Retrieval of atmospheric trace gases from satellite infrared limb sounding data

Shaomin Cai¹, Weijia Zhang²,*, and Yizhou Zhao³

¹ School of Medicine, Shaoxing University
312,000, Zhejiang Province, China

² Shaoxing University
No.508, Huancheng West Road, Shaoxing 312000, China

³ Department of Astronautics
Nanjing University of Aeronautics and Astronautics
No. 169, Sheng Tai West Road, Nanjing 211106, China

Corresponding author E-mail: zhangw@usx.edu.cn

(Manuscript received in final form June 7, 2022)

Abstract— The Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) instrument which operated in the near-to-mid infrared on the Envisat satellite from 2002–2012 is a Fourier-transform spectrometer for the measurement of high-resolution atmospheric emission spectra at the Earth’s limb. The initial operational products are profiles of temperature, H₂O, O₃, CH₄, N₂O, HNO₃, and NO₂, and this list is extended to include the important GHG (greenhouse gases) and ODS (ozone depleting substances) species using MORSE (orbital retrieval processor).

This paper discusses retrievals of minor trace species of HCFC-22, HOCl, OCS, CH₃I, N₂O, HNO₃, and NO₂. Preliminary zonal mean results of these retrievals are satisfactory except for HCN and CF₄ species. We then use MIPAS to estimate the total F and Cl budget in the atmosphere. Comparisons with SLIMCAT are also discussed, with the results consistent.

Finally, we focus on improving the retrieval method for minor trace species, COF₂, by adjusting correlation length and a priori constraint, as well as the chi-square method introduced to analyze noise. Tests have shown that our improvement could get more information from the measurements.

Key-words: MIPAS, atmospheric trace gases, HCFC-22, SLIMCAT (chemical transport model), chi-square test
1. Introduction

1.1. Ozone depleting substances, ODS and greenhouse gases, GHG

It is known that the stratospheric fluorine and chlorine both have important effects on the atmosphere, especially the chlorine on the ozone depletion and the fluorine on the long-lived greenhouse gases. Accurate knowledge of the distribution and trends relevant atmospheric trace gases is crucial for the validation of numerical model simulations, which are used to study the climate system. However, until now, not much stratospheric data has been available.

In this study, the measurement of trace gases for the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) is extended to include the important GHG and ODS species, whose retrieval is challenging using an Oxford-specific state-of-the-art retrieval scheme, known as MORSE. The instrument, MIPAS, is a Fourier-transform spectrometer for the measurement of emission spectra from the Earth’s limb operating in the infra-red spectrum region (Fischer et al., 2008). It will be introduced in Section 2. With the ESA MICAL processor, the MIPAS L1 dataset has been available since 2019. MORSE is known as MIPAS optimal estimation using sequential estimation, which will be introduced in Section 3.

In this paper, the minor trace species, which will be investigated, are hydrogen cyanide (HCN), carbon tetrafluoride (CF₄), HCFC-22, carbonyl fluoride (COF₂), carbon tetrachloride (CCl₄), sulphur hexafluoride (SF₆), carbonyl sulphide (OCS), hypochlorous acid (HOCl), and ethane (C₂H₆). Concerning these species, work has been done for the retrieval as well as validation of HCN (Wiegele et al., retrieval, 2012; Glatthor et al., 2009), HCFC-22, CCl₄, SF₆, OCS, HOCl (Clarmann et al., 2009) and C₂H₆ (Wiegele et al., 2012; Glatthor et al., 2009).

The preliminary results are presented in Section 4, together with initial validation with the SLIMCAT model. Section 5 establishes the results of the total chlorine and fluorine budgets measured by MIPAS, which including these newly retrieved minor trace species and all the chlorine or fluorine-containing species that have been retrieved using MORSE.

An investigation of improving the COF₂ retrieval has been made in Section 6. Diagnostics and chi-square tests are introduced to analyze the retrievals.

