GOME/ERS-2

GDP 5.0

Upgrade of the GOME Data Processor for Improved Total Ozone Columns

ALGORITHM THEORETICAL BASIS DOCUMENT

Prepared by	R. Spurr
Contributors	M. van Roozendael, D. Loyola, C. Lerot, J. van Geffen, J. van Gent, C. Fayt, J.-C. Lambert, W. Zimmer, A. Doicu, S. Otto, D. Balis, M. Koukouli, C. Zehner
Project Coordinator	Diego Loyola - DLR/IMF
Technical Officer	Claus Zehner – ESA/ESRIN
Document	DLR/GOME/ATBD/GDP5
Iss./Rev	1B
Date	August 9th, 2012

DLR | BIRA-IASB | RT SOLUTIONS | AUTH

Distribution

Organisation	**Name**
ESA	C. Zehner, T. Fehr, A. Dehn
DLR	D. Loyola, W. Zimmer, A. Doicu
BIRA-IASB	M. van Roozendael, C. Fayt, J. van Gent, C. Lerot, J.-C. Lambert
AUTH	D. Balis, M. Koukouli
RT Solutions Inc.	R. Spurr

DOCUMENT CHANGE LOG

Issue	Rev.	Date	Pages	Changes	Status
1	0	02.10.09	all	Initial issue	Draft
1	0	14.12.09	all	Revision of Initial Issue	Draft
1	A	31.10.11	all	First Version	Draft
1	B	09.08.12	all	Final Issue	Final

Summary

This document is the *streamlined* ATBD for the GDP5 (Version 5 of the GOME Data Processor) Project. This project was instigated at the end of the GODFIT (GOme Direct FITting) Phase B project in May 2007. GDP5 uses the GODFIT direct fitting algorithm for retrieval of total ozone column amounts – a departure from all earlier GDP total column retrievals, which were based on the DOAS concept.

The initial Phase A of the GODFIT project was completed in January 2004 following an ESA review of three new total ozone algorithms funded under the AO/1-4235/02/I-LG ITT. GODFIT Phase A also encompassed a new improved DOAS-style algorithm called GDOAS, which became the basis for the operational GOME Data Processor (GDP) Version 4.0 algorithm in 2004. Phase B of GODFIT was started in March 2005, and a number of improvements were reported in the Phase B ATBD issued in summer 2007. The GDP5 ATBD is based in part on this Phase B documentation. The GDP5 project started soon after GODFIT Phase B was completed, and in its first phase, there were a number of new developments (in particular the practical implementation of a T-shifting scheme and the test of the NNORSY O_3 climatology). The first version of the GDP5 algorithm was ready in summer 2009 following the decision to stick with the TOMS Version 8 O_3 profile climatology.

Following the Introduction (section 1), an overview of the algorithm is presented in section 2. Section 3 describes the forward model set-up, i.e. the choice (TOMS Version 8 vs. NNORSY) and usage of O_3 profile climatology, the use of a shift parameter to adjust the temperature profile, the use of albedo closure terms, and the derivation of trace gas cross-sections. Section 4 deals with the radiative transfer part of the forward modeling, i.e. the generation of optical properties and the simulation of earthshine radiances and analytic Jacobians using the LIDORT model. Section 5 describes the inverse model, in GDP5 a variation of the Levenberg-Marquardt method. Section 6 deals with ancillary algorithms – application of a semi-empirical molecular Ring correction developed for GODFIT, and the derivation and implementation of cloud parameters using the OCRA/ROCINN algorithms. Section 7 contains notes on practical aspects of the GDP5 algorithm, including the UPAS implementation and performance issues, plus configuration and state vector settings.

The present document is a condensation of the internal full ATBD document for GDP5 – the latter contains a lot of additional material which is not strictly part of the final algorithm.

TABLE OF CONTENTS

1. INTRODUCTION 9

1.1. HISTORICAL OVERVIEW; OZONE INSTRUMENTS 9
1.2. BACKGROUND TO GDP AND THE GODFIT ALGORITHM 9
1.3. GDP5 PROJECT OVERVIEW; ATBD SCOPE 10

2. OVERVIEW OF THE GODFIT ALGORITHM 12

3. FORWARD MODEL SETUPS 15

3.1. OZONE PROFILES 15
3.1.1. Introductory remarks 15
3.1.2. The TOMS V8 ozone climatology 15
3.1.3. Ozone profiles: the NNORSY data 16
3.1.4. NNORSY vs. TOMS V8 comparisons 18
3.2. PRESSURE AND TEMPERATURE; THE T-SHIFT PROCEDURE 20
3.2.1. P/T data sets: TOMS P/T baseline and ECMWF 20
3.2.2. The temperature shifting algorithm 21
3.2.3. Retrievals using the T-shift in GODFIT 24
3.3. TRACE GAS CROSS-SECTIONS 26
3.3.1. Ozone cross-section data sets 26
3.3.2. Ozone cross-sections: Sensitivity studies 26
3.4. ALBEDO CLOSURE 28
3.4.1. Internal closure 28
3.4.2. External closure 29
3.4.3. Albedo data sets 29
3.5. OTHER ATMOSPHERIC CONSTITUENTS 30

4. FORWARD MODEL RADIATIVE TRANSFER 32

4.1. FORWARD MODEL: OPTICAL PROPERTY SETUPS 32
4.1.1. Atmospheric optical property inputs 32
4.1.2. Surface and cloud setups; the IPA and CRB 33
4.2. FORWARD MODEL: LIDORT SCATTERING CODE 34
4.2.1. Introductory remarks on LIDORT and VLIDORT 34
4.2.2. Solving the RTE in LIDORT 35
4.2.3. Jacobian output from LIDORT 36
4.2.4. Sphericity corrections 37

5. INVERSE MODEL 39

5.1. GODFIT PHASES A AND B: OPTIMAL ESTIMATION 39
5.2. GDP5: LEVENBERG-MARQUARDT AND LINE-SEARCH 40
5.3. AVERAGING KERNEL 42

6. ANCILLARY ALGORITHMS 44

6.1. THE MOLECULAR RING EFFECT 44

6.1.1. Introduction *44*
6.1.2. Empirical Ring correction in GODFIT *44*
6.1.3. The LIDORT-RRS model and RRS correction factors *46*
6.2. CLOUD PRE-PROCESSING 48
6.2.1. Introductory remarks *48*
6.2.2. OCRA algorithm *49*
6.2.3. ROCINN algorithm *49*
6.2.4. Validation of OCRA and ROCINN algorithm *51*
6.2.5. Total ozone sensitivity to cloud inputs *52*
7. THE GDP5 GODFIT ALGORITHM 56
7.1. RETRIEVAL STATE VECTOR FOR GODFIT IN GDP5 56
7.1.1. Baseline sets of fitting parameters *56*
7.1.2. Earthshine shift *57*
7.2. CONFIGURATION SETTINGS FOR GDP5 57
7.2.1. Forward model settings *57*
7.2.2. Inverse model configurations *58*
7.2.3. Re-sampling issues *58*
7.3. THE GODFIT SOFTWARE PACKAGE 60
7.3.1. Overall capability *60*
7.3.2. Performance issues *60*
7.3.3. Algorithm verification *62*
7.4. THE UPAS INSTALLATION 62
7.4.1. UPAS implementation issues *62*
7.4.2. Level 1 issues *64*
REFERENCES 65

ACRONYMS

AMF	Air Mass Factor
AUTH	Aristotle University of Thessaloniki
BIRA-IASB	Belgian Institute for Space Aeronomy
BRDF	Bi-directional Reflectance Distribution Function
CHEOPS	Climatology of Height-resolved Earth Ozone and Profiling Systems
DLR	German Aerospace Center (Deutsches zentrum für Luft- und Raumfahrt)
DM	Daumont-Malicet
DOAS	Differential Optical Absorption Spectroscopy
DU	Dobson Unit
ECMWF	European Center for Medium-range Weather Forecasting
ENVISAT	Environmental Satellite
EO	Earth Observation
ERS-2	European Remote Sensing Satellite-2
EOS-AURA	(NASA's) Earth Observing System Aura
ESA	European Space Agency
ESRIN	European Space Research Institute
FD	Finite Difference
GAW	Global Atmospheric Watch
GDP	GOME Data Processor
GODFIT	GOme Direct FITting
GOME	Global Ozone Monitoring Experiment
IPA	Independent Pixel Approximation
LIDORT	LInearized Discrete Ordinate Radiative Transfer
LUT	Look Up Table
METOP	Meteorological Operational
MSG	Meteosat Second Generation
NDACC	Network for the Detection of Atmospheric Composition Change
NDSC	Network for the Detection of Stratospheric Change
NNORSY	Neural Network Ozone Retrieval System
OMI	Ozone Monitoring Instrument
PMD	Polarization Measurement Device
RRS	Rotational Raman Scattering
RT	Radiative Transfer
SAO	Smithsonian Astrophysical Observatory
SAOZ	Système d'Analyse par Observation Zénithale
SCIAMACHY	SCanning Imaging Absorption spectroMeter for Atmospheric CartograpHY

SGP	SCIAMACHY Ground Processor
SZA	Solar Zenith Angle
TOA	Top of Atmosphere
TOMS	Total Ozone Mapping Spectrometer
UPAS	Universal Processor for UV/VIS Atmospheric Spectrometers
UV	Ultra Violet
WMO	World Meteorological Office
WOUDC	World Ozone and Ultraviolet Data Center

1. INTRODUCTION

1.1. Historical overview; ozone instruments

It is vital to maintain accurate global ozone records for trend analysis and climate change studies. The program for long-term global monitoring of total ozone amounts from satellite-borne UV spectrometers is now entering a new phase following the launch of second-generation remote sensing instruments and the development of more accurate total ozone retrieval algorithms.

The GOME (launched on board the ERS-2 satellite [*ESA*, 1995] in April 1995) was in operations for 16 years (ERS-2 decommissioning in July 2011). Data has been sporadic since 2003 due to problems with tape storage. The previous official total ozone product was generated via GDP Version 4.x, using a DOAS-style algorithm. See [*van Roozendael et al.*, 2006] for an algorithm description, and [*Balis et al*., 2007] for the 10-year validation.

The TOMS total ozone record dates back to 1978, but has now ceased. In December 2006, contact was lost with the final instrument (EP TOMS) and the project has wound down. At NASA, the new-generation ozone sensor is the joint USA/Dutch/Finnish OMI (Ozone Monitoring Instrument) [*Stammes et al*., 1999, *Levelt et al*., 2006]. Launched on the EOS-AURA platform in July 2004, OMI has been operating successfully for 8 years.

In Europe, SCIAMACHY was launched in March 2002 on the ENVISAT platform (for a mission overview, see [*Bovensmann et al*., 1999]) and it provided 10 years of data until the contact with the satellite was lost in April 2012. The current official ozone product is based on the SDOAS algorithm; see [*Lerot et al.*, 2009].

The GOME-2 instrument was launched successfully in October 2006 on board the first METOP satellite [*Munro et al*., 2006]. The instrument is working nominally after 6 years of operation, and O_3 total columns have been processed routinely using the GDP 4.x algorithm; see [*Loyola et al*., 2011].

1.2. Background to GDP and the GODFIT algorithm

The main operational Level 2 product from GDP (GOME Data Processor [*Loyola et al*., 1997]) is the global distribution of total vertical column amounts of ozone. Ozone column retrieval for GOME has (until now) been done using DOAS-type algorithms, comprising spectral fitting for effective slant columns followed by AMF computations (chiefly from look-up tables) for conversions to vertical column amounts. Version 3.0 of the GDP algorithm was validated in 2002 and written up in 2004 [*Spurr et al.*, 2005]. Partly as a result of this validation [*Lambert et al.*, 2002] and partly from an error analysis [*van Roozendael et al*., 2002], it was recognized that a number of issues needed to be addressed in order to obtain an improved record of total ozone.

In 2002, ESA sent out an invitation to tender (ITT) for a study on improved ozone algorithms for the GOME instrument. Three consortia were funded and all three studies came up with improved algorithms that were subsequently used to reprocess the entire 10-year GOME ozone record from July 1995. Off-line records were generated using DOAS-style algorithms by the University of Bremen Group [*Coldewey-Egbers et al.*, 2005; *Weber et al.,* 2005] and by the KNMI Group [*Eskes et al*., 2005].

The third group, led by Michel van Roozendael (BIRA-IASB) and R. Spurr (then at SAO), developed a new *direct fitting* algorithm called GODFIT (GOme Direct FITting). The GODFIT Phase A also contained an improved DOAS-style algorithm called GDOAS. The ATBD for this part of the project is found in [R1].

In 2004, GDOAS was selected to become the GDP (GOME Data Processor) Version 4.0 algorithm. GDP 4 has achieved accuracy at the "percentage level" compared to ground-based measurements. The 16-year GOME total ozone record has been re-processed with GDP 4.4. For a description of the GDP Version 4.0 GDOAS algorithm, see [R3] and [*van Roozendael et al.*, 2006], and for the 10-year validation, see [*Balis et al.*, 2007].

Direct fitting is a new-concept algorithm for total column retrieval based on a direct least squares inversion of radiances using on-the-fly radiative transfer simulations of both radiances and total column sensitivity functions. The heritage can be traced back to the analysis of ozone column measurements obtained from the continuous scan NIMBUS-7 data over the period 1979 to 1986 [*Joiner and Bhartia*, 1997]. Although not selected for the GDP 4.0 reprocessing, the GODFIT algorithm was recommended for phase-B funding by the December 2003 ESA-convened ITT committee.

The Phase B work was inaugurated in March 2005, and completed in May 2007. Particular attention was paid to high-latitude total ozone retrieval; Phase B was complemented by work done under the SAUNA validation exercise in spring 2006. A number of improvements and upgrades were made to the basic GODFIT algorithm, and a more thorough error analysis performed. For a description of this part of the project, see the ATBD [R2]. At the end of Phase B, GODFIT was selected for the next GDP upgrade to Version 5.0.

A proposal to ESA for the GDP5 work was accepted in summer 2007, and this project is based on the implementation of GODFIT into UPAS and its installation therein. This phase also includes the validation of GDP 5.0 and the complete reprocessing of the entire GOME total ozone record. The joint consortium is led by D. Loyola of DLR, and features scientists from BIRA-IASB (M. van Roozendael, C. Lerot, J. van Geffen, J. van Gent, C. Fayt, J.-C. Lambert, J. Granville), RT Solutions Inc. (R. Spurr), DLR (D. Loyola, W. Zimmer, A. Doicu) and AUTH (M. Koukouli, D. Balis). C. Zehner is the ESA project manager.

The present document is the ATBD for GODFIT in GDP 5.0; it is based in part on the GODFIT Phase B ATBD [R2]. The first draft was prepared in 2009 and the first issue was ready in 2011.

1.3. GDP5 Project Overview; ATBD Scope

The GDP5 project had three phases. Phase 1 was concerned with extensions of the GODFIT Phase B algorithm to deliver all options necessary for the operational implementation of GDP5 in the UPAS system. This has included as a new feature the T-shift implementation and testing, the full testing of the NNORSY climatology, the validation of NNORSY against TOMS-V8, and upgrades to the OCRA/ROCINN algorithms. Most of the work, including the further ozone climatology validation, was ready in 2009.

Phase 2 was concerned first with the implementation and verification of the GODFIT algorithm within the UPAS environment at DLR. The first draft of this document was a deliverable upon completion of this phase in 2009.

Phase 3 includes the geophysical validation of the GDP5 operational product, involving multi-year comparison against ground-based network data and results from other satellites, and the reprocessing of the entire GOME data record. This phase was initiated in 2010.

The project was extended in 2011 with the objective of computing the averaging kernels and including the NO_2 results in the HDF5 output product. After reprocessing and validation of the complete data record, the GDP5 team found out degradation issues with recent GOME data due to the internal closure option. A final Delta-validation in 2011 showed that the use of external closure gives the best results.

Following the algorithm overview in section 2, we discuss the forward model set-up in section 3, i.e. the choice (TOMS Version 8 vs. NNORSY) and usage of O_3 profile climatology, the use of a shift parameter to adjust the temperature profile, the use of albedo closure terms, and the derivation of trace gas cross-sections. Section 4 deals with radiative transfer: generation of optical properties and the simulation of earthshine radiances and analytic Jacobians using LIDORT. The inverse model is described next in Section 5 and Section 6 deals with ancillary algorithms – application of a semi-empirical molecular Ring correction developed for GODFIT, and the derivation and implementation of cloud parameters using the OCRA/ROCINN algorithms. Section 7 contains notes on practical aspects of the GDP5 algorithm (configuration and algorithm settings, performance, UPAS implementation).

Some details in this ATBD have not changed from the Phase B ATBD [R2], so there is some overlap in the documentation. Product Validation is the subject of a separate document to be released concurrently with this ATBD.

2. OVERVIEW OF THE GODFIT ALGORITHM

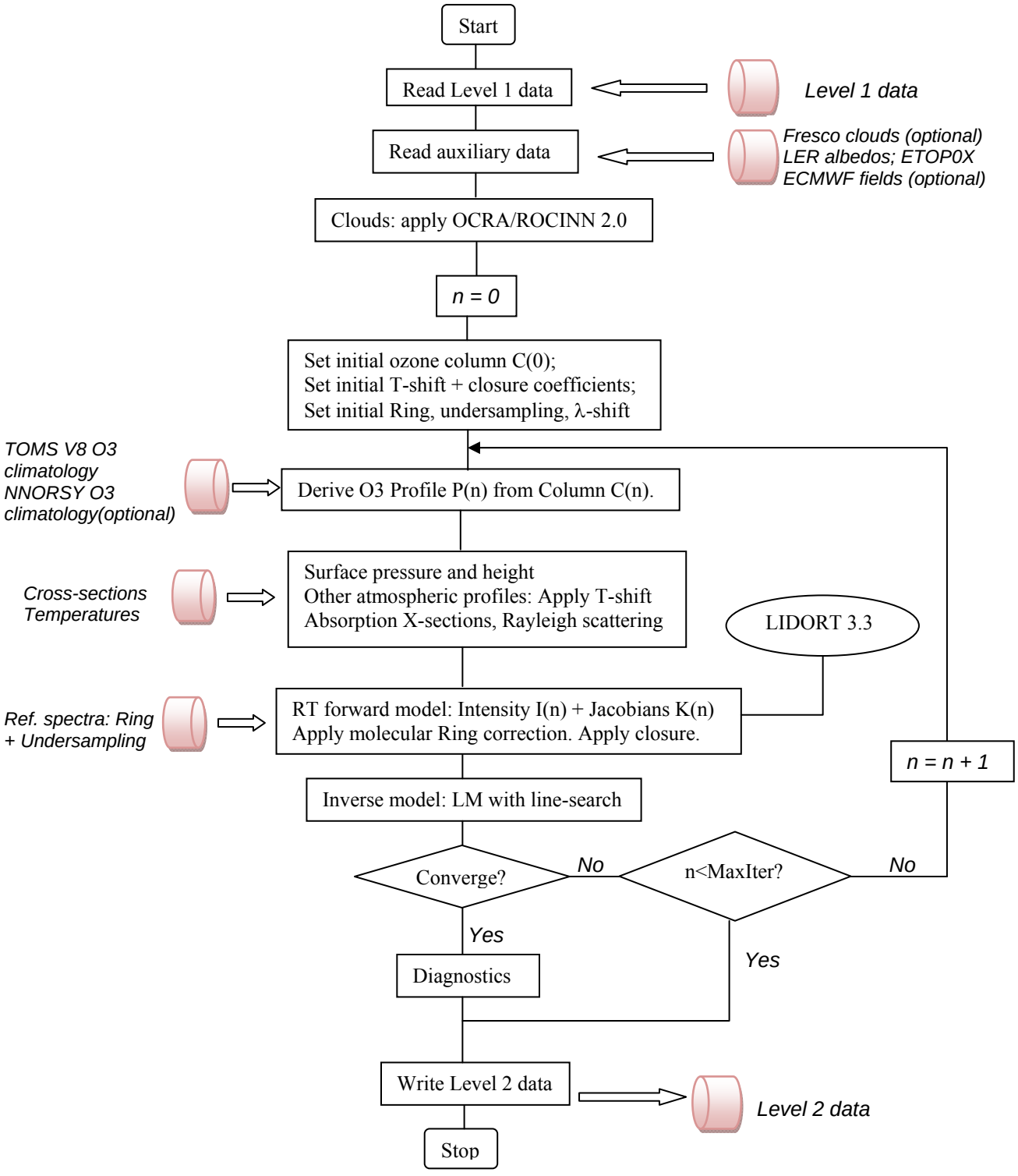


Figure 1. Flow Diagram of the GDP5/GODFIT direct fitting retrieval algorithm.

