncertainties in the retrieved properties, particularly ash cloud height. These uncertainties do not seem to be any larger than those encountered in other methods of initializing the model, and in many cases they may be smaller.

There are several limitations to consider when using an observed initialization. For ongoing eruptions, forecasters need to produce forecasts for ash which has not yet been emitted as well as ash which is observed. It is simple to combine model output from multiple HYSPLIT runs, so model output produced using an observed initialization as described here could be combined with model output produced using an estimated source term for the future ash. Forecasting the position of ash not yet emitted is difficult since the duration and future plume height must be estimated and eruptions are unpredictable. Consequently, it may be desirable to show these forecasts separately as well as combined so that areas covered only by predicted ash can be assigned a higher uncertainty.

The method of creating an observed initialization described here is broadly applicable and there are many ways it could be modified and applied. For instance *Wilkins et al.* [2014, 2016] use a similar technique with a different transport and dispersion model (NAME), meteorological model (Met Office Unified Model), volcanic eruption (Eyjafjallajökull May 2010), and retrieval algorithm [*Francis et al.*, 2012] applied to data from a different imager (SEVIRI sensor aboard the geostationary satellite Meteosat). Since their data were from a geostationary satellite, their passive IR satellite retrievals had a much higher temporal resolution.

Wilkins et al. [2016] used six passive IR retrievals that spanned a 35 h time range to create "data insertion forecasts" which used an initialization similar to our observed initialization. In addition to using a single passive IR retrieval to create an initialization, they also looked at ways of combining passive IR retrievals to create an initialization. For verification, they compared model output to one passive IR retrieval which occurred 3 h after the last passive IR retrieval used for the data insertion. They used CSI (or figure of merit in space) and a quality measure called SAL (structure amplitude location) for the comparison. They found that as measured by SAL and CSI, the data insertion forecasts performed as well or better than a NAME forecast produced using a vertical line source similar to what might be run operationally, and NAME and FLEXPART forecasts produced using an inversion algorithm to create a source term. In their case, using the most recent passive IR retrieval to produce a model initialization tended to produce better results as measured by the CSI and SAL. The eruption was ongoing, so the later passive IR retrievals captured ash which had not yet been emitted when the earlier passive IR retrievals occurred. A notable exception was that the second passive IR retrieval did not produce a better forecast than the first passive IR retrieval which they conjectured might be due to the second passive IR retrieval being of poorer quality. Initializations created with multiple passive IR retrievals tended to produce more conservative forecasts which had worse CSI and SAL scores.

Wilkins et al. [2014, 2016] point out that comparisons of passive IR retrievals and model output would be of more value if uncertainties in the measurements and/or model output were better known.

As a next step, we plan to incorporate knowledge of the location of meteorological cloud to indicate areas where the passive IR retrieval may miss ash. Creating an ensemble of model outputs may be a useful way to incorporate model uncertainty. Future work will focus on investigating the use of observed initializations for

a wide range of meteorological and eruptive conditions. One of the advantages of utilizing the automated passive IR retrievals is that verification may be automated to a certain extent which may allow for the analysis of more eruptions.

More statistical measures will need to be utilized. For this case, all forecasts produced relatively good overlap with the observations which made the CSI an appropriate measure. However, to differentiate between forecasts which have little or no overlap with the observations, a statistical tool, such as the fractions skill score [*Harvey and Dacre*, 2015; *Roberts and Lean*, 2007], may be used. An object-based approach may also be appropriate [*Wolff et al.*, 2014; *Miller et al.*, 2014; *Wilkins et al.*, 2016], particularly since the algorithm which identifies ash from the passive IR measurements identifies ash objects based on temporal-spatial as well as spectral information [*Pavolonis et al.*, 2015a, 2015b]. This type of approach may also better distinguish between cases where false alarms lie close to the observed area and cases where false alarms cover a small but significant area far from the observed area, such as we saw at T5. An object-based approach which utilizes multiple ash retrievals from geostationary satellites which have high temporal resolution may also allow this method to be adapted for situations in which retrievals are incomplete due to meteorological cloud cover.

Appendix A: Changing the Estimated Thickness of the Initialization

As discussed in section 3.3, the initial altitude range of the observed initializations must be estimated since the satellite retrieval only gives information about the ash cloud top height and not the ash cloud thickness. In addition to the 2.5 km thickness (with the top at the passive IR satellite-retrieved height) described in the main text, we tried initializations with the following estimated thicknesses.

1. Particles were released from positions 3 km above the satellite-retrieved top height to a point 3 km below the satellite-retrieved top height. This was done for satellite retrievals at T1 and T2.
2. Particles were released from positions 2 km above the satellite-retrieved top height to a point 1 km above the ground. This was done for satellite retrievals at T1 and T2.
3. Particles were released only at the height of the satellite retrieval of ash cloud top height. This was done for the satellite retrieval at T2.

