Satellite Derived Volcanic Ash Product Inter-Comparison in Support to SCOPE-Nowcasting

Final report

Version 1.0

Prepared by

Gareth Thomas and Richard Siddans

RAL Space

Remote Sensing Group

Harwell Oxford

Didcot, OX11 0QX

United Kingdom

EUMETSAT Contract: EUM/C0/13/ 4600001276/PDW (Rider 1)

2016-04-01

1 CONTENTS

2 Acronyms Used in this document 2
3 Introduction 3
4 Background 3
5 Study cases 4
6 Intercomparison approach 8
6.1 Product regridding 9
6.2 Pair-wise comparisons at 0.5° 10
6.3 Ensemble ash detection maps 17
6.4 Comparison with lidar curtain plots 19
6.5 Pair-wise comparison at instrument resolution 22
7 Expert scene analysis 23
8 Comparison with active sensors and geometric height determination 26
9 Concluding remarks 30
10 References 33

2 ACRONYMS USED IN THIS DOCUMENT

Table 1: List of acronyms used in this report

CALIOP	Cloud-Aerosol Lidar with Orthogonal Polarization
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation
RAL	Rutherford Appleton Laboratory
RGB	Red-Green-Blue
RL	Lower Right (image coordinate)
SEVIRI	Spinning Enhanced Visible/Infrared Imager
UL	Upper Left (image coordinate)

3 INTRODUCTION

This document provides detailed findings of the study "Satellite Derived Volcanic Ash Product Inter-Comparison in Support to SCOPE-Nowcasting" (undertaken in response to the EUMETSAT RFQ 14/210178), under the auspices of the WMO SCOPE-Nowcasting Pilot Project 2: "Globally consistent Volcanic Ash Products". The overall aims of the SCOPE-Nowcasting activity were:

1. Using pre-selected cases, quantify the differences between satellite-derived volcanic ash cloud properties derived from different techniques and sensors.
2. Establish basic validation protocol for satellite-derived volcanic ash cloud properties.
3. Document the strengths and weaknesses of different remote sensing approaches as a function of satellite sensor.
4. Standardize the units and quality flags associated with volcanic cloud geophysical parameters.
5. Provide recommendations to Volcanic Ash Advisory Centers (VAACs) and other users on how to best to utilize quantitative satellite products in operations.
6. Create a "road map" for future volcanic ash related scientific developments and inter-comparison/validation activities that can also be applied to SO_2 clouds and emergent volcanic clouds.

As described in the statement of work (RD-2), the aim of the study reported here was to perform the inter-comparison work needed to support the overall aims above. The report therefore addresses points 1, 2 and 4, with some initial findings on 3. This work was carried out to provide results for discussion at the "Intercomparison of Satellite-based Volcanic Ash Retrieval Algorithms with WMO SCOPE-Nowcasting" workshop, which was held in Madison, WI, USA from 29 June – 2 July 2015. Findings of the workshop are given in RD-3.

4 BACKGROUND

The Intercomparison of Satellite-based Volcanic Ash Retrieval Algorithms with WMO SCOPE-Nowcasting activity (hence forth referred to the SCOPE-ash) is motivated by the need to ensure that high quality volcanic ash products are available to improve the ash advisories provided to aviation users. There has been significant evolution in the quantitative remote sensing of volcanic ash clouds by satellite over the past decade, and especially since the costly disruption to aviation caused by the Eyjafjallajökull eruption in 2010. There now exist a plethora of different ash products from a wide range of satellite sensors and employing diverse approaches to the characterisation of ash. Although most of these products have been individually assessed, and there have been some limited inter-comparison exercises, a wide ranging assessment of the available ash products had not been attempted. This work redresses this omission.

This study represents a first attempt to define standards for the geophysical parameters, and their representation, in satellite ash products and a validation approach for satellite volcanic ash products. The results, combined with the discussions of satellite retrieval experts and VAAC representatives at the workshop, will be used to help VAACs and other users better utilise satellite based ash products, with the aim of improving the accuracy of volcanic ash advisories.

The data format specification provided for SCOPE-ash was included in the project Work Plan, and was based on that developed for the EUMETSAT project "Development of OCA type processors to volcanic ash detection and retrieval" (RD-1). Products formatted according to this specification were uploaded directly to a password protected FTP repository supplied by RAL Space. These data were then processed through the inter-comparison software developed during this project, producing a large number of plots and summary statistics. These were, in turn, made available to the SCOPE-ash organizing committee and data contributors ahead of the Madision workshop, at which the inter-comparison was discussed in the context of the six objectives listed above.

5 STUDY CASES

The volcanic eruption study cases used in the inter-comparison exercise were defined in the Work Plan and are summarised in Table 2, while the satellite products submitted for the inter-comparison are listed in Table 3. The exact scenes to include in the study varied by satellite and instrument, depending on the spatial and temporal coverage provided by each. A summary of the number of data files provided for each day of each eruption case, for each product included in the inter-comparison is given in Figure 1.

Eruption	Date range	Comment
Eyjafjallajökull	2010-04-16 – 2010-05-18	
Grimsvötn	2011-05-21 – 2011-05-23	
Kelut	2014-02-13 – 2014-02-14	
Kirishimayama	2011-01-27	
Puyehue-Cordon Caulle	2011-06-05 – 2011-06-18	
Sarychev-Peak	2009-06-15 – 2009-06-17	

Table 2: Eruption cases used in this study

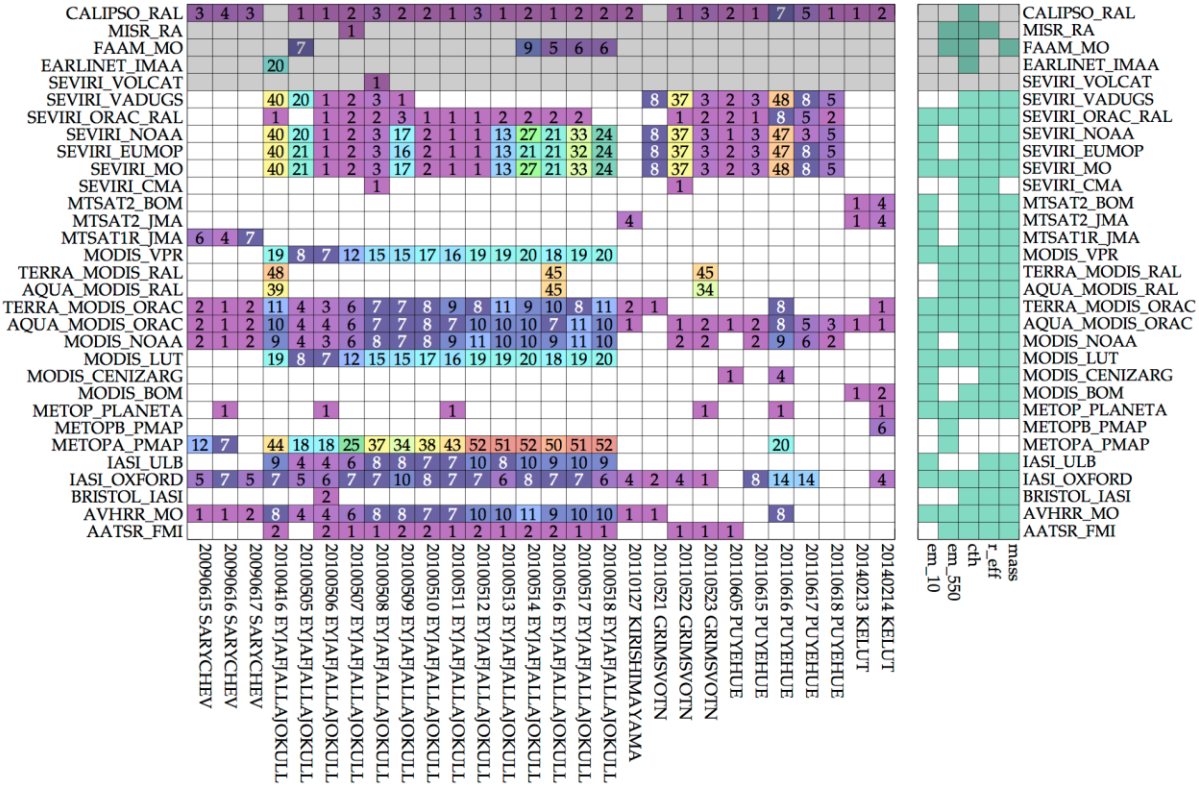


Figure 1: Overview of the data files provided by comparison product for study case day. The green boxes on the right indicate which ash properties are provided by each product. The products tinted grey are those considered to be validation data for the study. Figures in each box give the number of files provided in each case (and the colour-code reflects this). It should be noted that the number of files should not be taken as an indicator of the data volume or coverage as this depends also on the granularity of the products (chosen by the provider, usually following the granularity of the instrument L1 data). The statistic served during the project to cross-check that the correct number of files had been received and processed at RAL. The main purpose of the table here is to indicate which eruptions were covered by which sensor and which products are provided.

Product identifier	Source institution	Contact person	Comments
CALIPSO_RAL	NASA/RAL	Charles Trepte (NASA) charles.r.trepte@nasa.gov Richard Siddans richard.siddans@stfc.ac.uk	Regridded CALIOP level 1b attenuated backscatter. Used for height validation.
MISR_RA	NASA	Ralph Kahn ralph.kahn@nasa.gov Jim Limbacher jim.limbacher@nasa.gov	Stereo-parallax based ash height retrieval. Used for height validation.
FAAM_MO	UK Met Office	Franco Marenco franco.marenco@metoffice.co.uk	Extinction data from the Leosphere ALS450 lidar system on board the NERC FAAM aircraft. Used for height validation.
EARLINET_IMAA	Various	Gelsomina Pappalardo Gelsomina.pappalardo@imaa.cnr.it	Ground based lidar measurements. Used for height validation.
SEVIRI_VOLCAT	National Oceanic and Atmospheric Administration (NOAA)	Mike Pavolonis michael.pavolonis@nasa.gov	Expert classified SEVIRI scene. Used for evaluating ash detection.
AATSR_FMI	Finnish Meteorological Institute (FMI)	Timo Virtanen timo.h.virtanen@fmi.fi	Stereo-parallax ash height retrieval.
AVHRR_MO	UK Met Office	Pete Francis pete.francis@metoffice.gov.uk Mike Cooke michael.cooke@metoffice.gov.uk	
BRISTOL_IASI	University of Bristol	Luke Western luke.western@bristol.ac.uk	
IASI_OXFORD	University of Oxford	Lucy Ventress ventress@atm.ox.ac.uk	
IASI_ULB	Université Libre de Brusselles (ULB)	Lieven Clarisse lieven.clarisse@ulb.ac.be	
METOPA_PMAP	EUMETSAT	Ruediger Lang ruediger.lang@eumetsat.int	Combined GOME-2/AVHRR product.
METOPB_PMAP	EUMETSAT	Ruediger Lang ruediger.lang@eumetsat.int	Combined GOME-2/AVHRR product.
METOP_PLANETA	Russian State Research Center “Planeta”	Alex Rublev alex.rublev@mail.ru	AVHRR product.
MODIS_BOM	Australian Bureau of Meteorology (BOM)	Chris Lucas c.lucas@bom.gov.au	
MODIS_CENIZARG	Argentine National Commission for Space Activities (CONAE)	Guillermo Toyos gtoyos@conae.gov.ar	
MODIS_LUT	Italian Istituto Nazionale di Geofisica e Vulcanologia (INGV)	Stefano Corradini stefano.corradini@ingv.it Luca Merucci luca.meruci@ingv.it	
MODIS_NOAA	National Oceanic and Atmospheric Administration (NOAA)	Mike Pavolonis mike.pavolonis@noaa.gov	

