Impact of Spaceborne Observations on Tropospheric Composition Analysis and Forecast (ISOTROP)

Synthetic Observation Product Specification (SOPS)
Deliverable D4 / WP3

prepared by

Jukka Kujanpää, FMI
Albert Oude Nijhuis, KNMI
Henk Eskes, KNMI
Johan de Haan, KNMI
Pepijn Veefkind, KNMI
Johanna Tamminen, FMI

12 August 2015

ESA CONTRACT RECORD

<table>
<tr><td>ESA Contract No:
4000105743</td><td>SUBJECT:
Synthetic Observation Product Specification (SOPS)

TITLE:
Impact of Spaceborne Observations on Tropospheric Composition Analysis and Forecast (ISOTROP)</td><td>CONTRACTOR:
FMI

MAIN CONTRACTOR:
KNMI</td></tr>
<tr><td>ESA CR() No:</td><td>No. of Volumes: 1
This is Volume No: 1</td><td>CONTRACTOR'S REFERENCE:
ISOTROP/WP3/D4</td></tr>
<tr><td colspan="3">ABSTRACT:
This document is the deliverable D4 of the work package WP3.</td></tr>
<tr><td colspan="3">The work described in this report was done under ESA Contract. Responsibility for the contents resides in the author or organisation that prepared it.</td></tr>
<tr><td colspan="3">Names of authors:
Jukka Kujanpää (FMI)
Albert Oude Nijhuis (KNMI)
Henk Eskes (KNMI)
Johan de Haan (KNMI)
Pepijn Veefkind (KNMI)
Johanna Tamminen (FMI)</td></tr>
<tr><td colspan="2">NAME OF ESA STUDY MANAGER:
Dirk Schuettemeyer
DIV: EOP SMS</td><td>ESA BUDGET HEADING:</td></tr>
</table>

DOCUMENT CHANGE RECORD

Issue	Date	Modified items / Reason for change
0.9	22.11.2013	Initial revision of the full document
1.0	13.12.2013	Upated to reflect the latest data
2.0	24.4.2015	Major rewrite adding contributions from KNMI to form this one document, and updated to reflect the latest data
2.1	12.8.2015	Reformulated section 1.1. to explain that S5P orbits are used for synthetic LEO data. Replaced references to S5 with S5P where appropriate. In table 2.2. and the text in sect. 2.2.2. clarified that the errors for cloud pressure, cloud fraction and surface albedo are taken from the NO2 ATBD for S5P.

Contents

1 Introduction

1.1 Purpose and scope

The main aim of the ISOTROP (Impact of Spaceborne Observations on Tropospheric Composition Analysis and Forecast) project is to assess the impact of individual or combined low earth orbit (LEO) and geostationary orbit (GEO) satellite observation system measuring in the UV, visible, near infrared and short wave infrared at nadir on a data assimilation scheme for tropospheric composition monitoring and forecast with a focus on Europe [AD1]. This is to be achieved by performing an Observing System Simulation Experiment (OSSE). Simulated level 2 observational values and their errors are drawn from a representation of reality (the "Nature Run" taken from a state-of-the-art chemistry transport model simulation at high resolution) and the characteristics of the satellite retrievals. These data are fed into a data assimilation system of a second model and the impact is evaluated. This document describes the simulated level 2 trace gas observations. The trace gas products of interest are: tropospheric column of nitrogen dioxide (NO_2), vertical profile of ozone (O_3) and total columns of carbon monoxide (CO) and formaldehyde (HCHO). The following three main elements are needed to create the level 2 data files: 1.) the state vector of the atmosphere which contains the trace gas concentration profile, temperature profile, cloud pressure, cloud radiance fraction and surface albedo, 2.) an orbit simulator providing the solar and instrument measurement geometry and 3.) a look-up table and an interface to estimate the retrieval errors and averaging kernels. The error estimation procedure is described in detail in this document.

The satellite system to be studied comprises of the future Copernicus Sentinel-4 (S4) geostationary, Sentinel-5 (S5) polar-orbiting, and Sentinel-5 precursor (S5P) missions. These missions are dedicated to monitoring the composition of the atmosphere for the Copernicus Atmospheric Service (CAS). The S4 and S5 missions will be carried on meteorological operational satellites operated by EUMETSAT in the 2020 time frame: the S4 will be on the sounder satellite (MTG-S) of the Meteosat Third Generation programme performing hourly observations from a geostationary orbit while the S5 will be embarked on the polar-orbiting Metop-SG A satellite performing trace gas observations from a morning orbit (09:30 LT). Before the S5 mission, atmospheric composition for the CAS will be monitored by the TROPOMI instrument on the Sentinel-5 Precursor (S5P) mission, launch scheduled for 2016. However, the S5P will be put on an afternoon orbit (13:30 LT) to continue the observations made by the Dutch-Finnish OMI instrument on a similar orbit.

In ISOTROP, synthetic LEO and GEO observations have been generated based on the S5P instrument model and characteristics, because this one is the most developed of the three instruments. The instrument assumptions made are in line with the draft ATBDs of S5P-TROPOMI. Also, for the LEO simulations the overpass time of S5P has been adopted. The synthetic GEO-S4 orbits are generated with the S4 geometry and with an hourly measurement frequency. In order to cover both Copernicus low earth orbit missions (S5P and S5), the hourly S4 data is exploited to study the impact of different overpass times of S5P and S5 by assimilating observations for only one hour, namely the 13:00 UT and the 9:00 UT in separate OSSE experiments. The approach will be described in more detail in the final report and the reports describing the OSSE results. Note that CO is not a product of S4, and CO datasets have only been generated for the S5P geometry/orbits.

1.2 References

1.2.1 Applicable Documents

[AD1] Eskes, H., Veefkind, P., Strunk, A., Attié, J.-L., El Amraoui, L., Ricaud, P., Curier, L., Builtjes, P., Timmermands, R., Segers, A. and Lahoz, W., Impact of spaceborn observations on tropospheric composition analysis and forecast (ISOTROP) Proposal Volume I Technical Proposal, Proposal for ITT AO/1-6845/11/NL/AF, 2011.

[AD2] ESA Mission Science Division, GMES Sentinels 4 AND 5 Mission Requirements Document, EOP-SMA/1507/JL-dr, issue 3, revision 2, date 21.09.2011.

[AD3] Nijhuis, A. O., de Haan, J. F., Eskes, H. and Veefkind, P., ISOTROP synthetic observation generation by using look-up tables, Final version 1.0, 3 May 2013.

[AD4] Nijhuis, A. O., Eskes, H. and de Haan, J. F., Synthetic Observations Product Specification Level-2 user manual, Draft, 19 February 2012.

[AD5] Eskes, H. et al., Choice of model domains, resolution, study periods and further clarification of the approach, Final version, 28 September 2012.

[AD6] Kujanpää, J., Eskes, H., Veefkind, P., de Haan, J. and Segers, A., Synthetic retrievals: overview of algorithm verification, version 25 April 2014.

[AD7] Williams, J., Eskes, H. and Nijhuis, A. O., Comparison of cloud parameters calculated using ECMWF and ARPEGE/MOGACE input variables for use in retrievals, 08 January 2013.

1.2.2 Reference Documents

[RD1] de Haan, J. F., DISAMAR Algorithms and background, RP-TROPOMI-KNMI-066, KNMI, January 2012.

[RD2] de Haan, J. F., TROPOMI ATBD Ozone profile and tropospheric profile, S5P-KNMI-L2-0004-RP, issue 0.10.0, 2014-04-15, in review.

[RD3] van Geffen, J. H. G. M., Boersma, K. F., Eskes, H. J., Maasakkers, J. D. and Veefkind, J. P., TROPOMI ATBD of the total and tropospheric NO2 data products, S5P-KNMI-L2-0005-RP, issue 0.11.0, 2014-10-02, in review.

[RD4] de Smedt, I., van Roozendael, M., Theys, N., van Gent, J. and Danckaert, T., S5P/TROPOMI HCHO ATBD, S5P- BIRA-L2-400F-ATBD, Issue: 0.10.0, 2014-04-15, in review.

[RD5] Landgraf, J., van de Brugh, J., Scheepmaker, R. A., Vitas, N., Borsdorff, T., Houweling, S. and Hasekamp, O. P., Algorithm Theoretical Baseline Document for Sentinel-5 Precursor: Carbon Monoxide Total Column Retrieval, SRON-S5P-LEV2-RP-002, issue 0.10.0, 2014-04-15, in review.

[RD6] Rodgers, C. D., Inverse methods for atmospheric sounding - Theory and practice, vol. 2 of Atmospheric, Oceanic and Planetary Physics, World Scientific, Singapore, 2000.

[RD7] Boersma, K.F., Eskes, H.J. and Brinksma, E.J., Error analysis for tropospheric NO_2 retrieval from space, Journal of geophysical research, 109, 2004.

[RD8] Taylor, J. R., An introduction to error analysis: The study of uncertainties in physical measurements, second edition, University Science Books, Sausalito, California, 1997.

[RD9] OMI website - Instrument Characteristics, 2009, `http://www.knmi.nl/omi/research/instrument/characteristics.php?tag=full`

[RD10] Kleipool, Q. L., Dobber, M. R., de Haan, J. F. and Levelt, P. F., Earth surface reflectance climatology from 3 years of OMI data, J. Geophys. Res., 113, D18308, 2008, doi:10.1029/2008JD010290.

[RD11] Vandaele, A.C., Hermans, C, Simon, P.C., Carleer, M., Colin, R., Fally, S., Mérienne, M.F., Jenouvrier, A. and Coquart B., Measurements of the NO2 absorption cross-section from 42 000 cm-1 to 10 000 cm-1 (238-1000 nm) at 220 K and 294 K, J. Quant. Spectrosc. Radiat. Transfer, 59, 171-184, 1998. doi:10.1016/S0022-4073(97)00168-4

[RD12] Meller, R., and G. K. Moortgat (2000), Temperature dependence of the absorption cross sections of formaldehyde between 223 and 323 K in the wavelength range 225-375 nm, J. Geophys. Res., 105(D6), 7089-7101, doi:10.1029/1999JD901074.

[RD13] Migliorini, S., Piccolo, C. and Rodgers, C. D., 2008: Use of the Information Content in Satellite Measurements for an Efficient Interface to Data Assimilation. Mon. Wea. Rev., 136, 2633-2650. doi:10.1175/2007MWR2236.1

2 Simulation of observations

2.1 Overview

The simulation of retrieved vertical columns $\hat{N}_v$ for weak absorbers is based on the formulation

$$\hat{N}_v = N_v + \sigma_v \varepsilon = \frac{1}{M}(N_s + \sigma_s \varepsilon) = \frac{1}{M}(\sum_l m_l n_l + \sigma_s \varepsilon) = \frac{1}{M}(M \sum_l A_l n_l + \sigma_s \varepsilon) \tag{2.1}$$

where n_l is the vertical column amount in the nature run, m_l is the altitude resolved airmass factor and A_l is the element of the averaging kernel in layer l, respectively, N_v and σ_v are the total column and its error as seen by the instrument ($N_v \neq \sum n_l$), N_s and σ_s are the slant column and its error and ε is a gaussian random number. The airmass factor and averaging kernel are defined by eqs. 4.2 and 4.7, respectively. The latter two formulas provide a connection between the vertical profile of the trace gas in the nature run and the simulated vertical column. The formulas indicate that an estimate of the slant column error and the altitude resolved airmass factor, or the averaging kernel, are needed to simulate the observations for weak absorbers.

The exact computation of columns is different for the weak (NO_2 and HCHO) and the strong (CO) absorber as explained in section 2.2.1. A vertical total column is retrieved for HCHO and CO while a tropospheric column is retrieved for NO_2. Section 4.1.1 explains how the troposheric column and error are extracted for NO_2. The simulation of ozone observations is different from the other gases because vertical profile is retrieved for ozone. The procedure for ozone is detailed in section 7.

The retrieval results could be computed for each observation with the Disamar [RD1] package (fig. 2.1, left). This brute-force approach is however computationally intensive task. Instead, a look-up table approach (sect. 2.2) was selected for the project (fig. 2.1, right). The use of look-up tables allows for much faster generation of the synthetic observations than the brute-force approach. In addition, the values stored in the look-up tables provide an insight on the dependency of the error and the altitude resolved airmass factor on the input parameters. The brute-force approach is used to validate the observations generated with the look-up tables. During the project, draft ATBDs for the Sentinel 5 precursor instrument, TROPOMI, become available to the team ([RD3], [RD4]). It was decided to proceed consistently with the approximations made in the ATBDs for the weak absorbers NO_2 and HCHO: a constant value for the slant column error is now assumed as given in sections 4.2 and 5.2, respectively. For CO, the vertical column error is obtained from simulations with the Disamar package and stored in look-up tables, separately for clear and cloudy atmospheres as explained in section 6.1.

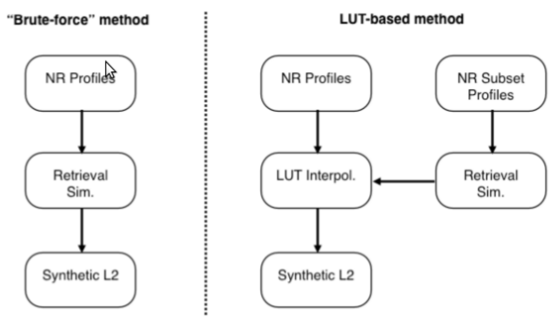


Figure 2.1: Schematic views of the brute force (left) and look-up table (right) approaches.

2.2 Error estimation using look-up tables

2.2.1 Weak and strong absorbers

The error estimation differs for weak (NO_2, HCHO) and strong absorbers (CO, O_3). For the weak absorber the trace gas itself does not affect the light path so much, and the altitude resolved air mass factor is about the same in the fit window. This simplifies the retrieval, and hence the differential optical absorption spectroscopy (DOAS) method can be applied. For strong absorbers, the altitude resolved air mass factor (AMF) is altered via the trace gas itself, and hence that the effective air mass factor changes within the fit window. As a consequence the DOAS method can not be applied for the strong absorbers. In this case optimal estimation ([RD6]) has to be applied, which typically take much more computational time.