We used only the ERA-Interim meteorology for these model runs. The CSI changed by less than 0.10 when one of the above thicknesses were used in place of the 2.5 km thickness. For the thicker ash clouds, the POD and FAR both tended to increase as would be expected. In a few cases the POD or FAR decreased slightly or stayed almost the same. This was due to output mass loading of some areas dropping below threshold for the thicker ash cloud initialization since the same amount of mass was now initially spread out over a larger volume. For the thinner initialization, the POD and FAR both decreased slightly but the decrease in CSI was less than 0.04.

Acknowledgments

The authors are grateful to the Federal Aviation Administration (FAA) for funding part of this study. The views expressed are those of the authors and do not necessarily represent the official policy or position of the FAA or NOAA. CALIOP data were obtained from the NASA Langely Research Center Atmospheric Science Data Center. ECMWF data were obtained using their publicly available web and python api's http://apps.ecmwf.int/datasets/. The HYSPLIT model is available at http://ready.arl.noaa.gov. The MODIS satellite data and HYSPLIT model output and input files are available from the authors (alice.crawford@noaa.gov). We are grateful to Jaime Kibler, Greg Gallina, and Grace Swanson for many conversations regarding VAAC operations; to Justin Sieglaff for providing technical assistance on interpreting and utilizing the satellite retrievals; and to Roland Draxler, Ariel Stein, and Tianfeng Chai for their insightful comments on the use of the HYSPLIT model and this manuscript.

References

Corradini, S., L. Merucci, A. J. Prata, and A. Piscini (2010), Volcanic ash and SO_2 in the 2008 Kasatochi eruption: Retrievals comparison from different IR satellite sensors, *J. Geophys. Res.*, *115*, D00L21, doi:10.1029/2009JD013634.

Costa, A., A. Folch, and G. Macedonio (2013), Density-driven transport in the umbrella region of volcanic clouds: Implications for tephra dispersion models, *Geophys. Res. Lett.*, *40*, 4823–4827, doi:10.1002/grl.50942.

Dare, R. A. (2015), Sedimentation of volcanic ash in the HYSPLIT dispersion model, *Tech. Rep. No. 079*, The Centre for Australian Weather and Climate Research, Melbourne, Victoria, Australia.

Dee, D. P., et al. (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, *137*(656), 553–597, doi:10.1002/qj.828.

Devenish, B. J., P. N. Fancis, B. T. Johnson, R. S. J. Sparks, and D. J. Thomson (2012), Sensitivity analysis of dispersion modeling of volcanic ash from Eyjafjallajökull in May 2010, *J. Geophys. Res.*, *117*, D00U21, doi:10.1029/2011JD016782.

Draxler, R. R., and G. D. Hess (1997), Description of the HYSPLIT_4 modeling system, *Tech. Rep. arl-224*, NOAA Air Resources Laboratory, Silver Spring, Md.

Draxler, R. R., and G. D. Hess (1998), An overview of the HYSPLIT_4 modeling system for trajectories, dispersion, and deposition, *Aust. Meteorol. Mag.*, *47*, 195–308.

Draxler, R. R., and G. D. Rolph (2012), Evaluation of the Transfer Coefficient Matrix (TCM) approach to model the atmospheric radionuclide air concentrations from Fukushima, *J. of Geophys. Res.*, *117*, D05107, doi:10.1029/2011JD017205.

Draxler, R. R., P. Ginoux, and A. F. Stein (2010), An empirically derived emission algorithm for wind-blown dust, *J. Geophy. Res. Atmos.*, *115*, D16212, doi:10.1029/2009JD013167.

Elíasson, J., and J. Yoshitani (2015), Airborne measurements of volcanic ash and current state of ash cloud prediction, *Disaster Prevention Res. Inst. Ann. B*, *58*, 35–41.

Folch, A. (2012), A review of tephra transport and dispersal models: Evolution, current status, and future perspectives, *J. Volcanol. Geotherm. Res.*, *235*, 96–115.

Francis, P. N., M. C. Cooke, and R. W. Saunders (2012), Retrieval of physical properties of volcanic ash using Meteosat: A case study from the 2010 Eyjafjallajökull eruption, *J. Geophys. Res.*, *117*, D00U09, doi:10.1029/2011JD016788.

Harvey, N. J., and H. F. Dacre (2015), Spatial evaluation of volcanic ash forecasts using satellite observations, *Atmos. Chem. Phys. Discuss.*, *15*, 24,727–24,749.

