Authors' response to Anonymous Referee #1 of "Retrieval of volcanic SO$_2$ from HIRS/2 using optimal estimation" by Georgina M. Miles et al.

The manuscript presents a new scheme for the retrieval of atmospheric SO$_2$ total column amounts after volcanic eruptions from HIRS/2 observations. An optimal estimation retrieval using radiances from three HIRS/2 channels in the mid-infrared region is presented. Retrieved parameters are cloud top and the total column amounts of water vapour and SO$_2$. Major error sources, which are identified by synthetic observations, are cloud/ash interference and the assumptions on the altitude and vertical extend of the SO$_2$ plume. Through simulations it is further demonstrated that the new scheme is superior compared to simple brightness difference methods. The method has been applied to the case of the eruption of the Cerro Hudson volcano in 1991. This is an important piece of work since it presents an improved retrieval scheme to obtain SO$_2$ from TOVS measurements and, thus, opens the possibility to obtain climatological time series of this important trace species. After some modifications/extensions as detailed below, I strongly support its publication in AMT.

General comments:

The optimal estimation scheme is used but not explained nor referenced. I would propose to add a paragraph with the main formulas (adapted to the actual retrieval problem) and add the main references. It would be very helpful to add a table or some graph summarizing the major error terms which have been investigated and how those are handled (some explicitly, some implicitly as part of the measurement error).

We thank the Reviewer 1 in particular for pointing out the missing identification of the optimal estimation scheme. This is a clear oversight and some indication of inverse method used is important to be included in the paper, even if it is only a mathematical tool. We have now referred in the text to Rodgers (2000) generally and Miles *et al.,* (2014) specifically, the latter of which used identical retrieval methodology in terms of the inverse method and cost minimisation part of the retrieval used with the forward model. In that text the methodology is explained quite exhaustively. The following text and references have been added:

"Retrievals are obtained using the Levernburg-Marquart minimisation method after Rodgers (2000), and the full optimal estimation scheme used here is described in detail in Miles et al., (2015)."

Rodgers, C., "Inverse methods for atmospheric sounding: theory and practice", 1 edn, World Scientific, 2000.

5 Miles, G. M., Siddans, R., Kerridge, B. J., Latter, B. G., and Richards, N. A. D.: Tropospheric ozone and ozone profiles retrieved from GOME-2 and their validation, Atmos. Meas. Tech., 8, 385-398, doi:10.5194/amt-8-385-2015, 2015.

The issue of the handing of errors can be elucidated, indeed it could be made clearer that some sources of error can be handled by the retrieval (presence of cloud/ash, SO2/H2O covariance, measurement

10 noise and an attempt at FM error), and others can only be explored to obtain a general indication of the sort of confidence that may be placed on the results (such as height and plume thickness uncertainty).

The following text has been added/amended in the discussion on error study, section 3:

"There are some sources of error that can be incorporated and dealt with by the retrieval. These include measurement noise, the presence of cloud or ash, SO_2/H_2O covariance and an estimate of forward

15 model error discussed above. The main sources of error that cannot be adequately represented in the forward model are errors that impact ill-posed nadir SO_2 column retrievals in general. These are incorrect height assignment of the SO_2 plume, incorrect thickness in the plume represented in the forward model and particularly in the case of infrared measurements and sensitivity to the presence of cloud and/or water vapour. Their relative impacts vary and the sensitivity of the solution to them can be

20 quantified using simulations. It should be noted that some of these errors (plume height and profile shape) cannot often be known at the time of retrieval, and as such the actual impact on the retrieval result also cannot be known. They are investigated here in order to give a general indication as to the potential error that can be associated with the results, to give a window of confidence. Others, such as the impact of cloud or ash on the retrieved SO_2 error can be investigated for use in quality control."

25 *Specific comments:*

P1L20: 'detection method': The method presented here is more than pure 'detection' – it's quantification.

Text has been changed to "..detection and quantification method…".

2

P2L6: could you specify more precisely the channel boundaries for the different HIRS/2 instruments. How much does this affect the retrieval scheme?

There are 15 instruments on just the TiROS/NOAA platforms (1978-2005 before MetOp and HIRS/4) each instrument had similar but slightly different channel configurations but where channels are considered to be in common the central wavenumber is similar. It would be a distraction and exhaustive to describe each of the channel boundaries for each of the instruments, but it is sufficient in the authors' view to mention that the channels can vary between platforms/instruments, but all broadly have the three channels used in the retrieval. It has not been fully investigated for the OE scheme but other HIRS/2 instruments have been used by the Prata fit method in the literature.

This instrument is a broadband radiometer rather than a spectrometer, so small differences in the central wavenumber of channels between instruments is not considered sufficient to appreciably alter or limit this approach between instruments for the OE scheme. They were designed to respond to specific, principal absorbing species relevant to the purpose of the instrument, which was to characterise temperature profiles, water vapour, total ozone, cloud top pressure and surface reflectance, and as such do not change much between instrument as they were designed to be similar (see for example NOAA 2008). The width of the channels is pertinent if there are multiple absorbing species within the envelope of the instrument response function. All such species are required to be taken into account in the forward modelling of the atmosphere if they impact the measurement such that their absence would contribute to model error.

https://docs.lib.noaa.gov/noaa_documents/NESDIS/NOAA-N_Prime_Booklet_12-16-08-1.pdf

P4L24: It would help the reader if a higher resolved spectrum of the single major contributors to the radiance could be provided, overlaid by the channel-boundaries. Are there any other gases contributing in each channel?

Other gases do indeed contribute to the channels, but the predominant species (the absence of which would contribute appreciable error in modelled channel radiance) are included and modelled in RTTOV at a climatological value if they are not retrieved. The error of including at climatological value was

investigated in great detail using the RFM as a way of estimating forward model error (in addition to the impact of modelling the atmosphere at the lower spectral resolution of RTTOV compared to the RFM). Other potential contributions to forward model error that can be estimated this way include spectral resolution of the forward model, including in the FM non-retrieved gases at their climatological value, excluding other gases which are known to exist in the real atmosphere and representing the vertical atmosphere at limited height resolution.

Error contributions deemed significant in the discussion of the RFM used to estimate forward model error (FME) are those that for a channel stand out from the others in terms of magnitude, and that exceed the noise equivalent brightness temperature difference for a given scene temperature. The absolute difference between a test case and the reference case is taken as the channel contribution of FME. The simulations were performed for all 19 of the HIRS/2 channels, irrespective of the fact that only three are used in the column retrieval. Extensive simulations with the RFM were performed that tested the sensitivity of the channel brightness temperatures to variation of the elements listed above. The RFM was run with spectral resolution ranging from 0.001 cm−1 to 0.1 cm−1 to quantify the effects of a reduction in spectral resolution. This only appreciably impacted channels not used here. The change in simulated channel brightness temperature for gaseous species at their climatological level and 1 standard deviation from it were used to quantify the individual impacts of non-retrieved trace gas variability. This only really impacted the channel used to detect column ozone and not used in the retrieval. The RFM was also used to simulate the impact of including all of the minor species such as SF6 and F12-14 (anthropogenic halides). They showed that provided one accepted that their variability was low in the real atmosphere, there was very little sensitivity to them and they did not require inclusion in the background profile used in the forward model, but by this method their exclusion contributed quantitatively to the FME. All elements of forward model error were combined in quadrature. The forward model error as defined by calculations with the RFM is not definitively appropriate to a forward model based on RTTOV. It will contribute to the estimate of the total FME for the purposes of this method development in the absence of an equivalent term being evaluated for RTTOV (which is considerably more challenging to obtain), and broadly constitute a minimum envelope of FME for the HIRS/2 channels. This exercise has only been performed for HIRS/2

4

NOAA11, but since the channels in the retrieval are very similar, the FME contribution is not expected to change appreciably, but could be the subject of future work.

The above is mentioned but very succinctly in section 2.3 for the sake of brevity and concentrating on the channels used in the retrieval.

5   The following text has been added to section 2.1 to be more clear:

"Other atmospheric gases not retrieved but contribute appreciably to channel brightness temperature are represented in the forward model by a climatological value. The potential error that this can introduce is incorporated into the estimate of forward model error."

The following text and figure have been added:
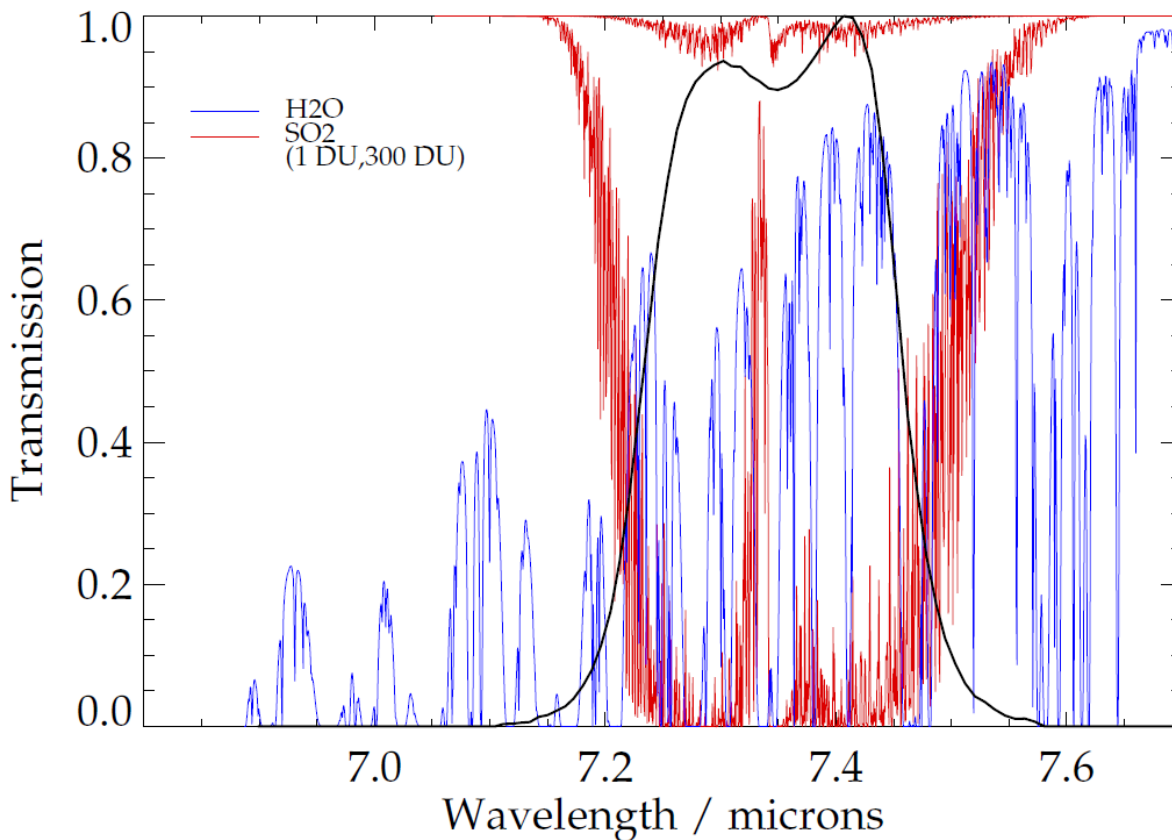
10

"Channel 11 from HIRS/2 aboard NOAA11, centred on 7.2 μm, is shown in Fig. 1. Also shown are simulated transmission spectra for water vapour (which this channel was designed to detect) and SO_2, for two column amounts. It demonstrates both that the channel and spectral feature coincide well, and for large column amounts of SO_2 the channel would be strongly affected."

It is not thought necessary to plot the other two channels, since they are a further water vapour channel and window channel and are not particularly illuminating.

The following has been added to the text:

"Further information about the principle absorbers of the other channels not used in the retrievals can be found in NOAA (1981)."

NOAA (1981), NOAA Technical Report NESS 83: Atmospheric Sounding User's Guide, Technical report, National Oceanic and Atmospheric Administration. NESS 83.

P5L19, 'up to 300 DU': Could you give examples from literature how this number covers the upper limit of volcanic eruptions. In addition, the spectral plots (see comment P4L24) should include lines of SO_2 for different column amounts.

It is not stated that this number (300 DU) covers the upper limit of volcanic eruptions. The limit was chosen in this case to be appropriate for the case study eruption (and those smaller), where a priori knowledge existed (e.g. from TOMS, references given in text) to suggest that in nearly all instances this would be sufficient. This would be the case for the 2008 eruption of Kasatochi which was of a similar magnitude. The training limit is very important due to the way in which RTTOV calculates layer transmittances for gases because some species require higher order terms in their predictor coefficients that are challenging to characterise. To train the model for higher SO_2 column amounts is no doubt possible, but would be the subject of further, future work as non-linearities in the behaviour of the model at very high SO_2 loadings, sensitivity to profile shape and saturation effects would all have to be adequately examined.

In order to put column amounts discussed into context, the following text has been added:

"100 DU represents an SO2 column from a large, explosive volcanic eruption. Pinatubo, for example, yielded column amounts of 350-500 DU (depending upon instrument) after 24 hours which reduced to 100 DU after 7 days (Carn et al., 2005). The OMI instrument (see Table 1) captured column amounts of around 200 DU after the 2008 eruption of Kasatochi (Prata et al., 2010)."

Spectral transmittances for two column amounts are now given in a new Figure 1.

P5L24 'see Fig. 1': Fig. 1 has not been described up to this point, but only later in the text. Further, it does not show the cost 'up to the training limit' of 300 DU, but only to 200DU.

This has been amended to reflect the actual scale used.

P5L30 'calculated numerically': Could you explain this more in detail. Are the analytic Jacobians used at all?

They are evaluated numerically by successive calls to the forward model for fractional perturbations of the state vector. The analytical Jacobians aren't used directly in the retrieval.

The text has been changed as follows: "As a result, these are evaluated numerically in the forward model by successive FM calls where each element of the state vector is fractionally perturbed in turn"

P5L31, 'manually': What does this mean? How large are the limits for H_2O, SO_2? How does it work when it is mentioned 'The weighting functions are allowed to make linear extrapolations. . .'? Is this only valid for the last iteration step?

The word manually has been removed to avoid confusion. It was used in an attempt to convey how the limits were imposed/hard-coded. The text has been modified as follows:

"…constrained in the FM by the physical limits that RTTOV will accept, or that are appropriate for the forward model. These are 0.01 to 800 DU for SO_2, $1e^{-6}$ to 16 times the column water amount predicted by ECMWF and a maximum cloud top height of 16 km (a conservative upper limit for tropopause height."

These limits necessarily apply to all retrieval steps to enable the forward model to function, because the forward model is used to evaluate the weighting functions. The limits, particularly in the case of water vapour, are generally extreme, and on such pixels where it is necessary (a handful out of a week's worth of orbits) there are typically other issues with the measurement that the forward model has a problem replicating. Such pixels are removed at point of quality control as they undoubtedly lead to either non-convergence or very large errors.

P6L25, 'The estimate accounts for inaccuracies that arise due to modelling the atmosphere at reduced spectral resolution, limited vertical resolution, inclusion of nonretrieved trace gases at a climatological level or their preclusion entirely, relative to a reference case.': What is meant with 'limited vertical resolution' and 'modelling .. at reduced spectral resolution'? How has this error been derived (line-by-line compared to band model)? How strong does this error depend on the atmospheric situation?

Please also see response to P4L24 comment above.

Reduced spectral resolution refers to the fact that the RFM is a line by line model, but may be run at poorer resolution to represent something close to the way in which RTTOV represents spectral transmittances.

It is acknowledged that FME contributions may change depending upon the state, but even though the changes are expected to be small, characterising FME is a non-exhaustive process that can only estimate contributing sources of error. In this case, there are larger sources of error from elsewhere (such as incorrect height assignment, errors in representation of SO_2 profile in forward model or the presence of multi-layer optically thin cloud or ash) that are expected to be considerably more dominant.

Section 2.3 now states: "…limited vertical resolution (100 m versus 1 km as used in the forward model outside the region of the SO_2 purturbation),"

P7L9, '100 DU': Could you put this number in perspective of the typical maximum column amounts e.g. after Pinatubo?

The following has been added to the text:

"100 DU represents an SO_2 column from a large, explosive volcanic eruption. Pinatubo, for example, yielded column amounts of 350-500 DU (depending on instrument) after 24 hours which reduced to 100 DU after 7 days (Carn et al., 2005). The OMI instrument (see Table 1) captured column amounts of around 200 DU after the 2008 eruption of Kasatochi (Prata et al., 2010)."

P7L16: How are the a-priori errors of the state vector element for water vapour set? Is this error only considered in the measurement space? Further, could you explain how the off-diagonal elements of the a-priori covariance matrix are set.

The following text has been added to the relevant section: "The a priori error for water vapour is based on the variance of water vapour the ECMWF atmospheric training profiles discussed above relative to the mean."

The error covariance of water with SO_2 is only considered in the measurement space, since it is applied to a channel that is sensitive to both water vapour and SO_2 it effectively absorbs the error covariance – it is mapped onto measurement space. As such there are no off-diagonal elements specified in the a priori covariance matrix. Provided the QC described is applied, they are not applicable.