1.2. Trace gases from MIPAS

A brief introduction of the minor trace species retrieved via Oxford MORSE is presented with a summary table, Table 1.
<table>
<thead>
<tr>
<th>Species</th>
<th>Source</th>
<th>Sink</th>
<th>Tropospheric VMRs [pptv]</th>
<th>Stratospheric Lifetime</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCFC-22</td>
<td>air conditioning, refrigeration <em>(Domanski, 1999)</em></td>
<td>reaction with OH and photolysis <em>(Midgley and Fisher, 1993)</em></td>
<td>200 (AGAGE for 2010 March)</td>
<td>12 years <em>(Reimann et al., 2013)</em></td>
</tr>
<tr>
<td>CF₄</td>
<td>primary aluminium Production <em>(Harnisch and Eisenhauer, 1998)</em></td>
<td>reaction with OH <em>(Watson et al., 1992)</em></td>
<td>78 (AGAGE for March 2010)</td>
<td>25,000-50,000 years <em>(Zander et al., 1996)</em></td>
</tr>
<tr>
<td>CCl₄</td>
<td>anthropogenic emission <em>(Altshuller, 1976)</em></td>
<td>photolysis <em>(Helas and Wilson, 1992)</em></td>
<td>90 (AGAGE for 2010 March)</td>
<td>44 years <em>(Reimann et al., 2013)</em></td>
</tr>
<tr>
<td>OCS</td>
<td>biomass burning <em>(Notholt et al., 2003)</em></td>
<td>dry soil <em>(Yi et al., 2008)</em></td>
<td>520 ±15 <em>(Mahieu et al., 2003)</em></td>
<td>4-5 years <em>(Mahieu et al., 2003)</em></td>
</tr>
<tr>
<td>C₂H₆</td>
<td>biomass burning and natural gas emissions <em>(Rinsland et al. 2005)</em></td>
<td>reaction with OH <em>(Rinsland et al., 2005)</em> <em>(Wiegele et al., 2012)</em></td>
<td>400 (AGAGE for 2010 March)</td>
<td>50 days <em>(Glatthor et al., 2009)</em></td>
</tr>
<tr>
<td>COF₂</td>
<td>degradation product of CFCs <em>(Muhle et al., 2010)</em></td>
<td>photolysis <em>(Szé, 1978)</em></td>
<td>150 <em>(Kaye et al., 1991)</em></td>
<td>12 minutes <em>(McCann et al., 2000)</em></td>
</tr>
<tr>
<td>HCN</td>
<td>biomass burning <em>(Singh et al., 2003)</em></td>
<td>ocean take up, atmospheric oxidation and photolysis <em>(Hickson et al., 2007)</em></td>
<td>300 <em>(Tereszchuk et al., 2013)</em></td>
<td>a few years <em>(Cicerone and Zellner, 1983)</em></td>
</tr>
<tr>
<td>SF₆</td>
<td>anthropogenic emission and natural minerals <em>(Harnisch and Eisenhauer, 1998)</em></td>
<td>electron absorption <em>(Singh et al., 1977)</em></td>
<td>7 (AGAGE for 2010 March)</td>
<td>3200 years <em>(Forster et al., 2007)</em></td>
</tr>
<tr>
<td>HOCl</td>
<td>reaction of ClO and HO <em>(Hickson et al., 2007)</em></td>
<td>photolysis <em>(Hickson et al., 2007)</em></td>
<td>150 <em>(Clarmann et al., 2009)</em></td>
<td>1 hour <em>(Mishalanie et al., 1986)</em></td>
</tr>
</tbody>
</table>
2. The MIPAS Instrument

The Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) is a Fourier-transform spectrometer for the measurement of emission spectra from the Earth’s limb (Fischer et al., 2008). MIPAS was operated from July 2002 to March 2004 at full spectral resolution (0.025 cm⁻¹) with 17 measurements for a complete limb scan. After numerous interferometer anomalies, it operated at a reduced spectral resolution (0.0625 cm⁻¹) from January 2005 until April 2012, thereby providing 10 years of data of species for investigation. The altitude coverage for the optimized resolution measurement mode, what we used in this study, is 5 to 70 km at the poles and 12 to 77 km, 7 km higher at the equator. For a complete limb scan at reduced spectral resolution, the number of measurements increased to 27 with a vertical spacing of approximating 4.5 km at higher altitudes and 15 km at lower altitudes. Fig. 1 shows a typical day’s sampling profile locations, which is useful for studying global climatology. There are 14 orbits within one day’s operation, producing large sample of measurements, about 1300 profiles per day. The major instrumental characteristics of MIPAS are summarized in Table 2.

![Fig. 1](http://www.atm.ox.ac.uk/group/mipas/L2OXF/20100321/)

**Fig. 1.** One day’s sampling profile locations on March 21, 2010. Different colours represent different orbits with the number list below, while symbols are profile locations. Open symbols represent daytime measurements while solid symbols are nighttime measurements. (Source from http://www.atm.ox.ac.uk/group/mipas/L2OXF/20100321/)
Table 2. MIPAS instrumental characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral range</td>
<td>685 – 2410 cm⁻¹ (infrared)</td>
</tr>
<tr>
<td>OR mode spectral resolution</td>
<td>0.0625 cm⁻¹</td>
</tr>
<tr>
<td>Tangent altitude range</td>
<td>5–150 km</td>
</tr>
<tr>
<td>Vertical resolution</td>
<td>~ 3 km</td>
</tr>
<tr>
<td>Field-of-view</td>
<td>3 km × 30 km</td>
</tr>
<tr>
<td>Orbit</td>
<td>sun-synchronous polar orbit</td>
</tr>
<tr>
<td>Orbit inclination</td>
<td>inclination of 98.40 degrees</td>
</tr>
<tr>
<td>Orbital period</td>
<td>about 100 minutes with a descending node mean</td>
</tr>
<tr>
<td>Operational period</td>
<td>local solar time (LST) of 10:00 am</td>
</tr>
<tr>
<td></td>
<td>July 2002 to April 2012</td>
</tr>
</tbody>
</table>

3. Retrieval method

3.1. Forward model and RFM

The atmosphere emits, absorbs, and scatters electromagnetic radiation, and the detailed spectrum depends on the gas molecules and temperature present. In the infrared spectral region, most of the molecular species have their own characteristic spectrum. Therefore, measurements of the infrared emission can be used to determine the atmospheric composition via radiative transfer.

The Reference Forward Model (RFM) is a line-by-line radiative transfer model, that originally developed by Oxford as FORTRAN program used to generate monochromatic limb emission radiances. It is a derivative of GENLN2 (Edwards, 1992), which is a general-purpose line-by-line atmospheric transmittance and radiance model. RFM is used as the forward model for Oxford MORSE retrieval algorithm.

