DLR

Deutsches Zentrum
für Luft und Raumfahrt e.V.

GOME

Level 1 to 2 Algorithms Description

TECHNICAL NOTE

ER–TN–DLR–GO–0025

Iss./Rev. 3/A

July 31, 2002

Deutsches Zentrum für Luft und Raumfahrt e.V. – DLR

Institut für Methodik der Fernerkundung – IMF

Oberpfaffenhofen, Germany

prepared by R. Spurr, DLR (now SAO), Date

W. Thomas, DLR–IMF, Date

D. Loyola, DLR–IMF, Date

Distribution list

Available to the public via Internet

Document Change Log

Iss./Rev.	Date	Section	Description of Change	approved by	Sign
A	16.03.89	all	Completely new		
1/A	30.09.94	all	Completely new, based on individual algorithm technical notes		
2/A	09.08.96	all	Update of all algorithms reflecting the 1st operational version GDP L12 2.0		
2/B	18.12.00	all	Update of all algorithms reflecting the operational version GDP L12 2.7		
3/A	31.07.02	all	Update of all algorithms reflecting the operational version GDP L12 3.0		

Table of Contents

0 Document Overview and History

The first issue of this document (Iss./Rev. 1/A) was released in August 1994, some 8 months before the launch of the ERS–2 satellite in April 1995. Following the instrument commissioning phase, the data processing validation campaign took place in the period October 1995 to March 1996, followed by a three–month period of analysis and trouble–shooting. In July 1996 the GDP 1–2 system reached a plateau, with the public release of level 2 products planned for end July. This was a good time to take stock, and at the final meeting of the ”Tiger Team” validation group at DLR on June 24/25 1996, a high priority recommendation was made to update GDP documentation. One of the main results was version 2/A of the present document, an update of the initial Algorithms Description Technical Note.The basic algorithm descriptions in the original issue 1/A are still relevant.

0.1 Version 2/A – GDP 2.0

First released version of the current document. In order to avoid excessive re–writes of large amounts of text, it has been decided to add an additional section ***at the end of the main chapters to include summaries of changes and improvements*** that have been added to the algorithms since the first issue release. Many of the items listed in the ’Open Issues’ sections have now been addressed, and points mentioned under this heading are discussed in the Update sections.

0.2 Version 2/B – GDP 2.7

The first minor update of GDP was in 1998 and changes from version 2.0 to 2.4 are summarized in version 2/B of this document. The next GDP version has been released in summer 1999 (GDP 2.7) and all GOME data have been reprocessed in fall 1999 using GDP 2.7. Changes from version 2.4 to 2.7 are also summarized in version 2/B of this document.

The introduction (Chapter 1) contains a new section summarizing the algorithm changes and improvements introduced in the interim periods, and described in following chapters. Chapters 2–5 (pre–processing, ICFA, DOAS and AMF algorithms respectively) each contain new description update sections. There have been few changes to the VCD algorithm (Chapter 6), but section 6.6 on quality control has been re–written.

0.3 Version 3/A – GDP 3.0

The next operational version of GDP will become operational in Summer 2002. All GOME level 2 data will be reprocessed in 2002. Changes from version 2.7 to 3.0 are also summarized in version 3/A of this document.

1 Introduction

1.1 Purpose and Scope

The German Remote Sensing Data Center (DFD) plays a major role in the design, implementation and operation of the GOME Data Processor (GDP). GDP is the off–line and NRT ground segment for the GOME instrument which flies on ERS–2. It incorporates a Level 0 to 1 processing chain, the complete GOME data archive, the DOAS trace gas total column retrieval process (Level 1 to 2 processing), and an image processing chain for the generation of higher level products.

Level 1 to 2 processing is concerned with the retrieval of column amounts from the calibrated geolocated radiances derived from the Level 0 to 1 processing. The operational Level 1 to 2 algorithms employs the DOAS spectral fitting of trace gas column amounts, and the emphasis is on the generation of a *total vertical ozone and nitrogen dioxide columns* for the Level 2 product.

Level 1 to 2 algorithms contain the bulk of the geophysical science required for the retrieval of atmospheric constituent quantities. The DOAS retrieval method has been successfully employed in many measurement applications, from ground–based spectrometers to ship, balloon and air–borne instruments. GOME represents the first application of this technique to measurements obtained from a passive remote sensing instrument.

It is a noteworthy feature of this retrieval technique that the spectral least–squares fitting for the *Effective Slant Column* (ESC) amounts is completely separated from the associated radiative transfer calculation of the *Air Mass Factor* (AMF) – the division of the slant column amount by the AMF yields the required *Vertical Column Density* (VCD). These two main algorithms are discussed in detail in chapters 3 and 4 respectively. The Initial Cloud Fitting Algorithm (ICFA) is supplementary to the main DOAS and AMF algorithms, and provides essential information on cloud parameters for totally and partially cloudy scenes (chapter 2). The computation of vertical columns is summarised in chapter 5, which also includes remarks on Level 2 product content and quality control.

In addition to the calibrated spectra and geolocation information extracted from the Level 1 Data Product, a large amount of auxiliary information (climatological databases) is required, especially for the radiative transfer (AMF) calculations. These Level 1 to 2 data bases have been described in an accompanying document [A3]; where appropriate, the data base requirements will be discussed in the algorithm descriptions given here. For convenience, summaries of the required input measurement data sets and the climatological data bases are given in chapter 6.

Requirements for the Level 1 to 2 Algorithms were first laid down in document [R2], which contains the recommendations of the GOME Data and Algorithm Subcommittee of GSAG. The development of algorithms and their subsequent software implementation represents an enormous amount of work, both from the scientists and the software engineers at DLR. The present technical note updates and collates three technical documents (documents [A5], [A6] and [A7]) produced informally from the scientific side; the intention is to provide a more formal mathematical description of the level 1–2 algorithms. The present document is supplementary to the main ADD for the GDP [A4]; it will not contain discussion of the software implementation and system engineering involved in Level 1–2 processing.

The authors would like to thank K. Chance, J. Burrows, H. Frank, E. Mikusch, D. Diebel, U. Platt, V. Rozanov, W. Balzer, P. Stammes and T. Kurosu for their contributions.

1.2 Definitions, Abbreviations and Acronyms

A list of all abbreviations and acronyms which are used throughout the SRDs for the GDP is given below:

BBM	Bread Board Model
BSDF	Bi–directional Scattering Distribution Function
CU	Calibration Unit
DFD	Deutsches Fernerkundungsdatenzentrum
DLR	Deutsche Forschungsanstalt für Luft– und Raumfahrt e.V.
DOAS	Differential Optical Absorption Spectroscopy
ERS	European Remote Sensing Satellite
ESA	European Space Agency
ESOC	European Space Operation Centre
ESTEC	European Space Centre of Technology
FM	Flight Model
FPA	Focal Plane Assembly
FPN	Fixed Pattern Noise
FSM	Flight Spare Model
GDP	GOME Data Processor
GOME	Global Ozone Monitoring Experiment
IMF	Institut für Methodik der Fernerkundung
HK	House Keeping (Data)
LED	Light Emitting Diode
MMCC	Mission Management and Control Center
PMD	Polarisation Measurement Device
PPG	Pixel–to–Pixel Gain
SRD	Software Requirements Document
TPD/TNO	Technisch Physische Dienst
UV	Ultra–Violet (spectral range)
VIS	VISible (spectral range)

Other abbreviations used in the present technical note are :

AAIA	Absorbing Aerosol Indicator Algorithm
AMF	Air Mass Factor
AOT	Aerosol Optical Thickness
AZM	Azimuth
BISA	Belgian Institute of Space Aeronomy
ESC	Effective Slant Column
FOV	Field of View
ICFA	Initial Cloud Fitting Algorithm
IUP	Institut für Umweltphysik der Universität Bremen
LOS	Line–Of–Sight zenith angle
LUT	Look–Up Table
PBL	Planetary Boundary Layer
PMD	Polarisation Measurement Device
RTM	Radiative Transfer Model
SZA	Sun Zenith Angle
TOA	Top Of Atmosphere
TOMS	Total Ozone Monitoring Spectrometer
VCD	Vertical Column Density

1.3 Documents

1.3.1 Reference Documents

The following reference documents are relevant:

[R1] ”The Interim GOME Science Report”, September 1993

[R2] ”Report of the GOME Scientific Advisory Group, Data and Algorithm Subcommittee”, February 1992

[R3] ”Report of the GOME Scientific Advisory Group, Calibration and Characterization Subcommittee, Scientific Requirements for the Calibration and Characterization of the Global Ozone Monitoring Experiment”, November 1992

1.3.2 Applicable Documents

The following documents are directly applicable to the present work: in particular note the three technical documents [A5], [A6] and [A7] and the GDP update document [A13] reflecting in detail all the changes that have been made in GDP version 2.7.

[A1] System Requirements Document of the GOME Data Processor, ER–SR–DLR–GO–0020, Issue 1, July 1993

[A2] Functional Software Requirements of the GOME Data Processor (Level 2), ER–SR–DLR–GO–0009, Issue 1, July 1993

[A3] GOME Data Bases (Level 1 to 2 Processing), ER–TN–IFE–GO–0018, Iss./Rev. 3/A, July 2002

[A4] Architectural Design Document of the GOME Data Processor (level 2), ER–AD–DLR–GO–0012, Iss./Rev. 1/A, April 1994

[A5] Internal Technical Note, ’DOAS Slant Column Retrieval in GOME Data Processing’, R. Spurr, University of Bremen, January 1994

[A6] Internal Technical Note, ’Use of Air Mass Factors in GOME Software’, R. Spurr and H. Frank, University of Bremen, University of Heidelberg, January 1993

[A7] Internal Technical Note, ’Initial Cloud Fitting Algorithm for GOME’, K. Chance, Smithsonian Astrophysical Observatory, Cambridge, Mass., December 1993

[A8] A Study of Methods for Retrieval of Atmospheric Constituents, Final Report, ESA/SERCO, December 1993

[A9] Numerical Recipes, William H. Press et al, Cambridge University Press, Second Edition (1992)

[A10] Interface Specification Document of the GDP, ER–IS–DLR–GO–0004, Iss./Rev. 2/C, December 2000.

[A11] Product Specification Document of the GDP, ER–PS–DLR–GO–0016, Iss./Rev. 3/D, May 2000.

[A12] Internal Technical Note, ’A tessellation algorithm for GOME and SCIAMACHY’, R. Spurr, Harvard Smithsonian Astrophysical Observatory, SAO, November 1998.

[A13] GOME Data Processor, Update Report for GDP 0–to–1 Version 2.0 and GDP 1–to–2 Version 2.7, ER–TN–DLR–GO–0043, Iss./Rev. 1/A, August 1999

[A14] ERS–2 GOME Data Products Delta Characterisation Report 1999, Lambert J.–C. and P. Skarlas, IASB, Brussels, Issue 0.1, November 1999

[A15] Brasseur G. and S. Solomon, Aeronomy of the middle atmosphere, second edition, D. Reidel Publishing Company, Dordrecht, Holland, 1986.

[A16] Burrows J.P., A. Dehn, B. Deters, S. Himmelmann,A. Richter, S. Voigt and J. Orphal, Atmospheric Remote–Sensing Reference Data from GOME: Part 1. Temperature–dependent Absorption cross–sections of NO_2 in the 231–794nm range, J. Quant. Spec. Radiat. Trans., 60, 1025–1031, 1998.

[A17] Burrows J.P., A. Richter, A. Dehn, B. Deters, S. Himmelmann, S. Voigt and J. Orphal, Atmospheric Remote–Sensing Reference Data from GOME: Part 2. Temperature–dependent Absorption cross–sections of O_3 in the 231–794nm range, J. Quant. Spec. Radiat. Trans., 61, 509–517, 1999.

[A18] Cantrell C.A., J.A. Davidson, A.H. McDaniel, R.E. Shetter, and J.G. Calvert, Temperature–dependent formaldehyde cross sections in the near–ultraviolet spectral region, J. Phys. Chem., 94, 3902–3908, 1990.

[A19] Chance K.V. and R.J.D. Spurr, Ring Effect Studies, Appl. Opt., 36, 5224–5230, 1997.

[A20] Chance K.V., Analysis of BrO Measurements from the Global Ozone Monitoring Experiment, Geophys. Res. Lett., 25, 3335–3338, 1998.

[A21] Eisinger M., personal communication, ESA–ESTEC, 2000.

[A22] Fortuin J.P.F. and H. Kelder, An ozone climatology based on ozonesonde and satellite measurements, J. Geophys. Res., 103, 31709–31734, 1998.

[A23] Harwood M.H. and R.L. Jones, Temperature dependent ultraviolet–visible absorption cross sections of NO_2 and N_2O_4: Low temperature measurements of the equilibrium constant for 2 NO_2 <— N_2O_4, J. Geophys. Res., 99, 22955–22964, 1994.

[A24] Koelemeijer R.B.A. and P. Stammes, Effects of clouds on the ozone column retrieval from GOME UV measurements, J. Geophys. Res., 104, 8281–8294, 1999.

[A25] Loyola D., Using Artificial Neural Networks for the Calculation of Air Mass Factors, ESAMS'99 – European Symposium on Atmospheric Measurements from Space, 1999.

[A26] Loyola D., Combining Artificial Neural Networks for Parameterization of Radiative Transfer Models, IEEE International Geoscience and Remote Sensing Symposium, IGARSS'2000, 2000.

[A27] McPeters R.D., P.K. Bhartia, A.J. Krueger, J.R. Herman, B.M. Schlesinger, C.G. Wellemeyer, C.J. Seftor, G. Jaross, S.L. Taylor, T. Swissler, O. Torres, G. Labow, W. Byerly and R.P. Cebula, Nimbus–7 Total Ozone Mapping Spectrometer (TOMS) Data Products User's Guide, NASA Reference Publication, 1996.

[A28] Platt U., D. Perner and H.W. Paetz, Simultaneous Measurement of atmospheric CH_2O, O_3, and NO_2 by differential optical absorption, J. Geophys. Res., 84, 6329–6335, 1979.

[A29] Richter A., personal communication, IUP, 2000.

[A30] van Roozendaal M., Technical Note: Ring effect study: Test of available data sets for DOAS fitting of GOME spectra in the O_3 and BrO intervals, BISA, May 2000.

[A31] Rothman L.S., R.R. Gamache, A. Goldman, L.R. Brown, R.A. Toth, H.M. Pickett, R.L. Poynter, J.–M. Flaud, C. Camy–Peret, A. Barbe, N. Husson, C.P. Rinsland and M.A.H. Smith, The HITRAN Database; 1986 Edition, Appl. Opt., 26, 4058–4097, 1987.

[A32] Rozanov V., D. Diebel, R.J.D. Spurr and J.P.Burrows, GOMETRAN : Radiative Transfer Model for the Satellite Project GOME, the Plane–Parallel Version, J. Geophys. Res., 102, 16683–16695, 1997.

[A33] Schneider W., G.K. Moortgat, G.S. Tyndall and J.P. Burrows, Absorption cross sections of NO_2 in the UV and visible region (200–700nm) at 298K, J. Photochem. Photobiol. A: Chem., 40, 195–217, 1987.

[A34] Slijkhuis S., A. v. Bargen, W. Thomas and K.V. Chance, Calculation of Undersampling correction spectra for DOAS spectral fitting, ESAMS'99 – European Symposium on Atmospheric Measurements from Space, ESA WPP–161, Noordwijk, The Netherlands, 1999.

[A35] Solomon S., A.L. Schmeltekopf, and R.W. Sanders, On the Interpretation of Zenith Sky Absorption Measurements, J. Geophys. Res., 92, 8311–8319, 1987.

[A36] Spurr R.J.D, T.P. Kurosu, and K.V. Chance, A Linearized Discrete Ordinate Radiative Transfer Model for Atmospheric Remote Sensing Retrieval, J. Quant. Spec. Radiat. Trans., 68, 689–735, 2001.

[A37] Spurr R.J.D, Improved climatologies and new air mass factor look–up tables for O_3 and NO_2 column retrievals from GOME and SCIAMACHY backscatter measurements, ESAMS'99 – European Symposium on Atmospheric Measurements from Space, Noordwijk, The Netherlands, ESA WPP–161, 277–284, 1999.

[A38] Stephens, G.L., Optical Properties of Eight Water Cloud Types; Commonwealth Scientific and Industrial Research Organisation, Division of Atmospheric Physics, Technical Paper No. 36, 35pp, 1979.

[A39] Voigt S., J. Orphal, K. Bogumil, and. J.P. Burrows, The temperature dependence ($203^{-}293$ K) of the absorption cross sections of O_3 in the 230–850 nm region measured by Fourier–transform spectroscopy, J. Photochem. Photobiol. A: Chem., 143, 1–9, 2001.