P8L1, Chapter 3.1, and Fig.1: Regarding the error bars shown in Fig. 1: Could you summarize which errors they contain? Have these errors been incorporated in the synthetic observations? (I assume no or only partially, otherwise the retrieval results should somehow scatter around these errors). Further, do the error bars represent 1 or 2-sigma values? The maximum value tested here seems to be 150 DU. Could you extend this range? You should also show, at which values the method fails and at which column amount of SO_2 the channel signal becomes saturated.

This figure is used to demonstrate deficiencies in the Prata method and the linear behaviour of the OE column retrieval, as appropriate for the case study in particular.

The error bars show the retrieved error. The simulations were performed with simulated measurement noise (which is small) and FME. That is why an OE retrieval is so much more useful than a band model, such as the Prata model shown, which has no possibility of estimating error or quantifying uncertainty. The figure caption has been amended as follows:

"Retrievals based on simulations by a line-by-line model (RFM), with synthetic measurement noise. The error bars for the column retrieval are the retrieved errors."

It is not the case that the maximum value tested is 150 DU, as can be seen from inspecting the far edge of the axis at 200 DU – a limit appropriate for the case study presented. The model fails above the training limit and in such a non-linear way that the authors feel it would be a distraction to show or

9

dwell on this matter, as it is behaviour related to the complex inner mechanics of RTTOV which is itself a comprehensive and complicated system. The channel becomes truly saturated above 1000 DU, far beyond the training limit, and as discussed elsewhere it would be the topic of future and non-trivial work to qualify the model behaviour for significantly larger eruptions.

5   This work is intended as a proof of concept, which the authors feel it demonstrates, rather than the definitive or comprehensive examination of the use of this instrument for all eruptions and for all HIRS/2 instruments. It is a worked demonstration applied to an eruptive event of significance, and as such the simulations and testing are all suitable for supporting both concept and case study.

10 *P8L25, 'Measurements were simulated for a plume at a range of altitudes from 8-18 km': However the caption in Fig. 2 says vice-versa: 'A 2 km thick triangular profile centred at 12 km is used to simulate measurements. The profile is then used in a retrieval with the retrieved height assigned to a range of altitudes.' Could you tell in which way the test retrievals have really been performed?*

This section (referring to what is now figure 3) now reads:

15 "Measurements were simulated for a plume at a range of altitudes from 8-18 km. Figure 3 shows the impact on the retrieved SO_2 column at a specified, fixed altitude of 12 km as a fraction of the true column at these altitudes. Errors range from typically ±0-30 % for most column amounts up to 100 DU, and increase for larger amounts, and for particular altitudes. While the specific error may be state dependent (upon meteorological conditions, specifically the water vapour profile), these simulations do

20 give a general indication as to the magnitude of error that can result from incorrect height assignment of the volcanic plume in the forward model. This is the largest source of error in the OE column retrieval (and the Prata-fit method) and is made more challenging because there is a dependency of the error on column amount. Since height assignment errors cannot be known such simulations can at least give a general indication of potential uncertainty of retrieved amounts, depending on the quality of information

25 available regarding altitude of volcanic SO_2. It is clear therefore that good prior knowledge of the SO_2 plume altitude is necessary for accurate retrieval or fit of SO_2 column amounts from HIRS/2."

The caption for what is now figure 3 has been amended to read:

"A measurement was simulated for a volcanic plume of triangular profile centred at a range of altitudes, for a range of total column amounts. A retrieval is then performed where the plume is assumed to be at 12 km. The fractional difference, or error, is plotted."

5 *P8L28: In this paragraph it is only referred to the Figure, however the results are not described. Please give also in the text at least some quantification of the resulting errors.*
Please see above.

P9L1-3: It is not clear what is different here compared from the paragraph before.
10 The text has been modified to make its purpose more clear: "The performance of the column fit was also directly assessed against a line-by-line model (RFM) for plume altitudes from 8 to 18 km (where the plume height assignment used in the retrieval was the same as that used in the measurement simulated by the RFM) and it was found that…". Specifically, it refers to a test of precision/accuracy of RTTOV vs RFM forward models, to show that it behaves in a similar way compared to the line-by-line
15 model irrespective of SO_2 plume altitude.

P9L6-13: Also here in the text some numbers (%error) should be mentioned. Further, could you explain, why there is such a large difference between the errors when the plume thickness is over-versus underestimated. Would this result not speak for application of a rather sharp profile in the
20 *retrieval to minimize the errors*
It is true that this suggests an underestimate of plume thickness would imply smaller errors than an overestimate. Further work would be required to establish an optimum thickness if there is one, particularly in relation to the vertical grid of the forward model, which may be a limiting factor. It is sufficient here to state that a profile that most resembles the true profile should be the best. In this case
25 there is plenty of ancillary information to give some indication of both plume thickness and plume altitude, namely lidar data. The actual error plume thickness cannot be known, but as with error in height assignment these simulations are a useful indicator to give confidence windows to whatever values might be retrieved.

Indicative numbers have been added to the text:

"The retrieval simulations suggest that errors are larger when the plume thickness is overestimated (typically 13 %), with only small inaccuracies introduced when the plume thickness is under-estimated (less than 2 %). It is therefore possible that an underestimate of plume thickness would result in smaller errors."

P9L29: 'water vapour clouds' should perhaps read 'liquid water clouds' ?

Text has been amended accordingly.

P9L29, 'above 5 km': However in Fig. 4 the retrieval seems to be OK up to 8-9 km. Can you give an explanation why the retrieval has problems to fit cloud heights above a certain altitude. How much does it depend on the atmospheric situation (tropics vs mid/high latitudes)?

The pertinent panel in figure 4 is the bottom right, which shows that deviations (errors) in retrieved water vapour column begin when a cloud is at 6 km. Poor fitting of water vapour leads to errors in the retrieval of SO_2, because the 7.3 micron channel is sensitive to both water vapour and SO_2. A cloud at 5 km shows no perceptible deviation, and it is likely that the threshold is somewhere in between in this case, which is the origin of the 5-6 km warning. The H_2O weighting function of this channel peaks at 700hPa (but as the reviewer mentions this may vary slightly depending on the state, which is why a mid-latitude profile was used). The 6.8 micron channel weighting function peaks at 500 hPa. In the abstract 6 km was stated as this will definitely contribute an error to the retrieval of SO_2.

For clarity this has been corrected to 5 km since it is stated as "…above.." the given level.

P10, chapter 3.4: What is the upper limit of the retrieved SO_2 (e.g. due to saturation effects)? This could also affect the total mass calculation in very dense plumes. Also some information about the convergence criteria of the retrieval and how many iterations are necessary are missing.

This has not been tested with RTTOV. It would require the regression coefficients to be trained for much larger column amounts. Exploratory work with the RFM found that channel brightness temperature differences for a given change in SO_2 column amount become increasingly small over 600

DU for this channel with HIRS/2 on NOAA11. Since this is above column amounts observed even for Pinatubo, it has not been investigated exhaustively and is not referred to in the text. This may change slightly depending on the altitude of the SO_2, but would require further simulations and further work to investigate if this method were to be applied to a an eruption with very high SO_2 column amounts (Pinatubo or larger).

The text in section 2.1 has been modified as follows:

"The 7.3 μm channel is sensitive to both water vapour and SO_2. This channel may be said to saturate for SO_2 columns above 600 DU where significant increases in SO_2 result in small changes in channel BT below the envelope of the channel noise and other error terms."

P12L28: Could you state which SO_2 altitude profile has been used for the case study and how it has been derived. Is the resulting SO_2-altitude error included in the column errors in Fig. 5?

More work was done than has been stated to identify the plume altitude from Hudson, and some of it bears repeating to explain the origin of the SO_2 profile used in the model, before it is stated. The following has been added to the text:

"In addition, contemporary lidar measurements of the Hudson plume were made at the CSIRO (Commonwealth Scientificand Industrial Research Organisation) Division of Atmospheric Research, at Melbourne, Australia (38 S, 145 E) (Young et al., 1992, Barton et al. 1992). These measurements are sensitive to ash, sulphate aerosol and meteorological (water) cloud. The backscatter profiles tend to indicate peaks at around and above 20 km, and frequently at 10-13 km. The higher peak is attributed to aerosol from the Pinatubo eruption. Young et al. (1992) interpret the majority of observations that are thought to include Hudson material as the feature at 12 km in October, with variable cirrus at 10 km. It is reported by the authors that the plume was observed consistently from 28th August until December 1991 between 10 and 13 km, with a decreasing scattering ratio. The relative proportions that contribute to the backscatter measured are expected to be dominated by ash in the first few weeks after the eruption. Little ash is expected to be present after a month beyond the eruption, but by this time the vast majority of the SO_2 will have oxidised into aerosol. Whilst lidar is not sensitive to the presence of gaseous SO_2 inferences can be drawn from the height of the aerosol it eventually becomes. In this case

the lidar information is considered to be a valuable starting point as a guide for estimating the cloud height of the SO$_2$, in the context of other information."

"Using all of this information, the Hudson plume is modelled as a triangular peaked profile with a baseline of 2 km between 11 and 13 km, peaking at 12 km."

5   This is considered to be a very reasonable estimate of the true shape of the SO$_2$ profile, given the considerable ancillary information available.

As stated, Figure 5 shows only the retrieval error. Given the amount of information about the plume height, it is unlikely that the plume height used in the retrieval is more than 1 km out. Figure 5 suggests that this would result in errors of not more than 10% for column amounts between 40-150 DU, and

10 slightly more for smaller column amounts. However, given that the altitude error is unknowable (though thought in this case to be small), it would not be appropriate to assign a value to represent this. It would amount to a guess. This is unfortunately a problem for many nadir viewing techniques where only one piece of information for SO$_2$ is available. This work has gone further than most in attempting to explore the potential errors that could result from incomplete height assignment, and stated often that

15 this element has the potential to be the largest source of error in plume mass estimate. This is why the authors sought as much information as possible on the matter.

P14L5: To make this calculations more clear and give the reader a better feeling for the derived e-folding times and its possible uncertainties it would be necessary to plot derived daily masses after 17th

20 *August and show the fitted exponential decay line.*

This would be the case, but as mentioned in the text, due to the narrowness of the swath and the rate of motion of the plume, in addition to the presence ash in the first day after eruption, of total mass estimates on successive days do not follow a smooth curve. There are also several ways of estimating plume total mass, each have inherent issues associated with them that introduce error. Adding up total

25 mass of the area represented by the footprint adds no information as other approaches like Kriging might, but in the case of HIRS/2 the calibration scanlines are missing, sometimes the plume is only partially sampled by a given orbit and there is movement between orbits. TOMS had the benefit of a wider swath, but HIRS/2 was able to sample both day and night so had more opportunity to monitor the

plume. We found that gridding, which in theory might get around the problem of incomplete sampling resulted in total masses that were heavily dependent upon grid size and in most cases under-estimated the maximum plume mass compared to summing the areas represented by the satellite footprint. Just summing footprints results in a fairly noisy representation of the decay. It is a concern that to venture too far into this discussion is a distraction from the main point of the paper, which is to introduce the technique and demonstrate its effectiveness using a case study. We feel that adding a figure would require significantly more discussion about this issue than is warranted here. Furthermore, while the majority of the SO_2 was released by Hudson on 12th August, about 30 % was erupted at various intervals in the 7 days leading up to it which makes estimating e-folding time for the total mass even more challenging since in the following weeks there is no way to distinguish which material was erupted when.

We have now mentioned this point in this section:

"…be overly-generous bounds by this method. This case is complicated by the fact that about 30 % of the SO_2 released by Hudson was erupted over the 7 days before the main eruption on 15th August, making the calculation of the decay subject to further uncertainty."

The e-folding time in the text was estimated using mass totals on two days where the plume was captured well by the HIRS/2 instrument and may be considered reliable, which is hopefully made clear in the text. The following text has been added:

"In reality the total mass observed does not decay smoothly, but has noise due to the fact that the plume is not always perfectly sampled, and the number of retrieved pixels excluded due to the presence of high or thick cloud or ash varies."

Also:

"More recently, Carn et al. (2016) estimated the e-folding time of Cerro Hudson to be ~7 days, based on mass estimates from TOMS (Constantine et al., 2000). They attribute this anomalously short e-folding time to the late southern hemisphere winter timing of the eruption. However, since Constantine et al., (2000) estimate nearly twice the total mass (4000kT) than that observed by HIRS/2 in this work (and the subsequent TOMS algorithm discussed here) it is possible that the inconsistency in e-folding times could be due to an over-estimate of initial erupted mass from the original TOMS algorithms. Total mass

estimates (and therefore e-folding time estimate) would be improved greatly in accuracy if the HIRS/2 instruments aboard NOAA10 and NOAA12 that were also present were used to result in very comprehensive sampling of this eruption."

5

P14K19, '2300 +- 600 kT': How has the error of 600 kT been calculated?

This is the average retrieval error to appropriate significant figures. The text has been amended as follows:

"This OE column retrieval finds a new total erupted mass estimate for the 1991 eruption of Cerro
10   Hudson of 2300 ± 600 kT from the HIRS/2 instrument aboard NOAA11, where the error is the retrieved error."

Table 3 now also explicitly mentions this in its caption (see below).

*P14L22: Please give also the total masses (including errors) of TOMS, Carn et al., 2016 and Prata et
15   al., 2003.*

The text has been modified to highlight the mass and origin of mass calculated, with further discussion added regarding the inconsistency in e-folding time that is thought to result from the Carn comaparison in particular. No errors are given in that text, but as their origin is the Constantine paper, it is mentioned elsewhere in the text that they estimated their retrieval error to be of the order of 30%. This paper does
20   not detail how that estimate is derived of what it consists of.

The Prata et al., (2003) method estimates an error of 5%, but simulations with line by line models show that this is generally a significant under-estimate. Indeed, no formal estimate of error is possible with the Prata fit method as it currently stands.

The following text has been added to the introduction of the Prata method: "Indeed, it is not possible to
25   formally quantify error of mass estimates from this method as it currently stands."

The caption to what is now Figure 2 has had the following added: "No error estimates are possible for the Prata fit method."

Table 3 (showing comparative mass estimates) has had two comments regarding errors added:

16

"[1]Constantine et al. (2000), with errors estimated to be circa 30 %."

….

"[4]This work, with retrieved error."

Technical comments:

P1L31, 'The TOVS instrument': But in the sentence before it is explained as a suite of instruments.

This has been clarified in the text.

P1L32, 'TIROS': Is written 'TIrOS' in L28.

These have been unified.

P4L21, 'Table 1': Shouldn't this read 'Table 2'?

This has been corrected.

P11L7: '(Constantine et al. 2000)' -> '(Constantine et al., 2000)'

This has been corrected.

P12L14: 'verses' -> 'versus'

This has been corrected.

P13L4: '317 channel' -> '317 nm channel'

Corrected.

P13L5: '340 to' -> 'channel at 340 nm to'

Corrected.

P14L3: Should the formula not read: N(t) = N_0 exp(- lambda t)

Corrected.

P15L15: 'satellites' -> 'satellite'

Corrected.

P23Fig3: '11.5, 12, 13 km' should read '11.5, 12, 12.5 km'

Corrected.

Authors' response to Anonymous Referee #2 of "Retrieval of volcanic SO$_2$ from HIRS/2 using optimal estimation" by Georgina M. Miles et al.

NOAA-9, 10, 11 and 12 were in orbit and operational at the time of the Cerro Hudson eruption. NOAA-11 was selected to pilot the technique principally because of its simple channel configuration. Using the 8.6 micron channel, which is potentially sensitive to both ash and SO$_2$, would be considered to be an extension to this work since it would require the forward model to additionally be able to simulate ash, and proof of concept with just a single channel sensitive to SO$_2$ was the first goal. Furthermore, NOAA11 benefits from an extra window channel (which on the other instruments is the 8.6 micron channel) that can be used for offline detection by means of BT difference or ratio flags should further diagnostics be required. This use was explored but is considered to be beyond the scope of the work, since the alternative and more reliable simulations using a cloud and aerosol model were perused to investigate the limitations of the retrieval under cloudy or ash filled FOVs.

We have modified the text as follows:

"This instrument was selected to demonstrate the capability of this version of the instrument with only one channel that is sensitive to SO_2 and two window channels that have some potential to be used to flag cloud and under some circumstances ash if required."

5 *P5 L19-24: It may be beneficial to train the model for amount larger than 300 DU, as (much) higher values occasionally occur in the most powerful eruptions (e.g. Nabro in 2011). Is there a specific reason to limit the procedure to 300DU? The reference to Figure 1 seems premature, as this figure is discussed only later in the paper and shows total SO_2 amounts up to 200 DU only. I suggest not mentioning the figure or to explicitly state that this figure is to be discussed in more detail later in the*
10 *text.*

It would be beneficial to train the model for higher amounts than 300DU to accommodate the larger and more intense eruptions. The limit was chosen in this case to be appropriate for the case study eruption, where a priori knowledge existed (e.g. from TOMS) to suggest that in nearly all instances this would be sufficient. The training limit is very important due to the way in which RTTOV calculates layer
15 transmittances for gases because some species require higher order terms in their predictor coefficients that are challenging to characterise. To train the model for higher SO_2 column amounts is probably possible, but would be the subject of further, future work as non-linearities in the behaviour of RTTOV at very high SO_2 loadings, sensitivity to profile shape and saturation effects would all have to be adequately examined.
20 The reference to figure 1 has been amended to reflect that it will be discussed in detail later on.

