Can we improve our detection of volcanic clouds by creating more representative covariance matrices?

Rebecca Tichford

Supervisors: Dr. Isabelle Taylor, Dr. Anu Dudhia

August 2024

Abstract

This project aimed to investigate whether the near-real-time detection of volcanic clouds with the Infrared Atmospheric Sounding Interferometer (IASI) on the MetOp satellite platforms could be improved with the creation of more representative covariance matrices. Fast linear retrievals are used to identify elevated levels of sulfur dioxide (SO_2) and volcanic ash in infrared spectra measured by satellite instruments. This technique avoids slow and computationally intensive calculations by using total error covariance matrices to represent the background spectral variability caused by clouds, surface and atmospheric temperatures, instrument noise and the presence of interfering species like water vapour and carbon dioxide in the atmosphere. Currently, the SO_2 retrievals developed by the Earth Observation Data Group at the University of Oxford, use covariance matrices constructed from ensemble spectra from over Europe and the North Atlantic, potentially introducing systematic errors in regions with different climatologies. This project aimed to assess whether more representative covariance matrices tailored to particular land surface types and cloud cover, could improve the accuracy with which SO_2 is detected.

In total 20 covariance matrices were constructed. Analysis found significant variation between the different covariance matrices. The main cause of this variation was suggested to be cloud cover, with smaller contributions from the latitude bands and climate types used to divide the ocean and land surfaces respectively. Some of the covariance matrices were tested within the Oxford linear SO_2 retrieval for the 2021 La Soufrière eruption, with results illustrated as scatter plots. The v_1 SO_2 absorption band retrievals were more affected by the total error covariance matrix used than the v_3 SO_2 absorption band retrievals. In particular, there was a distinct difference in the number and location of pixels flagged as containing SO_2 in the v_3 absorption band retrieval when different covariance matrices were used. Retrievals using covariance matrices constructed from ensemble spectra over the ocean appeared to give more false positives over land, whilst retrievals using covariance matrices constructed from ensemble spectra over the ocean tended to have reduced sensitivity to lower SO_2 column amounts. There were also some small differences between the SO_2 column amounts calculated using different covariance matrices. Future work should investigate the response of retrievals using different parts of the infrared spectra, and assess various eruption scenes to determine if there are certain conditions under which more representative covariance matrices may significantly improve retrieval accuracy.

Contents

1 Introduction and Background

1.1 Why monitor volcanic clouds?

Volcanic eruptions are geological hazards with the potential to have a range of local and global impacts on people and the environment. Emissions of volcanic gases and ash during eruptions can pose an immediate threat to population health and infrastructure [1], whilst potentially also leading to climatic cooling in the longer term [2]. Short-term impacts include the destruction of crops, land and water contamination and respiratory and cardiovascular human morbidity [1].

Volcanic gases contain aerosol precursors such as sulfur dioxide (SO_2). SO_2 is a precursor of sulfate aerosols such as sulfuric acid (H_2SO_4), which is formed via the atmospheric oxidation of SO_2 in either the gas or aqueous phases citet. Sulfate aerosols can both scatter incoming short-wavelength solar radiation and absorb outgoing long-wavelength infra-red (IR) radiation, with the balance between these effects determined by the size of the aerosol particle [1]. Aerosols can also impact radiative forcing by altering both the amount of cloud and the micro-physical properties of clouds [1]. Aerosols act as cloud condensation nuclei (CCN) upon which water droplets may form. A greater amount of aerosols will result in more CCN and an increased total cloud surface area, increasing the light-scattering efficiency of the cloud. This also impacts Earth's hydrological cycle [3].

SO_2 has a greater potential to result in cooling when injected into the stratosphere, where the limited potential for deposition means that sulfate aerosol particles have a longer lifetime. SO_2 is removed more quickly from the troposphere via chemical reactions and dry and wet deposition [1]. During the 1991 eruption of Mt. Pinatubo, 18 Tg of SO_2 were released into the stratosphere [4], leading to the production of around 30 Tg of sulphate aerosols, which subsequently caused sufficient back-scattering of solar radiation to increase the Earth's albedo and cause global mean air temperatures to reduce by 0.5^oC [2]. Volcanic ash can also affect radiative forcing, but these effects tend to be local due to the short atmospheric lifetime of volcanic ash [5].

In the immediate aftermath of a volcanic eruption, one of the industries at risk is aviation. Volcanic ash, containing sharp fragments of rock, mineral particles and volcanic glass, can not only result in reduced visibility and navigation systems failure but can also cause decreased engine performance [6]. Ash particles entering turbine engines melt due to the extremely high temperatures in the combustion chamber, before re-solidifying and accumulating in other parts of the engine, obstructing fuel and cooling systems and leading to a decrease in engine performance or failure [6]. Additionally, H_2SO_4 can damage both the exteriors of aircraft and the interiors of aircraft engines via corrosion [7]. Aircraft vulnerability to volcanic ash and gases was demonstrated by the closures of European airspace following the 2010 Eyjafjallajökull eruption in Iceland. Airspace closures and the subsequent cancellations of over 100, 000 flights caused economic losses of USD 1.7 billion [8].

1.2 Using satellite instruments to monitor volcanic clouds

Real-time monitoring of volcanic clouds is subsequently significant in alerting the aviation industry of volcanic hazards, facilitating timely warnings and the rerouting of flights to minimise risk to aircraft and passengers [8]. It is also important to estimate the total emission of SO_2 which can be used within climate models to provide a first estimate of any climate impacts. Near Real-time (NRT) monitoring can be achieved through the use of satellite observations [8]. Satellite-based monitoring systems have the advantage of facilitating global coverage, in contrast to ground-based monitoring, which can also be dangerous, expensive, and logistically challenging. Satellite instruments commonly use the UV and infrared regions of the electromagnetic spectrum for monitoring volcanic emissions, with infrared instruments being particularly useful because of their ability to operate at night and during high-latitude winters [9].

Despite comprising just 5% of total volcanic gas emissions [1], SO_2 is one of the most commonly measured gases as it has a very low background abundance, aside from a small number of anthropogenic point sources such as metal smelters and power plants [1]. This means that SO_2 is significantly easier to measure than volcanic emissions of species such as water vapour and carbon dioxide which, whilst being emitted in larger amounts during eruptions, are minor contributions relative to atmospheric background levels [10]. Additionally, the lifetime of SO_2 is long enough to enable detection but short enough to prevent significant atmospheric accumulation, which would interfere with the measurement of emission

events. SO_2 has distinct spectral signatures in both the UV and infrared regions of the electromagnetic (EM) spectrum. In the infrared these spectral signatures occur as three absorption bands illustrated in Figure 1 below: the ν_1, ν_3 and $\nu_1+\nu_3$ bands [1].

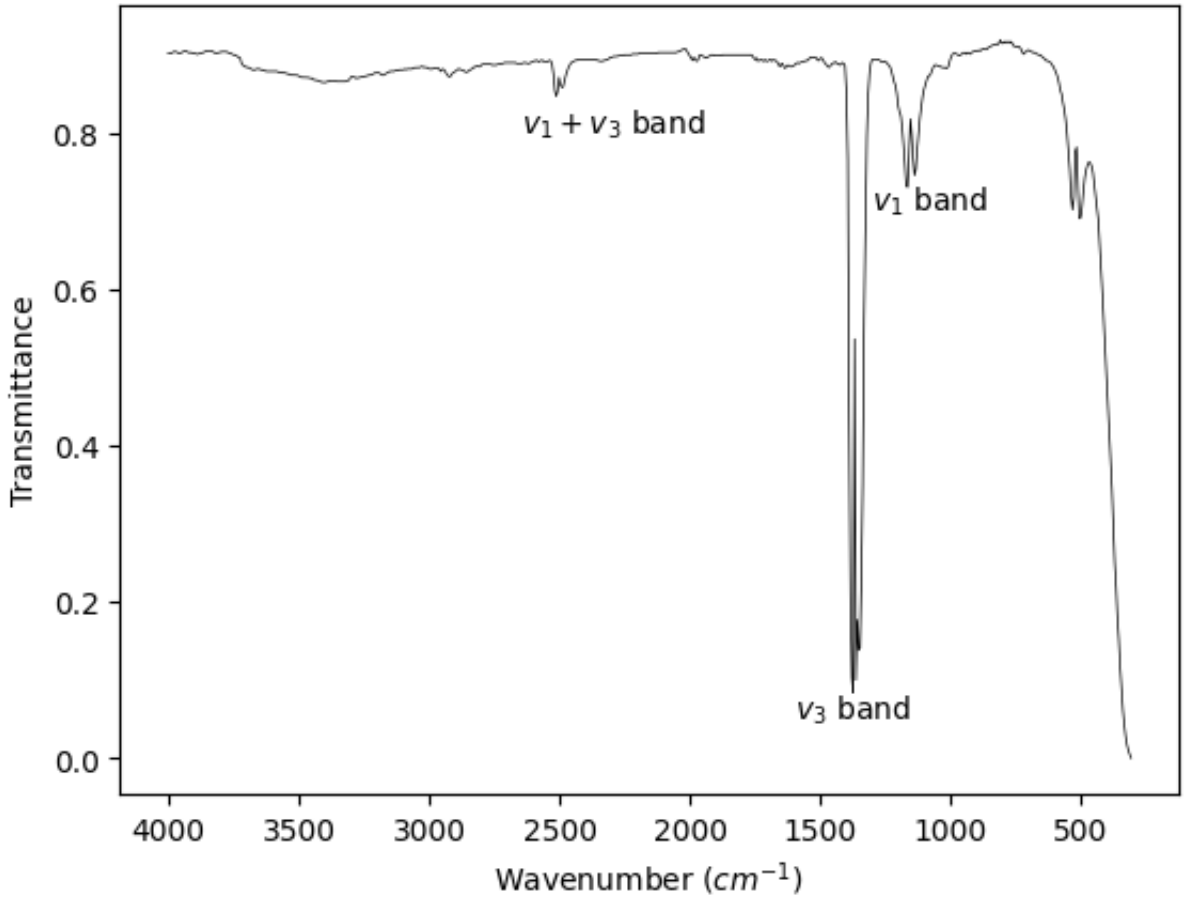

Figure 1: SO_2 IR spectra with the ν_1, ν_3 and $\nu_1+\nu_3$ absorption bands annotated. Spectral data is sourced from the National Institute of Standards and Technology (NIST) Atomic Spectra Database [11]

As can be seen in Figure 1, the anti-symmetric stretch ν_3 band at 7.3 μm (1362 cm^{-1}) is the strongest absorption feature. However, this feature is located within a region of strong water vapour absorption (see Figure 2) resulting in reduced sensitivity to SO_2 in this band in regions of the atmosphere with large quantities of water vapour (e.g. the Earth's boundary layer) [9]. The symmetric ν_1 band at 8.7 μm (1152 cm^{-1}), whilst being a smaller feature, is located in an atmospheric window, though it can be affected by the presence of both volcanic ash and sulfate aerosols. The final band, the combination $\nu_1+\nu_3$ band at 4 μm (2500 cm^{-1}), is the weakest feature and can be affected by reflected solar radiation. Consequently, this band tends to be used only when other channels are saturated by exceptionally large quantities of SO_2 [12].