[A40] Vountas M., V.V. Rozanov, J.P. Burrows, Ring Effect: Impact of Rotational Raman Scattering on Radiative transfer in Earth's Atmosphere, J. Quant. Spec. Radiat. Trans., 60, 943–961, 1998.

[A41] Wilmouth D., T.F. Hanisco, N.M. Donahue, and J.G. Anderson, Fourier Transform Ultraviolet Spectroscopy of the $A^2\,\Pi_{3/2} \leftarrow X^2\,\Pi_{3/2}$ Transition of BrO, J. Phys. Chem. A, 103, 8935–8945, 1999.

1.4 Overview

The present technical note describes the following five algorithms (chapters 2 to 6) :

- pre–processing algorithms
  For the generation of AAIA values, snow/ice flags and the tessellation results. Note, that the first two mentioned algorithms are not used operationally up to GDP 2.7, respectively.
- ICFA algorithm
  For the generation of cloud–top parameters.
- DOAS algorithm
  For the spectral least squares fitting of optical density to generate effective slant columns and diagnostics.
- AMF algorithm
  For the generation of trace gas air mass factors and simulated intensities.
- VCD algorithm
  Vertical column computation based on output from the previous three algorithms.

The last Chapter summarises the required input data sets for the successful execution of the complete sequence of Level 1–2 algorithms, and lists the appropriate climatological and auxiliary data bases required for the execution of these algorithms.

1.5 Summary of Algorithm Updates

1.5.1 Issue 2/A – GDP 2.0

The most significant changes from GDP 1.6 to GDP 2.0 and additions to the GDP algorithms are summarized as follows :

- ICFA algorithm
  - Inclusion of the Cloud Clearing Algorithm. The results of this algorithm are written to the L2 product, but are not used in the AMF and VCD calculations
  - Cloud–top pressure is now taken from the ISCCP data base
  - Cloud–top reflectance now contains the effect of surface albedo in the computation of the transmission loss term (escape function correction)
  - Extension of transmission templates to solar zeniths up to 90 degrees, including the effect of spherical geometry
- DOAS algorithm
  - Inclusion and use of GOME FM 1996 O_3 and NO_2 cross sections
  - Inclusion and use of GOME FM 1996 differential Ring cross–section. GOME solar spectrum is now excluded from the linear part of the DOAS fit.
- AMF algorithm
  - Generation and implementation of a look–up table of multiple scattering correction factors, for solar zeniths up to 92 degrees.
  - Operational calculation of a single scatter AMF from first principles, multipled by a multiple scatter correction extracted from the look–up table.

- VCD algorithm
  - Implementation of the individual quality check for potential ozone hole situations mentioned in section 6.6.
- ERS–2 propagator
  - One important change was made in the derivation of geolocation information (applies also to Level 1 data). All geometrical angles, surface positions, satellite heights and earth radii are now calculated with the aid of the ESA propagator for ERS–2; this was implemented in March 1996.

1.5.2 Issue 2/B – GDP 2.7

For a detailed description of changes from GDP 2.4 to GDP 2.7 see mainly [A 14].

- pre–processing algorithms
  - Tessellation algorithm : A new algorithm has been implemented (see [A12]) to account for surface inhomogeneities (albedo, ground height) in the GOME footprint. Large changes in the height above sea level and the ground albedo occur along the coast sides and for mountainous terrain. Up to GDP 2.4 the selection of valid values from the data bases was carried out using “nearest neighbour” of data base entries and centre co–ordinates of ground pixels as selection criterion. The new algorithm calculates an area–weighted quantity which is derived from all data base entries that falls into the GOME footprint. The algorithm works only for ground pixels with an integration time of 1.5s.
  - Absorbing Aerosol Indicator Algorithm : A new algorithm has been implemented to calculate an indicator for the presence of absorbing aerosol in the atmosphere. However, the results of the algorithm are not validated and a decision has been made not to release the results until the required validation has been carried out. Even more, the public availability of AAIA results requires a product change of GOME Level 2 products. The algorithm is currently switched off in the operational retrieval.
  - Snow/ice recognition algorithm : Two simple modules have been developed as a prerequisite of a first version of a snow/ice recognition algorithm for GOME. The sun–normalized (cosine–weighted) PMD reflectance is calculated for each PMD sub–pixel and three PMD channels. The second module performs a similar operation on the regular GOME backscatter measurements and calculates a sun–normalized (cosine–weighted) earth–shine reflectance. Both modules are currently not used in the operational retrieval.
- ICFA algorithm
  - ICFA slit function type changed from “rectangular” to “simple hyberbolic”.
- DOAS algorithm
  - Calculation of the so–called Bass–Paur temperature for ozone cross–sections as function of the maximum number density of ozone (was based on maximum volume mixing ratio before).
  - Update of data bases for O_3 and NO_2 cross–sections (flight model measurements – FM 1998) on most recent revision provided by IFE Bremen (see [A 3]). The updated NO_2 cross–sections are used operationally.

- Parallel usage of two ozone spectra at different temperatures implemented for minor trace gas fitting (BrO, HCHO). Not implemented for ozone retrieval.
- Inclusion of undersampling spectra for the O_3, NO_2, BrO and CH_2O standard fitting windows to improve the DOAS fitting. Operational use of undersampling correction spectrum for NO_2 retrieval in the visible window.
- Inclusion of theoretical Ring spectrum provided by SAO ([A19]).
- Implementation of a check module for valid wavelength calibration of both sun spectrum and backscatter spectrum in the DOAS module chain. For each channel the first wavelength entry is compared against a fixed wavelength and an error is issued if this wavelength does not fall within a spectral intervall of ± 0.16 nm around the fixed wavelength.
- Error–weigthed fitting now used operationally
- Gauss–Jordan elimination replaced by LU decomposition as matrix solver

- AMF algorithm
  - Generation and implementation of a look–up table of multiple scattering correction factors, for solar zenith angles up to 92 degrees including the polar view mode with enhanced LOS angles. Furthermore, Mie phase functions have been used for aerosol scattering.
  - The reference grids for the AMF multiple scattering correction factors have been changed, in order to account for the higher variability of natural conditions.
  - The theoretical top of atmosphere (TOA) has been changed from 60 km to 70 km. This was done also for the correction factors stored in the Look–up table.
  - Removal of the erroneous three months shift in the retrieval of trace gas profiles on the Southern hemisphere from the MPI data base.
  - Implementation of a combined linear time/latitude interpolation scheme for selected temperature/pressure/concentration profiles from climatologies. The profiles can be selected from different climatologies.
  - The value for the constant snow albedo has been set to 0.95 (was 0.75 formerly) for land surfaces.

1.5.3 Issue 3/A – GDP 3.0

- General
  - The product philosophy has changed, in order to provide maximum information to the user community. Thus, a failure in e.g., the ozone retrieval will not suppress the following NO_2 retrieval. Instead, the ozone entries will be filled with zeros but all other reliable retrieval results will be written to the product.
- ERS–2 propagator
  - The calculation of footprint coordinates (4 corners, centre position) for “static–view” pixels (static scan mirror) were calculated erroneously. The inter-

polation scheme in place assumed subsequent ground pixels in across–track direction, while subsequent static–view pixels follow in along–track direction. The interpolation scheme is now able to handle also static–view pixels correctly.

- The minimum line–of–sight angle has changed from 0.1 degree to 0.001 which is in line with the value used in GDP L01 processing.

- pre–processing algorithms
  - As a consequence of recent problems with improperly calibrated spectra (mainly earth–shine spectra) the earth–shine wavelength grid is not used anymore; instead, the sun wavelength grid is assigned to backscatter measurements on a pixel–to–pixel basis.
  - The derivation of cloud–top reflectances from data base entries has been corrected for negative azimuth differences (between the Sun and the satellite's position) which may occur as valid input for the radiative transfer model in place.
  - A trace gas profile climatology with enhanced tropospheric loading of O_3, NO_2, SO_2 and HCHO has been implemented. It is partially based on scenarios defined in [A8]. This data base is not used operationally.
  - A bi–modal undersampling correction spectrum for BrO fitting ([A20]) has been implemented. This spectrum is not used operationally.
- ICFA
  - A new flag is implemented that indicates a failure of the cloud fitting routine. If the normalized cloud coverage is greater than 1. or less than 0., the corresponding flag is set and a warning message is generated. This flag is written to the ICFA flag array and is part of the GOME level 2 product.
  - The calibration check module mentioned in section 1.5.2 is now called also in the algorithm chain of ICFA. If the check fails, a warning message is generated. Another flag is implemented that indicates the failure and this flag is again written to the GOME level 2 product.
- DOAS
  - NO_2 at 241K ([A15]) has been added as interfering species in the ozone fitting window in the UV.
  - An undersampling correction spectrum ([A34]) has been added to the ozone fitting window in the UV.
  - The updated GOME FM ozone cross–sections ([A15], [A17]) are now used in the standard UV fitting window.
  - The GOME FM Ring spectrum has been superseded by a bi–modal theoretical Ring spectrum provided by SAO ([A19]) in the UV fitting window. Only it's first component is applied to O_3 fitting. The spectrum was provided by [A30].
  - A recent Ring spectrum provided by SAO in 1997 that was used for NO_2 fitting has been superseded by the above–mentioned bi–modal theoretical Ring

spectrum. Only it's first component is applied to NO_2 fitting. The spectrum was provided by [A30].

- All shifts/squeeze operations for reference spectra are now switched off for ozone and nitrogen dioxide retrieval, except for the static undersampling correction spectra that were derived following [A34].
- A "warm" ozone spectrum and an ozone difference spectrum calculated from the difference of ozone cross–sections at different temperatures are now fitted simultaneously in the UV fitting window ([A29]). This option is used operationally.
- The fitted ozone temperature could be used as a diagnostic variable but is not part of the GOME level 2 product.
- Extraction of reference spectra is now done only once per processing order.
- The theoretical Doppler–shift of the sun spectrum can now be calculated for the centre wavelength of each fitting window and can be used to limit shift/squeeze operations of the sun spectrum. This option is not used in the operational context.
- It is recommended to apply a pre–shift of +0.02nm to measured GOME–FM reference spectra (O_3, NO_2) in Ch2 ([A30]). However, lowest fit errors (residuals) were found for a pre–shift of +0.017nm. Therefore, all GOME–FM spectra (O_3, NO_2) in Ch2 are pre–shifted by about 0.017 nm towards longer wavelengths for ozone fitting in the UV. The pre–shift is set in the static initialisation file.

- AMF
  - The TOMS V7 ozone profile climatology ([A27]) has been implemented and can be used for off–line AMF computation. Interfaces for LIDORT ([A35]) and GOMETRAN ([A32]) have been generated. The climatology is not used for the on–line calculation of single scattering AMFs.
  - Another ozone profile climatology published by [A22] has been implemented and can be used for off–line AMF computation. Interfaces for LIDORT ([A35]) and GOMETRAN ([A32]) have been generated. It is used operationally for the on–line calculation of single scattering AMFs for NO_2 in the NO_2 fitting window at 437.5 nm.
  - A LUT of AMFs for ozone at 325 nm has been generated using LIDORT ([A35]). It is based on TOMS V7 ozone profiles, i.e. total column content and latitude; other variables are ground/cloud–top albedo, ground/cloud–top height, land/sea mask (i.e., aerosol type), and viewing geometry (SZA, LOS, rel. AZM). The LUT is not part of the operational data bases.
  - A neural network approach was implemented to calculate AMFs for ozone at 325 nm, as function of the above–mentioned variables. The LUT was used as training data set for the neural network.
  - A new formula for calculating the Rayleigh scattering coefficient as suggested by ([A19]) replaces the formula from Brasseur and Solomon. This formula is used for both pre–calculated ozone AMFs building the training data set and on–line single scattering AMFs for NO_2 in the VIS fitting window.

- Geometric AMFs are calculated now for other species besides O_3 and NO_2.
- The cut–off parameter was re–set to SZA = 90° (92° in previous versions) because AMFs in the new LUT are not available under twilight conditions (SZA > 90°).

- VCD
  - An iterative scheme following [A37] has been implemented to derive the ozone vertical content.
  - A new flag has been implemented that indicates the usage of the iterative scheme for total ozone computation. This flag will be written to the AMF flag array and is part of the GOME level 2 product.
  - The intensity weighting of AMFs across the footprint is now controlled by a flag in the initialisation file. It is switched on by default.

2 Pre–processing Algorithms

2.1 Introduction

There are three pre–propcessing algorithms in the operational chain which are summarized in this chapter. The Absorbing Aerosol Indicator Algorithm (not used up to GDP 2.7), the tessellation algorithm (firstly used in GDP 2.7), and an algorithm for the determination of snow–covered scenes using GOME backscatter measurements and PMD measurements which is still in a premature state.

2.2 Absorbing Aerosol Indicator Algorithm

The Absorbing Aerosol Indicator Algorithm (AAIA) is a new stand–alone algorithm in GDP that is currently not called in the operational chain. Therefore, a detailed description of this algorithm is postponed until the code is used. As stated in the introduction, the results are currently unvalidated and an additional change of the product specification and the level 2 product is required.

2.3 Tessellation Algorithm

A new algorithm (tessellation algorithm) is used to derive an area–weigthed value for the ground albedo and the height above sea level of the GOME footprint. The area–weighted quantities take into account the effect of surface inhomogeneities in the GOME footprint. Application of the algorithm is confined to ground pixels with 1.5s integration time. If larger pixels are to be processed (6s or 12s integration time), the standard nearest neighbour technique is applied (nearest neighbours in terms of latitude/longitude of the centre of a GOME footprint compared to fixed latitude/longitude co–ordinates of data base entries).

The calculation of areas is carried out taking into account the sphericity of the Earth's surface. This is especially important over high latitudes where a large number of sub–areas (> 50) may occur in a single GOME footprint. This is due to the common resolution of data bases of topography and albedo which is typically 1 x 1 degree. It is therefore obvious, that a GOME footprint closer to the poles may comprise a larger number of such 1 x 1 degree areas.

A detailed description of the algorithm is laid down in [A 12].

2.4 Snow/Ice Detection Algorithm

A simple snow/ice detection algorithm using GOME backscatter measurements and GOME PMD measurements is currently under investigation. A first step is to calculate the sun–normalized PMD reflectances (cosine–weigthed) for each sub–pixel and channel.

$$PMD_{refl}\,(pixel, ch) \;=\; PMD_{meas}\,(pixel, ch) \cdot \pi \;/\; \cos(SZA) \qquad (1)$$

For each PMD channel we can define a threshold value (e.g. 0.4) for the reflectance above which a sub–pixel is denoted as snow–covered. To avoid confusion with cloudy pixels we make use of the ICFA–derived cloud coverage that must be below a second threshold value (e.g. 0.2). If both conditions are fulfilled, we see an almost cloud–free but bright pixel and we can flag such a pixel as snow–covered.

The simple algorithm is not used in the operational chain. However, first results over Mongolia showed its usefuleness to detect snow–covered scenes under low cloudiness conditions. More tests are necessary to tune the thresholds mentioned above for various surface and illumination conditions, especially over bright desert surfaces.

3 Initial Cloud Fitting Algorithm

3.1 Introduction

In order to determine ozone amounts more accurately, a correction is required in the Level 1 to 2 processing for cloudy and partially cloudy scenes. In its first operational version, the ICFA algorithm is confined to stand–alone GOME measurements; synergy with ATSR measurements from ERS–2 will not be considered at this stage.

Information about clouds will be extracted from GOME measurements both outside and within the well–known O_2 A–band (around 760 nm). The average transmittance through this band defines a relationship between cloud–top height and fractional cloud cover. If a canonical cloud–top height is specified, then the fractional cover is determined (this will be method employed here).

The fitting algorithm is based on the least squares comparison of GOME–measured atmospheric reflection functions and their simulated equivalents (sections 3.2 and 3.3). Simulated atmospheric reflectances are generated using a simple radiative transfer model. O_2 band absorption is the dominant feature, and templates of high–resolution atmospheric transmittances in the A–band are pre–calculated using suitable line–by–line code (section 3.4).

No information about the scattering properties of the clouds is assumed; instead, the algorithm approximates clouds as (bi–directionally) reflecting lower boundaries. This algorithm is only valid for optically thick water–droplet clouds, and cloud–top heights are therefore restricted to the troposphere. Bi–directional cloud–top reflectance depends on the optical depth of the underlying cloud; this is an external parameter to the algorithm (it cannot be inferred from GOME data in the present retrieval scheme). In the operational algorithm, clouds will be assumed semi–infinitely optically thick (see remarks in section 3.4).

3.2 Simulated and Measured Reflection Functions

The *measured* mean reflection function from the atmosphere in GOME wavelength bin $\Delta\lambda$ is :

$$R_{meas}(\lambda) = \left(\frac{\pi}{\mu_0}\right) . \frac{I(\lambda)}{F(\lambda)} \qquad (2)$$

where $I(\lambda)$ = radiant intensity as measured by GOME (radiance units).
$F(\lambda)$ = solar irradiance as measured by GOME.
μ_0 = cosine solar zenith angle.