Section 2.2: P6 L10-12: The later assessment of retrieval sensitivity to uncertainties in plume altitude and thickness is introduced here. The vertical extend of the plume is said to be derived from ancillary information. I think it would be good to state that all parameters involved here are effective values,
25 *certainly when using a pre-described triangular profile shape. For example, in reality the SO_2 profile may show multiple peaks at different altitudes. Knowing this, the assumption of a triangular shape is as good as any other.*

A broadly triangular profile may be considered a better representation than any other for a short eruption where material could be expected to gather at a height of neutral buoyancy in the stratosphere where vertical sheer in the advection profile is more limited than in the troposphere over a short altitude range.

5 In tandem with a point from Reviewer 1, the text has been modified to elaborate on the profile used and the motivation for its selection. It is very probable that multiple peaks in the SO_2 profile may exist. This wasn't explicitly tested. Specifically in this case however, the profile used is considered to be a reasonable representation of the plume observed in the case study. As figure 3 demonstrates, there is some sensitivity to the thickness of the modelled plume – modelling it to be too thick introduces more
10 error than under-estimating its thickness in the case tested here, although this is small compared to errors in height assignment. Often (e.g. for IASI and GOME-2) retrievals are performed assuming the material is at 3 altitudes, and the result which best fits the measurements is generally considered to be the 'best', but some human judgement (often based on ancillary information) is also required when looking at results for specific eruptions from these instruments or when considering total erupted mass.
15 In summary, an effective profile is often the best that can be done in the absence of any other information, but there may be some errors associated from getting it wrong – some of which needed to be explored here. As such, every effort must be made to make the profile and height as realistic as possible.

That being said, we appreciate the point the reviewer is making here. Section 2.2. has had the following
20 added to the start:

"In the absence of any further information, an effective SO_2 profile must be represented in the forward model."

More work was done than has been stated to identify the plume altitude from Hudson, and some of it bears repeating to explain the origin of the SO_2 profile used in the model, before it is stated. This is
25 discussed in response to a point below, under Case Study.

3. Error study. Overall a clear, to the point chapter.

Section 3.2.1 took me a bit longer to understand. P8 L25-28: The texts states that retrievals are performed on simulated spectra, with a fixes plume altitude of 12 km assumed in the retrieval. However , Figure 2 suggests that the peak altitude is fixed in the RFM simulations. Which is correct?

This section now reads (referring to what is now Figure 3):

5   "Measurements were simulated for a plume at a range of altitudes from 8-18 km. Figure 3 shows the impact on the retrieved SO_2 column at a specified, fixed altitude of 12 km as a fraction of the true column at these altitudes. Errors range from typically $\pm$0-30 % for most column amounts up to 100 DU an increase for larger amounts, and for particular altitudes. While the specific error may be state dependent (upon meteorological conditions, specifically the water vapour profile), these simulations do

10   give a general indication as to the magnitude of error that can result from incorrect height assignment of the volcanic plume in the forward model. This is the largest source of error in the OE column retrieval (and the Prata-fit method) and is made more challenging because there is a dependency of the error on column amount. Since height assignment errors cannot be known such simulations can at least give a general indication of potential uncertainty of retrieved amounts, depending on the quality of information

15   available regarding altitude of volcanic SO_2. It is clear therefore that good prior knowledge of the SO_2 plume altitude is necessary for accurate retrieval or fit of SO_2 column amounts from HIRS/2."

The caption for what is now figure 3 has been amended to read:

"A measurement was simulated for a volcanic plume of triangular profile centred at a range of altitudes, for a range of total column amounts. A retrieval is then performed where the plume is assumed to be at

20   12 km. The fractional difference, or error, is plotted."

P9 L1-3: What is meant here? Where retrievals also performed using the RFM as forward model, with the conclusion that it performed less well for plumes > 17 km than RTTOV?

25   The text has been modified to make its purpose more clear:

"The performance of the column fit was also directly assessed against a line-by-line model (RFM) for plume altitudes from 8 to 18 km (where the plume height assignment used in the retrieval was the same

as that used in the measurement simulated by the RFM) and it was found that…". Specifically, it refers to a test of precision/accuracy of RTTOV vs RFM forward models.

Section 3.2.2 P9 L9-10. Any idea as to why an underestimation of the plume thickness has significantly less impact on the result than an overestimate? Would this depend on the peak altitude of the plume, bringing it in another temperature/water vapour domain?

We are not certain, but as suggested it probably has something to do with where it is in the atmosphere and how this relates to the water vapour. It does suggest that so long as the plume is located (and modelled) above most of the atmospheric water vapour, it is better to underestimate the plume thickness or else use a very narrow profile in similar cases.

The text has been modified as follows:

"The retrieval simulations suggest that errors are larger when the plume thickness is overestimated (typically 13 %), with only small inaccuracies introduced when the plume thickness is under-estimated (less than 2 %). The modelled cloud top height was 3 km in all cases. It is therefore possible that an underestimate of plume thickness would result in smaller errors."

Section 3.3: The seem to be some little inconsistencies here. The section states that care should be taken with clouds above 5-6 km. A threshold of 5 km is used further on in the paper, whereas the abstract mentions 6 km. Yet, figure 4 suggests that one can go as far as 9 km without any significant problems.

We thank the reviewer for pointing out these inconsistencies. The pertinent panel in figure 4 is the bottom right, which shows that deviations (errors) in retrieved water vapour column begin when a cloud is at 6 km. Poor fitting of water vapour leads to errors in the retrieval of SO_2, because the 7.3 micron channel is sensitive to both water vapour and SO_2. A cloud at 5 km shows no perceptible deviation, and it is likely that the threshold is somewhere in between, which is the origin of the 5-6 km warning. In the abstract 6 km was stated as this will definitely contribute an error to the retrieval of SO_2. For clarity this has been corrected to 5 km since it is stated as "…above.." the given level.

4. Case study.

Please add a few words on why this particular eruption was selected to demonstrate the new algorithm. Also, this eruption is compared to Kasatochi in the abstract, something that you may want to repeat here.

The following text has been added:

5   "In this sense, as well as being a non-equatorial eruption, it has similarities to the 2008 Kasatochi eruption in the Northern hemisphere. It is selected here as a case study because it was a relatively large eruption that has not been studied exhaustively, and a very good example of an eruption in recent satellite history which only TOMS observed with any significance, that can benefit from application of this technique."

10

The assumed plume altitude and thickness of the plume is not mentioned in the text. From the description of previous studies (putting the Cerro Hudson aerosols at 11-13 km) I assume that you used the same 12 km plume as used for Section 3, but this is not clear.

This has now been clearly stated, as well as the reasons for it:

15   "In addition, contemporary lidar measurements of the Hudson plume were made at the CSIRO (Commonwealth Scientificand Industrial Research Organisation) Division of Atmospheric Research, at Melbourne, Australia (38 S, 145 E) (Young et al., 1992, Barton et al. 1992). These measurements are sensitive to ash, sulphate aerosol and meteorological (water) cloud. The backscatter profiles tend to indicate peaks at around and above 20 km, and frequently at 10-13 km. The higher peak is attributed to

20   aerosol from the Pinatubo eruption. Young et al. (1992) interpret the majority of observations that are thought to include Hudson material as the feature at 12 km in October, with variable cirrus at 10 km. It is reported by the authors that the plume was observed consistently from 28th August until December 1991 between 10 and 13 km, with a decreasing scattering ratio. The relative proportions that contribute to the backscatter measured are expected to be dominated by ash in the first few weeks after the

25   eruption. Little ash is expected to be present after a month beyond the eruption, but by this time the vast majority of the SO_2 will have oxidised into aerosol. Whilst lidar is not sensitive to the presence of gaseous SO_2 inferences can be drawn from the height of the aerosol it eventually becomes. In this case

the lidar information is considered to be considered a starting point as a guide for estimating the cloud height of the SO$_2$, to be considered in the context of other information."

"Using all of this information, the Hudson plume is modelled as a triangular peaked profile with a baseline of 2 km between 11 and 13 km, peaking at 12 km."

5. Discussion

Despite the remaining uncertainties in the new algorithm, the authors manage to demonstrate the added value of the new system in comparison to previous methods. It would be nice to see a short summary here of what drawbacks of the Prata methods have been resolved by the new OE schemes and which issues remain, such as the dependence on plume altitude information.

The list has been slightly expanded upon to include some more points:

"They include a quantified error on individual pixel retrieved values, latitudinal variation in accuracy, diagnostic indicators of the retrieval performance and goodness-of-fit and treatment of cloud and water vapour consistent to the retrieval of SO$_2$. When summing mass over a large number of pixels, the precision that these afford becomes increasingly important. Issues that remain are those endemic to ill posed problems where there is only one piece of information on SO$_2$ available and only limited information about the height or shape of the profile of a volcanic plume. It is conceivable that further progress might be made by using HIRS/2 aboard NOAA10 and 12 with the addition of the 8.6 µm channel in ash-free pixels."

I very much liked the clear understanding by the authors that the presented work can be seen as a mere first step toward extending and improving the long-term data series of volcanic SO$_2$ measurements from satellite. I certainly hope that (part of) the proposed future work will be realized.

We thank the reviewer for their support of what we consider to be useful work. This work has so far not had the benefit of any direct sfunding, but it is hoped that it may contribute to a future case for support for funding. We welcome any opportunities for collaboration.

Cosmetics:

24

P2 L28: TiROS –> TIROS; Or keep TiROS and used this spelling consistently throughout the paper.

These have been unified to read TIrOS.

P13 L1-5: The wavelength unit (nm) is missing a few times.

This has been corrected.

5 *P14 Eq 1: minus symbol missing in the exponent.*

This has been corrected.

P23: Caption: 11.5, 12, 13 –> 11.5, 12., 12.5

This has been corrected.

10

15

20

25

Retrieval of volcanic SO_2 from HIRS/2 using optimal estimation

Georgina M. Miles[1,2], Richard Siddans[2], Roy G. Grainger[1], Alfred J. Prata[1,3], Bradford Fisher[4], Nickolay Krotkov[5], Brian Kerridge[2]

[1]Atmospheric, Oceanic and Planetary Physics, University of Oxford, Oxford, UK
[2]Remote Sensing Group, STFC Rutherford Appleton Laboratory, Harwell Oxford, UK
[3]Nicarnica AS, Lysaker, Norway
[4]Science Systems and Applications, Inc, 10210 Greenbelt Road, Suite 600, Lanham, Maryland, USA
[5]Goddard Space Flight Center, Greenbelt, Maryland, USA

Correspondence to: G. Miles (georgina.miles@stfc.ac.uk)

Abstract. We present an optimal estimation (OE) retrieval scheme for stratospheric sulphur dioxide from the High Resolution Infrared Radiation Sounder 2 (HIRS/2) instruments on the NOAA and MetOp platforms, an infrared radiometer that has been operational since 1979. This algorithm is an improvement upon a previous method based on channel brightness temperature differences developed by Prata et al. (2003), which demonstrated the potential for monitoring volcanic SO_2 using HIRS/2. The Prata method is fast but of limited accuracy. This algorithm uses an optimal estimation retrieval approach yielding increased accuracy for only moderate computational cost. This is principally achieved by fitting the column water vapour and accounting for its interference in the retrieval of SO_2. A cloud and aerosol model is used to evaluate the sensitivity of the scheme to the presence of ash and water/ice cloud. This identifies that cloud or ash above 6 km limits the accuracy of the water vapour fit, increasing the error in the SO_2 estimate. Cloud top height is also retrieved. The scheme is applied to a case study event, the 1991 eruption of Cerro Hudson in Chile. The total erupted mass of SO_2 is estimated to be 2300 kT $\pm$ 600 kT. This confirms it as one of the largest events since the 1991 eruption of Pinatubo, and of comparable scale to the Northern Hemisphere eruption of Kasatochi in 2008. This retrieval method yields a minimum mass per unit area detection limit of 3 DU, which is slightly less than that for the Total Ozone Mapping Spectrometer (TOMS), the only other instrument capable of monitoring SO_2 from 1979–1996. We show an initial comparison to TOMS for part of this eruption, with broadly consistent results. Operating in the infrared (IR), HIRS has the advantage of being able to measure both during the day and at night, and there have frequently been multiple HIRS instruments operated simultaneously for better than daily sampling. If applied to all data from the series of past and future HIRS instruments, this method presents the opportunity to produce a comprehensive and consistent volcanic SO_2 timeseries spanning over 40 years.

1 Introduction

Volcanic eruptions are important for climate and climate change. They perturb atmospheric chemistry and radiative transfer. Their signal in climatic records must be accurately quantified before any attribution of climate change to anthropogenic

sources. Furthermore, by studying the response of the atmosphere to volcanic eruptions in terms of climate sensitivity this can test ideas relating to climate prediction.

The monitoring of volcanic SO_2 emissions, the main precursor to sulphate aerosols, is crucial for accurately characterising total emission estimates but also for understanding plume evolution. Until the mid-1990's, only one principal instrument
5 (the Total Ozone Mapping Spectrometer, TOMS) has been able to observe eruptions for an adequate period to generate something approaching a climate relevant record. The sensitivity of TOMS limits it to detecting only the larger, explosive eruptions rather than effusive ones where material remains predominantly in the troposphere. Satellite instruments that have been used to measure volcanic SO_2 are given in Table 1. From 1996, with the advent of the Global Ozone Monitoring Experiment (GOME) class instruments (UV-vis spectrometers) sufficient spectral resolution (and spatial resolution) has
10 enabled the detection of lower amounts of SO_2 with higher accuracy from increasingly smaller eruptions. This has improved further still with instruments such as the Infrared Atmospheric Sounding Interferometer (IASI), from which SO_2, sulphate aerosol and ash may be derived simultaneously due to its high spectral resolution and broad spectral coverage (Karagulian et al., 2010). Total erupted mass estimates for volcanic eruptions can often differ by greater than 100% between instruments, as a result of sampling, geometry, differences in sensitivity and assumptions that contribute to algorithms, such as plume
15 height. For example, Thomas et al., (2009) present a multi-sensor comparison of the 2005 eruption of Sierra Negra (Galapagos Islands), using concomitant observations by TOMS, OMI and MODIS. They found a wide estimate of total erupted SO_2 calculated from the three instruments, ranging from 60 kT to 1800 kT.

It is still the case that the operational period of these more sensitive, recent instruments is not yet long enough to constitute a
20 climate-relevant record. Here we present the methodology for a relatively fast and accurate volcanic SO_2 detection and quantification method for an instrument originally designed to operationally measure water vapour and temperature profiles.

HIRS/2 has the potential to have captured stratospheric emissions from explosive eruptions continuously since 1979, but with significantly higher temporal sampling and greater sensitivity than TOMS. This enables the 35 year volcanic SO_2
25 emission record from satellites to be significantly enhanced, with potential uses for constraining models and examining in detail individual eruptions and plume evolution.

1.1 HIRS/2 Instrument

HIRS/2 is one of three instruments that originally constituted the Television Infrared Observation Satellite (TIrOS) Operational Vertical Sounder (TOVS), designed to provide atmospheric profile measurements of temperature and water
30 vapour structure (Smith et al., 1979). The other TOVS instruments were the Stratospheric Sounding Unit (a radiometer) and the Microwave Sounding Unit (a scanning microwave spectrometer). The TOVS suite of instruments was first launched in 1979 aboard the new NOAA satellites based on the TIrOS-N design, and evolved in to the Advanced TOVS (ATOVS) system. Subsequent replacements have been deployed for the last 30 years aboard NOAA satellites (NOAA 6-17) (JPL,

2003), and more recently European platforms including most recently MetOp-A and B as HIRS/4. Throughout its deployment there have been at least two instruments (and occasionally three) orbiting simultaneously. HIRS/2 has 19 detector channels in the infrared and one in the visible part of the spectrum for cloud detection during the day. These channels are relatively broad, spanning between 0.1 and 0.5 μm depending upon wavelength. The key instrument parameters

5   are given in Table 2.

Two HIRS/2 channels coincide with SO_2 spectral absorption features, these being 7.3 μm (a strong asymmetric stretch vibration band) and 8.6 μm. The precise central wavenumber is dependent upon instrument version, and only HIRS aboard NOAAs 10 and 12 featured an 8.6 μm channel. These channels were originally chosen to be sensitive to water vapour for

10   use in sounding and applying corrections for the CO_2 and window channels. The 8.6 μm channel is also reported to be sensitive to volcanic ash and other aerosols (Kearney and Watson, 2009).

Channel 11 from HIRS/2 aboard NOAA11, centred on 7.2 μm, is shown in Fig. 1. Also shown are simulated transmission spectra for water vapour (which this channel was designed to detect) and SO_2, for two column amounts (1 and 300 DU). It

15   demonstrates both that the channel and spectral feature coincide well, and for large column amounts of SO_2 the channel would be strongly affected.

1.2 Previous efforts to retrieve of SO_2 with the HIRS instrument

Prata et al., (2003) demonstrated a method to detect volcanic SO_2 from HIRS, providing the SO_2 perturbation is strong enough, and located above any significant sources of water vapour. It is based on a synthesis of the expected clean

20   atmosphere brightness temperature for the channel, and the observed deviation from it when contaminated by SO_2. This method, hereafter referred to as either the Pratafit method or after Prata et al., (2003), uses a linear interpolation between the brightness temperatures of adjacent channels. It also assumes a fixed height of erupted volcanic SO_2, since theoretically only one piece of information can be obtained from one channel, and column amount is not insensitive to the height of the plume. The technique requires the SO_2 to be located in the upper troposphere/stratosphere above most of the atmospheric water

25   vapour, and there is no information about the height of the plume from the instrument itself. This information may be gleaned from other types of observations, but the fit is reliant upon the accuracy of this independent information.