The GODFIT (GOme Direct FITting) algorithm employs a classical inverse method of iterative least squares merit-function minimization. GODFIT is based on a linearized forward model, that is, a full multiple scatter radiative transfer simulation of GOME earthshine radiances and associated weighting functions with respect to state vector elements. These elements are the total O_3 column, several albedo closure coefficients, a temperature shift, a Ring correction amplitude, and a wavelength registration shift, and other ancillary parameter. In contrast with the DOAS approach, there is no need for a separation of slant column fitting and Air Mass Factor conversion to obtain the vertical column amount.

The main product is the O_3 vertical column itself and the main error diagnostic is the solution variance of this retrieved parameter. The total column and other parameter errors emerge naturally from the chi-square minimization, thanks to the ability of the forward model to deliver Jacobians with respect to these parameters. The basic product is thus simpler than that for DOAS; there are no intermediate quantities such as slant column densities and AMFs.

In order to obtain vertical column density (VCD) in DOAS, the AMF conversion requires knowledge of species profile. The profile should match the VCD, and in order to maintain consistency, it is better to compute the AMF and VCD iteratively. In GDP Version 3.0, an iterative scheme for AMF determination was introduced, along with a neural-network training scheme for fast extraction of AMF results [*Spurr et al.*, 2005]. In GDP 4.0, all AMFs are calculated directly using LIDORT [*van Roozendael et al.*, 2006].

The flowchart in Figure 1 gives an overview of the GDP5-GODFIT direct fitting algorithm. The algorithm is straightforward, with one major decision point. Following the initial read of GOME radiance and irradiance data, and the input of auxiliary data (topography fields, cloud information if pre-calculated, optional ECMWF temperature profiles), the cloud information (fractional cover from OCRA, cloud-top height and albedo from ROCINN) is then derived. The iteration counter is set (n=0), an initial guess is made for the state vector (total ozone amount, temperature shift, closure coefficients, etc.), and the *a priori* information is established.

A unique ozone profile $P(n)$ is then constructed from the total column estimate $C(n)$, using a 1-1 column-profile map based on column-classified ozone profile climatology (either the TOMS Version 8 data, or the NNORSY data set). Next, pressure, temperature and height profiles are constructed; this is where the current value of the temperature shift is applied. Spectral reference data is then created for the fitting window (trace gas cross-sections, Rayleigh cross-sections and depolarization ratios). These set-ups depend on GOME-derived cloud information (fractional cover, cloud-top pressure).

The algorithm then enters the radiative transfer step, in which optical properties are created and the LIDORT Version 3.3 model is called to deliver top-of-atmosphere (TOA) radiances $I(n)$, and the associated ozone column, albedo, and T-shift weighting functions $\boldsymbol{K}(n)$, at each iteration step n. In direct fitting, all simulated backscatter radiances and weighting functions are calculated "on-the-fly" using LIDORT Version 3.3. These simulated quantities are then corrected for the molecular Ring effect.

Next, the inversion module (variable-regularization Levenberg-Marquardt least squares, with line-searching) yields a new guess for the ozone column and other state vector fitting parameters. The iteration stops when suitable convergence criteria have been satisfied, or

when the maximum number of iterations has been reached (in which case, there is no established convergence and final product). Once the converged solution has been found, fitting parameter errors may be determined from the solution variance-covariance matrix; the results are then written to file.

In DOAS algorithms, it is difficult to characterize thoroughly all error sources. Error studies on the slant column fitting have shown that there are some fortuitous cancellations of systematic biases [*van Roozendael et al*., 2002]. The GODFIT direct fitting technique is conceptually simpler, and offers a more logical approach to error handling.

One of the things that emerged from early studies was the requirement for a more suitable treatment of telluric filling-in effects in ozone absorption bands due to rotational Raman scattering (RRS) by air molecules (the so-called molecular Ring effect). This is perhaps the single most important source of error in the older algorithms. In the GODFIT phase A work, a new empirical correction was developed for both the GODFIT and GDOAS algorithms; the GODFIT correction is a standard feature of the GDP5 algorithm. Also because of the GDP 4.x work, several improvements were made to the cloud property retrieval required as a pre-processing step in both the GDOAS and GODFIT algorithms. These improvements (in particular for the ROCINN cloud height and albedo retrieval) are described here.

A good retrieval algorithm stands on the strength of its forward model, and GODFIT is no exception. The forward model must simulate the earthshine radiance spectrum, together with the sensitivities (Jacobians or partial derivatives) of this spectrum to atmospheric and surface parameters that are to be retrieved or are sources of model parameter error. Computers are now so powerful that the use of "on-the-fly" full radiative transfer simulations of radiances and Jacobians is a viable proposition from a data turnover standpoint.

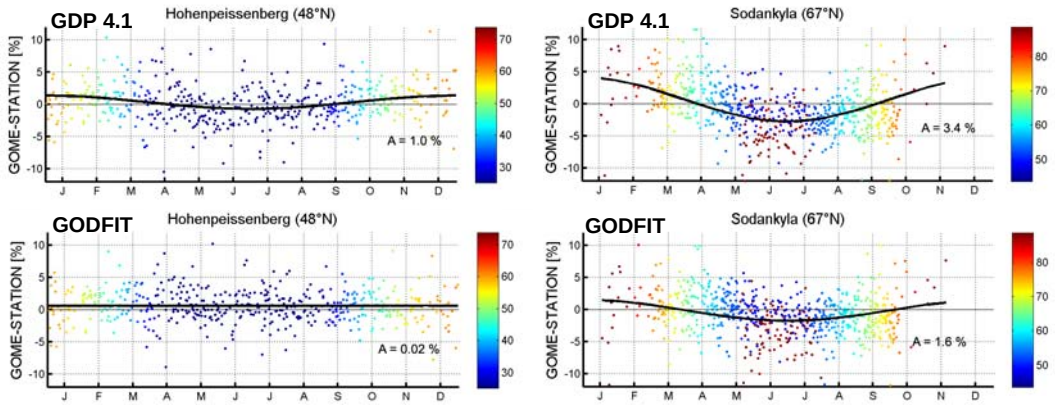


Figure 2. GODFIT (GDP5) and GDOAS (GDP4.1) validated against two northern hemisphere Brewer stations.

GDP5 is a not just an upgrade of GDP4, it is a new concept. We give here one example of the GODFIT capability, to show that this algorithm really is an improvement over the GDOAS results obtained within GDP4.x. Figure 2 presents two validations against Brewer stations in the northern hemisphere, one for Hohenpeissenberg (mid-latitude) and the other for Sodankyla (arctic). In both cases, the residual cyclic (seasonal) signature is markedly improved, particularly for the Arctic station.

3. FORWARD MODEL SETUPS

3.1. Ozone profiles

3.1.1. Introductory remarks

In GODFIT, ozone total column fitting can be thought of as a stripped-down version of a nadir-view ozone profile retrieval algorithm. In a multilayer atmosphere, the forward model will require the specification of a complete ozone profile, and in GODFIT, the ozone profile is parameterized by a single quantity – the total column. Hence, GODFIT needs a unique 1-1 correspondence between the ozone profile and its total column. A suitable form for this mapping was developed for GOME and first incorporated in GDP 3.0 [*Spurr et al*., 2005].

The use of total column as a proxy for the ozone profile was recognized a number of years ago by scientists at NASA, and column-classified ozone profile climatologies were created for the TOMS Version 7 [*Wellemeyer et al.*, 1997] and more recently for Version 8 (V8) retrieval algorithms [*Bhartia*, 2003]. The TOMS Version 7 profile set was the baseline in GDOAS and GODFIT Phase A, while Version 8 became the GDP4.x and GODFIT Phase B baseline. The GODFIT atmosphere-layering scheme uses the TOMS V8 pressure gridding.

Recently, new ozone climatology has been developed from a multi-year record of retrieved GOME ozone profiles, and the data set has become part of a neural network (NN) scheme for the extraction of ozone profiles. This is the NNORSY climatology [*Müller et al*., 2003; *Kaifel et al*., 2008]. GODFIT and GDOAS interfaces for the first version of the NNORSY climatology were developed in the Phase B work, but no NNORSY-based total column retrievals were performed, mainly due to some remaining limitations in the climatology itself. However, these limitations were subsequently resolved during the first phase of GDP5, and the NNORSY climatology has now been used for extensively testing its impact on ozone column retrieval.

The choice of O_3 climatology is an important decision for GDP5 operational implementation, and there have been a lot of NNORSY/TOMS retrieval comparisons. These are summarized in section 3.1.4; for a detailed comparison, see the Full-version ATBD and the various Progress Meeting minutes, After this lengthy comparison exercise, it was decided to stay with TOMS Version 8 pending further improvements in the NNORSY data set.

Remark. A new implementation of NNORSY (based on the NIWA ozone profile database [*Hassler et al*., 2008] was presented at PM5 (18 September 2009) by A. Kaifel. Also presented at this meeting was a preliminary assessment of the NNORSY output from the BIRA team.

3.1.2. The TOMS V8 ozone climatology

The TOMS profile is a set $\{U_j\}$ of partial columns in [DU]; the total column is $C = \sum_j U_j$. For two adjacent profiles $\{U_j^{(1)}\}$ and $\{U_j^{(2)}\}$ with total columns $C^{(1)}$ and $C^{(2)}$ we define an intermediate profile with column amount C according to:

$$U_j(C) = \left(\frac{C - C^{(1)}}{C^{(2)} - C^{(1)}} \right) U_j^{(2)} + \left(\frac{C^{(2)} - C}{C^{(2)} - C^{(1)}} \right) U_j^{(1)} . \qquad (1)$$

This defines a *linear* profile-column map. Total column weighting functions are related to profile Jacobians by means of chain rule differentiation, and from Eq. (1), the partial derivatives with respect to the total column are:

$$\frac{\partial U_j(C)}{\partial C} = \frac{U_j^{(2)} - U_j^{(1)}}{C^{(2)} - C^{(1)}}. \qquad (2)$$

This map allows us to interpolate smoothly between profile entries in the climatology; the shape will vary continuously. In effect, we are drawing on an ensemble of possible profiles of which the climatology is a sample. It is possible to define a quadratic map based on three adjacent profiles, and assuming gradient continuity instead of the piecewise step derivatives in equation (2). A 4-point Lagrangian (cubic) interpolation scheme was found to work well for the TOMS profiles, and this is the current default mapping for TOMS V8.

The 11-layer pressure grid is based on scale heights. Latitude variation is incorporated by combining two sets of profiles with a cosine-latitude weighting.

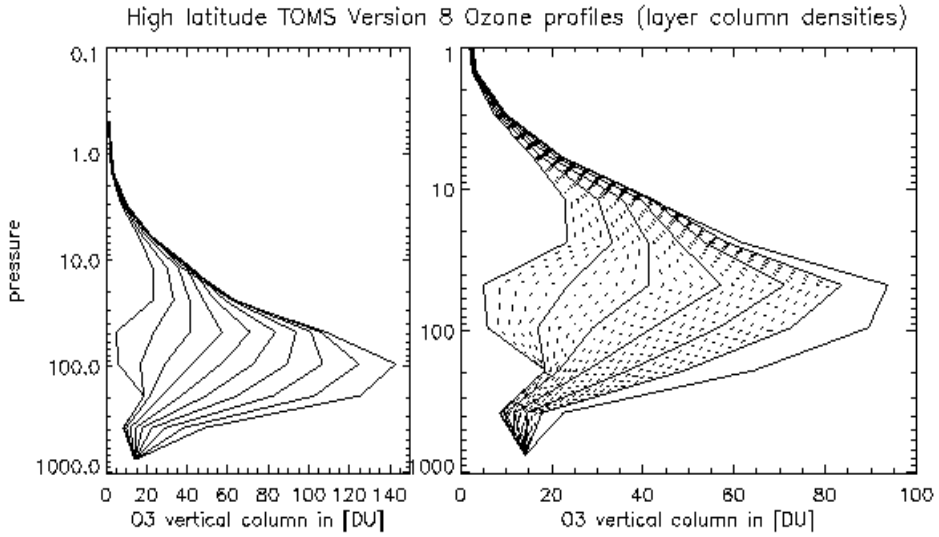


Figure 3. (Left) TOMS Version 8 high latitude profiles (partial column in [DU]). (Right) Intermediate profiles (dotted) for total columns at 10 DU intervals between climatology values.

TOMS Version 8 profiles [*Bhartia*, 2003] are defined on the same pressure grid. Profiles are specified for 18 latitude bands from pole to pole (10° intervals), and for each month of the year. Latitude and time variations are treated using a bilinear interpolation scheme. Figure 3 shows the ozone profile-column map for one set of 10 high-latitude TOMS V8 profiles.

Adjustments are made to lowest Umkehr profile elements to account for the GOME scenario surface pressure or the ROCINN-retrieved cloud-top pressure. This adjustment assumes that the partial column in any layer is proportional to the logarithm of the layer pressure drop. This adjustment does not destroy the integrity of the column-profile map.

3.1.3. Ozone profiles: the NNORSY data

The CHEOPS (Climatology of Height-resolved Earth Ozone and. Profiling Systems for GOME) project was designed to set up complete collection of GOME-dedicated profiles, as

retrieved over the entire length of the GOME data record using the NNORSY neural network algorithm [*Müller et al.*, 2003; *Kaifel et al.*, 2008]. However, there was a delay in the delivery of the data set, and it was impossible to use the data in time for the completion of GODFIT Phase B. This situation has changed for the GDP 5.0 project, and the NNORSY data became a serious contender for the operational choice of ozone profile climatology.

Integration of the NNORSY climatology is straightforward. For given latitude, longitude, time, total column amount and choice of input pressure grid, a stand-alone subroutine will return the complete ozone profile on that grid. Thus, the profile-column map is an intrinsic feature of the data set. Profiles have been compiled for total columns ranging at least from 150 DU to 575 DU, similar in scope to that for the TOMS V8 data set.

The profile climatology computation is based on the use of tangent hyperbolic functions employing a series of neural network weights and biases created off-line as part of the NN training by the NNORSY team [*Kaifel et al.*, 2008]. D. Loyola developed a module to remap the original NNORSY profiles to ensure that the integrated output profile is equal to the input total ozone; the profiles are additionally converted from number densities on a 61-level vertical grid to Umkehr layers in [DU] on the TOMS pressure grid. The NNORSY library contains additional climatologies (e.g. taking also temperature profile as input) that were not further tested in the framework of GDP 5.

During the trials with NNORSY in GODFIT, a number of issues emerged. First, it was found that the original (or "Mode 0") NNORSY-derived profiles were inconsistent. For an input total column C, the NN routine will generate a profile U_j, and it was then found that the *ex post facto* profile sum $C' = \sum U_j$ was sometimes significantly different to the original total column amount. It is then necessary to remap the profile by searching in the NNORSY climatology for a profile for the given location and time with the property that its integrated total column is equal to the amount C. A scaling factor $f = C/C'$ is applied only when the NNORSY climatology fails to yield such a profile. In the latter case, we have $\sum U'_j = \sum fU_j = fC' = C$. This leads to the NNORSY *remapped* or "Mode 1" profiles.

This remapping has some implications for scaling the profile derivative. For input column C delivering profile U_j, and profile derivative $V_j = \partial U_j / \partial C$, it follows that

$$V'_j \equiv \frac{\partial U'_j}{\partial C} = fV_j + \frac{U_j}{C'}\left[1 - f\sum V_j\right] \tag{3}$$

Upon investigation, it was found that the original profile derivatives $V_j = \partial U_j / \partial C$ were not properly computed. Although these "Mode 1" derivatives did not lead to any really significant differences in the total ozone product, the least-squares fitting convergence was much slower. A work-around solution for the NNORSY derivatives was implemented - these are now computed using a two-step finite difference scheme. That is, for column C resulting in (remapped) profile P_j, profiles $P_j^{\pm}$ are computed corresponding to columns $C^{\pm} = C \pm \delta C$, and the finite difference (FD) derivative is then

$$\frac{\partial P_j}{\partial C} \approx \frac{P_j^+ - P_j^-}{2\delta C} \tag{4}$$

Use of this FD scheme is equivalent to the linear profile-column map as used for TOMS V8. Results were not sensitive to the choice of difference step δC ; a default of 5 DU was used. This scheme generates derivatives with very similar shapes to those for TOMS; convergence is then similar to that for TOMS.

3.1.4. NNORSY vs. TOMS V8 comparisons

The first GODFIT comparisons were reported at PM2 in September 2008. Here, 4 different set of retrieved total ozone were compared with ground-based data, corresponding to 3 NNORSY modes and the phase-B TOMS V8 baseline. The 3 NNORSY modes are:

Mode 0: Original NNORSY profiles.
Mode 1: Remapped profiles, i.e. the integrated profile equals input total ozone.
Mode 2: Remapped profiles with matched TOMS-V8 tropospheric content.

"Mode 2" NNORSY profiles have a tropospheric/stratospheric adjustment designed to ensure that tropospheric content of all profiles matches that seen in TOMS Version 8.

Validations were made for all pixels within a radius of 500 km from 114 ground stations (85 in northern hemisphere and 29 in the southern) between 1996 and 2007. For GODFIT with a 325-335 nm fitting window, ECMWF temperature fields were used without T-shift fitting, and the algorithm used internal albedo closure, with cloud parameters from FRESCO+.

Time-latitude dependencies comparing NNORSY Modes with TOMS-derived total ozone were examined. Large differences at all latitudes were observed with NNORSY Mode 0. There was good agreement for Modes 1 and 2, with Mode 2 correction mainly affecting high latitudes. Similar results were noted for the seasonal-latitude dependency. For the time/SZA analysis, it was found that Mode 2 led to improvements with TOMS-V8 results at high SZA in northern latitudes, but agreements were poorer for high SZA in southern Polar Regions.