Product identifier	Source institution	Contact person	Comments
TERRA/AQUA_MODIS_ORAC	University of Oxford	Greg McGarragh g.mcgarragh1@physics.ox.ac.uk	
TERRA/AQUA_MODIS_RAL	RAL Space	Richard Siddans richard.siddans@stfc.ac.uk	
MODIS_VPR	Italian Istituto Nazionale di Geofisica e Vulcanologia (INGV)	Stefano Corradini stefano.corradini@ingv.it Luca Merucci luca.meruci@ingv.it	
MTSAT1R_JMA	Japanese Meteorological Agency (JMA)	Daisaku Uesawa d-uesawa@met.kishou.go.jp	
MTSAT2_JMA	Japanese Meteorological Agency (JMA)	Daisaku Uesawa d-uesawa@met.kishou.go.jp	
MTSAT2_BOM	Australian Bureau of Meteorology (BOM)	Chris Lucas c.lucas@bom.gov.au	
SEVIRI_CMA	China Meteorological Administration (CMA)	Lin Zhu zhulin@cma.gov.cn	
SEVIRI_EUMOP	EUMETSAT	Hans-Joachim Lutz hansjoachim.lutz@eumetsat.int	
SEVIRI_MO	UK Met Office	Pete Francis pete.francis@metoffice.gov.uk Mike Cooke michael.cooke@metoffice.gov.uk	
SEVIRI_NOAA	National Oceanic and Atmospheric Administration (NOAA)	Mike Pavolonis mike.pavolonis@noaa.gov	
SEVIRI_ORAC_RAL	RAL Space	Richard Siddans richard.siddans@stfc.ac.uk Gareth Thomas gareth.thomas@stfc.ac.uk	
SEVIRI_VADUGS	Deutsches Zentrum für Luft- and Raumfahrt (DLR)	Kaspar Graf kaspar.graf@dlr.de	

Table 3: Key to data products included in this study.

6 INTERCOMPARISON APPROACH

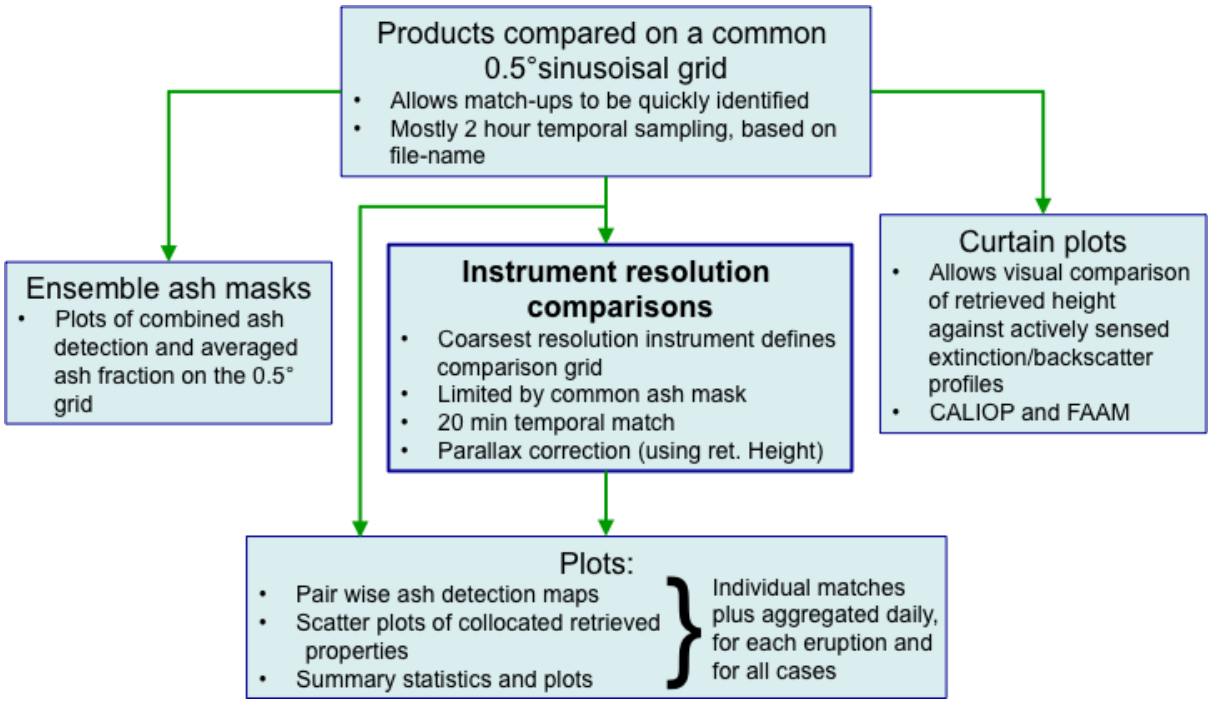


Figure 2: Overview of the intercomparison methodology.

The SCOPE-ash study involved the comparison of products from instruments with a wide range of spatial resolutions and spectral sensitivity, as well as an equally wide range of algorithm approaches. The "fair" comparison of all of these products is thus not straightforward. The approach taken in the study was to use a hierarchy of comparisons at different spatial and temporal resolutions, starting with a "lowest common demoninator" 0.5° sinusoidal grid with a lower resolution than any included product, and working up to pixel-by-pixel comparisons at instrument resolution for products from the same sensor. The comparison methodology is summarised in Figure 1 and is given in more detail in the following sections.

All plots produced in the intercomparison are automatically organised into a file-structure, which is online via password protected FTP: ftp://ftp.rsg.rl.ac.uk/ with the user-name "scopeftp" and password "Sc0pe2015(Eve23)". Table 4 provides details of the locations of the plots described in the following sections.

Directory	Reference	Description
reprojected_pngs/cv0p1/no_parallax_0p5deg	Section 6.1	Maps of the regridded products on the 0.5° grid.
reprojected_pngs/cv0p1/inst_res	Section 6.1	Maps of the products regridded to the instrument resolution grids. There is one image of each scene for each instrument grid used (see Section 6.5)
matches_pngs		Contains all the pair-wise comparison plots in a series of sub-folders:

matches_pngs/mv0p1-filterSEVIRI_NOAA	Section 6.2	Contains comparison plots filtered so that only pixels flagged as ash by the SEVIRI-NOAA product are included.
matches_pngs/mv0p1-filterSEVIRI_NOAA-min_em=0p05	Section 6.2	Contains comparison plots filtered so that only pixels flagged as ash by the SEVIRI-NOAA product and for which the emissivity at 10 µm is greater than 0.05 are included.
matches_pngs/mv0p1-filterSEVIRI_NOAA-min_em=0p1	Section 6.2	Contains comparison plots filtered so that only pixels flagged as ash by the SEVIRI-NOAA product and for which the emissivity at 10 µm is greater than 0.1 are included.
matches_pngs/mv0p1-min_em=0p05	Section 6.2	Contains comparison plots filtered so that only pixels for which the emissivity at 10 µm is greater than 0.05 are included.
matches_pngs/mv0p1-min_em=0p1	Section 6.2	Contains comparison plots filtered so that only pixels for which the emissivity at 10 µm is greater than 0.1 are included.
matches_pngs/*/no_parallax_0p5deg	Section 6.2	Sub-directory of each of the mv0p1 folders: contains the comparisons on the 0.5° sinusoidal grid.
matches_pngs/*/no_parallax_inst_res	Section 6.5	Sub-directory of each of the mv0p1 folders: contains the instrument resolution comparisons without parallax correction.
matches_pngs/*/inst_res	Section 6.5	Sub-directory of each of the mv0p1 folders: contains the instrument resolution comparisons including parallax correction.
ensemble_mask_pngs/cv0p1	Section 6.3	Contains the ensemble ash mask plots
calipso_curtain_pngs/mv0p1	Section 6.4	Contains the comparison plots against the CALIOP attenuated backscatter profile.
faam_curtain_pngs/mv0p1	Section 6.4	Contains the comparison plots against the FAAM aircraft lidar extinction profile.

Table 4: Overview of the directory structure used for the comparison plots.

6.1 PRODUCT REGRIDDING

The initial step in performing the inter-comparison is to pre-process all products, averaging each onto a 0.5° sinusoidal grid defined on an eruption-by-eruption basis. The grid for each eruption is defined such that its central point is located at the centre of the region defined for each eruption, which ensures that the grid cells are close to square boxes on the Earth's surface (the grid cells of a sinusoidal grid become increasingly skewed quadrilaterals on the surface as one approaches the edge of the grid).

This gridding was performed not only on the products evaluated in the study, but also on the validation data sets; namely CALIOP and FAAM lidar profiles, EARLINET ground based lidars and the expert classified SEVIRI scene. These products are thus included in the 0.5° and instrument resolution comparisons described below.

In order to minimise the influence of the different instrument resolutions and differences in the fraction of detected ash in each grid cell between products, the averaging of retrieved ash properties is weighted:

- Ash optical depth are converted to "emissivity", defined as:

$$\varepsilon_\lambda = 1.0 - \exp(-\tau_\lambda) \qquad 1$$

  where τ_λ is the optical depth at wavelength λ.

- Ash cloud-top height and effective radius are averaged weighted by the emissivity at 10 µm or, if the product does not include a 10 µm optical depth, the 550 nm emissivity. If optical depth is not defined at either wavelength, an unweighted mean is calculated.
- The unweighted ash column mass density is calculated, including pixels with no ash (i.e. zero mass).

