SCIAMACHY LEVEL 1B-2 DATA PROCESSING:

UPDATE OF OFF-LINE DATA PROCESSOR TO VERSION 3.0

A. von Bargen[1], T. Schröder[1], A. Doicu[1], K. Kretschel[1], Christophe Lerot[2], M. Van Roozendael[2], A. Kokhanovsky[3], M. Vountas[3], H. Bovensmann[3], M. Hess[1], B. Aberle[1], F. Schreier[1]

[1]*German Aerospace Center, Wessling, Germany, Email: Albrecht.von-Bargen@dlr.de*
[2]*Belgian Institute for Space Aeronomy, Brussels, Belgiums, Email: michelv@oma.be*
[3]*University of Bremen, Bremen, Germany, Email: heinrich.bovensmann@iup.physik.uni-bremen.de*

ABSTRACT

The new operational level 1b-2 Off-line data processor for SCIAMACHY contains several updates and evolutions in comparison to the former operational version. The retrieval from nadir measurements had been extended with the cloud algorithm SACURA to obtain the cloud-top height and cloud optical thickness as additional product. The retrieval of total columns for O3 and NO2 is now in-line with the scheme applied in the version 4 of the GOME data processor. In contrast to GDP, the cloud-top height and cloud optical thickness from SACURA are used instead of ROCINN. In addition, the retrieval of O_3 and NO_2 profiles from limb measurements is now based on the retrieval package DRACULA. The presentation provides an overview about the evolution of the processor and will close with an outlook to coming versions.

1. INTRODUCTION

Until summer 2006 the off-line processing of Level 2 data products was based on the operational version 2.5 of the SCIAMACHY Level 1b-2 Off-line processor (SGP25). This version contained algorithms for the calculation of the total column density of ozone and NO2 from measurements in the UV and visible spectral region carried in nadir observation geometry, a first algorithm version for the retrieval of ozone and NO_2 profile information from measurements recorded in limb observation geometry, and algorithms for the determination of cloud coverage and the absorbing aerosol index. The determination of total column densities was based on the GOME data processor (GDP) in its version 2.3 and 2.7 which was in between substituted in the operational processing of GOME data by version 3.0 and since the beginning of 2005 by version 4.0. Both versions brought a substantial evolution to the data quality of the GOME total column products. From that, it was decided to transpose the GOME version 4.0 algorithm to the operational processing of SCIAMACHY data and to share the advantage of a similar algorithm baseline for the operational processing of GOME and SCIAMACHY data of total columns. Furthermore, the only cloud product of SGP25, fractional cloud coverage based on OCRA algorithm, had been supplemented by the determination of cloud-top height and cloud optical thickness utilising a sub-set of the SACURA algorithm package developed at the University of Bremen.

The derivation of ozone and NO_2 profile information had undergone a full revision with the result that the originally approach was very limited in its potential for further improvements and evolution. Thus, this approach was substituted by the retrieval package DRACULA which was available as inhouse development and as a retrieval tool box for *non-linear discrete mathematically ill-posed problems* for SCIAMACHY at DLR.

The proceeding addresses first for completeness a brief outline of the operational processing chain. After that, the algorithm update for nadir measurements in the UV/VIS spectral region is presented beginning with an overview followed by the cloud and aerosol products and completing with the determination of total column densities of ozone and NO_2. The following section is then dedicated for the presentation of the limb algorithm update. The proceeding is finalized with an outlook to further developments.

2. BRIEF OUTLINE OF OPERATIONAL PROCESSING CHAIN

In Figure 1 we present a sketch of the full processing chain from level 0 to level 2. The details on processing of level 0 to level 1b are addressed in other parts of the proceedings and not subject to this part. But it should be noted that the operational level 0-1b data processing is carried out in two different modes with the Instrument Processing Facility (IPF) version 6.02: in near-real time mode (NRT) and off-line mode. The main difference between both is the treatment of the calculation of the leakage current which may be a minor source for differences in the determination of column densities in the level 1b to level 2 processing. Operational level 1b to level 2 processing is currently carried out in off-line mode, only. Note that the off-line data processor

Proceedings of the Third Workshop on the Atmospheric Chemistry Validation of Envisat (ACVE-3)
4-7 December 2006, ESRIN, Frascati, Italy (ESA SP-642,February 2007)

(SGP25 and SGP30 as well) is total different software as the IPF which contained in former versions (lower than 6) the NRT level 2 processing.