For the *simulation*, the reflection functions are assumed to be linear combinations of reflectances from the ground and from the cloud–top. In the computation of reflectances, it is assumed that radiative transfer is dominated by O_2 band absorption, and that all other scattering and absorption may be approximated by a closure term. Thus if f is the fractional cloud cover, we may write :

$$R_{sim}(\lambda) = f . R_{cloud}(\lambda) + (1-f) . R_{ground}(\lambda) + R_{closure}(\lambda) \qquad (3)$$

where $R_{cloud}(\lambda)$ is the convolution over the slit function $\Phi(\lambda'-\lambda)$ of the cloud–top reflectance $\varrho_{\lambda'}(\mu,\mu_0)$ and the transmittance $T_{\lambda'}(p_c,\mu,\mu_0)$ due to O_2 absorption from the top of the atmosphere to the cloud–top (where the pressure is p_c) :

$$R_{cloud}(\lambda) = \int \Phi(\lambda'-\lambda) . \rho_{\lambda'}(\mu,\mu_0) . T_{\lambda'}(p_c,\mu,\mu_0) . d\lambda' \qquad (4)$$

Here, μ is the line–of–sight zenith angle to the satellite. A similar expression can be written for the ground reflection term.

The following assumptions are made to simplify the model still further :

- The cloud–top and ground reflectances are assumed constant with wavelength over the fitting window chosen to cover the O_2 A–band. We call these quantities $\alpha(\mu,\mu_0)$ and β respectively. The ground reflectance β is assumed Lambertian, but the cloud–top reflectance $\alpha(\mu,\mu_0)$ is bi–directional and will depend on the viewing geometry.

- The O_2 transmittances depend on the geometry via the geometric path factor :

$$S = \left(\frac{1}{\mu_0}\right) + \left(\frac{1}{\mu}\right) \tag{5}$$

  where μ_0, μ are the cosines of the solar zenith angle and the line–of–sight angle, respectively. Pre–computed tables of transmittances will be prepared using a specially–written accurate line–by–line code (see section 3.4).

- The closure term varies slowly with wavelength in a linear fashion over the window taken for the fitting.

With these assumptions, the total simulated atmospheric reflectance may be written :

$$\begin{aligned} R_{sim}(\lambda) &= f\,\alpha(\mu,\mu_0) \int \Phi(\lambda-\lambda_0)\, T_\lambda(p_c,\mu,\mu_0)\, d\lambda \qquad (\textit{cloud}) \\ &+ (1-f)\,\beta \int \Phi(\lambda-\lambda_0)\, T_\lambda(p_g,\mu,\mu_0)\, d\lambda \qquad (\textit{ground}) \\ &+ \gamma\,(1-\lambda/\lambda_0) \qquad (\textit{closure}) \end{aligned} \tag{6}$$

Here, p_g is the pressure at ground level.

3.3 Least Squares Fitting of the Reflection Function

Least squares fitting involves the minimisation of the chi–squared merit function :

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{R_{meas}(\lambda_i) - R_{sim}(\lambda_i)}{\epsilon(\lambda_i)} \right)^2 \tag{7}$$

for the set of fitting parameters $\{f, p_c, \alpha, \beta, \gamma\}$. The fitting window contains N observations through a given O_2 absorption band, and $\varepsilon(\lambda_i)$ are the errors on the individual measurement reflections $R_{meas}(\lambda_i)$.

The simulated reflectance is non–linear in these fitting parameters, and research has shown that it is hard to achieve non–ambiguous results for the simultaneous determination of two or more of the above set of parameters. In order to achieve sensible results for the main parameters of interest (cloud–top pressure p_c and fractional cover f), three further assumptions will be made :

(1) Define three new *linear* fitting parameters P_1, P_2 and P_3 :

$$P_1 = f\,\alpha(\mu,\mu_0) \;; \qquad P_2 = (1–f)\,\beta \;; \qquad P_3 = \gamma$$

(2) Fix a canonical value of the cloud–top pressure. This will be done according to the following empirical equation (p_c in units of bars, ϕ is the latitude, A and C are constants) :

$$p_c = A + C * (1.0 - cos(2\phi)) \quad (8)$$

(3) Assume that the cloud top reflectance $\alpha(\mu,\mu_0)$ is taken from a climatological database, and that cloud fractional cover is derived from the first fitted parameter P_1 via the expression :

$$f = \frac{P_1}{\alpha(\mu,\mu_0)} \; ;$$

With these assumptions, the simulated reflectance is now :

$$R_{sim}(\lambda) = P_1\langle T_{cloud}(\lambda,\mu,\mu_0)\rangle + P_2\langle T_{ground}(\lambda,\mu,\mu_0)\rangle + P_3(1-\lambda/\lambda_0) \quad (9)$$

where the $\langle\ \rangle$ symbol denotes slit function convolution. This last expression is linear in the three fitting parameters P_1, P_2 and P_3. The corresponding *linear least squares regression* will give unique answers for the fitting parameters, and from the third assumption above, a unique answer for fractional cover.

Linear least squares regression will be performed using the fast and stable single value decomposition technique – this fitting algorithm is well documented and the algorithm SVDFIT (reference [A9], Chapter 14) has been chosen for this task. A flag will be set for the detailed output of standard fitting diagnostics (chi–square, root–mean–square, correlation matrix, error variances on the fitting parameters, fitted and "rest" spectra). [See Section 4.4 for more discussion on least squares fitting diagnostics].

The root–mean–square error and the closure parameter P_3 will be used to flag poor fits and dubious results (these values are divided by the average GOME measurement across the window and the results compared with parameter values). The fractional cover must lie in the range [0,1]: a result of 5% or less for *f* is rounded to zero (clear pixel), 95% or more is rounded to 1 (totally cloudy situation).

Fitting parameter P_2 is not used. The choice of P_1 rather than P_2 to determine fractional cover is mainly due to greater *a priori* uncertainties in the surface reflection compared with cloud–top reflection. These uncertainties could be due to varying or poorly known wavelength dependence of the surface reflection, and the variability of the ground scene (snow/desert/ocean, etc.). In addition, the effect of tropospheric aerosols on O_2 band absorption should ideally be included in the simple model of equation (2.5) – this could be included in the albedo β, which would then no longer be just the surface reflectance.

3.4 Preparation and Usage of Data Bases

Calculation of the A–band transmittances

Two basic sets of transmittances are required for the simulation of reflectance – one to cloud–top, the other to the ground. These must be derived from an accurate (better than 1%) line–by–line computation, and dedicated radiative transfer transmittance code has been written for this task. The O_2 A–band has been chosen in preference to the B–band, partly because of stronger absorption showing up in the measurements (see also the comments in 3.6 below).

The line–by–line code uses molecular spectroscopic parameters derived from the HITRAN database. The model uses a 16–layer standard atmosphere, and calculates transmittances for an average

viewing geometry (solar zenith $\mu_0 = 60°$, line–of–sight zenith $\mu = 22°$). The initial data base will contain 7 transmittances, corresponding to the lowest 7 pressure levels in the model atmosphere. These templates will be pre–calculated (line–by–line models are computer intensive) and stored as one of the climatological databases needed for Level 1 to 2 processing. The resolution should be appropriate for the subsequent convolution with the GOME slit function – 0.0025 cm^{-1} has been chosen as representative (10000 points across the A–band window [759–780 nm]).

The convolution of template transmittances down to the GOME spectral resolution must be done for the calibrated wavelength grid appropriate to a given measurement set; the GOME Channel 4 slit function is also required. If the slit function and all the wavelength calibration choices are known in advance, then *pre–convoluted transmittances* can be stored and summoned as required without the need for lengthy convolution calculations.

For a given cloud–top pressure and ground–level pressure, the two (convoluted) templates required in Eq. (9) will be found by interpolation (over pressure) of the *convoluted* database transmittances. Templates can then be scaled for other viewing geometries.

Saturation levels are reached with many line absorptions in the A–band, particularly for transmittances down to the ground. A saturation criteria is chosen such that convoluted ground–level transmittances lower than a certain value are excluded – an initial choice of a maximum of 5% non–linearity in the convoluted ground transmittance has been implemented. This value is a parameter input to the algorithm – further studies (and in particular, testing during the commissioning phase) could result in another value being chosen. This saturation parameter has the effect of masking selected observations from the fitting window.

Cloud–top reflectance data base

Scattering studies show that all types of tropospheric clouds have very similar scattering properties in the wavelength range of the O_2 A–band (the single scattering albedo is practically unity). Nonetheless, cloud–top reflectances $\alpha(\mu,\mu_0)$ do depend on the optical depth of the underlying cloud; a correction factor should be added to the semi–infinite cloud–top reflectance to allow for transmission loss through the cloud. The formula for cloud–top reflectance is :

$$\alpha(\mu,\mu_0) \;=\; \alpha_\infty(\mu,\mu_0) \;-\; \left(\frac{A\,K(\mu)\,K(\mu_0)}{B \;+\; C\,\tau_{cloud}}\right) \tag{10}$$

$\alpha_\infty(\mu,\mu_0)$ is the reflectance from a semi–infinitely optically thick cloud layer. The transmission loss correction term assumes no absorption by cloud particles; A, B, and C are constants, $K(\mu)$ and $K(\mu_0)$ are the ”escape functions”.

The semi–infinite reflectances $\alpha_\infty(\mu,\mu_0)$ were calculated using one of the invariance principles in radiative transfer theory. The data base is classified according to cloud type (8 tropospheric types from Stratus to Cumulonimbus), wavelength (10 wavelengths covering GOME range), and geometry (9 direction cosines). There is also some azimuth dependence in the cloud–top reflectances, and this is reflected in the data base – however, in the operational environment only the dominant azimuth–independent term is used in Eq. (10). The transmission correction term in Eq. (10) is azimuth–independent. Tables of escape functions $K(\mu)$ have been prepared as part of the cloud climatological data base for Level 1–2 processing – the classification is the same as that for $\alpha_\infty(\mu,\mu_0)$.

As mentioned already, there is no *a priori* knowledge of cloud optical thickness τ_{cloud}, and in the operational algorithm, clouds will be assumed to be semi–infinitely optically thick. Although this assumption may result in values of fractional cover slightly lower than those obtained for clouds of finite optical thickness, additional errors on the total atmospheric ozone columns due to this uncertainty will not be significant. The main reason for this is that most atmospheric ozone lies in the

stratosphere, and it is worth stressing again that the ICFA algorithm is intended to provide a simple correction to the *tropospheric* column of ozone due to the presence of clouds. Note however, that clouds may have an impact also on the stratospheric ozone column via the albedo effect [A24].

3.5 Summary of Algorithm Processing

- *Input Requirements*

1. Fixed parameters (saturation criterion, closure and root mean square maximum parameters, reference geometry parameters, fitting window, etc.).
2. Extracted Level 1 data (channel 4 GOME back–scatter radiances and solar irradiances and their errors, plus appropriate geolocation information).
3. Data bases (global ground topography, cloud–top reflectances, O_2 A–band transmittance templates).

- *Algorithm Steps*

– Read in the fixed parameters required for the algorithm. [These parameters would not normally be changed during routine operational execution].

– From geolocation information, get viewing geometry. Buffer extracted GOME measurements to given fitting window, and compute measured reflectances $R_{meas}(\lambda_i)$ and their errors.

– If calibration choice is known, extract pre–convoluted transmittance templates. If not, convolute high–resolution template transmittances from database to λ–grid of fitting window. Interpolate to current ground–level pressure and cloud–top pressure, and correct for current viewing geometry.

– For given geometry, interpolate cloud–top reflectance $\alpha(\mu,\mu_0)$ from data base.

– Create mask for observations by applying saturation parameter. Compute basis functions for the linear fitting routine.

– Perform linear fitting and generate fitting diagnostics.

– Compute fractional cover and total error on it. Set quality flags for fitting. Set default values if fitting has failed.

- *Output*

– Fractional Cloud cover, plus error on this. Two flags for quality of fit. (Canonical cloud–top pressure is also passed on as output).

– (Optional). Detailed diagnostics from the fitting, including covariances and correlation matrix, r.m.s error and goodness–of–fit, spectral information, etc.

3.6 Open Issues (Issue 1/A)

Templates

The first version of the algorithm assumes plane parallel atmosphere for the viewing geometry correction. For high solar zeniths, the template transmittances should really be calculated using full spherical geometry – this complicates the line–by–line model considerably and off–line work will be required to extend the transmittance database to cover this contingency.

Slit Functions

Convolution software based on analytic expressions for the GOME slit functions has been prepared and implemented, but the final forms of the slit functions have not yet been received. These are expected following the completion of the Flight Model (FM) calibration exercise.

Cloud–top reflectances

Though comprehensive data bases of semi–infinite cloud–top reflectances and escape functions have been compiled from radiative transfer simulations, these quantities must be validated against real observations of cloud–filled scenes (commissioning phase and afterwards).

O_2 line parameters

Some of the spectroscopic parameters from the HITRAN data base are known to have high uncertainties; this is especially true of the A–band values (the B–band parameters are in general better known). The need for updated and more accurate molecular data (line strengths and half–widths, transition energies, pressure broadening coefficients) has been identified within the GOME project, and should new data become available, the template calculations will need to be repeated and validated before reprocessing GOME data.

Other parameters

The choice of average viewing geometry and saturation parameter could be fine–tuned during the commissioning phase. Experience with real data during the commissioning phase should help to fix the root–mean–square and closure parameters used to establish the quality and noisiness of the fit.

Fall–back Option

The present algorithm is based on the use of two templates. Should this fitting procedure fail to work, it is still possible to use a single template to determine an average reflecting height (pressure) in the atmosphere, and assume that this height is the lower reflecting boundary in subsequent Air Mass Factor calculations.

3.7 Algorithm Updates

3.7.1 Issue 2/A – GDP 2.0

Templates

The open issue regarding templates mentioned above has now been addressed, and a revised data set created to include the possibility of ray–traced slant paths for solar zenith angles up to 90 degrees.

It is not necessary to include ray–tracing code in the operational GDP to calculate slant path lengths; instead an auxiliary data set of slant path factors has been generated off–line, for the given model atmosphere used in ICFA and for a number of solar zenith angles up to 90 degrees. The ray–tracing code was that used in the forward model GOMETRAN, and permission to use the code in the present context was granted by the University of Bremen.

The new high–resolution template data base actually consists of 16 *vertical layer transmittances* specified at 11001 spectral points between 12780 cm^{-1} and 13220 cm^{-1}, at a resolution of 0.04 cm^{-1}. (Updated O_2 A–band spectroscopic parameters [A31] were used in the generation of this database, see [A3] for details). The *cumulative transmittance* C_j down to the lower boundary of layer j (j = 1, 16) is given by :

$$Ln(C_j) = \sum_{k=1}^{j} \left(S_k(\theta_0) + \frac{1}{cos(\theta_1)} \right) \cdot Ln\,(T_k) \qquad (11)$$

where T_k is the vertical transmittance through layer k, and $S_k(\theta_0)$ is the solar slant path factor for layer k interpolated to solar zenith angle θ_0 from the auxiliary data base. θ_1 is the line of sight zenith angle, for which the geometrical path factor $1/cos\,(\theta_1)$ has been assumed.

Cumulative transmittances to the lowest 7 layers of the model atmosphere are computed in this manner. These are then convoluted with the GOME slit function. The convoluted transmittances are then interpolated to the cloud–top height and ground height defined by the ICFA algorithm.

It should be noted here that Eq. (8) for the cloud–top pressure has now been superceeded. Cloud–top pressure is now taken directly from an ISCCP data base (classification is by month, by latitude and by longitude – see [A3] for the details of this data set).

Cloud–top reflectances

Eq. (10) for the cloud–top reflectance is still valid, but the constants A, B and C have been re–defined to include the effect of ground–reflected light on the transmission loss term (there is a reduction in the amount of light lost through the cloud when the underlying ground surface is bright). The new definitions are :

$$A = 4 \; ; \; B = 6q + \frac{4\beta}{(1-\beta)} \; ; \; C = 3\,(1-g)\,\tau$$

where q = 0.71392, β is the ground albedo, g is the asymmetry parameter of the cloud particles, and τ is the optical depth of the cloud. It should be stressed that these constants apply for cloud particles with single scattering albedos equal to unity.

The additional term 4β/(1–β) is new. A flag can be set for the inclusion of this term; the default is to include it.

Cloud Clearing Algorithm (CCA)

This alternative to ICFA was developed during the work done for the ESA Scattering Studies Group (see [R12]). It employs the sub–pixel information contained in the PMD readouts from PMDs 2 and 3, and is a simple decision–making algorithm based on thresholds. It also generates a cloud fraction cover. It became clear soon after launch that the PMD reflectances generated in the Level 1 product easily showed the presence of strong reflectors within the pixel scene, and it was decided in 1995 to incorporate the algorithm in the GDP level 1–2.