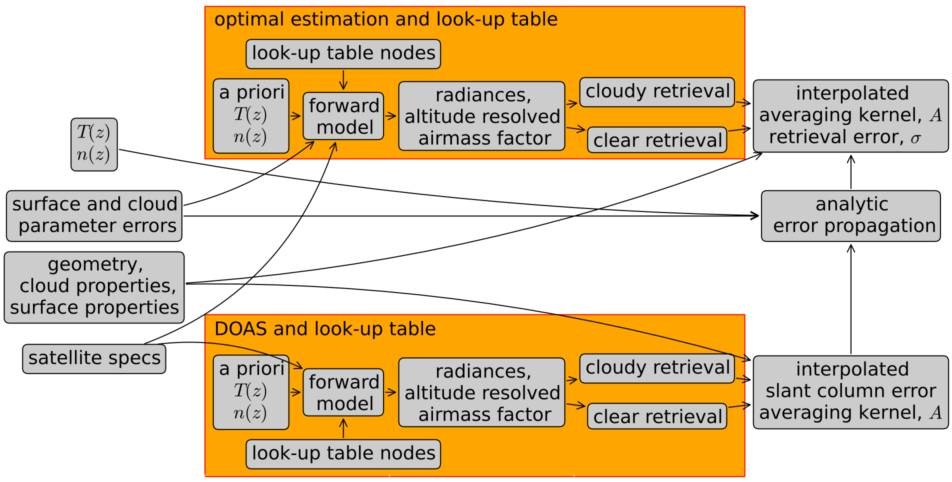


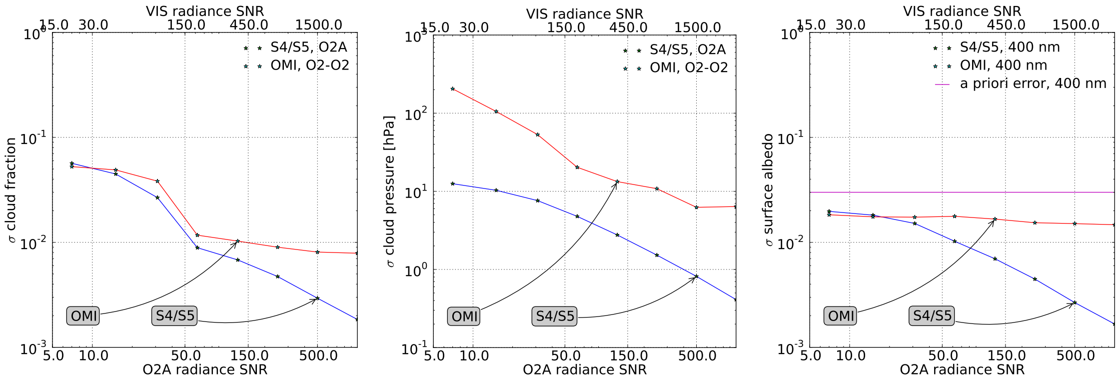
Figure 2.2: Schematic of estimating errors with a look-up table for the Differential Optical Absorption Spectroscopy (DOAS) approach and the optimal estimation (OE, [RD6]) approach. A retrieval error estimation starts with the boxes in the left: a trace gas profile $n(z)$, a temperature profile $T(z)$, surface and cloud parameter errors, geometry, cloud properties, surface properties, and satellite specifications. In the look-up tables errors for a clear and cloudy pixel are saved, which are interpolated with the cloud radiance fraction. For the DOAS method, error propagation is applied with the slant column error, to obtain the (tropospheric) column error.

The DISAMAR ([RD1]) tool can be used to simulate radiance spectra and to do retrievals. We will discuss two methods here: 1. Optimal Estimation (OE, [RD6]) and 2. the Differential Optical Absorption Spectroscopy (DOAS) method. Optimal estimation is an advanced method that can incorporate both the continous and differential part of the measured radiance spectrum. For OE, on-line radiative transfer calculations are needed to calculate the simulated radiance spectrum, and derivatives of the spectrum to the retrieval parameters. The derivatives of the spectrum are used to minimize the difference between the simulated spectrum and the measured spectrum, with which the retrieval parameters are optimized. This process of on-line raditive transfer calculations and optimisation is repeated until convergence occurs. The repeated on-line radiative transfer calculations makes optimal estimation slow. In the DOAS approach, only the differential part of the spectrum is used, and the slant column is determined. The DOAS approach can be applied with a least squares fit of the measured radiance spectrum and a look-up table containing the air mass factors. For error estimation of the retrieved (tropospheric) column cloud and surface parameter errors are used from *a priori*. By using error propagation, the error in the retrieved (tropospheric) column is then estimated.

A schematic of the two approaches to estimate errors with a look-up table is shown in figure 2.2. The look-up table nodes contain geometry, surface properties and cloud properties. One of the main differences between the approaches of error estimation is that for OE the complete error analysis is included in the retrieval. For DOAS error propagation is applied to the slant column error to obtain the vertical (tropospheric) column error. As a consequence, the look-up table approach for DOAS it less dependend on the *a priori* profiles, but for OE this may be much more critical.

2.2.2 Sensitivity to cloud and surface parameters

A key ingredient for error estimation is to characterize the uncertainty in the light path from the sun, through the atmosphere, via scattering or reflection back through the atmosphere to the satellite instrument. More convenient we can use the altitude resolved air mass factor (AMF), which describes how much light goes through each layer in the atmosphere. The AMF can be calculated with a radiative transfer model. The largest uncertainties in the AMF come from uncertainties in the surface albedo, cloud radiance fraction and cloud pressure. In this section we will estimate how much the uncertainties in the cloud and surface parameters will improve for the Sentinel 4/5 instruments.



(a) cloud fraction (b) cloud pressure (c) surface albedo

Figure 2.3: Retrieval errors of surface and cloud parameters, by using optimal estimation. DISAMAR ([RD1]) is used to do a simultaneous retrieval of (a) cloud fraction, (b) cloud pressure and (c) surface albedo, with the instrument specifications of the OMI and Sentinel 4/5 instruments. On the x-axis the signal-to-noise ratio for both the VIS and NIR is simultaneously varied with a 3:1 ratio. OMI uses the O_2-O_2 collision complex (460 - 490 nm) and Sentinel 4/5 uses the O_2A-band (758 - 770 nm) to retrieve cloud information. In the simulation an attempt is done to describe the atmosphere as real as possible: 1. a Henyey-Greenstein phase function, with optical depth of 32 and cloud fraction 0.5, is used to describe a cloud at 400-450 hPa., 2. aerosols with optical depth of 0.5 are added in the boundary layer (900-950 hPa), 3. the surface albedo ($\alpha = 0.05 \pm 0.02$) is described with a sixth degree polynomial, describing a differential structure in the surface reflection. In the retrieval a Lambertian cloud with an albedo of 0.8 is used and a first degree polynomial for the wavelength dependence of surface albedo.

In Fig. 2.3 it can be seen how well surface albedo, cloud radiance fraction and cloud pressure can be estimated in a simultaneous retrieval, for different noise levels. For the surface albedo a climatology is assumed in the retrieval algorithm. We have therefore assumed *a priori* errors for the surface albedo: 0.03 (1 standard deviation) for the visible and 0.10 (1 std) in the O2A band. In table 2.1 an overview is given for the errors of surface and cloud parameters, the signal-to-noise ratios, and fit window specifications. As a reference, the *a priori* errors from [RD7], that are used in tropospheric NO_2 retrievals are given. The uncertainties in cloud pressure, cloud fraction and surface albedo become much smaller for the Sentinel 4/5 instruments. Compared to DOMINO the improvement factors are $\sim 50\times$, $\sim 15\times$, $\sim 7\times$ for cloud pressure, cloud fraction and surface albedo respectively.

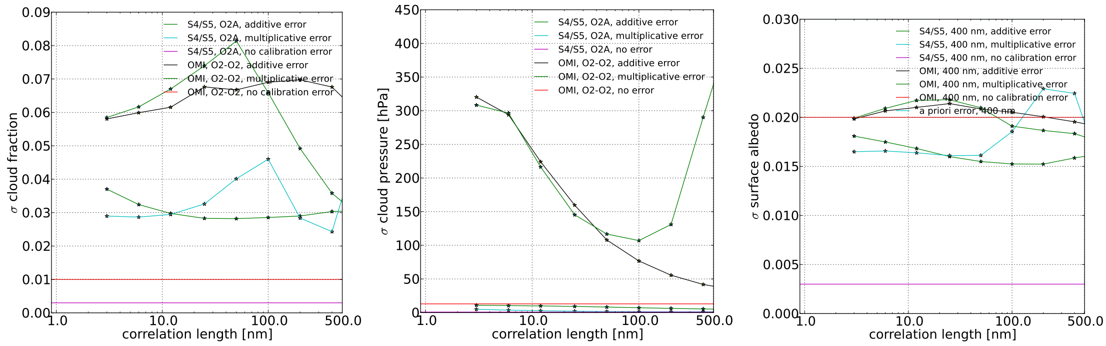
	DOMINO [RD7]	OMI OE simulation	S4/S5 OE simulation
cloud pressure [hPa]	50	13	0.8
cloud fraction	0.05	0.010	0.003
Surface albedo	0.02	0.017	0.003
Strat. Column [10^{-15} molec/cm2]	0.2	-	-
Irradiance SNR	-	5000	5000
Radiance SNR	-	375 (430 nm.)	1500 (430 nm), 500 (758 nm)
fit window [nm]	405-465	400 - 430, 460-490	400-430, 758-770
fit window, step [nm]	0.21 [RD9]	0.21, 0.21	0.20, 0.10
fit window, FWHM [nm]	0.63 [RD9]	0.63, 0.21	0.55, 0.50

Table 2.1: Surface and cloud parameter errors that are used in DOMINO and from an optimal estimation experiment for the OMI and Sentinel 4/5 instrument specifications. Signal to noise ratios of irradiance and radiance and fit window specifications are listed.

In the previous experiment we only added noise to the radiances and irradiances, according to the prescribed signal-to-noise ratios in the MRD ([AD2]). We can also look at how the errors for the cloud and surface parameters change, if we add an additive or multiplicative calibration error. The results of this experiment are shown in Fig. 2.4. In this experiment the spectral correlation length of the calibration error is varied. For a long correlation length ($>$ 1000 nm), the additive (multiplicative) radiance calibration error can be seen as a constant offset (factor). Such a calibration error has a large effect on the continuum. For a short correlation length ($<$ 1 nm), the additive (multiplicative) radiance calibration error has much more a fine structure. The correlation length of the calibration error determines which parameter is affected most, depending on wheter the parameter of interest is depending on the continuum part or the differential part of the spectrum. In Fig. 2.5 we can see that the addition of a straylight fit barely affects the results. Table 2.2 summarizes the simulation results and compares them with the values from the DOMINO study [RD7] and the values selected for ISOTROP. **The ISOTROP values are taken from the NO_2 ATBD for S5P** [RD3]. Most notably, the cloud pressure error is more conservative than given by the simulations with the S4/S5 settings. The cloud fraction and surface albedo errors are in line with the simulations assuming calibration errors.

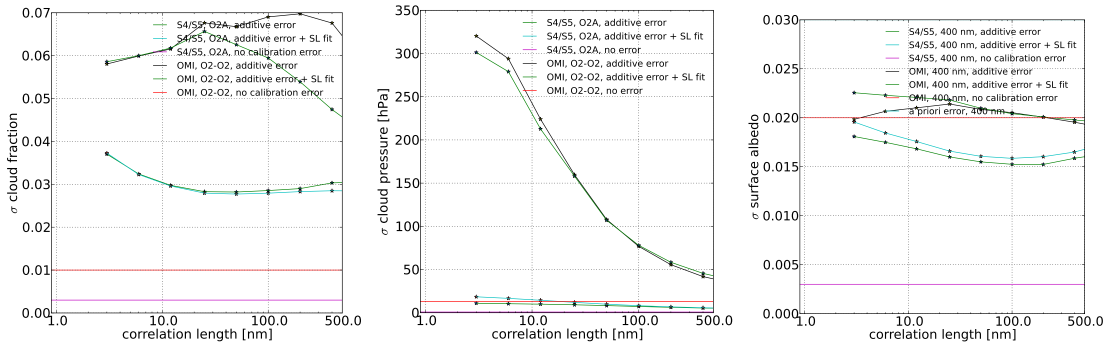
	DOMINO [RD7]	S4/S5 OE simulation			OMI OE simulation			**ISOTROP S5P** [RD3]
		-	+ add.	+ mul.	-	+ add.	+ mul.	
cloud pressure [hPa]	50	0.8	7	1	13	78	107	**50**
cloud fraction	0.05	0.003	0.029	0.046	0.010	0.069	0.066	**0.02**
surface albedo	0.02	0.003	0.015	0.019	0.017	0.021	0.019	**0.015**

Table 2.2: Surface and cloud parameter errors, when an additive or multiplicative calibration error is added to the reflection. In this table the results are listed for a correlation length of 100 nm, and a calibration error of 1%. As a reference also the DOMINO parameter errors are printed. The last column shows the values taken for ISOTROP from the NO_2 ATBD for S5P [RD3]. For more details of the experiment, see the caption in Fig. 2.3.



(a) cloud fraction (b) cloud pressure (c) surface albedo

Figure 2.4: Retrieval errors of surface and cloud parameters, by using optimal estimation. DISAMAR ([RD1]) is used to do a simultaneous retrieval of (a) cloud fraction, (b) cloud pressure and (c) surface albedo, with the instrument specifications of the OMI and Sentinel 4/5 instruments. In this experiment a 1% additive or 1% multiplicative calibration error is added to the reflection. On the x-axis the spectral correlation length is varied. For more details of the experiment, see the caption in Fig. 2.3.



(a) cloud fraction (b) cloud pressure (c) surface albedo

Figure 2.5: Retrieval errors of surface and cloud parameters, by using optimal estimation. DISAMAR ([RD1]) is used to do a simultaneous retrieval of (a) cloud fraction, (b) cloud pressure and (c) surface albedo, with the instrument specifications of the OMI and Sentinel 4/5 instruments. In this experiment a 1% additive added to the reflection and stray light is fitted. On the x-axis the spectral correlation length is varied. For more details of the experiment, see Fig. 2.3.

2.2.3 Overview of the look-up tables

For all look-up tables symmetry in the model of the atmosphere is used. Both the cloud and the surface are assumed to be Lambertian reflectors. Hence that the air mass factors and averaging kernels are therefore the same, when the albedo and pressure are the same. The final result from the look-up table is then obtained by taking a weighted average of the clear and the cloudy part. The slant column error will also be the same in the DOAS approach because in the slant column error the uncertainties in the cloud and surface parameters is not yet accounted for. Hence that for the weak absorbers (NO_2, HCHO) no distinction is made between a cloudy and clear slant column error.