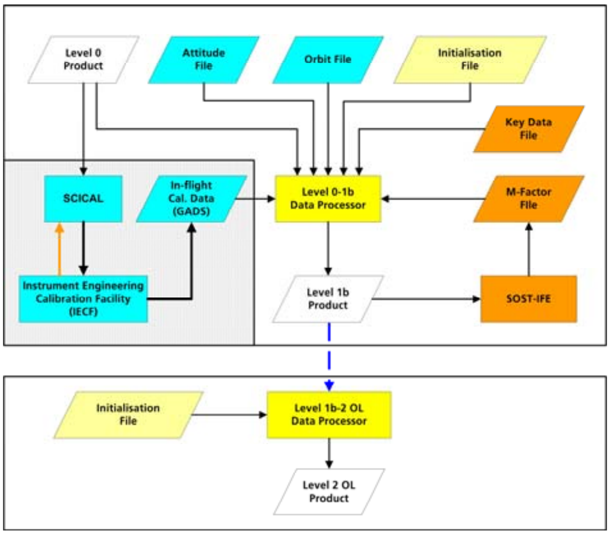

Figure 1. Sketch of the operational level 0 to level 2 processing. Details see text.

Beside the operational processing chain, the data user of level 1b is able to apply the calibration steps individually and optionally with the tool SciaL1c. The results of this tool are calibrated level 1 data which are the called level 1c. Note that in the operational processing the generation and distribution of level 1c products is not foreseen.

3. NADIR UV/VIS ALGORITHM UPDATES

Several products are generated from SCIAMACHY measurement in nadir observation geometry in the UV/VIS spectral region. Those are derived utilising several algorithm steps and different algorithms which can be dependent from each other. Thus, in Figure 2 we show an overview of the algorithm flow in this part of the data processor. The unit “Climatological Pre-processing” combines the processing of all cloud parameters and of the absorbing aerosol index. First, the derivation of cloud coverage from the PMD measurements is initiated utilising the Optical Cloud Recognition Algorithm (OCRA) [1]. The resulting product cloud fraction (CFR) for each spatial pixel based on the shortest integration time of the particular state is provided on one side for the output and additionally as input to the SACURA algorithm [2] for the determination of cloud-top height (CTH) and cloud optical thickness (COT). The result is written into the product, but also provided for the processing of the total column density of ozone and NO_2. For that, cloud-top height is converted to cloud-top pressure (CTP) and cloud optical thickness to cloud-top albedo (CTA). If no result from SACURA is available for the further processing, regress to the ISCCP data base is taken.

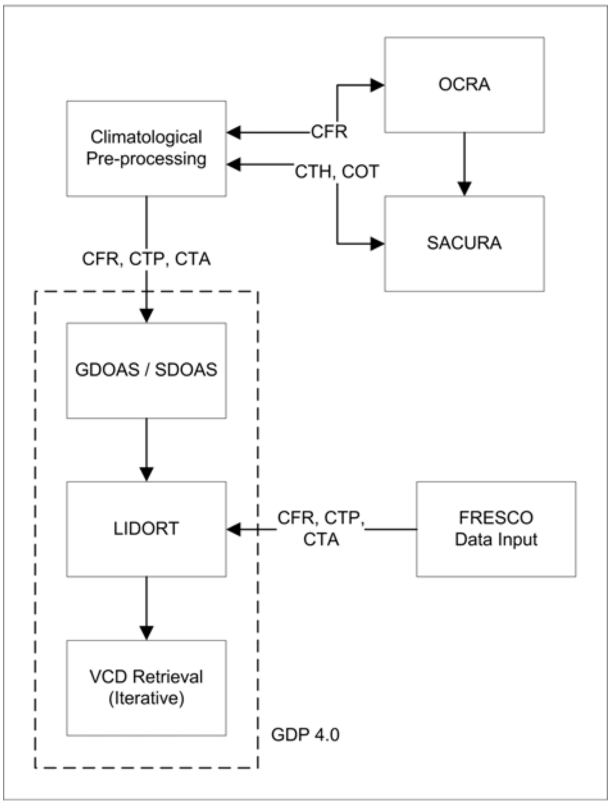

Figure 2. Sketch of the algorithm flow for the nadir UV/VIS algorithms. Details see text.