As RFM is a line-by-line model, the transmission through the atmosphere is built up by subdividing the atmospheric path into a number of path segments. Each path segment can be considered to be homogeneous with respect to temperature, pressure, and gas concentration. Absorption coefficients are obtained by summing over a number of transitions for each molecule:

\[ k(v) = \sum_i S_i(T) f\left((v - v_{0i}, p, T)\right), \]  

where \( S_i(T) \) is the line strength, \( f \) is the line-shape, and \( v_0 \) is the central position of line \( i \).
3.2. Inverse method and MORSE

The Levenberg-Marquardt method is applied to the Oxford MORSE retrieval algorithm. The Levenberg-Marquardt method can be used when the iterative process is unstable, because there is a significant degree of nonlinearity or when the iterative process begins far from the true solution. The method introduces a Levenberg-Marquardt parameter into the variation of the unknown, to stabilize the iteration. The solution is given by:

$$x_{i+1} = x_i + [(1 + \gamma)S_a^{-1} + K_i^T S_y^{-1} K_i]^{-1}[K_i^T S_y^{-1}(y - F(x_i)) - S_a^{-1}(x_i - a)]$$

where $i$ is the iteration number; $\gamma$ is the tuneable parameter which can be adjust to achieve local minimum of cost so the retrieval may be generated towards the true solution.

In the optimal sequential estimation of the MORSE algorithm, the state vector retrieved from one measurement set is adopted as the \textit{a priori} estimate during the inversion of the next set. Given that an equivalent solution can be derived without the simultaneous inversion of all measurements, the weighting function matrix is reduced, thereby enhancing memory efficiency. The solution is calculated sequentially at the profile level. Starting from the top of the atmosphere, the \textit{a priori} estimate is updated by using values retrieved at previous tangent heights.

For MIPAS, the MORSE code takes a series of measurements from a single limb scan from the apodized Level 1b radiances, to retrieve vertical profiles of atmospheric parameters (e.g., temperature, VMRs, etc.). The \textit{a priori} estimates required for the optimal estimation approach are taken from the climatology of Remedios et al. (2007). The internal forward model is based on the RFM model but using pre-tabulated monochromatic absorption coefficients rather than a full line-by-line calculation. The radiative transfer equation for MORSE additionally applies spectral and spatial convolutions to the radiance:

$$R(v_i, z_i) = \int \int L(v) \psi(v - v_i) \phi(z - z_i) dz dv,$$

where $\psi$ is the (apodized) instrument line shape (AILS), $\phi$ is the field-of-view, $v$ is the spectral point, and $z$ is the nominal tangent height.

Retrievals in this study were performed using v1.3 of the MORSE algorithm, ESA v5 L1B radiance data, together with the cloud index technique developed by Spang et al. (2004) to identify cloud-contaminated scans. The minor trace gases are retrieved with their corresponding microwindow lists, as shown in Fig.2. A microwindow is a small spectrum range containing molecular features used for retrieval.
4. Zonal mean databases comparison

4.1. MIPAS zonal mean

With the initial retrievals generated as stated in Section 3.2, one month of data (March 2010) of the minor trace gases are analyzed in this paper. A smoothing method is applied to the retrieval results in order to eliminate the anomaly data, which in this case is defined as cloud contamination, or a significant \textit{a priori} contribution (the random error greater than 70\% of the retrieved value). A $5\sigma$ global spike test is then applied to each species to all the remaining data in order to eliminate anomalous retrievals. The percentage of profiles removed by the spike test is shown in Table 3. The largest percentage is 22.3\% from CF4. The removed profiles appear to be randomly located, independent of latitudes. Reasons for these spike test failures have yet to be investigated. As the total number of profiles for each species for one month is over 32,000, there still remain over 25,000 profiles, sufficient for statistical analyses.

Using the screened MIPAS data listed in Table 3, the monthly zonal means of the species are constructed, together with a standard deviation of profiles contributing to each grid point, which is a combination of random error and atmospheric variability. Here we only considered a monthly zonal mean for
March 2010, with 5° latitude grid and pressure range from 0.1 hPa to 1000 hPa with 6 points/decade. The modeled data from the SLIMCAT chemical transport model, which will be introduced in Section 4.3, are provided in the same grid conditions.

<table>
<thead>
<tr>
<th>Species</th>
<th>Spike (%)</th>
<th>Max VMR (pptv)</th>
<th>Min VMR (pptv)</th>
<th>Pressure range (hPa)</th>
<th>spectral range (cm⁻¹)</th>
<th>spectral range (cm⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCFC-22</td>
<td>0.79</td>
<td>260</td>
<td>9</td>
<td>0.1–560</td>
<td>803–839</td>
<td>-</td>
</tr>
<tr>
<td>CF₄</td>
<td>22.31</td>
<td>170</td>
<td>40</td>
<td>0.1–316</td>
<td>1256–1286</td>
<td>-</td>
</tr>
<tr>
<td>CCl₄</td>
<td>2.10</td>
<td>200</td>
<td>0.01</td>
<td>4–316</td>
<td>792–795</td>
<td>-</td>
</tr>
<tr>
<td>OCS</td>
<td>1.98</td>
<td>1000</td>
<td>0.01</td>
<td>1–400</td>
<td>840–857</td>
<td>2050–2053</td>
</tr>
<tr>
<td>C₂H₆</td>
<td>14.00</td>
<td>1000</td>
<td>0.0007</td>
<td>0.1–560</td>
<td>819–845</td>
<td>-</td>
</tr>
<tr>
<td>COF₂</td>
<td>3.79</td>
<td>430</td>
<td>0.09</td>
<td>0.1–560</td>
<td>772–775</td>
<td>1230–3750</td>
</tr>
<tr>
<td>HCN</td>
<td>5.68</td>
<td>440</td>
<td>110</td>
<td>0.1–560</td>
<td>711–762</td>
<td>1370–1372</td>
</tr>
<tr>
<td>SF₆</td>
<td>13.73</td>
<td>8</td>
<td>2</td>
<td>0.1–380</td>
<td>938–951</td>
<td>-</td>
</tr>
<tr>
<td>HOCl</td>
<td>0.00</td>
<td>280</td>
<td>0.01</td>
<td>0.1–560</td>
<td>1226–1229</td>
<td>-</td>
</tr>
</tbody>
</table>

Spike is representing the percentage of profiles that removed by the global 5σ spike test, max and min VMRs are the maximum and minimum values of the zonal mean after the spike test, and spectral ranges are those used in the Oxford MORSE retrieval.