For the strong absorbers (CO, O_3) however, optimal estimation is used, and the column error or error profile is saved, where the surface and cloud parameter uncertainties are taken into account. Here a distinction is made between a cloudy and a clear error. The cloud radiance fraction is used to interpolate between complete clear and complete cloudy cases.

3 Verification of common input data

3.1 Measurement geometries

As a first test, it was confirmed that the simulated S4 and S5P data cover the European domain (15W-35E, 35N-70N) used as the largest assimilation area [AD5]. The data maps are plotted on this area.

The look-up tables are given as a function of solar zenith, viewing zenith and relative azimuth angles. For each gas product, it was verified that the solar and viewing angles were correctly computed. The solar angles are computed from time and location while the viewing angles are computed by the orbit propagator. In the final revised version of the look-up table interpolation code, the relative azimuth angle $\Delta\phi$ is computed from

$$\Delta\phi = \begin{cases} |\phi_s - \phi_v| \bmod 360, & \text{if } |\phi_s - \phi_v| \bmod 360 < 180 \\ 360 - |\phi_s - \phi_v| \bmod 360, & \text{otherwise} \end{cases} \tag{3.1}$$

where ϕ_s and ϕ_v are the solar and viewing azimuth angles, respectively.

Unfortunately, the relative azimuth angle is not stored in the output file. It has to be verified indirectly from the solar and viewing azimuth angles. As an example, figures 3.1 and 3.2 show the solar and viewing angles for S4 and S5P test files, respectively. The viewing zenith angle of S4 reaches ca. 70 degrees in Helsinki (60N), as it should. Also, the azimuth angles are defined in the range [0,360] clock-wise from the North and eq. 3.1 gives the correct relative azimuth for table look-up.

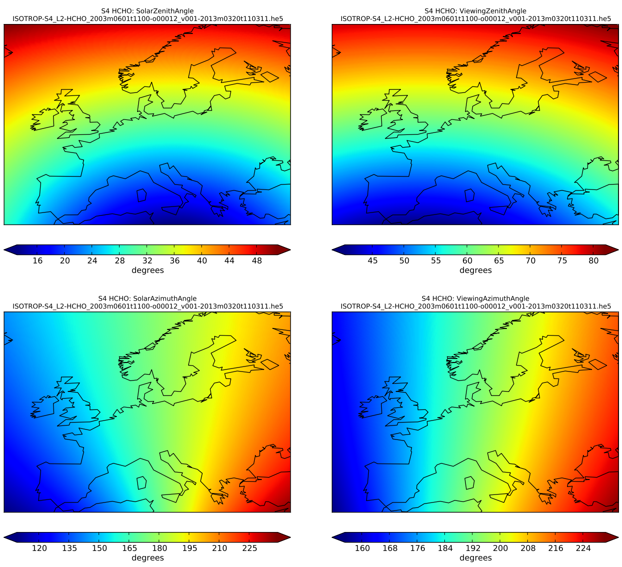


Figure 3.1: Examples of S4 solar (left, top) and viewing zenith (right, top) angles together with solar (left, bottom) and viewing azimuth (right, bottom) angles.

3.2 Surface pressure, cloud top pressure and cloud radiance fraction

Figure 3.3 shows, as an example, surface pressure, cloud top pressure and cloud radiance fraction for a test S5P product indicating realistic values for all parameters. The surface pressure is

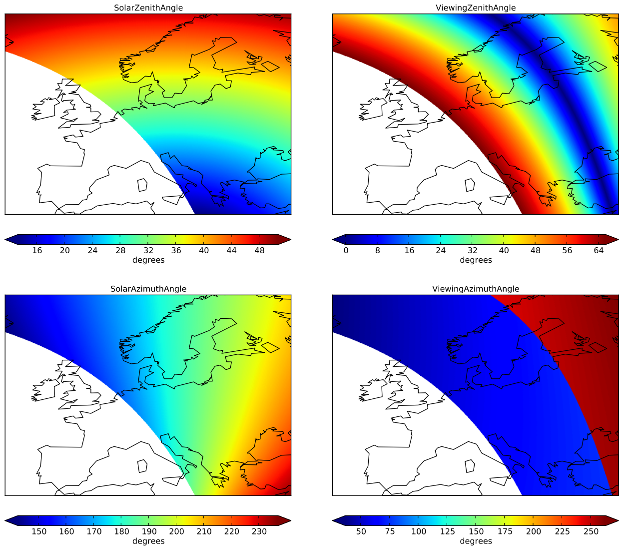


Figure 3.2: Examples of S5P solar (left,top) and viewing zenith (right,top) angles together with solar (left,bottom) and viewing azimuth (right,bottom) angles.

obtained from the ECMWF model input. In a real retrieval, the cloud parameters are retrieved from the O_2-A band measurements. In this simulation study, however, cloud parameters are not retrieved but obtained from the nature run and the errors are estimated. The approach to compute the effective cloud top height and effective cloud fraction from the ECMWF model is described in detail in [AD7]. It was decided to use only the ECMWF model clouds, since the MOCAGE analysed cloud properties showed unrealistic distributions.

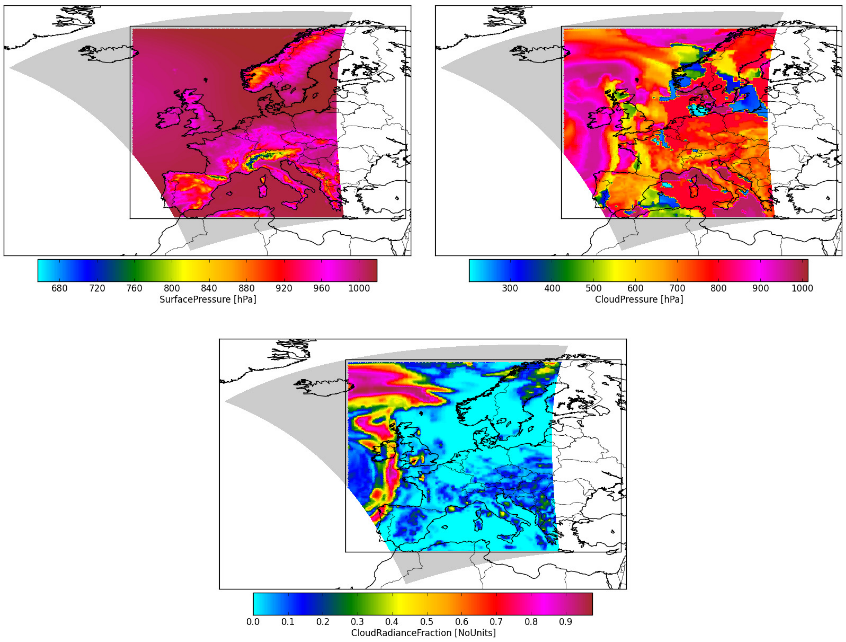


Figure 3.3: Surface pressure, cloud top pressure and cloud radiance fraction for S5P, orbit 21, 1 June 2003, time 12:34. Grey: fill values.

4 Tropospheric NO_2 column

4.1 Theory

4.1.1 Error propagation for NO_2 and averaging kernel

In this subsection it is shown how to get a tropospheric NO_2 column error, starting from a slant column error, that is obtained with the DOAS method. Error propagation is used, just as this is done in DOMINO. The slant column error comes from the look-up table, the error in the stratospheric column is assumed constant, and the errors in the effective air mass factors are estimated with finite differences.

Given a measurement of the total slant column amount of NO_2, N_s, and an *a priori* stratospheric column, $N_{v,str}$. The tropospheric column, $N_{v,tro}$, can then be calculated as:

$$N_{v,tro} = (N_s - M_{str}N_{v,str}) \big/ (M_{tro}) \; . \tag{4.1}$$

The air mass factors (AMFs), either for the stratosphere, M_{str}, or for the troposphere, M_{tro}, is calculated as:

$$M = \sum_l m_l n_l c_l \big/ \sum n_l , \tag{4.2}$$

with $m_l = \partial N_s / \partial n_l$. Here n_l is the vertical column amount in layer l. c_l is the temperature correction, from [RD7]:

$$c_l = \frac{221 - 11.4}{T_l - 11.4} , \tag{4.3}$$

where T_l is the temperature in Kelvin of layer l. Rules for error propagation are used from [RD8]. Assuming uncorrelated errors, the error in the tropospheric column is propagated as:[1]

$$\begin{aligned} \sigma^2_{N_{v,tro}} &= (\frac{\partial N_{v,tro}}{\partial N_s})^2 \sigma^2_{N_s} + (\frac{\partial N_{v,tro}}{\partial M_{str}})^2 \sigma^2_{M_{str}} + (\frac{\partial N_{v,tro}}{\partial N_{v,str}})^2 \sigma^2_{N_{v,str}} + (\frac{\partial N_{v,tro}}{\partial M_{tro}})^2 \sigma^2_{M_{tro}} \\ \sigma^2_{N_{v,tro}} &= (\frac{\sigma_{N_s}}{M_{tro}})^2 + (\frac{N_{v,str}\sigma_{M_{str}}}{M_{tro}})^2 + (\frac{M_{str}\sigma_{N_{v,str}}}{M_{tro}})^2 + (\frac{N_{v,tro}\sigma_{M_{tro}}}{M_{tro}})^2 . \end{aligned} \tag{4.4}$$

The error in the tropospheric AMF is calculated by taking the sum of squares of the all the components that influence the air mass factor.

$$\sigma^2_{M_{tro}} = \sum_* \sigma^2_{M_{tro},*} \tag{4.5}$$

These components are cloud pressure, cloud fraction and surface albedo. E.g. for surface albedo, α_s, the error in the tropospheric AMF is calculated with finite differences:

$$\sigma_{M_{tro},\alpha_s} = \frac{M_{tro}(\alpha_s + \Delta\alpha_s) - M_{tro}(\alpha_s)}{\Delta\alpha_s} \tag{4.6}$$

[1] Equation (4.4) is the same as equation (5) in [RD7], although here the errors the slant stratospheric column $\sigma_{N_{s,str}}$ are explicitly written out in terms of $\sigma_{M_{str}}$ and $\sigma_{N_{v,str}}$.

4.1.2 Averaging kernel

The averaging kernel for the total vertical column amount can be calculated as:

$$A_l = m_l c_l / M \tag{4.7}$$

Where m_l is the air mass factor at layer l, c_l the temperature correction and M the total air mass factor.

The averaging kernel for the troposhperic vertical column amount is calculated as:

$$A_{tro,l} = \begin{cases} m_l c_l / M_{tro} , & \text{if } p_l > 200. \text{ hPa} \\ 0. & \text{if } p_l < 200. \text{ hPa} \end{cases} \tag{4.8}$$

Where the total air mass factor is replaced with the tropospheric air mass factor, and it is zero in the stratosphere.

The tropospheric air mass factor is calculated as:

$$M_{tro} = \sum_l m_l n_l c_l \Big/ \sum n_l \text{ ,with} p_l > 200\text{hPa}. \tag{4.9}$$

with $m_l = \partial N_s / \partial n_l$, the air mass factor at layer l, n_l the vertical column amount in layer l, and c_l the temperature correction from [RD7].

4.1.3 Simulation of the retrieval for weak absorbers

Here we describe how the retrieval values are calculated in the L2 product, from the profiles from the nature run. In a simulated retrieval we can distinguish between truth values and perturbed values for the relevant parameters. For tropospheric NO_2 we start with perturbations on the parameters that are relevant in the retrieval:

$$\text{surface albedo:} \quad \tilde{\alpha}_s = \overline{\alpha}_s + \varepsilon \sigma_{\alpha,s} \tag{4.10}$$

$$\text{effective cloud pressure:} \quad \tilde{P}_e = \overline{P}_e + \varepsilon \sigma_{P,e} \tag{4.11}$$

$$\text{cloud radiation fraction:} \quad \tilde{\omega} = \overline{\omega} + \varepsilon \sigma_{\omega} \tag{4.12}$$

$$\text{vertical stratospheric column:} \quad \tilde{N}_{v,str} = \overline{N}_{v,str} + \varepsilon \sigma_{N,v,str}. \tag{4.13}$$

For example $\overline{\alpha}_s$ is the climatological surface albedo, interpolated from a climatology surface albedo map. For each pixel it is perturbed with a Gaussian random number, ε, multiplied with the uncertainty in the surface albedo $\sigma_{\alpha,s}$. This is also done for the other parameters but their origin differs. The parameters $\overline{P}_e$, $\overline{\omega}$ come from the ECMWF model and $\overline{N}_{v,str}$ comes from the nature run. Consequently we can interpolate the averaging kernels and air mass factors from the look-up table for the truth and perturbed values:

$$\overline{A}_l = A_l(\overline{\alpha}_s, \overline{P}_e, \overline{\omega}) \quad , \quad \tilde{A}_l = A_l(\tilde{\alpha}_s, \tilde{P}_e, \tilde{\omega}) \tag{4.14}$$

$$\overline{M} = M_l(\overline{\alpha}_s, \overline{P}_e, \overline{\omega}) \quad , \quad \tilde{M} = M_l(\tilde{\alpha}_s, \tilde{P}_e, \tilde{\omega}). \tag{4.15}$$