Successfully flagging pixels which contain elevated amounts of SO_2 requires the separation of the SO_2 signal from the signals of unwanted parameters that may affect the spectral background such as interfering species like water vapour; cloud; and atmospheric and surface temperature. This is achieved using a fast linear retrieval (outlined in section 2.2), a method developed by Walker et al. [13] for the Infrared Atmospheric Sounding Interferometer (IASI; described in further detail in section 2.1). This method utilises a total error covariance matrix formed from an ensemble of SO_2-free pixels. Covariance is a statistical measure of how two variables change together, with the total error covariance matrix containing information about the relationship between the different spectral channels of the IASI instrument. The total error covariance matrix represents the spectral variability due to the unwanted parameters, which consequently do not have to be retrieved themselves. Instead, the effects of these variables are approximated as noise. This means that the spectra of pixels which do contain SO_2 can be quickly distinguished, with the linear retrieval returning an effective SO_2 column amount. This enables the scanning of large

data sets in near real-time; this has important applications in hazard assessments, and can also be used to filter for significant volcanic events to be analysed further.

Ideally, the covariance matrix used in a retrieval process should be calculated from an ensemble of SO_2-free spectra from the same geographic location and in the same meteorological cloud conditions. This is because much of the spectral variability observed by IASI in different regions is due to variations in IR long-wave Planck emission by different surface types and temperatures. For example, a desert region has very different emission spectra than an ice sheet. The level of cloud cover also plays a key role, as the presence of opaque clouds may prevent some of the spectral features associated with surface characteristics from being observed in the spectra obtained by IASI. In practice, a general covariance matrix has been used. The general covariance matrix used in the Walker method was formed from pixels over the North Atlantic and Europe [13], [9]. However, its use may produce systematic errors in regions with significantly different climatologies. This study aims to assess whether using a more specialised covariance matrix based on a specific land surface type and cloud cover at a certain time of year could improve the accuracy with which the Fast Linear Retrieval calculates the effective SO_2 column amounts. Were this to be the case, an eventual aim would be to construct several different covariance matrices which represent different conditions to use during different times of the year and in different geographical regions.

2 Methodology

2.1 The Infrared Atmospheric Sounding Interferometer (IASI)

The spectral data used in this project is from IASI [14], a hyper-spectral cross-nadir Fourier Transform interferometer on board the polar-orbiting MetOp A, B and C satellites launched in 2006, 2012 and 2018 respectively. Fourier transform IR spectrometers are based on Michelson interferometers, which split collimated infrared radiation into two beams, reflecting each beam by a mirror before recombining them. One mirror is moved backwards and forwards along the optical axis, creating a path difference and leading to the production of a time-varying signal when the beams recombine. IASI passively detects the reduction in the radiance of the Earth's Planck emission due to atmospheric absorption by molecular vibration bands. As a hyper-spectral interferometer, IASI can detect absorptions by minor absorbing molecules including ammonia and chlorofluorocarbon (CFCs), as well as major absorbers such as water vapour and carbon dioxide, providing more detailed spectra than previous infrared sounders such as the Atmospheric Infrared Sounder (AIRS). The Level 0 (raw data) can be transformed into calibrated absorption spectra by applying an inverse Fourier transform from the time domain to the frequency domain [1]. This is Level 1 IASI data. Retrieval techniques can be executed to provide Level 2 data such as SO_2 column amounts.

Infrared instruments like IASI utilise the Planck radiation emitted by the Earth, the spectrum of which is affected by absorption and interaction due to certain gases in the atmosphere. The wavelength-dependent absorption of infrared radiation by volcanic ash and sulfate aerosols produces a negative difference in brightness temperature [8]. This temperature difference is opposite in sign to that caused by most other meteorological clouds. SO_2 has distinct spectral signatures in the UV and infrared regions of the EM spectrum. Each IASI instrument provides near-global coverage twice a day and has nadir viewing geometry with a resolution of 12.3 km [14], providing continuous coverage in the thermal infrared from wavenumbers of 645-2760 cm^{-1} sampled at a resolution of 0.25 cm^{-1} [14]. This supplies 8461 spectral channels; this large number of channels can be utilised in advanced retrieval techniques. Whilst IASI's predominant purpose was to provide meteorological information, since data became available in 2007 IASI has also been utilised to measure atmospheric composition, facilitating the monitoring of events like volcanic eruptions, wildfires, dust storms and pollution, all of which can impact atmospheric composition and the spectra observed by IASI.

The IASI spectra can be displayed in units of brightness temperatures (K). The range of channels used in this study is 700 to 1400 cm^{-1}. This range was chosen to include the absorption features of both SO_2 and volcanic ash and encompasses two atmospheric windows between bands of water vapour absorption. The presence of atmospheric windows enables the sensing of surface properties and trace species. An example spectra is shown below in Figure 2, with the key absorption bands illustrated. Using this large spectral range allows the covariance to be adapted for the retrieval of column amounts of multiple species. As can

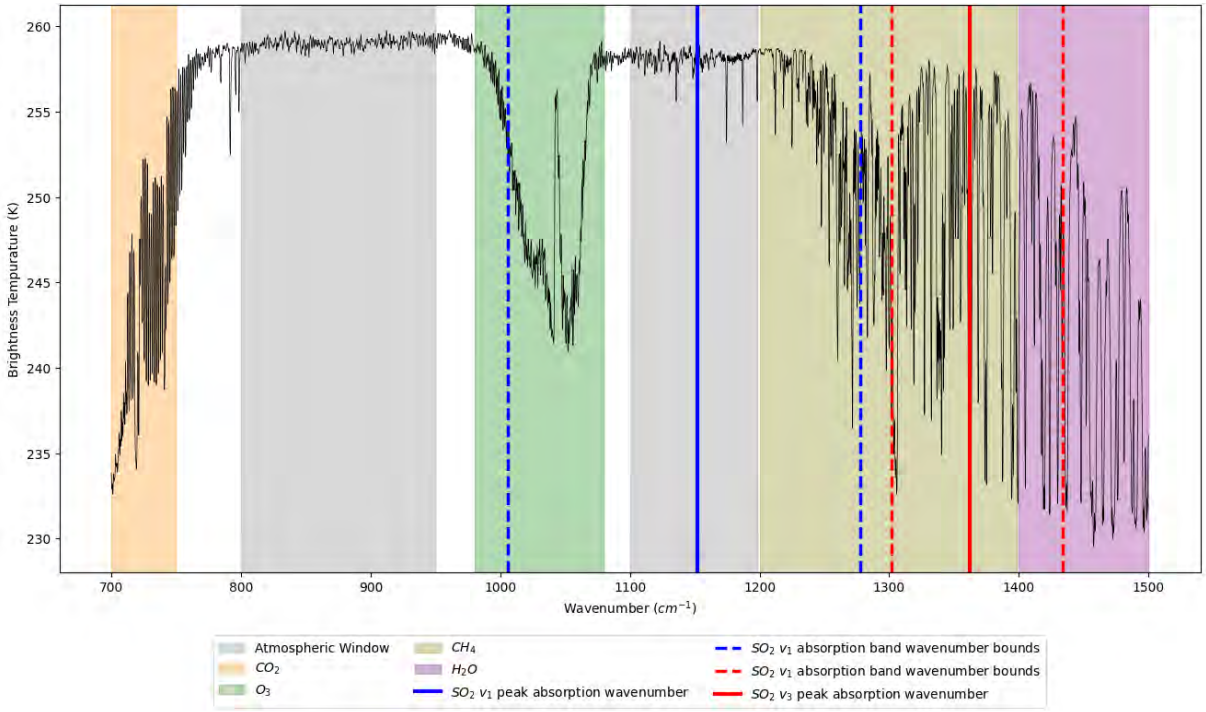

Figure 2: An example IASI spectrum in brightness temperature between wavenumbers of 700 cm^{-1} and 1500 cm^{-1} in clear-sky conditions over the mid-latitude ocean. The main absorption bands and atmospheric windows are illustrated; the SO_2 ν_1 and ν_3 absorption bands are shown by vertical lines blue and red lines respectively. The positions of the SO_2 absorption features were obtained from [15]. The IASI spectral data was made available by EUMETSAT.

be seen in Figure 2, the ν_1 band is located in an atmospheric window whilst the stronger v_3 band is close to the water vapour absorption band and is also collocated with the methane absorption band. IASI can also be used to detect the presence of ash, with the relevant wavenumber regions shown in Figure 3.

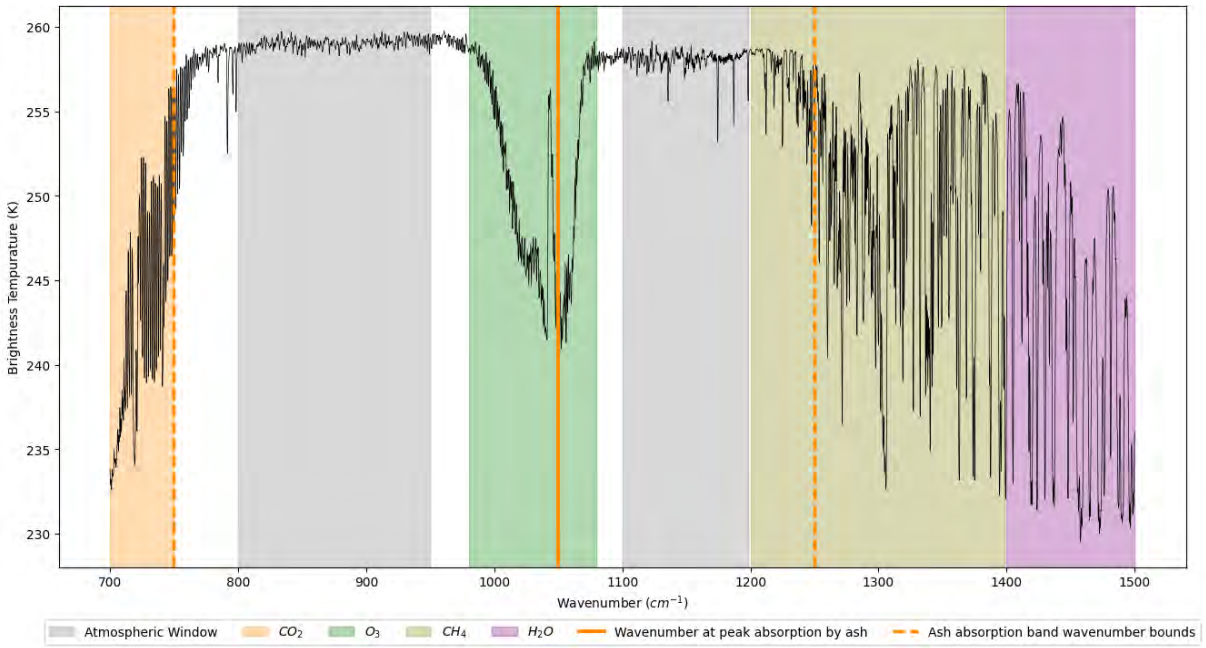

Figure 3: The same example IASI spectra with the ash absorption feature illustrated, shown by vertical grey lines. The wavenumbers of the ash absorption features were taken from Clarisse et al. (2010a; [16]). The IASI spectral data was made available by EUMETSAT.