MIPAS results for HCFC-22, CCl₄, C₂H₆, OCS, and SF₆ are shown in Figs. 3–7. The greatest concentrations occur roughly over the equator. Lower stratospheric concentrations are significantly higher over the equator compared to higher latitudes. Zonal mean plots of species HCFC-22, CCl₄, C₂H₆, OCS, and SF₆, shown in Figs.3–7, all have this pattern. This pattern can be attributed to the Brewer-Dobson circulation (a slow upwelling of stratospheric air in the tropics, followed by a pole-ward drift through the mid-latitudes, and a descent in the mid- and high latitudes). The general appearance is typical of long-lived gases such as CFCs that have tropospheric sources and are destroyed in the stratosphere. The standard deviation plots also show similar structure, which implies that MIPAS errors are approximately proportional to the retrieved values under the assumption that atmospheric variability is negligible. Fig. 3 shows that the tropospheric concentration for HCFC-22 is roughly 200pptv (red color), which is consistent with the value shown in Table 1. Fig. 4 shows that the concentration of CCl₄ in troposphere region is roughly 100pptv with a standard deviation of 10 pptv, i.e., 100±10pptv, which is consistent with the value (90pptv) of CCl₄ shown in Table 1. The zonal mean retrievals of C₂H₆ have smaller VMR in the troposphere,
of about 200pptv (shown in Fig. 5), than the value of 400 shown in Table 1. The tropospheric VMR of SF₆ is about 6±0.8 pptv in Fig. 7, which is close to the value shown in Table 1. The retrievals for these 5 minor species are reasonable, however, further validations are required.

Fig. 3. MIPAS zonal mean and standard deviation of HCFC-22 in March 2010.

Fig. 4. MIPAS zonal mean and standard deviation of CCl₄ in March 2010.

Fig. 5. MIPAS zonal mean and standard deviation of C₂H₆ in March 2010.
The plot of COF$_2$ (Fig. 8) shows a maximum of VMR around the middle of the stratosphere over the equator. This occurs, because COF$_2$ is an intermediate product in the degradation of CFCs in the stratosphere, which results in an increase in concentration. Also, because COF$_2$ is destroyed by photolysis and reaction with O(1D) in the upper stratosphere, the concentration decreases at higher altitudes. There is a second maximum of VMR in the stratosphere over the southern hemisphere (SH) polar region. As the data we obtained for this zonal mean plot is March 2010, which is the end of summer in SH. Then photochemical production of COF$_2$ extends to the pole in the middle stratosphere (i.e., in polar day).
The VMR of HOCl shown in Fig. 9 shows almost symmetric structure with high VMRs over mid-latitude regions at about 5hPa. HOCl is a product of ClO and HO, but destroyed by photolysis. As March is springtime, both north and southpoles have the same amount of sunlight which may be the reason of the symmetry between the north and south hemispheres. In March, when most of the sunlight is over the equator region, where there is more HOCl, the photolysis results in relatively small VMR. However, the South Pole region has higher VMRs at about 30hPa than the north. The reason requires further investigation.

Fig. 9. MIPAS zonal mean and standard deviation of HOCl in March 2010.

4.2. The SLIMCAT model

SLIMCAT is a 3D off-line chemical transport model (CTM) using winds and temperatures from meteorological analyses (e.g., from the UK Met Office or ECMWF) to specify the atmospheric transport and temperatures, and calculates the abundances of chemical species in the troposphere and stratosphere (Chipperfield, 2006). The model also uses a specified surface VMR of the source gases as boundary condition. These surface VMRs are time-dependent monthly values and based on surface measurements. The data used here for comparisons are provided by Sandip Dhomse and Martyn Chipperfield from output of a standard SLIMCAT stratospheric chemistry run for March 2010.

4.3. Zonal mean comparisons

SLIMCAT also models the species HCFC-22, CCl4, COF2, HOCl, and HCN. However, the HCN tropospheric concentration is not yet sorted out for SLIMCAT model, leaving four species. These are shown in Figs. 10–13.

For species HCFC-22, SLIMCAT and MIPAS agree well with each other. The maximum difference is 30 pptv, about 15%, around 20 hPa and 100 hPa in the southern hemisphere.
Differences in CCl₄ between MIPAS and SLIMCAT are very small over the equator in the stratosphere, but about 50 pptv, about 33%, in the lower troposphere, with MIPAS larger. Also, MIPAS is also larger in high latitude regions in the troposphere and has more details of variation of the VMRs. Since the SLIMCAT is the modeled with stratospheric chemistry scheme, better results will be obtained for stratosphere than for troposphere. The variation of CCl₄ VMRs in the troposphere is not well modeled.