	O_3 climatology	**T° profile**	**Effective cloud pars.**
Scenario 1 (*PM2 baseline*)	TOMS v8	ECMWF	OCRA/ROCINN
Scenario 2 *(PM3 baseline)*	NNORSY (mode 1)	TOMS shifted	OCRA/ROCINN
Scenario 3	NNORSY (mode 1)	ECMWF	OCRA/ROCINN
Scenario 4	TOMS v8	TOMS shifted	OCRA/ROCINN
Scenario 5	NNORSY (mode 1)	TOMS shifted	FRESCO+

Table 1. NNORSY/TOMS V8 comparisons: 5 scenarios.

Comparisons with OMI total ozone were made using Version 8.5 of the TOMS total ozone algorithm applied to Collection 3 OMI data, where for each station and day, ozone columns from the two closest pixels are compared (maximum distance of 250 km, largest time difference of 5h). Season-latitude dependencies of the GODFIT/OMI ozone differences were examined. Again, Mode 0 comparisons were found unacceptable. For Mode 2 NNORSY, the

agreement with OMI was better except after September where differences increased at high latitudes in both hemispheres. TOMS V8 and NNORSY Mode 1 show similar agreement with OMI. Mode 1 was chosen as the NNORSY baseline.

The second set of GODFIT comparisons were reported at PM3 in February 2009. This time, the focus was on NNORSY Mode 1 and TOMS V8. The main object is to test the two ozone data sets with and without (ECMWF) the temperature shift applied (see Section 3.2.3 for an example of T-shift testing). Cloud algorithms were also tested. The sampling sets are the same as those used in the first comparisons.

Temperature and climatology impacts are contrasted and compared in Figure 4. Depending on the region, the new baseline (scenario 2) leads to total O_3 columns in good agreement or 1-2 % larger than the values derived with the old baseline (scenario 1). It is seen that the T-shift leads to higher values in the tropics and in ozone hole conditions.

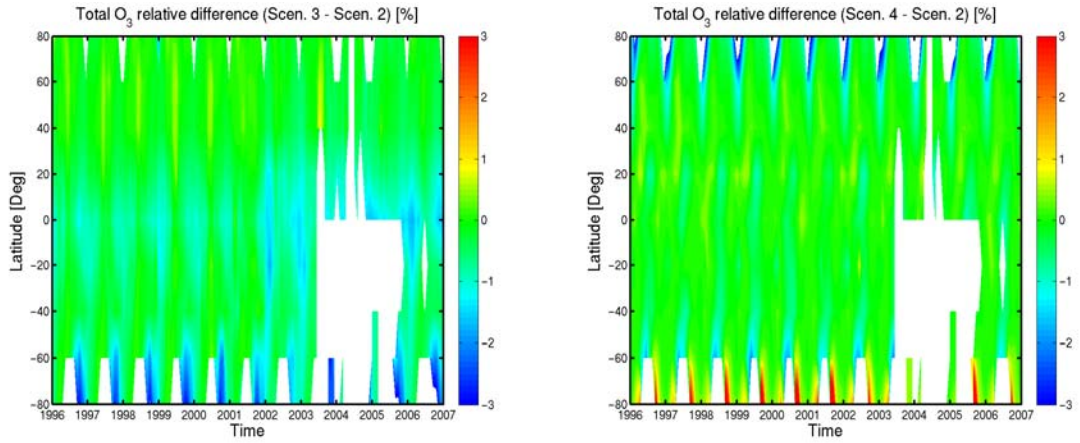


Figure 4. GODFIT comparisons for GOME total ozone from 1997-2007. (Left) Impact of temperature (ECMWF vs. T-shift) for the NNORSY mode 1 data. (Right) Impact of climatology for results using T-shifting.

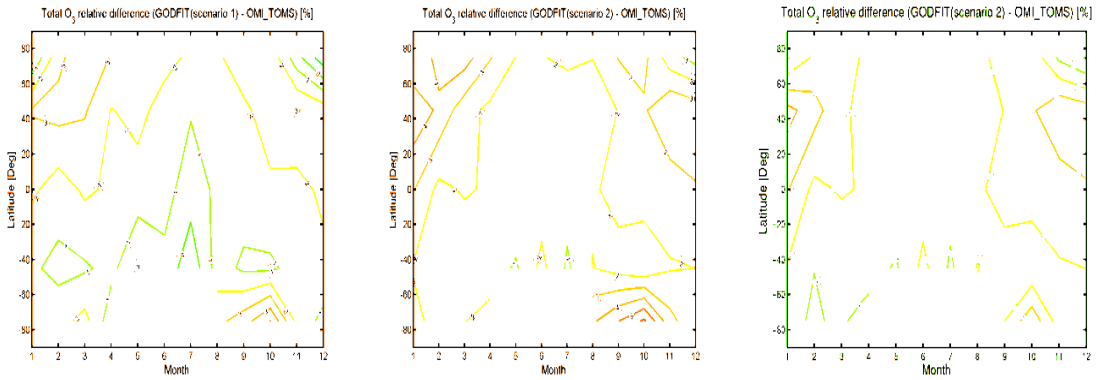


Figure 5. GOME-OMI comparisons for total ozone. (Left) Old baseline with ECMWF temperature and TOMS climatology; (middle) New baseline with NNOSRY Mode 1 and T-shifting; (right) New baseline but with no ghost columns in ice mode.

In Figure 5, we compare the new and old baselines (scenarios 1 and 2) against OMI results. Comparing the left and middle panels, we see that there is less structure with the new baseline. The question of overall bias is related to the choice of cross-sections (Section 3.3).

Note that with the new baseline (T-shift), there appears to be a systematic overestimation in ozone hole conditions. Investigations have shown that this is due in large part to the erroneous use of a ghost column over snow and ice surfaces (see the discussion in 4.1.2). In the right panel of Figure 5, we see the improvement when the ghost column is switched off over such surfaces.

Remark. It is possible to add a great many more examples comparing TOMS V8 with NNORSY. A number of validation studies were done for these choices, both against the WOUDC network and the NDACC stations. Validation results were obtained by three teams working in concert, one at DLR (GOME data), another at AUTH (WOUDC validation), and a third at BIRA-IASB (NDACC validation). Detailed results can be found in the presentation material from Progress Meetings 2, 3 and 4. Different cloud options were also tested. We give just one example from this volume of work. Figure 6 validates the GODFIT algorithm against the WOUDC network for 10 years of data; differences are averaged into latitude bands. The plots are very similar, though the TOMS V8 results validate better in the Antarctic (left panel).

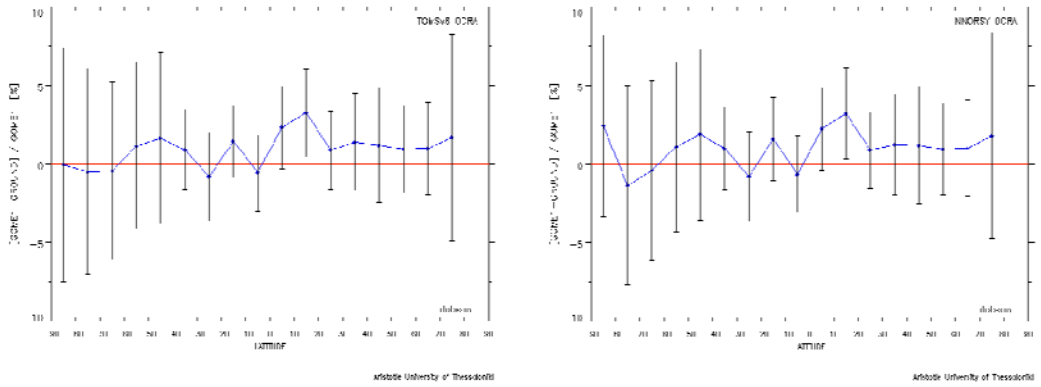


Figure 6. WOUDC validation of GODFIT using TOMS Version 8 climatology (left), and using the NNORSY model 1 climatology (right).

3.2. Pressure and temperature; the T-shift procedure

3.2.1. P/T data sets: TOMS P/T baseline and ECMWF

Pressure levels defined by the TOMS V8 climatology have been used to set vertical layering for the GODFIT forward model simulations; pressures are halved for each successive atmospheric boundary. The scale height is typically around 5.0 to 5.7 km. There are 14 levels in all. TOA was set at 0.03 hPa, with surface pressure allowed to vary according to suitable input values determined by geographical position.

Temperature profiles are required for hydrostatic balance and determination of ozone cross sections. In GODFIT and GDP5, temperature profiles are always taken from an external source. In the GODFIT Phase A baseline, temperatures were taken from an auxiliary TOMS temperature data set defined on the TOMS V7 pressure grid; temperature profiles were used without interpolation. For TOMS V8 in Phase B, a latitude-zone monthly temperature climatology supplied with the ozone profiles [*Bhartia*, 2003] has been used. Here, one T-

profile is given per 10° latitude bin and per month and the atmosphere is divided in 13 layers. Bi-linear interpolation in cosine latitude and time is performed for scenario adjustment.

In these TOMS-based atmospheric setups, height levels are determined by hydrostatic equilibrium based on a reference height input (the surface topographical height). The acceleration due to gravity is allowed to vary with latitude and height according to the specification in [*Bodhaine et al*., 1999]. Layering is set up for clear sky and cloudy scenes, with the lowest layer adjusted to fit the boundary pressure (surface or cloud-top); height and temperature adjust linearly with the pressure logarithm.

For this step of the physics setup, we need to assign for each GOME footprint, a surface height and surface pressure. We used the ETOP05 topographical database for surface height [*ETOP05*, 1988]. In GODFIT Phase A, surface pressures were either taken from the FRESCO [*Koelemeijer et al.*, 2001] or ICFA [*Kuze and Chanc*e, 1994] auxiliary data product, or from ECMWF surface pressure analysis fields, or interpolated against the surface height using a standard atmosphere. In Phase B, surface pressure was either interpolated as before, or taken from ECMWF or from the FRESCO+ [*Wang et al*., 2008] auxiliary data.

Given the temperature sensitivity of ozone absorption, it is desirable to use historical analysis fields from Numerical Weather Prediction (NWP) models, such as those from ECMWF data center (ERA-40). The ECMWF implementation was done at the end of GODFIT Phase A. ECMWF pressure and temperature fields are incorporated as a standard option in the Phase B and GODFIT software environments. One ECMWF profile is available every 6 hours with a spatial resolution of 3.75° X 5° (Lat/Long). Though the ECMWF profiles are specified at 37 levels, temperatures are interpolated (linearly with the logarithm of the pressure) to the TOMS pressure grid. The current baseline is to select the spatially and temporally closest profile for each pixel.

3.2.2. The temperature shifting algorithm

The ECMWF dataset is regarded the most accurate available for temperature fields. However, it is desirable to adjust the temperature profile to better reflect the dependence of the ozone absorption signature on temperature. The GDOAS algorithm has a temperature adjustment, and here we show that a temperature shift adjustment can be defined for GODFIT.

In DOAS-style algorithms, an “effective temperature” is determined as an adjustment factor from the use of two reference O_3 absorption cross sections at two different temperatures in the fitting [*Richter and Burrows*, 2002; *van Roozendael et al*., 2006]. In GODFIT, we define an effective temperature from the input temperature profile, through weighting with the layer total ozone column amounts:

$$T_{eff} = \frac{\sum T_n U_n}{\sum U_n}. \qquad (5)$$

Figure 7 (top left) shows orbit differences in GODFIT effective temperature using TOMS and ECMWF T-profiles. In Figure 7 (top right), the same orbit of effective temperature differences is plotted against the corresponding total ozone difference (ECMWF vs. TOMS).

There is a remarkable linear correlation; compared with ECMWF results, a difference of 10 K in effective temperature results in a 3% change in total ozone.

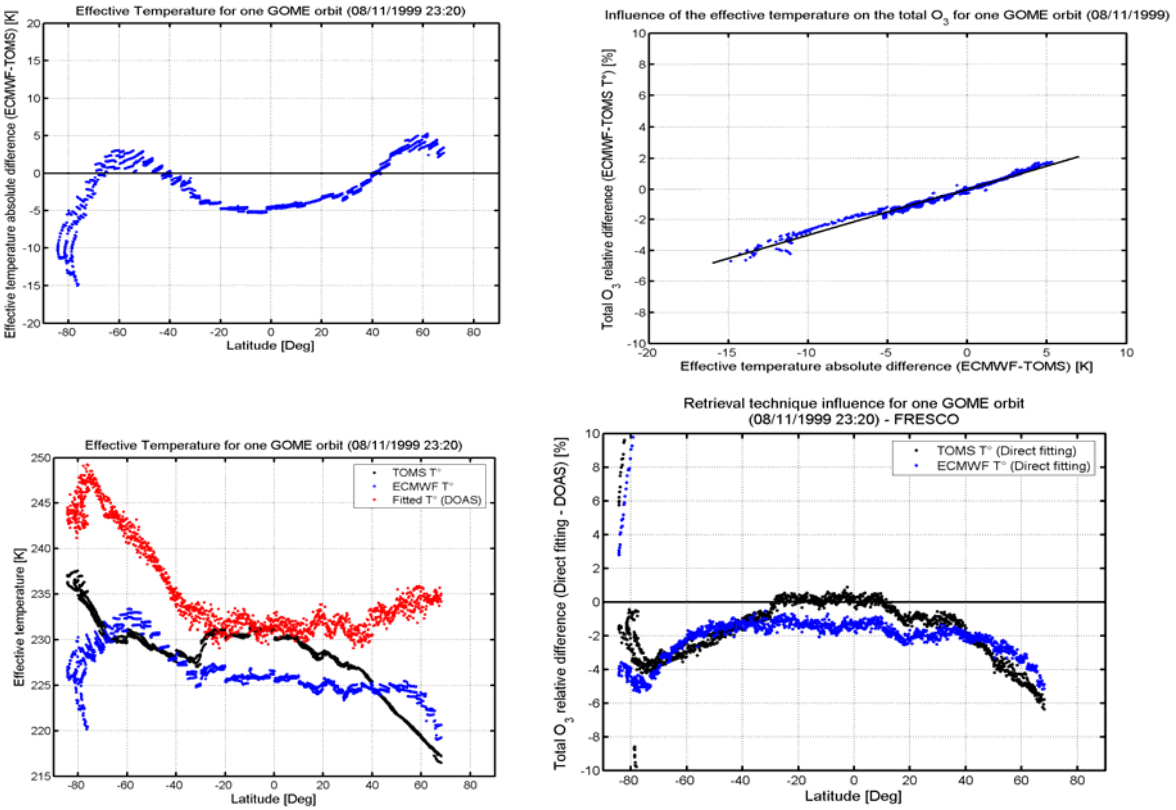


Figure 7. (Top left) ECMWF/TOMS effective temperature differences computed for one GOME orbit. (Top right) The same differences plotted against relative total ozone. (Bottom left) Effective temperatures estimated from TOMS and ECMWF temperature profiles in direct fitting and effective temperature adjusted during DOAS fit for one GOME orbit. (Bottom right) Total O_3 relative differences for the same orbit between the DOAS and Direct fitting methods.

In Figure 7 (lower right), ECMWF temperature profiles for one orbit lead to values for the total O_3 amount that are generally 1-2 % lower than those obtained with TOMS temperature profiles (other things being equal). Note also the negative offset in total ozone compared with GDOAS results. In Figure 7 (lower left), the corresponding absolute effective temperatures show negative bias compared with the DOAS effective temperature result. These plots strongly suggest the value of developing an adjustment procedure for temperature profiles in the direct fitting, and we now describe this adjustment for TOMS V8 and NNORSY.

Given an ozone profile consisting of layer Umkehr columns $\{U_n\}$, $n = 1....N_L$ (the number of layers) depending on input total column C, and a corresponding set of temperatures $\{t_n\}$ specified at associated pressure levels $\{p_n\}$, $n = 0,.... N_L$. We assume these temperatures have already been registered (that is, suitably interpolated using the hydrostatic balance equations) on the (TOMS) pressure grid. We allow the temperature profile to be written

$$\theta_n = t_n + S\phi_n . \tag{6}$$

Here, S is a ***shift amplitude*** which is treated as a free parameter to be retrieved, and ϕ_n is a pre-specified *temperature shape function*.

Column air density and its partial derivative with respect to the T-shift S are then given by:

$$A_n = \frac{d_n R}{2}\left[\frac{p_{n-1}}{\theta_{n-1}} + \frac{p_n}{\theta_n}\right]; \tag{7a}$$

$$\frac{\partial A_n}{\partial S} = -\frac{d_n R}{2}\left[\frac{p_{n-1}\phi_{n-1}}{\theta_{n-1}^2} + \frac{p_n \phi_n}{\theta_n^2}\right]. \tag{7b}$$

Here, d_n is the layer height thickness and R is a constant related to Loschmidt's number. The air density is required for the computation of IOPs (inherent optical properties) which are the inputs to the LIDORT radiative transfer model. [IOP setup is discussed below in section 4.1]. Since we are retrieving S, LIDORT must also produce Jacobians with respect to S, and for these we require the ***linearized*** derivatives of the IOPs – hence the need for the air density derivative. For trace gas absorption, ozone cross-sections α_n are also dependent on the T-shift S. Indeed, for a parameterized quadratic temperature dependence based on a set $\{\sigma_0, \sigma_1, \sigma_2\}$ of cross-sections, we have (θ^*=273.15 K):

$$\alpha_n = \sigma_0 + \sigma_1(\theta_n - \theta^*) + \sigma_2(\theta_n - \theta^*)^2 ; \tag{8a}$$

$$\frac{\partial \alpha_n}{\partial S} = \phi_n\left[\sigma_1 + 2\sigma_2(\theta_n - \theta^*)\right]. \tag{8b}$$

Again, the T-shift derivative in Eq. (8b) is necessary for the linearized IOP setup. It is straightforward to write down a derivative if linear or Lagrangian interpolation is used for the temperature dependence of the cross-sections.

The baseline default is to use a constant shape function $\phi_n \equiv 1$ applied to all temperatures within a specified altitude range in the atmosphere. This range might be the whole atmosphere itself. A second idea uses a concept called *shift triangulation*. It is known that the largest sensitivity to temperature occurs at the ozone number density maximum. This level is different for each column-classified profile. In shift triangulation, we assume for each profile that the shift tails off linearly in both directions from the peak maximum value S, decreasing to zero at height z_u above the peak level, and zero at height z_b below the peak level. This shape function is:

$$\left.\begin{aligned} S_j(z) &= S\frac{z_u - z}{z_u - z_{p,j}}, \quad z_u \geq z \geq z_{p,j} \\ S_j(z) &= S\frac{z - z_b}{z_{p,j} - z_b}, \quad z_{p,j} \geq z \geq z_b \\ S_j(z) &= 0, \quad z \geq z_u, z_b \geq z \end{aligned}\right\} \tag{9}$$

Equation (9) defines the shift factor entries ϕ_{inj} used in the foregoing analysis; S is the "shift maximum" - a free parameter. For the j^{th} column-classified profile we assign peak height $z_{p,j}$ to be that boundary level p of the input altitude grid for which the corresponding layer contains the largest partial column of ozone (evaluated at zero temperature shift). The limits z_u and z_b may be chosen by examination of the effect of temperature changes on ozone profiles. Typically z_b can be taken as the surface level, while z_u might be 35-40 km.