The ash band is a wider v-shaped absorption feature spanning regions of ozone absorption as well as two atmospheric windows [16].

2.2 The Fast Linear Retrieval Method (Walker et al 2011)

Fast linear retrievals are much more efficient at scanning large data sets than computationally intensive, slow and inefficient fully quantitative retrievals requiring radiative transfer calculations. The fast linear retrieval method was designed to take advantage of IASI's large number of channels and high spectral resolution to provide a Near Real Time (NRT) indication of elevated SO_2 signals in the atmosphere, assisting in determining whether a full iterative retrieval is needed for those pixels [13]. In a given wavenumber channel, changes in the spectral background due to unwanted parameters including interfering species, surface temperature, atmospheric temperature and cloudiness can overwhelm the spectral signal of the target species, in this case SO_2. The Fast Linear Retrieval Method method aims to avoid the direct retrieval of these unwanted parameters.

In the linear retrieval, the spectral measurements made by IASI, $\mathbf{y}$, are represented by a forward model F and the sum of the random and systematic measurement error:

$$y = F(\mathbf{x}, \mathbf{u}) + \xi_{rnd} + \xi_{sys} \tag{1}$$

Where the forward model is a function of the true state vector $\mathbf{x}$ and $\mathbf{u}$. $\mathbf{u}$ is the best estimate of other parameters related to the instrument, atmosphere and the Earth's surface that affect the spectra observed by IASI. $\mathbf{x}$ is constituted of the SO_2 column amount x_c and the wavenumber-independent brightness temperature offset, off. The spectral offset describes uncertainty in $\mathbf{u}$ corresponding to the spectral contributions caused by any differences between the actual and modelled Earth surface temperature, the presence of volcanic ash and the cloudiness of the atmosphere. The random measurement error ξ_{rnd} is the error due to instrument noise. Instrument noise is caused by random variations in the spectra unrelated to atmospheric parameters, and can be due to various effects including electronic and thermal phenomena within the instrument itself. The systematic error ξ_{sys} is due to uncertainties in $\mathbf{u}$ caused by uncertainties in the forward model. The forward model used is the Reference Forward Model (RFM), a line-by-line radiative transfer model that can calculate transmittance through the atmosphere given atmospheric gas profiles[17]. The uncertainties are due to mismatches between the forward model parameters that have been assumed and atmospheric variability in temperature, pressure and composition [17]. In general, the systematic errors are spectrally correlated.

F is linearised about a reference state, the linearisation point, with ambient levels of SO_2 and the best estimate of $\mathbf{u}$.

$$F(\mathbf{x}, \mathbf{u}) \approx F(\mathbf{x_0}, \mathbf{u}) + K(\mathbf{x} - \mathbf{x_0}) \tag{2}$$

The Jacobian K is the matrix of the partial derivatives of F with respect to $\mathbf{x}$, evaluated at $\mathbf{x_0}$. This linearised function can be substituted into Equation (1).

$$y \approx F(\mathbf{x_0}, \mathbf{u}) + K(\mathbf{x} - \mathbf{x_0}) + \xi_{rnd} + \xi_{sys} \tag{3}$$

Next, the total error can be defined as the difference between the spectra observed by IASI (y), and the spectra simulated by the RFM for ambient levels of SO_2.

$$error = y - F(\mathbf{x_0}, \mathbf{u}) \tag{4}$$

From Equations 3 and 4, it can be seen that the total error is given by the following expression.

$$error \approx K(\mathbf{x} - \mathbf{x_0}) + \xi_{rnd} + \xi_{sys} \tag{5}$$

To make an optimal unconstrained least-squares estimate $\hat{\mathbf{x}} = [\hat{x_c}, \hat{off}]^T$ of the true state vector $\mathbf{x}$, a gain matrix G is used to minimise the error.

$$\hat{\mathbf{x}} = \mathbf{x}_0 + \mathbf{G} * error \tag{6}$$

The gain matrix or measurement contribution function G represents the weights allocated to a given spectrum in determining the existence of SO_2. The error term from Equation 4 can then be substituted

into Equation 6 to give an expression for the estimate $\hat{\mathbf{x}}$ of the true state vector $\mathbf{x}$, in this case the column amount of SO_2.

$$\hat{\mathbf{x}} = \mathbf{x}_0 + \mathbf{G}(\mathbf{y} - F(\mathbf{x_0}, \mathbf{u})) \tag{7}$$

Equation 7 shows that the optimal unconstrained least-squares estimate of the column amount of SO_2 is given by the sum of the ambient levels of SO_2 and the measurement contribution function multiplied by the difference between the spectral measurement y and the results of the forward model F at the linearisation point. The gain matrix G is a function of the inverse of the total measurement error covariance matrix, $\mathbf{S}_\xi^{\mathbf{tot}}$, the covariance of the sum $\xi_{rnd} + \xi_{sys}$, and the Jacobian K.

$$\mathbf{G} = (\mathbf{K}^T \mathbf{S}_\xi{}^{tot^{-1}} \mathbf{K})^{-1} \mathbf{K}^T \mathbf{S}_\xi^{\mathbf{tot^{-1}}} \tag{8}$$

The diagonal elements of $\mathbf{S}_\xi^{\mathbf{tot}}$ contain the total random and systematic variance and weight the spectral points against each other, giving less weight to points with a larger total error. The Jacobian K represents the sensitivity of IASI to changes in atmospheric parameters and links the measured spectra to changes in atmospheric temperature and concentrations of trace gases. The RFM is used to simulate atmospheric spectra and produce K, with the i^{th} row of the Jacobian calculated using the partial derivative of the i^{th} simulated spectra in units of brightness temperature with respect to the SO_2 column amount. The derivatives are calculated numerically by applying a perturbation to SO_2 column amounts in the RFM [17] and quantify how changes in SO_2 column amount affect the measured infrared spectra [13].

For the Fast Linear Retrieval to be successful, the total measurement error covariance matrix, which provides information on possible spectral inter-correlation, needs to be as realistic as possible. S_ξ^{tot} can be calculated in two ways, either using perturbation theory in combination with the forward model or via a large ensemble of spectra with ambient levels of SO_2 that are representative of the atmospheric conditions in which the detection is applied. The latter method is used in this investigation. The ensemble method is suitable for volcanic SO_2 as, in the absence of any volcanic events, the signal from SO_2 is substantially below the noise. This is significant as spectra containing enhanced levels of SO_2 mistakenly included in the ensemble can reduce the sensitivity of SO_2 detection. While not investigated in this project, future studies could exclude regions around persistent sources of SO_2 in an attempt to minimise their contribution to the covariance matrix.

2.2.1 Calculating a Covariance Matrix

An ensemble of N measurement spectra are used, with measurements in a range of channels. This range should include areas of the infrared spectrum with both strong and minimal contributions from SO_2 to maximise contrast with the spectral background. Correlations in S_ξ^{tot} mean that all channels used add useful information to the matrix. The number of channels used will be denoted as n_c.
The mean spectrum in each channel $\bar{y}$ should be calculated and stored as an array with dimensions (1,n_c).

$$\bar{\mathbf{y}} = \frac{1}{N} \sum_{i=1}^{N} y_i \tag{9}$$

The total error covariance matrix can then be calculated using:

$$\mathbf{S}_\xi^{\mathbf{tot}} \approx \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{y_i} - \bar{\mathbf{y}})((\mathbf{y_i} - \bar{\mathbf{y}})^T \tag{10}$$

In practice, the most efficient way to calculate this was found to be utilising the method outlined in Smith (2012 [18]) to find the covariance S of N measurement vectors stored in an array $\mathbf{Y}$ [18]. N is the number of pixels. Smith's equations use the matrices with dimensions that are the transposes of those used in this investigation and so the equations have been adapted accordingly.

$$\mathbf{S}_\xi^{\mathbf{tot}} = \frac{1}{N-1} (\mathbf{Y}^T \mathbf{Y} - N \bar{\mathbf{y}}^T \bar{\mathbf{y}}) \tag{11}$$

$\mathbf{Y}^T\mathbf{Y}$ is the mean square matrix and is calculated by taking the matrix multiplication of the transposed spectral matrix and the spectral matrix. In matrix multiplication, each element ij in the resulting matrix is the dot product of the row i in the first matrix and the column j in the second. The spectral matrix and transposed spectral matrix have dimensions (N, n_c) and (n_c, N) respectively. The mean square matrix has dimensions (n_c, n_c), as an m x n matrix multiplied by an n x k matrix will give an m x k matrix.
$\bar{\mathbf{y}}^T\bar{\mathbf{y}}$ is equal to the outer product of $\bar{\mathbf{y}}\bar{\mathbf{y}}$ and has dimensions (n_c, n_c). Consequently $S = S_\xi^{tot}$ will also have dimensions (n_c, n_c) as expected.

2.2.2 Removing pixels containing SO_2

As the covariance matrices must be constructed from SO_2 free pixels, the corresponding linear retrieval data was loaded and used to remove SO_2 flagged pixels from the construction of the covariance matrices. The linear retrieval data were previously generated using the fast linear retrieval method [13]. The present version of this retrieval uses the ν_3 absorption band and so may not successfully remove pixels which contain SO_2 within regions where there is a large quantity of water vapour.