A description of how the Prata method operates is detailed in Prata et al., (2003). While useful in itself, its most significant shortcoming is that due to its simplicity, the model is unable to capture atmospheric variability (other than potentially that of

30   SO_2). This particularly alludes to the variability of cloud, temperature and water vapour. Without independent height information of the SO_2 the radiance relationships are subject to potentially significant error. Indeed, it is not possible to formally quantify error of mass estimates from this method as it currently stands. Its strengths are that the operations required are computationally inexpensive and straightforward, as it is based on the principles of a band model. It has also

performed well against other observational data sets, although the previously mentioned uncertainties that contribute to error make quantifying overall uncertainty difficult. It uses a minimum offset threshold in brightness temperature for the channel affected by SO_2 in order to predict the presence of SO_2 and yet excludes the effects of atmospheric water vapour variability. As such, its sensitivity to low amounts of SO_2 is limited.

Guo et al. (2004) presented a re-evaluation of the 1991 Pinatubo eruption using SO_2 derived from HIRS/2 using the Prata-fit method, and compared it to SO2 derived from TOMS measurements. They were found to be broadly consistent. The Prata-fit method works sufficiently well to suggest that the 7.3 μm SO_2 feature it uses is robust enough to make further exploitation more refined. Use of information arising from other HIRS channels would constitute an improvement to the Prata fit method, as multiple wavelength information can be used to diagnose attributes of the atmospheric profile such as temperature and the presence of cloud. This problem is well suited to an optimal estimation retrieval, which would incorporate a forward model (FM) of sufficient complexity to represent these atmospheric attributes. As with the Prata fit, unavoidably it will require some estimate of the altitude of an SO_2 plume.

1.3 Outline of paper

In Section 2, an Optimal Estimation (OE) retrieval algorithm methodology to extend the Prata-fit method is presented. Section 3 comprises an error study and presents results of retrievals from simulated measurements in order to understand the sensitivity of the algorithm and potential sources of error. Section 4 presents a case study of the 1991 Cerro Hudson eruption, where the algorithm is applied to real data and new eruption mass estimates are evaluated, and compared to existing mass estimates from other instruments/methods. In Section 5 the results are discussed and further work is suggested.

2 Methodology

2.1 Retrieval algorithm and forward model

The HIRS/2 measurements used here are all-sky brightness temperatures from the instrument aboard NOAA11. This was selected to demonstrate the capability of this version of the instrument with only one channel that is sensitive to SO_2 and two window channels that have some potential to be used to flag cloud and under some circumstances ash if required (although only one is used here directly). The brightness temperatures are a product derived from the raw voltage measurements via a radiance and brightness temperature conversion and have been subject to calibration factors and some basic quality control. Further information about the instrument is available from NOAA (1981) and elsewhere. The data format contains the time in seconds from midnight of the measurement, the solar zenith angle, 19 IR channel brightness temperatures, one visible channel albedo, latitude, longitude, satellite altitude, line number for each orbit and the scan position (see Table 2).

Retrievals are obtained using the Levernburg-Marquart minimisation method after Rodgers (2000), and the full optimal estimation scheme used here is described in detail in Miles et al., (2015). The retrieval uses three HIRS/2 channels to derive three products: the SO_2 column, a scaling factor for a water vapour profile and effective cloud top pressure. The 7.3 μm channel is sensitive to both water vapour and SO_2. This channel may be said to saturate for SO_2 columns above 600 DU where significant increases in SO_2 result in small changes in channel BT below the envelope of the channel noise and other error terms. The weighting function for water vapour of the 6.8 μm channel peaks at around 500 hPa (around 5 km), and as such would have some sensitivity to the region where the vast majority of the water vapour in the column resides. To represent both channels accurately, some knowledge of cloud is required, which may be gleaned from the 11.1 μm channel window channel. This channel is highly sensitive to the emitting temperature of the lowest surface it observes (be it cloud or the surface), thus with some knowledge of the surface and atmospheric temperature profile it is possible to obtain an estimate of cloud top height. Other atmospheric gases not retrieved but contribute appreciably to channel brightness temperature are represented in the forward model by a climatological value. The potential error that this can introduce is incorporated into the estimate of forward model error.

Radiative Transfer for TOVS (RTTOV) is a radiative transfer model (developed by the UK Met Office, Saunders et al., 1999, ECMWF 2001) designed to simulate the instruments of TOVS including HIRS/2, and is used extensively (particularly for assimilation) because of its speed. It calculates layer transmittances for a variety of trace gas species using look-up tables of parameterised regression coefficients for a range of temperatures and pressures. It has been further developed since the TOVS system was first deployed, and version 10 is used here. RTTOV also has the functionality to compute partial derivatives.

RTTOV estimates channel brightness temperature based on pre-calculated coefficients for layer transmittances that are generated for a range of atmospheric profiles. As such, it is extremely fast, but as it stands it does not incorporate any representation of SO_2 other than at a very low climatological value. To alter the transmittance model to include SO_2 would require substantial re-working of program code. It is possible to calculate a set of predictor coefficients for SO_2 and incorporate them within RTTOV by replacing the properties of another gaseous species that has negligible impact on the total column transmittance within the selected HIRS/2 channels (in this case, carbon monoxide). The coefficients were generated by a `training' methodology using an extensive range of specimen atmospheric profiles, where the SO_2 was represented from very low/background levels to very large perturbations, after Matricardi (2008, 2010) and Siddans (2011). This approach retains the speed and accuracy offered by RTTOV and enables the model to be used to represent atmospheric gases for future instruments not already catered for (ECMWF, 2001).

For this work, the predictors were trained using profiles with up to 300 DU. Some care is required in the generation of these coefficients for SO_2. They are required to be limited to those that represent a first order relationship with SO_2 since the more complicated (higher order) predictors caused erroneous results. This is thought to be a result of both the dynamic range SO_2 can exhibit in a volcanically perturbed atmosphere, and the fact that RTTOV was not explicitly designed to model SO_2 for
5 this instrument. The cost in terms of accuracy over this range of SO_2 is shown to be small, as demonstrated in Fig. 2 to be discussed in detail later.

The column retrieval developed here uses atmospheric profiles from the ECMWF ERA-Interim product (Dee et al., 2011) to represent atmospheric properties other than SO_2, or as a first guess in terms of the water vapour profile. These contain
10 profiles on a pressure grid of 37 levels from 1000 hPa to 1 hPa. RTTOV is capable of generating weighting functions, but they refer to the sensitivity of the simulated measurements to perturbations in the atmospheric profile, rather than directly to changes in state vector. As a result, these are evaluated numerically in the forward model by successive FM calls where each element of the state vector is fractionally perturbed in turn. RTTOV has certain physical limits for its input values, and when occasionally the predicted updated state lies outside these they are constrained in the FM by the physical limits that
15 RTTOV will accept, or that are appropriate for the forward model. These are 0.01 to 800 DU for SO_2, $1e^{-6}$ to 16 times the column water amount predicted by ECMWF and a maximum cloud top height of 16 km (a conservative upper limit for tropopause height). The weighting functions are allowed to make linear extrapolations beyond these limits, allowing the retrieval more freedom, but unphysical profiles are suppressed with quality control of the derived products (discussed later).

20

2.2 Profile definition in forward model

In the absence of any further information, an effective SO_2 profile must be represented in the forward model. The three-element state vector comprises a scaling factor for the SO_2 profile, a scaling factor for a water vapour profile and a cloud top pressure. A volcanic SO_2 perturbation is represented by a vertically localised triangular profile. This triangular profile is
25 normalised to have an integrated mass of 1 DU. This was partly done to ease interpretation, since the retrieved scaling factor would be approximately equal to the total amount of SO_2 in the column. The rest of the profile is prescribed by a background SO_2 volume mixing ratio climatology, the total column mass of which is less than 1 DU. In the forward model, a scaling factor applies to a specified height region of the SO_2 profile, scaling all elements within and none outside this. The expected region of the volcanic plume is estimated using ancillary information, such as lidar or results from modelling of the
30 eruption available in the literature. The sensitivity to how well the altitude and thickness of an SO_2 plume is evaluated using retrievals from simulated measurements, and detailed in Section 3.

In an analogous way to SO_2, H_2O is represented in the state vector by a profile scaling factor, but it applies to the entire profile rather than a localised height region. The profiles used for retrieval are those collocated from the ECMWF ERA-

Interim product for a given HIRS/2 pixel (which represents the best guess for the state), but in principle any climatological profile can be used. In the case where a scaling factor is close to one, it would indicate that the H_2O profile is similar to that which produced the measurement.

The third element of the state vector is cloud top height (CTH), or specifically the geopotential height at an equivalent pressure level. It was found that the speed of convergence was significantly reduced if the initial guess of cloud top pressure was reasonably accurate. As such, this is derived before the retrieval using interpolation between calls to a radiative transfer model that simulates the 11.1 μm channel brightness temperature (BT) for 0-10 km (using associated ECMWF ERA Interim temperature profile), and included a test for temperature inversions.

2.3 Error

An estimate of forward model error was calculated using the Reference Forward Model (RFM) — a line-by-line radiative transfer model (Dudhia, 2002), discussed further in Section 3. The estimate accounts for inaccuracies that arise due to modelling the atmosphere at reduced spectral resolution, limited vertical resolution (100 m versus 1 km as used in the forward model outside the region of the SO_2 perturbation), inclusion of non-retrieved trace gases at a climatological level or their preclusion entirely, relative to a reference case. This yields a channel quantity (in brightness temperature) that is combined in quadrature with the noise equivalent differential radiance for each channel, and is thus incorporated into measurement noise for the purposes of the retrieval. The a priori error associated with cloud height is 10 km. The a priori error for water vapour is based on the variance of water vapour in the ECMWF atmospheric training profiles discussed above relative to the mean.

2.4 Estimation of SO_2 and H_2O covariance for HIRS/2

Establishing an appropriate SO_2 a priori error is potentially a non-trivial issue with regard to a retrieval problem where the measurements have relatively little sensitivity. A volcanically perturbed SO_2 profile can contain 2 or 3 orders in magnitude more than a background profile, and at the centre of a large plume this can be even more. A good a priori error gives the retrieval the freedom to find a correct minimum in cost space, and can restrict it from converging on a solution that is un-physical. The variance for a background profile would be very small, as opposed to a profile where SO_2 is expected, which would be very large. If there is sufficient information contained within the measurements, one would conventionally use a variance that spans both scenarios. This results in a poor constraint for an ill posed problem but is necessarily used here, where a first guess/a priori error of 100 DU is used and a prior variance is the first guess squared. 100 DU represents an SO_2 column from a large, explosive volcanic eruption. Pinatubo, for example, yielded column amounts of 350-500 DU (depending upon instrument) after 24 hours which reduced to 100 DU after 7 days (Carn et al., 2005). The OMI instrument (see Table 1) captured column amounts of around 200 DU after the 2008 eruption of Kasatochi (Prata et al., 2010).

32

Early results of the retrieval scheme run with real measurements revealed that there were many `false positives' of SO_2 retrieved. Their structure indicated that they were related to the presence of water vapour, or errors in the fit for water vapour. This indicated the degree of covariance between SO_2 and water vapour which had to be incorporated into the retrieval since the 7.3 μm channel is sensitive to both water vapour and SO_2.

5

The retrieval was applied to one day of `clean' measurements in the Southern Hemisphere where no volcanically perturbed profiles were expected. The retrieval was forced not to retrieve SO_2 by artificially constraining the a priori variance, but none-the-less small amounts of SO_2 are retrieved from that channel because of inadequacies in characterising the water vapour. The brightness temperature fit residuals in the SO_2 channel were very small, but it is expected that nearly all of the

10 SO_2 being retrieved on this day is being falsely attributed. The standard deviation of the 7.3 μm channel brightness temperatures fit residual in the retrieval of 0.92 K constitutes an estimate of the 'real world' error covariance of water vapour with SO_2 for this instrument. This is incorporated by adding it in quadrature to the forward model error for this channel and resulted in a significant reduction in the occurrence of false positives.

15 **3 Error study: Retrievals from simulated measurements**

There are some sources of error that can be incorporated and dealt with by the retrieval. These include measurement noise, the presence of cloud or ash, SO_2/H_2O covariance and an estimate of forward model error discussed above. The main sources of error that cannot be adequately represented in the forward model are errors that impact ill-posed nadir SO_2 column retrievals in general. These are incorrect height assignment of the SO_2 plume, incorrect thickness in the plume represented in

20 the forward model and, particularly in the case of infrared measurements, sensitivity to the presence of cloud and/or water vapour. Their relative impacts vary and the sensitivity of the solution to them can be quantified using simulations. It should be noted that some of these errors (plume height and profile shape) cannot often be known at the time of retrieval, and as such the actual impact on the retrieval result also cannot be known. They are investigated here in order to give a general indication as to the potential error that can be associated with the results, to give a window of confidence. Others, such as

25 the impact of cloud or ash on the retrieved SO_2 error can be investigated for use in quality control.

3.1 Spectral precision of forward model

In order to assess the accuracy of the RTTOV-based fast column retrieval forward model, it is compared to simulations from a model with a higher accuracy. The RFM is a line-by-line radiative transfer model (Dudhia, 2002) capable of modelling the

30 atmosphere at a spectral resolution of up to 0.0001 cm^{-1}. The RFM is not suitable for the forward model because it is

computationally expensive and it does not inherently represent any effects of cloud or ash. Figure 2 shows the results of column retrievals from HIRS/2 channel BTs simulated by the RFM, using a sample ERA-Interim cloud-free meteorology (temperature and water vapour profiles) at 0 and 60°S latitude and 0°W longitude, where only the column amount of SO_2 is changed in the simulation. It also shows the SO_2 fit by the Pratafit method. The Pratafit method does not fit SO_2 below 5 DU, which depending upon the atmospheric state can be equivalent to an observed brightness temperature difference of up to 4 K. The bias of the Prata fit has a dependence upon latitude, primarily because of the different amount of water vapour in the profile at the two latitudes shown here. The column retrieval has a very small bias that only becomes perceptible at SO_2 loadings approaching 200 DU, at which point it is of the order of <5 DU.

3.2 Sensitivity to forward model representation of SO_2 plume

Both the altitude and amount of SO_2 affect the 7.3 μm channel brightness temperature but as there is only one channel sensitive to SO_2 on NOAA11 considered here, there is at most one piece of information that can be retrieved for SO_2. Therefore, for an accurate retrieval of SO_2 column, it is important to have some knowledge of the plume altitude or its vertical profile. The column retrieval developed here requires some information of the height of the SO_2, but this can be subject to uncertainty and may change with time. As such, the sensitivity of the retrieval to errors associated with plume height and specification must be examined.

3.2.1 Altitude

Measurements were simulated for a plume at a range of altitudes from 8-18 km. Figure 3 shows the impact on the retrieved SO_2 column at a specified, fixed altitude of 12 km as a fraction of the true column at these altitudes. Errors range from typically ±0-30 % for most column amounts up to 100 DU, and increase for larger amounts, and for particular altitudes. While the specific error may be state dependent (upon meteorological conditions, specifically the water vapour profile), these simulations do give a general indication as to the magnitude of error that can result from incorrect height assignment of the volcanic plume in the forward model. This is the largest source of error in the OE column retrieval (and the Prata-fit method) and is made more challenging because there is a dependency of the error on column amount. Since height assignment errors cannot be known such simulations can at least give a general indication of potential uncertainty of retrieved amounts, depending on the quality of information available regarding altitude of volcanic SO_2. It is clear therefore that good prior knowledge of the SO_2 plume altitude is necessary for accurate retrieval or fit of SO_2 column amounts from HIRS/2.

The performance of the column fit was also directly assessed against a line-by-line model (RFM) for plume altitudes from 8 to 18 km (where the plume height assignment used in the retrieval was the same as that used in the measurement simulated by the RFM) and it was found that for altitudes of over 17 km the column fit was unable to retrieve SO_2 columns less than 30 DU, but in all other cases true clear-sky column amounts were retrieved accurately from simulated measurements.

5

3.2.2 Profile shape and plume thickness

Figure 4 shows the consequences that can result from retrieving the volcanic plume with a fixed profile shape that represents the thickness of the plume incorrectly. Measurements were simulated using a triangular profile centred at 12 km but with baselines of 1 and 4 km. They were then used in the retrieval with a fixed profile shape with a triangular perturbation also centred at 12 km, but with a baseline of 2 km (thought to be the best representation of the plume used in the case study in Section 4). The retrieval simulations suggest that errors are larger when the plume thickness is overestimated (typically 13 %), with only small inaccuracies introduced when the plume thickness is under-estimated (less than 2 %). The modelled cloud top height was 3 km in all cases. It is therefore possible that an underestimate of plume thickness would result in smaller errors.