Based on the magnitudes of the reflectances R_2 and R_3 from PMDs 2 and 3, the sub–pixel is deemed cloudy or clear or undetermined, depending on whether R_2 or R_3 or the ratio R_2/R_3 exceed or fall beneath certain threshold values. Further refinements are necessary when the initial threshold test yields an undetermined answer. There are distinctions between land and sea surfaces, and the sub–pixel status cannot be assigned when the underlying surface is covered with snow or ice. The fractional cover is defined as (the notation is clear) :

$$f_{CCA} = \frac{N_{cloudy}}{N_{clear} + N_{cloudy}} \qquad (12)$$

The original set of thresholds were simulated for a number of solar zenith angles and a limited number of reflecting ground surfaces. These thresholds are static; it was envisaged that a large and dynamically–updated set of thresholds would eventually be used for this algorithm. However, the ex-

tension of the threshold data base has not been attempted for the present deployment in GDP, and there some severe processing problems in the creation of dynamic thresholds.

Output from the algorithm is written to the Level 2 product; this comprises the fractional cover, and 16 numbers indicating the CCA status of each subpixel (0 = clear, 1 = cloudy, 2 = undetermined). These CCA results are not used in any other part of the GDP 1–2 system. Only the ICFA result is used in the AMF and VCD algorithms to allow for cloud contamination in the geophysical retrieval. One can thus regard the CCA results as a useful diagnostic tool.

Remarks

The ”average viewing geometry” mentioned in the previous section no longer applies, as the transmittances are calculated cumulatively with correct slant path factors.

It was found that the best fits were obtained when the saturation parameter was set as low as possible (0.01); effectively, no points have been masked in the fit. A wavelength range of 758.4 to 778.5 nm has been the default.

It has not been necessary to rely on the ”fall–back option” mentioned above in the Open Issues section. The ICFA has proved stable in operational running.

3.7.2 Issue 2/B – GDP 2.7

The ICFA slit function type has been changed from “rectangular” to “simple–hyperbolic”. This change was already in place for version 2.4. It shall be mentioned, that the usage of a rectangular slit function was never considered but happened due to wrong settings in the static initialisation file.

3.7.3 Issue 3/A – GDP 3.0

It has been realised recently that the spectral calibration of both earth–shine and sun spectra in Ch4 changes with time. The beginning of Ch4 may differ by about 1 detector pixel, i.e. around 0.23 nm. Typically, older spectra (with respect to the GOME lifetime) start with lower wavelength. This is obviously an artefact of L01 processing using the calibration lamp lines for spectral calibration. At the beginning of the GOME lifetime a sufficient number of lamp lines was available for each channel and the final polynomial fit was working satisfactory. Due to the aging process of the calibration lamp an increasing number of lamp lines is filtered out as being unreasonable and the polynomial fitting becomes poorer and my finally lead to a different assignment of the wavelength to the first detector pixel.

A small data base of the spectral channel limits (i.e., start and end wavelength of the channel) and a spectral interval around these limits have been integrated, in order to detect these spectral shifts during the processing. To pass the test, the actual first wavelength of each channel must lie inside the spectral interval around the fixed wavelength taken from the data base. The static channel limits are optimized for spectra from years 1998 and before. During the processing of more recent orbits an enhanced number of ground pixels, say spectra, are recognized as being miscalibrated.

However, the impact on cloud coverage results is low, typically the observed changes are around several percent (random). Therefore, the processing of such pixels will not be aborted but a corresponding (new) flag is set accordingly.

Under high sun zenith angle conditions, the spectral fitting of simulated to measured spectra may lead to normalized cloud coverage results below 0. or higher than 1. Users will be informed about these unphysical results by a new flag which is part of the ICFA flag output that is written to the GOME level 2 product.

4 DOAS Spectral Fitting Algorithm

4.1 Introduction

The Differential Optical Absorption Spectroscopy (DOAS) technique as described e.g. in [A28] is used for the retrieval of atmospheric trace gas effective slant column (ESC) amounts from moderately high–resolution spectral data in the UV and visible regions of the spectrum. GOME represents the first application of the technique to passive remote sensing instruments in space. The first operational algorithm will focus on the retrieval of atmospheric columns of ozone.

DOAS involves the least–squares fitting of ratioed measurement spectra to a set of reference spectra (absorption cross–sections or instrument–specific reference measurements). The fitted ESCs are converted to geometry–independent vertical columns by division with an appropriate Air Mass Factor.

Section 4.2 deals with the simulated spectrum, sections 4.3 and 4.4 with the fitting and diagnostic generation. The choice of DOAS windows is important for GOME, and this is dealt with in section 4.5. Reference spectra are discussed in section 4.6, which is followed by algorithm implementation and a summary of open issues (sections 4.7, 4.8).

4.2 The Simulated Spectrum

The trace gas absorption in modeled on the Lambert–Beer law: An incremental decrease of intensity $dI(\lambda)$ at wavelength λ through a slant path distance ds is proportional to the absorption coefficient $\sigma(\lambda)$ times the incident intensity $I(\lambda)$ and the absorber column amount $C(s)ds$:

$$dI(\lambda) = -I(\lambda) \cdot \sigma(\lambda) \cdot C(s) \cdot ds \tag{13}$$

When there are several absorbers, the contributions are additive. If now $I_0(\lambda)$ is the incoming radiation at the top of the atmosphere, and there are no other radiative transfer processes in operation, then we can integrate Eq. (13) to obtain an expression involving the sum of optical densities :

$$ln\left(\frac{I(\lambda)}{I_0(\lambda)}\right) = -\sigma_1(\lambda)\, A_1(s) - \sigma_2(\lambda)\, A_2(s) - .. \tag{14}$$

where :

$$A_j(s) = \int_0^s C_j(s)\, ds \tag{15}$$

is the effective slant column density of absorber j over path length s.

Absorption coefficients are usually taken from data sets of absorption cross–sections expressed in units of [$cm^2.mol^{-1}$]; if the concentrations $C_j(s)$ are in [$mol.cm^{-3}$] then slant columns are in units of [$mol.cm^{-2}$]. When reference data are given in laboratory measured intensities or counts, the effective slant columns must be normalised *a posteriori* by an optical density additional to the reference data.

In the real atmosphere, incident solar light will be further attenuated by molecular (Rayleigh) scattering, and by absorption and scattering due to other particulates (aerosols, clouds); there is also the surface reflecting property of the earth's surface to be accounted for. These effects contribute

broad–scale spectral features to the back–scattered spectra; in the DOAS approach, such effects are filtered out from the highly structured and more finely resolved *differential* trace gas spectral signatures. The broad scale features are approximated by a low order polynomial in wavelength. Thus, Eq. (14) can now be extended :

$$Y_{sim}(\lambda) \equiv ln\left(\frac{I(\lambda)}{I_0(\lambda)}\right) = -\sigma_1(\lambda)A_1(s) - \sigma_2(\lambda)A_2(s)..- B_1 - B_2(\lambda-\lambda_0) - B_3(\lambda-\lambda_0)^2 \quad (16)$$

where a polynomial of degree 2 has been assumed. The quantity $Y_{sim}(\lambda)$ is the simulated optical density.

4.3 Least Squares Fitting of Optical Density

The simulated quantity must be fitted to the measured optical density given by :

$$Y_{meas}(\lambda) \equiv ln\left(\frac{I_{nadir}(\lambda)}{I_{sun}(\lambda)}\right) \quad (17)$$

Least squares fitting involves the minimisation of the chi–squared merit function :

$$\chi^2 = \sum_{i=1}^{N}\left(\frac{Y_{meas}(\lambda_i) - Y_{sim}(\lambda_i)}{\sigma_{meas}(\lambda_i)}\right)^2 \quad (18)$$

to establish the set of *fitting amplitudes* or *regression coefficients*, which are here the slant column amounts $\{A_i\}$ and the polynomial coefficients $\{B_j\}$ in Eq. (16). The fitting window contains N observations, and $\sigma_{meas}(\lambda_i)$ are the individual errors on $Y_{meas}(\lambda_i)$.

Optical densities are dimensionless numbers. If the absolute errors $\varepsilon_{nadir}(\lambda_i)$ and $\varepsilon_{sun}(\lambda_i)$ on $I_{nadir}(\lambda_i)$ and $I_{sun}(\lambda_i)$ respectively are known, then we shall define the relative optical density measurement error $\sigma_{meas}(\lambda_i)$ in Eq. (18) as :

$$\sigma_{meas}(\lambda_i) = ln\left\{\frac{1 + \frac{\epsilon_{nadir}(\lambda_i)}{I_{nadir}(\lambda_i)}}{1 + \frac{\epsilon_{sun}(\lambda_i)}{I_{sun}(\lambda_i)}}\right\} \quad (19)$$

Eq. (18) is then a *weighted* least squares sum, with weights $\sigma_{meas}(\lambda_i)^{-2}$. Individual measurement errors are often not known in DOAS spectral fitting, and it is then customary to set $\sigma_{meas}(\lambda_i) = 1$ for all points in Eq. (18); the fitting is then unweighted**.**

The simulated optical density $Y_{sim}(\lambda)$ is *linear* in the fitting parameters $\{A_i, B_j\}$ – the number of such parameters equals the number of reference spectra used in the fit plus the degree of the filtering polynomial. If no further adjustment of the reference spectra is assumed, then the fitting is linear least squares (linear multiple regression). [In most applications of DOAS, little useful information about particulate properties or surface reflectance can be gleaned from the filtering amplitudes $\{B_j\}$. These parameters are usually discarded, but their cross–correlations with amplitudes $\{A_i\}$ sometimes yield information about the fitting].

The fitting can be improved when the reference spectra are adjusted to allow for uncertainties in the wavelength calibration of the instrumental measurements. An entire spectrum can be translated in wavelength by a single value (shift); it can also be stretched or compressed by a single value (squeeze) about some reference point. A shift and squeeze together will move the wavelength registration from λ to $\lambda*$:

$$\lambda^* = \lambda + <shift> + (<squeeze> - 1).(\lambda - \lambda_{ref}) \quad (20)$$

where λ_{ref} is a reference wavelength (usually taken to be the middle value of the fitting window).

For each trace gas reference spectrum, the shift and squeeze can be allowed to vary in the fitting. The simulated optical density is still given by Eq. (16) above, but this now depends *non–linearly* on the shift/squeeze pairs (which must therefore be fitted using a non–linear least squares fitting algorithm.). However, linear fitting still applies for the amplitudes $\{A_i, B_j\}$; the merit function χ^2 is the same for both fitting processes.

Non–linear fitting of shifts and squeezes follows an iterative procedure; the algorithm will search for the smallest χ^2 in fitting–parameter space using a modified "steepest descent" method for each guess of the shift/squeeze pair. For every such iteration, linear least squares fitting must be performed for the amplitudes $\{A_i, B_j\}$ – the linear fit is a prelude to each iteration of the non–linear fit. Once the non–linear fitting has found the optimum shift/squeeze values (maximum number of iterations should be 20), a final linear fit is taken to confirm the output values of the fitting amplitudes $\{A_i, B_j\}$.

Both fitting algorithms are adapted from Numerical Recipes routines ([A9], Chapter 14) – the single value decomposition routine SVDFIT for the linear fitting, and the Levenberg–Marquand algorithm MRQMIN for the non–linear fitting.

4.4 Fitting Diagnostics

Fitting diagnostics indicate the quality of the fit. It is usual to assume that the measurement data are independent, and the propagation of measurement errors then leads to the following equation for the variance of any fitted parameter F :

$$Var_F = \sum_{i=1}^{N} \sigma_i^{\,2} \left(\frac{\partial F}{\partial Y_{meas}(\lambda_i)} \right)^2 \quad (21)$$

where σ_i are the individual errors on the measurement values $Y_{meas}(\lambda_i)$. $\{Var_F\}^{0.5}$ is then the error on the fitted amplitude F. These variances are the diagonal elements of the Covariance matrix – the off–diagonal elements express the correlations between pairs of parameters. In the DOAS algorithm, the correlation matrix will be the standard diagnostic, with off–diagonal elements between –1 and +1. Of particular interest are correlations between fitted slant columns for different gases, and correlations between the broad–scale polynomial coefficients and the trace gas slant columns.

The final minimum value of the merit function (Eq. (18)) will be specified in the diagnostics. Also useful is the root–mean–square (RMS) error :

$$RMS = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left\{ Y_{meas}(\lambda_i) - Y_{sim}(\lambda_i) \right\}^2 \quad (22)$$

For the linearly–fitted parameters, one can define the goodness–of–fit statistic :

$$R = Q\left(\frac{\nu}{2}, \frac{\chi^2}{2} \right) \quad (23)$$

This is the probability that a value of χ^2 as poor as the value in Eq. (18) should occur by chance. [The chi–square probability density function is the incomplete gamma function Q, with number of degrees of freedom ν =N –M (number of data points minus number of fitted parameters)]. If R is larger

than 0.1, the goodness–of–fit is generally acceptable; values of R consistently less than 0.001 usually indicate an unacceptable fitting model.

When the fitting is unweighted, errors on the fitting amplitudes $\{A_i, B_j\}$ must be multiplied by the factor $\sqrt{(\chi^2/\nu)}$, where ν = N–M as before. This is equivalent to the *assumption* of a good fit, so the goodness–of–fit criterion does not apply. For GOME, errors on the data should be known from the Level 0–1 processing; the fitting should be weighted least squares, with an independent goodness–of–fit statistic available.

A visual examination of fitted and measured spectra is often the best test of the DOAS fit, and detailed spectral output will be produced during the commissioning phase. It will not be possible to generate this detailed spectral information routinely during normal operational running of the instrument. In the Level 2 Product, output information from the DOAS fitting will be confined to the fitted trace gas slant columns and their errors, plus the values of chi–square, RMS and goodness–of–fit and the number of iterations required for convergence of the non–linear fitting.

4.5 Choice of Fitting Windows

The following considerations govern the choice of fitting windows for the retrieval of ozone column amounts.

- Windows should include distinct O_3 absorption features strong enough to be detected in all measured spectra from GOME. In this respect, windows must lie in the O_3 Huggins bands (310–350 nm) and/or the Chappuis bands (430–550 nm). Measurements are thus restricted to GOME channels 2 and 3.
- Interfering species should be avoided wherever possible. O_4 and H_2O are present at certain places in the Chappuis bands – their spectral properties are not known to a high degree of accuracy. NO_2 will always be present as an interfering species.
- Strong Fraunhofer lines should be avoided. The back–scattered spectrum will show partially filled Fraunhofer lines (the Ring effect), and this unwanted interference can be partially compensated for by the inclusion of a GOME–measured or theoretical Ring spectrum as one of the reference spectra.
- Windows should not cross channel boundaries (different wavelength registrations and spectral resolutions).
- Experience with ground–based DOAS retrieval has shown that the number of points N should not be too high (instability in the fitting algorithm) or too low (fitting becomes poorly determined). The range $50 < N < 500$ is suitable.
- Air Mass Factors should not vary much over a given window (only one representative AMF per window is calculated for the conversion to vertical column amounts). This limits the size of the UV window because the AMF increases quickly with wavelength over the range 310–340 nm (especially for high solar zenith angles); it is also sensitive to climatological inputs in this region.

Two ”strawmen” windows have been selected for the first operational algorithm. These choices are provisional at this stage, and should be confirmed during the commissioning phase. They are

(1) GOME channel 2 (UV ozone) 323–335 nm

(2) GOME channel 3 (Visible ozone) 430–535 nm

The UV window contains 4 strong Ozone absorption bands; it has limited interference from other trace gas absorbers (NO_2, BrO, occasionally SO_2 and OClO), and a smaller Ring effect interference. The most critical factor determining window size is the AMF variability. Wavelengths below 320 nm were not considered because of stray–light corruption and lower signal–to–noise.

For the visible window, the AMFs are smoother and less sensitive to climatological inputs. The main problem here is the presence of O_2–O_2 and H_2O; regions with larger O_2–O_2 absorption should be avoided (masked out) because of additional pressure dependencies in the spectra.

Though these windows are optimised for ozone retrieval, effective slant columns for interfering trace gas species will also be generated from the fitting. In addition to ozone slant columns and their errors, the Level 2 Product will contain other fitted trace gas slant columns and respective errors. [Pointers in the Product Header records will indicate which gas has been fitted in which window]. Beside O_3, slant columns of NO_2, BrO and (occasionally) SO_2 and OClO could be fitted in the UV, with O_3, H_2O and NO_2 columns retrievable in the visible window.