2.2.3 Parameters Investigated

Covariance matrices are calculated from SO_2-free spectra for a number of conditions. Land cover is considered first, with the pixels divided into land and ocean. The ocean pixels are then classified into five latitude bands: 60 to 90 ^{o}N; 30 to 60 ^{o}N; 30 ^{o}S to 30 ^{o}N; 30 to 60 ^{o}S and 60 to 90 ^{o}S. Meanwhile, the land pixels are classified into five biomes based on climate classifications. Finally, cloud fraction is considered. Any pixel with a cloud fraction of greater than 10% is considered to be cloudy, whilst any pixel with a cloud fraction of less than 10% is considered cloud-free. This results in a total of 20 categories. The biome classifications used are sourced from the June 2006 version of the Köppen-Geiger Climate Classification [19]. The climate map has a spatial resolution of 0.5 degrees and is based on data sets from the University of East Anglia Climate Research Unit and the Global Precipitation Climatology Centre at the German Weather Service. The data used is representative of the 50-year period 1951-2000, with classes based mainly on vegetation and surface type. Further subclasses based on precipitation and heat are available but not used within the scope of this study. However, they could be used in the future to improve analysis.

The biome data did not have a regular grid due to containing only land coordinates and so a direct comparison between IASI latitudes and longitudes and biome latitudes and longitudes could not be used to find the closest biome to each IASI pixel. Initially, the distance between every IASI pixel and each biome coordinate was calculated, with the IASI pixel being assigned the biome of the closest pixel. However, this method proved to be too computationally intensive within the timeframes of this study and instead, the latitudes and longitude coordinates of each of the five biomes used were converted into Polygons in Python. Figure 4 illustrates a plot of these Polygons.

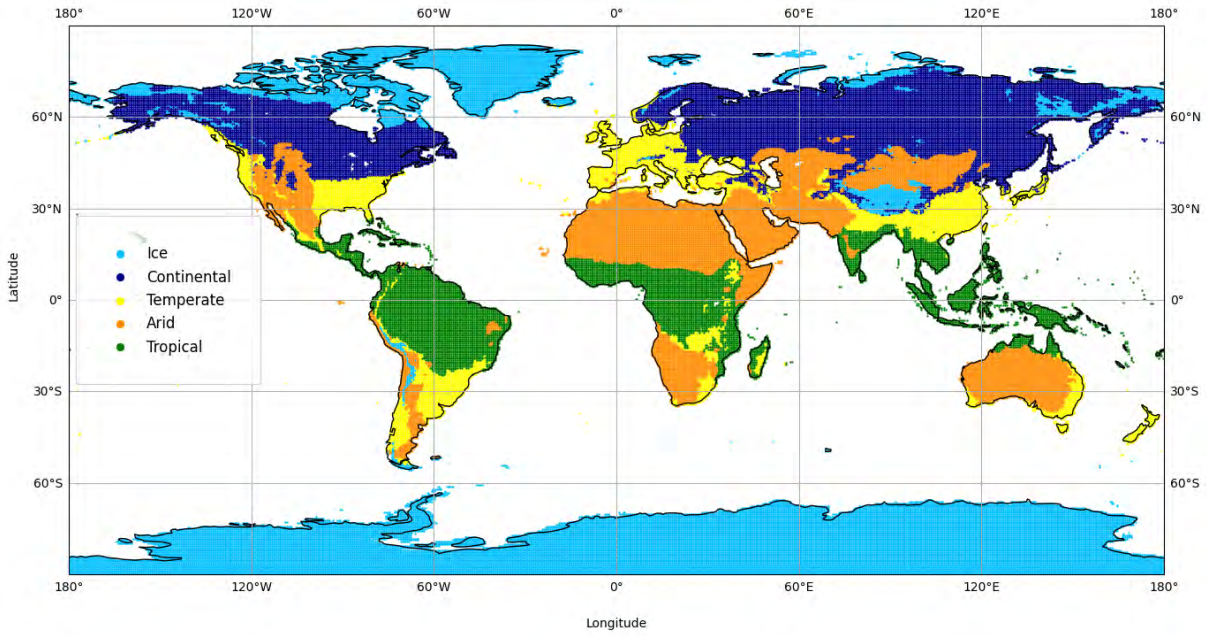

Figure 4: A map of five biomes based on the Köppen-Geiger Climate Classification [19] and created from polygons made using the shapely and geopandas libraries in Python.

Each IASI pixel was assigned to either a biome polygon or an ocean latitude category. The method used had some limitations which should be investigated further: visual inspection showed that some IASI pixels were not assigned to a biome, possibly due to boundary and precision issues. Polygon buffering of 0.05 degrees was used to reduce non-classifications by including pixels within 0.05 degrees of the biome

boundaries, though this may lead to some loss of accuracy and false classifications due to overlapping boundaries. In the future, land-surface emissivity climatologies derived directly from IASI data could be used to provide more accurate surface-type and temperature classifications [20].

This investigation constructed covariance matrices from spectral data from 2020. All the IASI spectral files from 2020 were used, with each pixel sorted into the 20 categories, and the mean square matrix, number of pixels and mean spectrum were calculated and stored for each category for each file. The mean square matrices, pixel numbers and mean spectra for each file were then combined into totals for each category. The 20 covariance matrices, one for each category, were then calculated.

2.2.4 Constructing Gain Matrices and finding SO_2 Column Amounts

When the ensemble method is used, the mean spectrum $\bar{\mathbf{y}}$ calculated in Equation 5 should be used in place of the forward model term in Equation 2 as it is likely to be a more representative estimate of the climatological mean spectrum. This means that equation 7 becomes:

$$\hat{\mathbf{x}} = \mathbf{x}_0 + \mathbf{G}(\mathbf{y} - \bar{\mathbf{y}})) \tag{12}$$

The gain matrix G is calculated using Equation 8.

Column amounts of SO_2 are calculated in Dobson units (DU), a unit measuring the amount of a particular gas in a vertical atmospheric column. The magnitude of a gas in DU can be calculated from a column amount given in molecules per cm^2 using the equation below:

$$1DU = 2.69 \times 10^{16}\ molecules\ cm^{-2} \tag{13}$$

The total column amounts $\hat{x_c}$ of SO_2 retrieved in the Walker method are not accurate retrievals of the SO_2 column amounts; rather they are used to establish whether levels of SO_2 are enhanced with regards to ambient atmospheric levels. If column amounts of SO_2 at a certain pixel are above a threshold value, those pixels are 'flagged' as containing SO_2 and a full quantitative retrieval may be used.

2.3 Applying the Representative Covariance Matrices to the Fast Linear Retrieval for the La Soufrière 2021 Eruption

La Soufrière volcano on the Caribbean Island of St. Vincent entered a phase of explosive eruptive activity during April 2021, with multiple explosive events occurring from the 9^{th} to the 22^{nd} April during which time significant ashfall affected much of St Vincent and many neighbouring islands [21]. A minimum estimate of the total emission of SO_2 to the atmosphere during this time period was calculated to be 0.63 $\pm$ 0.5 Tg based on IASI measurements [7].

In this study, four of the covariance matrices created are used in both the ν_3 and ν_1 SO_2 absorption band fast linear retrievals to facilitate the comparison of the number of flagged pixels. The four covariance matrices used are those calculated for the Cloud Free Ocean Latitudes -30 to 30°, Cloudy Ocean Latitudes -30 to 30°, the Cloud Free Tropical Land Surface and the Cloudy Tropical Land Surface. Maps of SO_2 column amounts obtained during each retrieval are plotted. These categories of covariance matrices were chosen to mirror the geographical location of the La Soufriére eruption and cloud.

3 Results

3.1 Covariance Matrices

Mean square matrices, mean spectra, and total number of pixels were calculated for the 20 surface type and cloud cover categories for a year (2020) of IASI data from all three instruments. This enabled the calculation of 20 covariance matrices representative of the background spectral covariance of each category. Figure 5 below illustrates 20 example covariance matrices calculated from 10 files in February 2020.

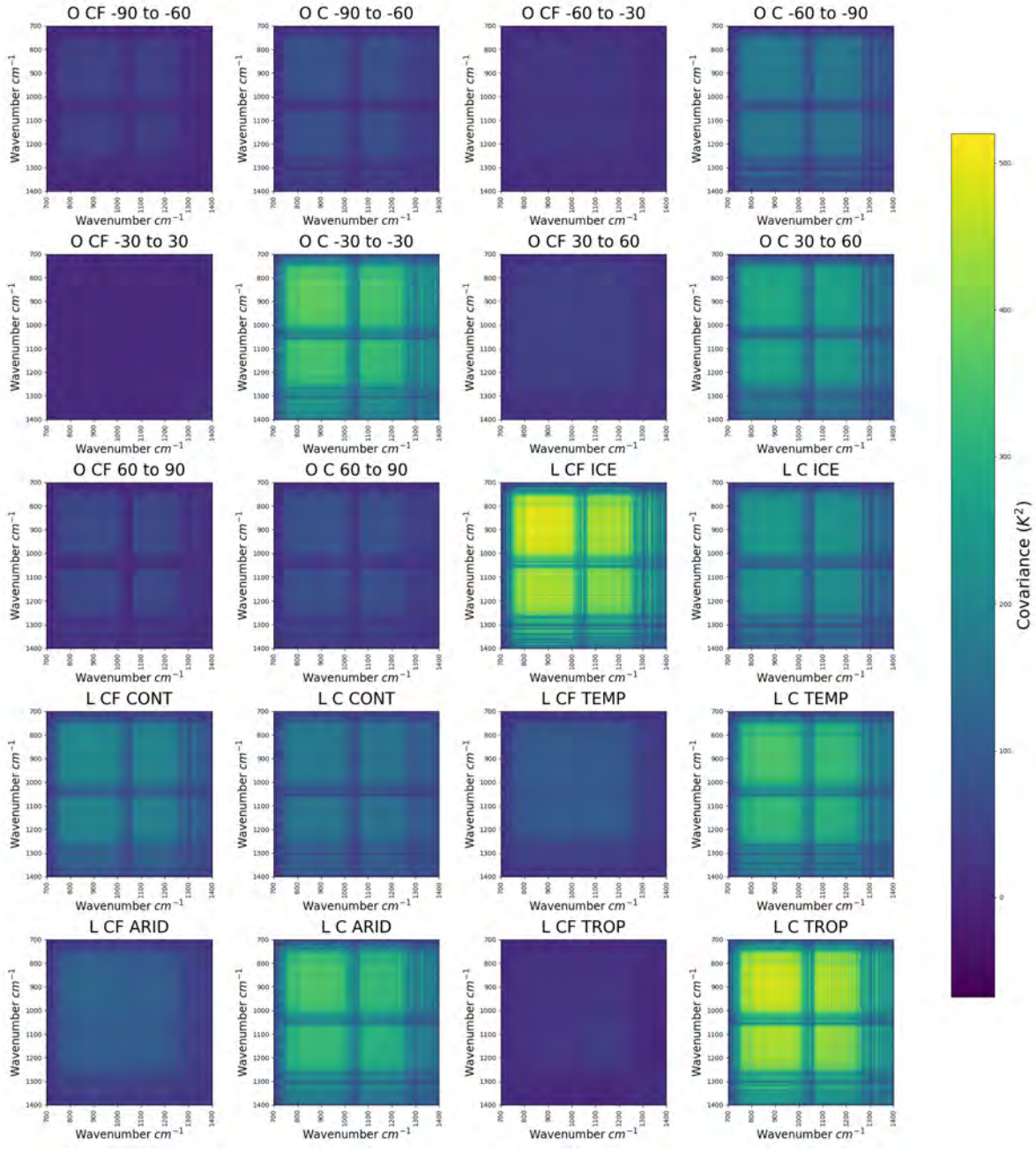

Figure 5: Covariance matrices for wavenumbers between 700 and 1400 cm^{-1} for each of the 20 land surface type and cloud categories created using 10 IASI-a files from February 2020
L: Land O: Ocean C: Cloudy, CF: Cloud Free