3.3 Sensitivity of Retrieval Scheme to Cloud and Ash

Some understanding must be obtained of how the column retrieval forward model behaves in the presence of ash and cloud of different type. The forward model fits a cloud top pressure using the 11.1 μm channel, which is expected to work well for most scenes with cloud in the troposphere. The effect of cloud on the other channels is examined here using a cloud model, the Oxford-RAL Retrieval of Aerosol and Cloud (ORAC) model. The model is described in detail by Poulsen et al. (2012), where it was used as part of an optimal estimation retrieval of cloud properties for the Along Track Scanning Radiometer (ATSR) by simulating radiances in a combination of visible, near infrared (NIR) and IR channels. The model parameterises a cloudy scene by ascribing cloud phase, effective radius of a size distribution, the 0.55 μm optical depth and a cloud top pressure. It uses the plane parallel approximation and models cloud as a single layer. The model represents trace gases at a background climatological level. The system can also be used to retrieve ash plume properties: plume height, optical thickness and ash particle effective radius (McGarragh et al., in preparation, 2017)

HIRS/2 measurements were simulated for a range of liquid and ice cloud and ash optical depths, effective radii and at a range of altitudes when no volcanic SO_2 is present. These channel brightness temperatures were then used to retrieve SO_2 to identify where this resulted in an erroneous fit.

An example is shown in Fig. 5, which shows that for liquid water clouds above 5 km, the column retrieval erroneously retrieves some SO_2 when there is none, the water vapour and cloud top height become inaccurate and the fit cost begins to

35

increase. The results indicated that low optical depth or effective radii for cloud or aerosol can result in poor fitting of the measurements, both resulting in an underestimate of cloud top pressure with false positives of SO_2 and an over-estimation of water vapour. This yields a crucial quality control threshold where retrieved cloud top altitudes of greater than 5-6 km should not be trusted, as they are likely to result in spurious detection of SO_2 and a high retrieval cost. This may imply that very thin cloud beneath 5 km (or incorrectly retrieved to be) could still contribute to poor fitting of the measurements.

3.4 Quality Control

The results of the column retrieval must be subject to some quality control. In addition to the disregard of non-converged and converged pixels with cloud retrieved at an altitude greater than 5 km, a retrieved column is only considered useful if the error is less than the retrieved amount. Quality control becomes very important when erupted plumes are used to calculate total erupted mass, where even a small amount of noise can yield a biased mass total. For the purposes of gridding or summing pixels for deriving a global/plume mass estimate, a minimum retrieved SO_2 threshold may be applied in deference to the lower detection limit of the retrieval, in order to avoid spurious low values that the retrieval should not be sensitive to, such as those relating to water vapour or cloud that are not accounted for in either the error covariance or the forward model. An effective way of obtaining this quantitatively is to apply a 2 or 3 sigma test, where sigma is the standard deviation of the retrieved SO_2 on a day when no volcanic SO_2 is expected to be present. This threshold gives statistical confidence that a value above it is significantly distinct from the noise above the 95 or 97 percentile. The sigma threshold for 6th August 1991 (a day when there was no SO_2 present in the region relating to the case study in Section 4) was 2.7 DU, and is probably a lower estimate of the detection limit of the HIRS/2 SO_2 column retrieval in the mid-latitudes. Multiples of this value indicates confidence that a retrieval result is dominated by signal rather than noise.

4 Case Study: Cerro Hudson Eruption in 1991

Cerro Hudson (45.54°S, 72.58°W, elevation 1905 m) is a stratovolcano in the south Chilean Andes that erupted explosively in August 1991, two months after the Pinatubo eruption. The eruption was estimated to be 10-20 times smaller than Pinatubo in terms of SO_2 that was expected to be emitted. In this sense, as well as being a non-equatorial eruption, it has similarities to the 2008 Kasatochi eruption in the Northern hemisphere. It is selected here as a case study because it was a relatively large eruption that has not been studied exhaustively, and a very good example of an eruption in recent satellite history which only TOMS observed with any significance, that can benefit from application of this technique.

At the time of the 1991 eruption, the only satellite available that could detect SO_2 with any demonstrated accuracy was TOMS. The Microwave Limb Sounder, a contemporaneous instrument that observed SO_2 from Pinatubo at a higher altitude, produced noisy results in the lower stratosphere at this latitude (Read et al, 1993). In addition, contemporary lidar measurements of the Hudson plume were made at the CSIRO (Commonwealth Scientificand Industrial Research

5   Organisation) Division of Atmospheric Research, at Melbourne, Australia (38 S, 145 E) (Young et al., 1992, Barton et al. 1992). These measurements are sensitive to ash, sulphate aerosol and meteorological (water) cloud. The backscatter profiles tend to indicate peaks at around and above 20 km, and frequently at 10-13 km. The higher peak is attributed to aerosol from the Pinatubo eruption. Young et al. (1992) interpret the majority of observations that are thought to include Hudson material as the feature at 12 km in October, with variable cirrus at 10 km. It is reported by the authors that the plume

10   was observed consistently from 28th August until December 1991 between 10 and 13 km, with a decreasing scattering ratio. The relative proportions that contribute to the backscatter measured are expected to be dominated by ash in the first few weeks after the eruption. Little ash is expected to be present after a month beyond the eruption, but by this time the vast majority of the SO_2 will have oxidised into aerosol. Whilst lidar is not sensitive to the presence of gaseous SO_2, inferences can be drawn from the height of the aerosol it eventually becomes. In this case the lidar information is considered to be a

15   valuable starting point as a guide for estimating the cloud height of the SO_2, in the context of other information. As well as some ground observations, the Hudson eruption was sensed remotely by AVHRR (ash), lidar (sulphate aerosol) and incidentally by an aircraft (Barton et al. 1992). Hofmann et al. (1992) reported possible exacerbation of Antarctic ozone depletion of 10-20% of total column due to the presence of Hudson aerosol in the lower stratosphere for September 1991. The anomalous depletion occurred within the polar vortex predominantly at 11-13 km and 25-30 km, the respective altitudes

20   of the Hudson and Pinatubo aerosols.

The transport of the Hudson volcanic plume was first numerically modelled by Barton et al. (1992), to reasonably good agreement with satellite and lidar observations. The plume was also modelled using an isentropic trajectory model, initiated by TOMS observations of SO_2 (Schoeberl et al. 1993). These models showed good spatial agreement with observations for

25   the first eight days after the eruption which is an indication that the height assignment of the erupted plume was accurate within the models. The most explosive eruption began and ended on 15th August. It was at this stage of its eruptive phase that the majority of the material was injected into the stratosphere (Constantine et al., 2000).

4.1 Results

Using all of this information, the Hudson plume is modelled as a triangular peaked profile with a baseline of 2 km between

30   11 and 13 km, peaking at 12 km. Figure 6 shows an example of the SO_2 retrieval applied to a day of data on 15th August 1991, and its associated retrieval error. Figure 7 shows results for the same day as Figure 6, but for the other elements of the state vector: the retrieved water vapour scaling factor and cloud top height (with their associated retrieved errors). Only high cost and convergence criteria have been applied. In general, the retrieved values of cloud top height have very small errors.

For the water vapour scaling factor, the largest errors occur in the presence of high or thick cloud, which is expected. As shown in Section 3, the cloud model simulations suggested that the retrieval struggles in the presence of high cloud and can on occasion fit spuriously enhanced SO_2, potentially because it results in a poor estimate of water vapour in the correspondingly colder scene. Regions of very high water vapour scaling factor result in very high errors in retrieved SO_2, and data with cloud top height greater than 5 km are not considered reliable for SO_2.

Figure 8 shows nine days of retrieved SO_2 from the 1991 Cerro Hudson eruption following the largest eruption phase on 15th August. The eruption began on 8th August emitting smaller amounts of SO_2 into the upper troposphere lower stratosphere, which can be seen as already present in the path of the main plume on subsequent days. The multiple sampling of the plume by successive orbits (day and night) is quite apparent, particularly as the plume becomes more distorted after 20th August.

4.2 Plume Mass Estimate

The simplest method to estimate the total erupted mass or mass present in a volcanic plume is to take the sum of the representative footprint areas of the satellite that measured SO_2. This method presents several problems relating to sampling of a volcanic plume, particularly with an infrared instrument that measures both night and day that could sample the plume more than once, orbits may partially sample the plume in any one swath and the plume will move constantly between sampling. Alternatively, gridding averages the data into grid boxes on a latitude and longitude grid. Some care must be taken to account for whether or not the gridded data are representative of the data resolution, and keeping track of bins with no data can be a way to estimate under-sampling. Guo et al. (2004) used two methods of gridding data, that of kriging for TOMS data and nearest neighbour interpolation for HIRS/2 (Pratafit method) to account for larger spatial gaps between points. These methods either impose statistical methods or manually introduce information based on assumptions. While both can be utilised in such a way as to indicate an estimate of the error or uncertainty that this introduces, mass estimates presented here are only based on the sum of equivalent contiguous footprints represented by each HIRS ellipse.

Furthermore, if gridding is used, in order to ensure that the data are sampled fairly, the orbits should first be split into ascending and descending nodes, with care taken regarding where a plume is in relation to the date line. This is in an effort to minimise recording the same data point twice when the plume has moved by the time the region is sampled again. Other methods are available but often require a model or further ancillary information.

4.3 Comparative measurements of SO_2

The plume mass estimate for the HIRS/2 SO_2 retrievals for the Cerro Hudson eruption may be qualitatively compared to the figures for TOMS within Constantine et al. (2000). Total erupted mass estimates given can be directly compared, as shown

in Table 3, although the methodology by which the estimates were derived differs. Spatially, HIRS/2 has the advantage of smaller footprint than that of TOMS, (IFOV 1.25° x 1.25°/17.4 km x 17.4 km versus 3° x 3°/ 50 x 50 km) but the TOMS swath is 50% wider (3000 km). For a case such as the Hudson plume, TOMS is more likely to capture the entire plume in one orbit swath and sample it only once, which on the one hand greatly reduces ambiguity in deriving total plume mass but
5   on the other hand the frequency of observation is reduced and sometimes only part of the plume is captured. As reported by Constantine et al. (2000), this was sometimes the case, and a 'best' estimate of the TOMS data was used to contribute to the values in Table 3.

The erupted mass estimates given in Table 3 that relate to HIRS/2 are the sum of equivalent footprint areas , from nodes that
10   capture the most of the SO_2 plume present each day. Figures are rounded to reflect probable accuracy. For the total eruptive period, this method has yielded a total erupted SO_2 mass estimate of 2300 kT with an averaged retrieved error of 27 %. This error does not incorporate error that arises from uncertainty in the height of the SO_2 in the forward modelled plume (as demonstrated in Section 3), or error that might arise from discounting pixels where SO_2 was retrieved below the 3-sigma threshold. It does not account for absent scanlines due to instrument calibration, so should be considered a lower limit. As
15   previously discussed, a good estimate of plume height is an unavoidable requirement in SO_2 detection with an instrument with only one channel sensitive to atmospheric SO_2. In the case of this work, height assignment error of ± 1 km introduces a mass dependent bias of between 5 and 20% for a given pixel depending upon where in the atmosphere the plume is located. For TOMS, the approximate error suggested for the total erupted mass estimate is 30% (Krueger et al, 1995, Constantine et al. 2000).
20   The TOMS algorithms used in Constantine et al. (2000) have been recently updated, and a brief comparison is presented here to some initial data from an updated TOMS algorithm. This algorithm exploits the way ozone and sulphur dioxide both strongly absorb UV radiation. The new TOMS algorithm builds on the early heritage of BUV algorithms (Krueger et al., JGR, 1995). These algorithms retrieve both O_3 and SO_2 by taking advantage of the large SO_2/O_3 cross section ratio (CRS) differences in the gas absorbing bands. This approach constructs radiance tables using a forward model that accounts for
25   both the O_3 and SO_2 cross sections. The new algorithm uses the 317 nm channel to retrieve SO_2 (CRS ~ 2.5), the 331 nm channel to retrieve O3 (CRS ~0.15), and the channel at 340 nm to retrieve the spectral dependence, dR/dλ. This methodology further applies a small second order step2 correction that accounts for non-orthogonality between the SO_2 and O_3 channels.

A one week composite of retrieved SO_2 for both instruments is shown in Fig. 9 where SO_2 from the main eruptive phase can
30   be seen circumnavigating the hemisphere. There is clear complementarity between the instruments in terms of absolute amount retrieved and characterisation of the plume. The smaller pixel size of HIRS and more frequent sampling enables the plume to be observed in finer detail; however the wider swath of TOMS frequently captures more of the plume in one swath.

For a more detailed comparison, two orbits during the 1991 Hudson eruption are considered where the plume is almost fully sampled by both instruments, as shown in Fig. 10. The pixels in the region of the plume were also relatively cloud-free or had low cloud during the observation.

5   The geographical bounds considered for the mass estimate are between -53° and -45° in latitude and 10° to 60° in longitude. Using the method of summing over mass and area discussed previously, the mass of the plume represented here by HIRS/2 and TOMS is calculated to be 1398 and 1540 kT respectively, after quality control has been applied. The missing four scan lines due to a HIRS calibration phase that coincide with the plume in the region of high concentration suggests the HIRS estimate is an underestimate. It is apparent that HIRS/2 is potentially more sensitive to lower amounts of SO_2. It is

10   challenging to directly compare the SO_2 retrieved by two instruments with differing footprint sizes. Gridding might offer an alternative method of plume mass estimate, but selection of the most appropriate grid box size relative to the pixels of each instrument coupled with the small size of the plume with a strong SO_2 concentration gradient make it a challenge for such a comparison to be equitable and account for instrument attributes. A comparison involving gridding for a larger eruption (c.f. Pinatubo) would be less problematic.

15

4.4 E-folding time

The e-folding time for erupted SO_2 is a measure of the residency of the material in the atmosphere, and is affected by the height the material reaches and in the case of very large eruptions, the amount itself. It is also affected by wind shear (horizontal and vertical) and humidity, which affects the rate at which the SO_2 is oxidised and sulphate aerosols grow. The

20   measure is more suited to large eruptions (e.g. El Chichón in 1982 or Pinatubo in 1991), in terms of inferring effects upon radiative forcing, about which Miles et al. (2004) and other works are concerned. This is because the amount and height that such eruptions reach in the stratosphere gives the SO_2 sufficient time to become globally mixed, and as such affect the radiative forcing globally. Equation 1 describes the process of exponential decay, where $N(t)$ is a quantity at time t, N_0 is the initial quantity at time $t=0$ and λ is the decay constant.

25    $$N(t) = N_0 e^{-\lambda t} \hspace{10cm} (1)$$

The e-folding time, the time in which the initial quantity is reduced to 1/e of its initial value, is given by the reciprocal of the decay constant. Using approximate values from the mass estimates derived from Fig.9 where the total SO_2 can be said to drop from around 1500 kT (the total mass present on 17th August 1991 associated with main plume) to 500 kT 18 days later, this yields an e-folding time of around 16 days. Two days after the largest plume was erupted is used here to minimise

30   potential obscuration of the plume by the coincident presence of thick ash. In reality the total mass observed does not decay smoothly, but has noise due to the fact that the plume is not always perfectly sampled, and the number of retrieved pixels excluded due to the presence of high or thick cloud or ash varies. The variability of the mass estimates and the associated

retrieval error make only an estimate appropriate for this approach, but it is not considered to be an unreasonable one. If the e-folding time is calculated for the extremes of the retrieved error bounds of the mass estimates, the e-folding time is 10 days at a minimum, and 35 days at its shallowest descent, but these are considered to be overly-generous bounds by this method. This case is complicated by the fact that about 30 % of the SO_2 released by Hudson was erupted over the 7 days before the main eruption on 15th August, making the calculation of the decay subject to further uncertainty. The e-folding time for this SO_2 plume as estimated by Constantine et al. (2000) is around 15 days, but they state that this is algorithm dependent. These estimates are somewhat smaller than the e-folding times for the larger eruptions (e.g. Pinatubo), which is to be expected due to the considerably lower altitude of the Hudson plume. More recently, Carn et al. (2016) estimated the e-folding time of Cerro Hudson to be ~7 days, based on mass estimates from TOMS (Constantine et al., 2000). They attribute this anomalously short e-folding time to the late southern hemisphere winter timing of the eruption. However, since Constantine et al., (2000) estimate nearly twice the initial total mass (4000kT) than that observed by HIRS/2 in this work (and the subsequent TOMS algorithm discussed here) it is possible that the inconsistency in e-folding times could be due to an over-estimate of initial erupted mass from the original TOMS algorithms. Total mass estimates (and therefore e-folding time estimate) would be improved greatly in accuracy if the HIRS/2 instruments aboard NOAA10 and NOAA12 that were also present were used to result in very comprehensive sampling of this eruption.

5 Discussion

This OE column retrieval finds a new total erupted mass estimate for the 1991 eruption of Cerro Hudson of 2300 ± 600 kT from the HIRS/2 instrument aboard NOAA11, where the error is the retrieved error. This does not incorporate any error from plume altitude estimation but the potential impact has been quantified by forward model simulations. This total mass estimate is lower than that of TOMS (Constantine et al. 2000) and that of Carn et al. (2016) but higher than that derived in a similar way using the methodology of Prata et al. (2003) for HIRS/2. Reasons for this include (but are not limited to) differences in sampling, height sensitivity, instrument differences and attributes or accuracies of the forward model or fit employed in SO_2 detection. From the comparison with the new TOMS algorithm, the HIRS/2 results presented here are highly consistent, and further quantitative comparison, for this eruption in particular, is desirable.