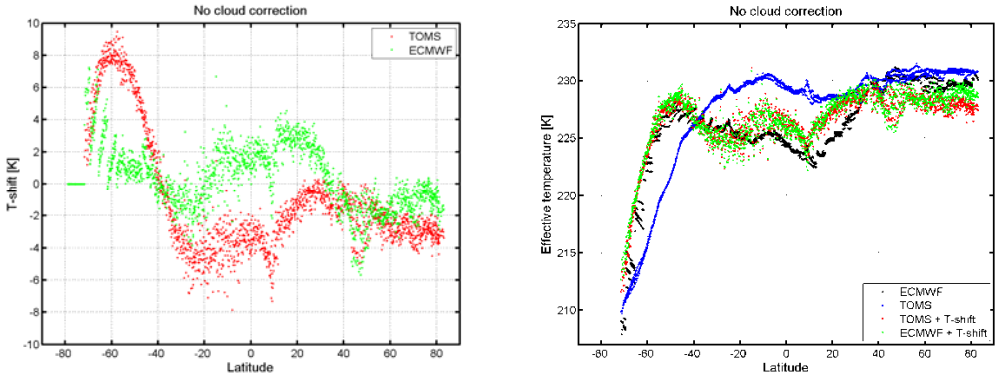


Figure 8. (Left) retrieved T-shift parameters for one GOME orbit. (Right) Effective temperatures for the same orbit, with and without retrieved T-shifts.

3.2.3. Retrievals using the T-shift in GODFIT

Here we present some results for the GODFIT retrieval of ozone, with and without the T-shift parameter added to the state vector as an ancillary retrieval parameter. Results for a single orbit have shown that the T-shift is very weakly dependent on any *a priori* value; in other words, the fitting is robust. Figure 8 shows that in general, only small shift adjustments are obtained for ECMWF profiles, larger values are apparent for TOMS profiles.

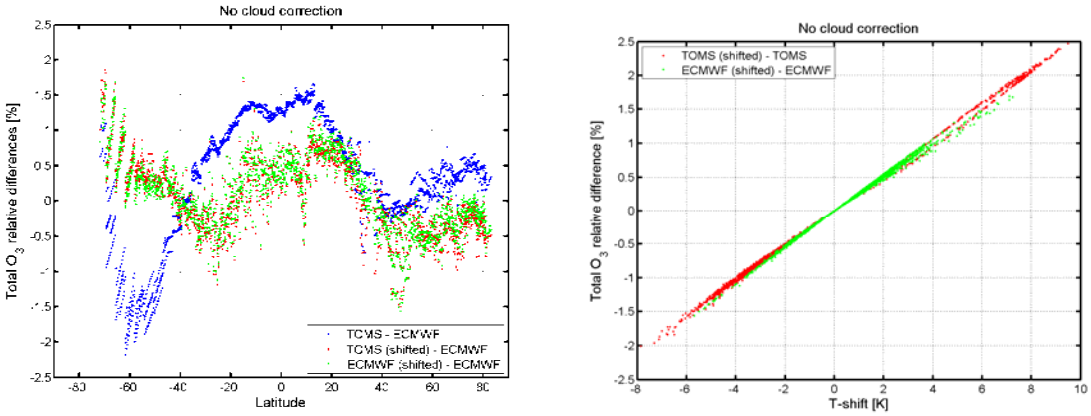


Figure 9. (Left) total ozone differences for one orbit (as in previous figure), comparing results against a non-shifted ECMWF baseline. (Right) T-shift linear correlation with ozone differences, for the same data set.

As far as the impact on total ozone is concerned, Figure 9 shows that the T-shift procedure applied to TOMS temperature profiles generally leads to small differences with respect to the

ozone columns derived using ECMWF temperatures without the application of a T-shift. Thus, ozone columns derived using the T-shift process are consistent, no matter the choice of *a priori* profile. This is an important remark, as it means that the ECMWF profiles are not required as a *sine-qua-non* for accurate total columns. Indeed, the use of ECMWF profiles for temperatures is questionable; there are a number of concerns about bias in the retrieved ozone, quite apart from questions over the use of an external database in the retrieval. Figure 9 (right panel) also demonstrates the quasi-linear correlation between T-shift and total O_3 differences. The quality of the fit using the T-shift is just as good if not better than that obtained using the Phase-B baseline (ECMWF, no shift).

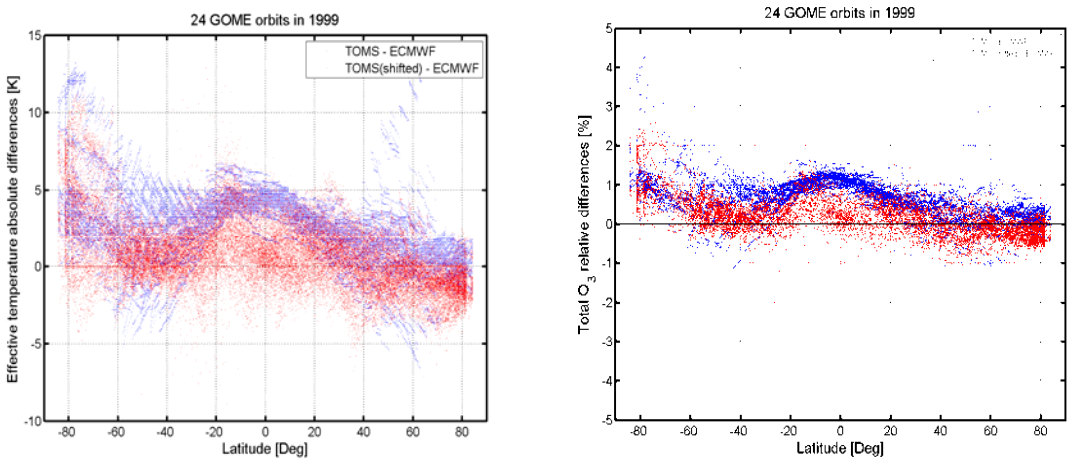


Figure 10. (Left) 24-orbit comparisons of effective temperature differences (left) and total ozone differences (right), with TOMS–unshifted (blue) and TOMS T-shifted (red) results compared against the un-shifted use of ECMWF profiles.

Next, we present some results for a wider sample (24 orbits in 1999). Figure 10 compares all results for this sample, showing ozone differences for TOMS (with and without the T-shift), and for shifted-TOMS against un-shifted ECMWF. The improvement in total ozone results with the T-shift is apparent.

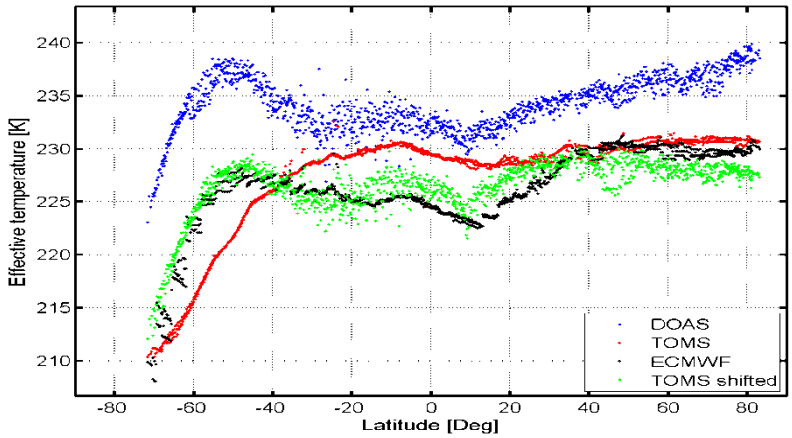


Figure 11. Effective temperatures for one orbit, comparing DOAS results with three GODFIT results as indicated. GODFIT used the DM cross-sections (see next section).

Finally, we return to effective temperatures, comparing GDOAS adjusted values with results from GODFIT, with and without T-shifting for TOMS, and without shifting for ECMWF (Figure 11). There is a persistent high offset (10 K) for GDOAS results compared with those obtained with un-shifted ECMWF. Note again the good agreement (apart from this offset) when comparing T-shifted TOMS in the direct fitting mode with GDOAS values. The DOAS approximation (using the Beer-Lambert law) neglects the wavelength dependency of the photon path length. This shortcoming leads to a misfit of ozone absorption structures, which incidentally appears to be compensated by an overestimation of the effective temperature.

GDP5 baseline. Uniform shape function (same shift at all levels), with initial value $S = 0$.

3.3. Trace gas cross-sections

3.3.1. Ozone cross-section data sets

In Phase A of GODFIT and in the subsequent work with GDOAS in GDP 4.0, ozone cross sections were taken from the GOME FM 98 data sets for ozone [*Burrows et al.*, 1999] and NO_2 [*Burrows et al.*, 1998]. In the Huggins bands range, the baseline used re-sampled data which included the solar I_0 correction [*Aliwell et al.*, 2002], and which were pre-shifted (the pre-shift was optimized at +0.016 nm). For more discussion on these data sets, see [*Orphal*, 2003]. The re-sampling of cross-sections is treated in more detail in section 7.2.3.

Cross-section uncertainties are still a crucial source of retrieval error in DOAS and no less so in direct fitting retrieval algorithms. Hitherto in ozone algorithms, the choice of ozone cross-sections has focused on the older Bass/Paur laboratory data from the 1980s [*Bass and Paur*, 1984] versus the use of GOME [*Burrows et al.*, 1999] or SCIAMACHY [*Bogumil et al.*, 2003] flight model data. The latter data sets are both derived from instrument measurements taken by the respective instrument in its pre-flight calibration phase.

Recently, it has become clear that newer high-resolution laboratory data (the so-called "Daumont-Malicet [DM]" cross sections [*Daumont et al.*, 1992; *Malicet et al.*, 1995]) give lower uncertainties than those from flight model data. See for example the investigation by [*Liu et al.*, 2005]. The DM cross sections were introduced in GODFIT Phase B. The following sensitivity analysis indicates that DM should be chosen as the baseline for GDP5.

At the IGACO May 2009 meeting, discussion focused strongly on the evaluation of the DM data set, in comparison to the current standard (Bass/Paur). NASA and ESA are both inclined to switch to DM as the new cross-section baseline for TOMS Version 9 and GDP 5 respectively. More laboratory data is required: at the meeting, it was noted that there is a need for a large scale project dedicated to improving spectroscopy for atmospheric chemistry satellite missions.

3.3.2. Ozone cross-sections: Sensitivity studies

First, we look at DOAS total ozone. Use of the DM data gives rise to high quality fits for GOME and SCIAMACHY, and to a good consistency for both instruments in terms of effective temperature as fitted in the DOAS method. This is clearly seen in Figure 12, where DOAS effective temperatures are computed for GOME and SCIAMACHY overpass data during the SAUNA campaign. Effective temperatures agree closely with the DM data – this is not the case with Bass and Paur, and the flight model data sets.

In addition, O_3 columns based on DM data are in good agreement with reference values obtained from GDP 4.0 with effective temperature forcibly constrained to the ECMWF value. For GOME, these reference values are obtained with the FM98 data [*Burrows et al.*, 1999] by fitting the temperature, whereas for SCIAMACHY spectra, they are obtained using the FM data [*Bogumil et al.*, 2003] with a +3% scale factor. Figure 13 illustrates the good agreement between these reference values and DM-based O_3 columns with ECMWF-constrained effective temperatures. [The -2% scale factor applied to SCIAMACHY O_3 cross-sections other than [*Bogumil et al.*, 2003] is required due to the presence of intensity offsets in the Level 1c v6 extracted irradiance and radiance spectra].

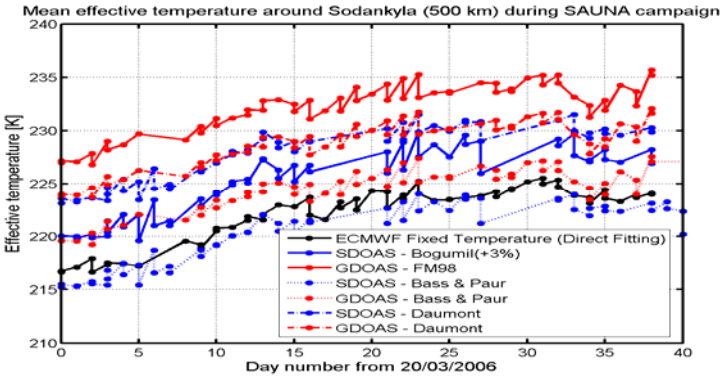


Figure 12. Effective temperature adjusted in the DOAS fit for GOME and SCIAMACHY pixels in the Sodankylä area during the 2006 SAUNA campaign. Results are computed for all available O_3 cross-section data sets, and compared with the ECMWF GODFIT result.

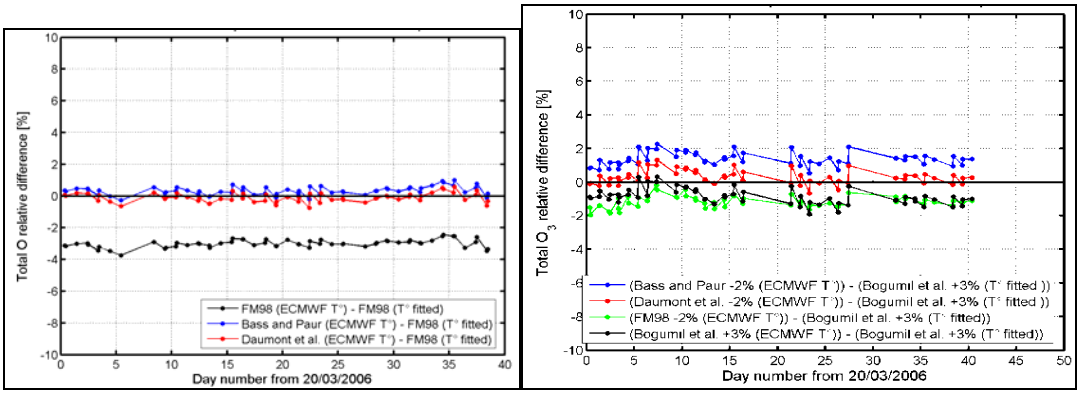


Figure 13. Total O_3 relative differences between results issued from various O_3 cross-section data sets by forcing the effective temperature to the ECMWF value and the reference values (details in the text) for GOME (left panel) and SCIAMACHY (right panel) pixels in the Sodankylä area during SAUNA campaign.

For GDOAS in GOME, the DM/FM98 bias is clear. Using FM 98 data leads to O_3 columns 3% *lower* than the columns derived using the DM data. Using the DM data can still lead to differences between the GDOAS and GODFIT results, as there is still a bias between the effective temperature adjusted in DOAS and the more realistic ECMWF temperatures. This could possibly be the result of a misfit effect related to the approximate Ring correction scheme used in the DOAS part of the inversion (this is currently under investigation).

Returning to Figure 11, it would appear that in GDOAS, the ~10K effective temperature offset should produce ozone columns on average 3% *higher* than expected. However, the GDOAS use of the GOME FM 98 cross sections leads to total ozone values ~3% *lower* than those obtained using DM data. The two effects compensate; a typical fortuitous happenstance for DOAS. In GODFIT, there is a bias in the retrieved effective temperature when using FM98, suggesting that the T-interpolation used in FM98 is not accurate enough. There is no such bias observed with the DM data. Fitting residuals based on the FM 98 and DM cross-sections are similar in magnitude.

The above results indicate that the use of DM cross-sections in direct fitting not only reduces bias but also improves the mean consistency with past GDOAS results (and therefore with correlative data).

3.4. Albedo closure

Scattering from tropospheric aerosol is very difficult to decouple from surface reflectivity in the UV ozone range considered here (315-340 nm). In DOAS, slant column fitting, all broadband radiative effects (all aerosol effects along with surface reflection and molecular scattering) are filtered out using a low-order polynomial. This filtering is an *external closure*; given the highly simplified forward model assumption inherent in the Beer-Lambert extinction law, the resulting fitted parameters have no physical interpretation. In the AMF computation, aerosols are then reintroduced in an *ad hoc* manner; there is no real justification for this, and they become a hard-to-quantify source of error.

3.4.1. Internal closure

In GODFIT GDP5, it is possible to use an internal or an external closure for the surface albedo. For the internal closure, tropospheric aerosol scattering and absorption and surface reflectivity are brought together in an *internal albedo closure term* that is fitted internally in the sense that coupling between the surface and tropospheric scatterers is treated properly in a full multiple scattering context. Direct fitting makes a simultaneous determination of an *effective* wavelength-dependent albedo in a *molecular* atmosphere, thus avoiding introduction of a lot of extraneous (and highly uncertain) aerosol information into the modeling. Assuming the surface albedo $R(\lambda)$ is a 3-parameter quadratic function of wavelength λ, we write:

$$R(\lambda) = \gamma_0 + \gamma_1(1-\lambda/\lambda_0) + \gamma_2(1-\lambda/\lambda_0)^2; \qquad (10)$$

This allows us to define Jacobians with respect to fitted parameters γ_0, γ_1 and γ_2:

$$K_{\gamma_0}(\lambda) = K_R(\lambda);\ K_{\gamma_1}(\lambda) = (1-\lambda/\lambda_0)K_R(\lambda);\ K_{\gamma_1}(\lambda) = (1-\lambda/\lambda_0)^2 K_R(\lambda) \qquad (11)$$

where $K_R(\lambda)$ is the LIDORT albedo weighting function output. We assume first guess values $\gamma_1 = \gamma_2 = 0$, and an initial value for γ_0 is taken from a suitable database (see section 2.3.3). Sensitivity results in phase A have shown that the use of albedo closure makes the algorithm insensitive to the presence of aerosols, even with absorbing aerosols in the lowest layer.

GODFIT uses the independent pixel approximation for partially cloudy scenes. Clouds are treated as Lambertian reflecting surfaces (CRB), with cloud albedo and cloud-top height derived from the ROCINN cloud pre-processing algorithm (section 6.2.3). The use of an internal closure becomes problematic for strongly cloud-contaminated scenes (cloud fraction

> 0.85) using the CRB model, because of the lack of signal from the surface and lower atmosphere below cloud-top. The solution in this case is to ignore any surface contributions, treating the scene as fully cloudy (cloud fraction 1.0) with the cloud-top treated as a reflecting surface characterized by the internal albedo closure. In this case, the first parameter in Eq. (10) is the *retrieved* cloud-top albedo.

In addition, it has been shown that the surface albedo closure becomes unstable under snow and ice conditions. The cloud fraction delivered by the OCRA algorithm (section 4.1) is unreliable in this case, as OCRA cannot really distinguish between snow/ice and clouds. To remedy this, a "snow/ice mode" has been introduced in OCRA/ROCINN: this mode is activated when the UV surface albedo $\alpha_{UV} > 0.2$ (α_{UV} is taken from a dedicated albedo data set – see below section 3.4.3). In this mode, the cloud fraction is set to 1.0, and the cloud-top albedo and the cloud-top height are the retrieved "scene" albedo and height respectively. See also the discussion on the snow/ice mode in section 4.1.2.

3.4.2. External closure

In GODFIT Phase B and GDP5, there is now an option to use an ***external albedo closure term***. This was absent in Phase A, though it has always been required for DOAS algorithms. The main motivation is due to the new requirement for the GODFIT system to ingest data with significant radiometric calibration uncertainty or even with unknown radiometric calibration; this applies particularly to ground-based data.