4.6 Reference Data

In the algorithm development phase, trace gas cross–sections have been taken from the literature. A data base of cross–sections forms part of the GOME Level 1–2 climatological data base. It may be more satisfactory during operational running to use cross–sections derived from GOME Bread–board model (BBM) measurements: it is proposed to replace the O_3 Chappuis and part of the NO_2 spectra with GOME–derived equivalents. (In this case, reference data would be converted to cross–section values before use in the DOAS algorithm).

A Ring spectrum can be computed from zenith sky measurements taken with GOME (already done for the BBM, will be performed during Flight Model calibration). The Ring spectrum can be used as an effective absorption spectrum in the DOAS fitting. The etalon spectrum can also be fitted – the reference values will also be derived from pre–flight calibration measurements.

O_3 cross–sections in the Huggins bands are temperature dependent, and a representative temperature has to be selected from climatology (this is the single atmospheric input to the DOAS algorithm). The current default chooses a temperature corresponding to the maximum concentration in an appropriate ozone climatological profile. The cross–section temperature dependence is expressed through the (empirical) quadratic interpolation formula of Bass & Paur :

$$\sigma_\lambda(T) = \sigma_\lambda(T_0) \cdot \left(1 + a_\lambda (T–T_0) + b_\lambda \left(T–T_0\right)^2 \right) \quad (24)$$

where the reference values $\sigma_\lambda(T_0)$ and the quadratic coefficients a_λ and b_λ are tabulated together in the data set, and temperature T is in K (T_0 = 273.15 K).

H_2O cross–sections have been calculated from line spectroscopic parameters ([A31]) using a dedicated line–by–line code. To avoid sampling problems, the cross–sections were computed at 0.01 nm resolution.

All cross–sections will be convoluted with the appropriate GOME slit functions. With the exception of H_2O, convoluted trace gas cross–sections will be preserved on the original wavelength grids used in the data bases. For the majority of trace gas absorbers, convolution will not radically change the cross–sections, due to the relatively high resolution of the instrument. H_2O cross–sections at 0.01 nm resolution are not suitable for fitting, and these will be convoluted down to the GOME resolution. [A surplus of fine detail in the high–resolution data set will obscure the fitting]. To highlight the differential features of the measured spectrum, it is also useful to filter out (linear fitting of low–order polynomial) the broad–scale continuum in the cross–sections.

If the choices of wavelength calibration are known in advance, then the cross–sections can be pre–convoluted, thus avoiding the tedious on–line repetition of convolution calculations. Pre–convolved quantities will be stored as part of the data base. [This convolution exercise can only be performed during the commissioning phase, when the calibration choices become known, and the final form of the slit function has been determined].

Reference spectra derived from instrument measurements should be first de–convoluted with the appropriate slit function, then re–convoluted with the GOME slit functions. This would not be so much of a problem for Ring spectra and CATGAS measurements taken with the GOME BBM and/or FM (though strictly speaking, the deconvolution should be performed).

It should be noted that the ”squeeze” operation is in a sense a convolution. Thus strictly speaking, convolutions on cross–sections should be performed after each new shift and squeeze assignation. The necessity for this repeated convolution should be determined in the testing phase.

4.7 Summary of Algorithm Processing

- *Input requirements*

1. Extracted Level 1 Data – radiances and their absolute errors, solar irradiances and their absolute errors.
2. Parameters controlling the fitting
    - Definition of fitting windows (how many, start and finish wavelengths for each window).
    - Choice of reference spectra to be used in fitting.
    - Degrees of fitted polynomials (broad–scale features).
    - Flags controlling fitting (use of weighted merit function, use of shift/squeeze options in non–linear fitting).
    - Shift and squeeze extremes. Specified *fixed* shifts and squeezes (linear fitting only).
    - Representative temperature for O_3 Huggins bands cross–sections.
3. Data bases (cross–sections, other reference spectra, slit function parameters, ozone profile climatology).

- *Algorithm Steps*
    - Extract parameter information from file (this would not normally be changed during operational running). Check parameter information and write to file.
    - Extract Bass–Paur temperature using ozone profile climatological database.
    - Buffer Level 1 data according to given fitting window, and compute measured optical density and individual errors (Eqs. (17) and (19)).
    - Buffer reference spectra to given window choice. Perform convolution with GOME slit function if required.
    - Apply low pass filter to emphasise differential features of reference data if required. Scale reference spectra and compute second derivatives for spline interpolation.

- Perform fitting (linear – single call to SVDFIT; nonlinear – repeated iteration using MRQMIN until convergence attained).
- Compute fitting diagnostics (fitting parameter variances and correlations, RMS, goodness–of–fit, chi–square, etc.).

- *Output*
  - Trace gas effective slant columns (fitted coefficients), errors on these columns, RMS, chi–square, goodness–of–fit, number of iterations (non–linear fitting).
  - (Optional). Detailed diagnostics from the fitting, including correlation matrix, residual spectra, fitted spectrum, fitted polynomial coefficients, fitted shifts and squeezes, etc.).

4.8 Open Issues (Issue 1/A)

Reference Spectra & Cross–sections

The implementation of GOME–derived cross–sections (in place of literature spectra) has not been decided. The implementation of Flight Model Ring and Etalon spectra awaits completion of the FM calibration exercise. The NO_2 cross–section data base is expected to be updated before launch.

Slit Functions

Convolution software based on analytic expressions for the GOME slit functions has been prepared and implemented, but the final forms of the slit functions have not yet been received. These are expected following the completion of the FM calibration exercise.

Non–linear vs. Linear fitting

Both options must be implemented in the operational code. Fixed shifts and squeezes can only be determined after the in–flight wavelength calibration choices are determined (commissioning phase). Policy on the operational checking of shift and squeeze parameters has yet to be decided.

Windows

Choice of fitting windows is as yet provisional, awaiting testing on real data during the commissioning phase.

Other parameters

The choice of a representative temperature for the O_3 Huggins bands cross–sections is provisional.

4.9 Algorithm Updates

4.9.1 Issue 2/A – GDP 2.0

The changes and improvements to the DOAS algorithm that have been implemented since the previous issue of this Technical note are described below. Many of the open issues noted in section 4.8 have been closed; this section lists and describes the changes, and discussion of the above mentioned open issues will be dealt with in the course of the text.

GOME FM Reference Spectra & Cross–sections

At the GOME Data and Algorithm Meeting on 9 January 1996, a recommendation was made to include the available GOME FM O_3 and NO_2 cross–sections in the level 1–2 databases, and to use these values in the DOAS fitting, in preference to the literature cross–sections. The GOME FM measurements were taken towards the end of the FM calibration phase in early 1995, and the data processing was completed in December 1995. (A short description of these data sets may be found in the update section of the GOME 1–2 databases technical note, document [A3]).

The DOAS algorithm is then using cross–sections derived from the same instrument; the slit function is the same for cross–section measurements and observations from space, and the spectral resolutions are comparable (though the wavelength registrations will be different). There is therefore no need to convolve the FM reference data with the GOME slit function. [Following the GOME FM calibration phase, a data set of parameters was derived for the generation of the analytic expressions describing the GOME slit functions in the four channels].

The GOME FM measurements were performed at a number of temperatures. A Bass–Paur–type quadratic fitting formula was applied to the O_3 cross–sections for use in computing the temperature dependence. 221 K was the representative temperature selected for O_3. No such formula is available for the NO_2 data. NO_2 vmr peaks in the stratosphere at 30–35 km, and the data set derived from measurements made at 241 K was used in the DOAS algorithm.

Off–line work has found that the DOAS slant column fitting was slightly improved in channel 2 using the GOME FM cross–sections instead of the Bass–Paur literature data. A substantial improvement was noticed in the visible window 425–450 nm in the NO_2 slant columns, and the old data set of NO_2 [A33] has now been superceded. An improved NO_2 data set has recently appeared in the literature [A23] and this too has been added to the reference spectra library (see [A3]). All options have been retained for the DOAS algorithm, which is therefore still able to use the old data in a consistent manner.

Measured GOME FM differential Ring cross–sections have also been derived from zenith sky spectra taken at the end of the FM calibration phase. These were incorporated in GDP in January 1996, with a recommendation for their use in DOAS. It has been shown that this differential Ring cross–section will fit the Fraunhofer structure in the L1 data more accurately than a GOME sun spectra, and it was therefore necessary to exclude the GOME solar spectrum from the linear part of the DOAS fit (the solar spectrum shifts and squeezes are still computed). Measured Ring cross–sections are only available in channels 2 and 3 down to about 325 nm, and new theoretical work has shown that it is possible to generate these Ring spectra from model studies of Raman scattering the atmosphere.

Strategy for DOAS Algorithm

Experience with real GOME data has shown that it it is not feasible to compute a fixed set of shifts and squeezes for a given set of wavelength calibrations and use theses fixed values in a straightforward linear fit. Therefore the shifts and squeezes options are always turned on in routine operational running, and the shifts and squeezes calculated from scratch for each retrieval.

The choice of fitting window in the UV (325–335 nm) has not altered following extensive testing. For the routine generation of an O_3 column from this window, only the GOME FM O_3 and Ring cross–sections are used in the fitting. The main level 2 product is derived from the retrieved O_3 column from this window.

The original specification in Channel 3 was 425–450 nm. It was seen that this window was suitable for NO_2 retrieval, but the O_3 results were not reliable. In an attempt to retrieve O_3 from the Visible, windows 450–495 nm and 510–550 were tried. For the first of these, it was found that the L1 radiances contained large irregular structures due to the polarisation sensitivity, and it proved difficult

to include these in the fitting. For the second window that polarisation sensitivity behavior is smooth enough to be filtered out with the low–order additive polynomial used in DOAS, but there remained substantial interference from water vapor. It was decided not to use a window in the visible for the routine production of O_3 column amounts, pending further research. [Note that run–time considerations also mitigated against the (somewhat time–consuming) retrievals in the visible].

4.9.2 Issue 2/B – GDP 2.7

The more physical error–weigthed fitting is now carried out in both operational DOAS fitting windows. The error–weighted fitting is disabled if the ratio of earth–shine data and sun spectrum is fitted (not used operationally).

A module for checking the wavelength calibration of both the sun and the backscatter spectrum has been added in the DOAS module chain. We look for the first spectral point of each channel and compare that wavelength to fixed wavelength values which are known to be representative for the first detector pixel. If the difference is greater than a certain threshold ($\pm$ 0.16nm) the processing of that window is aborted and an appropriate error message is generated. In contrast to previous GDP versions the processing continues with the next fitting window (if available).

All smoothing methods are switched off for all spectra during the operational processing.

Two ozone spectra at different temperature can now be fitted simultaneously for each fitting window. However, this is is only possible if ozone is not the main species of interest in a given fitting window because there is no calculation of the total ozone content and its error. This method is not used operationally.

The Bass–Paur temperature can be determined now either in the static initialisation file (fixed value for each species with temperature dependency) or it is taken from that layer where the ozone number density is maximum. The method is controlled by flag setting in the initialisation file.

An alternative matrix solver (LU decomposition) has been added to the standard Gauss–Jordan matrix solver in the DOAS core module (MRQMIN).

Updated GOME FM cross–sections are available now from IFE Bremen for both O_3 and NO_2 ([A15], [A17]). Based on the outcome of the delta validation period in Spring and Summer 1999 it has been decided to use the NO_2 spectrum in the VIS window but not the new FM data for O_3 in both windows, as these lead to an almost constant bias of –3% for all ozone values of a GOME orbit.

The NO_2 fitting now takes into account a theoretical Ring spectrum (SAO), and the interfering species O_4 (O_2–O_2 collision complex) and H_2O. Furthermore, an undersampling correction spectrum based on the work of Slijkhuis et al. (see [A34]) is applied simultaneously. The water vapor cross–sections are used in the same way as any other species because the water vapor continuum absorption is negligible in the spectral range of the standard VIS window.

4.9.3 Issue 3/A – GDP 3.0

Besides the application of other reference spectra and additional interfering species (see in the introduction) the main changes in DOAS fitting are with the application of an ozone difference spectrum, i.e. the difference of ozone cross–sections at different temperatures ([A29], [A30]). The temperature dependency of ozone cross–sections in the Huggings bands can be expressed by:

$$\sigma(T) = \sigma(241K) + \Delta T \cdot \frac{\partial \sigma}{\partial T} \qquad (25)$$

where the first derivative can be assumed as constant over the range of temperatures experienced and simply given (as an example) by the difference between the cross–sections at 241K and 221K (operational default). Thus, the ozone spectrum at 241K and the difference spectrum are fitted simultaneously and the final ozone slant column content (and it's error) is simply the fitted value A_1, while the second term A_2 can be associated with an effective ozone temperature.

$$T_{eff} = T_{241} + \frac{A_2(241K - 221K)}{A_1(241K)} \cdot \Delta T \tag{26}$$

As stated already in the introduction, a number of further changes have been implemented in this version. However, these changes do not affect the core algorithm and are not repeated here.

5 Air Mass Factor Algorithm

5.1 Introduction

The Air Mass Factor (AMF) algorithm is the second major algorithm in the Level 1–2 processing of GOME spectral data. AMFs are required for the conversion of effective slant column densities of trace gas absorbers (as produced by the DOAS spectral fitting algorithm) to vertical column densities, which are independent of viewing geometry (solar and line–of–sight). The AMF represents the enhancement of the absorption of a given trace gas along slant paths of transmitted light in the atmosphere (see e.g., [A35]).

As stated before, the two major algorithms are essentially separated in the DOAS retrieval scheme. AMFs require the calculation of absorption paths in the atmosphere – there is no instrumental input other than viewing geometry extracted from the geolocation information, and the AMF computation is therefore a *pure radiative transfer simulation*.

Following the definition of the AMF in section 5.2, single scattering radiative transfer theory is presented in section 5.3. Sections 5.4 and 5.5 summarise the ray tracing formalism and the atmospheric scattering, respectively. *A priori* climatological and reference data requirements are discussed in the text, and summarised in section 5.6 (see also section 6.3 and reference document [A3]).

Note, that the operational system calculates only a single scattering AMF using online RT simulations while the multiple scattering contribution is added afterwards uisng LUT of multiple scattering correction factors. More information is given in section 5.9.

The separation of DOAS spectral fitting and AMF computation is very convenient in the DOAS retrieval of vertical column amounts, but it disguises a central conceptual problem, namely, that in order to retrieve a column of ozone, it is necessary to know the *a priori* (climatological) profile of ozone accurately. The rationale behind the DOAS approach is the approximation that the AMF is insensitive to absolute amounts of the trace gas in question, and it is only the shape of the profile which is really significant in the AMF computation.

5.2 Definition of the Air Mass Factor

The AMF for trace gas *g* requires calculation of the optical densities $\tau_{slant}(g)$ and $\tau_{vert}(g)$ along the slant and vertical paths respectively. The relation to slant and vertical column amounts (S_{slant} and S_{vert}) is given by :

$$AMF(g) = \frac{\tau_{slant}(g)}{\tau_{vert}(g)} = \frac{S_{slant}(g)}{S_{vert}(g)} \tag{27}$$

where $\tau_{slant}(g)$ and $\tau_{vert}(g)$ originate from AMF radiative transfer computation (Eq. (28)) and Eq. (29) given below. The conversion to vertical columns $S_{vert}(g)$ is then just the division of the slant column amount $S_{slant}(g)$ by the appropriate Air Mass Factor AMF(g).

The slant density τ_{slant} is given from Beer's Law :

$$\tau_{slant}(g) = ln\,(R_{2g}) - ln\,(R_1) \tag{28}$$

where R_{2g} is the back–scattered reflectivity including all absorptions *except* the trace gas g of interest, and R_1 is the corresponding reflectivity including *all* absorptions. Here, reflectivity means the ratio of the back–scattered irradiance to the incoming solar irradiance (the absolute value of the solar flux is not needed).

The vertical optical density is given by the integral over atmospheric height of the vertical concentration profile $C_g(z)$ and the cross–section $\sigma_g(z, \lambda)$:

$$\tau_{vert}(g) = \int_0^{z_0} \sigma_g(z,\lambda)\, C_g(z)\, dz \tag{29}$$

It is particularly important to make a precise calculation of the AMF for long paths (solar zenith angles above 80°). For a *plane–parallel non–refracting atmosphere of relatively low optical depth*, the Air Mass Factor is well approximated by the geometrical factor (θ_0 and θ_1 are the solar and line–of–sight zenith angles) :

$$AMF_{geoemtric} = sec\,\theta_0 + sec\,\theta_1 \tag{30}$$

Note, that the application of geometric AMFs is also not recommended if there is a remarkable concentration of the trace gas of interest in the lower troposphere. This is especially true for e.g., biomass burning scenarios (HCHO, NO_2) and other events where an enhanced tropospheric loading of trace gases may occur (e.g. enhanced NO_2, SO_2 loading over cities, industrialized regions).