6.2 PAIR-WISE COMPARISONS AT 0.5°

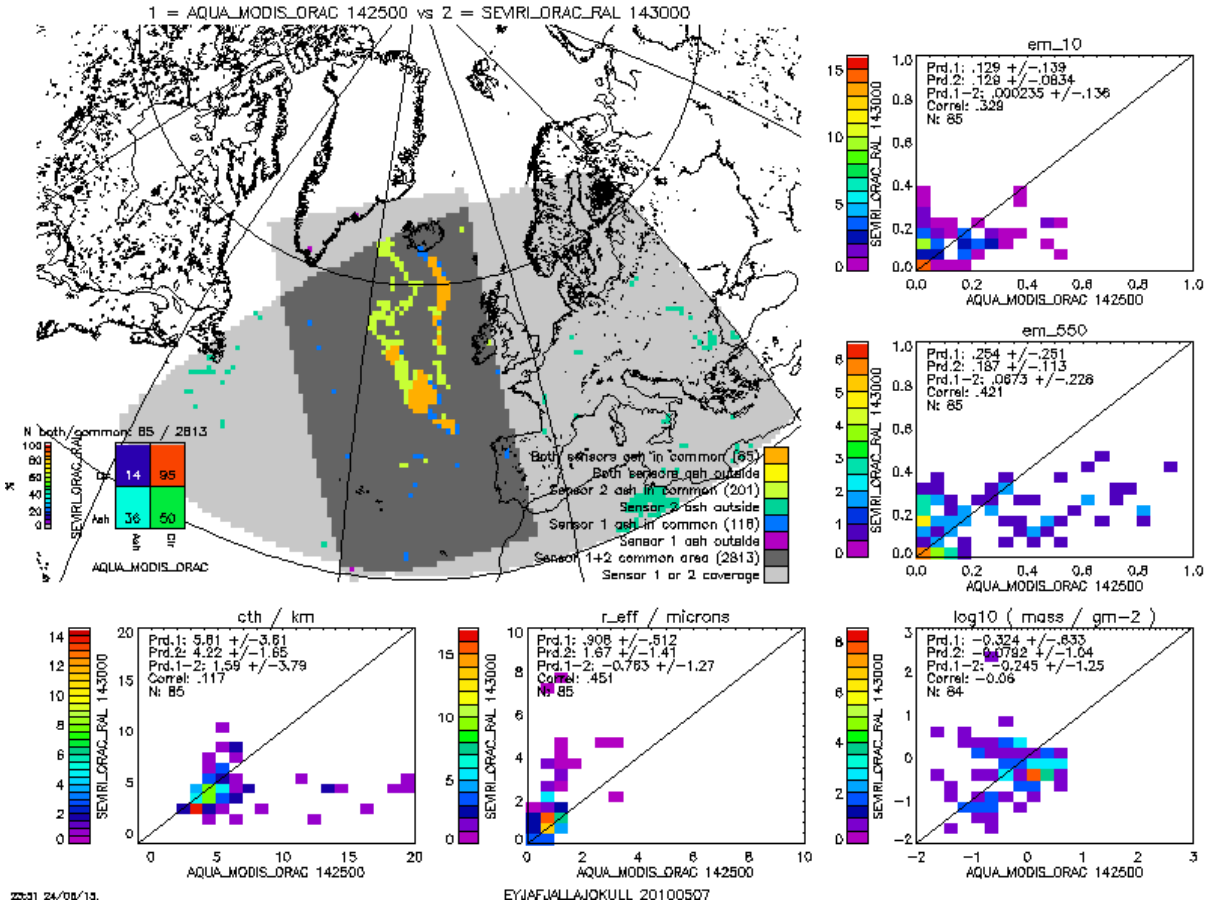


Figure 3: An example of a pair-wise comparison on the 0.5 grid, between the MODIS-AQUA ORAC product from University of Oxford and the SEVIRI ORAC product from RAL, for Eyjafjallajöjull at approximately 14:30 on 7 May 2010. In the map panel, the common area of the two products appears as dark grey, while grid cells containing ash in both products are coloured orange. Grid cells containing ash in the SEVIRI product within the overlap area but not detected as ash by the MODIS product are lime-green, while those detected by MODIS and not SEVIRI are blue.

Once regridded, each product pair was compared, using a ±1 hour temporal match criteria based on the time specified in the product file name (i.e. temporal matching of 0.5° gridded products did not require each file to be read, just a list of file names). From these matches, a series of plots were generated (see Figure 3):

- Pair-wise detection maps, and associated confusion matrices (top left panel of Figure 3), showing where each product pair agrees/disagrees on the presence of ash and the number of pixels:
  - Where both products have detected ash
  - Where both products are present, but only one has detected ash
  - Where both products are present and neither has detected ash
  - Where only one product is present and has, or has not, detected ash
- Scatter density plots of retrieved ash properties (for grid cells where both products detect ash): ash emissivity (as defined above) at 550 nm and 10 µm, ash cloud top height, ash effective radius and column ash mass density. Each of these includes associated statistics: mean and standard deviation of each product, mean and standard deviation of the pixel-wise difference between each product, and Pearson correlation coefficient of the two products.

The confusion matrices produced for each product pair provide a quantitative visual indication of the level agreement in ash detection as summarised in Table 5.

% of grid points where: $\left[\frac{\text{Sensor "Y" says clear}}{\text{Either sensor detects ash}}\right]$	% of grid points where: $\left[\frac{\text{Both sensors say clear}}{\text{Either sensor detects ash}}\right]$
% of grid points where: $\left[\frac{\text{Both sensors detect ash}}{\text{Either sensor detects ash}}\right]$	% of grid points where: $\left[\frac{\text{Sensor "X" says clear}}{\text{Either sensor detects ash}}\right]$

Ideally	
0	100
100	0

Table 5: Description of ash detection confusion matrices, with an example of the ideal confusion matrix.

In addition to plotting each individual temporal matchup, equivalent plots as also produced for aggregated matches on a daily basis, as well as across all matches found for a particular eruption. In addition, maps of the aggregated retrieval products for common pixels between the product pair are also plotted on a daily basis and for the whole eruption, as shown in Figure 4.

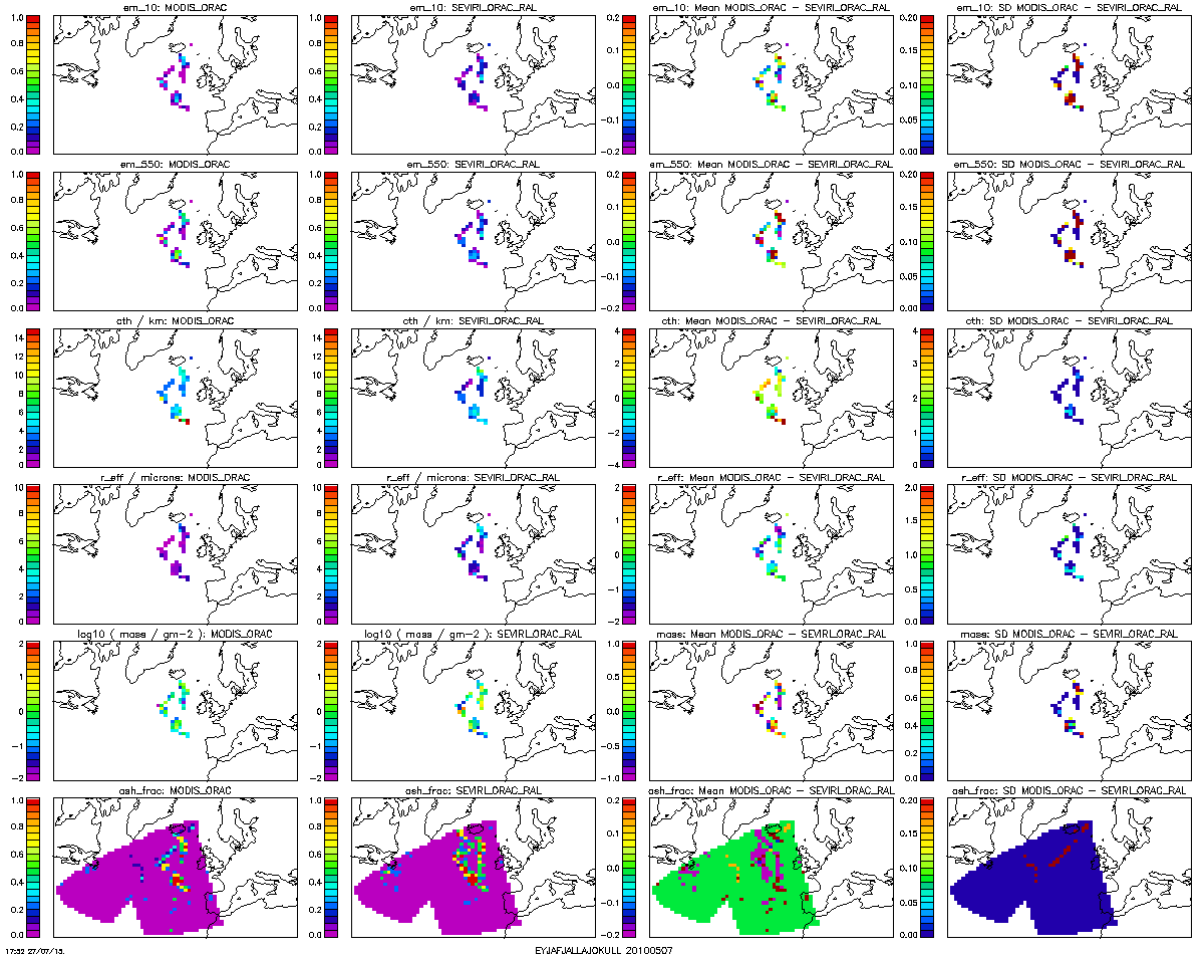


Figure 4: The daily summary maps for the same product pair as shown in Figure 3, on the same date.

Finally, the individual comparisons were combined into summary tables including all products for each eruption, both on a daily basis and for the entire eruption case. These tables include:

- Confusion matrices for all product pairs
- Detection maps
- Scatter plots of each retrieval parameter (ash cloud-top height, emissivity at 550 nm and 10 µm, effective radius and column mass density), along with the PDF of each parameter derived from each individual product, as shown in Figure 5.
- Matrix plots of the comparison statistics for each product pair (products X and Y) for ash detection (Figure 6):
    - Percentage miss-matched detection: number of ash pixels only detected in X / Number of ash pixels in X or Y.
    - Percentage consistent detection: number of ash pixels in X and Y / Number of ash pixels in X or Y (percentage)
    - The common ash cloud area detected by X and (X or Y)
    - The common ash cloud area detection by (X and Y) and (X or Y)

- Matrix plots of the comparison statistics for each product pair (products X and Y) for the retrieved parameters (Figure 1):
    - The mean difference (X – Y)
    - Standard deviation of the difference
    - Pearson correlation
    - Number of matching points

In addition to the basic comparison, where each product is compared as-is, comparisons have also been performed with additional constraints applied to the data:

- The SEVIRI product provided by NOAA has been used as a master ash flag, so that only pixels which are flagged as ash in this product are included in the comparisons. The NOAA SEVIRI ash detection has been found to generally detect a larger extent of contiguous ash clouds than the other products included in the study, while also not displaying the scattered, isolated false-positive ash detections apparent in most hi-sensitivity ash detection schemes. Thus, using this product to limit the ash areas included in other products reduces the obvious false positives included in the comparisons, while minimising the exclusion of true ash pixels. (Obviously this constraint also limits cases to those within the SEVIRI field of view.)
- The data has been filtered by setting minimum value thresholds on the ash emissivity at 10 µm. This excludes optically thin ash from the comparisons, which can be expected to provide poorly constrained retrievals of ash properties, with sensitivity to underlying water cloud being a particular example. Emissivity thresholds of 0.1 and 0.05 have been used (which essentially correspond one-to-one with the 10 µm optical depth at these low values).