The external closure works in a similar manner to that in DOAS, except that a full radiative transfer simulation replaces the Beer-Lambert model. First, the LIDORT simulation is performed with a fixed albedo A_0; there is no Jacobian with respect to this variable. Then the molecular Ring correction is added to the LIDORT output – call this result the corrected LIDORT radiance $I_{LC}(\lambda;A_0)$. The simulated TOA radiance is then written:

$$I_{sim}(\lambda) = I_{LC}(\lambda; A_0)R(\lambda) \tag{12}$$

We take the polynomial $R(\lambda)$ in this equation to be quadratic as in Eq. (10) above, with external fitted parameters γ_0, γ_1 and γ_2 having the following weighting functions:

$$K_{\gamma_J}(\lambda) = (1-\lambda/\lambda_0)^{J-1} I_{LC}(\lambda; A_0) \tag{13}$$

Note that this external closure is applied after the molecular Ring correction, but before the addition of undersampling.

GDP5 baseline. The original intention was to use internal closure for operational processing. However, there have been some calibration and degradation issues with recent GOME data before the ERS-2 decomissioning in summer 2011. Following the final Delta-validation to test the closure choice, it was found that the use of external closure gives the best results overall. This is therefore the baseline.

3.4.3. Albedo data sets

The static surface albedo climatology used in GDP 3.0 [*Matthews*, 1983; *Bowker et al.*, 1985] has been replaced in GDP 4.x and GODFIT with a dynamic albedo data set derived from accumulated satellite reflectance data. In GODFIT Phase A, the GOME Lambertian equivalent reflectivity (LER) data set of albedos prepared from 5.5 years of reflectivity data

[*Koelemeijer et al*., 2003] was used. In GDP 4.x and now in GDP5, this data is combined with Nimbus-7 TOMS LER values (N7-TOMS) prepared from 14.5 years of data from 1978 [*Herman and Celarier*, 1997], and valid for 340 and 380 nm.

The GOME LER data has monthly and yearly entries on a 1°x1° latitude/longitude grid, at 12 different wavelengths spanning the GOME range; the N7-TOMS data is also monthly. We use GOME LER data at 335 and 380 nm, and N7-TOMS LER data at 380 nm; the desired combination albedo is $a(\lambda) = s(\lambda)a_{TOMS}(380)$, with scaling is $s(\lambda) = a_{GOME}(\lambda)/a_{GOME}(380)$, and λ = 335 nm for total ozone fitting [*Boersma et al*., 2004]. In this way, the strengths of both data sets are combined: the long duration of the TOMS record (1978-1992) and the spectral information (11 wavelengths) of the shorter GOME record (1995-2001).

For the snow/ice mode activation in OCRA/ROCINN, we use the "UV albedo" data set from [*Tanskanen*, 2004]. This is a global UV range climatology with fine spatial (1°x1°) and temporal (daily) resolution and realistic high-latitude values. It has been created from the TOMS 360 nm reflectivity time-series data (NIMBUS-7) by application of the "moving time-window" method [*Tanskanen et al.*, 2003].

3.5. Other atmospheric constituents

Nitrogen Dioxide. In GDP5, NO_2 is considered as an interfering trace species and it has been modeled in the GODFIT algorithm. Number density profiles were taken from climatology prepared from stratospheric HALOE data [*Lambert et al*., 2000]. GOME FM 98 NO_2 cross-sections [*Burrows et al.*, 1998] are appropriate, with temperature interpolation done linearly. The climatology is used as a first guess for the total column, if this is considered worthy of fitting. In this case, the easiest way to define the profile-column map is to ensure that the profile scales equally at all levels when deriving a total column Jacobian; profile shape will not change. Thus, the retrieval parameter is the profile-scaling factor, not the actual column itself (the latter can be restored easily enough).

It is possible to use data from a chemistry model (for example, the GEOSCHEM 3-D chemistry model) to develop parameterized tropospheric profiles of NO_2 to be grafted smoothly on to the stratospheric HALOE data. This is important for the extension of GODFIT to visible wavelength windows and minor trace species. However, the GODFIT baseline for validation testing does not include NO_2 as an interfering species in the total ozone direct fitting algorithm, as there have been some issues with the T-shift fitting in the presence of NO_2. This baseline also applies to GDP5 - no NO_2 in the fitting.

Other trace species. Although the focus is on ozone, the GODFIT algorithm has the capability to retrieve other trace species absorbing in the UV and visible. In particular, this includes the species BrO, HCHO and OClO in the UV-B range and in the range below 320 nm, SO_2 can be retrieved additionally to ozone. As with NO_2, the retrieval is based on a profile-column map that is essentially a uniform scaling of a given profile of the trace species in question. It is worth noting again that retrievals of NO_2, BrO, HCHO and OClO have so far been done using DOAS options within GODFIT.

The situation with SO_2 is a little different. The atmosphere is not optically thin at wavelengths below 320 nm at the beginning of GOME Band 2a. Although DOAS-style fitting for SO_2 and O_3 in this region has been done with GOME data (see for example

[*Thomas et al*., 2005]), it has been shown that retrieval accuracy for SO_2 is seriously compromised because of the Beer-Lambert assumption. GODFIT offers the possibility to circumvent this problem with the use of full radiative transfer simulations of backscatter radiances. A combined direct fitting of SO_2 and O_3 columns is currently under investigation.

Aerosols. There is limited information on aerosol properties to be gleaned from column ozone retrieval in the UV (absorbing aerosols are an exception in that their presence can be detected, but even here it is sometimes difficult to derive their optical properties). As noted already, the baseline for GODFIT is to dispense with explicit knowledge of any aerosols in the fitting, and perform RT simulations in a Rayleigh atmosphere with an adjustable surface albedo closure polynomial. However, aerosol data are required for any investigation of aerosols as sources of model parameter error. In GODFIT, we take the MODTRAN aerosol data sets [*Kneizys et al*., 1989] as a reference; these provide aerosol loading and optical properties. All aerosol inputs are linearly interpolated to the clear sky and cloudy sky height grids; aerosol scattering distributions were assumed to follow the Henyey-Greenstein phase function law.

4. FORWARD MODEL RADIATIVE TRANSFER

4.1. Forward model: Optical property setups

4.1.1. Atmospheric optical property inputs

In GDP5, the simulation of earthshine radiances and parameter Jacobians is done using the multi-layer multiple scattering radiative transfer code LIDORT version 3.3 ([*Spurr*, 2008] and references therein). Like the DISORT discrete ordinate scattering code [*Stamnes et al.*, 1988; *Thomas and Stamnes*, 1999], LIDORT is a pure scattering code: it does not distinguish individual atmospheric distributions. LIDORT requires as input a set of inherent optical properties (IOPs) of the atmosphere. The atmosphere is assumed stratified, with a number of optically uniform sub-layers. The total phase function is assumed to have a Legendre polynomial expansion in the cosine of the scattering angle; the phase function is normalized to 4π.

The IOP inputs for each layer n are the total extinction optical thickness Δ_n, the total layer single scatter albedo ω_n, and the set of Legendre expansion coefficients β_{ln} for the total phase function. In GDP5, the IOPs are constructed from atmospheric profiles of temperature, pressure, air density and trace gas distributions and knowledge of Rayleigh (molecular) scattering parameters plus trace gas absorption cross-sections. [Aerosol loadings and optical properties are ignored; see the discussion on internal albedo closure in the previous section].

The two free parameters are the total column C and the T-shift S. In hydrostatic equilibrium, the air column density A_p in layer p will depend on S through the formula in section 3.2.2. For both the TOMS and NNORSY implementations, ozone partial columns U_n depend only on C, and temperature-dependent ozone cross sections σ_p^{O3} depend only on S. The Rayleigh scattering coefficient σ^{Ray} will be determined from a standard formula (both GODFIT and GDOAS use the recent formulation as given in [*Bodhaine et al.*, 1999]); there is no dependence on C and S. We can then write:

$$\Delta_n(C,S) = \sigma^{Ray} A_n(S) + \sigma_n^{O3}(S) U_n(C)\,; \tag{14}$$

$$\omega_n(C,S) = \frac{\sigma^{Ray} A_n(S)}{\Delta_n(C,S)}\,; \tag{15}$$

For molecular (Rayleigh) scattering, the phase function has a $cos^2\Theta$ dependence on scattering angle Θ, with 3 phase function Legendre moments $\beta_0 = 1$, $\beta_1 = 0$, $\beta_2 = (1+\rho)/(2+\rho)$, where the depolarization ratio ρ is taken from [*Chance and Spurr*, 1997].

The GODFIT forward model requires partial derivatives of the TOA radiance with respect to the parameters C and S. Since the choice of C uniquely determines a partial column profile $\{U_n\}$, and the choice of S determines the air density and ozone cross-section profiles, then we will require a set of *linearized IOP inputs* to LIDORT for the generation of Jacobians with respect to these parameters. Differentiating equations (14) and (15), we find:

$$\left.\begin{aligned}\frac{\partial \Delta_n(C,S)}{\partial C} &= \sigma_n^{O3}(S)\frac{\partial U_n(C)}{\partial C};\\ \frac{\partial \Delta_n(C,S)}{\partial S} &= \frac{\partial \sigma_n^{O3}(S)}{\partial S}U_n(C) + \sigma^{Ray}\frac{\partial A_n(S)}{\partial S}.\end{aligned}\right\} \quad (16)$$

$$\left.\begin{aligned}\frac{\partial \omega_n}{\partial C} &= -\frac{\omega_n}{\Delta_n}\frac{\partial \Delta_n}{\partial C};\\ \frac{\partial \omega_n}{\partial S} &= \frac{1}{\Delta_n}\left[\sigma^{Ray}\frac{\partial A_n}{\partial S} - \omega_n\frac{\partial \Delta_n}{\partial S}\right].\end{aligned}\right\} \quad (17)$$

In the Rayleigh-only case (no clouds/aerosols), there are no derivatives of β_{lp} (this is not the case for an atmosphere with clouds or aerosols). Linearized IOP inputs may be defined for other model parameters (such as the aerosol profile) that are sources of uncertainty in the fit; for examples appropriate to ozone modeling in the UV, see [*van Oss and Spurr*, 2002].

4.1.2. Surface and cloud setups; the IPA and CRB

For all LIDORT calculations, the other major input to LIDORT is the ground surface optical property. The GODFIT/GDP5 default assumes a Lambertian surface characterized by a total albedo *R*, but a full BRDF formulation for surface reflectivity has been written into the LIDORT code and can be used for sensitivity and error study. Lambertian surface albedo weighting functions have been part of LIDORT from the outset. The GODFIT implementation was discussed in section 3.4. Further details on surface property weighting functions (including BRDFs) may be found in [*Spurr*, 2004].

The GODFIT and GDP5 baseline is to use the independent pixel approximation (IPA), in which TOA radiance in a partially cloudy scenario is simulated as a linear combination of backscatter from clear and fully cloudy scenes, weighted by the *effective* cloud fractional cover f_c. Additionally, clouds are treated as Lambertian reflecting boundary (CRB) surfaces. While it is well known that the IPA+CRB is not a realistic treatment of partially cloud-filled scenarios, it should be remembered that most ozone is above cloud-top level. Clouds are therefore a first-order correction to the basic ozone retrieval.

Cloud optical properties (in GDP 4.4 and GDP5) come from the OCRA/ROCINN algorithm Version 2.0 [*Loyola et al.*, 2009]. OCRA provides the cloud fraction, with cloud top albedo and height obtained through ROCINN (see Section 6.2). In this approximation, clouds are treated as Lambertian reflectors. However, it has recently been shown that the use of a clouds-as-layers scheme (treating clouds as Mie scattering water droplets) in UPAS GDP4.x total ozone retrieval can lead to improved total ozone products.

It is well known that the use of a ghost column (climatological ozone below cloud-top) in the DOAS total ozone algorithms can lead to a significant overestimation of the total column. In addition, there is a vertical ozone intra-cloud correction (VIC) to the ghost column that has been implemented in UPAS. This is an empirical correction designed to allow for intra-cloud ozone absorption [*Loyola*, 2007]. It has been shown that the use of this correction virtually eliminates any cloud dependency in the DOAS total ozone product.

A separate "ghost column" calculation is not needed in GDP5/GODFIT (except for snow and ice mode, see below), since the unique column/profile map ensures that the tropospheric part of the profile column below cloud-top is automatically adjusted in the fit. Consequently, the VIC is not applicable to GODFIT. Instead, we define effective OCRA/ROCINN products as follows. If X is the cloud-top albedo (CTA) and Y the cloud fraction (CF), then the effective values are:

$$\left.\begin{array}{l} CTA^* = X \\ CF^* = Y \end{array}\right\} \text{ for } X > 0.8; \qquad \left.\begin{array}{l} CTA^* = 0.8 \\ CF^* = \dfrac{Y.X}{0.8} \end{array}\right\} \text{ for } X \le 0.8 \tag{18}$$

One other thing deserves to be noted here. For scenes that have been designated as snow- or ice-covered, the cloud algorithm cannot distinguish between the ground and the cloud. Thus in "ice mode", the cloud fraction is set to 1.0, and ROCINN retrieves the effective scene albedo and reflecting surface height. Because of the highly reflecting surface, the effective light path in the cloud is enhanced by multiple reflections, and there is a danger of ozone column overestimation if the ghost column is used. The use of a ghost column for the ice mode was disabled in GDP 4.x, and following some ozone-hole studies with GODFIT, the ghost column was also omitted in this mode.

4.2. Forward model: LIDORT scattering code

4.2.1. Introductory remarks on LIDORT and VLIDORT

The linearized discrete ordinate radiative transfer code LIDORT is the *radiative transfer scattering model* for GODFIT forward model simulations. The original Phase A algorithm was based on earlier "pseudo-spherical" versions of this model [*Spurr* 2002], while in Phase B, a single scatter correction was used additionally for the outgoing line-of-sight, based on LIDORT Version 2.2+ [*Spurr*, 2003]. The key innovation for GODFIT was the development of an analytic total column weighting function. Recently, LIDORT has been given a number of performance enhancements, and the latest version (3.3) is based on the specification in a recent review paper [*Spurr*, 2008].

In *ozone profile* retrievals, the polarization correction applied to Level 1 data is an important source of error; the inclusion of polarization in the RT simulations is also an important consideration. Polarization correction look-up tables for the RT simulations for OMI and GOME ozone profile algorithms have been created [*Hasekamp et al*., 2002, *Liu et al.*, 2005]. However, in a DOAS fitting with narrow windows in the range 325-335 nm, the polarization signature is subsumed in the low-filter external closure polynomial; polarization is neglected in AMF RT calculations. In GODFIT, the polarization default is the same – any polarization effects are subsumed by the albedo closure treatment.

However, it is desirable to look at polarization effects in more detail, for two reasons. First, it is possible in direct fitting to use windows with wavelengths down to 312 nm at the beginning of GOME Band 2a, and this would require inclusion of polarization effects. Secondly, neglect of polarization is an important source of forward model error (the scalar approximation can give up to 10% errors under favorable conditions [*Mischenko et al*., 1994; *Lacis et al*., 1998]).

GODFIT partner Robert Spurr has developed a complete linearized vector discrete model VLIDORT [*Spurr*, 2006]. This is the counterpart to the scalar LIDORT code, and VLIDORT Version 2.2 has all the capabilities of its scalar counterpart LIDORT Version 3.3; VLIDORT can be executed in scalar (intensity-only) mode and is then identical in function to LIDORT. For the most up to date implementation, see [*Spurr*, 2008].

In 2006, the GODFIT environment at BIRA-IASB was extended to include an interface to VLIDORT and it is now possible to make a complete scalar/vector forward model error characterization for the algorithm. In addition, VLIDORT was used to generate reflectance templates for the ROCINN version 2.0 cloud-property retrieval algorithm; this development is reported in section 5.2.3. VLIDORT is not part of GDP5 in UPAS; there is only an interface for LIDORT Version 3.3 (see section 6.7 for more details on the UPAS implementation).

4.2.2. Solving the RTE in LIDORT

The RTE for unpolarized light of intensity I in a single layer is

$$\mu \frac{\partial I(\Omega, x)}{\partial x} = I(\Omega, x) - \frac{\omega}{4\pi} \int P(\Omega, \Omega') I(\Omega', x) d\Omega' - \frac{\omega}{4\pi} P(\Omega, \Omega_0) FTe^{-\eta x}. \qquad (19)$$

We use x for optical thickness measured downwards from the top of the layer. $\Omega = \{\mu, \phi\}$ is the directional variable, where μ is the cosine of the polar angle, ϕ is the azimuth angle between planes containing incident and scattering beams, and P is the phase function for scattering from direction Ω' to Ω. The integral term in the source function is the multiple scatter contribution (diffuse radiation), and the primary scattering of direct sunlight is proportional to flux F. The solar direction is $\Omega_0 = \{-\mu_0, \phi_0\}$ with angles defined at the bottom of the atmosphere; the factor η in the primary scatter attenuation is an average secant determined by ray tracing through a spherical-shell atmosphere (the pseudo-spherical approximation), and T is the solar beam transmittance to the top of the layer.

The first step in the solution of Eq. (19) is the expansion of I and P in terms of a Fourier series in the cosine of the relative azimuth $\phi - \phi_0$. In the resulting equation for the Fourier component $I_m(x,\mu)$, the multiple scatter integral is replaced by a summation using a double Gauss-Legendre quadrature scheme. This results in a set of $2N$ coupled *linear* differential equations for the discrete ordinate intensities $I_m(\pm\mu_j)$, where N is the number of quadrature streams in the half-space and $\pm\mu_j$, $j = 1 \ldots N$ are the discrete ordinate polar directional cosines with quadrature weights a_j, $j = 1 \ldots N$:

$$\mu_j \frac{\partial I_m(\mu_j)}{\partial x} = I_m(\mu_j) - \frac{\omega}{2} \sum_{k=\pm 1}^{\pm N} a_k \Pi(\mu_j, \mu_k) I_m(\mu_k) - \frac{\omega F}{4\pi} \Pi(\mu_j, -\mu_0\} FTe^{-\eta x}. \qquad (20)$$

The homogeneous equations in Eq. (20) are solved by standard eigenvalue methods [*Stamnes et* al, 1988], while the particular integral due to the solar beam source term is determined either by exponential substitution [*Thomas and Stamnes*, 1999], or by Green's function methods [*Siewert*, 2000; *Spurr*, 2002].

Solutions for all layers are then fixed through the application of three boundary conditions: (1) no downwelling diffuse radiation at TOA; (2) a surface reflection condition at the lower

boundary linking upwelling and downwelling intensities there; and (3) continuity of the radiation field at all intermediate layer boundaries. This results in a sparse linear matrix algebra problem AX = B for the unknown vector of integration constants X; solutions are found using the LAPACK numerical packages. This completes the discrete ordinate solution.

To obtain the field at arbitrary direction μ, we "post process" the solution by means of the source function integration technique [*Chandrasekhar*, 1960]. Here, we substitute the discrete ordinate solution into the original RTE and (depending on which level in the atmosphere we require the intensity) integrate over layer optical thickness values. For TOA radiances:

$$I(0,\Omega) = I_{surf}(\Psi_n,\Omega)e^{-\Psi_n/\mu} + \sum_{p=1}^{n}\Lambda_p(\Omega)e^{-\Psi_p/\mu}, \qquad (21)$$

where Λ_p are the integrated layer source terms for layer p, $\exp(-\Psi_p/\mu)$ is the line-of-sight attenuation factor for cumulative optical depth Ψ_p, and I_{surf} is the upwelling radiation at the bottom of the atmosphere (n layers; total optical thickness Ψ_n). I_{surf} follows directly from the surface boundary condition. Summing the Fourier azimuth cosine series then completes the solution; an accuracy criterion is applied to limit the number of Fourier terms.