5.3 Single Scattering Algorithm

For a *local* solar zenith angle θ at a given atmospheric height z, we may define the total optical density of the attenuation along the (slant) light path to that height as :

$$\tau_{total}(\theta, z) = \tau_{Rayleigh}(\theta, z) + \tau_{Aerosol}(\theta, z) + \tau_{gases}(\theta, z) \tag{31}$$

where :

$$\tau_{gases}(\theta, z) = \int \sum_j C_j(z)\, \sigma_j(z)\, ds \tag{32}$$

for concentrations C_j and cross–sections σ_j for trace gas absorber j, and $_{Rayleigh}$, $_{Aerosol}$ the contributions of molecular and aerosol extinction to the total optical density.

Ray tracing in a spherically curved refracting atmosphere is used to determine the integrations over path length. The transmittances required for the Ozone Air Mass Factor are :

$$T_1(\theta, z) = exp\,(-\tau_{total}(\theta, z)) \tag{33}$$

$$T_2(\theta, z) = exp\,(-[\,\tau_{total}(\theta, z) - \tau_{gas}(\theta, z)\,]) \tag{34}$$

for all absorbers and for all absorbers *except* the trace gas of interest, respectively.

The phase function $P(\Theta, z)$ for single scattering is a weighted sum of contributions from particulates (Mie–type aerosols) and those from molecular (Rayleigh) scatterers. Here, Θ is the angle of scatter between the line–of–sight and the solar path (see Eq. (45) below).

$$P(\Theta, z) = \frac{P_{rayleigh}(\Theta)\,\tau_{Rayleigh}(\theta, z) + P_{Aerosol}(\Theta)\,\tau_{Aerosol}(\theta, z)}{\tau_{Rayleigh}(\theta, z) + \tau_{Aerosol}(\theta, z)} \tag{35}$$

Single scattered flux contributions will be further attenuated along the line of sight by factors :

$$U_1(\theta_1, z) = exp\,(-\tau_{total}(\theta_1, z)) \tag{36}$$

$$U_2(\theta_1, z) = exp\left(-[\tau_{total}(\theta_1, z) - \tau_{gas}(\theta_1, z)]\right) \quad (37)$$

defined analogously to T_1 and T_2 in Eqs. (33) and (34). Here, θ_1 is the local line–of–sight zenith angle.

For satellite viewing, we must also consider rays of sunlight singly scattered downwards to the lower boundary (ground or tropospheric reflecting cloud), and then reflected back into the line of sight direction. In addition, direct flux sunlight may be reflected off the lower boundary surface, then singly scattered into the line of sight. A direct reflection into the line of sight is also present. If the lower boundary surface is Lambertian (uniform reflector) then the flux boundary condition is :

$$F_{reflected}(\theta, z_s) = F_{incident}(\theta, z_s) \cdot \left(\frac{A}{2\pi}\right) \quad (38)$$

for surface height z_s and Lambertian reflectance A.

The total back–scattered intensity from all scattering layers (a discretisation of the altitude coordinate z is assumed) is given by :

$$I_1 = \sum_z T_1(\theta, z) \cdot P(\Theta, z) \cdot U_1(\theta_1, z) \quad (39)$$

$$I_2 = \sum_z T_2(\theta, z) \cdot P(\Theta, z) \cdot U_2(\theta_1, z) \quad (40)$$

It is these quantities that will be used in the definition of the slant optical densities (Eq. (28) above).

5.4 Ray Tracing

Parallel rays of sunlight entering the atmosphere must be traced through the atmosphere to allow for the geometrical effects of the earth's curvature and the refraction of light due to varying air density. A number of equally separated parallel rays are traced through an atmosphere with equally separated vertical height layers (1 km thickness, typically 70 layers between 0 and 70 km).

For the single–scattering AMF computation, rays are assumed to start at the appropriate local solar zenith angles at the top of the atmosphere and reach the nadir, where the local solar zenith is that given at the subsatellite position. The algorithm is designed to calculate AMFs for a number of input solar zenith angles, and in this case, rays begin at those angles reaching a nadir close to the lowest given value, and are then traced through the atmosphere over a number of local nadir values separated by angles of 0.5°, until the largest input solar zenith is covered.

The relevant formulae for a refracting spherically–curved atmosphere are :

$$sin(\theta_r) = N(H_{scale}, \lambda)\, sin(\theta_{ir}) \qquad (Snell's\ law\ of\ refraction) \quad (41)$$

$$sin(\gamma) = \frac{(R_1 + R_e)}{(R_2 + R_e)}\, sin(\theta_{ir}) \quad (42)$$

$$p^2 = R_1^2 + R_1^2 - 2R_1R_2cos(\gamma) \quad (43)$$

where R_e is the earth radius, R_1 and R_2 heights of successive layers, with incident and reflected zenith angles θ_i and θ_r, earth centred angle γ and path length p. The refractive index of air $N(H_{scale},\lambda)$ depends on the temperature T(z) and pressure P(z) through the scale height H_{scale}, and weakly on the wavelength. [A standard reference formula (e.g. that due to Edlen) is used].

There is no direct tracing of rays from the top of the atmosphere to a local nadir with zenith greater than 90° – tracing is only then possible to a given level above the surface (the shadow height, below which the atmosphere is in twilight). In addition, incoming rays with zeniths greater than 90° only penetrate to a certain depth (the tangent height), which must be found by iteration.

5.5 Molecular and Particulate Scattering and Extinction

Rayleigh scattering

The molecular (Rayleigh) scattering phase function $P_{rayleigh}(\Theta)$ including polarization is given by :

$$P_{Rayleigh}(\Theta) = \frac{3}{2(2+\delta)} \cdot (1+\delta+(1-\delta)\cos^2\Theta) \quad (44)$$

where δ is the depolarisation factor and the scattering angle Θ is expressed in terms of the solar and line–of–sight zeniths θ and θ_1 and the relative azimuth $(\phi–\phi_1)$ between the planes containing the solar and line–of–sight rays, through the spherical geometry formula :

$$\cos\Theta = \cos\theta \,.\, \cos\theta_1 + \sin\theta \,.\, \sin\theta_1 \,.\, \cos(\phi–\phi_1) \quad (45)$$

If unpolarized radiation ($\delta = 0$) is assumed, Eq. (44) becomes the well–known

$$P_{Rayleigh}(\Theta) = \frac{3}{4}(1+\cos^2\Theta) \quad (46)$$

Rayleigh scattering at height z is calculated from :

$$e_{Rayleigh}(\lambda, z) = \varrho_{air}(z) . \sigma_{rayleigh}(\lambda) \quad (47)$$

for air density $\varrho_{air}(z)$ and scattering coefficient $\sigma_{rayleigh}(\lambda)$. The formula for $\sigma_{rayleigh}(\lambda)$ is taken from Brassseur and Solomon [A15]:

$$\sigma_{Rayleigh}(\lambda) = 3.93 \,.\, \frac{C}{\lambda^r} \quad (48)$$

where C and r are defined as follows :

$$C = \frac{6+3\,.\,\delta}{6-7\,.\,\delta} \quad where\ \delta = 0.0295\ (depolarization\ factor) \quad (49)$$

$$r = H1 + H2\,.\,\lambda + \frac{H3}{\lambda} \quad where\ H1 = 3.916,\ H2 = 0.074,\ H3 = 0.05 \quad (50)$$

The wavelength λ is given in metres. The depolarization factor can be specified in the initialisation file. Note, that this formula is clearly a refinement of the standard λ^{-4} law for Rayleigh scattering).

Aerosols

For aerosols, the scattering properties of spherical particles are assumed (Mie aerosols). A data base of such optical properties (scattering, absorption, extinction coefficients) is required, and this has been taken from the LOWTRAN7 Mie data base. There are two phase function types available: Henyey–Greenstein (HG) and Mie–type phase functions. The operational default is to use HG–type aerosol phase functions.

An extinction profile at 550 nm is first specified, and scattering and extinction coefficients at other wavelengths are determined from a table of normalisation factors. The aerosol optical thickness (AOT) at 550 nm is then calculated as the integrated extinction over height. The atmosphere is divid-

ed into 4 regimes [boundary layer 0–2 km, free troposphere 2–10 km, stratosphere 10–30 km, and mesosphere (30 km upwards)], and there are options for the boundary layer aerosol types (continental, maritime, urban, etc.) and the stratospheric aerosols (background, fresh volcanic, etc.). Phase functions are taken directly from an associated table, classified in the same fashion. Application of aerosol models in the planetary boundary layer (i.e., "rural" or "maritime") depends on the underlying ground. A vegetation index data base provides (besides others) a land/sea mask which is used to switch between the aerosol models. A standard visibility of 23 km in the boundary layer is applied but over sandy surfaces (e.g. Saharian desert, Gobi desert) the horizontal visibility is decreased to 5 km while the visibility is increased to 50 km over snow–covered scenes. A higher visibility goes along with lower aersosol loading and vice versa.

Clouds

There are two treatments of clouds in the AMF algorithm. The first assumes that tropospheric water droplet clouds are sufficiently optically thick to act as reflecting boundaries – in this case, the cloud top height (or pressure) is regarded as the lowest level in the altitude grid, and a bi–directional reflecting boundary condition then applies. A data base of cloud–top bi–directional reflectances including escape function correction terms has been specially prepared for the GOME applications, and this is described in more detail chapter 2, section 3.

Clouds can also be treated as layers of particulate scatterers. Tables of optical properties must be generated in a data base so that the extinction and scattering coefficients and the phase functions can be returned from these look–up tables. Clouds are then treated as for aerosols – the AMF model requires an extinction profile and phase functions at those heights where cloud is present. For water droplet clouds, a Mie scattering programme was used to generate these tables of cloud optical properties.

It should be noted that the option to use clouds as layers of particulates will not be used in the routine operational running of the AMF algorithm. Instead, clouds will be treated in the reflecting approximation as bi–directionally reflecting lower boundaries, and AMFs calculated down to cloud–top. The atmosphere above such cloud boundaries will be assumed clear. The reason for this is that stand–alone GOME measurements cannot supply enough information on clouds to make a meaningful simulation of radiation within cloud.

Operationally, the cloud–top height will be supplied from the ICFA algorithm as a fixed parameter input – this height will serve as the lower boundary in the AMF calculation to cloud–top. The cloud type and the cloud optical thickness however are given in the static parameter input file of GDP. Here, an operational default of 20 is used for the cloud optical thickness and the cloud type is STRATUS II, following the scheme presented in [A38].

5.6 Reference and Climatological Data

For the computation of trace gas optical densities, suitable cross–section data are required. For ozone in the Huggins bands wavelength range, a profile of cross–sections must be constructed incorporating the temperature dependence of the cross–sections (the quadratic Bass–Paur representation). High resolution H_2O cross–sections computed from molecular parameters should be convoluted with an appropriate slit function, when H_2O is a contributing species in the RT simulation.

The Rayleigh optical properties are calculated explicitly (see Eqs. (44) and (48) for example). As mentioned above, aerosol data is taken from the LOWTRAN7 data set. For clouds as particulate layers, a data set of optical properties (scattering and extinction parameters, asymmetry parameters, phase functions) was created for the GOME application. This includes data for eight water droplet cloud types, and for two high cloud (ice crystal) types.

For the earth's surface as the lower atmospheric boundary, the ground height and surface reflectance are required. Operationally, these are extracted from global look–up tables of topography and Lambertian total albedo, dependent on geographical location. Options exist to invoke wavelength dependence of the surface reflection, and the implementation of a "glitter" albedo for specular sea surfaces. Another data set has been created for the extraction of cloud–top bi–directional reflectances, classified according to incident and reflected zenith angles, relative azimuths for the eight standard water–droplet cloud types (see section 3.4 for a discussion of this data base).

The profile data base required for AMF comprises pressure, temperature and trace gas concentrations classified by season and latitude zone; at present (GDP 2.0) these are from two sources (MPI model climatology, US standard atmospheres dataset). In test mode, options exist for the input of user–defined special profiles (ozone hole, Gaussian profiles, etc.).

5.7 Summary of Algorithm Processing

The AMF software in GOME Level 1–2 processing has been adapted from a stand–alone research model. For verification purposes, the AMF algorithm is capable of dealing with ground–based as well as satellite viewing conditions. The test environment has been constructed so that extensive tables of AMFs can be generated, should the operational need for look–up tables prove necessary.

It should also be noted that several options not used in the operational running of the AMF algorithm can be switched on in an off–line test environment. These include the treatment of clouds as particulate layers, and the correction of cloud–top reflectances for finite optical depth. Further, the issue of multiple scattering correction tables has not been decided yet (1995).

The wavelength and viewing geometry input parameters will be determined operationally from the choice of DOAS fitting windows (AMF to be calculated at one representative wavelength for each fitting window) and the geolocation information (viewing angles, geographical latitude/longitude, time). In the off-line mode, the model can be run simultaneously for a wide range of input wavelengths and viewing angles.

In operational Level 1–2 processing, the cloud cover fraction will be determined by the ICFA algorithm. When this value is non–zero (partial or total cloud), then the AMF algorithm will be executed twice, once to the ground and once to cloud–top (see also section 6.2).

- *Input requirements*
  - Number and values of wavelengths for calculation.
  - Numbers and values of line–of–sight zenith angles, range of values of solar zenith angles, number and values of relative azimuth angles.
  - Flags for
    (i) intensity output (diagnostic information)
    (ii) use of multiple scattering formalism
    (iii) use of ray–tracing geometry
    (iv) presence of cloud reflecting boundary
    (v) access to climatological data bases
    (vi) use of Gaussian profile information (test mode only)
    (vii) use of ground–based viewing geometry (test mode only)
    (viii) presence of cloud layers (test mode only)
  - Flags for contributing molecules (for which AMFs will be found)
- *Data base requirements*

- Cross–sections for all trace gases
- Temperature, pressure and concentration profiles
- Aerosol and cloud optical properties
- Cloud–top reflectance data
- Surface reflectance data, global topography data set

- *Summary of algorithm steps (operational mode)*
  - Extract parameter information from file (this would not normally be changed during operational running).
  - Extract geolocation information and convert to required input for the AMF module (pixel location, viewing geometry, time).
  - Perform interface function (extract information from data bases and prepare all climatological inputs).
  - Establish local height grid, compute vertical optical density (Eq. (29)).
  - Establish local geometrical grid, carry out ray tracing to generate field of attenuation factors (Eqs. (33) to (37)).
  - If multiple scattering flag, either extract multiple scattering contributions from look–up tables or calculate these using Monte–Carlo simulation.
  - Compute complete back–scattered intensities into line–of–sight direction.
  - Generate AMFs and intensities (Eqs. (27) and (28)).
- *Output*
  - AMFs for each retrieved trace gas, one for each fitting window wavelength, down to ground and cloud–top.
  - Total back–scattered intensity (both from ground and from reflecting cloud–top), and for each representative fitting window wavelength.

5.8 Open Issues (Issue 1A)

Multiple scattering operational use

A scheme is under consideration to implement look–up tables of Montecarlo–derived multiple scattering correction factors for the AMF intensities. The Montecarlo module wil be retained in the operational software. The default will be the single scattering AMF calculation from scratch.

Re–processing and depleted ozone profiles (see also section 6.5)

If an individual result for the vertical column density for one ground pixel is deemed unsatisfactory, then a better result can be obtained by using a better guess for the *a priori* ozone climatology and re–computing the AMF.

Validation against another model

The AMF radiative transfer model must be validated against another radiative transfer code with the same capability. The GOMETRAN model can be used for this validation, and early results have shown good agreement with the AMF values given by the above algorithm. Any model validating

the operational AMF software must have the ability to handle the climatological conditions and viewing scenarios used in the operational data processing.

5.9 Algorithm Updates

5.9.1 Issue 2/A – GDP 2.0

Multiple scattering operational use

Owing to a set of unfortunate circumstances, it has not proved possible to use the Monte Carlo multiple scattering algorithm originally written for the AMF code. [The Monte Carlo module has been retained in the operational software]. Instead, the AMF module is run in single scattering mode, and a correction is made to the AMF using a multiplicative correction factor interpolated from a look–up table.

The look–up table was created using the forward radiative transfer model GOMETRAN (property of the University of Bremen). AMFs were calculated both in the single scattering and multiple scatter modes, and the correction factors computed as the ratios of these quantities. It was found that the correction factors can be accurately parameterised according to zenith angles and line–of–sight angles. The polar view mode of GOME is excluded in the version presented here.

A double parameterisation scheme was adopted whereby the dependence of the correction factors on the cosine of the solar zenith angle and the dependence on the cosine of the line–of–sight zenith angle are both specified by low–order polynomials. A linear least squares routine (SVDFIT, see chapter 3) was used to determine the polynomial parameter coefficients. Sets of such coefficients are then determined for a number (6) of surface albedos, a number (9) of different ground heights, and a number (14) of different atmospheric profiles reflecting the latitudinal and seasonal variation of atmospheric conditions. The table thus has 3 classifications in addition to the double parameterisation. Albedos were always taken as Lambertian. Two different aerosol types in the boundary layer (rural, maritime) were taken as default aerosols taking into account different aerosol scattering properties over land and over the oceans. 'Rural' was taken for the latitudinal belts in the tropics and the northern midlatitudes, 'maritime' was used for the latitudinal belts in the midlatitudes of the southern hemisphere and both the polar regions. No longitudinal variations of aerosol types is included, however correction factors are less sensitive to different aerosol loading than the AMF itself. The azimuthal dependence was neglected because it is below 2% for most scenarios.