The diagonal elements of the covariance matrices represent variances (the covariance between the same wavenumber channel), whilst the off-diagonal elements represent covariance between different spectral channels. Several features can be observed in these matrices. Firstly, for each matrix, regions of higher

covariance are observed in channels corresponding to the atmospheric windows between the CO_2 and O_3 absorption bands (800 - 950 cm^{-1}) and between the O_3 and H_2O absorption bands (1100 - 1400 cm^{-1}). In atmospheric window regions, the spectrum is dominated by the Planck emission from the Earth's surface. Variations in the Plank emission affect the whole spectrum, resulting in a higher covariance between channels. In contrast, there are areas of lower covariance with other channels in the absorption bands. In regions of absorption, the photons detected originate from the atmosphere itself. This is because of the absorption and re-emittance of specific wavelengths of infrared radiation from the Earth by absorbing molecules. For example, in the O_3 absorption band, the photons detected will originate from the stratosphere where there are high concentrations of long-wavelength absorbing ozone molecules, rather than from the surface. This results in the central cross feature in each covariance matrix. Similarly, in the water vapour absorption bands, where transmittance is minimal, the photons reaching IASI originate from the point in the atmosphere at which the water molecules are located. This results in the reduced covariance along the bottom and right of each matrix.

When comparing the matrices calculated for different categories, there is a noticeable difference in the covariance between spectral channels for the cloudy and cloud-free matrices. Spectra of pixels classified as 'cloudy' are found to have higher covariances. As clouds tend to have a uniform effect over a range of wavelengths, many spectral channels will exhibit similar features. For instance, clouds can cause strong scattering in multiple spectral channels, leading to higher covariance between them. This effect is compounded by the fact that when clouds are present in the atmosphere, the thermal emission detected in multiple channels will be influenced by the same cloud properties, such as cloud top temperature and composition, leading to a higher covariance between channels. The exception to this is the cloud-free covariance matrix for the ice biome, which exhibits very high covariance between spectral channels. This could be due to the different emissivity properties of ice compared to other surfaces. The cloud-free ocean covariance matrices for 60 to 90^oN and 60 to 90^oS display similarly higher covariances than the cloud-free ocean covariances at lower latitudes; this could also be due to the presence of sea ice.

Additionally, spectra over land regions tend to have higher covariance than spectra over ocean regions. Land surfaces experience more significant diurnal temperature fluctuations, resulting in more complex and variable infrared spectra, whilst the higher heat capacity of the ocean means that the spectra detected above oceans tend to be more consistent.

3.1.1 How different are the covariance matrices for various land surface and cloud fraction categories?

In order to ascertain whether the use of more representative covariance matrices can improve the detection of volcanic clouds, it is first necessary to investigate whether there are significant differences between the covariance matrices constructed from pixels belonging to the various land surface and cloud fraction categories. Initially, the means, minima and maxima of each covariance matrix were compared, with the results displayed in Figures 6 and 7 below.

From a visual inspection of Figures 6 and 7, a significant variation in the covariance matrices can be seen both between the land and ocean covariance matrices and within the different categories. For the ocean covariance matrices, the mean covariances of the 'cloudy' matrices were greater than those of the 'cloud-free' matrices in each latitude band, with the differences greatest for the lower latitude bands. The 'cloudy' covariance matrices also had a greater spread than the 'cloud-free' covariance matrices. Again, the lower latitude bands saw the greatest spread in covariance. Overall, there is a clear indication that the lowest latitudes close to the equator see the greatest covariance in 'cloud-free' conditions, whilst the highest latitudes close to the poles see the greatest covariance in 'cloudy' conditions.

Trends are slightly different for the land covariance matrices, with less distinct trends. Whilst the temperate, arid and tropical surface types saw higher means and greater spread in their covariances matrices for 'cloudy' conditions than 'cloud-free', the opposite is the case for the ice and continental surface types. Overall, the mean and spread of the 'cloud-free' covariance matrices tend to be smaller for land surface types that are, in general, located closer to the Equator. Trends in the mean and spread of the 'cloudy' covariance matrices are harder to identify, with perhaps slightly greater means and spread for the covariance matrices constructed from pixels above land surface types generally located closer to the Equator, such as the tropical, temperate and arid biomes.

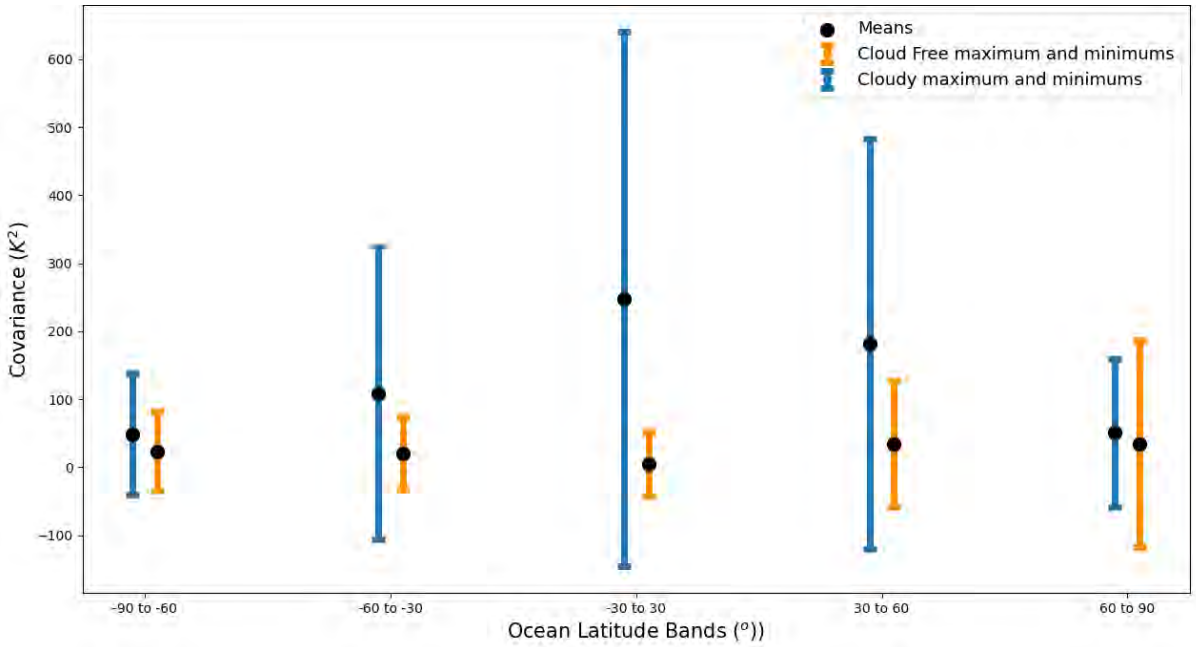

Figure 6: Figure illustrating the mean, maximum and minimum statistics for cloudy and cloud-free covariances for each of the ocean latitude bands

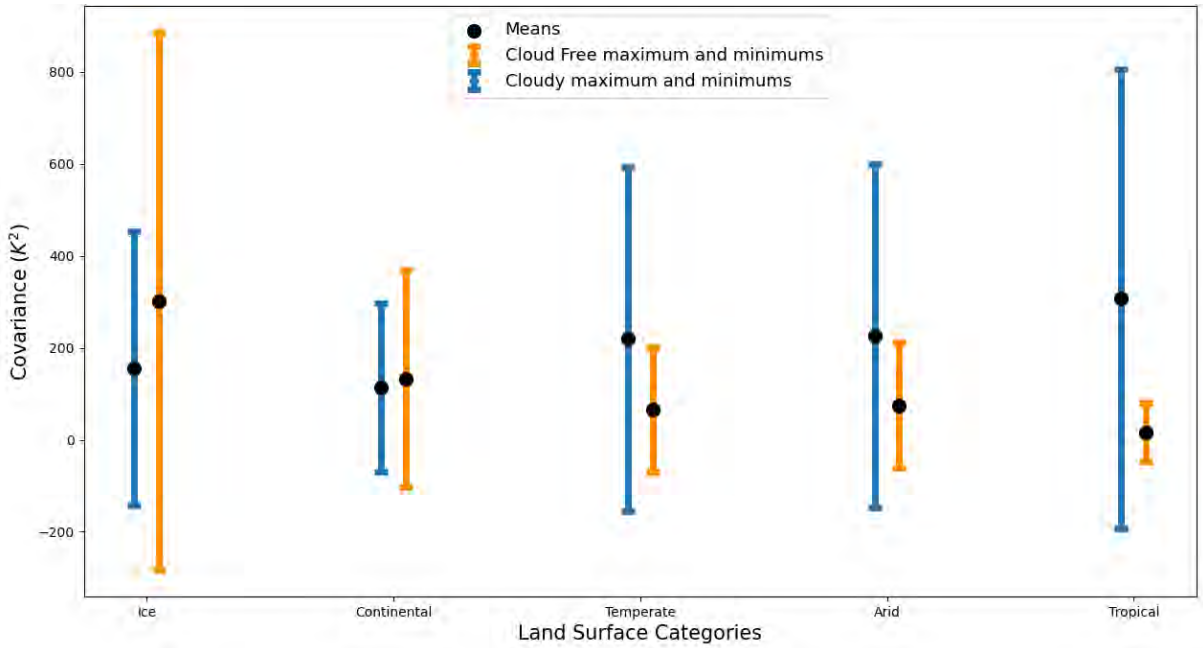

Figure 7: Figure illustrating the mean, maximum and minimum statistics for cloudy and cloud-free covariances for each of the land surface types

For both ocean and land surface categories, cloud cover seems to play a greater role in influencing the covariance matrices than either the latitude band or the land surface type. The exception would be the ice surface type, which appears to have a greatly different covariance to the other land surface types under 'cloud-free' conditions.