SLIMCAT and MIPAS values of HOCl have larger differences than the other three species in the March 2010 zonal mean plots. Zonal means from SLIMCAT and MIPAS only consistent with each other at large VMR in the upper stratosphere region and around 50 hPa in northern high latitude region.
The differences in COF$_2$ between MIPAS and SLIMCAT are very small over the equator in the stratosphere, but about 0–5 pptv, about 5%, in the lower troposphere, with MIPAS larger. Larger differences occur mainly in the stratospheric region with the percentage of about 40%.

5. Some preliminary results

5.1. Fluorine and chlorine budgets

As MIPAS can measure many of the halogenated compounds in the stratosphere, it is useful to study the fluorine and chlorine budgets with MIPAS data. The fluorine species measured by MIPAS are COF$_2$, CFC-11 (CCl$_3$F), CFC-12 (CCl$_2$F$_2$), CF$_4$, HCFC-22 (CHClF$_2$), and SF$_6$ from which the total fluorine measured by MIPAS can be computed as:
\[ [F]_m = 2[COF_2] + [CCl_3F] + 2[CCl_2F_2] + 4[CF_4] + 2[CHClF_2] + 6[SF_6], \]  
\[ \text{(4)} \]

where ‘[ ]’ represents the volume mixing ratio (VMR) of each species.

Mahieu et al. (2008) show that most of the stratospheric fluorine species eventually form hydrogen fluoride [HF], which is not measured by MIPAS. Other fluorine containing species are negligible compared with these. Hence the total fluorine considered here is represented as following:

\[ [F] = [F]_m + [HF]. \]  
\[ \text{(5)} \]

*Fig. 14* shows the vertical profiles with global average of all the fluorine species and total fluorine measured by MIPAS. In the lower stratosphere and upper troposphere, the atmospheric fluorine is mainly contributed by CFC-12. COF₂ contributes nearly half of the fluorine at the 10 hPa pressure level, while CF₄ dominates in the upper stratosphere.

The total \([F]_m\) decreases with altitude as fluorine ends up as HF. The expected HF zonal plot for March 2010 is constructed by the following:

\[ [HF]_m = [F_m]_{\text{max}} - [F]_m, \]  
\[ \text{(6)} \]

where \([F_m]_{\text{max}}\) is the maximum value of \([F]_m\).
Constructing the monthly zonal mean for expected HF from MIPAS, \([\text{HF}]_m\), is compared with the zonal mean from SLIMCAT, \([\text{HF}]_a\) in Fig. 15. The plots are consistent with each other. Adding the vertical profile of HF modeled by SLIMCAT (see Fig. 14) should result in constant total \([F]\).

Similarly, for chlorine budget, the species measured by MIPAS are CFC-11, CFC-12, CFC-22, CCl₄, HOCl, and ClONO₂, from which the total chlorine measured by MIPAS can be computed as:

\[
[\text{Cl}]_m = 3[\text{CCl}_3\text{F}] + 2[\text{CCl}_2\text{F}_2] + [\text{CHClF}_2] + 4[\text{CCl}_4] + [\text{ClONO}_2] + [\text{HOCl}]. \quad (7)
\]

There are three major chlorine species, HCl, CH₃Cl, and CFC-113 that are not measured by MIPAS. However, CH₃Cl has spectral feature at about 700 cm⁻¹ and CFC-113 has features at 780–995 and 1005–1232 cm⁻¹ spectral ranges, and so they are the potential retrieval candidates of MIPAS. Fig. 16 shows the global averaged vertical profiles of all the chlorine species and total chlorine \([\text{Cl}]_m\) measured by MIPAS. The solid line, representing the amount of \([\text{Cl}]_m\), is decreasing with altitude, starting with about 2.4 ppbv in the troposphere. In the troposphere, around half of the total chlorine is contributed by CFC-12 and around one third of it is contributed by CFC-11. ClONO₂ dominates the chlorine VMR at around the 20 hPa pressure level without considering the concentrations of HCl, CH₃Cl, and CFC-113.

The approach of constructing expected [HF] can be applied for expected \([\text{HCl}]+[\text{CFC-113}]+[\text{CH}_3\text{Cl}]\). Data provided by ACE-FTS will be included for comparison in the future.
6. COF$_2$ MIPAS retrieval method study and improvement

6.1. Modifications of the retrieval method

The problem of retrieving minor traces gases with Oxford MORSE algorithm and current parameters set for MORSE code is that these gases have relatively small signal-to-noise ratio in the emission spectrum compared with atmospheric major gases, such as O$_3$ and CH$_4$, which are retrieved by Oxford MORSE. The minor trace gas retrieval problem can be shown mathematically as follows:

From Section 3.2, the retrieval covariance general solution for optimal estimation is $S_x = (K^T S_y^{-1} K + S_a^{-1})^{-1}$. Let

$$S_n^{-1} = K^T S_y^{-1} K,$$

where $S_i$ is the instrument noise in the level 1 spectrum, $S_n$ is the measurement error transferred into VMR space by the Jacobian, i.e., $S_n$ is the covariance of the measurements.

Eq.(8) can be simplified as:

$$S_x = (S_n^{-1} + S_a^{-1})^{-1}.$$  (9)

For major atmospheric gases, who have strong emission signals, retrieval errors, $S_x$, is dominate by the measurement noise, $S_n$, covariance of a priori is negligible. Hence the retrieval is mainly contributed by the measurements.
However, for minor trace gases, for example, COF$_2$, whose emission signal is small, the measurement errors are at the same magnitude scale as the \textit{a priori} errors. From Eq.(9) we can see that in order to obtain better retrievals, one can either reduce the measurement noise or increase the \textit{a priori} covariance, in other word, reduce $S_n$ or increase $S_a$.