Heffter, J. L, and BJ. B. Stunder (1993), Volcanic ash Forecast Transport and Dispersion (VAFTAD) model, *Comput. Tech.*, *8*, 533.
Johnson, C. G., A. J. Hogg, E. H. Herbert, R. S. J. Sparks, J. C. Phillips, A. C. Slim, and M. J. Woodhouse (2015), Modeling intrusions through quiescent and moving ambients, *J. Fluid Mech.*, *771*, 370–406.
Kristiansen, N. I., A. Stohl, A. J. Prata, A. Richter, S. Eckhardt, P. Seibert, A. Hoffmann, C. Ritter, L. Bitar, T. J. Duck, and K. Stebel (2010), Remote sensing and inverse transport modeling of the Kasatochi eruption sulfur dioxide cloud, *J. Geophys. Res.*, *115*, D00L16, doi:10.1029/2009jd013286.
Kristiansen, N. I., et al. (2012), Performance assessment of a volcanic ash transport model mini-ensemble used for inverse modeling of the 2010 Eyjafjallajökull eruption, *J. Geophys. Res.*, *117*, D00U11, doi:10.1029/2011jd016844.
Leadbetter, S. J., and M. C. Hort (2011), Volcanic ash hazard climatology for an eruption of Hekla Volcano, Iceland, *J. Volcanol. Geotherm. Res.*, *199*, 230–241.
Mastin, L., A. R. Van Eaton, and J. B. Lowenstern (2014), Modeling ash fall distribution from a Yellowstone supereruption, *Geochem. Geophys. Geosyst.*, *15*, 2474–2495, doi:10.1002/2014GC005469.
Mastin, L. G., et al. (2009), A multidisciplinary effort to assign realistic source parameters to models of volcanic ash-cloud transport and dispersion during eruptions, *J. Volcanol. Geotherm. Res.*, *186*(1–2), 10–21, doi:10.1016/j.jvolgeores.2009.01.008.
Miller, S. D., C. E. Weeks, R. G. Bullock, J. M. Forsythe, P. A. Kcuera, B. G. Brown, C. A. Wolff, P. T. Partain, A. S. Johnes, and D. B. Johnson (2014), Model-evaluation tools for three-dimensional cloud verification via spaceborne active sensors, *J. Appl. Meteorol. Climatol.*, *53*, 2181–2195.
Moxnes, E. D., N. I. Kristiansen, A. Stohl, L. Clarisse, A. Durant, K. Weber, and A. Vogel (2014), Separation of ash and sulfur dioxide during the 2011 Grimsvötn eruption, *J. Geophys. Res. Atmos.*, *119*, 7477–7501, doi:10.1002/2013JD021129.
Pavolonis, M. J., A. K. Heidinger, and J. Seiglaff (2013), Automated retrievals of volcanic ash and dust cloud properties from upwelling infrared measurements, *J. Volcanol. Geotherm. Res.*, *118*, 1436–1458, doi:10.1002/jgrd.50173.
Pavolonis, M. J., J. Sieglaff, and J. Cintineo (2015a), Spectrally enhanced cloud objects: A generalized framework for automated detection of volcanic ash and dust clouds using passive satellite measurements: 1. Multispectral analysis, *J. Geophys. Res. Atmos.*, *120*, 7813–7841, doi:10.1002/2014JD022968.
Pavolonis, M. J., J. Sieglaff, and J. Cintineo (2015b), 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, *J. Geophys. Res. Atmos.*, *120*, 7842–7870, doi:10.1002/2014JD022968.
Peng, J., and R. Peterson (2012), Attracting structures in volcanic ash transport, *Atmos. Environ.*, *48*, 230–239.
Pouget, S., M. Bursik, P. Webley, J. Dehn, and M. J. Pavolonis (2013), Estimation of eruption source parameters from umbrella cloud or downwind plume growth rate, *J. Volcanol. Geotherm. Res.*, *258*, 100–112, doi:10.1016/j.jvolgeores.2013.04.002.
Prata, A. J., and A. T. Prata (2012), Eyjafjallajökull volcanic ash concentrations determined using spin enhanced visible and infrared imager measurements, *J. Geophys. Res.*, *117*, D00U23, doi:10.1029/2011JD016800.
Prata, A. T., S. T. Siems, and M. J. Manton (2015), Quantification of volcanic cloud top heights and thicknesses using A-train observations for the 2008 Chaitén eruption, *J. Geophys. Res. Atmos.*, *120*, 2928–2950, doi:10.1002/2014JD022399.
Roberts, N. M., and H. W. Lean (2007), Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events, *Mon. Weather Rev.*, *136*, 78, doi:10.1175/2007MWR2123.1.