For a given pixel scene, the correction factor is recovered by first using the polynomial coefficients to compute all possible factors for the given pair of zenith angles (SZA and LOS), and then by linear or Akima interpolation over albedo, height and latitude to end up with the correct value. Results have shown that for a given atmospheric scenario, the full multiple scattering AMF at 325 nm is recovered to within 1% for all solar zenith angles up to 92 degrees and all appropriate line–of–sight zeniths.

It should be noted that there are some questions concerning the mixing of results from two very different radiative transfer codes (a single–scatter AMF from the AMF model is multiplied by a correction factor from GOMETRAN output). However, it was seen in the validation that the error introduced by this mismatch remained small (less than 5%). This was checked by comparing the single scatter AMFs from both models. It was recognized that it is more consistent and safer to use one only radiative transfer model, and in principle GOMETRAN has the capability to fulfill this requirement, still using the *ab initio* single scatter computation with a correction factor. Unfortunately, it was not possible to complete this task in time for the public release of Level 2 data.

AMF at 325 nm in UV window

It was seen during the commissioning phase validation that total O_3 columns derived from the 325–335 nm fitting window showed a marked dependence on viewing geometry (solar zenith angle) away from tropical regions, with up to 30% discrepancies observed at solar zeniths around 90 degrees. It was concluded that this was due to the wrong choice of a representative wavelength for the AMF.

It was found necessary to move the representative wavelength to 325 nm, at the lower end of the fitting window range, where O_3 absorption is strongest. For high sun zenith angles the enhancement of ozone absorption along the slant path is no longer described correctly by the Beer's law. Unrealistic high AMFs occur if the representative AMF is calculated at low absorption (e.g. 330nm) but are lower if computed at shorter wavelengths. It couldn't finally be clarified which wavelength is best suited (even the next relative absorption maximum around 322 nm outside the fitting window was discussed) because application of a single AMF cannot account in any way for the large wavelength dependency of AMFs in the wavelength region of interest. However, it was recognized at the June 24/25 1996 meeting that this switch to 325 nm is a work–around solution that could be implemented easily.

Using simulated data, it has been demonstrated (M. Buchwitz and V. Rozanov, University of Bremen) that the approach works sufficiently well, to describe the absorption along the average light path in the fitting window. The following end–to–end test has been carried out: GOMETRAN [A 15] was used to simulate a TOA backscatter spectrum using a known amount of ozone and a standard ozone profile. The spectrum was subsequently feeded into the DOAS kernel and the resulting slant column was converted into a vertical column using again GOMETRAN (and the same ozone profile) to calculate the according AMF. Finally, the input ozone content was compared to the derived vertical content and agreement was found better than 2% up to SZAs of 85 degrees.

Two other solutions were proposed; to move the UV window further towards the visible (where the O_3 cross–section trend is flatter, but the absorption weaker); the second method uses modified cross–sections in the DOAS algorithm :

$$\sigma^{*}_{o3}(\lambda) \; = \sigma_{o3}(\lambda) \; \cdot AMF_{o3}(\lambda) \tag{51}$$

Thus DOAS retrieves an effective Vertical Column. This second approach has bee tried successfully with simulated measurements, but remains unproven with real GOME data. Even more, it requires excessive time–consuming AMF calculations for each wavelength in the fitting window (or, if possible, giant AMF look–up tables) and it is therefore unlikely, that the method will be implemented in the operational chain.

Data Base Changes

(i) For consistency with the DOAS algorithm, the AMF module now has the capability to use GOME FM O_3 and NO_2 cross–sections, plus new literature cross–sections for NO_2 (see [A 3]). The new NO_2 data includes both temperature dependence, so that cross–sections input to the AMF module are calculated for each layer of the model atmosphere (as is the case for O_3).

(ii) A more consistent and fuller use of the surface albedo data sets has now been implemented. Depending on surface type derived from the 'vegetation index' data base and sun zenith angle, the following now holds :

- For snow surface, a constant total Lambertian albedo of 0.75 is assumed.
- For Ocean surface, an appropriate Glitter albedo (depends on wavelength and solar zenith) is extracted from data base.
- For other land surfaces, a spectrally dependent Lambertian albedo appropriate to the given land surface type is assumed.

Re–processing the AMF

When the quality check flag is turned on, and a potential O_3 hole situation flagged, the single scatter AMF computation is repeated using a better guess for the O_3 input profile (see Chapter 6 for details on the generation of this profile). The multiple scatter correction is unchanged – this is not strictly speaking correct (the correction factor used the old profile), but the effect will be second order.

5.9.2 Issue 2/B – GDP 2.7

The AMF look–up table of multiple scattering correction factors has been recalculated using more appropriate reference grids for albedo, ground height and time. There are now four albedo values (0.02, 0.2, 0.5, 0.99), seven height values (0.0, 2.0, 4.0, 6.0, 8.0, 10.0, 12.0 km) and two days of the year (15th January/15th July instead of 15th April/15th Ocotber) reflecting the higher natural variability of albedo, cloud–top height and profile shape (thus correction factor). In addition, the single scattering AMF (and the basic input AMFs for the MS correction factor table) are calculated now for an atmospheric height (TOA) of 70 km (60 km formerly). The erroneous three months shift on the Southern hemisphere in extracted trace gas profiles from the MPI climatology data set has been removed and Mie phase functions were used to simulate aerosol scattering properties. The table has been extended to cover the range of line–of sight zenith angles which occur under polar view mode conditions (40 to 60 deg).

The snow albedo of snow/ice surfaces over land has been changed to 0.95 (formerly 0.75) reflecting the generally higher albedo of fresh snow and the higher albedo of snow–covered surfaces at high latitudes. The necessary input land/sea mask is available from a global vegetation index data base.

5.9.3 Issue 3/A – GDP 3.0

- Rayleigh scattering

The Rayleigh scattering formula has been changed according to ([A19]) which has the advantage of a wavelength–dependent depolarization factor.

$$\sigma_{Rayleigh}(\lambda) \;=\; A \;\cdot \frac{1000}{\lambda^4}\left(1 - \frac{B}{\lambda^2} - \frac{C}{\lambda^4}\right) \textit{with}\ \lambda\ \textit{in}\ [nm] \tag{52}$$

where A = 3.9992662E–04, B = 1.0689770E–02, C = 6.6814090E–05 and the depolarization factor δ as follows :

$$\delta \;=\; \frac{6 \cdot (F - 1)}{3 + 7 \cdot F}\ \textit{and} \tag{53}$$

$$F \;=\; H_1 \;+\; H_2 \cdot \frac{1000}{\lambda^2} \;+\; H_3 \cdot \frac{1000}{\lambda^4}\ \textit{with}\ \lambda\ \textit{in}\ [nm] \tag{54}$$

where H_1 = 1.0469541, H_2 = 3.2503153D–04, H_3 = 3.8622851D–05, respectively.

- AMF retrieval

A completely new algorithm for calculating the AMF (for ozone at 325nm) from a number of geophysical variables has been established. It is based on neural network techniques and is described briefly in the following paragraph. A detailed description of the AMF parameterization with neural networks is given in [A25] and [A26].

A forward backpropagation network is trained using a very large dataset of AMFs covering all possible geopysical scenarios for GOME. Maximum differences between AMFs from radiative transfer

calculations and AMFs retrieved by the network are below 2%, in uppermost cases below 1%. The interpolation and extrapolation capabilities of the neural network were verified against radiative transfer calculations for standard scenarios. Thus, the network computes very accurate AMF values while the processing time and the size of required databases are drastically reduced.

Several data sets that contain ozone AMFs and corresponding variables for a number of geophysical scenarios have been compiled for the training phase of the neural net. More precisely, instead of having one large data base there are single LUT's for three latitude bands (tropic, midlatitude, arctic), two aerosol types (rural, maritime), and two view modes (normal, polar). Thus, a total of 12 LUT's was prepared.

The ozone AMF ensemble used to train the neural networks were calculated from TOMS V7 T–p–conc profiles using LIDORT [A27] taking into account multiple scattering, refraction and the sphericity of the earth's atmosphere. Polarisation was not taken into account in the forward simulations. The layer total content ozone profiles were first outsplined to a fine vertical resolution that is required for radiative transfer simulations. Other input parameters are albedo (0.01, 0.1, 0.3, 0.5, 0.75, 0.98), height above sea level (0.0, 2.0, 4.0, 6.0, 8.0, 10.0, 12.0 km), sun zenith angles (15.0, 30.0, 45.0, 55.0, 65.0, 70.0, 75.0, 80.0, 82.0, 84.0, 86.0, 88.0, 89.99 deg), line–of–sight zenith angles (0.0, 5.0, 10.0, 15.0, 20.0, 25.0, 30.0, 36.0, 40.0, 45.0, 50.0, 55.0, 60.0 deg), azimuths (0.0, 30.0, 60.0, 90.0, 120.0, 150.0, 180.0 deg), and two aerosol modes (maritime, rural using HG phase functions) as given by the Lowtran aerosol scheme. A constant horizontal visibility of 23 km is applied for all AMFs while clouds are not explicitly taken into account. Instead, as in all previous versions of GDP, AMFs to cloud–top are simulated assuming clouds as reflecting boundaries and taking the cloud–top height and the cloud–top reflectance (besides all other variables) as input for the AMF calculation.

Each single AMF table is divided in a training, test and validation dataset. Different perturbed training sets are generated using the bagging technique. The neural networks training was carried out independently for each LUT and results have been stored for the operational processing. Note, that the AMF is finally computed analytically, using the results from the network. No additional LUTs of neither AMFs nor ozone profiles need to be stored.

The actual latitude of the footprint centre, the underlying ground (land or water) and the view mode are used to select and combine the corresponding neural networks.

AMFs for a complete GOME orbit are computed with a feedforward network in a few seconds, i.e. the processing time needed by a neural network is several orders of magnitude lower than the on–line radiative transfer model calculations.

6 Vertical Column Calculation

6.1 Introduction

As has already been noted, the basic formula for the generation of vertical column densities (VCDs) involves the division of the effective slant column amount by the Air Mass Factor for the trace gas in question :

$$VCD \;=\; ESC \;/\; AMF \tag{55}$$

The situation becomes more complicated if clouds are present. It is necessary to distinguish between clear, totally cloudy and partially cloudy pixel scenes. Fractional cloud cover F_c is generated by the ICFA algorithm. The basic algorithm for the generation of VCD in the presence of clouds is described in section 6.2.

A better estimation of the VCD may be obtained by allowing the AMF to vary with viewing geometry across the pixel. This is especially true for low sun (high solar zenith angles, long atmospheric paths). The available geolocation information from extracted GOME Level 1 data allows the calculation of three AMFs across one pixel (three different viewing geometries) and a scheme is outlined in section 6.3 for this extended field–of–view improvement to the VCD algorithm.

Section 6.4 contains the summary of the algorithm processing. Section 6.5 gathers together the various quantities to be included in the GOME Level 2 data product from the three main algorithms and the vertical column computation. Quality control and commissioning phase validation are discussed in section 6.6.

6.2 Vertical Column Density and Cloud Cover

If the fractional cloud cover is non–zero, then it is necessary to calculate two AMFs – one for the clear atmosphere down to the ground (AMF_{clear}), the other down to cloud top H_{ctop} (AMF_{cloudy}). Though there may be two AMFs, there is only one ESC from the DOAS fitting – this latter is representative of the entire ground pixel. We may define a total AMF as the linear combination of AMF_{cloudy} and AMF_{clear} weighted with the fractional cloud cover :

$$AMF_{total} \;=\; F_c\, AMF_{cloudy} \;+\; (1–F_c)\, AMF_{clear} \tag{56}$$

For the cloudy parts of the pixel, the AMF is only known to cloud–top, and we have no knowledge of the vertical column between cloud–top and ground. This unknown quantity is called the *ghost vertical column* (GVC) and must be estimated if the total column is to be produced. Using the proportionality of cloud cover, we can define the total vertical column density as :

$$VCD_{total} \;=\; \frac{\left[ESC \;+\; F_c \cdot GVC \cdot AMF_{cloudy}\right]}{AMF_{total}} \tag{57}$$

and the corresponding quantity down to the cloud–top as :

$$VCD_{cloudtop} \;=\; \frac{\left[ESC \;-\; (1 \;-\; F_c) \cdot GVC \cdot AMF_{clear}\right]}{AMF_{total}} \tag{58}$$

Note, that the contribution under the clouds (f_c x GVC x AMF_{cloudy}) actually denotes a virtual slant column. This is of course an approximation for the total slant column density because the estimated slant column ESC contains already information about the absorption in layers below the cloud top, if

the cloud coverage is less than 1. Other formulas have been tried with less success and Eq. (57) still holds in the operational chain.

In the case of totally cloudy scene ($F_c = 1$) and the clear scenario ($F_c = 0$), Eq. (57) reduces to the following :

$$VCD_{total} = GVC + \frac{ESC}{AMF_{cloudy}} \quad \text{(total cloud cover)} \tag{59}$$

$$VCD_{total} = \frac{ESC}{AMF_{clear}} \quad \text{(no cloud; clear)} \tag{60}$$

Ghost column computation

The ghost vertical column can be estimated from climatology; a suitable trace gas concentration profile C(z) is taken (appropriate for given season of year and geographical position) and integrated over height :

$$GVC = \int_{H_{ground}}^{H_{ctop}} C(z)\, dz \tag{61}$$

This may be unsatisfactory, because the tropospheric burden of Ozone could be vastly different from the climatological ghost column. However, this definition is used for the computation of the GVC.

It would be better to estimate GVC using quantities estimated from the DOAS retrieval. When two effective slant columns ESC(1) and ESC(2) are available from two DOAS fitting windows well separated in wavelength (for example 323–325 nm and 430–470 nm for ozone), then the GVC may be estimated by :

$$GVC = \frac{ESC(2)}{AMF_{clear}(2)} - \frac{ESC(1)}{AMF_{clear}(1)} \tag{62}$$

This formula is only provisional at the present stage and due to instrumental problems, the required successful fitting of ozone in the VIS spectral range was never achieved.

6.3 Extended Field–of–View Calculation

The use of a mean AMF for the whole pixel (implied in the above calculation) may be inaccurate in some cases (especially long paths). The AMF algorithm will return individual Air Mass Factors for a number of independent sets of viewing geometries – three such sets can be derived from the Level 1 extracted geolocation information. The formulae below apply to any trace gas.

Assume that we have now three spectra integrated over three parts of the pixel, and for each part, the viewing geometry angles are constant; we can then compute three Air Mass Factors $\{AMF_i, i=1,2,3\}$. Each spectrum has its own effective slant column S_i defined by :

$$S_i = \frac{ln\,[R^*_i] - ln\,[R_i]}{\sigma_{gas}} \tag{63}$$

where σ_{gas} is the trace gas cross–section, R^*_i is the back–scattered intensity without the absorption of the trace gas included, and R_i the intensity including the trace gas absorption. Adding up the three spectra gives the real spectrum, and the effective slant column retrieved from the DOAS fitting :

$$ESC = S = \frac{ln\,[R^*] - ln\,[R]}{\sigma_{gas}} \tag{64}$$

$$where \quad R^* = \sum_{i=1}^{3} R^*_i \quad and \quad R = \sum_{i=1}^{3} R_i$$

In order to get a relation between S and S_i using the information implied in the division of 3 sub–pixels, we make two assumptions :

- Optical densities of trace gases are much smaller than 1. This implies :

$$S_i = \frac{1 - (R_i/R^*_i)}{\sigma_{gas}} \quad and \quad S = \frac{1 - (R/R^*)}{\sigma_{gas}} \tag{65}$$

This gives the following relation between S and S_i :

$$S = \frac{\sum_{i=1}^{3} S_i R^*_i}{\sum_{i=1}^{3} R^*_i} \tag{66}$$

- The Vertical Columns are constant over the entire ground pixel. Then :

$$VCD = \frac{S_i}{AMF_i} \quad \text{for each i=1,2,3} \tag{67}$$

Using Eqs. (66) and (67), the relation between the total vertical column for the whole pixel and the total effective slant column S is :

$$VCD = \frac{S}{\sum_{i=1}^{3} Q_i \; AMF_i} \tag{68}$$

where the factors Q_i are given by :

$$Q_i = \frac{R^*_i}{R_i} \tag{69}$$

The intensities R^*_i and R_i are automatically calculated in the AMF algorithm and can easily be passed on to the Vertical Column algorithm along with the corresponding Air Mass Factors. Indeed the intensity R_i is currently produced as a diagnostic from the AMF algorithm – after multiplication with the solar irradiance, R_i can be compared with the GOME back–scatter measurements.