The slant column in the retrieval can then be calculated as:

$$N_s = \tilde{M} * \sum_l \tilde{A}_l n_l + \varepsilon \sigma_{N,s}, \tag{4.16}$$

with $\sigma_{N,s}$ the uncertainty in the slant column in the retrieval, interpolated from the simualed retrieval in the look-up table. Note that the slant column now contains all the perturbations. Consequently the vertical and tropospheric columns are calculated as:

$$N_v = N_s/\overline{M} \tag{4.17}$$

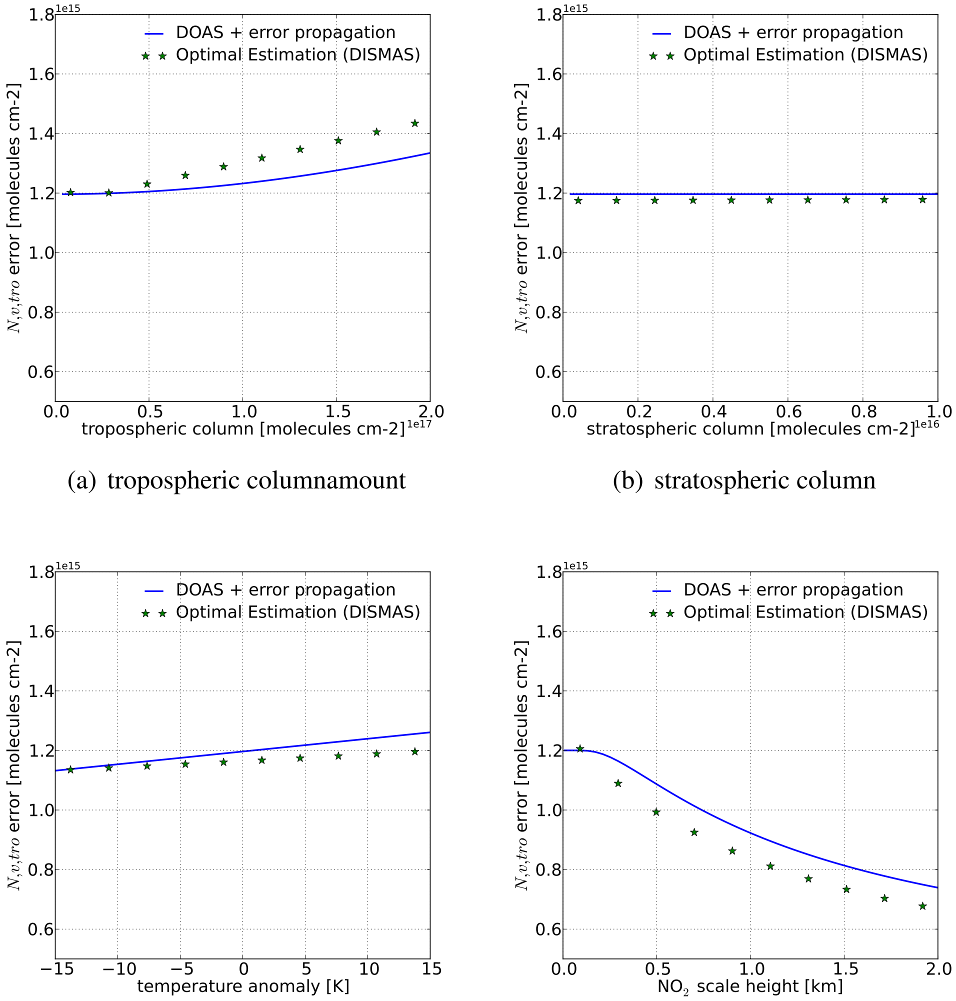
$$N_{v,tro} = \left(N_s - \overline{M}_{str} N_{v,str}\right) / \left(\overline{M}_{tro}\right) . \tag{4.18}$$

Here the unperturbed air mass factors are used. In the L2 product this leads to a consistent set of variables.

4.1.4 Comparison of error propagation with Optimal Estimation

We can ask the question what the influence is of a chosen algorithm on the retrieval error. In this document we started with the estimation of parameter errors of cloud and surface properties. These errors can be used for error propagation, which is typically done for weak absorbers, such as NO_2 in DOMINO. These errors can also be used for an optimal estimation retrieval algorithm as *a priori* errors. Both error propagation and optimal estimation deliver a tropospheric column error estimate. We started estimating surface and cloud parameter errors, by using an optimal estimation technique. As these errors are used as *a priori* in an optimal estimation technique, they should not differ much from the *a posteriori* parameter errors. As a consequence, the two error estimation techniques should entail more or less the same error estimate.

In figure 4.1, it is shown that DOAS+error propagation produce about the same error estimates for the tropospheric NO_2 column as optimal estimation. The only thing we had to do to achieve this, was to estimate the parameter errors first for cloud and surface properties by using optimal estimation. The anwser to the question wheter which retrieval algorithm is best to use is intrinsicly linked to the treatment of surface and cloud parameter errors. If the cloud and surface parameters are obtain with optimal estimation (section 2.2.2), the tropospheric column error is not depending much on the applied technique.



(a) tropospheric columnamount

(b) stratospheric column

(c) temperature

(d) profile shape

Figure 4.1: The retrieval error in the tropospheric NO_2 column by using two techniques: 1. DOAS and error propagation and 2. optimal estimation. In each plot a different variable is varied. (a) Varying the tropospheric NO_2 column amount. (b) Varying the stratospheric NO_2 column amount. (c) Varying the temperature profile. An offset is added to the *a priori* temperature profile that is used in the look-up table. (d) Varying the profile shape. A profile that decays exponentailly with height is used, where the scale height is varied. The solar zenith angle is 30.0 degrees, the viewing zenith angle is 50 degrees, the relative azimuth angle is 90.0 degrees, the surface pressure is 1100 hPa and the surface albedo is 0.03. The test is done for a clear pixel. In this experiment no calibration errors are added.

4.2 Input data and assumptions

4.2.1 Error assumptions

Surface albedo at wavelength 432.5 nm is interpolated from the OMI surface albedo climatology, constructed using 3 years of Ozone Monitoring Instrument (OMI) measurements obtained between October 2004 and October 2007 at 23 wavelengths between 328 and 500 nm for each calendar month in a 0.5° by 0.5° longitude-latitude grid [RD10]. Example surface albedo is shown in Fig. 4.6 (bottom, right). Cloud top pressure and fraction are obtained from ECMWF model (sect. 3.2). Table 4.1 lists the error assumptions used.

Table 4.1: Error assumptions for tropospheric NO_2

Parameter	Error assumption
Cloud top pressure [hPa]	50
Surface albedo	0.015
Cloud fraction	0.02
Stratospheric column [molec/cm^2]	$1.5 \cdot 10^{14}$

4.2.2 Slant column error

Initially, a scene-dependent slant column error was obtained from a look-up table. This approach was changed for the final version of the synthetic data in order to be consistent with the TROPOMI ATBD of the total and tropospheric NO_2 data products [RD3]. A constant slant column error approximation of $7 \cdot 10^{14}$ molec/cm^2 is used for all measurement scenes.

4.2.3 Look-up table for altitude resolved air mass factor

The altitude resolved air mass factors for NO_2 are obtained from a look-up table. One issue which arose was the extrapolation of the AMF (Kernel) to pressures near the surface. This extrapolation is improved by adding a row of "below-surface" values which are consistent with the above-surface AMF values.The axes of the look-up tables are *CloudSurfaceAlbedo*, *CloudSurfacePressure*, *CosSolarZenithAngle*, *CosViewingZenithAngle* and *RelativeAzimuthAngle*. The altitude resolved air mass factor is the same for a cloudy and a clear pixel, when the albedo and the pressure of the Lambertian reflector are the same. Slices of the look-up table are shown in figure 4.3. Figure 4.2 shows the prior profile used for the altitude resolved airmass factor look-up table.

Table 4.2: Characteristics of the look-up table for tropospheric NO_2

Parameter	Notes/Values
Retrieval approach	weak absorber, DOAS approach
Prior information	
Temperature profile	U.S. Standard atmosphere temperature profile
NO_2 profile	Slant column error not depending on *a priori* parameter errors. Only one profile needed. European polluted NO2 profile from CAMELOT study
NO_2 absorption cross-section	Vandaele et al. 1998 [RD11]
Other trace gas profiles	-
Spectral and radiometric settings	
Spectral range [nm]	400 - 465
Spectral resolution (FWHM) [nm]	0.55
Spectral sampling [nm]	0.2
SNR for earth radiance	1500 (not relevant for airmass factor)
SNR for solar irradiance	5000 (not relevant for airmass factor)
Node points	
Cos(SZA)	0.1 (0.10) 1.0
Cos(VZA)	0.3 (0.05) 1.0
Relative azimuth diff. [degrees]	0, 90, 180
Cloud/Surface pressure [hPa]	1100 (100) 200
Cloud/Surface albedo	0.02, 0.04, 0.06, 0.1, 0.2, 0.3, 0.4, 0.8, 0.9
Output	
Altitude resolved air mass factor	at 432.5 nm
AMF pressure levels [hPa] (AMF stored at layer average pressures)	1100, 1000, 900, 800, 700, 600, 500, 400, 300, 200, 137.50, 68.75, 34.38, 17.19, 8.59, 4.30, 2.15, 1.07, 0.54, 0.27, 0.13, 0.07

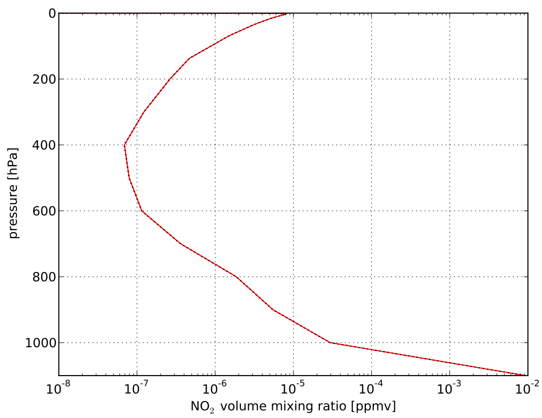
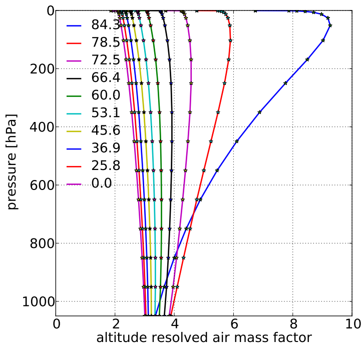
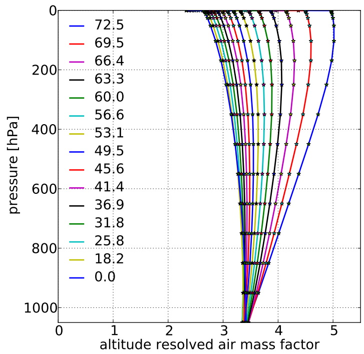


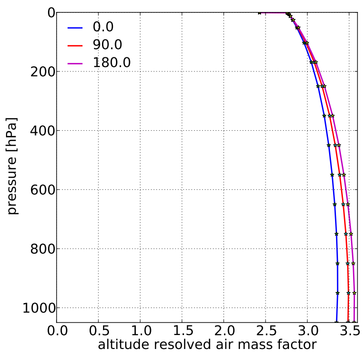
Figure 4.2: Prior profile used in generating the NO_2 altitude resolved air mass factor look-up table.



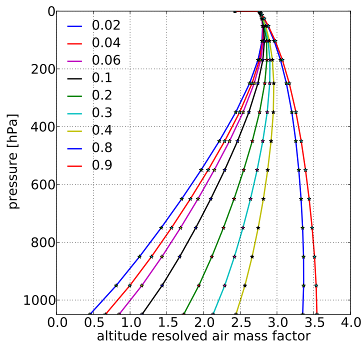
(a) SZA



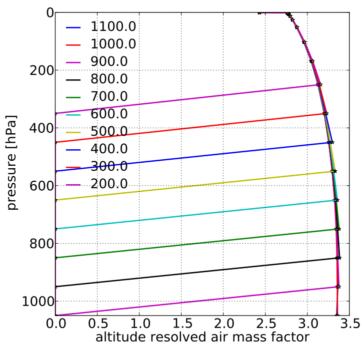
(b) VZA



(c) RAA



(d) albedo



(e) surface/cloud pressure

Figure 4.3: Altitude resolved air mass factors from the look-up table for NO_2. The averaging kernel is obtained from this profile, by multiplying with a temperature correction and air mass factor, see equation 4.7. Layer values are marked with a star. In each figure one ax of the LUT is varied: (a) solar zenith angle, (b) viewing zenith angle, (c) relative azimuth angle, (d) albedo and (e) surface/cloud pressure. Default values are: solar zenith angle 53 degrees, viewing zenith angle 26 degrees, relative azimuth angle 0 degrees, cloud/surface pressure 1050 hPa and albedo 0.8.

4.3 Validation of synthetic data

Figure 4.4 shows the mean RMS of the applied perturbations and the reported error of the retrieved tropospheric columns as a function of the NR tropospheric column for an S5P test orbit. The values are consistent and also agree with the error range of 15-25 % given in the TROPOMI ATBD [RD3].

Random samples of 20 observations were selected for 11 tropospheric column bins covering the range of columns in the Nature Run. Disamar was run for these selected test cases using the measurement geometry, surface albedo, surface pressure and cloud radiance fraction given in the output file of the LUT method, and NO_2 profile from MOGACE given in the intermediate template file. Pressure and temperature profiles were not, however, replaced by the ECMWF data but a priori data were used. Figure 4.5 compares the mean tropospheric column error obtained from the Disamar runs for each bin to the corresponding mean error of the retrieved columns. The values are consistent considering that the slant column error is constant in the LUT method while it is derived from the measurement noise in the Disamar runs. Signal-to-noise ratios of 1500 and 5000 were used for the earth radiance and solar irradiance, respectively, with no calibration errors.

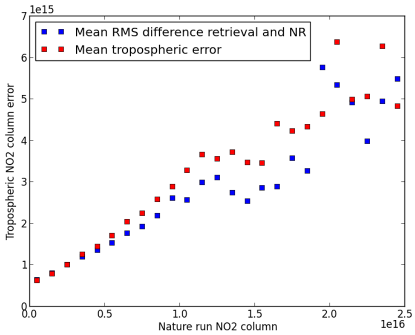


Figure 4.4: Check of the consistency between the mean RMS of the applied perturbations and the reported error of the retrieved tropospheric columns, as a function of the NR tropospheric column. Note that, because of the complicated dependence of the error on the profile and column, and because of the sample size, the two curves do not have to be identical. For the larger columns the error is about 25%. S5P, orbit 25, 1 June 2003, time 12:34. Note that the Tropomi ATBD for NO_2 estimates an error range of 15-25%.

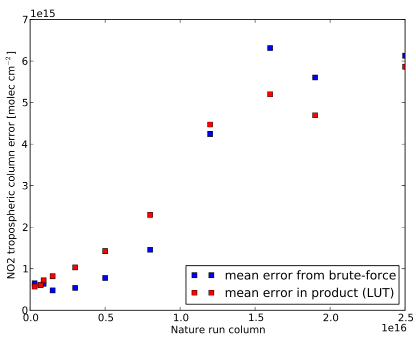


Figure 4.5: Check of the consistency between the mean error from brute-force Disamar runs and the corresponding mean error for the retrieved columns, as a function of the NR tropospheric column. S5P, orbit 25, 1 June 2003, time 12:34.