Rolph, G. D., R. R. Draxler, A. f. Stein, A. Taylor, M. G. Ruminski, S. Kondragunta, J. Zeng, H. Huang, G. Manikin, J. T. McQueen, and P. M. Davidson (2009), Description and verification of the NOAA smoke forecasting system: The 2007 fire season, *Weather Forecasting*, *24*, 361–378.
Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X. Y. Huang, W. Wang, and J. G. Powers (2008), A description of the advanced research WRF version 3, *NCAR Tech Note NCAR/TN-475+STR*, Mesoscale and Microscale Meteorology Division, National Center for Atmospheric Research, Boulder, Colo.
Stein, A. F., R. R. Draxler, G. D. Rolph, B. J. B. Stunder, M. D. Cohen, and F. Ngan (2015), NOAA's HYSPLIT atmospheric transport and dispersion modeling system, *Bull. Am. Meteorol. Soc.*, *96*, 2059–2077, doi:10.1175/BAMS-D-14-00110.1.
Stein, A. F., G. D. Rolph, R. R. Draxler, B. Stunder, and M. Ruminski (2009), Verification of the NOAA smoke forecasting system: Model sensitivity to the injection height, *Weather Forecasting*, *24*, 379–394.
Stohl, A., et al. (2011), Determination of time- and height resolved volcanic ash emissions and their use for quantitative ash dispersion modeling: The 2010 Eyjafjallajökull eruption, *Atmos. Chem. Phys.*, *11*(10), 4333–4351, doi:10.5149/acp-11-4333-2011.
Stunder, B., J. L. Heffter, and R. R. Draxler (2007), Airborne volcanic ash forecast area reliability, *Weather Forecasting*, *22*, 1132, doi:10.1175/WAF1042.1.
Thorsteinsson, T., T. J'ohannsson, A. Stohl, and N. Kristiansen (2009), High levels of particular matter in Iceland due to direct ash emissions by the Eyjafjallajökull eruption and resuspension of deposited ash, *J. Volcanol. Geotherm. Res.*, *186*(1–2), 22–31, doi:10.1016/j.jvolgeores.2009.02.015.
Vernier, J. P., et al. (2013), An advanced system to monitor the 3D structure of diffuse volcanic ash clouds, *J. Appl. Meteor. Climatol.*, *52*, 2125–2138, doi:10.11175/JAMC-D-12-0279.1.
Waythomas, C. F., W. E. Scott, S. G. Prejean, D. J. Schneider, P. Izbekov, and C. J. Nye (2010), The 7–8 August 2008 eruption of Kasatochi Volcano, Central Aleutian Islands, Alaska, *J. Geophys. Res.*, *115*, B00B06, doi:10.1029/2010JB007437.
Webley, P., B. J. B. Stunder, and K. G. Dean (2009), Preliminary sensitivity study of eruption source parameters for operational volcanic ash cloud transport and dispersion models—A case study of the August 1992 eruption of the Crater Peak vent, Mount Spurr, Alaska, *J. Volcanol. Geotherm. Res.*, *186*, 108–119.
Wilkins, K. L., I. M. Watson, N. I. Kristiansen, H. N. Webster, D. J. Thomson, H. F. Dacre, and A. J. Prata (2016), Using data insertion with the NAME model to simulate the 8 May 2010 Eyjafjallajökull volcanic ash cloud, *J. Geophys. Res. Atmos.*, *121*, 306–323, doi:10.1002/2015JD023895.
Wilkins, K. L., S. Mackie, I. M. Watson, H. N. Webster, D. J. Thomson, and H. F. Dacre (2014), Data insertion in volcanic ash cloud forecasting, *Ann. Geophys.*, *57*, 1–6.
Winker, D. M., Z. Liu, A. Omar, J. Tackett, and D. Fairlie (2012), CALIOP observations of the transport of ash from the Eyjafjallajökull volcano in April 2010, *J. Geophys. Res.*, *117*, D00U15, doi:10.1029/2011JD016499.
Winker, D. M., et al. (2010), The CALIPSO Mission: A global 3D view of aerosols and clouds, *Bull. Am. Meteorol. Soc.*, *91*, 1211–1229.
Winker, D. M., J. Pelon, M. A. Vaughan, A. Omar, Y. Hu, K. Powell, Z. Liu, W. H. Hunt, and S. A. Young (2009), Overview of the CALIPSO Mission and CALIOP data processing algorithms, *J. Atmos. Oceanic Technol.*, *26*, 2310–2323.
Wilks, D. S. (2011), *Statistical Methods in the Atmospheric Sciences*, 3rd ed., Elsevier, New York.
Wolff, J. K., M. Harrold, T. Fowler, J. H. Gotway, L. Nance, and B. G. Brown (2014), Beyond the basics: Evaluating model-based precipitation forecasts using traditional, spatial and object-based methods, *Weather Forecasting*, *29*, 1451–1472.