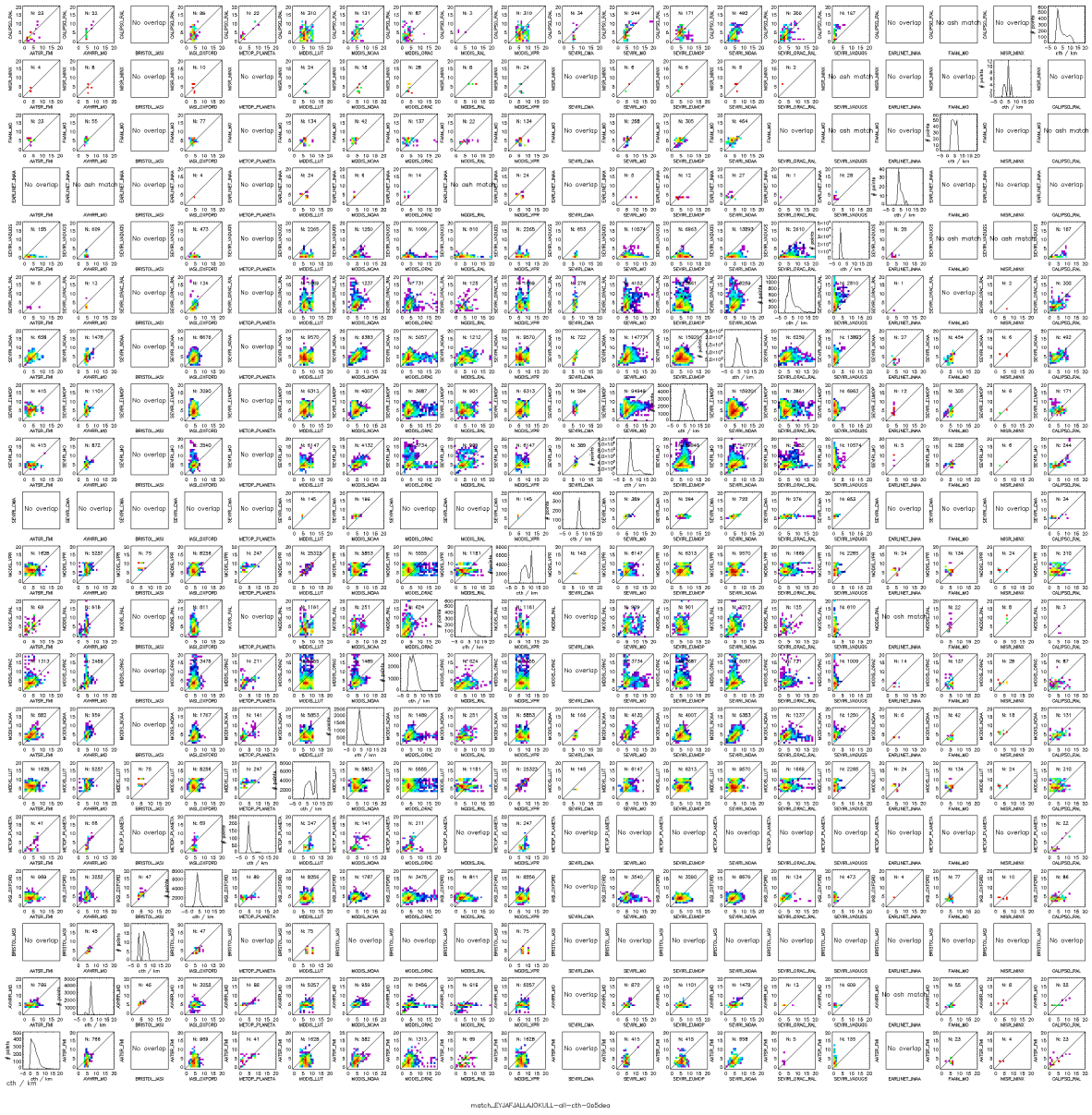


Figure 5: Summary plot table of ash cloud-top height comparisons for the Eyjafjallajökull eruption. Each panel shows the scatter plot for an product pair, with the PDF of ash cloud-top height for each product appearing along the diagonal.

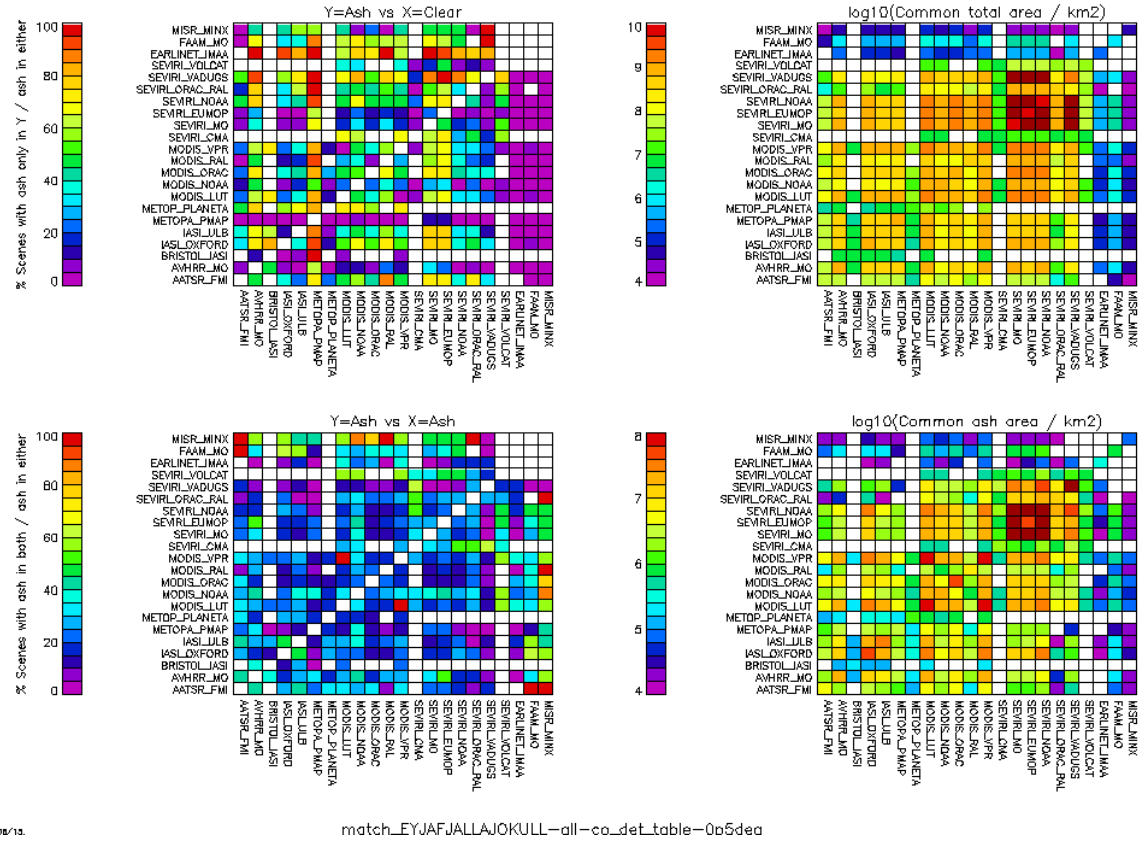


Figure 6: Statistics of the ash detection comparison for the Eyjafjallajökull eruption on the 0.5° grid. Clockwise from the top-left the panels show the percentage miss-matched ash detection between product pairs; the common area detected in product Y and (Y or X); the percentage consistent detection; the common area detected by products (Y and X) and (Y or X).

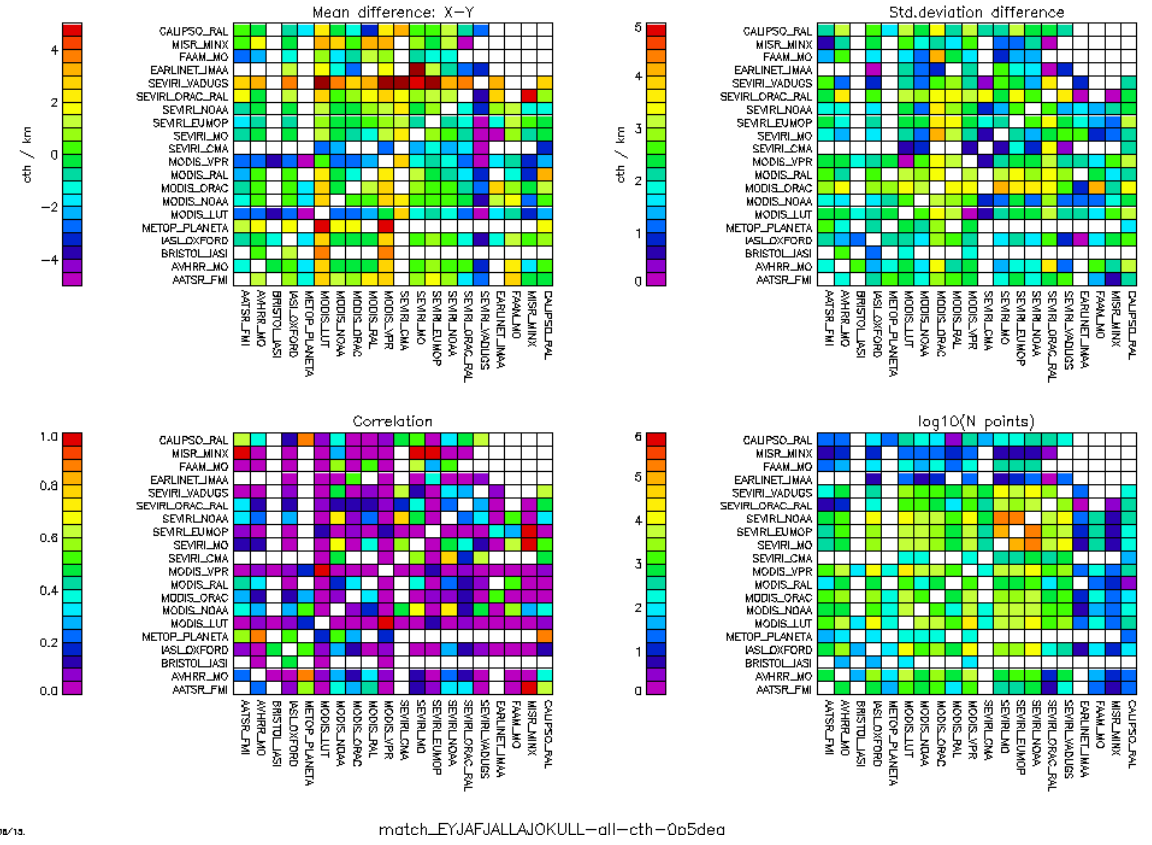


Figure 7: Summary statistics of the ash cloud-top height comparisons for the Eyjafjallajkull eruption, on the 0.5° grid.

6.3 ENSEMBLE ASH DETECTION MAPS

Using the matches defined above, ensemble ash detection maps were also produced, as shown in Figure 8. These plots show the number of products which provide data for each 0.5° grid cell and how many detected ash over a two hour window, as well as the average ash fraction (i.e. fraction of instrument pixels detected as ash within each 0.5°).

It was hoped that ensemble masks could be used to provide an additional constraint to the pair-wise comparisons of products, by limiting the pixels compared to the area where a majority of products agreed that ash was present. The complexity of this approach, due to the large variability in coverage by different instruments within a scene and the prevalence of false detections (over desert regions, for instance), means that it was not used in the pair-wise comparisons, with the SEVIRI-NOAA product playing a similar role instead.