4.2.3. Jacobian output from LIDORT

The discrete ordinate RT solution is analytically differentiable [*Spurr*, 2002]. We may carry out an explicit differentiation of the compete radiation field (both the discrete ordinate solution and the post-processed result) by repeated applications of the chain rule to each of the component parts of the solution, always expressing the answers in terms of the basic set of linearized IOP inputs.

First, differentiation of the discrete ordinate homogeneous and particular solutions may be done in a straightforward manner [*Spurr*, 2002]. Differentiation of the boundary value problem yields A.dX = dB + dA.X for the derivatives dX of the integration constant vector. This is essentially the same as the original linear problem: the inverse matrix has already been determined, and only the right hand source vector is different. It is possible to follow through the source function integration completely by explicit differentiation and repeated applications of the chain rule as before. In this manner, the weighting functions can be determined exactly and analytically with no additional *numerical* computation. Further details of the linearization and differentiation of the discrete ordinate solutions may be found in [*van Oss and Spurr*, 2002].

Older versions of LIDORT were written to deliver a set of *profile* Jacobians; a *column* weighting function would then be created *externally* from the following chain-rule sum:

$$K_{col}(\Omega) \equiv \frac{\partial I(0,\Omega)}{\partial C} = \sum_{p=1}^{n}\frac{\partial I(0,\Omega)}{\partial U_p}\frac{\partial U_p}{\partial C} \quad . \qquad (22)$$

Version 2.5 of LIDORT was written for the phase A GODFIT study - this performs the summation in Eq. (2.17) *internally*, and is much more efficient. This version of LIDORT is designed to output <u>*bulk property*</u> weighting functions (derivatives with respect to whole-atmosphere properties such as total column amounts. In this regard, the T-shift Jacobian is

also a bulk property weighting function. All later versions of LIDORT have a dual profile and bulk property Jacobian facility.

Finally, we note that for the albedo Jacobian, the layer discrete ordinate homogeneous and particular solutions have no dependence on the surface. Thus for the surface property Jacobian, we start with the linearized boundary problem and then work through differentiation of the post-processed solution.

4.2.4. Sphericity corrections

The use of the pseudo-spherical (P-S) approximation (solar beam attenuation treated for a curved atmosphere, but scattering remains plane-parallel) in the RT theory is essential for dealing with solar zenith angles in excess of 70-75° [*Dahlback and Stamnes*, 1991]. This is a standard feature of many RT models; for a discussion of the LIDORT and VLIDORT implementations, see for example [*Spurr*, 2008]. The P-S treatment is sometimes called the *incoming sphericity* correction for the solar beam. In the P-S approximation, scattering is assumed to take place along the nadir, with solar and line of sight viewing angles unchanged.

We now consider an *outgoing sphericity correction*. In a curved atmosphere, both solar and viewing angles vary along the line-of-sight path from the bottom of the atmosphere to the satellite. For atmospheric layer *n* traversed by the line-of-sight path, the upwelling radiance at the layer-top is (to a high degree of accuracy) given by:

$$\mathbf{I}^{\uparrow}(\Omega_{n-1}) \cong \mathbf{I}^{\uparrow}(\Omega_n)T(\Omega_n) + \mathbf{\Lambda}_n^{\uparrow}(\Omega_n) + \mathbf{M}_n^{\uparrow}(\Omega_n)\,. \qquad (23)$$

Here, $\mathbf{I}^{\uparrow}(\Omega_n)$ is the radiance at the layer bottom, $T(\Omega_n)$ the layer transmittance along the line of sight, and $\mathbf{\Lambda}_n^{\uparrow}(\Omega_n)$ and $\mathbf{M}_n^{\uparrow}(\Omega_n)$ are the single- and multiple-scatter layer source terms respectively. The transmittance and layer source terms are evaluated with geometry Ω_n at the bottom of the layer. Equation (23) is applied recursively, starting with the upwelling intensity $\mathbf{I}_{SURF}^{\uparrow}(\Omega_{SURF})$ evaluated at the surface, and finishing with the intensity field at top of atmosphere ($n = 0$). The transmittances and single-scatter layer source terms may be determined through an accurate single scatter calculation allowing for changing geometrical angles along the line of sight.

To evaluate the multiple scatter sources, we run LIDORT in "multiple-scatter mode" successively for each of the geometries from Ω_{SURF} to Ω_1, retaining only the appropriate multiple scatter layer source terms, and the surface upwelling intensity $\mathbf{I}_{SURF}^{\uparrow}(\Omega_{SURF})$. This is the *full outgoing sphericity correction*. To implement this, we need N_L separate calls to LIDORT for a total number of layers, and this is much more time consuming that a single P-S call with geometry Ω_{SURF} (this is the default in the absence of a line-of-sight correction).

However, since scattering is strongest near the surface, the first LIDORT call (with geometry Ω_{SURF}) is the most important. Thus, a simpler line-of-sight correction is to assume that *all* multiple scatter source terms are taken from this first LIDORT call; in this case, we require only the accurate single scatter calculation to complete $\mathbf{I}_{TOA}^{\uparrow}$. This approximation is known as the "poor man's" sphericity correction; it requires very little extra computational

effort compared to a single call with the regular P-S geometry. The sphericity correction can also be set up with just two calls to VLIDORT made with the start and finish geometries Ω_{SURF} and Ω_1; in this case, multiple scatter source terms at other geometries are *interpolated* at all levels between results obtained for the two limiting geometries.

Accuracies for these types of outgoing sphericity corrections were investigated in [*Spurr* 2003] for the scalar code; results for VLIDORT are similar. In general for line-of-sight view angles not too far from zenith (30° or less), the regular P-S approximation gives sufficient accuracy, but for wider off-nadir viewing the outgoing sphericity correction is necessary [*Caudill et al.*, 1997; *Spurr* 2003; *Rozanov et al*., 2000]. Thus for regular cross-track GOME viewing (maximum swath 960 km), the P-S approximation is accurate enough, but for the GOME polar view mode [*ESA*, 1995] and for other instrument such as GOME 2 (swath 1920 km) [*Munro et al*., 2006] and OMI ([*Levelt et al*., 2006], swath width 2600 km) the outgoing correction is necessary.

GDP5 Default. Both GDP4.4 and GDP5.0 have LIDORT Version 3.3 as the baseline. This has the "poor man's" sphericity as an automatic procedure. The model first makes an accurate single scatter calculation along a curved line of sight, taking the BOA (bottom of atmosphere) geometry as the starting point. Note that the single scatter computation uses a fine-layer subdivision of the coarse layering structure of the atmosphere in order to integrate the RTE; the default is to use 2 subdivisions, as this gives sufficient accuracy for GOME.

5. INVERSE MODEL

5.1. GODFIT Phases A and B: Optimal Estimation

GODFIT GDP5 is a direct fitting technique, using iterative non-linear least squares minimization with some degree of regularization. There are many variants for this kind of inverse function, and in GODFIT Phase A/B, the familiar "optimal estimation" methodology was used, with loose *a priori* constraints on the state vector elements. However, some faster alternatives were investigated during the first phase of the GDP5 project, and variant of the Levenberg-Marquardt least squares algorithm with line searching was selected and implemented for operational use in GDP5, and this algorithm is discussed in section 5.2. First, we summarize the optimal estimation method [*Rodgers*, 2000] for GODFIT Phase A/B.

Classical least squares fitting minimizes a chi-square merit (or cost) function. In an ill conditioned retrieval problem, the number of fitting parameters exceeds the number of pieces of independent information that can be extracted from the fit. We deal with this by adding a regularization term to the cost function. Thus:

$$\chi^2 = (\mathbf{Y}_m - F(\mathbf{X}))^{\mathbf{T}} \mathbf{S}_y^{-1} (\mathbf{Y}_m - F(\mathbf{X})) + (\mathbf{X} - \mathbf{X}_\mathrm{a})^{\mathbf{T}} \mathbf{S}_\mathrm{a}^{-1} (\mathbf{X} - \mathbf{X}_\mathrm{a}) . \tag{24}$$

Here, we have the measurement vector of TOA radiances $\mathbf{Y}_m$, the state vector $\mathbf{X}$, the forward model radiance simulations $F(\mathbf{X})$, and the error covariance matrix $\mathbf{S}_y$.. $\mathbf{X}_\mathrm{a}$ is the *a priori* state vector, with $\mathbf{S}_\mathrm{a}$ the corresponding covariance matrix. For a forward model with non-linear dependence on atmospheric parameters, the fitting is non-linear least squares (NLLS), proceeding iteratively via a series of linearizations about the atmospheric state at each iteration step:

$$\left.\begin{aligned} &\mathbf{X}_{i+1} = \mathbf{X}_\mathrm{a} + \mathbf{D}_y \left[\mathbf{Y}_m - F(\mathbf{X}_i) - \mathbf{K}_i (\mathbf{X}_i - \mathbf{X}_\mathrm{a})\right]; \\ &\mathbf{D}_y = \mathbf{S}_{i+1} \mathbf{K}_i{}^{\mathbf{T}} \mathbf{S}_y^{-1} \quad \text{and} \quad \mathbf{S}_{i+1} = \left(\mathbf{K}_i{}^{\mathbf{T}} \mathbf{S}_y^{-1} \mathbf{K}_i + \mathbf{S}_a^{-1}\right)^{-1} . \end{aligned}\right\} \tag{25}$$

$\mathbf{K}_i = dF(\mathbf{X}_i)/d\mathbf{X}_i$ is the matrix of Jacobians, $\mathbf{D}_y$ is the matrix of contribution functions, and $\mathbf{S}_{i+1}$ is the solution covariance matrix. The latter is the main diagnostic output. The iteration stops when one or more convergence criteria are met. The computation proceeds efficiently when one uses an SVD (singular value decomposition) on the (scaled) matrix Jacobians; for details, see for example [*Van Oss et al*., 2002]. In terms of the singular values Λ_n resulting from this decomposition, the DFS (degrees of freedom for signal) indicator is:

$$DFS = \sum_{n=1}^{NP} \frac{\Lambda_n^2}{1 + \Lambda_n^2} . \tag{26}$$

DFS is a useful measure of the number of independent linear combinations of state vector elements that can be retrieved. If N_P is the dimension of the state vector, then DFS = N_P when the measurement completely determines the state.

We express the optimal estimate in terms of a true value, plus errors [*Rodgers*, 2000]:

$$\mathbf{X} \approx \mathbf{X}_\mathrm{true} + (\mathbf{A} - \mathbf{I})(\mathbf{X}_\mathrm{true} - \mathbf{X}_\mathrm{a}) + \mathbf{D}_y \boldsymbol{\varepsilon}_y . \tag{27}$$

The matrix $\mathbf{A} = \mathbf{D}_y\mathbf{K}$ is the set of averaging kernels and it is an indicator of the sensitivity of the retrieval to the true state. The second term in (27) is the smoothing error. The remaining error $\boldsymbol{\varepsilon}_y$ can be divided into three components as follows.

Measurement errors. The measurement error due to random noise contributions is $\boldsymbol{\varepsilon}_{me} = \mathbf{D}_y\,\boldsymbol{\varepsilon}_y$ and this generates a solution covariance contribution $\mathbf{S}_{\text{noise}} = \mathbf{D}_y\mathbf{S}_y\mathbf{D}_y^{\mathbf{T}}$ which contributes to the overall retrieval solution covariance matrix.

Model parameter errors. The retrieval error due to this source of uncertainty is $\boldsymbol{\varepsilon}_{\text{par}} = \mathbf{D}_y\mathbf{K}_b\boldsymbol{\Delta}\mathbf{b}$, where $\mathbf{K}_b$ is the sensitivity (Jacobian) of the forward model to parameter $\mathbf{b}$, and $\boldsymbol{\Delta}\mathbf{b}$ is the error on $\mathbf{b}$. If the error is random with covariance $\mathbf{S}_b$, then the associated solution covariance contribution is $\mathbf{S}_{\text{param}} = \mathbf{D}_y\mathbf{K}_b\mathbf{S}_b\mathbf{K}_b^{\mathbf{T}}\mathbf{D}_y^{\mathbf{T}}$, which contributes to the overall retrieval solution covariance matrix. The estimation of these errors is greatly helped when the forward model is able to deliver Jacobians $\mathbf{K}_b$ in an efficient and accurate manner (LIDORT has this capability). Model parameters can be any atmospheric variables that are not fitted (cloud fraction and cloud top pressure, cross-section amplitudes, temperature profile entries, etc.).

Forward model errors. Here the retrieval error due to this source of uncertainty is given by $\boldsymbol{\varepsilon}_{\text{fwd}} = \mathbf{D}_y\boldsymbol{\Delta}\mathbf{F}$, where $\boldsymbol{\Delta}\mathbf{F}$ is the forward model error due either to incorrect physical assumptions in the model (neglect of polarization, omission of RRS) or to a certain level of mathematical approximation in the RT solutions (number of stratifications, number of discrete ordinates in the diffuse scattering quadrature approximation, plane-parallel scattering, etc.). This is a systematic error.

5.2. GDP5: Levenberg-Marquardt and Line-search

At DLR, several inversion methods were made available for GDP5 retrieval; these use Tikhonov regularization (constant and variable) over a series of iterations, and apply the so-called line-search algorithm at each iteration step [*Doicu et al*, 2004; *Doicu et al*., 2007]. Specifically, the method selected for GDP5 is related to the Levenberg-Marquardt algorithm, through minimization of the following functional at the k^{th} iteration step:

$$\mathsf{F}_k(\mathbf{x}) = \left\|\mathbf{G}(\mathbf{x}) - \mathbf{y}\right\|^2 + \alpha_k\left\|\mathbf{L}_k\cdot(\mathbf{x}-\mathbf{x}_k)\right\|^2 / \qquad (28)$$

Here, $\mathbf{L}_k$ is a constant and invertible square matrix acting as a constraint, and the factor α_k is a regularization parameter which determines the strength of this constraint contribution to the functional. Forward model $\mathbf{G}(\mathbf{x})$ is linearized about its value $\mathbf{F}_k = \mathbf{F}(\mathbf{x}_k)$ according to $\mathbf{G}(\mathbf{x}) = \mathbf{F}_k + \mathbf{K}_k\cdot(\mathbf{x}-\mathbf{x}_k)$, where $\mathbf{K}_k = \mathbf{K}(\mathbf{x}_k)$ is the weighting function matrix. Setting the first derivative of the Taylor series of $\mathsf{F}_k(\mathbf{x})$ around $\mathbf{x}_k$ to zero yields the next guess:

$$\mathbf{x}_{k+1} \equiv \mathbf{x}_k + \tau_k\mathbf{p}_k = \mathbf{x}_k + \left[\mathbf{K}_k^T\cdot\mathbf{S}_y^{-1}.\mathbf{K}_k + \alpha_k\mathbf{L}_k^T\cdot\mathbf{L}_k\right]^{-1}\cdot\mathbf{K}_k^T\cdot\mathbf{S}_y^{-1}.\left[\mathbf{y}-\mathbf{F}_k\right]. \qquad (29)$$

Here, subscript T again denotes matrix transpose. Eq. (29) allows us to find the search direction $\mathbf{p}_k$ (a vector with unit modulus), and the step length τ_k.

Given that the functional is minimized for a certain neighborhood around $\mathbf{x}_k$ in state vector space defined by the length τ_k, we can set up a whole series of minimum values

$\mathbf{x}_i \equiv \mathbf{x}_k + \tau_i \mathbf{p}_k$ defined by a series of step lengths $\tau_i \in (0, \tau_k)$. We then search this series for a special $\mathbf{x}_j$ such that $\mathsf{F}(\mathbf{x}_j) \leq \mathsf{F}(\mathbf{x}_i), \forall i \neq j$. This is the "line-search" procedure.

Although the line-search algorithm can in principle produce faster convergence through a smaller number of iterations, it can also lead to additional computational effort, since the line-search routine (and by default, the forward model) is usually called for all wavelengths in a given fitting window. However, performance gains can be achieved by application of line-search at a reduced number of wavelengths, and this was established for GDP5 by means of a sensitivity study.

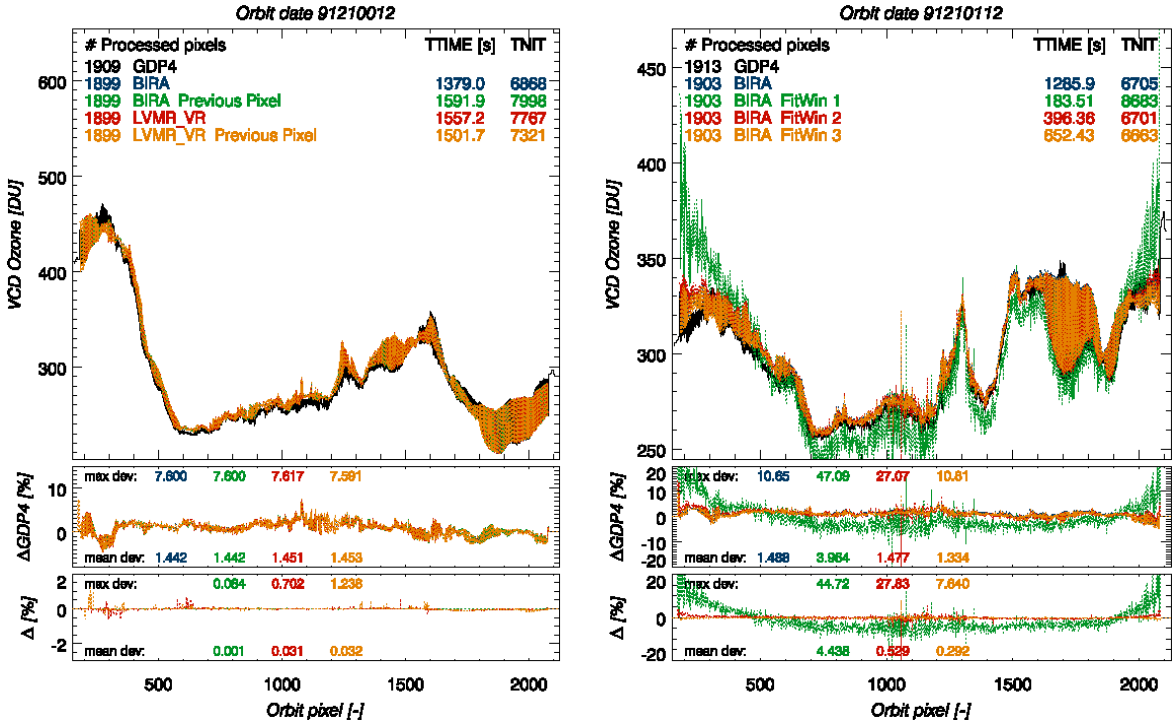


Figure 14. Examples of performance testing for inverse functions. (Left panel) comparison of GODFIT (BIRA) optimal estimation with LVMR_VR (GDP5 default) with and without previous pixel ozone value as first guess. (Right) fitting window optimization with GODFIT (no previous pixel); for windows, see text.