The advantage of Eq. (68) is that it uses ratios of intensities rather than absolute values. Another possibility is to assume that intensities are independent of the viewing geometry across the ground pixel, in which case the factors Q_i are equal to the percentages P_i of the ground pixel cover. The operational default will be Eq. (68).

The above analysis applies to clear ground pixels or totally cloudy pixels. For partially cloudy pixels, we assume further :

- Clouds are equally distributed over the ground pixel with the same cloud–top height.

Then Eq. (57) can be used as before, with the AMFs now replaced by weighted sums as in Eq. (68) .

6.4 Summary of Algorithm Processing

Algorithm Inputs/Outputs

The VCD algorithm inputs are simply the outputs from the three other main Level 1–2 Processing algorithms. From ICFA we get the fractional cloud cover and its error; from DOAS fitting, the slant columns and their errors, for each window and each trace gas. From the AMF algorithm, we get the AMFs themselves for each window and trace gas, down to ground and cloud–top, and the corresponding intensities (total, and without respective trace gas contributions).

The Level 2 product will contain only total vertical columns to the ground, plus their respective errors. In computing the errors on the vertical columns, we have clear contributions from the fitted parameters involved in the VCD computation (ESCs, fractional cover). The status of errors on the AMFs is less clear, and it has been suggested to compute an error by comparing the calculated intensities from the AMF algorithm to equivalent GOME measurements.

Summary of Steps

- If the extended field–of–view calculation is required, compute weights Q_i of intensities and compute weighted vertical density (Eqs. (69) and (68)).
- If total or partial cloud cover, compute fraction–weighted AMFs following Eq. (56).
- If total or partial cloud cover, compute ghost column (Eq. (61)).
- Compute vertical column amounts from Eqs. (57) to (60), respectively.

6.5 Open Issues (Issue 1A)

Ghost column

Both methods remain untried, and the best approximation to the ghost column will emerge during the commissioning phase, when there will be opportunities to validate against ground data.

Extended field–of–view

Remains essentially untried. There is a problem over the partially cloud scenario, as the extended FOV calculation really requires knowledge of where the cloud is present in the pixel. This cannot be found from ICFA – an alternative cloud recognition algorithm (like the threshold algorithm suggested for PMD data) could provide this information. Much testing of the AMF module is needed to establish the FOV requirement; that is, to know when the usual mean–value AMF becomes too unrepresentative to use in the retrieval.

6.6 Algorithm updates

6.6.1 Issue 2/A – GDP 2.0

The GVC is calculated from climatologies, because the second method (using results from different fitting windows) couldn't be used due to instrumental problems in the VIS spectral range. The extended file–of–view options is the operational default and the intensity–weighting scheme is applied. However, the problem of knowing the position of a cloud field inside the GOME footprint remains unsolved. As a work–around solution, a total of six AMFs (three cloud–free and three

cloudy, if clouds are present) is usually calculated and the total AMF in Eq. (50) originates from calculating a representative (intensity–weighted) AMF across the pixel.

6.6.2 Issue 2/B – GDP 2.7

An additional weighting method is introduced using a geometric weighting scheme. Three sets of geometric weights have been defined for the three positions (left, centre, right) of each GOME footprint where the geolocation and the sun and satellite zenith angles and the relative azimuth are known, respectively.

equal weighting (equal weights for each geometric position, i.e. no weighting at all)

linear weighting

parabolic weighting

The geometric weights are defined as follows:

$$Q_{geo} = \sum_{i=1}^{3} Q_{geo,i} \; and \; Q_{geo,i} = 1,1,1 \; or \; Q_{geo,i} = 1,2,1 \; or \; Q_{geo,i} = 1,4,1 \qquad (70)$$

Both the intensity–weighting and the geometric weighting are applied to AMFs in the operational system. The parabolic weighting scheme is used operationally. It is switched on by a keyword given in the initialisation file. Thus, the Q–factors from section 6.3 now read :

$$Q_i = \frac{R^*_i \cdot Q_{geo,i}}{R_i \cdot Q_{geo}} \qquad (71)$$

The parabolic weighting puts more weight on the centre value which is reasonable because the scanning is done with a constant velocity of the scan mirror across the pixel. Therefore, the measured spectrum is an average across the pixel and application of AMFs that have been calculated at the edges of the footprint lead to a higher contribution of the edges to the final result. As expected, especially the results of the backscan pixel benefit from the *parabolic weighting* scheme in the sense that it is now representative for the entire GOME swath width.

6.6.3 Issue 3/A – GDP 3.0

An iterative solution scheme for the total ozone content has been established according to ([A37]). For the first pixel of an orbit an initial total content of 250 DU is assumed and an appropriate AMF is calculated by the neural net, taking into account the actual viewing geometry, geographic location, reflection properties of the underlying reflecting surface which can be either a cloud layer or the ground, the height above sea level of the reflecting surface and the aerosol type. An updated total content is then calculated using Eq. (68), as long as the iteration condition that follows below is fulfilled:

$$\frac{|VCD_{old} - VCD_{new}|}{|VCD_{new}|} < 10^{-4} \qquad (72)$$

Note, that the slant column remains unchanged during the iteration process. A maximum number of iterations is given in the initialisation file, in order to avoid numerical instabilities. If the maximum number of iterations is reached, an error message is generated and the computation is aborted. In this case, only the slant column content will be written to the level 2 product.

Susbsequent pixels use the ozone vertical content calculated for the previous pixel as a starting value, thus reducing remarkably the number of iterations. Typically, there are less than 5 iterations necessary to fulfill the iteration condition.

For GOME ground pixels with more than 1.5 s integration time, a corresponding number of AMFs across the larger footprint is calculated and the vertical content is derived from the average AMF of the pixel.

In GDP V3.0 the ozone ghost column is calculated by the neural network using latitude, cloud–top height, and ozone total column as the basic input. The network first selects the appropriate TOMS profiles as function of latitude and total columns and performs a multidimensional interpolation to retrieve the ghost column up to the given cloud–top height. In fact, the network performs the profile search and the interpolation between several profiles while the ghost column content is still derived using Eq. (61). The NO_2 ghost column is in all GDP versions calculated from Eq. (61).

6.7 Level 2 Product

Certain algorithm results and diagnostics are included in the Level 2 product, along with the vertical columns. Formatting and definitions of the Level 2 product are dealt with in the Interface Specification and Product Specification documents respectively ([A10] and [A11]); here we will just summarise the relevant output.

For VCD, the product includes total vertical columns and their errors for a combination of 7 windows and trace gases (O_3 UV and visible, NO_2 visible as the operational default). Pointers identifying windows and trace gases appear in the Level 2 product header. Total 14 entries.

For ICFA, the cloud–top height and fractional cover will be printed, along with their total errors. Surface pressure will also be specified (5 entries in total). [Cloud–top height is actually an input in the first operational version, so its error will be zero].

For DOAS, we allow the same combination of up to 7 trace gas species and windows This makes for a total of 18 values (7 ESCs and errors, plus RMS, chi–square, iteration number and goodness–of–fit).

For AMF, we get AMF values and their errors for cloudy/clear and the same combination of 7 windows and trace gases (14 values). We also print total intensity for each window, plus representative measured back–scatter for error comparison (4 values, total 18 entries). Note, that due to the limitations of the on–line radiative transfer model the calculated intensities originate from single scattering simulations only. This limits, of cours, its usage for error comparison.

Only field–of–view averaged AMFs are given. Vertical columns to cloud–top can be inferred from the existing ancillary information in the product.

Additional diagnostics not derived from any physical algorithm in the Level 1–2 processing chain will also be specified; these are quantities derived from a purely statistical averaging of PMD and GOME nadir observations. They are : the pixel contrast numbers (mean & standard deviation of 16 sub–pixel PMD values); 16 sub–pixel ”colour” values (ratios PMD_3/PMD_2); and the pixel ”colour gradient” (linear regression gradient of GOME channel 3 measurements in the range 450–600 nm).

6.8 Quality Control and Validation

Discussion

Two approaches to quality control have been put forward. On a general level, it is possible to assign single–number ”success” scores to every retrieved vertical column density, based on the combined total error and perhaps on other diagnostics. However the only satisfactory way to test for retrieval accuracy is to examine patterns of retrieved data, in space and time, and look for areas or time series of identifiably bad results. Once the identification of bad cases has been completed and remedies

worked out off–line, the Level 2 data must then be re–processed in its entirety (as has happened several times with the TOMS algorithm).

In general we may say that individual results from least squares fitting algorithms cannot be re–computed; in DOAS one cannot change the cross–sections and reference spectra for one ground pixel without changing them for all other pixels (i.e. complete re–processing). The same applies to the ICFA results, though here the fitting depends on the assignation of a fixed cloud–top height; this height could be determined by examination of a number of already–minimised merit functions (the so–called chi–square grid searching technique).

Many possibilities exist for re–processing the AMF result, the main problem being to decide which climatological scenario input needs altering to achieve the "correct" result. Identification of such a change is extremely difficult on a short–term basis. Many atmospheric events are unpredictable and on too small a spatial scale. As far as O_3 is concerned, the retrieval of depleted columns in ozone hole scenarios may be possible, and the following suggestion has been recommended for implementation in the Level 1–2 operational algorithm.

6.8.1 Issue 2/B – GDP 2.7

Individual re–processing for potential O_3 hole situations

A potential Ozone hole situation is first flagged according to latitude and time of year. For this purpose some new input parameters are required from the operational static parameter input file :

- Beginning and end of Arctic Spring (days in the year)
- Beginning and end of Antarctic Spring (days in the year)
- Southernmost and Northernmost latitude limits for O3 hole.

The total vertical column result is then examined. If this is below a certain fixed value F (another parameter input, current default 250 DU), and if the possible presence of an O_3 hole has been flagged, then the *single scatter AMF will be re–computed once only with a depleted O_3 profile*. The depleted profile is constructed as follows.

First define a depletion factor D. This is the ratio of the original vertical column result V_0 and the original climatological Ozone column C_0 down to the ground. The depletion is assumed to take place only in the stratosphere between two heights H_1 and H_2.; the original profile is unchanged outside this altitude range.

The lower height H_1 is determined as that height in the original O_3 profile where the decrease in concentration becomes less than 5% between adjacent levels. This corresponds approximately to the tropopause level, where the stratospheric Ozone distribution starts. H_1 varies with the latitudinal variation of the MPI O_3 climatology, and is typically 16–18 km in the tropics and 10–12 km in the polar regions. H_2 is then defined as that height above the main stratospheric Ozone bulge where the concentration equals that at H_1. H_2 is typically 36 km in the tropics, falling to 30–31 km at the poles.

We construct an analytic depleted profile using a form of the generalized distribution function :

$$z \notin [H_1, H_2] \;:\; P^*(z) = P_0(z) \qquad (\textit{original profile outside height interval}) \tag{73}$$

$$z \in [H_1, H_2] \;:\; P^*(z) = \overline{P} + \frac{w_0\, e^{-(z-z_m)\beta}}{\left\{1 + e^{-(z-z_m)\beta}\right\}^2} \quad (\textit{analytic distribution}) \tag{74}$$

where $\overline{P}$, w_0 and β are all parameters characterizing the distribution, and z_m is the height at which the maximum concentration occurs (assumed known). This analytic expression models closely the

stratospheric Ozone bulge. The 3 distribution parameters may be found from the following conditions.

$$P^*(H_1) = P_0(H_1) \quad and \quad P^*(H_2) = P_0(H_2) \quad (continuity\ at\ z = H_1,\ z = H_2) \tag{75}$$

$$\int_{H_1}^{H_2} P^*(z)\,dz = D \cdot \int_{H_1}^{H_2} P_0(z)\,dz \qquad (Depletion\ of\ original\ stratospheric\ column) \tag{76}$$

The single scatter AMF is then recalculated using profile $P^*(z)$. The corresponding vertical column is then computed. If the new total column is greater than the fixed threshold parameter F, then the re–processing has failed and the new result is discarded, and a 'bad quality' flag is set. If the new result is less than F, then it is written to the Level 2 product and replaces the original result. The re–processing is only attempted once.

AMFs are calculated to cloud–top as well, whenever there are partially or totally cloudy pixel scenes. For these AMFs, depletion still takes place in the stratosphere between heights H_1 and H_2, but the depletion factor must be modified slightly because the original climatological column is only down to cloud–top. If this column to cloud–top is denoted by C_1, then the modified depletion factor is given by :

$$1 - D_{cloudtop} = 1 - \left(\frac{C_0}{C_1}\right) \cdot D_{ground} \tag{77}$$

The depleted profile down to cloud–top is computed as above using the continuity conditions and the stratospheric column depletion.

6.8.2 Issue 3/A – GDP 3.0

The updated version 3.0 computes the ozone AMF at 325 nm based on an iterative approach [A37] and a neural network. The basic input for the latter makes use of AMFs that were calculated using the TOMS V7 ozone profile climatology. This climatology contains also ozone profiles under ozone hole conditions. The re–processing loop as described in section 6.8.1 is therefore not used anymore. The ghost vertical column below clouds is now also computed by the neural network (see explanation in section 6.6.3).

7 Summary of Data Base Requirements

7.1 Extracted Level 1 Data

The necessary extracted Level 1 data for the operational Level 1–2 processing algorithms are as follows.

AMF algorithm (geolocation only)
* Solar zenith and azimuth angles at the spacecraft (left/centre/right of ground pixel)
* Line–of–sight zenith and azimuth angles at the spacecraft (left/centre/right of pixel)
* Latitude and longitude of centre of pixel and corner points
* Time values (year, month, day)

ICFA algorithm (some geolocation, GOME channel 4 spectra, PMD data)
* Solar zenith and azimuth angles at the spacecraft (centre of ground pixel)
* Line–of–sight zenith and azimuth angles at the spacecraft (centre of pixel)
* Latitude and longitude of centre of pixel
* GOME channel 4 earthshine and solar spectra (wavelengths, values, errors)
* 16 Subpixel reflectances from each of the 3 PMDs (for Cloud Clearing Algorithm only)

DOAS algorithm (spectra only)
* GOME channel 2 and 3 earthshine spectra (wavelengths, values, errors)
* GOME channel 2 and 3 solar spectra (wavelengths, values, errors)

In general the earthshine and solar spectra will be specified on different wavelength grids. A supplementary interpolation is needed to generate the solar spectra on the wavelength grid of the earthshine measurements. This is necessary for the calculation of ratioed spectra in DOAS and ICFA algorithms.

No extended field–of–view averaging is anticipated for the ICFA algorithm.

7.2 Extracted Level 2 Data

This data set consists of already–processed level 2 data. It has no direct use in any of the algorithms in Level 1–2 Processing, and will be used only for comparison and quality control. A decision on the individual re–processing of certain pixels will be made on the strength of this input (for example, if the level 2 result is significantly different from the nearest *already–known* Level 2 values in space and/or time, then a re–processing of the Air Mass Factor could be performed to improve the result).

Because of limitations on the Level 1–2 dispatcher, it is expected that the data set here will consist of the nearest–in–time complete orbit of processed level 2 data.

7.3 Climatological and Reference Data Sets

AMF algorithm
– Trace gas cross–sections
– Global surface topography data set
– Global total albedo and vegetation index data set
– Glitter albedo data set (sea surfaces only)
– Spectral albedo data set for different surface types
– Global data set of model profiles (temperature, pressure and trace gas concentrations)
– Data set of aerosol particle loadings

– Data set of aerosol optical properties (Mie extinction/scattering, phase functions)
– Data set of cloud–top reflectances

– Data set of multiple scattering correction factors

– Data set of neural net weights for a number of geophysical scenarios (12).

ICFA algorithm
– Data base of template transmittances in the O_2 A–band
– Data set of slit function parameters (for convolution)
– [Pre–convoluted transmittances]
– Topography and surface reflectance data sets (as for AMF)
– Data set of cloud–top reflectances
– Data set of cloud–top pressures (ISCCP)

DOAS algorithm (spectra only)
– Data set of slit function parameters (for convolution)
– Trace gas reference cross–sections (GOME FM O_3 and NO_2, other literature spectra)
– Other reference spectra (GOME FM Ring, theoretical Ring, undersampling correction)

Vertical Column
– Global data set of model profiles (temperature, pressure and trace gas concentrations)

Details of these data sets and their classifications can be found in document [A3]. The data set requirements listed above apply to the *stand–alone operation of the individual algorithms*, without reference to other algorithms. The complete Level 1–2 operational software will of course involve all algorithms in sequence, and the extraction of certain data need then only be done once (for example with the global topography data set, which is required for both ICFA and AMF).