After the unit “Climatological Pre-processing” a further module is implemented which is mainly based on the GDP 4.0 [3]. This module contains the DOAS implementation as originally developed at the Belgian Institute for Space Aeronomy (BIRA) for GOME (GDOAS) and adapted for SCIAMACHY (SDOAS). The determination of the total column is finally completed by the determination of Air Mass Factors (AMFs) utilizing the Radiative Transfer Model (RTM) LIDORT version 2.2+ and the iterative retrieval scheme for the calculation of vertical column densities. Note that in detail the ozone and NO2 total column derivation differs. For an overview, see below in subsection. A more detailed description can be found in.

Since the original approach of the total column determination at BIRA is based on cloud products generated with the FRESCO algorithm [4], an instance of the module is available for test and verification purposes. This module instance allows the direct input of the appropriate cloud parameters from file to the determination of the total column density.

3.1. Cloud Parameters and Absorbing Aerosol Index

The implementation of OCRA remained un-changed between version 2.5 and 3.0 of the data processor. The cloud-free reflectance database is based on GOME data; scaling factors which are adapting the storage of the cloud-free reflectance in RGB space to coverage scale from 0 to 100%, had been determined for version 2.5 from SCIAMACHY measurements and are also remained un-changed. Since of these two facts, the cloud fraction can be only an effective one. It is foreseen for further improvements, to recalculate the cloud-free reflectance database from SCIAMACHY PMD measurements and, hence, adjust the scaling factors.

The determination of cloud-top height and cloud optical thickness is a new feature in the operational processing. It is based on a reduced version of University of Bremen's SACURA algorithm which is widely documented (see for example in [2]). The cloud parameters are derived from oxygen A-band measurements. The reduction of the originally retrieval scheme is mainly with the forward modelling which is carried out only once per state since of performance reasons. Internal verifications had shown that this approach is sufficient. Since of the physical model on which SACURA is based, the scheme can be only applied to scenes with cloud fraction higher than 20%. This restriction can lead to a high part of spatial pixel for which then no SACURA results are available for the further processing. Thus, an upgrade of the approach is in implementation which allows the inclusion of scenes down to cloud coverage of 5%.

The algorithm for the Absorbing Aerosol Index (AAI) remains also unchanged between version 2.5 and 3.0. A cut-off for the solar zenith angle at 80° is contained since the look-up tables for the Rayleigh databases are limited. The AAIA product is written to the Cloud & Aerosol MDS which provides records at each shortest integration time per state. Note that the AAIA is computed at integration time which is mostly longer than the shortest integration time per state so that not all records are filled with information. Unfortunately, the AAIA is currently hampered by the quality of the reflectance of the level 1b product which means a general restriction to the quality of the product. This will also be subject to further algorithm evolution.

3.2. DOAS Algorithm Implementation

The goal of the evolution of the DOAS algorithm implementation is to achieve a substantial improvement for the product quality of the total column of ozone and NO_2. For that, benefit had been taken from the algorithm developments of GDP. Additionally, the compatibility to GDP was one requirement to achieve improvements. Thus, the GDOAS implementation adapted for SCIAMACHY at BIRA, now called SDOAS, had been built up as reference. It is worth to note that no algorithm change between GDOAS and SDOAS must be introduced to transfer the algorithm. The most intense transfer work laid in the determination of correct control parameters of the algorithm, namely the correct usage of reference spectra and wavelength shifts. In case of NO_2, the processing of 60 orbits even distributed over the year 2003 and compared to results of GOME data processing showed that calibration issues interfered in the slant column derivation. Thus, a post-processing for the slant column density had been introduced which is scientifically justified but will be subject to further investigations. Those investigations may lead to a correction of the post-processing scheme based on a broader data base.