For the sensitivity investigations, 24 GOME orbits from 1999 were processed with GODFIT, using the NNORSY climatology (mode 1). Results from a number of trial inversion techniques and a variety of line-search wavelengths were compared with results both from the current operational GDP4.1 data and the original GODFIT inversion method. Within the 325.5 to 335.5 nm fitting window, the "line-search" wavelengths were chosen either at equidistant spacing or at values around maxima of the Huggins bands absorption cross sections. It was found that:

- When decreasing the number of wavelengths in the line-search algorithm, the number of calls to the forward model is lowered and the total processing time is reduced.
- Regularization with line-search can enhance speed of ozone column retrievals by up to 20 % without loss of accuracy.

- Accuracy and speed of processing depend not only on the choice of wavelengths for the line-search, but also on the inversion technique applied and the use of constant or variable regularization.

The sensitivity studies demonstrated that the Levenberg-Marquardt method with variable regularization (LVMR_VR) using 3 line-search wavelengths grouped closely around the 328 nm Huggins bands maximum gave the most satisfactory results. Results were found to be almost identical to those from the older GODFIT (optimal estimation) algorithm, while the processing time was lowered by 10%-15%.

Additional sensitivity studies were performed using LVMR_VR for accuracy and speed optimization. These studies included: (1) the influence of compiler options; (2) the use of external LAPACK libraries; (3) the use of the previous pixel total ozone value during orbit processing; and (4) the choice of fitting window. In general, the LVMR_VR speed can be increased by 4% (on average) due to optimized compiler options including the external library, 10% due to the total ozone previous pixel value as first guess, and up to 10% using a reduced fitting window. Thus, at least 14% and potentially 24% in processing time can be saved when applying LVMR_VR.

Figure 14 (left panel) shows results for one orbit for the previous pixel sensitivity test. Total ozone columns (here called VCDs or vertical column densities) are plotted for 5 retrievals, 2 with the original GODFIT optimal estimation inversion, 2 with the LVMR_VR algorithm and 1 with GDP4.1. Differences to the GDP 4.1 DOAS results are plotted below the main graph. Times and numbers of pixels processed are shown at the top right. The improvement with the LVMR_VR algorithm using the previous pixel is clear.

Figure 14 (right panel) is a sensitivity test for fitting window optimization, using the original GODFIT inverse model. 3 reduced fitting windows were considered: (1) 327.5 to 328.5 nm; (2) 326.5 to 329.5 nm; (3) 325.5 to 330.5 nm (half number of original grid points). Results are contrasted with the 88-point full window result. For (3), accuracy is better than 1 % compared to original BIRA result, and the mean decrease in orbit processing time is 49 % for optimal estimation and 44 % for LVMR_VR, over the 24 orbits treated in the study. A more systematic investigation for fitting window optimization is planned.

5.3. Averaging Kernel

In optimal estimation, the averaging kernel **A** is defined as the product of the contribution function matrix ($\mathbf{D}_y$ in Eq. (25)) and the Jacobian matrix **K**. It is a measure of the departure of the estimator from the truth and the dependence on *a priori* settings. The trace (diagonal sum) of **A** is the "degrees-of-freedom-for-signal" (DFS) for the retrieval; this is the same quantity as in Eq. (26). *A priori* regularization is essential for an ill-posed problem such as ozone profile retrieval from GOME measurements; the DFS (degrees of freedom for signal) for ozone in this inverse problem is typically 4.5-5.5 for a state vector with a 11-layer ozone profile [*Liu et al.*, 2005].

For the GDP5 total column retrieval, the problem is well-posed. Here, ozone is represented in the state vector as a single element for total vertical column amount. Accordingly, the averaging kernel matrix reduces to a vector that indicates the sensitivity of the retrieved total

column to changes in ozone concentration in different layers. The total column amount defines a unique profile with the column-classified TOMS climatology.

One way to establish this averaging kernel vector uses a DOAS-based approach for an optically thin absorber [*Eskes and Boersma*, 2003]. Here, AK vector values are approximated as the ratios of layer air-mass factors (AMFs) to the total AMF. For the determination of layer AMF values, we make one final call to LIDORT to generate a vector of profile layer Jacobians $\mathbf{K}^*$ for the ozone profile corresponding to the final total column. Layer AMFs are calculated at one wavelength for the optically-thin absorber case - this may not be accurate enough for absorption in the ozone Huggins bands. This procedure is independent of the inverse algorithm (either optical estimation or line-search Levenberg-Marquardt) used to retrieve the total column.

A better approach, and the one selected for GDP5, is to calculate the averaging kernel using a sum over all spectral points - this requires profile weighting functions to be calculated at every wavelength in our Huggins band window. As with the DOAS-based approach, we must call LIDORT to derive the ozone profile layer Jacobians $\mathbf{K}^*$ using that TOMS V8 ozone profile corresponding to the final retrieved value for the total column. In the off-line GODFIT algorithm using optimal estimation inversion, we use an external profile data set (either the NNORSY climatology [*Kaifel et al*., 2008], or the DOC climatology [*Weber et al*., 2005]) to supply the *a priori* profile and covariance required for the calculation of the contribution function matrix $\mathbf{D}_y$ in equation (25). For the operational line-search Levenberg-Marquardt inversion, the contribution function matrix is defined in the same way, but now the constraint term is provided by the Tikhonov regularization (c.f. Eq. (29)). The averaging kernel is then given by $\mathbf{A} = \mathbf{D}_y \mathbf{K}^*$.

6. ANCILLARY ALGORITHMS

6.1. The Molecular Ring effect

6.1.1. Introduction

The Ring effect (the filling-in of Fraunhofer solar spectral features and telluric absorption-band signatures due to rotational Raman scattering) induces a small-amplitude distortion in the earthshine spectrum [*Grainger and Ring*, 1962]. It is normally dealt with by applying an additive correction to the radiance. In the Huggins bands, the telluric component is significant, and the Ring effect distortion is large enough to seriously compromise the fitting accuracy [*Sioris and Evans*, 1999]. In DOAS-type algorithms, Ring structures are commonly dealt with by using pre-calculated *Ring spectra*, usually defined as (logarithms) of the ratios of radiances with and without rotational Raman scattering (RRS). Amplitudes for one or more selected Ring spectra (Fraunhofer and/or telluric effects) are included in the state vector; Ring scaling parameters are "pseudo absorbers". A zero-order RRS calculation can be carried out on the Fraunhofer reference spectrum at a fixed temperature [*Chance and Spurr*, 1997]; the resulting Ring spectrum is often sufficiently accurate for use as references in DOAS-type fitting of (minor species) trace gas columns. For total ozone, GDP versions up to and including 3.0 have only used a single Fraunhofer Ring spectrum.

The lack of a molecular Ring correction for GOME total ozone retrieval was recognized as one of the most important sources of error [*van Roozendael et al.*, 2002]. This was remedied in all three recent algorithms [*van Roozendael et al.*, 2006; *Eskes et al.*, 2005; *Coldewey-Egbers et al.*, 2005] arising out of the ESA ITT. At that time, the new Ring correction for the GODFIT Phase A was developed for both GDOAS and direct fitting, and the full descriptions first given in the GODFIT Phase A ATBD are still valid. In the Phase A work, the accuracy of the empirical Ring correction method was characterized by performing closed-loop tests using look-up tables of simulated radiances calculated with and without inelastic rotational Raman scattering (RRS). In Phase B, new work has focused on the use of the LIDORT-RRS RT model (including inelastic RRS treatment), not only for the validation of the empirical correction to be used operationally, but also for the provision of look-up tables to be used in the operational algorithm as an alternative to the empirical correction

6.1.2. Empirical Ring correction in GODFIT

The correction is based on a simplified forward model of the intensity $I(\lambda)$ at the satellite; this includes an explicit contribution due to RRS, as described by the following equation:

$$I(\lambda) = I^0(\lambda).\exp\left[-\sigma_{O3}(\lambda).E_{O3} - P^{\lambda}\right] + E_{Ring}.I_0^{RRS}(\lambda).\exp\left[-\sigma_{O3}(\lambda).E_{O3}^{RRS}\right]. \qquad (30)$$

The first term on the right-hand side describes elastic scattering and follows directly from the Lambert-Beer law, with $I^0(\lambda)$ the solar intensity, P_λ a low-order polynomial, and σ_{O3} and E_{O3} the ozone absorption cross-section and effective slant column respectively. The Ring effect is modeled by the second term. We may consider several approximations. First, the Raman light is assumed to be produced close to the surface, with a spectral shape given by a source spectrum for Raman scattering $I_0^{RRS}(\lambda)$. This source spectrum only treats the spectral smoothing effect of RRS on the solar intensity. In practice, it is calculated by convolution of a GOME irradiance spectrum using rotational Raman cross sections appropriate to inelastic

scattering into the wavelength of interest. The fractional intensity of Raman light (the E_{Ring} parameter) is freely adjustable; this may vary considerably and will depend on parameters such as cloud coverage, cloud altitude and surface albedo. Ozone absorption (the term $\sigma_{O3}(\lambda).E_{O3}^{RRS}$) is then treated consistently, assuming that Raman photons produced at the surface and/or above clouds travel upward to the satellite. Ozone absorption taking place in the incoming light is assumed to be fully smeared out in the inelastic process.

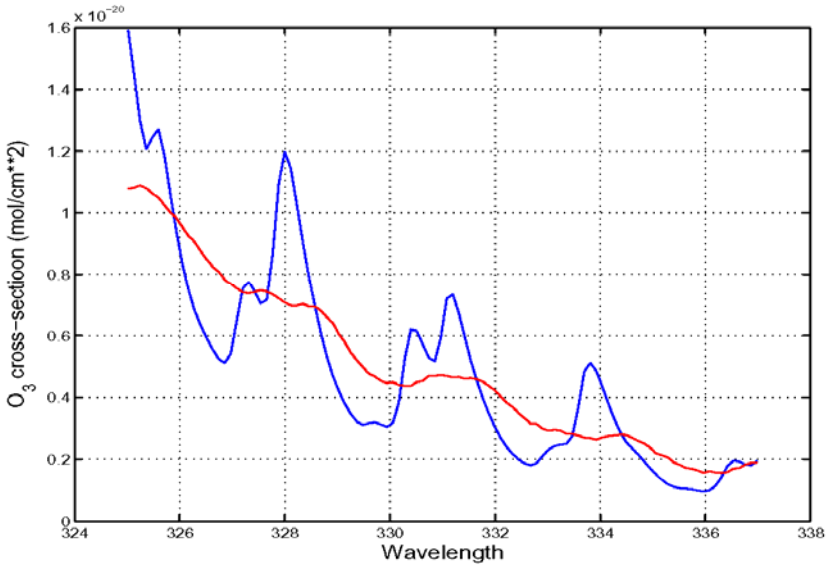


Figure 15. Regular ozone absorption cross-section at 241°K (blue line); cross-section after smoothing by the RRS process (red line)**.**

In the direct-fitting algorithm, we replace the elastic Beer-Lambert term in equation (30) by a full LIDORT sun-normalized radiance calculation:

$$\frac{I(\lambda)}{I^0(\lambda)} = \left[\frac{I(\lambda)}{F(\lambda)}\right]^{LIDORT} + E_{Ring}.I_0^{RRS}(\lambda).\exp\left[-\tau_{O3}^{RRS}\right]. \qquad (31)$$

The term τ_{O3}^{RRS} represents the effective optical density of ozone in the Raman light along the line-of-sight path from the surface (where RRS is assumed to take place) to the satellite. To a first approximation, this term can be evaluated from a simple geometrical consideration:

$$\tau_{O3}^{RRS} = \tau_{O3}^{vertical} \cdot \sec(\theta_0). \qquad (32)$$

With this equation, we recognize that ozone absorption in the Raman light can be represented with sufficient accuracy by means of a simple geometrical enhancement factor (this is generally valid for a stratospheric absorber such as ozone). We also neglect the possible impact of ozone absorption that takes place in the incident beam before the generation of Raman photons, since O_3 absorption structures are expected to be largely scrambled in the RRS process. The latter approximation is valid for most observations. However, for large solar zenith angles (> 85°), ozone absorption in the incident beam is much stronger and its impact more noticeable. To deal with this, the following more accurate definition has been used with GODFIT:

$$\tau_{O3}^{RRS} = \tau_{O3}^{vertical} \cdot \left[\sec(\theta_0) + \Phi(\theta) \cdot \frac{\sigma_{O3}^{RRS}}{\sigma_{O3}} \right], \quad (33)$$

where $\Phi(\theta)$ is the geometrical enhancement factor of the incident beam (taking into account the Earth's sphericity), θ is the solar zenith angle, θ_0 the viewing zenith angle, and σ_{O3}^{RRS} an ozone absorption cross-section for Raman scattering representative of the ozone absorption cross-section after smoothing by the Raman process. The two ozone cross-sections as used in equation (33) are illustrated in Figure 3.1.

6.1.3. The LIDORT-RRS model and RRS correction factors

The empirical approach still relies on the fitting of a Ring correction factor. A more physically realistic approach is to correct calculated earthshine radiances for the effects of rotational Raman scattering (RRS) by using a proper radiative transfer simulation of inelastic scattering. This approach has been implemented in GODFIT code (but not in UPAS GDP5).

The RT approach uses a look-up table of RRS radiance correction factors calculated using the LIDORT-RRS model [*Spurr et al*., 2008].). Correction (or "filling") factors are defined as relative differences between backscatter radiances calculated with RRS and without RRS (elastic). In GODFIT, RRS corrections for a range of solar and viewing geometries are classified as functions of physical parameters such as surface albedo and total O_3 column. A related method has been developed for the weighting-function DOAS algorithm [*Coldewey-Egbers et al*., 2005]; their LUT is based on the SCIATRAN model [*Vountas et al*., 1998].

LIDORT-RRS is first-order in Raman scattering; for a given wavelength λ, direct beam and diffuse field photons are Raman scattered once into and out of this wavelength. For UV Huggins bands, the distribution of Raman Anti-Stokes and Stokes transitions is about ±2 nm from excitation wavelength λ; further details on Raman spectroscopy can be found in [*Chance and Spurr*, 1997]. Diffuse source terms for radiation that is to be Raman scattered are computed using zero-order (elastic) radiative transfer calculations (based on Rayleigh scattering) at λ and all Raman-shifted wavelengths. Discrete ordinate solutions in LIDORT-RRS are entirely analytic, and the model is able to generate output at arbitrary viewing geometry and optical depth. LIDORT-RRS is fully compatible with LIDORT Version 3.3: it has pseudo-spherical treatments of attenuation for the incoming solar beam and the outgoing line-of-sight path, plus an exact single scatter correction procedure.

In addition to a monochromatic calculation of inelastic scattering, the LIDORT-RRS model has a fast-track "binning realization". This is based on instrumental solar irradiance spectra – fluxes are defined for wavelength bins set by the instrument pixel resolution (about 0.105 nm for GOME in the UV), and all Raman cross-sections occurring within a given bin are summed. For instruments such as GOME or SCIAMACHY, the binning calculation is much faster than an equivalent monochromatic calculation for spectral regions in the UV

LIDORT-RRS is now available in Version 2.2 which includes a full profile, total column and surface Jacobian facility. This RT model has been installed in the GODFIT environment at BIRA-IASB, and will be used to perform on-the-fly direct fitting retrievals of O_3 and SO_2. In the binning realization, LIDORT RRS is about 4 times slower than LIDORT.

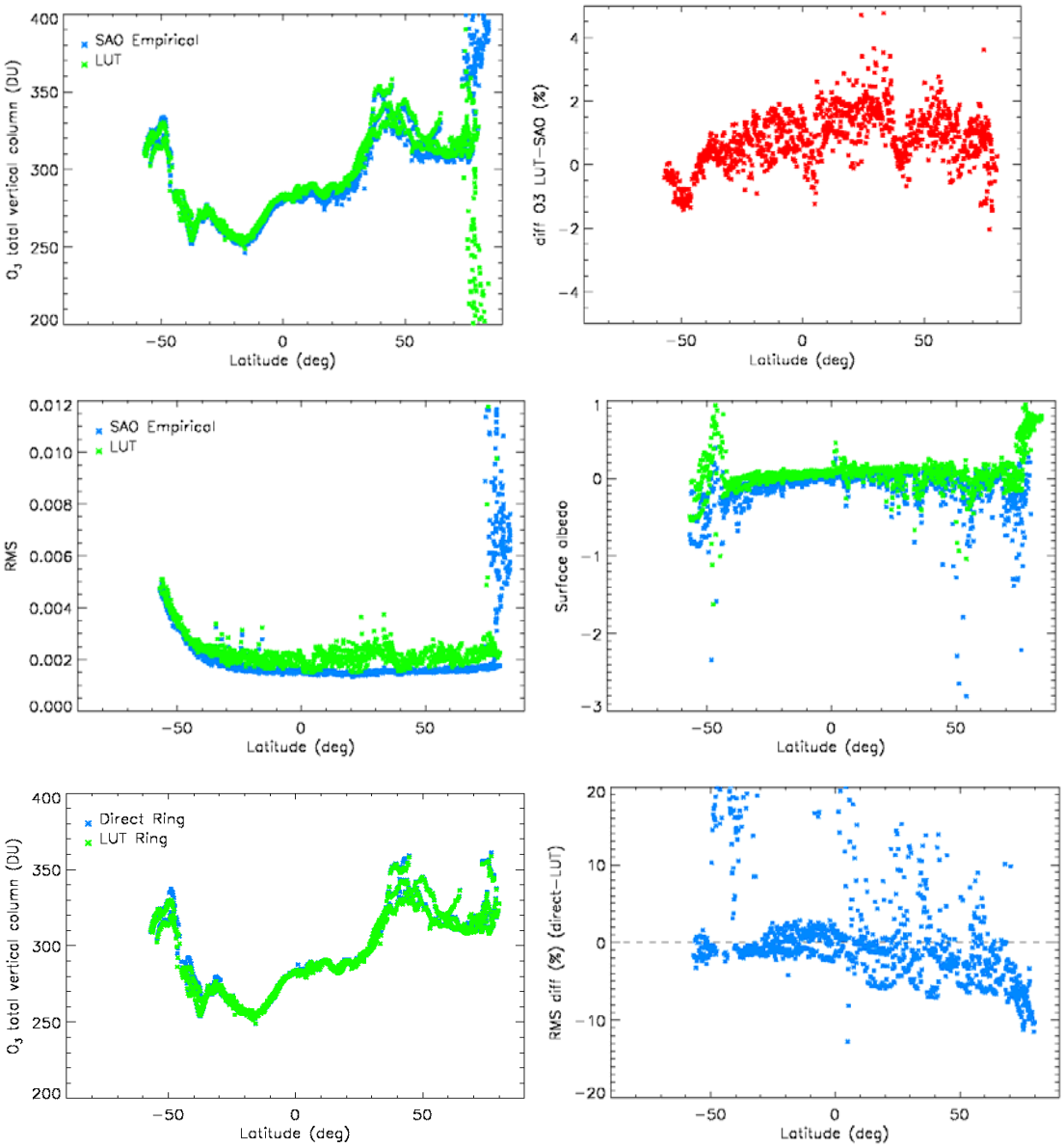


Figure 16. An orbit of GODFIT results comparing the empirical Ring correction procedure with the LUT correction (top and middle panels), and the LUT correction with an on-the-fly LIDORT-RRS simulation (lower panels). (Top left) Total ozone columns, (top right) column differences, (middle left) fitting RMS residuals, (middle right) internal closure (retrieved surface albedo), (lower left) total ozone column, and (lower right) RMS differences.