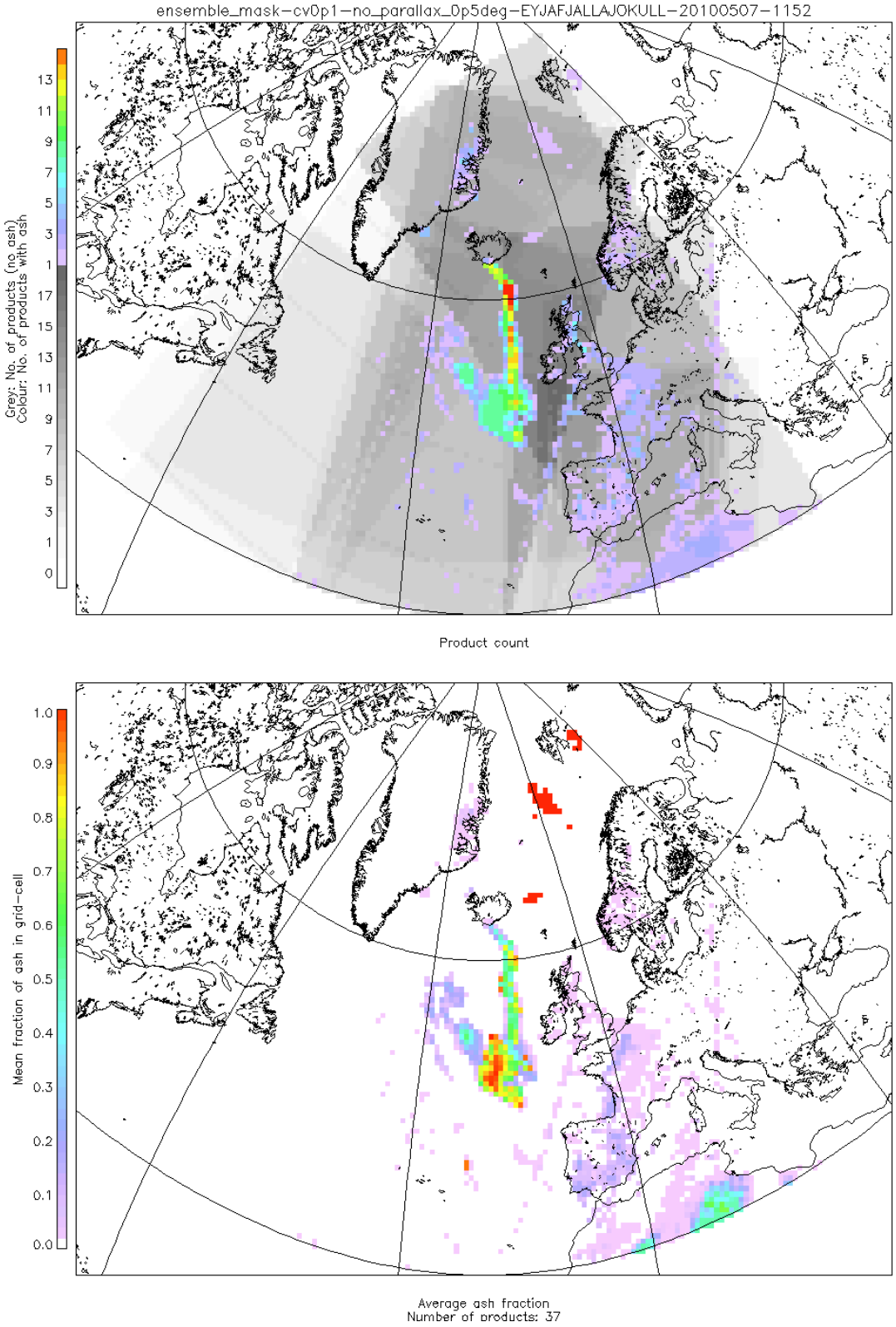


Figure 8: Ensemble ash mask for the Eyjafjallajkull eruption at approximately midday, 7 May 2010.

6.4 COMPARISON WITH LIDAR CURTAIN PLOTS

Again, using matches defined from the 0.5° gridded products, so-called curtain plots are created for each CALIOP and FAAM lidar matchup, as shown in Figure 9 and Figure 10 respectively. These plots provide a detailed view of a product in the region sampled by the lidar and are generated directly from the original product files (not from the regridded data). For each matchup the retrieved ash cloud-top height, collocated with the lidar track, is over-plotted on the lidar attenuated backscatter (from CALIOP) or extinction (from FAAM) profile along with its associated backscatter. In addition the satellite retrieval products in the region of the lidar measurements are also plotted. Due to the differences in the CALIOP and FAAM measurements, there are differences in how these plots are generated in each case:

- In the case of CALIOP, data was extracted along a 100 km wide swath, centred on the CALIPSO track and the following products were displayed (if available):
  - The 11-12 µm brightness temperature difference
  - The ash detection mask
  - The ash cloud-top height
  - The ash optical depth at 10 µm and 550 nm
  - The ash effective radius
  - The ash column mass density
  - In addition, the 8.7, 11 and 12 µm false-colour image provided by the Imaging Infrared Radiometer (IIR) on board CALIPSO is also plotted.
- In the case of FAAM, the aircraft track was broken into 15-minute segments, which were matched against the satellite products individually. Satellite data in a region centred on the FAAM track, with a 2 degree lat-lon margin[*], is plotted with the FAAM measurements over-plotted:
  - The ash detection mask, plotted on a map to provide geolocation for the scene
  - 11-12 µm brightness temperature difference (with FAAM ash detection over-plotted)
  - The ash cloud-top height (with the FAAM cloud-top height estimate over-plotted)
  - The ash optical depth at 10 µm and 550 nm (with the 550 nm optical depth estimated from the FAAM extinction over-plotted)
  - The ash effective radius (with FAAM ash detection over-plotted)
  - The ash column mass density (with mass density estimated from the FAAM extinction over-plotted)

  Note that, aside from the ash detection mask plot, all of the satellite imagery is plotted on the native grid supplied by the data product itself.

[*] Note that, unlike the CALIPSO orbit track, the FAAM tracks contain frequent changes of direction and sampling as the aircraft changed direction, speed and altitude.

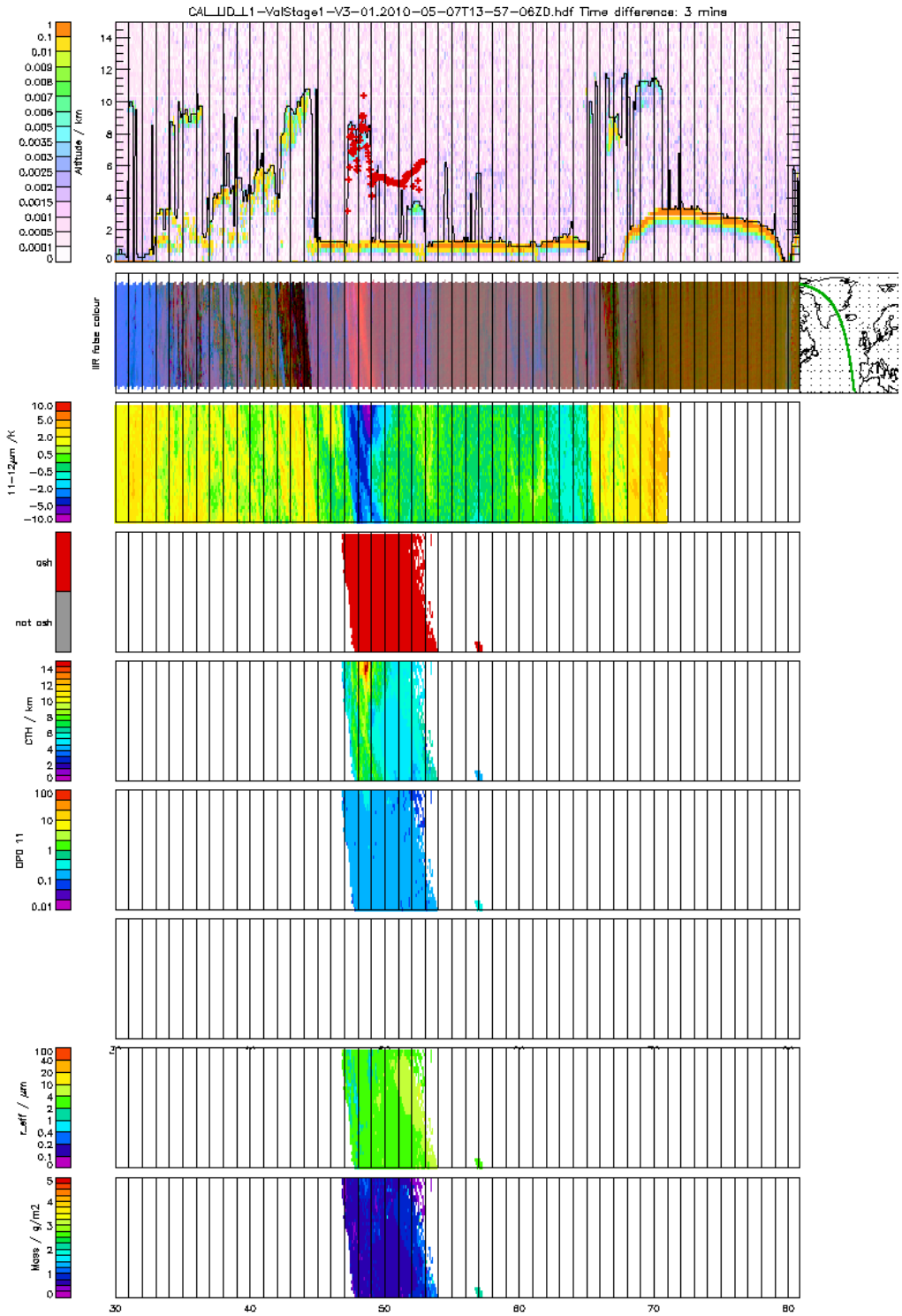


Figure 9: CALIOP curtain plot comparison of the NOAA SEVIRI product for the Eyjafjallajökull eruption at 14:30, 7 May 2010. Below the attenuated backscatter curtain plot the 8.7, 11, 12 µm false colour imagery from the IIR imager, the 11-12 µm BTD from SEVIRI, the NOAA SEVIRI ash mask, ash cloud-top height, 11 µm optical depth, a place holder for the missing 550 nm optical depth, ash effective radius and column mass loading.