In particular, the derivation of the slant columns for ozone is carried out in the spectral region of 325 to 335 nm. As solar irradiance spectrum the calibrated solar mean reference spectrum measured on the elevation scan mirror (ESM) is in usage. The wavelength calibration of the solar reference is optimized over the fitting interval by taking into account a wavelength shift. This shift had been determined over the fitting interval by means of non-linear least squares on the pre-convolved NEWKPNO atlas. The DOAS algorithm is carried out taking into account the absorption cross-sections of ozone [5] at 223 K and 243 K shifted by + 0.02 nm and scaled by 1.03, the absorption cross-section of NO_2 [5] at 243 K, the Ring effect spectrum convolved of the Kurucz solar atlas with RRS cross-sections of molecular N_2 and O_2, and finally a third-order polynomial. Note that the raw slant column is corrected for the molecular Ring effect as described in [3].

The determination of NO_2 slant columns is performed in the spectral interval of 426.5 to 451.5 nm. The solar irradiance spectrum measured via the azimuth scan mirror (ASM) is taken in un-calibrated mode. As for ozone, the wavelength calibration is extended by a shift for the solar irradiance. The absorption cross-sections of NO_2 [5] at 243 K, of ozone [5] at 243 K, of O_2-O_2 [6] (with a wavelength axis correction by Burkholder), and of water [7] are applied for the determination of the slant column. Furthermore, the Ring effect is taken during the fit as for ozone, and a polynomial of second order. Finally, the intensities are corrected for an off-set. The result for the slant columns are then corrected by adding an off-set of $+1.0x10^{15}$molecule/cm². This has been justified above.

3.3. Total column density derivation

The total column density derivation is driven by two parts, the AMF calculations utilising the RTM LIDORT version 2.2+ and the determination of the vertical column density (VCD). Note that the scheme to calculate the AMFs are as in GDP version 4.0 including the settings (see []). The AMF calculation for ozone is performed at a wavelength of 325.5 nm in combination with the iterative determination of the VCD. In contrast, the AMF of NO_2 is determined at 440 nm but not in the iterative scheme. In both cases, a geometrical weighting of simple parabolic (1-4-1) is applied for the AMF calculation. The atmospheric climatology in use is the TOMS version 8 ozone profile climatology for ozone and the stratospheric NO_2 climatology of Lambert et al. [3] for the NO_2 retrieval. Temperature profiles are taken from TOMS. Note that the tropospheric NO_2 profile climatology is clean for the NO_2 retrieval. The surface database is based on the GOME LER database [8] and ETOP05 [9]. Finally, the cloud information is prepared by radiance weighting.

The cloud parameters in the operational processing are determined with OCRA (cloud fraction) and SACURA (cloud-top pressure and cloud-top albedo). In case, no data from SACURA are available, values from the ISCCP database are taken. Note that the reference system SDOAS of BIRA makes use of the FRESCO algorithm for the determination of the cloud parameters.

3.4. Verification summary

Since the total column determination of the reference system (SDOAS) is based on the cloud parameter derivation with FRESCO, the verification is performed in two steps. First, reference data had been generated with the original SDOAS algorithm. Those had been provided together with the cloud parameters. These cloud data had then been used in the instance of the processing prototype which allows the input of external cloud parameter. With this verification step a proof of the correct implementation of the total column derivation with respect to the reference could be carried out. The second step contained then the inclusion of the cloud parameters in the retrieval. The results are then compared to the results of the first step. Finally, an inter-comparison to GOME data had been also performed, but this is presented in another contribution of these proceedings. Note that the verification test database consists of 60 orbits of 2003 even distributed over all seasons.

We can summarize the results of the first step (SDOAS original versus processor prototype instance in combination with FRESCO cloud data) as follows: One observes for the ozone VCDs a mean deviation of around -0.4% with a standard deviation of 0.35%. 97% of the test data match the 1% level. The differences between processor implementation and reference are easily explained (and verified). The climatological databases are extracted by nearest neighbour in season and geo-location for the reference system. In contrast the extraction of climatological databases in the operational processor is carried out utilising interpolation schemes for seasons and geo-location. In case of NO_2, the mean deviation is also around -0.4% with a standard deviation of 0.25%. Here, we can observe that most of the data match the 1% level.
The second verification step, i.e. the inter-comparison of the same retrieval scheme but based on the different cloud parameter input, can be summarized to: the mean deviation of ozone VCDs is around +0.4% with a standard deviation of 1.4%. In case of NO_2 VCDs, a mean deviation of around +0.02% with a standard deviation of around 0.25% is observed. It is not surprising that the deviations in case of NO_2 are smaller since the VCD retrieval is based on an atmospheric climatology with a clean troposphere.