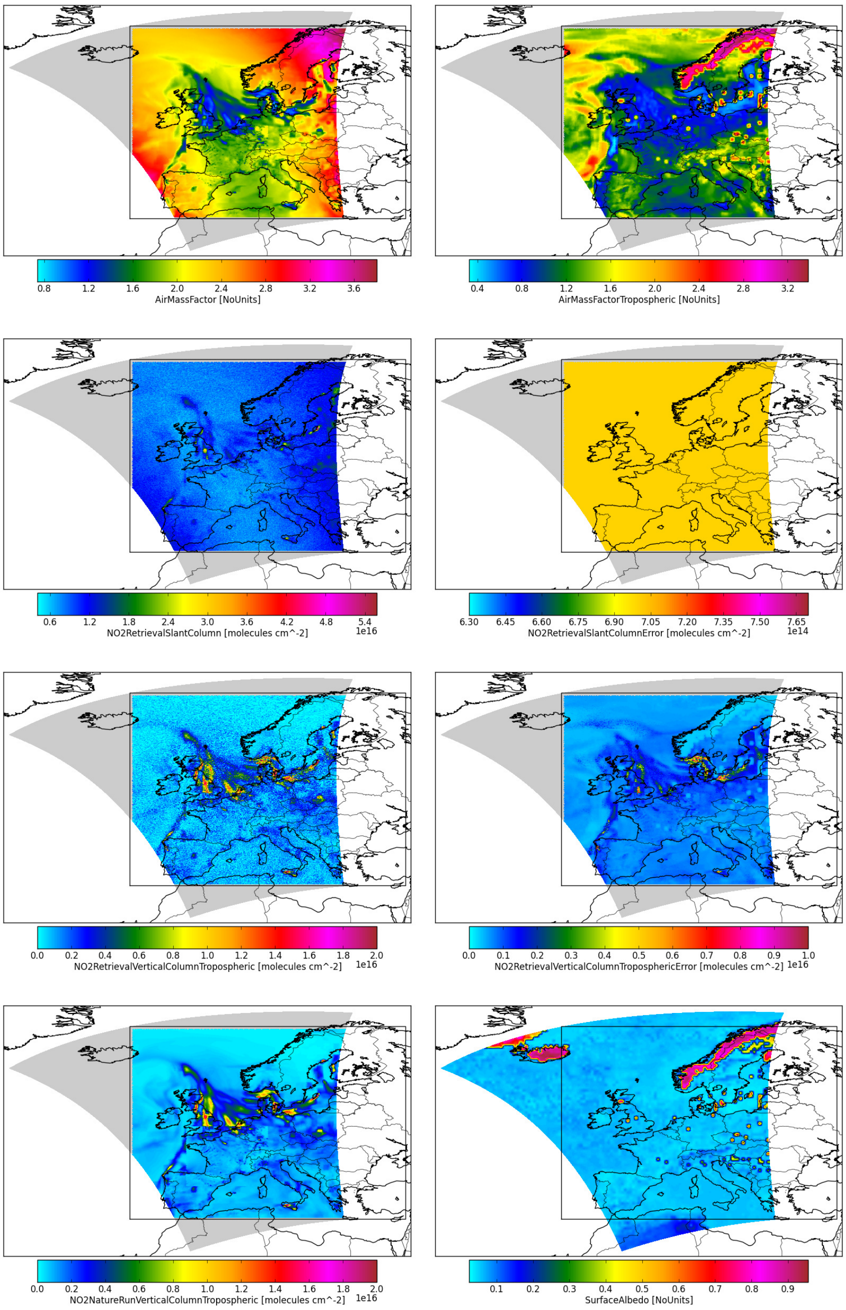


Figure 4.6: Top: Air mass factors for the NO_2 column (left) and the tropospheric column (right). S5P, orbit 21, 1 June 2003, time 12:34. Grey: fill values. 2nd row: Slant column and error. S5P, orbit 21, 1 June 2003, time 12:34. Grey: fill values. Third: Tropospheric vertical NO_2 column and error. Bottom: Nature run tropospheric column and surface albedo. S5P, orbit 21, 1 June 2003, time 12:34. Grey: fill values.

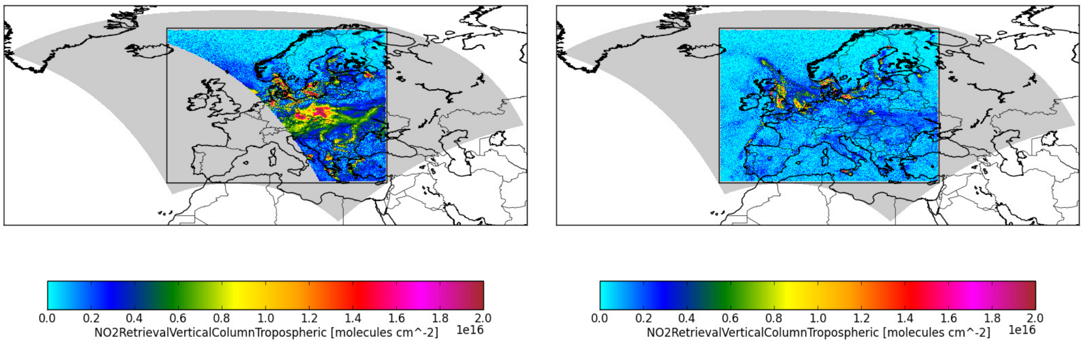


Figure 4.7: Example of two S4 orbits, early morning, 4:00, and mid-day, 12:00, on 1 June 2003. The plots show the large diurnal cycle present in the MOCAGE nature run. Data is plotted for SZA < 85 degree.

Figure 4.6 shows the retrieved tropospheric vertical NO_2 column and its error together with the applied air mass factors, the slant column and its error, the nature run tropospheric column and the surface albedo map at 432.5 nm for a selected S5P test orbit. The tropospheric air mass factor well reflects the cloud features west of the British Isles and the highly reflecting snow surface on the Scandinavian Mountains. As mentioned in Sect. 4.2.2, a constant slant column error of $7 \cdot 10^{14}$ molec/cm^2 is used for all measurements. The retrieved tropospheric columns are consistent with the nature run tropospheric columns.

Figure 4.7 shows the retrieved tropospheric vertical NO_2 column for two selected S4 test orbits: one for early morning and the other for mid-day measurement time. The large diurnal cycle present in the MOCAGE nature run is well captured by the retrieved data.

5 Total HCHO column

5.1 Theory

5.1.1 Error propagation and averaging kernel

The approach for formaldehyde is almost identical to the one for nitrogen dioxide. However, for formaldehyde the retrieval error is for the total column, and not for the tropospheric column. The temperature-independent *Meller* cross sections [RD12] are used for formaldehyde.

5.2 Input data and assumptions

5.2.1 Error assumptions

Surface albedo is interpolated from the same climatology as from NO_2 [RD10] but at 347.5 nm. The error assumptions for formaldehyde are the same as for NO_2 (Table 4.1) except the stratospheric column error is not relevant for the formaldehyde total column.

5.2.2 Slant column error

Similarly to NO_2, a constant slant column error approximation of $1.2 \cdot 10^{16}$ molec/cm^2 given in the TROPOMI ATBD for formaldehyde [RD4] is used for all measurement scenes.

5.2.3 Look-up table for altitude resolved air mass factor

Similary to NO_2, the altitude resolved air mass factors are interpolated from a look-up table. The prior profiles used in computing the air mass factors with Disamar are plotted in figure 5.1. Table 5.1 lists the characteristics and figure 5.2 shows slices of the look-up table.

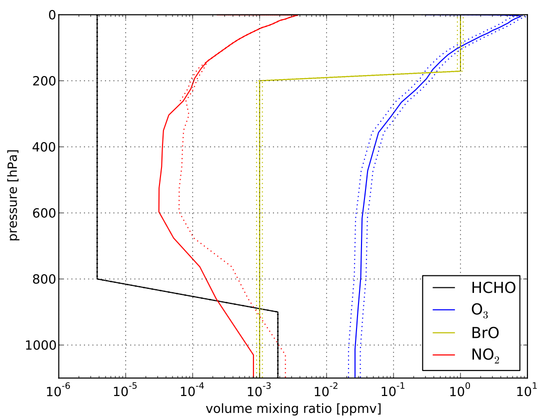
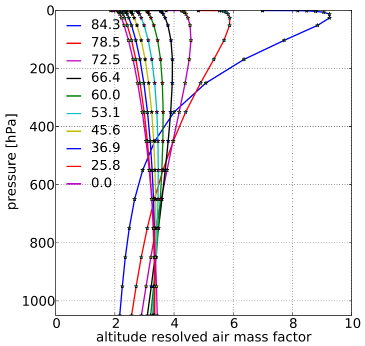


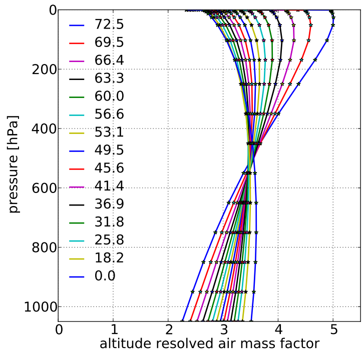
Figure 5.1: Prior profiles used in the formaldehyde retrieval.

Table 5.1: Characteristics of the look-up table for formaldehyde

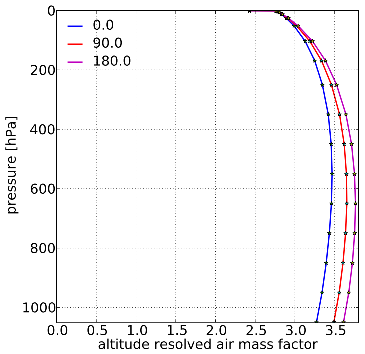
Parameter	**Notes/Values**
Retrieval approach	weak absorber, DOAS approach
Prior information	
Temperature profile	U.S. Standard atmosphere temperature profile
HCHO profile	Slant column error not depending on *a priori* parameter errors. Only one profile needed.
HCHO absorption cross-section	Meller et al. [RD12]
Other trace gas profiles	BrO, O3.
Spectral and radiometric settings	
Spectral range [nm]	335 - 360
Spectral resolution (FWHM) [nm]	0.55
Spectral sampling [nm]	0.2
SNR for earth radiance	1500
SNR for solar irradiance	5000
Node points	
Cos(SZA)	0.1 (0.10) 1.0
Cos(VZA)	0.3 (0.05) 1.0
Relative azimuth diff. [degrees]	0, 90, 180
Cloud/Surface pressure [hPa]	1100 (100) 200
Cloud/Surface albedo	0.02, 0.04, 0.06, 0.1, 0.2, 0.3, 0.4, 0.8, 0.9
Output	
Altitude resolved air mass factor	at 347.5 nm
AMF pressure grid [hPa] (AMF stored at layer average pressures)	1100, 1000, 900, 800, 700, 600, 500, 400, 300, 200, 137.50, 68.75, 34.38, 17.19, 8.59, 4.30, 2.15, 1.07, 0.54, 0.27, 0.13, 0.07



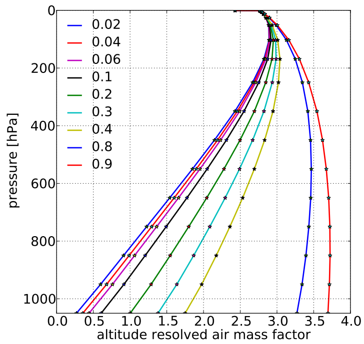
(a) SZA



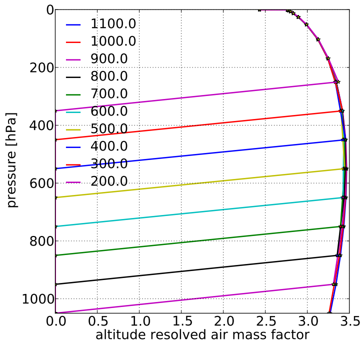
(b) VZA



(c) RAA



(d) albedo



(e) surface/cloud pressure

Figure 5.2: Altitude resolved air mass factors from the look-up table for HCHO. Layer values are marked with a star. In each figure one ax of the LUT is varied: (a) solar zenith angle, (b) viewing zenith angle, (c) relative azimuth angle, (d) albedo and (e) surface/cloud pressure. Default values are: solar zenith angle 53 degrees, viewing zenith angle 26 degrees, relative azimuth angle 0 degrees, cloud/surface pressure 1100 hPa and albedo 0.8.

5.3 Validation of synthetic data

Figure 5.3 (left) compares the RMS difference between the retrieval and the nature run column with the reported error in the retrieved columns as a function of the Nature Run column for an S5P test orbit. The values agree well for columns below $1.0 \cdot 10^{16}$ molec/cm^2 while deviations are observed for larger columns.

Random samples of 10 observations were selected for 11 surface albedo bins covering the whole surface albedo range from 0 to 1. Disamar was run in optimal estimation mode for these selected test cases using the measurement geometry, surface albedo, surface pressure, cloud radiance fraction and the nature run column given in the output file of the LUT method. A priori data were used for the pressure and temperature profiles, and for the formaldehyde vertical profile shape. Disamar was run with two different settings for the SNR of the earth radiance (1000 and 1500) while the SNR was 5000 for solar irradiance in both cases. No calibration errors were applied.

Figure 5.3 (right) compares the mean column error obtained from the Disamar runs for each bin to the corresponding mean error of the retrieved columns for the two different SNR settings as a function of surface albedo. The values are consistent considering that the slant column error is constant in the LUT method while it is derived from the measurement noise in the Disamar runs. As expected, the SNR mostly affects the retrievals over surfaces with low albedo due to a smaller signal than obtained over highly reflecting surfaces. The values obtained with the 1500 SNR setting agree better with the values obtained with the LUT method.

Figure 5.4 shows the retrieved vertical formaldehyde column and its error together with the applied air mass factors, the slant column and its error, the nature run column and the surface albedo map at 347.5 nm for a selected S5P test orbit. Similarly to NO_2, the air mass factor well reflects the cloud features west of the British Isles and the highly reflecting snow surface on the Scandinavian Mountains. As mentioned in Sect. 5.2.2, a constant slant column error of $1.2 \cdot 10^{16}$ molec/cm^2 is used for all measurements. The retrieved tropospheric columns are consistent with the nature run tropospheric columns. For formaldehyde, the noise is of the same order as the columns leading to negative values for the retrieved columns.