To analyse these differences quantitatively, the Frobenius norm was calculated for each of the 20 covariance matrices. The Frobenius norm is a square matrix norm which gives a scalar value; this scalar can be used to compare and quantify the similarity of different covariance matrices. The Frobenius norm of an m x n matrix is found by taking the square root of the sum of the absolute squares of the matrix

elements:

$$|| S_{\xi}^{tot} ||_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} | S_{ij} |^2} \tag{14}$$

The Frobenius norm was calculated for each of the 20 covariance matrices produced for the year 2020. A boxplot of the norms was then produced to quantitatively compare the norms of the different covariance matrices and is shown in Figure 8 below.

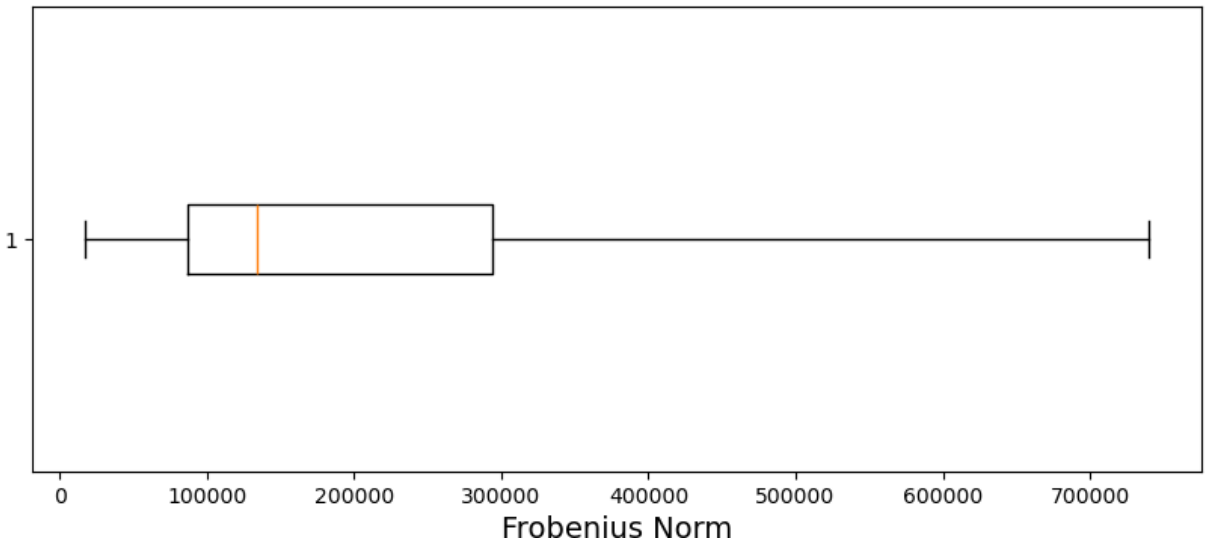

Figure 8: Box Plot of the distribution of the Frobenius norms of the 20 covariance matrices displayed above

The large IQR relative to the total range of Frobenius norms suggests high variability between the covariance matrices for different categories. This analysis confirms that using spectra from the varying land surface type and cloud fraction categories to construct covariance matrices can produce significantly different results.

3.1.2 Investigating how much information we need to calculate representative covariance matrices

In order to determine the maximum number of pixels needed to be added to the calculations to construct a representative covariance matrix, the determinant of the covariance matrix was recalculated after each file, contributing around 1-10,000 pixels, was added. However, this determinant was found to be very small, roughly equal to 0. This indicates a degeneracy in the covariance matrix and suggests that inverting the matrix may lead to large uncertainties in the retrieval.

Correlation matrices were then calculated from the covariance matrices to investigate this further using Equation 15 below.

$$S_{ij,cor} = \frac{S_{ij,cov}}{\sqrt{S_{ii,cov} S_{jj,cov}}} \tag{15}$$

These correlation matrices can be plotted.

As the covariance matrices are symmetric matrices, a property in which the matrix itself is equal to its transpose, the diagonal elements of the correlation matrices are equal to 1. The correlation matrices plotted in Figure 9 show a very high correlation. This was somewhat unexpected as the noise of the instrument is thought to reduce this correlation to a higher degree.

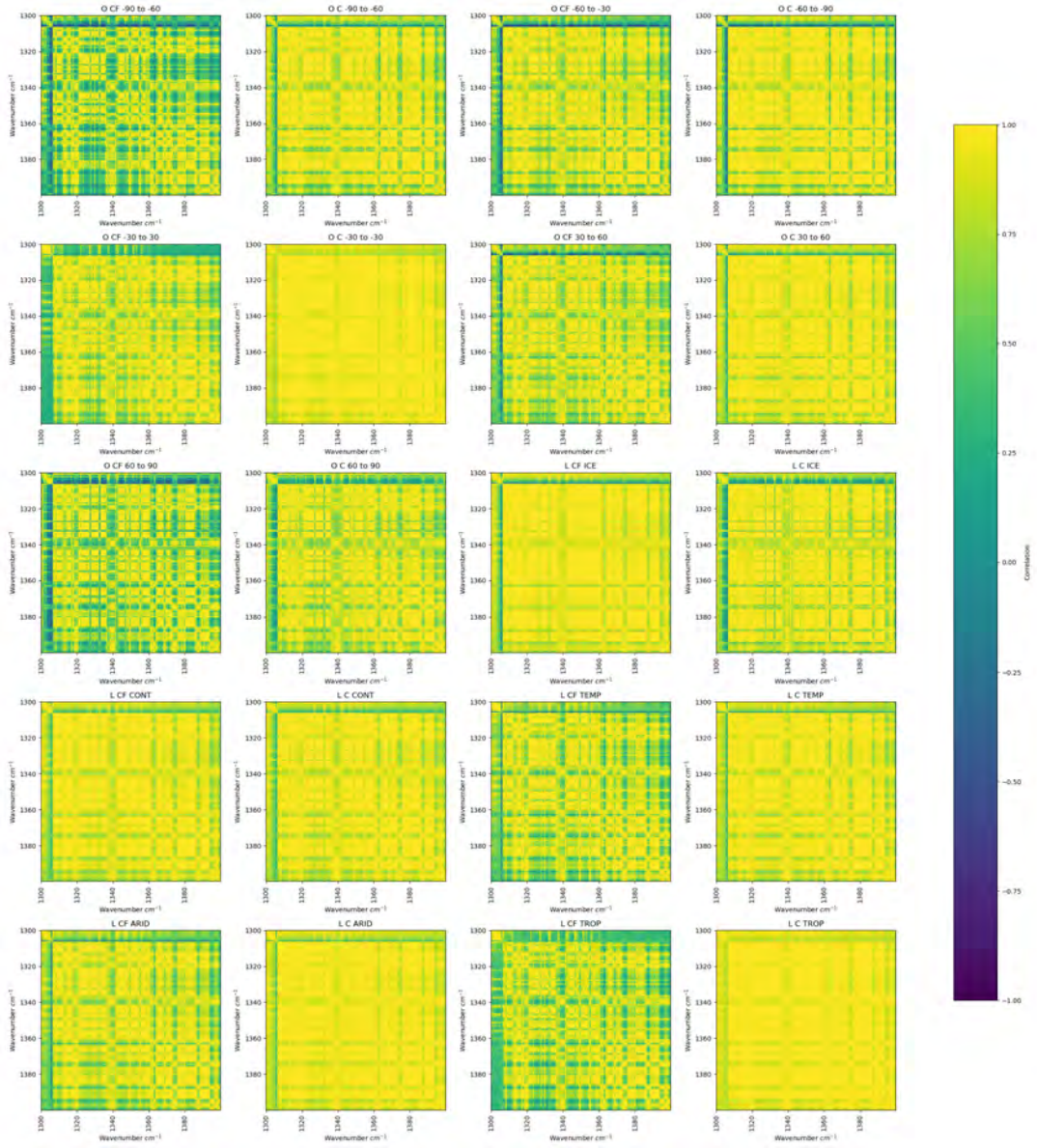

Figure 9: Correlation matrices for each of the 20 land surface type and cloud categories created using 10 IASI-A files from February 2020

To avoid the issue of the near-zero determinants, other information was used to determine the number of pixels required to construct a representative covariance matrix. As each new file was added to the covariance matrix calculation, the difference between the new total error covariance matrix and the previous total error covariance matrix was calculated. The maximum and mean differences of this difference matrix were plotted against the number of pixels used in the calculation. A logarithmic y-scale was used to clearly show the levelling off of the mean and maximum differences, indicating that there was little change in the total error covariance matrix with the further addition of new spectra.