In order to avoid a misinterpretation of the product results, we like to point out that the slant column density results are provided in the product without the Ring correction. Additionally, the operational level 2 product is based on the level 1b product which is computed in off-line mode. In general, the difference for ozone total columns is less than 0.01% but one can observe some cases where the difference is around 1% and higher. In case of NO_2 the difference between usage level 1b NRT and level 1b offline product is more pronounced and can vary around 1-4%.

4. PROFILE RETRIEVAL FROM LIMB MEASUREMENTS

In order to improve the performances of the ozone and NO2 profile retrieval in the UV/VIS spectral region, the former existing algorithm (version 2.5) has been replaced by a dedicated inversion tool including the retrieval package DRACULA (aDvanced Retrieval of the Atmosphere with Constrained and Unconstrained Least Squares Algorithms). DRACULA has been designed at DLR for solving in general inverse problems in atmospheric remote sensing. It has been utilized to a number of scientific applications, in particular to MIPAS and SCAMACHY data. In the following we introduce DRACULA before the details of the operational retrieval settings are presented. Finally, the implementation and verification baseline is outlined.

4.1. Retrieval tool box DRACULA

DARCULA is the only retrieval tool box for solving of *non-linear discrete mathematically ill-posed problems.* Its unique interface allows a versatile usage and opens the opportunity to apply the tool box for a large class of applications.

Specifically, DRACULA includes Tikhonov Regularization type methods and Iterative Regularization type methods. In the first category, the conventional formulation (with a priori, a posteriori, and error-free parameter choice methods) together with its iterated version is implemented. The computation of the regularized solution in the framework of Tikhonov

Regularization type methods is performed by using the Generalized SVD or iterative solvers as for instance CG (Conjugate Gradient), BICG (Bi-Conjugate Gradient), GMRS (General Minimal Residual Solution). As iterative methods we mention following methods: Iterative Regularized Gauss-Newton Method, Inexact iterative Newton Algorithms (REGINN), Regularized Levenberg-Marquardt Method, and Regularized Trust-Region Method. The advantage of iterative methods over standard approaches is the computational performance and the fact that no parameter choice method is required.

With DRACULA, we are able to cross-check with several independent methods the pre-optimization of the algorithm control parameters. For specific applications, we can select the optimal algorithm combination with respect to accuracy and computational performance.

4.2. Retrieval settings

The retrieval performance depends on the identification and optimization of several algorithm control parameters. They can be categorized into physical and mathematical (method) control parameters.

The physical control parameters contain the selection of a priori state vector, here the climatology of McLinden et al. provided by the University of Bremen [10]; the choice of the measurement noise variance (covariance matrix); the retrieval grid size; the tangent height pointing; and the surface albedo. The retrieval grid size had been set to a value of 3.3 km due to the equivalence of the measurement grid. The tangent height pointing is provided in the level 1b product. No additional correction is applied in the level 1c-2 processing. But it shall be noted that the tangent height pointing is a point of intensive discussions and investigations. According a recommendation of the University of Bremen [10], the surface albedo had been set to a value of 0.3.

The mathematical control parameters include the type of regularization matrix and the strength of regularization. In the case of iterative methods, the termination criterion is decisive to obtain a stable solution. The used regularization matrix is an exponential matrix of Markov random process. The correlation length is equivalent to the retrieval grid size. The strength of regularization is chosen automatically by using appropriate parameter choice method as for instance L-Curve, GCV (Generalized Cross-Validation), quasi-optimal criterion or Maximum-Likelihood-Estimation. The stopping rule for iterative methods is the discrepancy principle whereas the noise level is determined by analysing the residual of un-constrained least squares methods.

The retrieval is also influenced by the forward model control parameters. The treatment of the multiple scattering term in the forward model can be performed with pre-calculated look-up tables or by using a two-dimensional RTM based on Picard iteration. Since of computational resources, the first had been adopted for the operational processing. Additionally, the choice of the objective function may influence the accuracy of the retrieval. Here, the objective function is modelled as in DOAS-type applications. Hence, the inverse problem is nearly linear and the computational performance is substantially enhanced.