The retrieval precision demonstrated in this case study is slightly smaller (~3 DU) than that proposed for the TOMS instrument (6-7 DU). As such, with the increased sampling of the IR instrument it is apparent that HIRS/2 can offer a positive contribution to the atmospheric SO_2 emission record from explosive volcanic eruptions up to and beyond the launch of GOME and other satellites that followed. Moreover, benefits of the optimal estimation approach over and above the more rapid but limited brightness temperature difference method are significant. They include a quantified error on individual pixel retrieved values, latitudinal variation in accuracy, diagnostic indicators of the retrieval performance and goodness-of-fit and treatment of cloud and water vapour consistent to the retrieval of SO_2. When summing mass over a large number of

pixels, the precision that these afford becomes increasingly important. Issues that remain are those endemic to ill posed problems where there is only one piece of information on SO_2 available and only limited information about the height or shape of the profile of a volcanic plume. It is conceivable that further progress might be made by using HIRS/2 aboard NOAA10 and 12 with the addition of the 8.6 µm channel in ash-free pixels.

5

There are clear opportunities for extending this work. In particular, as the HIRS/2 instrument was present aboard a number of the NOAA platform series, and often simultaneously flown (NOAA 10, 11 and 12 were all in orbit at the time of the Cerro Hudson eruption) there is the possibility to fully characterise eruptions with very high temporal sampling. More rigorous methods for interpolation, sampling and gridding the data can also be used to reduce errors in the total mass estimates. The

10 application of further tools such as chemistry transport or trajectory models for understanding plume evolution would be better constrained by the availability of more measurements.

The first HIRS instrument was flown aboard TIrOS-N in 1978, and there are almost continual data available to the present, and for the foreseeable future of the Met-Op series of satellites, enabling a potential dataset spanning 40+ years. Generating an SO_2 dataset for the duration would be an opportunity to maximise the value and legacy of the satellite data. Such a data-

15 set, with an accompanying error covariance estimate could be used as input to a climate model to better assess the effects of large volcanic eruptions on the radiative balance of the atmosphere. For much of the latter half of that period, there are (and will be) other satellite instruments capable of measuring SO_2 in the limb and the nadir, in particular high resolution spectrometers with very much enhanced accuracy and precision, that will provide correlative information about the quality of the HIRS/2 SO_2 column retrievals that may be considered in retrospective terms. There is also a break in the TOMS

20 record during 1995–1996 that can be filled by HIRS/2 estimates.

It would be highly desirable to extend comparisons from this eruption with TOMS SO_2 in general, comparing a longer record by both instruments for other eruptions, since both provide a unique record of SO_2 potentially spanning many decades. Satellite records of this length for climatologically important trace gases are rare, and would also provide further constraint to volcanic SO_2 emissions in coupled chemistry climate models.

25

References

Barton, I., Prata, A., Watterson, I. & Young, S., 'Identification of the Mount Hudson volcanic cloud over SE Australia', Geophys. Res. Lett. 19, 1211–1214, 1992.

5

Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noel, S., Rozanov, V. V., Chance, K. V. and Goede, A. P. H., 'SCIAMACHY: Mission Objectives and Measurement Modes', J. Atmos. Sci. 56, 127–150, 1999.

Carn, S., Strow, L., Souza-Machado, S. D., Edmonds, Y. and Hannon, S., 'Quantifying tropospheric volcanic emissions with
10   AIRS; the 2002 eruption of Mt. Etna (Italy)', Geophys. Res. Lett. 32, L02301, 2005.

Carn, S. A., Clarisse, L., and Prata, A., J.,. "Multi-decadal satellite measurements of global volcanic degassing." Journal of Volcanology and Geothermal Research 311, 99-134, 2016

15   Chahine, M. T., Pagano, T. S., Aumann, H. H. et al., 'AIRS: Improving weather forecasting and providing new data on greenhouse gases', Bull. Am. Meteorol. Soc 87(7), 911–926, 2006.

Campion, R., Salerno, G., Coheur, P.-F., Hurtmans, D., Clarisse, L., Kazahaya, K., Burton, M., Caltabiano, T., Clerbaux, C. and Bernard, A., 'Measuring volcanic degassing of SO_2 in the lower troposphere with ASTER band ratios', J. Volcanol.
20   Geoth. Res. 194, 42–54, 2010.

Clerbaux, C., Coheur, P.-F., Clarisse, L., Hadji-Lazaro, J., Hurtmans, D., Turquety, S., Bowman, K., Worden, H. and Carn, S. A., 'Measurements of SO_2 profiles in volcanic plumes from the NASA Tropospheric Emission Spectrometer (TES)', Geophys. Res. Lett 25, L22807, 2008.

25

Clerbaux, C., Coheur, P.-F., Hurtmans, D., Barret, B., Carleer, M., Colin, R., Semeniuk, K., McConnell, J. C., Boone, C. and Bernath, P., 'Carbon monoxide distribution from the ACE-FTS solar occultation measurements', Geophys. Res. Lett. 32, L16S01, 2005.

30   Coheur, P.-F., Barret, B., Turquety, S., Hurtmans, D., Hadji-Lazaro, J. and Clerbaux, C., 'Retrieval and characterization of ozone vertical profiles from a thermal infrared nadir sounder', J. Geophys. Res. 110, D24303, 2005.

Constantine, E., Bluth, G. & Rose, W, 'TOMS and AVHRR observations of drifting volcanic clouds from the August 1991 eruptions of Cerro Hudson', Remote Sensing of Active Volcanism, AGU Geophysical monograph 116, 45–64, 2000.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F. (2011), The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q.J.R. Meteorol. Soc., 137: 553–597. doi:10.1002/qj.828

Dudhia, A. (2002), 'RFM Software User's Manual'. Issue 4.2, http://www.atm.ox.ac.uk/RFM/sum.html.

ECMWF, 'An improved general fast radiative transfer model for the assimilation of radiance observations.'. ECMWF Technical Memo No. 345, 2001.

Gottwald, M., Bovensmann, H., Lichtenberg, G., Noel, S., von Bargen, A., Slijkhuis, A., Piters, A., Hoogeveen, R., von Savigny, C., Buchwitz, M., Kokhanovsky, A., Richter, A., Rozanov, A., Holzer-Popp, T., Bramstedt, K., Lambert, J.-C., Skupin, J., Wittrock, F., Schrijver, H. and Burrows, J., SCIAMACHY, Monitoring the Changing Earth's Atmosphere, DLR, 2006.

Guo, S., Bluth, G., Rose, W., Watson, I. and Prata, A., 'Re-evaluation of SO_2 release of 15th June 1991 Pinatubo eruption using ultraviolet and infrared satellite sensors', Geochem Geophys Geosys 5(4), Q04001, 2004.

Hofmann, D., Oltmans, S., Harriss, J., Solomon, S., Dreshler, T. and Johnson, B., 'Observation and possible causes of new ozone depletion in Antarctica', Nature 359, 283–287, 1992.

Eisinger, M. and Burrows, J., 'Tropospheric sulfur dioxide observed by the ERS-2 GOME instrument', Geophys. Res. Lett. 25, 41774180, 1998.

JPL, 'Mission and Spacecraft Library (NOAA)'. http://samadhi.jpl.nasa.gov/msl/QuickLooks/tirosnQL.html, 2003.

Karagulian, F., Clarisse, L., Clerbaux, C., Prata, A., Hurtmans, D. and Coheur, P., 'Detection of volcanic SO_2, ash, and H_2SO_4 using the Infrared Atmospheric Sounding Interferometer (IASI)', J. Geophys. Res. 115, D00L02, 2010.

Kearney, C. and Watson, I., 'Correcting satellite-based infrared sulfur dioxide retrievals for the presence of silicate ash', J. Geophys. Res.114, D22208, 2009.

Kerr, J., McElroy, C. and Olafson, R., 'Measurements of ozone with the Brewer ozone spectrophotometer', Proc. Int. Ozone Symp. 1, 74–79, 1980.

Khokhar, M., Frankenberg, C., Roozendael, M. V., Beirle, S., Kuhl, S., Richter, A., Platt, U. and Wagner, T., 'Satellite observations of atmospheric SO_2 from volcanic eruptions during the time-period of 1996-2002', Adv. Space. Res. 36, 879–887, 2005.

Krotkov, N. A., Schoeberl, M. R., Morris, G. A., Carn, S., and Yang, K., 'Dispersion and lifetime of the SO_2 cloud from the August 2008 Kasatochi eruption', J. Geophys. Res. 115, D00L20, 2010.

Krueger, A., 'Sighting of El Chichón sulphur dioxide clouds with the Nimbus 7 Total Ozone Mapping Spectrometer', Science 220, 1377–1378, 1983.

Krueger, A., Walter, L., Bhartia, P., Schnetzler, C., Krotkov, N., I.Sprod and Bluth, G., 'Volcanic sulphur dioxide measurements from the Total Ozone Mapping Spectrometer (TOMS) instruments', J. Geophys. Res.100, 14057–14076, 1995.

Krueger, A., Krotkov, N. and Carn, S., 'El chichón: The genesis of volcanic sulphur dioxide monitoring from space', J. Volcanol. Geoth. Res., 2007.

Lee, C., Richter, A., Weber, M. and Burrows, J. P., 'SO_2 retrieval from SCIAMACHY using the weighting function DOAS (WFDOAS) technique: comparison with standard DOAS retrieval', Atmos. Chem. Phys. 8, 61376145, 2008.

Matricardi, M., The generation of RTTOV regression coefficients for IASI and AIRS using a new profile training set and a new line-by-line database, Technical report, ECMWF. Technical Memorandum 564, 2008.

Matricardi, M., 'A principal component based version of the RTTOV fast radiative transfer model', Q. J. R. Meteorol. Soc. 136 Part A, 1823– 1835, 2010.

Miles, G., Grainger, R. and Highwood, E., 'The significance of volcanic eruption strength and frequency for climate', Q. J. Roy. Meteor. Soc. 130, 2361–2376, 2004.

Miles, G. M., Siddans, R., Kerridge, B. J., Latter, B. G., and Richards, N. A. D.: Tropospheric ozone and ozone profiles retrieved from GOME-2 and their validation, Atmos. Meas. Tech., 8, 385-398, doi:10.5194/amt-8-385-2015, 2015.

5   NOAA (1981), NOAA Technical Report NESS 83: Atmospheric Sounding User's Guide, Technical report, National Oceanic and Atmospheric Administration. NESS 83.

Pieri, D. and Abrams, M., 'ASTER watches the world's volcanoes: a new paradigm for volcanological observations from orbit', J. Volcano. Geotherm. Res. 135, 13–28, 2004.

10

Poulsen, C., Watts, P., Thomas, G., Sayer, A., Siddans, R., Grainger, R., Lawrence, B., Campmany, E., Dean, S. and Arnold, C., 'Cloud retrievals from satellite data using optimal estimation, evaluation and application to ATSR', J. Atmos. Meas. Tech., 2012.

15   Prata, A., Rose, W., Self, S. and O'Brien, D., Volcanism and the Earth's Atmosphere, Geophysical Monograph Series, ed. A. Robock and C. Oppenheimer, Vol. 139, American Geophysical Union, Washington, DC, chapter Global, long-term sulphur dioxide measurements from TOVS data: a new tool for studying explosive volcanism and climate, pp. 75–92, 2003.

Prata, A. and Bernado, C., 'Retrieval of volcanic SO_2 column abundance from Atmospheric Infrared Sounder data', J.
20   Geophys. Res. 112, D20204, 2007.

Prata, A. and Kerkmann, J., 'Simultaneous retrieval of volcanic ash and SO_2 using MSG-SEVIRI measurements', Geophys. Res. Lett. 34, L05813, 2007.

25   Prata, A., Gangale, G., L.Clarisse and Karagulian, F., 'Ash and sulphur dioxide in the 2008 eruptions of Okmok and Kasatochi: Insights from high spectral resolution satellite measurements', J. Geophys. Res. 115, D00L18, 2010.

Read, W., Froidevaux, L. and Waters, J. W., 'Microwave limb sounder measurement of stratospheric SO_2 from the Mt. Pinatubo volcano', Geophys. Res. Lett. 20, 12991302, 1993.

30

Read, W., Froidvaux, L., Santee, M. and Livesey, N., 'Observations of volcanic SO_2 and HCl from Aura MLS'. American Geophysical Union, Fall Meeting 2009, abstract number V24B-03, 2009.

Rodgers, C., "Inverse methods for atmospheric sounding: theory and practice", 1 edn, World Scientific, 2000.

Saunders, R., Matricardi, M. and Brunel, P., 'An improved fast radiative transfer model for assimilation of satellite radiance observations', Q. J. R. Meteorol. Soc. 125, 1407–1425, 1999.

5 Siddans, R., 'Fast transmittance modelling of the MSG and MTG solar channels for cloud retrievals - Final Report, Technical report', STFC. Delivered in response to EUMETSAT RFQ 10/202521, 2011.

Schoeberl, M., Doiron, S., Lait, L., Newman, P. and Krueger, A., 'A simulation of the Cerro Hudson SO_2 cloud', J Geophys Res, 98, 2949–2955, 1993.

10

Smith, W. L., Woolf, H. M, Hayden, C. M., Wark, D., and McMillin, L. M., The TIROS-N operational vertical sounder,Bull. Amer. Meteorol. Soc., 60, 1177-1187, 1979.

N. Theys, I. De Smedt, H. Yu, T. Danckaert, J. van Gent, C. Hörmann, T. Wagner, P. Hedelt, H. Bauer, F. Romahn, M.
15 Pedergnana, D. Loyola, and M. Van Roozendael
Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-309, 2016

Thomas, H., Watson, I., Kearnet, C., Carn, S., Murray, S., 'A multi-sensor comparison of sulphur dioxide emissions from the 2005 eruption of Sierra Negra volcano, Galapagos Islands', Remote Sensing of Environment, 113, 1331-1342.

20

Thomas, H. E. and Prata, A. J.: Sulphur dioxide as a volcanic ash proxy during the April–May 2010 eruption of Eyjafjallajökull Volcano, Iceland, Atmos. Chem. Phys., 11, 6871-6880, doi:10.5194/acp-11-6871-2011, 2011.

Watson, I., Realmuto, V., Rose, W., Prata, A., Bluth, G., Gu, Y., Bader, C. and Yu, T., 'Thermal infrared remote sensing of
25 volcanic emissions using the moderate resolution imaging spectroradiometer', J. Volcanol. Geoth. Res. 135, 75 89, 2004.

Yang, K., Krotkov, N., Krueger, A., Carn, S., Bhartia, P. and Levelt, P., 'Retrieval of large volcanic SO_2 columns from the Aura Ozone Monitoring Instrument: Comparison and limitations', J. Geophys. Res. 112, D24S43, 2007.

30 Yang, K., Dickerson, R., Carn, S., Ge, C. and Wang, J., 'First observations of SO_2 from the satellite Suomi NPP OMPS: Widespread air pollution events over China', Geophys. Res. Lett. Vol. 40, 4957–4962, doi:10.1002/grl.50952, 2013.

Young, S., Manson, P. & Patterson, G. (1992), 'Southern hemisphere lidar measurements of the aerosol clouds from Mt Pinatubo and Mt Hudson', Extended Abstracts: 16th International Laser Radar Conference, Cambridge, Mass. July 20-24th 1, 3–6.
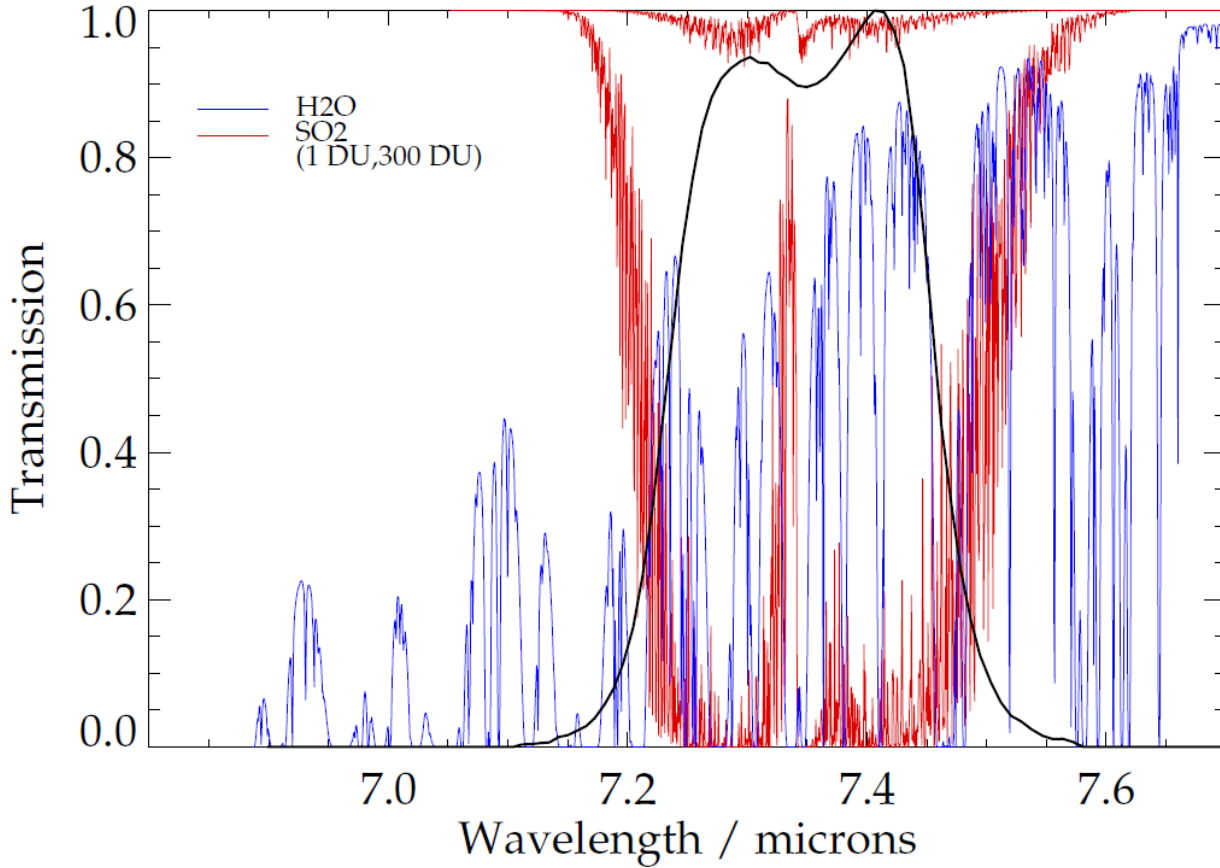
5

Figure 1: Transmission spectra of H_2O and SO_2 simulated from southern hemisphere midlatitude water ECMWF ERA Interim background vapour profile using the RFM (see text). The SO_2 spectra were simulated using triangular profiles to represent column amounts of 1 and 300 DU, as used in the forward model.