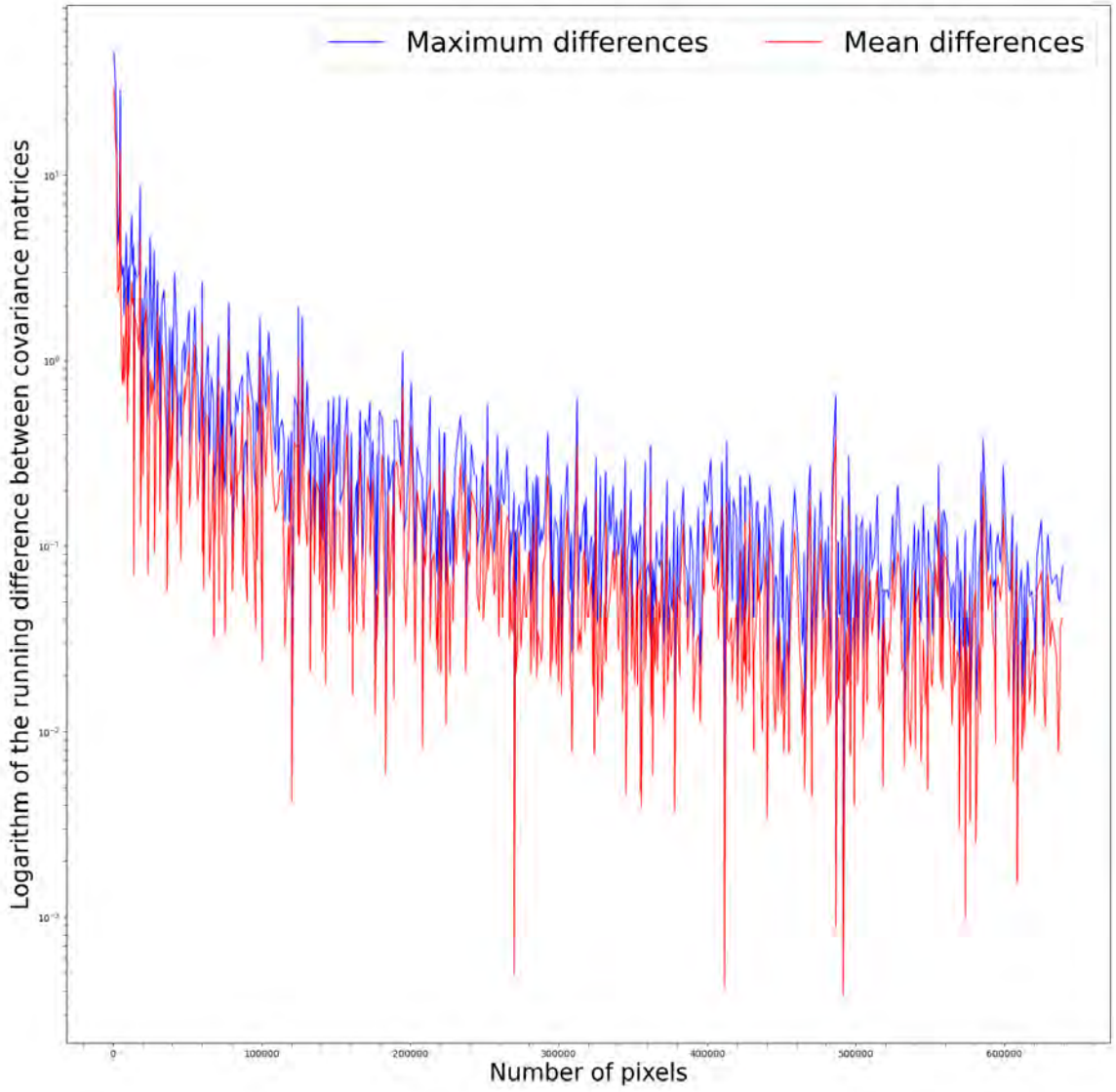

Figure 10: Mean and maximum differences between covariance matrices against number of pixels added for a cloud-free covariance matrix for ocean regions between 30 to 60^oN calculated using 1 month of spectral data from February 2020

Figure 10 shows that with the addition of further pixels, the difference between the old total error covariance matrix and the new, quantified by the mean and maximum of the difference matrix, decreases. The initial decrease is rapid up to 20,000 pixels, after which the mean and maximum differences decrease more slowly to close to zero, with minimal variation after 200, 000 pixels.

3.2 La Soufrière SO_2 Column Amounts using the ν_3 and ν_1 SO_2 Absorption Band Fast Linear Retrievals

Four of the 20 covariance matrices calculated were used within the ν_1 and ν_3 fast linear retrievals, demonstrating the utility of using a wider channel range when calculating the covariance matrices. Below, the number and distribution of the flagged pixels for each retrieval are discussed and compared. It is important to note that though some differences between results obtained using different covariance matrices have been identified, there is limited scope to comment quantitatively as only four categories of covariance matrices were tested for one scene.

Covariance Matrix Categories	Number of Pixels Flagged as Containing Elevated Levels of SO_2	
	ν_1	ν_3
Cloud Free Ocean Latitudes -30 to 30 °	450	1765
Cloudy Ocean Latitudes -30 to 30 °	407	1713
Cloud Free Tropical Land Surface	487	1761
Cloudy Tropical Land Surface	301	1674

Table 1: A table recording the number of pixels flagged for the La Soufrière scene by the Fast Linear Retrieval for two different absorption bands when using different covariance matrices

As illustrated in Table 1 above, the ν_3 band is much more sensitive to SO_2. The standard deviation for the ν_3 band results is 43.20 DU, almost half of the ν_1 band standard deviation of 80.4 DU, indicating that the use of more representative covariance matrices may have a significantly larger impact on retrievals using the ν_1 absorption band. This is reinforced by the maps of the SO_2 column amounts plotted following each retrieval, shown in Figures 11 and 12 below.

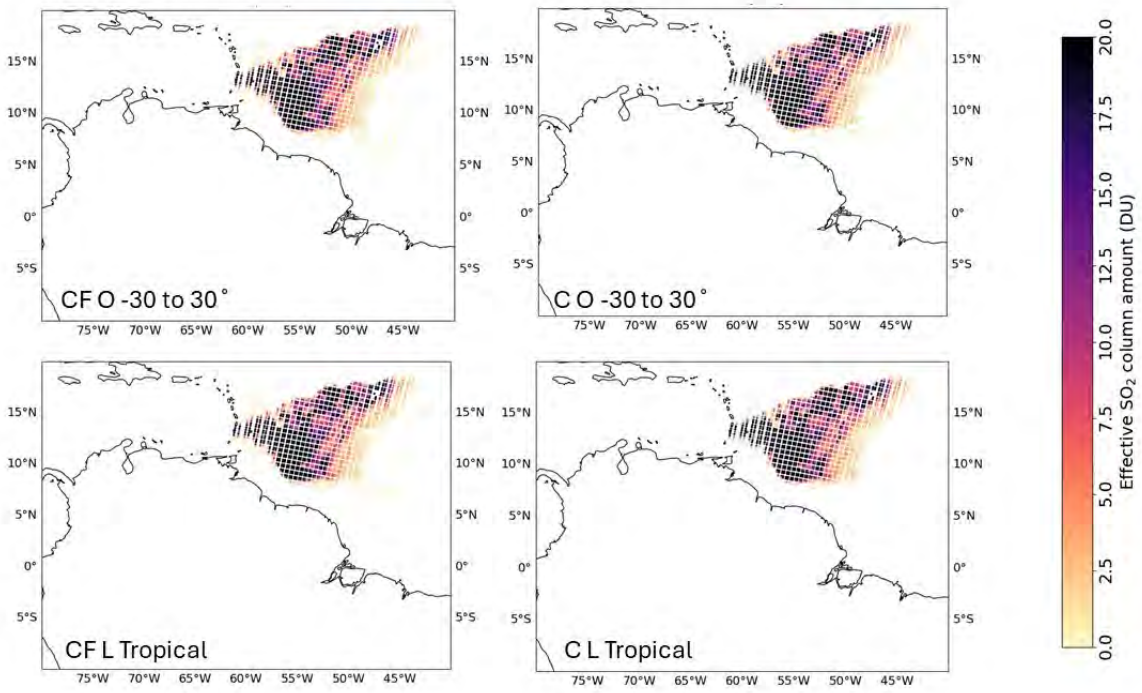

Figure 11: Maps showing the SO_2 column amounts for four different ν_3 band linear retrievals using covariance matrices for those categories detailed in Table 1

The cloud-free and ocean ν_3 band absorption band retrievals produce more scattering in Figure 11 above, with the cloudy, tropical land surface retrieval producing the least (potentially) false flagging of pixels. However, this does not mean that this retrieval produced the most accurate results as it may also have missed some pixels which did contain SO_2. The cloudy, ocean -30 to 30 ° gave the next least amount of false flagging, again indicating that cloud cover perhaps plays a more significant role than surface type in influencing the covariance matrices. Overall, however, the detections from the different retrievals are quite similar. Though the other 16 covariance matrices should also be tested, these preliminary results may suggest that for the ν_3 band, using more representative covariance matrices does not make a significant difference to the results of the fast linear retrieval.

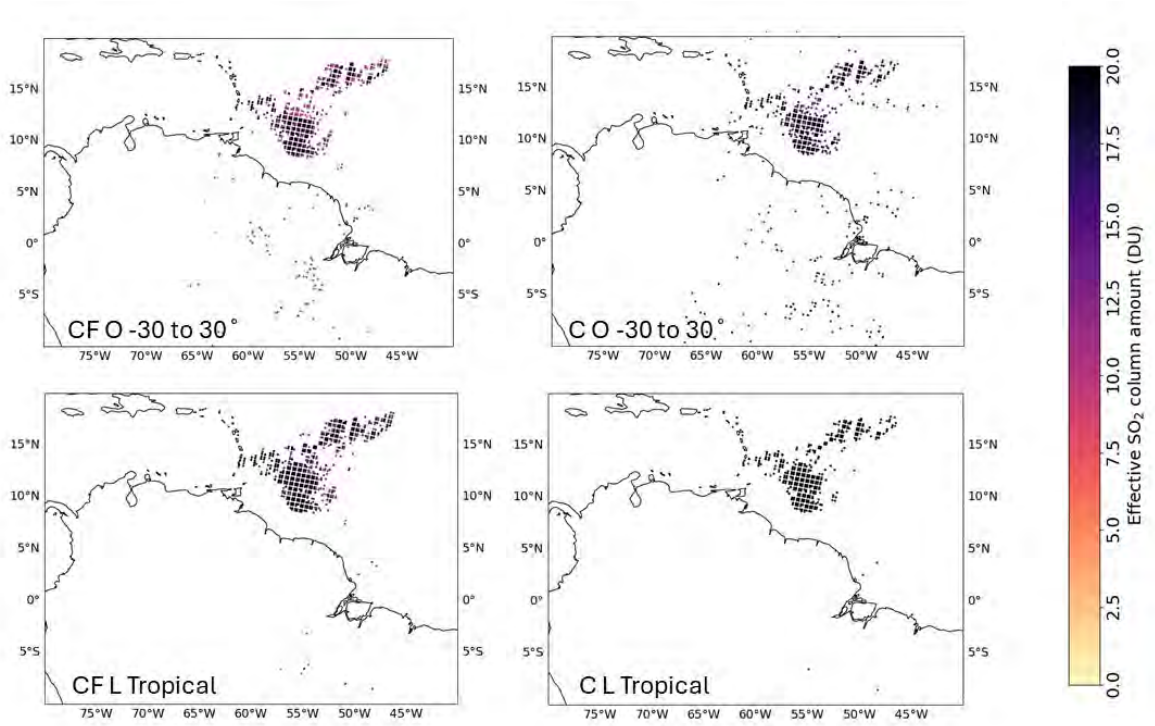

Figure 12: Maps showing the SO_2 column amounts for four different ν_1 band linear retrievals using covariance matrices for those categories detailed in Table

The use of different covariance matrices seems to have had a greater impact on the ν_1 band retrievals. Interestingly, the ocean covariance matrices gave a significant amount of potentially false flagging of pixels, particularly over land. This was most evident in the cloudy, ocean -30 to 30 ° retrieval, but was also significant for the cloud-free ocean -30 to 30 ° retrieval. The tropical land surface retrievals produced significantly less false positive flagging over land but also appeared to miss some pixels in the cloud flagged by other retrievals, including the ν_3 band retrievals, as containing SO_2. The two land retrievals, particularly that using a cloudy covariance matrix, appeared to only flag pixels with the very highest column amounts of SO_2, perhaps indicating reduced sensitivity of these land retrievals for volcanic clouds over the ocean.