4.3. Implementation and verification baseline

DRACULA had been interfaced to the operational level 1b-2 off-line processor in version 3.0. In this implementation, the retrieval is carried out for column densities per retrieval layer (partial columns). In close cooperation with the University of Bremen, the control parameters were determined. Ozone profiles are retrieved in the Chappuis band; NO_2 profile in the visible spectral region. The verification had been performed on hand a selected set of orbits out of 2005 by comparison with profile retrievals carried out at the University of Bremen. The verification had been finalized with the selection of the regularization method (Tikhonov Regularization). The selection took into account that with the method averaging kernels can be provided to the users. The implementation had been completed by the extension of the field “Additional diagnostics” in the limb MDS (Measurement Data Set) of the product model by the a priori profile, the averaging kernel, the profile results in number density, the conversion factors between number density, partial column and VMR representation to allow an easier access to those data.

Unfortunately, the product content is incorrect for the field “Measurement grid”, since there are the entries temperature and pressure inter-changed.

5. OUTLOOK

The update of the operational Level 1b-2 off-line data processor introduced a substantial improvement for all total column and profile products. However, the cloud and aerosol products are subject to further evolutions in the near future. Further evolutions to the total column and profile products are planned and envisaged in the context of the SCIAMACHY Quality Working Group. There are also intentions to bring new products in the pipeline, especially a first operational product from SCIAMACHY infrared measurements.

6. REFERENCES

1. D. Loyola and T. Ruppert, A new PMD cloud recognition algorithm for GOME, *Earth Observation Quarterly*, 58:45-47, 1998.
2. M. Vountas, SCIAMACHY Level 1b-2 OL Processing: Algorithm Theoretical Baseline Document: Semi-Analytical Cloud Retrieval Algorithm for SCIAMACHY, *Technical report, ENV-ATB-IFE-SCIA-0003*, Issue 1.0, 16 May 2006; and references herein.
3. M. Van Roozendael, D. Loyola, R. Spurr, D. Balis, J.-C-. Lambert, Y. Livshitz, P. Valks, T. Ruppert, P. Kenter, C. Fayt, C. Zehner, Ten years of GOME/ERS-2 total ozone data – The new GOME data processor (GDP) version 4: 1. Algorithm description, Journal of Geophysical Research, 11:D14311, 2006
4. R.B.A. Koolemeijer, P. Stammes, J.W. Hovenier, and J.F. de Haan, A fast method for retrieval of cloud parameters using oxygen A band measurements from Global Ozone Monitoring Experiment, *Journal of Geophysical Research*, 106:3475-3490
5. K. Bogumil, J. Orphal, T. Homann, S. Voigt, P.Spietz, O.C. Fleischmann, A. Vogel, M. Hartmann, H. Bovensmann, J. Frerick and J.P. Burrows, *Measurements of Molecular Absorption Spectra with the SCIAMACHY Pre-Flight Model: Instrument Characterization and Reference Data for Atmospheric Remote-Sensing in the 230-2380 nm Region*, J. Photochem. Photobiol. A., 157:167-184, 2003
6. G.D. Greenblatt, J.J. Orlando, J.B.Burkholder, and A.R.Ravishankara Absorption measurements of oxygen between 330 and 1140 nm, *J. Geophys. Res.*, 95:18 577–18 582, 1990
7. L. Rothman et al., The HITRAN molecular spectroscopic database: Edition of 2000 including updates throughout 2001, *J. Quant. Spectrosc. Radiat. Transfer*, 82:5-44, 2003
8. R.B.A. Koelemeijer, J.F. de Haan, J.W. Hovenier, and P.Stammes, A database of spectral surface reflectivity in the range of 335-772 nm derived from 5.5 years GOME observations, *J. Geophys. Res.*, 108(D2):4070, 2003
9. J.R. Herman and E.A. Celarier, Earth surface reflectivity climatology at 340-380 nm from TOMS data, *J. Geophys. Res.*, 102:28 003-28 011, 1997
10. C. von Savigny, pers. Communication, 2005