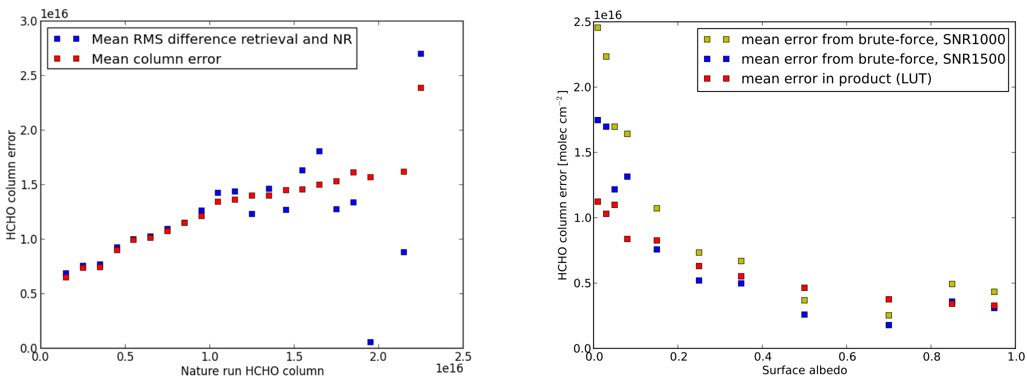


Figure 5.3: Left: check of the consistency between the mean RMS of the applied perturbations and the reported error of the retrieved columns, as a function of the NR HCHO column. Right: check of the consistency between the mean error from brute-force Disamar runs and the reported error of the retrieved columns as a function of the surface albedo for two different setting for the SNR of earth radiance: 1000 and 1500. S5P, orbit 25, 1 June 2003, time 12:34.

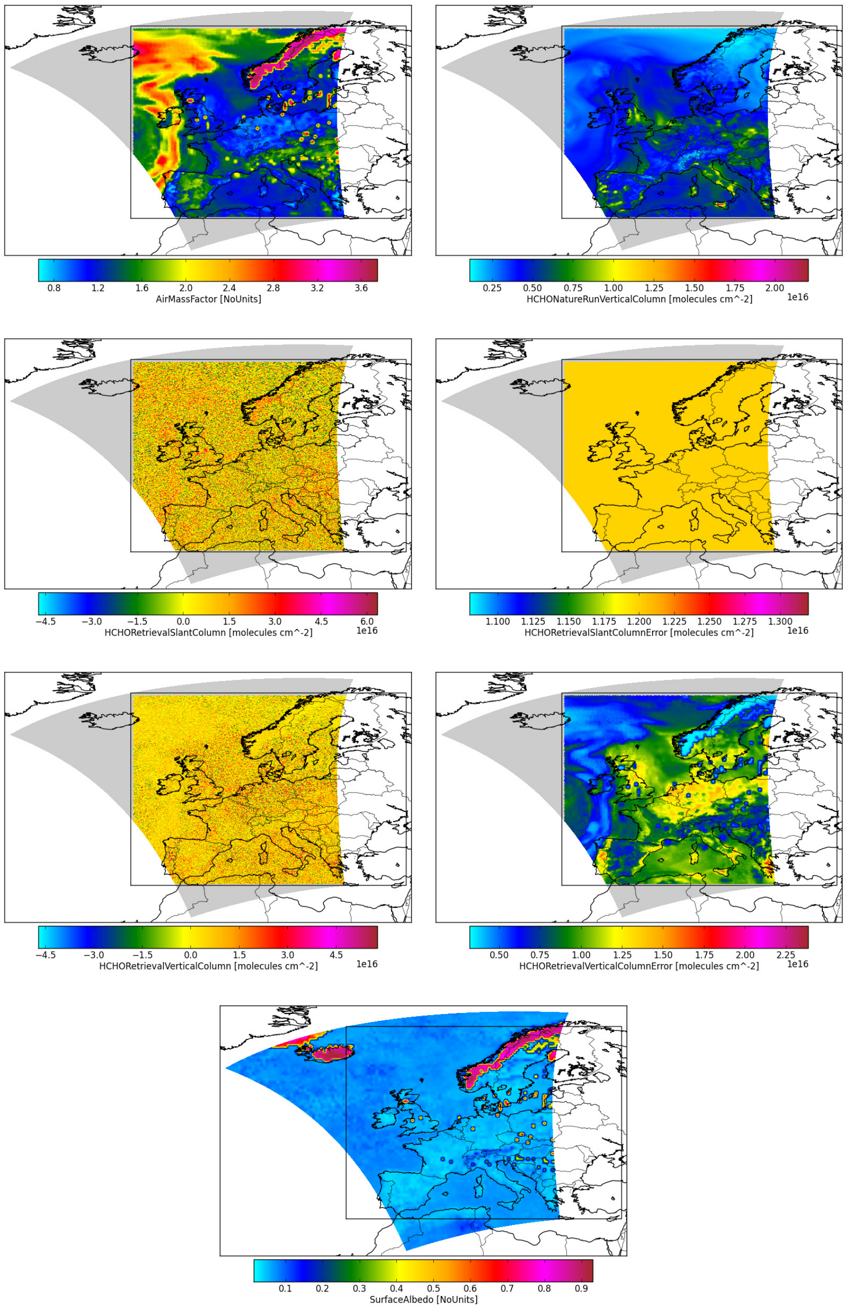


Figure 5.4: Air-mass factor, nature run, retrieval slant column, retrieval slant column error, retrieval vertical column, vertical column error and albedo for HCHO. S5P, orbit 21, 1 June 2003, time 12:34. Grey: fill values. The spots in the surface albedo map are due to missing values.

6 Total CO column

6.1 Theory

For the assimilation of the CO column we would like to know something about the sensitivity with height of the retrieved column. As CO is a strong absorber in the infrared we have to be aware of two things: 1. the altitude resolved air mass factor is strongly influenced by CO absorption itself, 2. scattering is not of importance in the infrared. In the look-up table the air mass factor is saved at 2344 nm, which has litte absorption of CO, H_2O and CH_4. This seems a good estimator of the sensitivity with altitude: It is barely influenced by CO absorption itself but it does capture the blockage of light pad by clouds. For the retrieval of the CO column we can write:

$$\hat{x} = \sum_l A_l n_l + \sigma_{N,v}\varepsilon (= N_v + \sigma_{N,v}\varepsilon) \qquad (6.1)$$

Where $\hat{x}$ is the retrieved CO column, n_l is the column amount in layer l, $\sigma_{N,v}$ is the amplitude of the vertical column error (1 std.), and ε is random number from a normal distribution. A_l is the averaging kernel, for which the air mass factor at 2344 nm is taken. The vertical column error is obtained as a weighted average

$$\sigma_{N,v} = \omega\sigma_{N,v,cld} + (1-\omega)\sigma_{N,v,clr} \qquad (6.2)$$

where $\sigma_{N,v,clr}$ and $\sigma_{N,v,cld}$ are the vertical column errors for the clear and fully cloudy scenes, respectively, and ω is the cloud radiance fraction. Similarly the airmass factor m_l at each layer l is obtained from

$$m_l = \omega m_{l,cld} + (1-\omega) m_{l,clr} \qquad (6.3)$$

where $m_{l,clr}$ and $m_{l,cld}$ are the altitude resolved airmass factors for clear and fully cloudy scenes, respectively. The total airmass factor is computed from

$$M = \frac{\sum m_l}{\sum n_l} \qquad (6.4)$$

using the prior profile, and the elements of the averaging kernel are obtained from

$$A_l = \frac{m_l}{M} \qquad (6.5)$$

The retrieved column is then computed from Eq. 6.1 using the nature run profile.

6.2 Look-up table for vertical column error and altitude resolved airmass factor

The vertical column error $\sigma_{N,v}$ and the altitude resolved air mass factor m_l for clear and fully cloudy scenes are interpolated from a look-up table. Table 6.1 lists the characteristics while figures 6.2 and 6.3 show slices of the look-up table. The prior profiles are plotted in figure 6.1.

6.3 Surface albedo

Surface albedo at 2300 nm is interpolated from a climatology provided by SRON (Paul Tol).

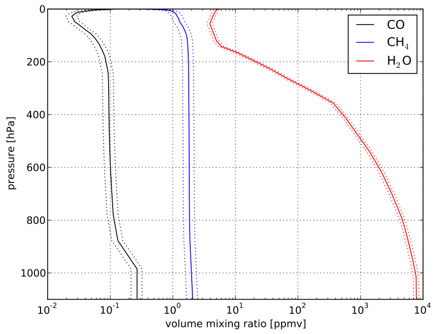
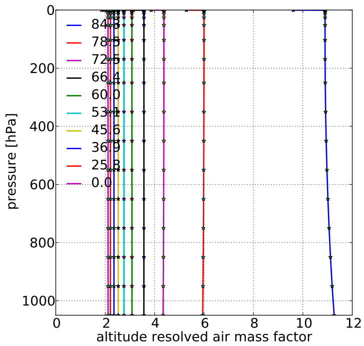


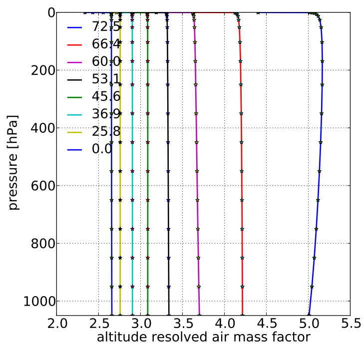
Figure 6.1: A priori profiles used in CO retrieval for the look-up table.

Table 6.1: Characteristics of the look-up table for CO

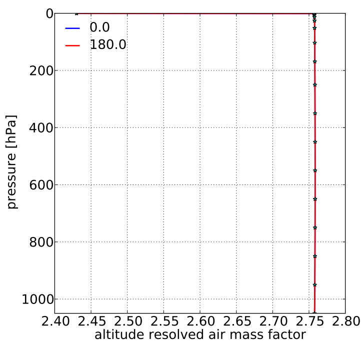
Parameter	Notes/Values
Retrieval approach	strong absorber, optimal estimation
Prior information	
Temperature profile	U.S. Standard atmosphere temperature profile
CO profile	Error in first order not dependent on profile. CO profile for china polluted used based on TM4
CO absorption cross-section	- (HITRAN lines used)
Other trace gas profiles	H_2O and CH_4 have a small effect. U.S. standard atmosphere H_2O profile and CH_4 CAMELOT polluted are used in the retrieval
Spectral and radiometric settings	
Spectral range [nm]	2330 - 2345
Spectral resolution (FWHM) [nm]	0.25
Spectral sampling [nm]	0.1
SNR for earth radiance	120
SNR for solar irradiance	5000
Additive calibration error (%)	1.0, correlation length 100 nm
Node points	
Cos(SZA)	0.1 (0.10) 1.0
Cos(VZA)	0.3 (0.10) 1.0
Relative azimuth diff. [degrees]	0, 180
Cloud/Surface pressure [hPa]	1100 (100) 200
Cloud/Surface albedo	0.0, 0.005, 0.01, 0.02, 0.04, 0.06, 0.1, 0.2, 0.3, 0.4, 0.8, 0.9
Output	
Altitude resolved air mass factor	at 2344.0 nm
AMF pressure grid [hPa] (AMF stored at layer average pressures)	1100, 1000, 900, 800, 700, 600, 500, 400, 300, 200, 137.50, 68.75, 34.38, 17.19, 8.59, 4.30, 2.15, 1.07, 0.54, 0.27, 0.13, 0.07



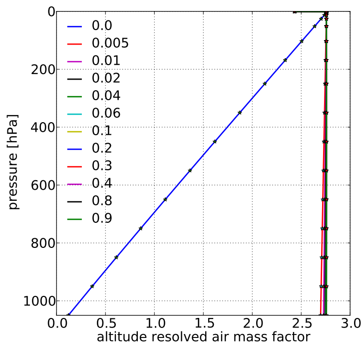
(a) SZA



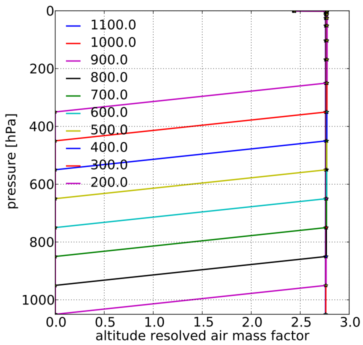
(b) VZA



(c) RAA

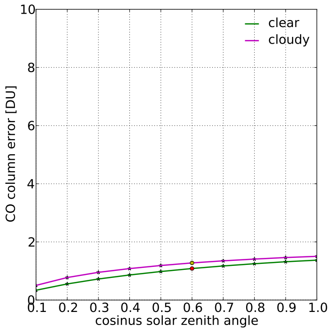


(d) albedo

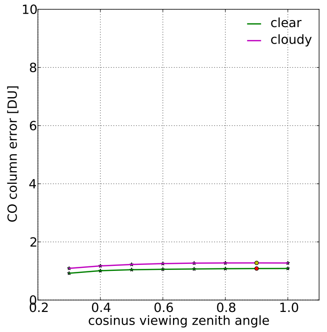


(e) surface/cloud pressure

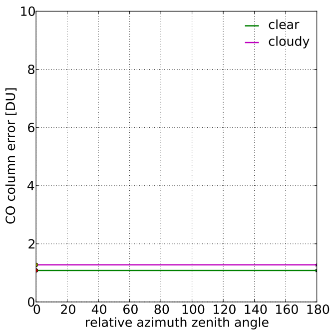
Figure 6.2: Altitude resolved air mass factors from the look-up table for CO. Layer values are marked with a star. In each figure one ax of the LUT is varied: (a) solar zenith angle, (b) viewing zenith angle, (c) relative azimuth angle, (d) albedo and (e) surface/cloud pressure. Default values are: solar zenith angle 53 degrees, viewing zenith angle 26 degrees, relative azimuth angle 0 degrees, cloud/surface pressure 1100 hPa and albedo 0.8.