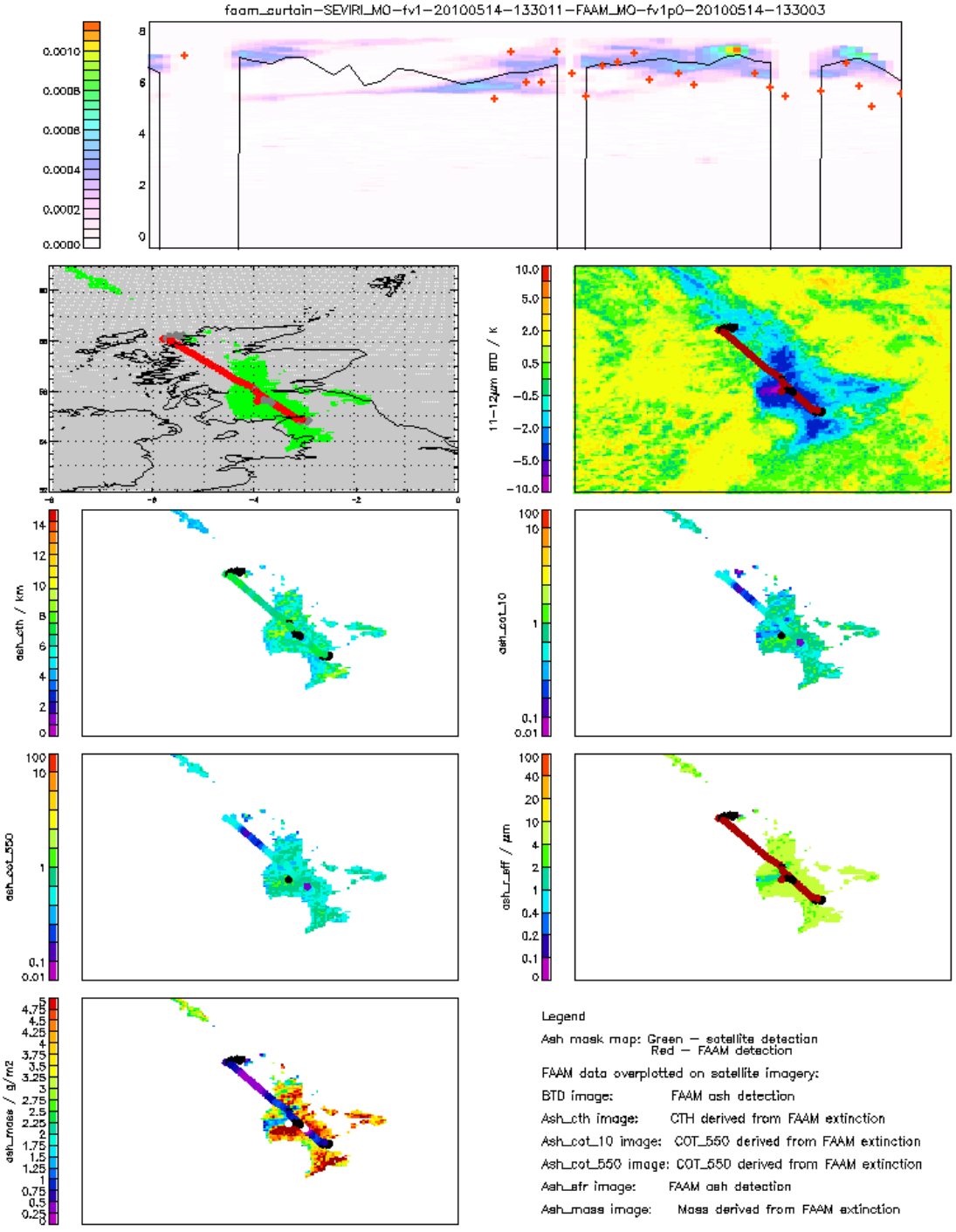


Figure 10: FAAM curtain comparison of the Met Office SEVIRI product for Eyjafjallajökull at 13:30, 14 April 2010. In the top curtain plot, the red crosses are CTH from the satellite retrieval and the black line is height derived from the FAAM measurements. Below the extinction curtain plot are (left-right, top-bottom): a map of the SEVIRI ash mask and imagery for the 11-12 μm BTD, ash cloud-top height, 11 μm optical depth, 550 nm optical depth, effective radius and column mass density. In each case the FAAM measurements are over-plotted, either with the equivalent data (where available) or with a simple measurement flag.

6.5 PAIR-WISE COMPARISON AT INSTRUMENT RESOLUTION

Finally, the 0.5° degree matched data were used as the basis to produce matches at close to the full instrument resolution, using the same methodology as described in point 2 above, but limiting matches to ±10 minutes. Each match-up was performed at the spatial resolution of the lowest resolution instrument in the pair, so that instrument resolution comparisons actually comprise a hierarchy of resolutions:

- If a match included the EUMETSAT PMAP product – which is on the GOME-2 instrument grid – then the native GOME-2 grid was used (rectangular scenes on the ground).
- If PMAP wasn't included, but an IASI product was, then the native IASI grid was used (12 km circular FOV on the ground).
- If neither of the above instruments were included, but a geostationary imager product (from either SEVIRI or MTSAT-1R or -2), then the appropriate geostationary grid is used (e.g. 3 km resolution at nadir for SEVIRI).
- If two different polar orbiting imagers (e.g. AATSR, AVHRR, MISR or MODIS) – or one of these instruments and an active sensor – formed the match, an approximately 4 km sinusoidal grid was used.
- If two products from the same polar orbiting imager formed the match, an approximately 1 km sinusoidal grid was used.

Note that the two final resolutions do not correspond directly to the measurement grids of the instruments involved. The 4 km sinusoidal grid was chosen for comparison of different polar imagers as it should minimise the sampling differences between the different instruments, while still provide reasonable spatial resolution. For comparing products from the same polar imager, a 1 km sinusoidal grid will ensure that individual instrument pixels are compared for the instruments included in the study, while allowing common gridding and mapping software to be used.

The plots produced from the instrument resolution comparisons mirrored those described in section 6.2, as shown in Figure 11, including the summary plots and statistics tables. In addition to the comparisons using the geolocation information provided by each product, the analysis has also been performed on parallax corrected data for imager data (unless comparing observations from the same platform), using the retrieved ash cloud-top height and instrument viewing geometry to calculate the nominal position of the ash if it were viewed vertically.

Note that generation of parallax corrected comparisons was complicated by the lack of pixel specific time and/or sub-satellite location in the data specification – this prevented the spectrometer products (PMAP and IASI products) from having parallax correction applied, as the viewing geometry could not be uniquely determined from the data products[†]. These parameters should be included future similar studies.

[†] In the case of imager based products, the satellite location is either fixed (for geostationary products) or can be assumed to lie at the centre of the imager swath.

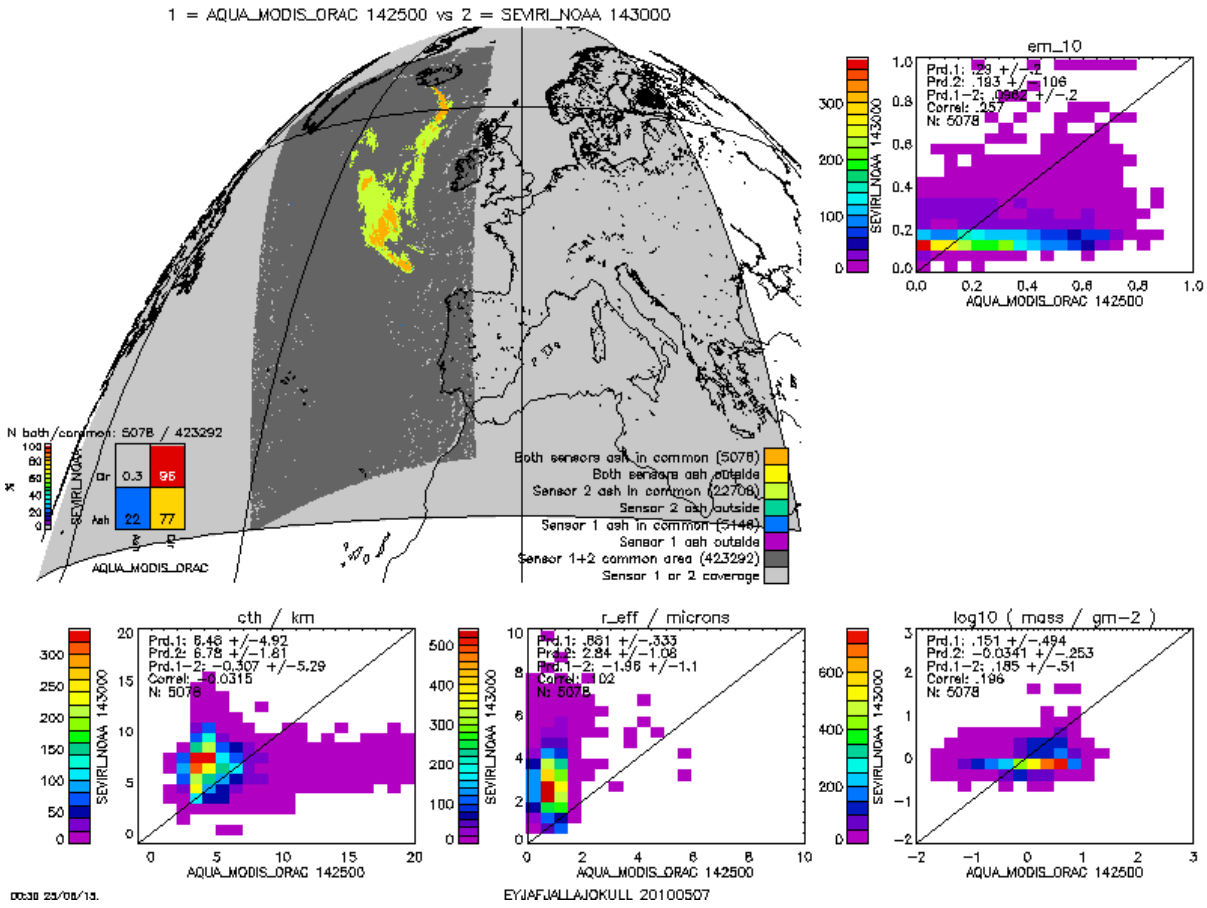


Figure 11: Instrument resolution (SEVIRI-pixel) comparison of MODIS-Aqua ORAC product from the University of Oxford and the SEVIRI product from NOAA at 14:30, 7 May 2010. This can be compared to the 0.5° grid plot shown in Figure 3. Note also that the NOAA product does not provide a ash optical depth at 550 nm, so the corresponding plot is missing.

7 EXPERT SCENE ANALYSIS

For this study a single SEVIRI scene, shown in false colour in Figure 12, from the Eyjafjallajökull eruption (SEVIRI reference time of 2010/05/08 04:12:41) was provided with a human expert ash detection as a reference ash mask, displayed in Figure 13. This scene is co-located with SEVIRI ash products and with an descending (nighttime) overpass of MODIS-Aqua[‡], as well as a CALIPSO overpass. This section provides a summary of the comparisons against this reference mask.

The confusion matrices for the ash detection for each algorithm are presented in Table 6. These data have produced using the SEVIRI instrument resolution, but been filtered so that only data which provide an ash fraction of greater than 1% on a 1° lat-lon grid are included. This filter removes many scattered false-positives, which are a common feature of many products, so that only contiguous regions of ash are included in the comparison, thus providing values which better represent how well the main ash plume defined in the expert analysis was captured by each product.

[‡] Thus, only thermal-only MODIS products provide a match to this scene.