10
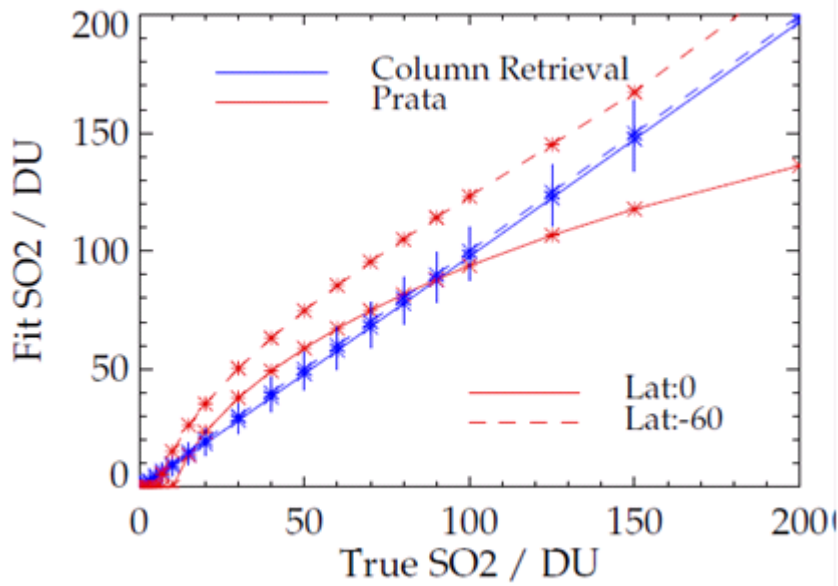
Figure 2: Retrievals based on simulations by a line-by-line model (RFM), with synthetic measurement noise. The error bars for the column retrieval are the retrieved errors. These simulations use temperature and water vapour from a cloud-free ECMWF ERA-Interim atmosphere on 15th August 1991, for a grid box centred at 0°E and both 0°N and -60°N and 0°E. The vertical bars show the retrieved error for the column retrieval. No error estimates are possible for the Prata fit method.
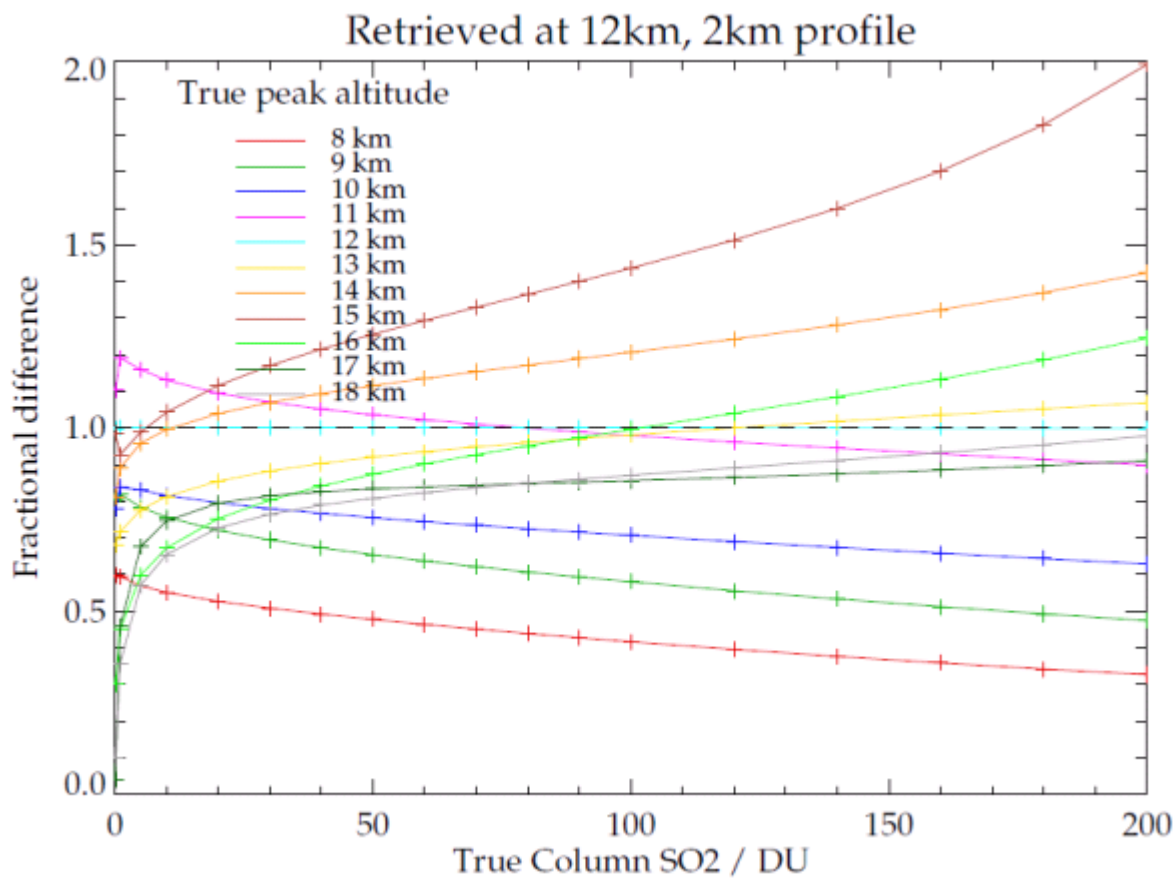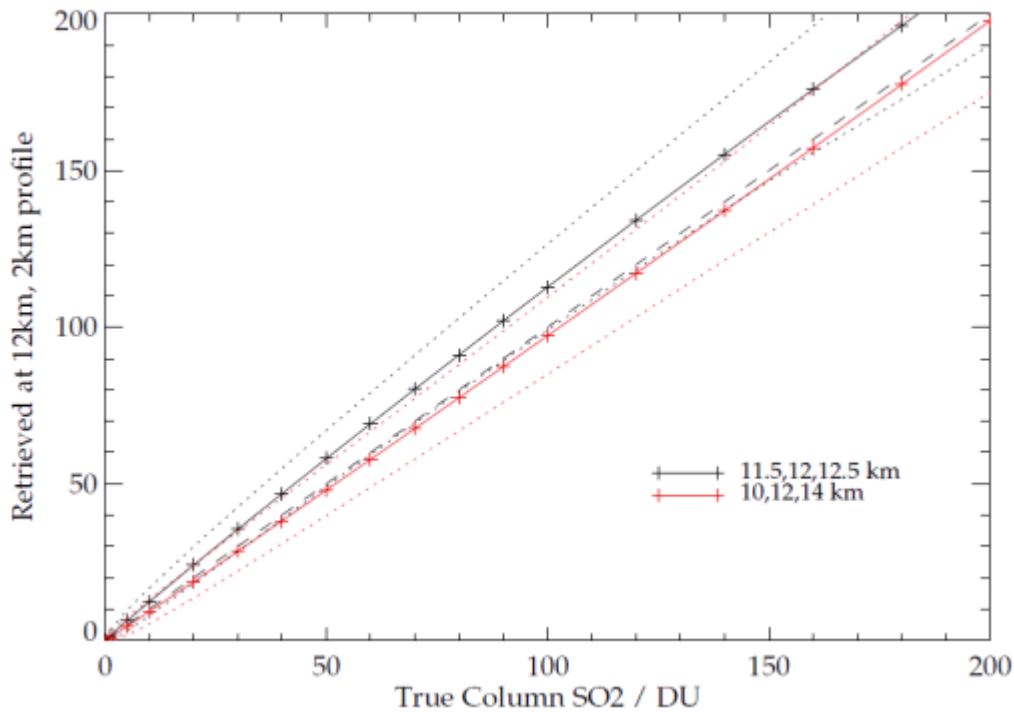
5

Figure 3: A measurement was simulated for a volcanic plume of triangular profile centred at a range of altitudes, for a range of total column amounts. A retrieval is then performed where the plume is assumed to be at 12 km. The fractional difference, or error, is plotted.

Figure 4: The black line indicates how columns from 0.1-200 DU are retrieved on a fixed grid with a scalable triangular profile with base, mid-point and top at 11, 12, 13 km respectively, when the true profile shape is given by a triangular profile at 11.5, 12, 12.5 km, effectively over-estimating the thickness of the plume. The red line shows the equivalent result for an underestimate of the plume thickness, the real profile given by 10, 12, 14 km. The dotted lines show the bounds of retrieved error in each case. The dashed line is x=y shown for clarity.
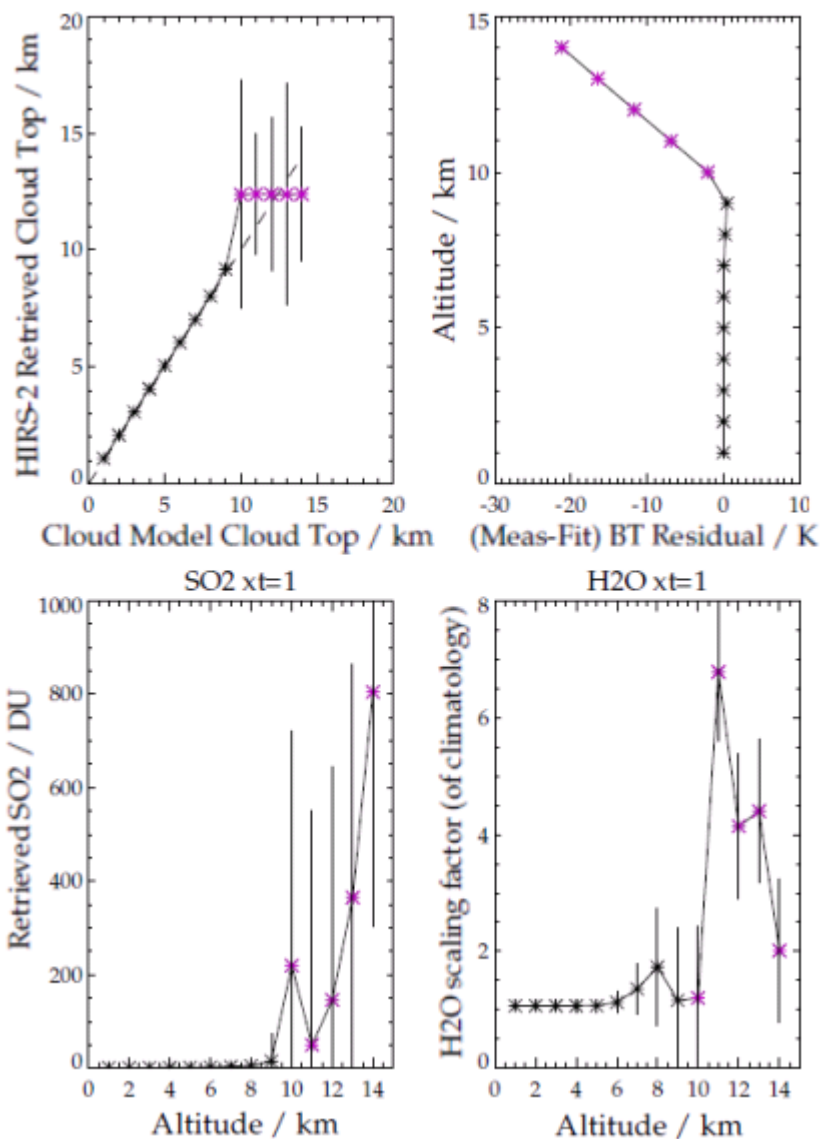
Figure 5. The top left plot shows retrieved cloud top height as a function of 'true' cloud top height as simulated by the cloud model. Black symbols indicate that the retrieval converged and purple indicates that it did not. The top right plot is of the fit residual (measurement minus fit) in the 11.1 μm channel. The bottom left plot shows the retrieved SO_2 as a function of the cloud top height in the cloud model, and the bottom right the equivalent for the water vapour scaling factor.
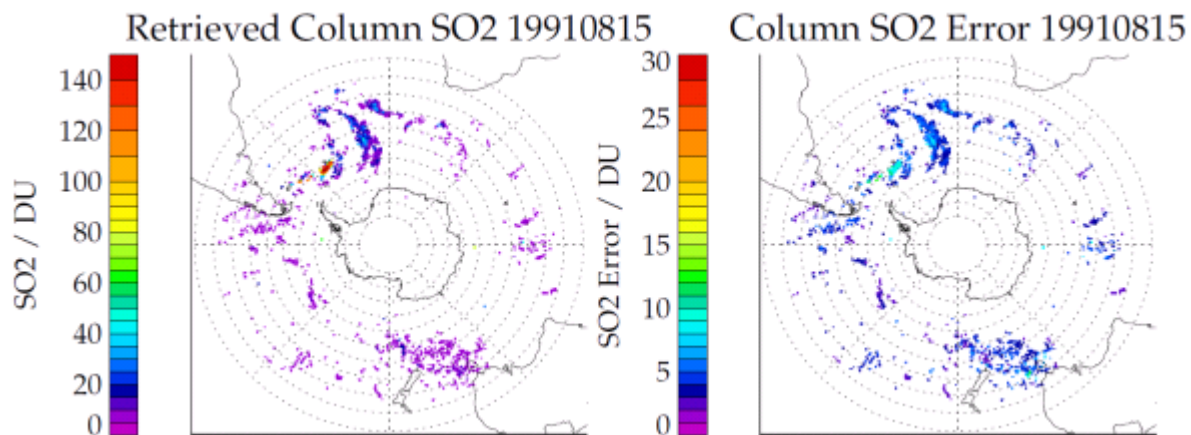
Figure 6: Retrieved SO$_2$ columns for 15th August 1991, and retrieved error for orbits that day. Erupted SO$_2$ from the start of the eruptive phase (from 8th August 1991) is evident ahead of the larger plume emitted on 15th August. Data are screened at the 2-sigma level (5.4 DU).
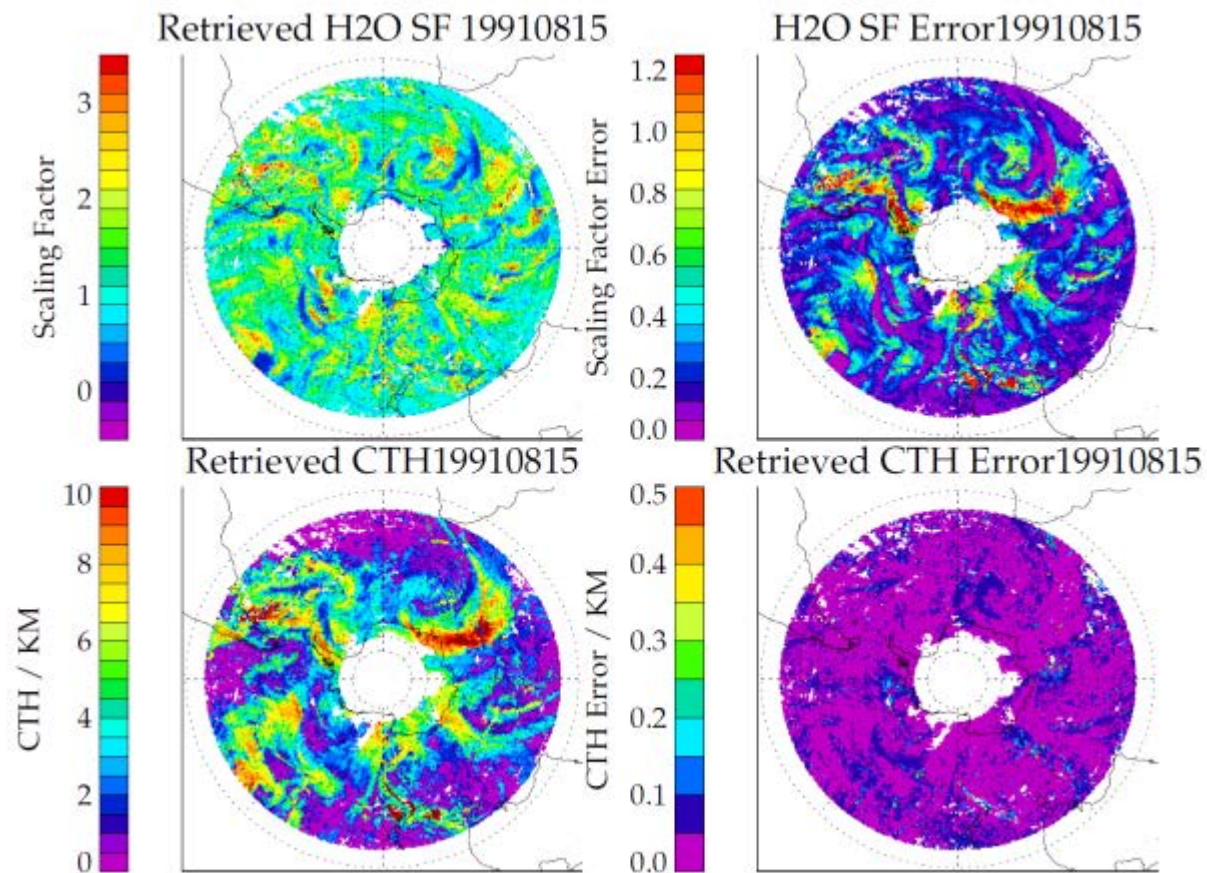
5