Reducing measurement noise can be achieved by retrieved with more measurements, for example, retrieved COF$_2$ with monthly spectrum instead of individual profiles. Increasing the \textit{a priori} covariance simply means relaxing the \textit{a priori} constrains in our study. In this paper, we focus on the second method.

Two ways can be considered to change the \textit{a priori} covariance matrix in the MORSE code: one is by changing the correlation length and the other is by changing the \textit{a priori} uncertainty for each gas. The correlation length can be considered a parameter representing the vertical smoothing of retrievals. Table 4 shows three different parameter settings for studying the effect of relaxing the \textit{a priori} constraints for COF$_2$ retrievals. The default setting (cur) has a strong correlation length of 50 km, which provides regularization at the expense of vertical resolution. The vertical resolution of COF$_2$ retrievals can be improved by reducing the correlation length to 5 km (cor5km). Initially, the \textit{a priori} uncertainty is set to 100% to avoid overconstraining the retrieval for major species. An \textit{a priori} uncertainty of 1000% (apun) is selected to test the effect of increasing \textit{a priori} assumed errors. These three settings of COF$_2$ retrievals are all generated with the same spectral microwindows, which are associated with different altitude ranges and are shown in Table 5.

\begin{table}[h]
\centering
\caption{Three different parameter settings for COF$_2$ retrievals}
\begin{tabular}{lll}
\hline
Name & Correlation length & \textit{A priori} uncertainty \\
\hline
cur (default) & 50km & 100\% \\
cor5km & 5km & 100\% \\
apun & 50km & 1000\% \\
\hline
\end{tabular}
\label{tab:parameter_settings}
\end{table}

\begin{table}[h]
\centering
\caption{The microwindow set and associated altitude ranges for COF$_2$}
\begin{tabular}{lll}
\hline
Microwindow & Spectral range (cm$^{-1}$) & Altitude (km) \\
\hline
COF20301 & 1230.3750 $\sim$ 1233.3750 & 12.0 $\sim$ 40.0 \\
COF20302 & 1233.6875 $\sim$ 1235.8125 & 7.5 $\sim$ 19.5 \\
COF20303 & 772.0000 $\sim$ 775.0000 & 18.0 $\sim$ 43.0 \\
COF20304 & 1225.7500 $\sim$ 1228.6875 & 16.5 $\sim$ 46.0 \\
COF20305 & 1222.4375 $\sim$ 1225.4375 & 10.5 $\sim$ 54.0 \\
\hline
\end{tabular}
\label{tab:microwindows}
\end{table}
6.2. Evaluation parameters

6.2.1. Averaging kernels

Averaging kernels describe how the true state ($\mathbf{x}$) of the atmosphere has been distorted by the retrieval ($\hat{\mathbf{x}}$):

$$\mathbf{x} = \mathbf{A}\mathbf{x}, \quad (10)$$

which is defined as the rows of matrix $\mathbf{A}$:

$$\mathbf{A}\mathbf{K} = \mathbf{I} - \mathbf{S}\mathbf{s}_{\mathbf{a}}^{-1}. \quad (11)$$

The rows of $\mathbf{A}$ are the averaging kernels (smoothing functions) that map the true state into the retrieval space. The width of the kernel is a measure of the retrieval resolution. The area of the averaging kernel is roughly unity where the retrieval is accurate and can be thought of as a rough measure of the fraction of the retrieval, that comes from the measurement rather than the a priori estimate (Rodgers, 2000).

6.2.2. Degrees of freedom (DOF)

The number of degrees of freedom (DOF) is a scalar measure of the number of independent quantities that can be measured. The largest number of degrees of freedom possible is clearly determined by the number of elements in the measurement vector (or, indeed, the state vector – whichever is the smaller), and it is defined as the trace of $\mathbf{A}\mathbf{K}$:

$$DOF = tr(\mathbf{A}\mathbf{K}) \quad (12)$$

6.2.3. Shannon information content

The Shannon information content is a scalar quantity originally conceived to describe the information carrying capacity of communications channels (Shannon and Weaver, 1962). It is also a useful quantity for the optimization of observing systems. The information content of a measurement can be defined qualitatively as the factor by which knowledge of a quantity is improved by making the measurement (Rodgers, 2000). It is defined as:

$$IC = -\frac{1}{2} \ln |\mathbf{I}_n - \mathbf{A}\mathbf{K}|. \quad (13)$$
6.3. Results

Data for the three aforementioned retrieval settings for 1 day are generated. Figs. 17–19 show the zonal mean of COF$_2$ for 1 day. These three figures all capture the maximum concentration at the mid-stratosphere over the equatorial region of COF$_2$ in March. A decrease in the correlation length with increasing COF$_2$ concentration is shown over the North Pole at around the 30 km altitude (Fig. 18). The same finding is also captured by increasing the a priori uncertainty, as shown in Fig. 19. Fig. 19 shows more fluctuations of retrievals at higher altitudes, especially over the equatorial region. The concentration of COF$_2$ in Fig. 18 is roughly smaller than that in Fig. 17 at the same altitude. Overall, the concentrations are consistent with one another.

Fig. 17. Zonal mean of COF$_2$ with 1 day of data under “cur(default)” setting.
Fig. 18. Zonal mean of COF$_2$ with 1 day of data under “cor5km” setting, which changes the correlation length to 5 km.