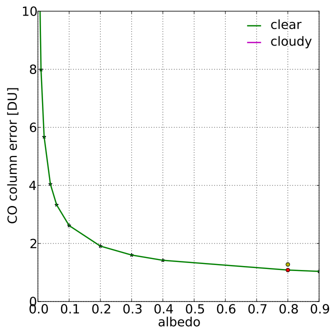
(a) SZA



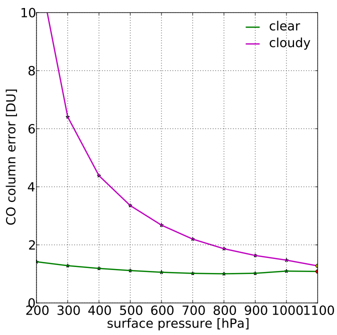
(b) VZA



(c) RAA



(d) albedo



(e) surface/cloud pressure

Figure 6.3: Vertical column errors from the clear and cloudy look-up tables for CO. Layer values are marked with a star. In each figure one ax of the LUT is varied: (a) solar zenith angle, (b) viewing zenith angle, (c) relative azimuth angle, (d) albedo and (e) surface/cloud pressure. Default values are: solar zenith angle 53 degrees, viewing zenith angle 26 degrees, relative azimuth angle 0 degrees, cloud/surface pressure 1100 hPa and albedo 0.8.

6.4 Validation of synthetic data

Figure 6.4 (left) compares the RMS difference between the retrieval and the nature run column with the reported error in the retrieved columns as a function of the nature run column for an S5P test orbit. The values agree well throughout the albedo range studied.

Random samples of 10 observations were selected for 8 surface albedo bins covering the surface albedo range from 0 to 0.5. Disamar was run in optimal estimation mode for these selected test cases using the measurement geometry, surface albedo, surface pressure, cloud radiance fraction and the nature run column given in the output file of the LUT method. A priori data were used for the pressure and temperature profiles, and for the CO vertical profile shape. SNR was set to 120 and 5000 for the earth radiance and solar irradiance, respectively. In addition, 1 % additive calibration error with a 100 nm correlation length was applied as for the generation of the look-up table.

Figure 6.4 (right) compares the mean column error obtained from the Disamar runs for each bin to the corresponding mean error of the retrieved columns as a function of surface albedo. The values agree well although the deviation between the values increases with decreasing surface albedo. The interpolation of the look-up tables give slightly larger errors than Disamar when the albedo is below 0.1. This is likely to result from the rapid increase of column error with decreasing surface albedo as the surface albedo approaches zero (Fig. 6.3(d)). There is very little Rayleigh scattering at these wavelengths and the signal comes mainly from the surface. Therefore, the CO retrieval is very error-prone over black surfaces, e.g. water surfaces.

Figure 6.5 shows the retrieved vertical CO column and its error together with the applied air mass factor, the nature run column and the surface albedo map at 2300 nm for a selected S5P test orbit. The surface level air mass factor and the column error well reflect the cloud features west of the British Isles. The column error also well reflects the larger land surface albedo over Spain. The retrieved tropospheric columns are consistent with the nature run tropospheric columns. As for formaldehyde, the noise is of the same order as the columns leading to negative values in the retrieved columns.

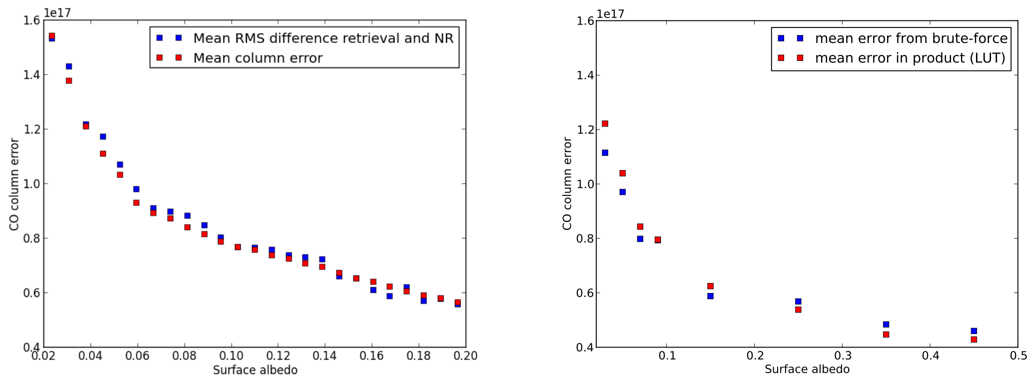


Figure 6.4: Left: check of the consistency between the mean RMS of the applied CO column perturbations and the reported error of the retrieved columns, as a function of the surface albedo. Right: check of the consistency between the mean error from brute-force Disamar runs and the reported error of the retrieved columns, as a function of the surface albedo. S5P, orbit 25, 1 June 2003, time 12:34.

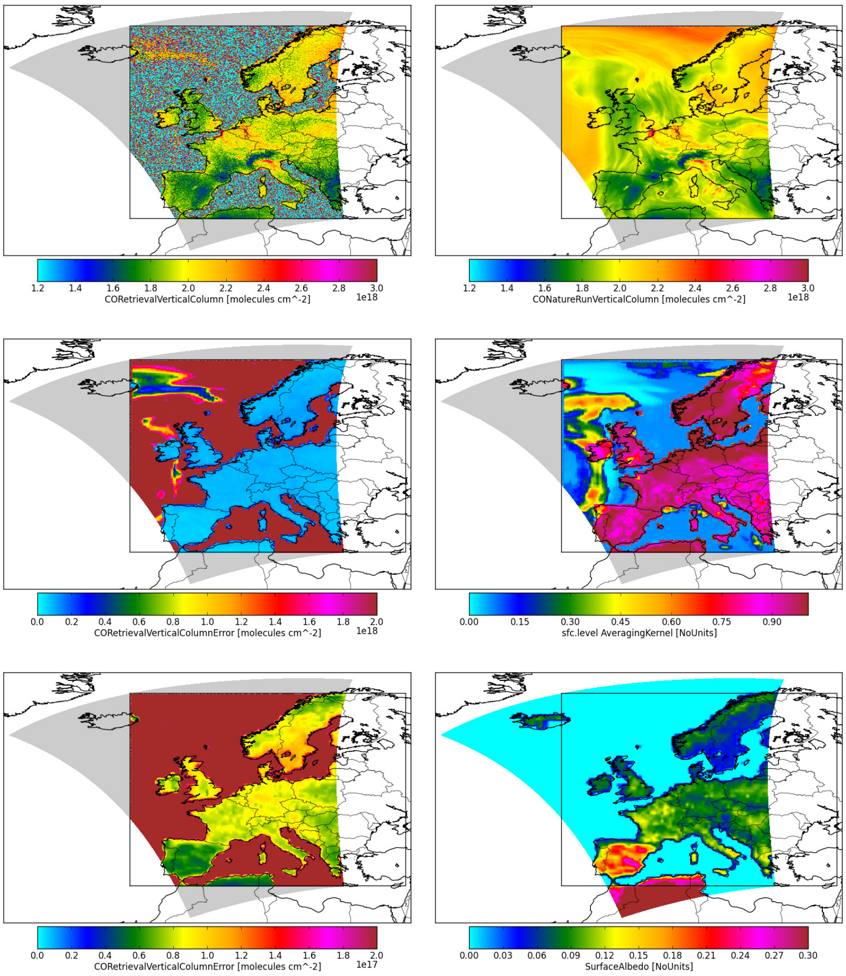


Figure 6.5: Retrieval vertical column, nature run vertical column, retrieval vertical column error showing larger errors, averaging kernel value at the surface, retrieval vertical column error showing smaller errors and surface albedo for CO. S5P, orbit 21, 1 June 2003, time 12:34. Grey: fill values.

7 O_3 profile

7.1 Theory

The ozone profile is retrieved with the optimal estimation (OE) method using the Disamar package. As described by Stefano Migliorini [RD13] the results from the OE can be passed on to a data assimilation system in a very compact way by expressing the retrieval vector in terms of the eigenvectors of the OE SVD. In this section, we translate his approach to the implementation of OE in Disamar.

In Disamar, a basis transformation known as pre-whitening is applied to make the error covariance matrices diagonal [RD1]. This leads to a transformed Jacobian

$$\overline{\mathbf{K}} = \mathbf{S}_\varepsilon^{-1/2}\mathbf{K}\mathbf{S}_a^{1/2} = \mathbf{U}\mathbf{W}\mathbf{V}^T \tag{7.1}$$

where the SVD has been applied. The diagonal matrix $\mathbf{W}$ contains the singular values w_k while the matrix $\mathbf{V}$ contains the singular vectors.

Starting point then are the retrieval, it's relation to the true world through the averaging kernel, and the covariance of the noise:

$$\mathbf{x} - \mathbf{x}_a = \mathbf{A}\left(\mathbf{x}_{true} - \mathbf{x}_a\right) + \mathbf{G}\varepsilon, \tag{7.2}$$

$$\mathbf{S}_{noise} = \mathbf{G}\mathbf{S}_\varepsilon\mathbf{G}^T = \mathbf{A}\mathbf{S} = \mathbf{S}_a^{1/2}\mathbf{V}(\mathbf{G}')^2\mathbf{V}^T\mathbf{S}_a^{1/2} \tag{7.3}$$

$$\mathbf{G}' = \mathrm{diag}\left\{\frac{w_k}{w_k^2+1}\right\} \tag{7.4}$$

From here, the following steps are introduced to come up with an efficient interface to data assimilation.

1. Instead of providing $\mathbf{x}$ one can provide a "shifted" state to the assimilation, with the same result, but removing the need to provide the a-priori profile:

$$\mathbf{x}^{(a)} = \mathbf{x} - \left[\mathbf{I} - \mathbf{A}\right]\mathbf{x}_a = \mathbf{A}\mathbf{x}_{true} + \mathbf{G}\varepsilon, \tag{7.5}$$

With the same covariance $\mathbf{S}_{noise}$ as $\mathbf{x}$.

2. Next, a new "rotated" solution can be defined which represents the result in the eigenspace:

$$\mathbf{x}^{(b)} = \mathbf{V}^T\mathbf{S}_a^{-1/2}\mathbf{x}^{(a)} = \mathbf{V}^T\mathbf{S}_a^{-1/2}\mathbf{A}\mathbf{x}_{true} + \mathbf{V}^T\mathbf{S}_a^{-1/2}\mathbf{G}\varepsilon, \tag{7.6}$$

With the diagonal covariance,

$$\Lambda = \mathbf{V}^T\mathbf{S}_a^{-1/2}\mathbf{S}_{noise}\mathbf{S}_a^{-1/2}\mathbf{V}^T = \mathrm{diag}\left\{\left[\frac{w_k}{w_k^2+1}\right]^2\right\}, \tag{7.7}$$

3. The storage of a covariance matrix can be avoided by scaling the solution with the eigenvalues

$$\mathbf{x}^{(c)} = \Lambda^{-1/2}\mathbf{x}^{(b)} = \Lambda^{-1/2}\mathbf{V}^T\mathbf{S}_a^{-1/2}\mathbf{A}\mathbf{x}_{true} + \varepsilon^{(c)} = \mathbf{A}^{(c)}\mathbf{x}_{true} + \varepsilon^{(c)}, \tag{7.8}$$

$$\mathbf{x}^{(c)} = \Lambda^{-1/2}\mathbf{V}^T\mathbf{S}_a^{-1/2}\left(\mathbf{x} - \left[\mathbf{I} - \mathbf{A}\right]\mathbf{x}_a\right) \tag{7.9}$$

Where the covariance of $\varepsilon^{(c)}$ is the identity matrix $\mathbf{I}$. The transformed averaging kernel is now

$$\mathbf{A}^{(c)} = \Lambda^{-1/2}\mathbf{V}^T\mathbf{S}_a^{-1/2}\mathbf{A} = \Lambda^{-1/2}\mathbf{G}'\mathbf{W}\mathbf{V}^T\mathbf{S}_a^{-1/2} \tag{7.10}$$

$$\mathbf{A}^{(c)} = \mathrm{diag}\left\{w_k\right\}\mathbf{V}^T\mathbf{S}_a^{-1/2}, \tag{7.11}$$

4. Finally, only the leading q eigenvectors may be provided by storing only the first q elements of $\mathbf{x}^{(c)}$ and the first q rows of $\mathbf{A}^{(c)}$.

Storage is now reduced from $2n + 2n^2$ (two states and two full matrices), to $q + qn$ (for the transformed kernel+state). For instance, for 40 vertical levels and DFS = 8, the size reduction is a factor of 10. Furthermore, the assimilation effort is significantly reduced because ony q observations are assimilated instead of n, and because there are no correlations between the observations. The noise-dominated retrieval components are removed. Also, the risk of ending up with negative eigenvalues of the covariance matrix is avoided.

7.2 Look-up table

Table 7.1 lists the characteristics the look-up table. We focus on the tropospheric ozone and therefore apply the 300-320 nm spectral window. The wavelength dependent SNR is defined at 300, 310 and 320 nm.

Figure 7.1 plots the prior ozone profile and covariance matrix used in the optimal estimation retrievals for the look-up table. The prior profile is from the US 1976 standard atmosphere with 30 % error bars. The off-diagonal elements of the prior covariance are compute using a correlation length of 6 km. Correlation between the troposphere and the stratosphere is removed.

Figure 7.2 shows examples of the six leading eigenvectors of the transformed averaging kernel for the ozone profile part. The first eigenvector provides weighting for the total column while the following vectors are sensitive to fine structures. In the troposphere, the sixth eigenvector is in the noise level and therefore it is sufficient to store only six eigenvectors from the transformed avaraging kernel. Figure 7.3 shows the eigenvector elements for the cloud fraction, cloud pressure and the surface albedo indicating also that six eigenvectors are sufficient.

We applied a postprocessing step to the lookup-table to avoid interpolation between kernel vectors of opposite sign. All nearest-neighbor pairs in the LUT were compared. In case of a large overlap between the vectors, but with opposite sign (negative inner product), one of the vectors is multiplied by -1.

The wavelength range used for the ozone retrievals is 300-320 nm. Profiles for each 7x7 high-resolution footprint are only available for this range. There is less information in the upper stratosphere as compared to a retrieval on the 270-320 nm range, and the total DFS is reduced from about 8 to about 5. In the troposphere the DFS is also reduced, but only a little bit, due to the fact that the photons in the range 270-300 nm are reflected predominantly in the middle-upper stratosphere. The first couple of observations (eigenvalues) mainly constrain the stratosphere, but observation 4-5 contain also information on the troposphere. These kernel vectors always show sensitivity both in the troposphere and lower stratosphere. This implies that a good stratospheric data assimilation analysis is needed in order to extract the tropospheric information. The 6 observations have been rescaled in such a way that the observation error equals 1 for each of the 6 retrievals, and there are no correlations between the 6 observations (see equations above). The kernel is provided for 21 vertical ozone levels, which are now the native levels for the retrieval in Disamar. The ozone profile unit is ln(vmr / 1ppmv), the natural log of the mixing ratio. This is the internal representation of the profile in DISAMAR. Interpolations

have been reduced as much as possible. We have used a profile interpolation scheme to map the NR profile on the retrieval grid. This scheme extrapolates outside the domain based on a constant mixing ratio assumption.