It is important to note that whilst running the linear retrieval it was found that some covariance matrices threw errors due to null values recorded for the mean spectra and mean square matrices in some channels. This suggests that there is an unidentified bug in the code which generates the covariance matrices.

4 Future Work

There is significant scope for future work pertaining to this study and related areas.

- The performance of the remaining 16 representative covariance matrices not as yet reviewed within Section 3.2 of this report should be investigated and compared.
- A more in-depth investigation into how different parts of the infrared spectra, for instance, the absorption bands of volcanic ash, respond to the use of more representative covariance matrices is required to make conclusions on the significance of the impact of the use of these matrices in linear retrievals.

- An assessment of different eruption scenes to ascertain whether there are special cases in which using a more representative covariance matrix makes more of a difference to the retrieval. For example, in a desert with the presence of dust clouds or in icy locations where the covariance matrix has been found to be particularly different to other locations.
- An exploration of any temporal aspect to variation in the covariance matrices by comparing the use of covariance matrices calculated from spectra obtained in various months of the year in the retrieval.
- A wider temporal range of spectral data could also be employed in the calculation of the various covariance matrices; this investigation has used data only from the year 2020.

Additionally, if further work in creating more representative covariance matrices is attempted, the use of IASI data to classify land surfaces into different regions based on emissivity is recommended, rather than the use of biome classifications. This will limit the uncertainties introduced. More generally, an investigation into why the determinants of the total error covariance matrices are so close to zero, and what this means for the accuracy of the SO_2 column amounts calculated using the linear retrieval is suggested. The cause of the unknown bug causing null value errors should also be established.

5 Conclusions

The study establishes that around 100,000 spectra are needed to calculate a representative covariance matrix (Section 3.1.2). It is shown that there are significant differences between the covariance matrices calculated for 20 different land surface types and cloud fractions (Section 3.1.1). The factor which appears to have the greatest role in causing variance between the covariance matrices is cloud cover, with both latitude bands and surface type playing a more minor role for the ocean and land surface types respectively. The exception to this is the cloud-free ice land surface type covariance matrix, which displays significantly different behaviour. This should be investigated further as a potential scenario in which using a more representative covariance matrix may be vital in achieving accurate results from the Fast Linear Retrieval.

It is suggested that the impact of the use of more representative covariance matrices on the results of the fast linear retrieval may vary greatly depending on which absorption band is used (Section 3.2). In this investigation, the results of the ν_1 band retrievals were found to vary significantly more than the results of the ν_3 band retrieval. There was, in particular, a distinct difference in performance between those retrievals using covariance matrices constructed from ensemble spectra over land, and those using covariance matrices constructed from ensemble spectra over the ocean. The former appeared to produce significant potentially false positive, scattered flagging over land whilst the latter had reduced sensitivity to lower SO_2 column amounts. Whilst it has been established that the total error covariance matrices vary greatly depending on the categories of the ensemble spectra used to construct them, an investigation into the impact of using these matrices on the various linear retrievals has not been completed. This is a key area for further exploration, as outlined in Section 4.

6 Acknowledgements

I would like to thank my supervisors Isabelle Taylor and Anu Dudhia and the other members of EODG at Oxford University for their support and guidance during my internship. IASI spectral data was made available by EUMETSAT.

References

[1] Oppenheimer, C., B. Scaillet, and R. S. Martin. 'Sulfur Degassing From Volcanoes: Source Conditions, Surveillance, Plume Chemistry and Earth System Impacts'. Reviews in Mineralogy and Geochemistry 73, no. 1 (2011): 363–421. https://doi.org/10.2138/rmg.2011.73.13.

[2] Parker, D. E., H. Wilson, P. D. Jones, J. R. Christy, and C. K. Folland. 'The Impact of Mount Pinatubo on World-Wide Temperatures'. International Journal of Climatology 16, no. 5 (1996): 487–97. https://doi.org/10.1002/(SICI)1097-0088(199605)16:5¡487::AID-JOC39¿3.0.CO;2-J.

[3] Harrison, R. M., De Mora, S. J. 'Introductory Chemistry for the Environmental Sciences' (2nd ed.). Cambridge University Press. (1996) https://doi.org/10.1017/CBO9780511809224

[4] Guo, Song, Gregg J. S. Bluth, William I. Rose, I. Matthew Watson, and A. J. Prata. 'Re-evaluation of SO 2 Release of the 15 June 1991 Pinatubo Eruption Using Ultraviolet and Infrared Satellite Sensors'. Geochemistry, Geophysics, Geosystems 5, no. 4 (2004): 2003GC000654. https://doi.org/10.1029/2003GC000654.

[5] Niemeier, U., C. Timmreck, H.-F. Graf, S. Kinne, S. Rast, and S. Self. 'Initial Fate of Fine Ash and Sulfur from Large Volcanic Eruptions'. Atmospheric Chemistry and Physics 9, no. 22 (2009): 9043–57. https://doi.org/10.5194/acp-9-9043-2009.

[6] Marianne Guffanti, John W. Ewert, Gregory M. Gallina, Gregg J.S. Bluth, Grace L. Swanson, Volcanic-ash hazard to aviation during the 2003–2004 eruptive activity of Anatahan volcano, Commonwealth of the Northern Mariana Islands, Journal of Volcanology and Geothermal Research 146, no. 1–3, (2005): Pages 241-255. https://doi.org/10.1016/j.jvolgeores.2004.12.011.

[7] Taylor, Isabelle A., Roy G. Grainger, Andrew T. Prata, Simon R. Proud, Tamsin A. Mather, and David M. Pyle. 'A Satellite Chronology of Plumes from the April 2021 Eruption of La Soufrière, St Vincent'. Atmospheric Chemistry and Physics 23, no. 24 (14 December 2023): 15209–34. https://doi.org/10.5194/acp-23-15209-2023.

[8] Mastin, Larry, Mike Pavolonis, Samantha Engwell, Rory Clarkson, Claire Witham, Greg Brock, Ian Lisk, et al. 'Progress in Protecting Air Travel from Volcanic Ash Clouds'. Bulletin of Volcanology 84, no. 1 (2022): 9. https://doi.org/10.1007/s00445-021-01511-x.

[9] Taylor, Isabelle A., James Preston, Elisa Carboni, Tamsin A. Mather, Roy G. Grainger, Nicolas Theys, Silvana Hidalgo, and Brendan McCormick Kilbride. 'Exploring the Utility of IASI for Monitoring Volcanic SO2 Emissions'. Journal of Geophysical Research: Atmospheres 123, no. 10 (2018): 5588–5606. https://doi.org/10.1002/2017JD027109.

[10] Carn, S.A., L. Clarisse, and A.J. Prata. 'Multi-Decadal Satellite Measurements of Global Volcanic Degassing'. Journal of Volcanology and Geothermal Research 311 (2016): 99–134. https://doi.org/10.1016/j.jvolgeores.2016.01.002.

[11] Kramida, A., Ralchenko, Yu., Reader, J. and NIST ASD Team (2023). NIST Atomic Spectra Database (version 5.11), [Online]. Available: https://physics.nist.gov/asd [04/08/2024]. National Institute of Standards and Technology, Gaithersburg, MD. DOI: https://doi.org/10.18434/T4W30F

[12] Karagulian, F., L. Clarisse, C. Clerbaux, A. J. Prata, D. Hurtmans, and P. F. Coheur. 'Detection of Volcanic SO 2 , Ash, and H 2 SO 4 Using the Infrared Atmospheric Sounding Interferometer (IASI)'. Journal of Geophysical Research: Atmospheres 115, no. D2 (2010): 2009JD012786. https://doi.org/10.1029/2009JD012786.

[13] Walker, J. C., A. Dudhia, and E. Carboni. 'An Effective Method for the Detection of Trace Species Demonstrated Using the MetOp Infrared Atmospheric Sounding Interferometer'. Atmospheric Measurement Techniques 4, no. 8 (2011): 1567–80. https://doi.org/10.5194/amt-4-1567-2011.

[14] 'ESA — About IASI'. Accessed 12 September 2024. `https://www.esa.int/Applications/Observing_the_Earth/Meteorological_missions/MetOp/About_IASI`.

[15] 'Zenith Absorption — Atmospheric Infrared Spectrum Atlas'. Accessed 9 August 2024. https://eodg.atm.ox.ac.uk/ATLAS/zenith-absorption.

[16] Clarisse, Lieven, Fred Prata, Jean-Lionel Lacour, Daniel Hurtmans, Cathy Clerbaux, and Pierre-François Coheur. 'A Correlation Method for Volcanic Ash Detection Using Hyperspectral Infrared Measurements'. Geophysical Research Letters 37, no. 19 (2010). https://doi.org/10.1029/2010GL044828.

[17] Dudhia, Anu. 'The Reference Forward Model (RFM)'. Journal of Quantitative Spectroscopy and Radiative Transfer 186 (2017): 243–53. https://doi.org/10.1016/j.jqsrt.2016.06.018.

[18] Smith, Andy. 'Thoughts about Covariance', (2012)

[19] Kottek, Markus, Jürgen Grieser, Christoph Beck, Bruno Rudolf, and Franz Rubel. 'World Map of the Köppen-Geiger Climate Classification Updated'. Meteorologische Zeitschrift 15, no. 3 (2006): 259–63. https://doi.org/10.1127/0941-2948/2006/0130.

[20] Martins, João, Isabel Trigo, Virgílio Bento, and Carlos Da Camara. 'A Physically Constrained Calibration Database for Land Surface Temperature Using Infrared Retrieval Algorithms'. Remote Sensing 8, no. 10 (2016): 808. https://doi.org/10.3390/rs8100808.

[21] Taylor, Isabelle A., Roy G. Grainger, Andrew T. Prata, Simon R. Proud, Tamsin A. Mather, and David M. Pyle. 'A Satellite Chronology of Plumes from the April 2021 Eruption of La Soufrière, St Vincent'. Atmospheric Chemistry and Physics 23, no. 24 (2023): 15209–34. https://doi.org/10.5194/acp-23-15209-2023.