Fig. 19. Zonal mean of COF$_2$ with 1 day of data under “apun” setting, which changes the \textit{a priori} uncertainty to 1000\%. 

The averaging kernels are derived within the 30–60N latitude band for these three sets of retrievals. The averaging kernels are the rows of matrix $A$ as mentioned in Section 6.2.1. Figs. 20–22 show the averaging kernel and vertical resolution of three different setting retrievals of COF$_2$. The resolution at each altitude is defined as the ratio of the grid spacing to the diagonal of the averaging kernel matrix, which is only meaningful where averaging kernels have distinct peaks at the tangent points.

**Fig. 20.** Left panel: Averaging kernels for retrievals with default setting over the 30–60N region on March 21, 2010. The measurement altitude of each averaging kernel is indicated by a cross with matching color, with the values of DOF= 4.69, IC=7.64. Right panel: Vertical resolution as a function of altitude of this averaging kernel. Resolutions are lower than the MIPAS field of view of ~3 km.
Fig. 21. Averaging kernels and vertical resolution for COF$_2$ retrieval after changing the \textit{a priori} correlation length to 5 km. DOF=7.03, IC=10.78.

Fig. 22. Averaging kernels and vertical resolution for COF$_2$ retrieval after changing \textit{a priori} uncertainty to 1000%. DOF=8.34, IC=15.13.
A direct comparison of the vertical resolution of the three retrievals is shown in Fig. 23. The best vertical resolution is achieved at around the $20\text{-}30$ km altitude for all three retrievals, which is as expected as the concentration of COF$_2$ is largest at these altitudes over the north mid-latitude region. Vertical resolutions for “cor5km” and “apun” retrievals are clearly better than that for “cur” retrieval. At around $20$ km altitude, the vertical resolution is about $4$ km, thereby indicating “cur” (default) retrieval. Meanwhile, vertical resolutions for the other two retrievals are about $3$ km, which is the best resolution that can be achieved, matching of MIPAS field of view height. From the perspective of resolution, reducing the correlation length to $5$ km or increasing the a priori uncertainty to $1000\%$ can achieve better retrieval results for COF$_2$.

Fig. 23. Vertical resolutions of the three different retrievals. The black dotted line shows the vertical spacing of the retrieval grid (which is also the measurement tangent height spacing) for the mid-latitude profile.

Table 6 lists the degrees of freedom (DOF) and information content (IC) for three different retrievals, which are derived from the corresponding AK. As it was expected, relaxing the a priori constraints (“cor5km” and “apun”) results in a higher DOF and in more information from measurements. In other words, the DOFs for “cor5km” and “apun” are higher than that for the default setting of retrievals. Similarly, the same trend is observed for IC.
Table 6. DOF and IC for the three different COF$_2$ retrievals

<table>
<thead>
<tr>
<th>settings</th>
<th>DOF</th>
<th>IC</th>
</tr>
</thead>
<tbody>
<tr>
<td>cur(default)</td>
<td>4.69</td>
<td>7.64</td>
</tr>
<tr>
<td>cor5km</td>
<td>7.03</td>
<td>10.78</td>
</tr>
<tr>
<td>apun</td>
<td>8.34</td>
<td>15.13</td>
</tr>
</tbody>
</table>

6.4. Chi-square test

The diagnostics only depend on the retrieval and *apriori* covariance, having little information about the consistency of the individual retrieved profiles. In theory, with relaxing the *apriori* constrains, the noise will increase linearly. But in reality, the noise always increases nonlinearly, and the chi-square is a quantity to measure the uncertainty of individual retrieved profiles in order to make sure that the noise is not amplified too much.

The chi-square test is introduced to analyze the consistency of the individual profiles and is defined as follows:

$$\chi^2 = \frac{1}{N} \sum_{i=0}^{n} \delta p_i^T \cdot S^{-1}_{\delta} \cdot \delta p_i ,$$  \hspace{1cm} (14)

where $\delta p_i$ is the difference between the retrieval profile $i$ and the reference profile; $S_{\delta}$ is the measurement noise of the difference between the retrieval profile and the reference profile; there are $N$ profiles in total.

There are different applications of the chi-square test by choosing different reference profiles. In this study, we consider the following applications.

Application 1: Reference profile = interpolation profile, with assumption that linear interpolation is valid for the three adjacent retrieval profiles. Then $\delta p_i$ and $S_{\delta}^{-1}$ are defined as follows:

$$p_{in} = \frac{(p_{i-1} + p_{i+1})}{2} ,$$  \hspace{1cm} (15)

$$\delta p_i = (p_i - p_{in}) ,$$  \hspace{1cm} (16)

$$\delta S_i = (S_i + S_{in}) = \frac{3}{2} S_i ,$$  \hspace{1cm} (17)

where $S_{in}$ is the covariance of interpolation profile, which is derived as $S_{in} = \frac{1}{4} (S_{i-1} + S_{i-1}) \approx \frac{1}{2} S_i$, where $S_i$ is the measurement covariance of profile $i$, the inverse of $S_i$ is the differences of the inverse retrieval covariance and the inverse
of the a priori covariance. In mathematical term, \( S_i^{-1} = S_a^{-1} - S_{ai}^{-1} \), where \( S_a \) and \( S_{ai} \) represent the covariance of retrieval and the a priori for profile \( i \).