For the assimilation the only fields in the synthetic L2 files needed are "AveragingKernel_transformed_lnvmr" and "O3RetrievalState_lnvmr". The "O3AprioriProfilePressures" define the pressure grid.

The meaning of the rotated kernel vectors can be understood in the following way:
- The mean size of the observation, or kernel, is a measure of the signal-to-noise ratio of the observation.
- The ratio between the observation and the kernel is a measure of the amount of ozone.
- The normalised kernel profile tells from which part of the profile the information results.

Table 7.1: Characteristics of the look-up table for O_3.

Parameter	**Notes/Values**
Retrieval approach	strong absorber, optimal estimation
Prior information	
Temperature profile	U.S. Standard atmosphere temperature profile
O_3 profile	Error profile hardly depends on the ozone profile. U.S. standard atmosphere ozone profile is used. *A priori* error is 30% for the whole profile. A priori correlation length of 6 km is used.
O_3 absorption cross-section	Brion et. al
Other trace gas profiles	-
Spectral and radiometric settings	
Spectral range [nm]	300 - 320
Spectral resolution (FWHM) [nm]	0.5
Spectral sampling [nm]	0.06
SNR for earth radiance	50 at 300 nm, 300 at 310 nm, 1000 at 320 nm, linear interpolation in between
SNR for solar irradiance	5000
Node points	
Cos(SZA)	0.1 (0.10) 1.0
Cos(VZA)	0.3 (0.10) 1.0
Relative azimuth diff. [degrees]	0, 60, 120, 180
Cloud/Surface pressure [hPa]	1050, 970, 890, 801, 701, 601, 501, 401, 301, 201
Cloud/Surface albedo	0.02, 0.04, 0.06, 0.10, 0.20, 0.30, 0.40, 0.8, 0.9
Output	
AveragingKernel_transformed_lnvmr	From Eq. 7.11: for ozone profile in the pressure grid defined below, cloud fraction, cloud top pressure and surface albedo (24 elements in total)
Pressure grid for ozone profile [hPa]	1050, 950, 850, 750, 650, 550, 450, 350, 250, 168.8, 103.1, 51.56, 25.78, 12.89, 6.445, 3.222, 1.611, 0.8055, 0.403, 0.2015, 0.1005

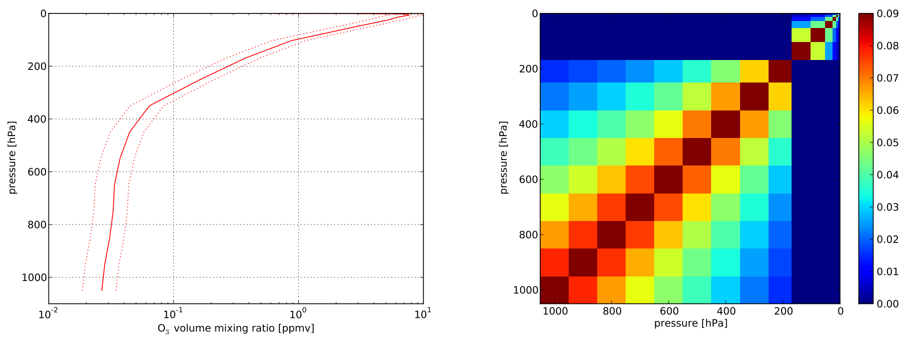


Figure 7.1: Prior ozone profile (left) and covariance (right) used in generating the O_3 LUT. 30 % error in the volume mixing ratio is assigned to the profile and 6 km correlation length is used to compute the off-diagonal elements in the prior covariance. The correlation between the troposphere and stratosphere is removed. The tropopause is at ca. 170 hPa.

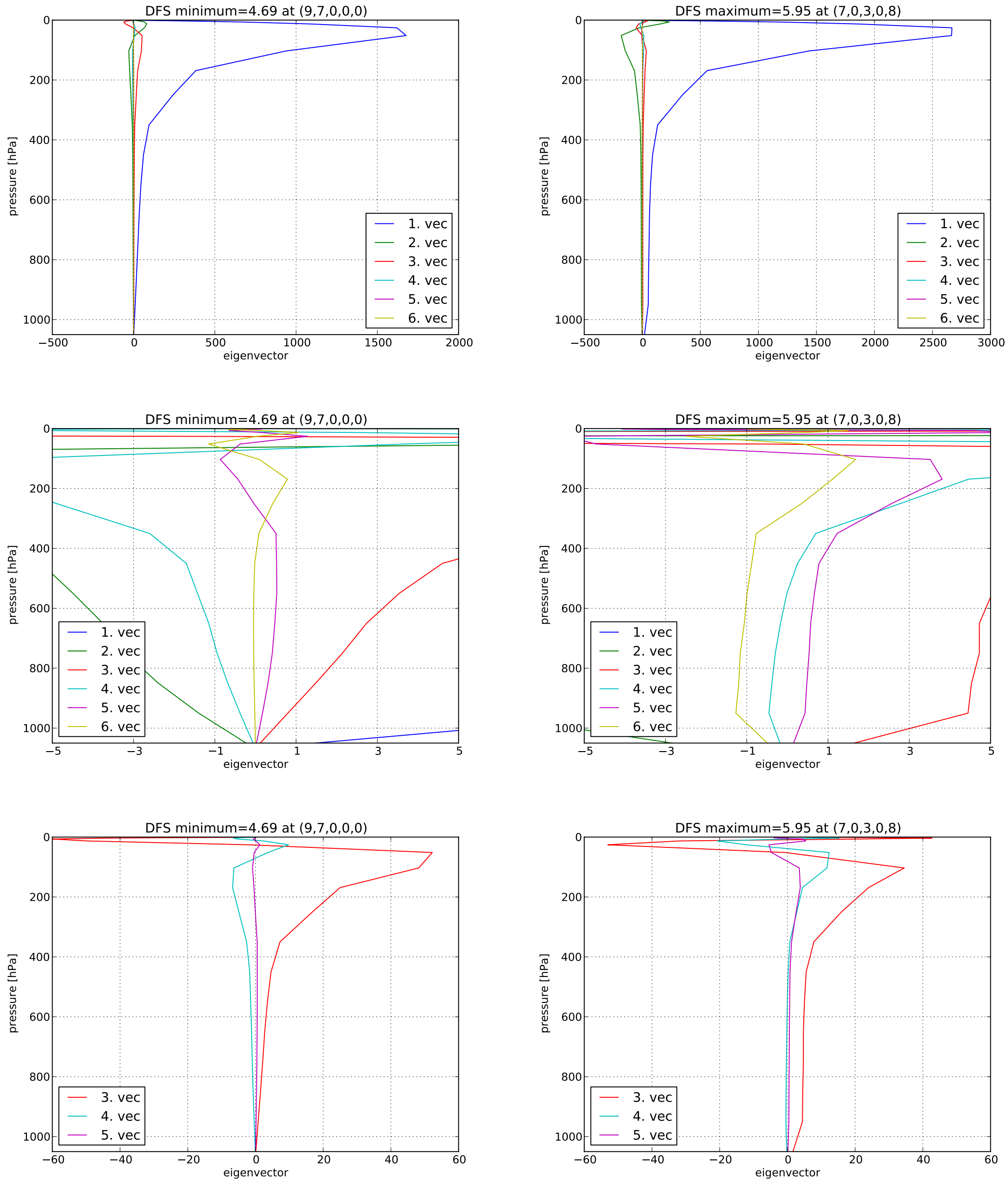


Figure 7.2: Ozone profile eigenvectors at the minimum (left) and maximum (right) degrees of freedom of signal (DFS) for 1050 hPa surface pressure on a full scale (top), zoomed to show the small values in the troposphere (middle) and showing only eigenstates from 3 to 5. The DFS minimum occurs for at LUT nodes: $\theta_s = 0.0°$, $\theta_v = 0.0°$, $\Delta\phi = 0.0°$ and $A_{surf} = 0.02$. The DFS maximum occurs for at LUT nodes: $\theta_s = 36.9°$, $\theta_v = 72.5°$, $\Delta\phi = 180.0°$ and $A_{surf} = 0.9$. In both cases the sixth eigenvector is in the noise level ($\pm$ 1). The states 4-5 have comparable values in the troposphere and stratosphere, and carry information from the troposphere.

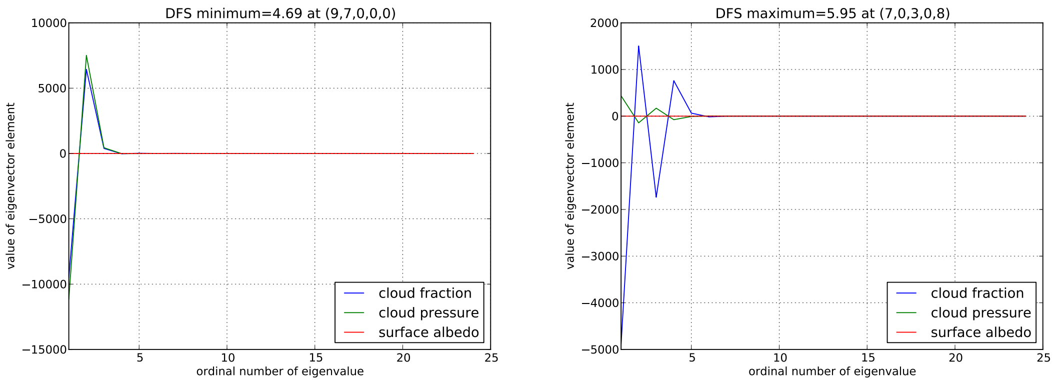


Figure 7.3: Value of the eigenvector element for the cloud fraction, cloud pressure and surface albedo as a function of the ordinal number of the eigenvalue in the same conditions as in Fig. 7.2. The values decrease rapidly after five leading eigenvalues.

Annexes

A Structure of the look-up tables

A hdfview snapshot of the NO2 look-up table, to illustrate the structure of the look-up table, can be seen in Fig. A.1.

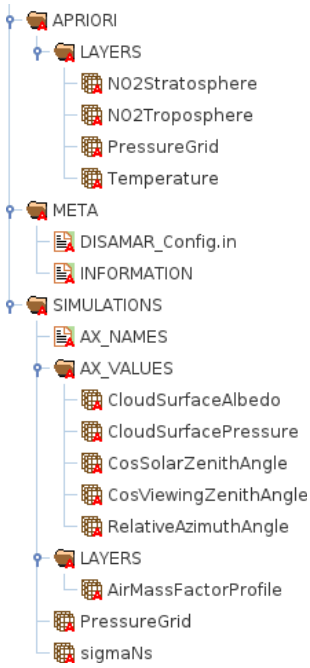


Figure A.1: Structure of the look-up table. The *a priori* profiles for temperature, tropospheric NO_2 and stratospheric NO_2 are given on layer values. In the simulations the retrieval error $\sigma_{N,v,tro}$ (*SigmaNvtro*) and the air mass factor profile M (*AirMassFactorProfile*) are calculated for the axes of the look-up table.

B Interface

An interface is written in Python that is able to generate L2 files from an intermediate file. It constists of two files: error_estimation.py and mod_interface.so. The first file is the executable and the second one is a dedicated fortran module. As input intermediate files are needed with a specific structure. Orbital information and trace gas profiles should be in here. See the example files for the details. The interface can be used as:

```
>>> error_estimation.py <intermediate file>
```

For example:

```
>>> error_estimation.py small_template.he5
```

will generated 4 new files: small_L2-CO.he5, small_L2-HCHO.he5, small_L2-NO2.he5, small_L2-O3.he5.

The settings of the interface can be adjusted in the Python code:

```
#ISOTROP input errors, used for error propagation in NO2 and HCHO
pro_error = {}
pro_error['cloudpressure']     = 10
pro_error['surfacealbedo']     = 0.015
pro_error['cloudfraction']     = 0.03
pro_error['sigma.N,v,str']     = 0.2E15

#adjust these settings to thy needs, and update this with the appropiate paths
settings = {}
settings['tracegases'] = ['NO2', 'HCHO', 'CO', 'O3']    #tracegases that will be process
settings['tracegase_files'] = {}
settings['tracegase_files']['CO']          = '../all_LUTs/ISOTROP_CO_LUT_v1.2.h5'
settings['tracegase_files']['HCHO']         = '../all_LUTs/ISOTROP_HCHO_LUT_v1.2.h5'
settings['tracegase_files']['NO2']          = '../all_LUTs/ISOTROP_NO2_LUT_v1.2.h5'
settings['tracegase_files']['O3']           = '../all_LUTs/ISOTROP_O3_LUT_v1.2.h5'

settings['intermediate-tag'] = 'template'

#fields names
settings['field_names'] = {}

#pressure field names
settings['field_names']['P_NO2']             = 'MOCAGE_Press_prof'
settings['field_names']['P_HCHO']            = 'MOCAGE_Press_prof'
settings['field_names']['P_CO']              = 'MOCAGE_Press_prof'
settings['field_names']['P_O3']              = 'MOCAGE_Press_prof'

#trace gas profile field names, all are assumed to by mixing ratios
settings['field_names']['prof_NO2']          = 'MOCAGE_NO2_prof'
settings['field_names']['prof_HCHO']         = 'MOCAGE_HCHO_prof'
settings['field_names']['prof_CO']           = 'MOCAGE_CO_prof'
settings['field_names']['prof_O3']           = 'MOCAGE_O3_prof'

#surface albedo field names
settings['field_names']['SA_NO2']            = 'SurfaceAlbedo_NO2'
settings['field_names']['SA_HCHO']           = 'SurfaceAlbedo_HCHO'
settings['field_names']['SA_CO']             = 'SurfaceAlbedo_CO'
settings['field_names']['SA_O3']             = 'SurfaceAlbedo_O3'

#temperature field names
settings['field_names']['P_temp']            = 'ECMWF_Pres_prof'
settings['field_names']['prof_temp']         = 'ECMWF_Temp_prof'
```