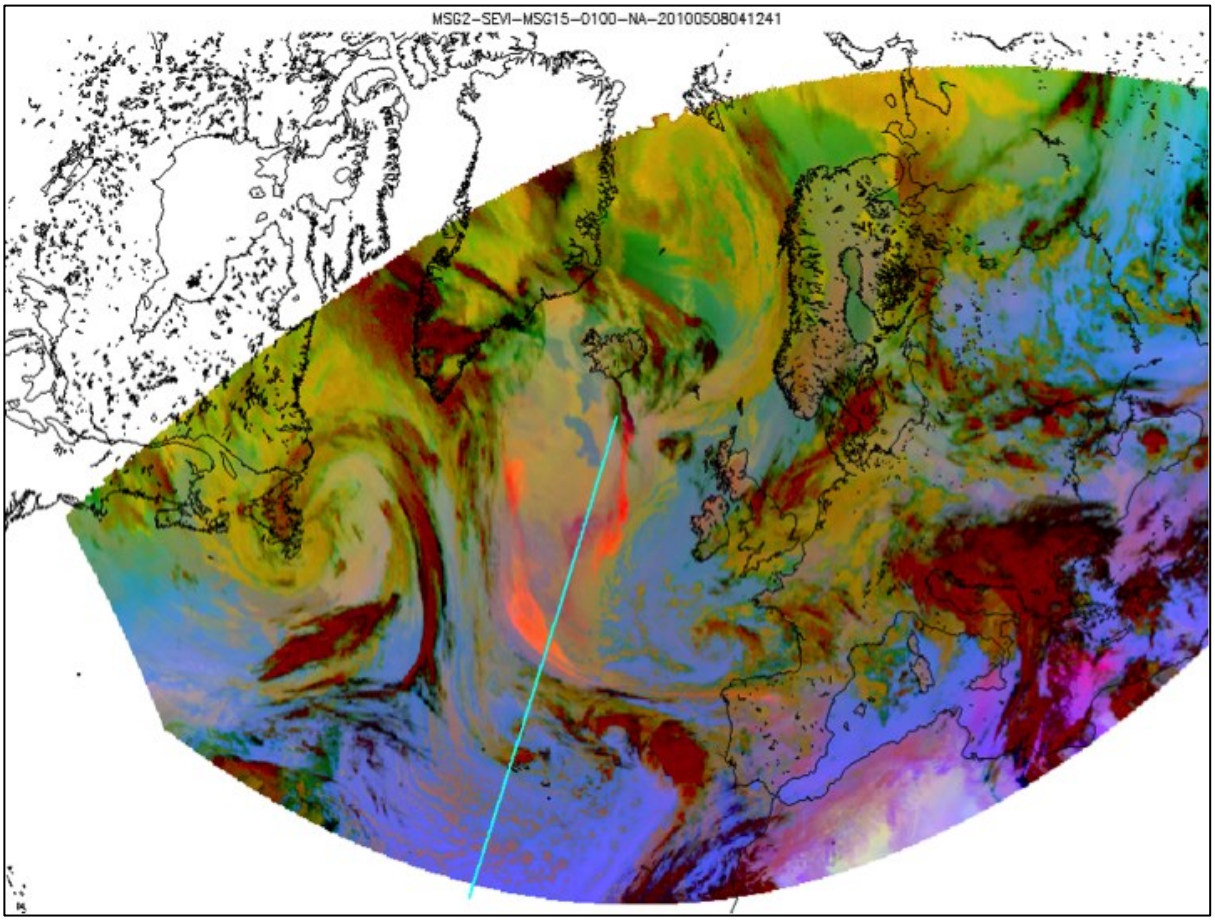


Figure 12: 8.7, 11, 12 µm false colour image of the VOLCAT expert analysis scene, with the corresponding CALIPSO overpass track indicated.

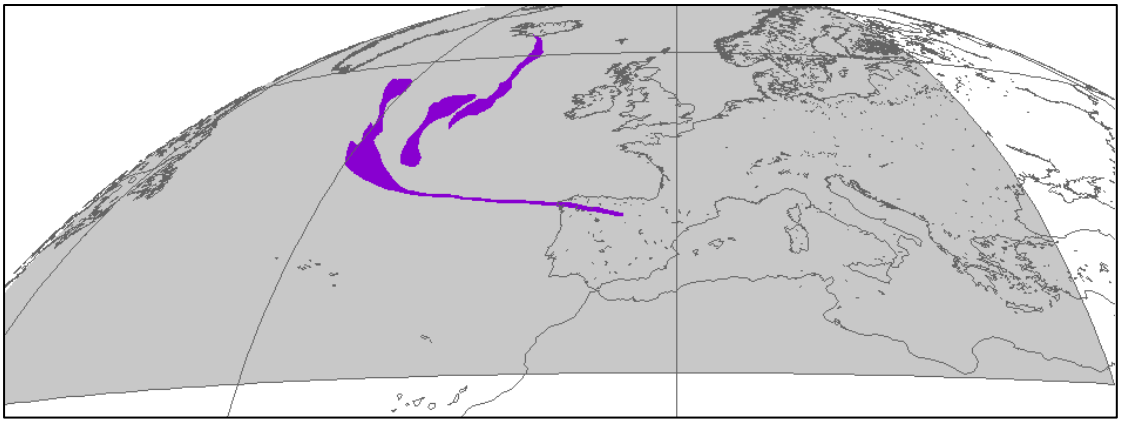


Figure 13: The ash cloud defined in the VOLCAT expert analysis scene.

Summarising these results, we can see that:

- The MODIS products are generally more conservative than the SEVIRI products and under-estimate the area of ash compared to the VOLCAT scene – in all three cases, the proportion of cases where VOLCAT classes a pixel as clear, while the MODIS product says it is ash is low (top left-hand box of the confusion matrices)
- The MODIS products are all missing approximately 30-35% of the expertly identified ash pixels (lower right confusion matrix box).
- The SEVIRI products generally detect more ash than is found in the VOLCAT scene, with the exception of the SEVIRI_EUMOP product, and detect a higher fraction of the ash cloud defined by VOLCAT than the MODIS products. In most cases, the extra ash detected in the SEVIRI products is due to false detection of dust over North Africa or the Mediterranean; the exceptions to this are the SEVIRI_NOAA product, in which the ash cloud is co-located with that from VOLCAT but has a larger extent, and the SEVIRI_CMA product, which detects additional ash near Iceland.
- The SEVIRI_EUMOP, SEVIRI_MO and SEVIRI_VADUGS products all provide quite consistent and, compared to the VOLCAT product, conservative detections of the ash plume, but the EUMOP product displays far fewer false detections than the other two products.

VOLCAT	3.2	96		5.4	96		3.2	96
	60	37		60	35		60	37
	MODIS_LUT			MODIS_NOAA			MODIS_VPR	
VOLCAT	50	82		48	95		4.6	98
	39	11		33	19		43	53
	SEVIRI_CMA			SEVIRI_MO			SEVIRI_EUMOP	
VOLCAT	43	97		39	96		68	91
	48	9.6		37	23		24	8.1
	SEVIRI_NOAA			SEVIRI_ORAC_RAL			SEVIRI_VADUGS	

Table 6: Confusion matrices for the MODIS and SEVIRI product comparisons against the VOLCAT expert analysis scene. Only pixels which contain greater than 1% of ash within a 1° box have been included.

It is important to note that sampling and coverage has undoubtedly played a role in these results (as is a common theme in this study) as the products are not consistent with each other. In particular:

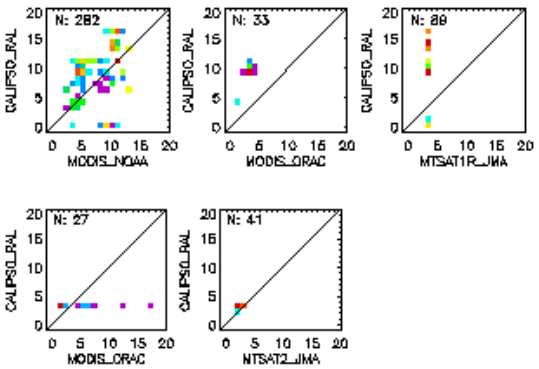
- The MODIS scenes do not include the Mediterranean or North Africa, so any possible confusion of ash and desert dust in these products does not effect their performance like it does the SEVIRI products.
- Similarly, the SEVIRI_CMA product is for a considerably smaller spatial area than the full scene as shown in Figure 12 and used by the other SEVIRI products, and also doesn't include the Mediterranean/North Africa area.
- The SEVIRI_ORAC_RAL product uses a sub-sample of every 4th SEVIRI pixel in both north-south and east-west directions (i.e. only 1 in 16 pixels is actually processed), which greatly reduces the sample size when calculating the confusion matrices.

8 COMPARISON WITH ACTIVE SENSORS AND GEOMETRIC HEIGHT DETERMINATION

Three active sensor datasets were available for this study, as well as one well established geometrical height retrieval scheme:

1. The CALIOP lidar on board the CALIPSO satellite in the A-train. Level 1b attenuated backscatter profiles were compared against, using the same methodology as previous studies (Thomas and Siddans, 2015 and references within). The use of attenuated backscatter, rather than higher-level CALIOP aerosol and cloud products, ensures that the CALIOP data is free from its own retrieval artefacts.
2. The Leosphere ALS450 lidar system on board the NERC FAAM aircraft. Ash extinction profiles of Eyjafjallajökull ash over the UK, derived by the UK Met Office, were compared, using a similar methodology to that used for the CALIOP backscatter data.
3. Measurements from ground based lidars in the European Aerosol Research Lidar Network (EARLIENT), which provided some limited detections of ash height from the Eyjafjallajökull eruption.
4. The MISR stereo ash height retrieval from JPL. The multi-view parallax based height estimation provided by MISR is expected to provide a more robust height estimate than the thermal emission based methods used in most of the other passively sensed products in this study[§].

These products provide the closest data available to "ground-truth" on ash cloud height for this study. In practice, the vast majority of the matchups between the passively sensed data and the validation data occur with CALIOP, which is the only one of the above products that provides global data, and even this is quite sparse. Scatter plots of ash cloud-top height verses each of the validation products are shown for each of the eruption cases in Figure 14, Figure 15, Figure 16 and Figure 17 respectively, while summary statistics are given in Table 7 to Table 10**Error! Reference source not found.**.



[§] The study also contains a stereo ash height retrieval from FMI using the AATSR instrument. However, this product is still considered to be at a relatively early stage of development and has thus not been treated as a validation product.

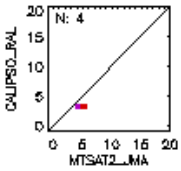


Figure 14: Scatter plots of retrieved ash cloud-top height against estimated height from the CALIOP lidar for the Sarychev (1st row), Kirishimayama (2nd row) and Kelut (3rd row) eruptions.

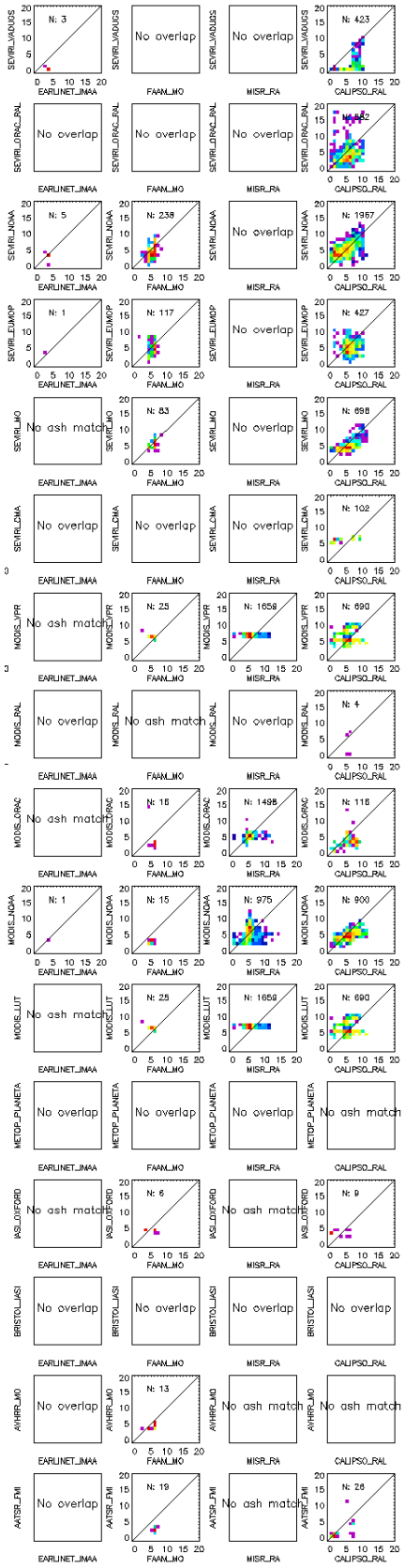


Figure 15: Scatter plots of retrieved ash cloud-top height against estimated height from the four validation datasets for the Eyjafjallajökull eruption. From left to right: EARLINET lidar, FAAM aircraft lidar, MISR stereo height, CALIOP orbital lidar.