Hence, chi-square value in Eq.(14) for this application is derived as follows:

\[
\chi^2 = \frac{1}{N} \sum_{i=0}^{n} \delta p_i^T \cdot \frac{2}{3} (S_{xi}^{-1} - S_{ai}^{-1}) \cdot \delta p_i = \\
= \frac{1}{N} \sum_{i=0}^{n} (p_i - p_{in})^T \cdot \frac{2}{3} (S_{xi}^{-1} - S_{ai}^{-1}) \cdot (p_i - p_{in}) \\
= \frac{1}{N} \sum_{i=0}^{n} [p_i - \frac{(p_{i-1} + p_{i+1})}{2}]^T \cdot \frac{2}{3} (S_{xi}^{-1} - S_{ai}^{-1}) \cdot [p_i - \frac{(p_{i-1} + p_{i+1})}{2}].
\]  

(18)

6.5. Chi-square test results

The chi-square value for application 1 in this report is constructed using 1 day (March 21, 2010) of retrieval data for COF2. As previously conducted, cloud accumulation profiles are eliminated. Fig. 24 shows the distribution of chi-square values for the three different retrieval settings. Most chi-square values are relatively small, except for some anomalies (large chi-square value). Distributions of the chi-square value fall roughly exponentially as the chi-square value increases for all three retrievals. Table 7 summarizes the mean and median values for the chi-square values in Fig. 24. The median value is a more sensible quantity to describe the chi-square value because of the anomalies in the chi-square values. The median of the chi-square value is larger for relaxing the a priori constraint retrievals. The mean of “apun” is almost three times larger than the mean of default retrievals, whereas the mean of “cor5km” is almost twice as much as that of „cur(default)‟.

<table>
<thead>
<tr>
<th>settings</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>cur(default)</td>
<td>12.65</td>
<td>7.45</td>
</tr>
<tr>
<td>cor5km</td>
<td>22.00</td>
<td>12.96</td>
</tr>
<tr>
<td>apun</td>
<td>38.13</td>
<td>20.24</td>
</tr>
</tbody>
</table>

To better understand the distributions, locations of anomalies need to be identified. As shown in Figs. 25(a) and 25(b), the distribution of chi-square values for default setting retrievals only has two chi-square values that are larger than the threshold. This result implies two anomaly profiles of the retrieval. From Fig. 25(b), these two anomaly profiles are in the south polar region. Figs. 25(c)
and 25(d) show the distribution of chi-square values and location of anomaly profiles for the 5 km correlation length. Anomaly profiles mainly lie within the 80–90S latitude band. One profile exists in the north polar region. Figs. 25(e) and 25(f) illustrate the distribution of chi-square values and locations of anomaly profiles for the 1000% a priori uncertainty retrieval. The chi-square values are more varied than the previous two retrievals and contain more anomalies. Anomaly profiles occur mainly at around the 60N latitude in the northern hemisphere, which may correspond to the oscillations at the same latitude range in the zonal mean plot (Fig. 19). In the southern hemisphere, half of the anomalies lie over the South Pole. Data are retrieved in March, which is summer in the southern hemisphere. Temperature is more varied in the south polar region, which would have influence as COF$_2$ is mainly from the temperature-dependent reaction of HCFC-22+OH. One possible reason for the anomalies is temperature uncertainty, which needs to be confirmed through further investigations.
Fig. 24. The distribution of chi-square values for (a) cur(default) setting, (b) cor5kmsetting, and (c) apun setting of retrievals on date March 21, 2010. The y axis is the frequency and the x axis is the chi-square value.
Fig. 25. (a), (c), and (e) are distributions of chi-square values with the band of 5 for the three different setting retrievals. The y axis is the frequency (i.e., number of counts) and the x axis is the chi-square value. (b), (d), and (f) are the locations of profiles that have chi-square value larger than the threshold of 60. The y axis is the latitude and the x axis is the longitude.
7. Conclusions and future work

In this paper, we discussed retrievals of minor trace species of HCFC-22, HOCl, OCS, C₂H₆, COF₂, HCN, CF₄, SF₆, and CCl₄. Preliminary zonal mean results of these retrievals are satisfactory except for HCN and CF₄ species, which have some anomalies requiring further investigation. Comparison with SLIMCAT results are also discussed. CCl₄, COF₂, HCFC-22, and HOCl zonal mean plots of MIPAS and SLIMCAT seem consistent. We then used MIPAS to estimate the total F and Cl budgets in the atmosphere, and also compared the result with SLIMCAT.

The expected zonal mean HF produced from MIPAS fluorine budgets agrees with the SLIMCAT results. The vertical profiles of both fluorine and chlorine budgets need to be compared with other satellite results, like ACE-FTS and other MIPAS processors.

Finally, we focus on improving the retrieval method for minor trace species, COF₂, by adjusting correlation length and a priori constraint, as well as chi-square introduced to analyze noise. Results shown that we did make improvement and could get more information from the measurements.

Comparison with ACE-FTS, other MIPAS processor, and model results are required for further validation.

Acknowledgments: I would like to take this chance to express my gratitude to those who assisted me in many ways during the writing of this study. My deepest thank goes to Professor Anu Dudhia from Department of AOP Physics, University of Oxford, for his explicit advice on sharpening my ideas and improving the accuracy of this paper. I shall extend my thanks to relevant MIPAS and SLIMCAT team members for their ground-breaking work. Also, I’m obliged to my family and colleges who accompanied me these years, giving me the support that I will never forget. Finally, my sincere gratitude goes to those who have devoted time reading this paper and given me many suggestions, which will benefit my study and work in the future.

References