Figure 7: The top left and top right show the retrieved water vapour scaling factor and its error from the column retrieval. The bottom left and right the equivalent for the retrieved cloud top height.
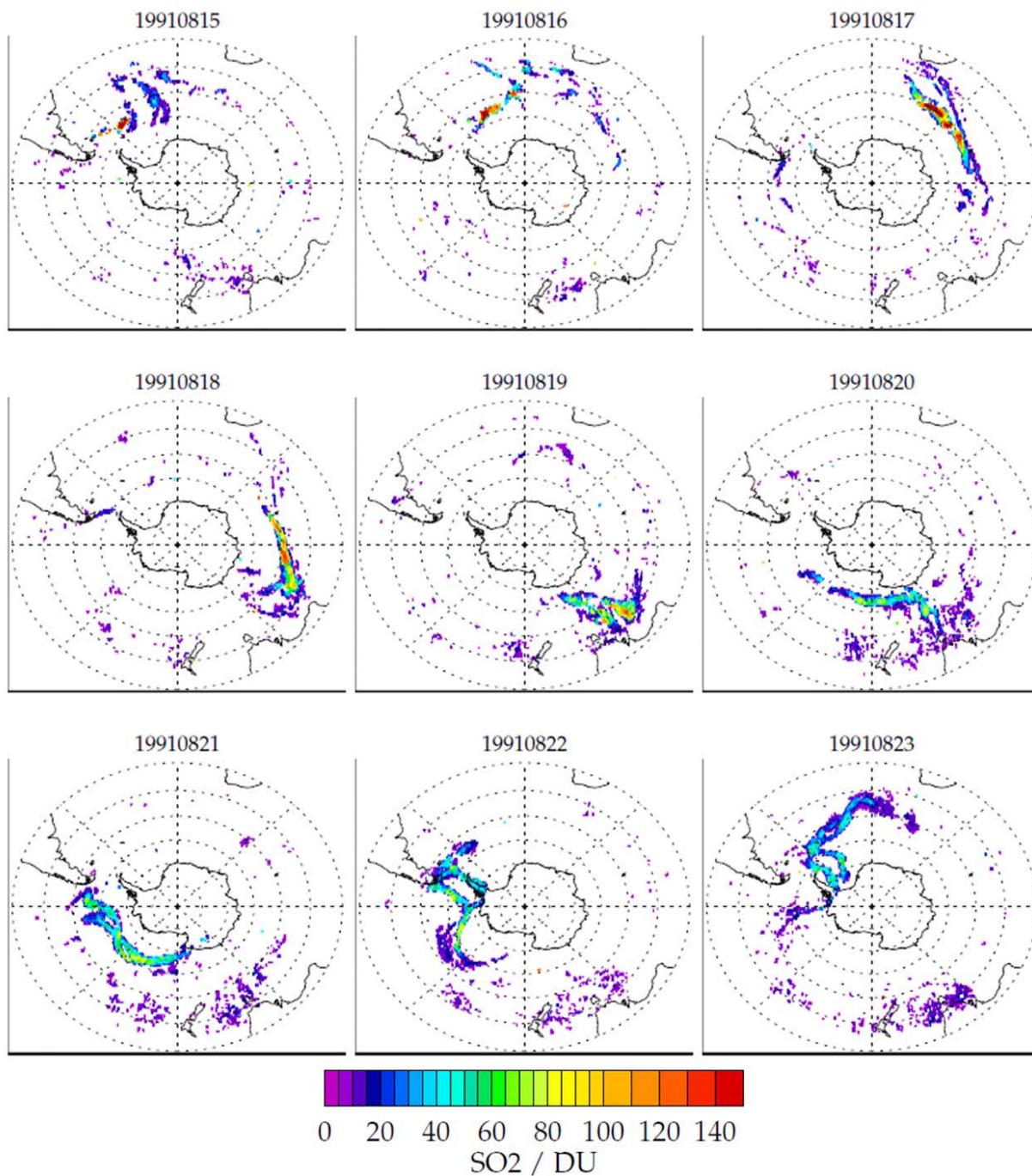
Figure 8: Progression of main erupted plume from 15th August 1991, using all orbits (day and night) from HIRS/2 NOAA11. The eruption began with smaller amounts emitted from 8th August, which are apparent on 15th and disassociated from the main plume. The plume's transport between observations is evident, particularly from 21st August, where it is captured multiple times by multiple swaths. Data have been screened at the 3-sigma level (8.1 DU) for clarity of the main plume.
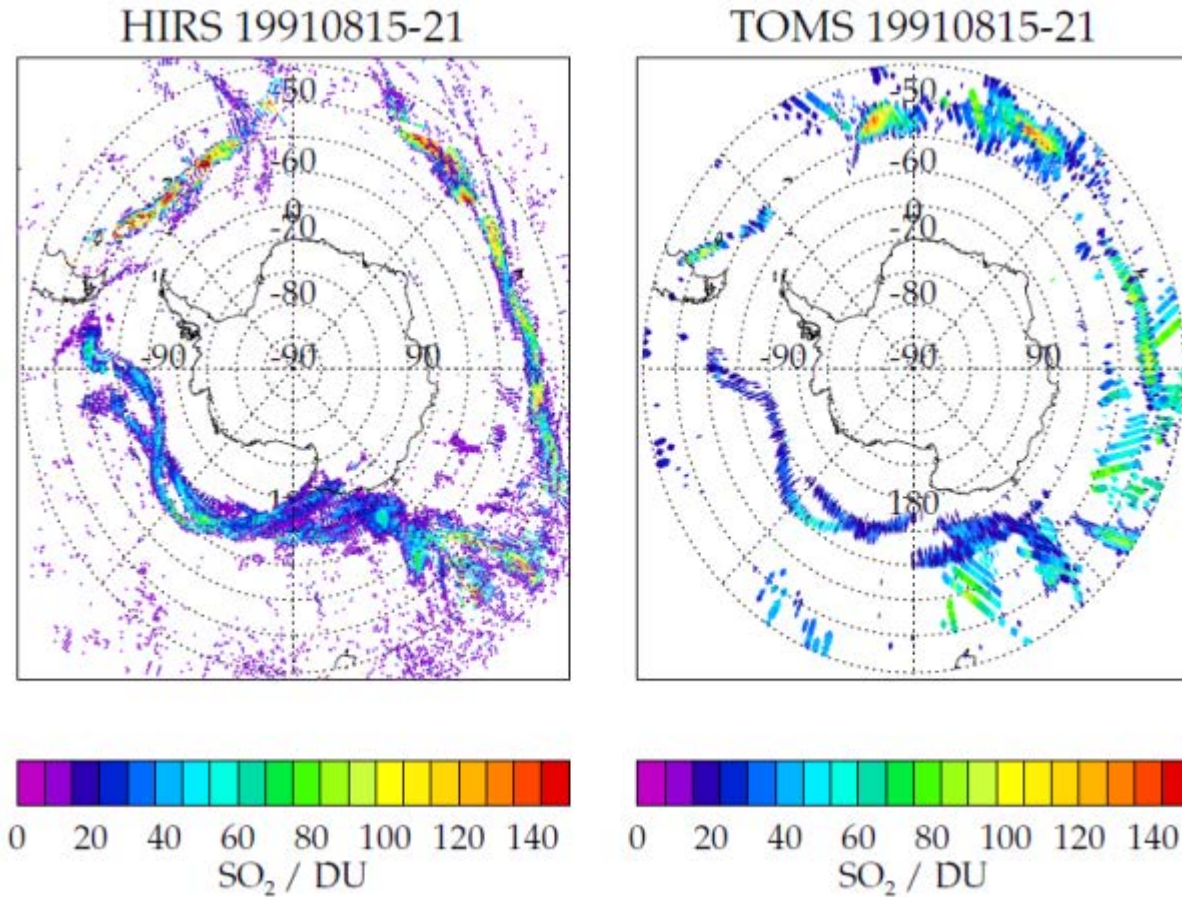
5

Figure 9: Seven day composite of retrieved SO$_2$ from 15-21st August 1991. For clarity in comparison, TOMS data are screened to have a minimum value of 15 DU and HIRS/2 data uses 3 sigma (7.1 DU)
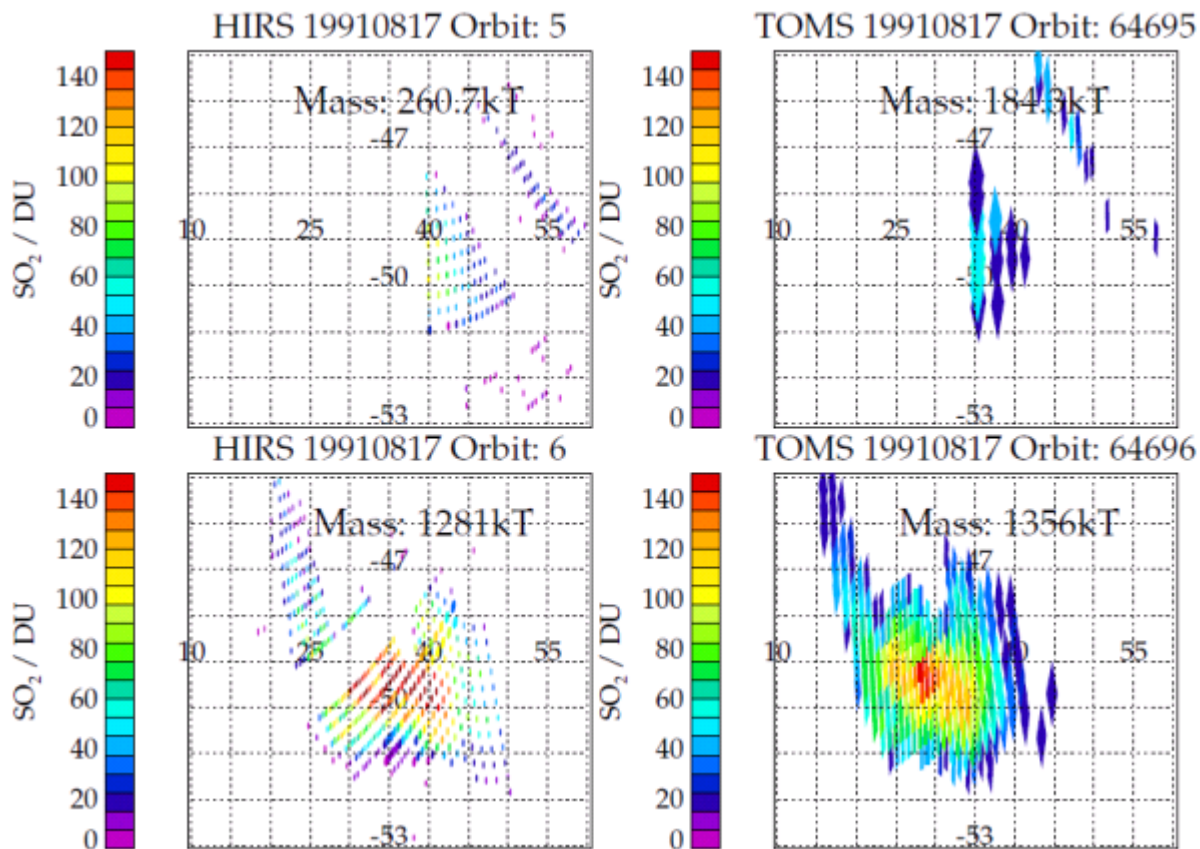
Figure 10: The main Hudson plume on 17th August 1991 as observed in orbits 5 and 6 by HIRS/2 and 64695 and 64696 by TOMS, two days after the main paroxysmal eruption that occurred on 15th August. Four scan lines in the HIRS/2 panels are missing due to routine a calibration phase in which no data are provided. HIRS and TOMS data are both screened at the quality level of 2 - sigma level (5.4. DU and 15 DU respectively).

Table 1. Instruments (many of which were flown aboard several different platforms which are not listed) that have been used to measure volcanic SO₂ in the atmosphere.

Instrument name	Viewing geometry, spectral region	Period of operation	Relevant reference
TOMS, TOMS-like instruments (e.g. SBUV/2)	Nadir, UV	1979+	Krueger (1983), Kerr et al. (1980); Krueger et al., (1995, 2007); Guo et al. (2004)
HIRS/2	Nadir, IR	1979+	Prata et al., 2003, this work.

MLS	Limb, IR	1991+	Read et al., (1993, 2009)
GOME, GOME-2	Nadir, UV-vis	1995+	Eisinger & Burrows (1998); Khokhar et al. (2005); Nowlan et al. (2011); Rix et al. (2011)
ASTER	Nadir, IR imager	1999+	Pieri & Abrams (2004); Campion et al., (2010)
MODIS	Nadir, IR Imager	1999+	Watson et al., (2004)
SCIAMACHY	Nadir/Limb, UV-vis	2002-2012	Bovensmann et al., (1999); Gottwald et al., (2006); Lee et al., (2008)
AIRS	Nadir, IR Spectrometer	2002+	Carn et al., (2005); Chahine et al., (2006); Prata & Bernado (2007); Prata et al. (2010)
TES	Nadir, IR FTS	2004+	(Coheur et al. (2005); Clerbaux et al. (2005, 2008))
SEVIRI	GEO, vis/NIR/IR imager	2005+	Prata & Kerkmann (2007); Thomas & Prata (2011)
IASI	Nadir, IR FTS	2006+	Karagulian et al. (2010)
OMI	Nadir, UV	2006+	Krotkov et al. (2010); Yang et al. (2007)
Suomi NPP OMPS	Nadir/Limb, UV	2011+	Yang et al., (2013)
TROPOMI	Nadir spectrometer UV/vis	2017+	Theys et al., (2016)

Table 2. HIRS/2 Instrument Parameters

Instrument Parameter

Cross-track scan   ± 49.5 ° (± 1125 km) nadir

Number of steps	56	
Optical Field Of View	1.25 °	
Step angle	1.8 °	
Ground resolution IFOV (nadir)	17.4 km diameter	
Ground resolution IFOV (end of scan)	58.5 km by 29.9 km	
Distance between IFOV's	42 km along track and nadir	

Table 3. Total erupted SO_2 rounded estimates for Cerro Hudson

Eruptive Phase	TOMS SO_2[1]	TOMS SO_2[2]	HIRS/2 Prata fit[3]	HIRS/2 OE[4]
8-9th August	700 kT	-	300 kT	500 ± 150 kT
12 August	600 kT	-	400 kT	300 ± 90 kT
15 August	2700 kT	2000 kT	1200 kT	1500 ± 400 kT

[1]Constantine et al. (2000), with errors estimated to be circa 30 %.

[2]This work, based on updated TOMS algorithm, for total mass as observed on 16th August (as region poorly observed on 15th) with consideration of pixel overlap within orbit

[3]After Prata et al. (2003) but data reproduced and sampled as OE HIRS/2 product is herein.

[4]This work, with retrievd error.

Authors' response to Associate Editor's Technical Correction requirements of "Retrieval of volcanic SO_2 from HIRS/2 using optimal estimation" by Georgina M. Miles et al.

1. I am not sure if a reference is possible in the abstract without resolving it in the abstract. Possibly it might be necessary to add the related bibliographic data in parentheses instead of referring to the references section of the paper.

This reference has been removed, since it becomes readily apparent with method is referred to in the text.

2. The term "retrieval error" in line 2 of the Discussion is somewhat ambiguous. In Clive Rodgers' terminology the retrieval error is retrieval noise AND smoothing error, while others use this term in a more generic sense. I suggest to remove the ambiguity by using a more specific wording here.

The retrieval error has been restated as "retrieval noise error".

5

By the way, MIPAS has also been used to measure volcanic SO2 (Hoepfner et al.

Sulfur dioxide (SO2) from MIPAS in the upper troposphere and lower stratosphere 2002–2012

Atmos. Chem. Phys., 15, 7017-7037, 2015.

Thank you for this reference. It has been added to the table and the text.