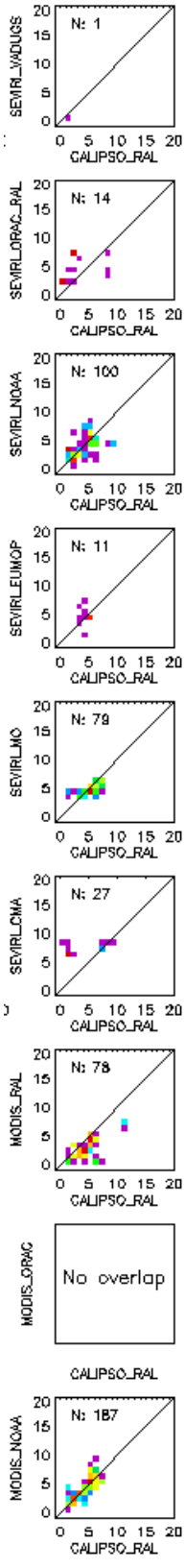


Figure 16: Scatter plots of retrieved ash cloud-top height against estimated height from the CALIOP lidar for the Grimsvötn eruption.

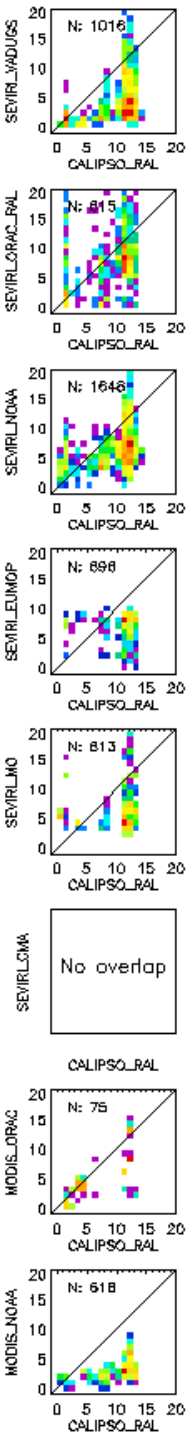


Figure 17: Scatter plots of retrieved ash cloud-top height against estimated height from the CALIOP lidar for the Puyehue eruption.

One aspect of these comparisons, which makes evaluation of the products difficult, is the paucity of matches for many of the instruments. This is particularly true for the EARLIENT data, where we are relying on coincidence of detectable ash without underlying cloud with a satellite overpass over a small number of sites which are far from the volcano itself. Similarly, the low coverage of the FAAM measurements also result in small sample sizes.

In the case of comparison with MISR and CALIOP, the density of matches depends strongly on the instrument being compared. For the geostationary platforms (where we have continuous coverage) and the MODIS instrument (MODIS-Aqua is part of the A-train formation along with CALIPSO, while MODIS-Terra is on the same platform as MISR), a large number of collocated pixels are available; for MetOp or ENVISAT based instruments far fewer matches are available. Finally, it should be noted that aside from the CALIOP product, all of the height validation data is specific to the Eyjafjallajökull eruption.

In general, simpler schemes tend to provide lower height correlations than more ambitious schemes, but show very little difference in RMS and mean/standard deviation difference; for instance the MODIS_LUT, MODIS_VPR and SEVIRI_EUMOP schemes versus the NOAA and ORAC based products. This can be attributed to the simpler schemes providing more self-consistent results – i.e. a fairly constant height is often retrieved across an individual image of a given cloud – but are more likely to produce spurious results. The more complex schemes, which apply a more comprehensive forward model to match the observations, are more robust, but provide much noisier results.

It is also obvious from the CALIOP comparisons in Figure 15 and Figure 17 (for the Eyjafjallajökull and Puyehue eruptions; the two which provide the most collocations) that the quality of retrieved ash height (and, by inference, the other retrieved parameters) varies from eruption to eruption. For example the SEVIRI_NOAA ash cloud-top height product provides a correlation approaching 0.6 when compared against CALIOP for the Eyjafjallajökull eruption, which drops to under 0.4 for the Puyehue eruption. A similar pattern is seen for the other products which provide results from both eruptions. Although some of this difference could be due to the large areas of optically thin ash associated with the Puyehue eruption, it is also probably a reflection of the focus on Eyjafjallajökull in the development and characterisation of most satellite ash retrieval schemes and differences in the optical properties of the ash from the two eruptions.

Product	No. of matches	Correlation	Mean difference (km)	St. Dev. difference (km)	RMS (km)
AATSR_FMI	26	0.511	1.015	2.508	2.66
IASI_OXFORD	9	-0.221	0.111	2.75	2.595
MODIS_CENIZARG	56	0	12.236	2.803	12.547
MODIS_LUT	690	0.038	-0.452	2.872	2.906
MODIS_NOAA	1987	0.328	2.723	3.65	4.553
MODIS_ORAC	251	0.476	2.212	3.203	3.887
MODIS_RAL	82	0.597	1.671	1.68	2.362
MODIS_VPR	690	0.038	-0.451	2.873	2.906
MTSAT1R_JMA	89	0.131	7.472	4.869	8.903

MTSAT2_JMA	45	-0.111	0.356	0.76	0.832
SEVIRI_CMA	129	0.639	-2.822	2.964	4.084
SEVIRI_EUMOP	1134	0.3	2.566	3.498	4.337
SEVIRI_MO	1385	0.336	2.125	3.829	4.377
SEVIRI_NOAA	3715	0.536	0.922	3.552	3.669
SEVIRI_ORAC_RAL	1159	0.408	0.988	4.396	4.504
SEVIRI_VADUGS	1440	0.574	4.794	3.752	6.087

Table 7: Overall statistics of ash cloud-top height comparisons with the CALIOP lidar.

Product	No. of matches	Correlation	Mean difference (km)	St. Dev. difference (km)	RMS (km)
MODIS_NOAA	1	-	0	-	-
SEVIRI_EUMOP	1	-	-0.6	-	0.6
SEVIRI_NOAA	5	-0.582	0.6	1.838	1.75
SEVIRI_VADUGS	3	-0.883	2.4	1.217	2.597

Table 8: Overall statistics of ash cloud-top height comparisons with EARLIENT lidar profiles.

Product	No. of matches	Correlation	Mean difference (km)	St. Dev. difference (km)	RMS (km)
AATSR_FMI	19	0.526	3.937	0.447	3.961
AVHRR_MO	13	0.568	1.123	0.988	1.471
IASI_OXFORD	6	-0.585	1.7	1.792	2.359
MODIS_LUT	25	-0.722	-0.656	1.29	1.424
MODIS_NOAA	15	0.038	2.12	0.824	2.265
MODIS_ORAC	16	-0.563	2.6	3.388	4.186
MODIS_VPR	25	-0.722	-0.656	1.29	1.424
SEVIRI_EUMOP	117	0.064	1.084	2.056	2.316
SEVIRI_MO	83	0.381	0.677	1.16	1.337
SEVIRI_NOAA	238	0.477	0.908	1.721	1.942

Table 9: Overall statistics of ash cloud-top height comparisons with FAAM aircraft measurements.

Product	No. of matches	Correlation	Mean difference (km)	St. Dev. difference (km)	RMS (km)
MODIS_LUT	1659	-0.06	-2.042	1.067	2.304
MODIS_NOAA	975	0.229	-0.879	2.085	2.261
MODIS_ORAC	1498	0.298	-0.002	0.873	0.873
MODIS_VPR	1659	-0.058	-2.041	1.066	2.302

Table 10: Overall statistics of ash cloud-top height comparisons with the MISR stereo ash-height product.

9 CONCLUDING REMARKS

This study has comprised the comparison of a large number of products, with varying spatial and temporal characteristics. Matching product pairs across all eruption cases has resulted in a large number of plots and summary statistics. Deriving firm conclusions from these results is very difficult, as:

- With the exception of the information on height provided by active sensors, and the expert classification ash mask for a single SEVIRI image, there is little ground truth with which to conduct a true validation.
- Different sensors, and even different products from the same sensor, provide different amounts of data in terms of coverage, spatial resolution and temporal coverage, which leads to strong sampling issues affecting comparisons of different product-pairs.
- Different products are available for each eruption case study.

Furthermore, when comparing ash detection between products, one is faced with the difficulty of what the goals each detection algorithm are. E.g. is the goal to identify “definite” ash pixels, the most likely total extent of the ash cloud, or those pixels suitable for ash property retrieval; all of these criteria will produce different ash detections and selecting the “best” becomes subjective.

One can anticipate that these problems could be addressed, at least in part, by:

- Providing more expert-classified ash images, ideally providing a large enough sample size to provide a common-pixel mask for the comparison of retrieved ash properties.
- Allowing more time for retrieval teams to produce a more complete set of results for inclusion in the inter-comparison.
- Adopting an iterative approach to the inter-comparison exercise, whereby retrieval teams can submit improved/more consistent products and the comparison approach can be refined, based on each revision of the exercise.
- Large scale comparisons such as presented here, could be complemented by focused case studies designed to reveal the reasons for differences between products:
    - Focus on some well understood test scenes.
    - Constrain common retrieval inputs (eg. Ash optical properties, ancillary data such as Met-fields).
- Make better use of CALIOP data:
    - Identification of scenes which provide a relatively simple retrieval problem (single ash layer without underlying cloud, for instance)
    - Identify ash within CALIOP

Despite these limitations, some qualitative conclusions can be drawn from this study:

- Height and mass tend to “validate”/inter-compare relative well compared to optical depth and effective radius. This is likely a reflection of the limited knowledge of, and relatively simplistic treatment of, ash optical properties in the retrieval algorithms. This problem is undoubtedly complicated by the variability of ash properties from eruption-to-eruption (and even over the course of a single eruption).

- Most schemes perform well in some situations, though it is not always straightforward to focus the comparison on these (beyond drilling down to specific days/scenes). There is little consistency between products in which scenes provide the best results.
- Difficult to validate height for Puyehue without more careful identification of ash in CALIPSO
- The MODIS and SEVIRI schemes from NOAA show an overall high level of maturity:
    - Their ash detection, while not conservative (the extent of the detected ash cloud tends to be larger than in most other products and the expert identified SEVIRI scene), it not prone to the scattered false detection prevalent in many other products.
    - They have good level of consistency with each other, and actively sensed data, for height and mass.
    - They also have tendency to correlate with other schemes (less so for optical depth). This is likely at least partially due to the removal of false-positives from other products by the co-location criteria with the NOAA ash mask.
- IASI schemes seem to provide robust ash detection. Characterising the quality of optical depth, effective radius and mass are hampered by the relatively low spatial resolution of the instrument.
- Tendency for simpler schemes to produce more consistent results. More ambitious schemes sometimes work better but also prone noise and to deviate more in “difficult” conditions – suggesting a trade-off between robustness and extracting maximum information.

10 REFERENCES

RD-1. Thomas, G. E and R. Siddans, “Development of OCA type processors to volcanic ash detection and retrieval:”, Final Report EUMETSAT Contract: EUM/C0/13/4600001276/PDW.

RD-2. Eumetsat, Satellite Derived Volcanic Ash Product Inter-Comparison in Support to SCOPE-Nowcasting, Statement of work EUMfTSS/SOW/14/778570, 28 October 2014

RD-3. WMO, SCOPE-Nowcasting Pilot Project 2 : Globally consistent Volcanic Ash Products, Meeting on the Intercomparison of Satellite-based Volcanic Ash Retrieval Algorithms, 2015, FINAL REPORT, http://www.wmo.int/pages/prog/sat/documents/SCOPE-NWC-PP2_VAIntercompWSReport2015.pdf