In GODFIT, the RRS corrections LUT is classified as follows; by solar zenith angle (20 values from 15 to 88°), by line-of-sight zenith angle (7 values from 0 to 55°), by relative azimuth angle (5 values from 0 to 180°), by albedos (7 values: 0.0, 0.05, 0.1, 0.2, 0.5, 0.8, 1.0), by ozone profile (TOMS V7 or V8 profiles), and by pressure of the lowest boundary (11 values from 1013 hPa to 100 hPa). All LIDORT-RRS calculations were done for a Rayleigh atmosphere; the binning realization with a GOME solar spectrum was used throughout. Although solar zenith angle and albedo are the major dependencies, lower boundary pressure is important because the bulk of inelastic scattering takes place in the lower troposphere.

Filling corrections were computed at 220 wavelength points between 315 and 335 nm. Interpolation from the LUT is done linearly (table entries are dense enough).

Up till the present, derivatives of these Ring corrections factors with respect to ozone column or albedo are determined from finite differences (these will be replaced by the analytic derivatives from Version 2.2). Results so far show that the use of look-up tables corrects for most of the RRS signal in the GODFIT fitting residuals. Initial results have shown that the quality of fit is not better than that obtained with the use of the empirical RRS correction. This may be due to intrinsic uncertainty introduced when interpolating between look-up table values, or inaccuracies due to the presence of clouds; the interpolation process is currently under further investigation .

This is seen clearly in Figure 16, where some initial results have been compiled for one orbit using the empirical approach to Ring correction ("SAO correction" in the figure) versus the LUT-derived correction. Ozone results compare well (Figure 16, top left panel), with differences generally in the 1-2% range. [The large differences for high SZA are due to an artifact in the LUT treatment, which has now been removed]. In the middle left panel of Figure 16, the RMS is slightly higher (there is one less parameter to fit). Finally the effect on surface albedo is generally positive (middle right panel) – clearly the empirical correction parameter is correlated to the fitted internal closure parameters that constitute the surface albedo (negative values are common here because of the lack of aerosol in the modeling). In the lower two panels, we compare the LUT results with an on-the-fly LIDORT-RRS calculation of the Ring correction for this orbit. It is clear that the LUT representation is doing a good job. However, there are some effects (probably cloud related) which need further investigation, as can be seen from the RMS differences.

Another option for the GODFIT direct fitting is to use a combination of the LUT and empirical approaches. The LUT correction is first applied to the LIDORT simulation of radiance, and then the remaining Ring signal is fitted using the empirical spectrum. Another possibility for combination is to use an exponential fit for a Ring spectrum amplitude, akin to the method used in the weighting-function DOAS algorithm [*Coldewey-Egbers et al*., 2005]. Both options are currently under consideration.

6.2. Cloud Pre-processing

6.2.1. Introductory remarks

All versions of GDP up to and including Version 3.0 have used the ICFA algorithm [*Kuze and Chance*, 1994] for cloud pre-processing in the IPA (independent pixel approximation). ICFA retrieves the effective cloud fraction f_c from O_2 *A* Band fitting of transmittances, but takes effective cloud-top pressure q_c from the ISCCP data base [*Schiffer and Rossow*, 1983]. In Phase A, GODFIT was programmed to ingest ICFA data. The FRESCO O_2 *A* Band least squares algorithm [*Koelemeijer et al*., 2001] is based on GOME reflectivity measurements in and around the O_2 *A* band; FRESCO supplies fitted values of f_c and q_c, plus errors on these two parameters. Results from the original FRESCO application to GOME data were supplied as data input to GODFIT in Phase A. ICFA and FRESCO assume a fixed cloud-top albedo of 0.8. However, the old Phase A options were replaced by OCRA/ROCINN.

For Phase B of the GODFIT work, the OCRA/ROCINN cloud algorithm version 1.0 first used in GDP 4.0 was employed to generate cloud properties. Version 2.0 of the algorithm was compiled successfully in the GODFIT environment in 2006. OCRA is an optical sensor cloud detection algorithm that uses the PMD devices on GOME to deliver geometric cloud fractions for GOME scenes. ROCINN takes the OCRA cloud fraction as input and uses a neural network training scheme to invert GOME reflectivities in and around the O_2 *A* band. ROCINN version 2.0 [*Loyola et at.*, 2009] retrieves cloud-top pressure and cloud-top albedo. Also in Phase B and subsequently in the sensitivity studies for GDP5, FRESCO was replaced by its upgrade called FRESCO+ [*Wang et al*., 2008].

In the next two sub-sections, the descriptions are for the most part taken from the GDP 4.0 paper [*Van Roozendael et al*., 2006], but the use of VLIDORT to generate reflectances based on full polarization scattering of light is new. In section 6.2.4, we report on sensitivity tests on total ozone from OCRA/ROCINN compared with other algorithms used for GOME and SCIAMACHY (FRESCO+, OCRA/SACURA).

6.2.2. OCRA algorithm

The basic idea in OCRA (Optical Cloud Recognition Algorithm [*Loyola and Ruppert*, 1998]) is to break down each optical sensor measurement into two components: a cloud-free background and a residual contribution expressing cloud influence. The key to OCRA is the construction of a cloud-free composite invariant with respect to atmosphere, topography and solar and viewing angles. For a given location (x,y), we define a reflectance factor $\rho(x,y,\lambda)$ measured by the PMDs of GOME at wavelength λ for the ground cover projection of the image. This reflectance is translated into normalized *rg*-color space via the relation:

$$r = \frac{\rho(x,y,\lambda_R)}{\sum_{i=R,G,B}\rho(x,y,\lambda_i)}, g = \frac{\rho(x,y,\lambda_G)}{\sum_{i=R,G,B}\rho(x,y,\lambda_i)}. \tag{34}$$

If M is the set of n normalized multi-temporal measurements over the same location (x,y), then a cloud-free (or minimum cloudiness) pixel rg_{CF} in M is selected with the brightness criterion $\|rg_{CF} - w\| \geq \|rg_k - w\|$ for $k = 1,..,n$, where $w = (1/3,1/3)$ is the *white point* in the *rg* chromaticity diagram. A global cloud-free composite is constructed by merging cloud-free reflectances $\rho_{CF}(\lambda)$ (corresponding to rg_{CF}) at all locations. The geometric cloud fraction is determined by examining separations between RGB reflectances and their cloud-free composite values:

$$c_f = \sqrt{\sum_{i=R,G,B_i}\alpha(\lambda_i)\max(0,[\rho(\lambda_i) - \rho_{CF}(\lambda_i)]^2 - \beta(\lambda_i))}. \tag{35}$$

Scaling factors α ensure that the cloud fraction is mapped to [0, 1], while offsets β account for aerosol and other radiative effects. A detailed description is given in [*Loyola*, 2000].

6.2.3. ROCINN algorithm

ROCINN [*Loyola*, 2004] is a relatively new algorithm based on O_2 *A* band reflectances from GOME: it delivers cloud-top pressure and cloud-top albedo. It assumes the IPA; cloud fraction c_f derived from the OCRA algorithm is a fixed input to ROCINN. For a scattering

atmosphere, reflectances in ROCINN version 2.0 are calculated by the LIDORT or VLIDORT RT model using line spectroscopic information for the O_2 *A* band (taken from the HITRAN 2004 database [*Rothman et al*., 2005]), before convolution with the instrument function. Forward model reflectance is then:

$$R_{sim}(\lambda) = c_f \langle R_c(\lambda, \Theta, c_a, c_z) \rangle + (1 - c_f) \langle R_s(\lambda, \Theta, s_a, s_z) \rangle \tag{36}$$

Here, $\langle R \rangle$ denotes the convoluted reflectance to cloud-top or surface for path geometry Θ (solar and line-of-sight zenith and azimuth angles), wavelength λ, surface albedo s_a and cloud-top albedo c_a, and lower boundary heights s_z (surface) and c_z (cloud-top). Quantities s_z and s_a are the surface height and albedo, taken from a suitable database and assumed known. ROCINN aims to retrieve cloud-top height c_z and the cloud-top albedo c_a. Reflectivity calculations based on Eq. (36) are used to create a complete data set of simulated reflectances for all viewing geometries and geophysical scenarios, and for various combinations of cloud fraction, cloud-top height and cloud-top albedo. High-resolution reflectances are computed for the range 758-772 nm at every 0.002 nm before convolution.

In ROCINN, the forward model function is represented by the set $S = \{(X_i, Y_i)\}$ for $i = 1,\ldots,s$. Inputs X are the parameters $\{ c_f, \Theta, s_a, s_z, c_a, c_z\}$. The outputs Y are the simulated radiances $\{R_{sim}(\lambda)\}$. To generate an inverse data set, we first add normal distributed Gaussian measurement noise ε to the simulated radiances: $R = R_{sim} + \varepsilon$. We may now generate the inverse data set $S^* = \{(X_i^*, Y_i^*)\}$ for $i = 1,\ldots,t$, where now the input set X^* comprises the parameters $\{R_{sim}(\lambda), c_f, \Theta, s_a, s_z\}$ and the output is now $Y^* = \{c_a, c_z\}$, the unknown cloud-top albedo and cloud-top height. A neural network NN_{INV} is finally trained with the *inverse* data set S^*, giving the result:

$$\{c_a, c_z\} = NN_{INV}(R_{sim}(\lambda), c_f, \Theta, s_a, s_z) \tag{37}$$

For more details on the use of neural networks to solve inverse problems, see [*Loyola*, 2004].

GOME is not able to distinguish between cloud cover and snow/ice surface conditions, as the instrument does not possess infrared channels. Furthermore, GOME is not sensitive to optically thin cirrus clouds. Despite these caveats, it is worth remarking that the cloud fraction is retrieved under all surface conditions; for snow/ice scenarios, OCRA will return an overestimate of the cloud fraction. Similarly, the cloud-top height and cloud-top albedo are retrieved for all pixels flagged as partially or fully cloudy. In GOME scenes for which OCRA has overestimated (underestimated) the cloud fraction, ROCINN will compensate by underestimating (overestimating) the cloud-top albedo and slightly overestimating (underestimating) the cloud-top height. For high surface albedo pixels, ROCINN can also be run in a newly-implemented snow/ice mode, in which the cloud fraction is set to 1, and the retrieved cloud-top albedo and cloud-top height are then regarded as effective values.

The ROCINN version 1.0 algorithm as used in GDP 4.0 was based on the use of transmittance-only reflectivity calculations, with molecular and aerosol scattering neglected. In 2005, a new set of templates were created using the LIDORT model (no polarization) with Rayleigh scattering in a 35-layer atmosphere with a single temperature/pressure profile. Two tables for reflectances to cloud-top (9 heights, 6 albedos), and surface (6 heights, 5 albedos) were prepared. The albedo dependence was established using the LER (Lambertian

Equivalent Reflectivity) formulation, and the first three Fourier components of the sun-normalized intensity fields were calculated and saved. For the initial runs, the table classifications were the same as those used for transmittance-only tables.

In 2006, the tables were upgraded using the VLIDORT model, to include polarization effects in the calculation of reflectances in an atmosphere with absorption and Rayleigh scattering. Classification for the cloud-top tables was extended to albedos as low as 0.3, and the highest cloud-top height was extended from 11 km to 14 km. Some 2,772,000 reflectance spectra were calculated for the template database; extra entries arise because of azimuth dependence in the scattering results. The wavelength range was also optimized to 758-771 nm.

The GODFIT Phase B implementation is similar to that in GDP 4.0, but is now called OCRA/ROCINN Version 2.0 as it is based on VLIDORT simulations. To reflect actual conditions, the surface albedo is taken the LER data set, while the surface height is taken from the ETOPO5 data set, with the template tables interpolated differently for these two quantities. This version 2.0 has been re-coded in plain C (formerly in C++) and developed into a stand-alone library with flexible porting (tested to date on Intel and GNU compilers). Version 2.0 comes with full documentation, and is designed to respond to GOME measurement types (nominal 1.5 or 6-second pixels, etc.).

6.2.4. Validation of OCRA and ROCINN algorithm

A verification of OCRA and ROCINN was described in the GDP 4.0 paper [*van Roozendael et al*., 2006]. These algorithms were verified by comparison with ATSR-2 data, obtained simultaneously with GOME (both instruments are on board ERS-2). ATSR has a much smaller field of view (pixel resolution ~1.1 km), so GOME narrow-swath orbits were selected. Cloud fraction and cloud-top height are determined from ATSR-2 data using infrared brightness temperature algorithms and the instrument's stereo viewing capability. ATSR-2 measurement data for one year (April 1998 to March 1999) were used in the verification (331 orbits) [*Siddans et al.*, 1999]. OCRA/ROCINN and FRESCO results were compared with this data set. Cloud-top albedo was not available from ATSR-2.

This ATSR-2 comparison confirmed the results reported in [*Tuinder et al*., 2004] where several algorithms for retrieving cloud fraction using GOME data were compared against synoptic surface observations: in this work, OCRA had a mean difference of −10% compared with synoptic data, compared to FRESCO (−19.7%) and ICFA (−38.9%). For cloud-top height, the ROCINN, FRESCO and ATSR-2 algorithms gave similar results for this study. Heights from the ROCINN algorithm show reduced scatter against ATSR-2 values compared with FRESCO. Both ROCINN and FRESCO tend to underestimate the cloud-top height. ROCINN is smoother and more stable than FRESCO and has fewer outliers. In general, these ATSR-2 comparisons show that OCRA/ROCINN gives better results on average than those from FRESCO. Differences in fraction and cloud-top height are smaller, and standard deviations of these differences are lower. For more details, see [*van Roozendael et al.*, 2006].

Here we report on comparisons that are more recent. An inter-comparison of GOME and ATSR cloud-top heights was performed for the ROCINN and the SACURA algorithms [*Rozanov et al*., 2006]. Here it was found that ROCINN was on average 0.5 km below ATSR, with SACURA 0.6 km above. Comparisons were restricted to high cloud-cover scenes. Cloud

fractions from OCRA and MERIS were compared in a study by [*Casadio et al.*, 2006]. In addition to the ATSR results, an extensive set of comparisons were made with cloud properties measured by OCRA/ROCINN and those measured by MSG (Meteosat Second Generation, data provided by Werner Thomas). For OCRA/ROCINN Version 1.0, the results were presented at the IGARSS 2006 conference. More recent MSG comparisons have been done with OCRA/ROCINN Version 2.0 [*Loyola et al.*, 2009], based on a larger sample set. In the histograms shown in Figure 17, we compare the performance of the OCRA/ROCINN versions against MSG, for cloud fraction and cloud-top height. The improvement in cloud-top height is noticeable (mainly the result of more representative reflectance templates in ROCINN). Results from the MSG comparisons are summarized in Table 3.

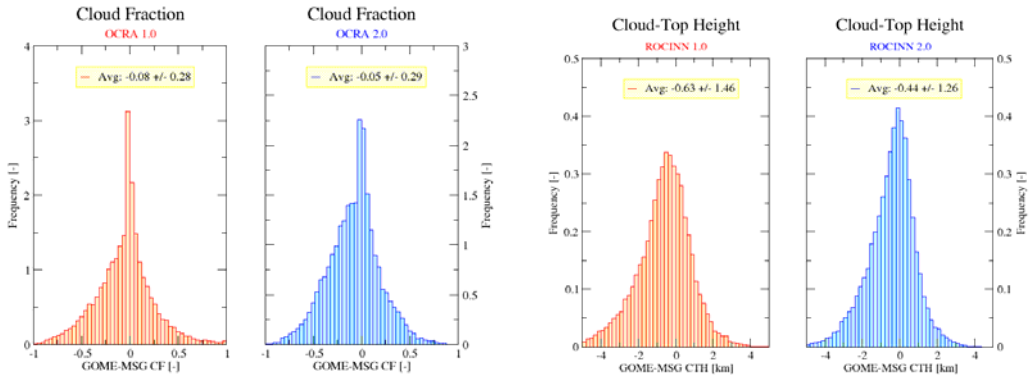


Figure 17. MSG and OCRA/ROCINN comparisons of cloud-fraction and cloud-height.

	Cloud Fraction	***Cloud-Top Albedo***	***Cloud-Top Height [km]***
Number of Pixels	30802	15649 (t>5) 20633 (t>3)	23944
OCRA/ROCINN 1.0	-0.08 ± 0.28	0.09 ± 0.10	-0.63 ± 1.46
OCRA/ROCINN 2.0	-0.05 ± 0.29	0.07 ± 0.09	-0.44 ± 1.26
Land (Version 2.0)	-0.01 ± 0.26	-0.08 ± 0.09	-0.59 ± 1.33
Sea (Version 2.0)	-0.11 ± 0.31	-0.06 ± 0.08	-0.31 ± 1.19

Table 3. MSG and OCRA/ROCINN comparisons of cloud-fraction and cloud-height.

6.2.5. Total ozone sensitivity to cloud inputs

Clouds have a significant influence on the retrieved ozone total column. In both the GODFIT direct fitting and GDOAS algorithms, the LIDORT-simulated intensity field and the Ring correction factor depend on the three cloud parameters (cloud fraction, cloud-top height and cloud-top albedo). In Figure 18 (taken from [*van Roozendael et al.*, 2006]) the effect of errors in cloud fraction from OCRA on the ROCINN parameters cloud-top height and cloud-top albedo is shown, along with the effect on the GDOAS-retrieved total ozone column itself. The top (bottom) row of three panels shows normalized histograms for errors induced by a 10% overestimation (underestimation) of cloud fraction. A 10% increase of the cloud fraction induces ~5% decrease for the cloud-top albedo and a ~5% increase for the cloud-top height. The ROCINN algorithm compensates a possible cloud-fraction overestimation by

underestimating the cloud-top albedo and overestimating the cloud-top height. Thus the net effect of combined OCRA/ROCINN uncertainties is to maintain the level of ozone total column error to the ±0.5% level.

Next, we report on some more recent studies comparing total ozone results for GOME obtained by the OCRA/ROCINN and FRESCO algorithms, and for SCIAMACHY obtained by the OCRA/SACURA and FRESCO algorithms. The cloud parameters from FRESCO (version GO-v4) and from OCRA/ROCINN (version 1.0) were compared for 24 GOME orbits from 1999. OCRA derives cloud fractions larger than FRESCO while the cloud top albedos issued from ROCINN are generally weaker than the FRESCO ones. However, the products of these two parameters (the cloud radiance fraction) derived from both algorithms are in good agreement. ROCINN derives higher clouds than FRESCO and it was shown that the total O_3 relative differences retrieved with FRESCO or OCRA/ROCINN (-2% to 5%) are mainly explained by the cloud top pressure differences (Fig. 19). Obviously, the total ozone is mainly affected by cloud top pressure differences for cloud radiance fraction not too low (> 0.2). For pixels in the FRESCO ice mode, the total O_3 relative differences retrieved with FRESCO or OCRA/ROCINN depend also on the OCRA cloud fraction as this FRESCO parameter is set to unity in this mode. A similar behavior was observed for SCIAMACHY when comparing OCRA/SACURA to FRESCO, except that cloud fractions from FRESCO and OCRA are much better correlated for this instrument. For low values of cloud fraction, correlation between cloud top heights from FRESCO and SACURA is poor.

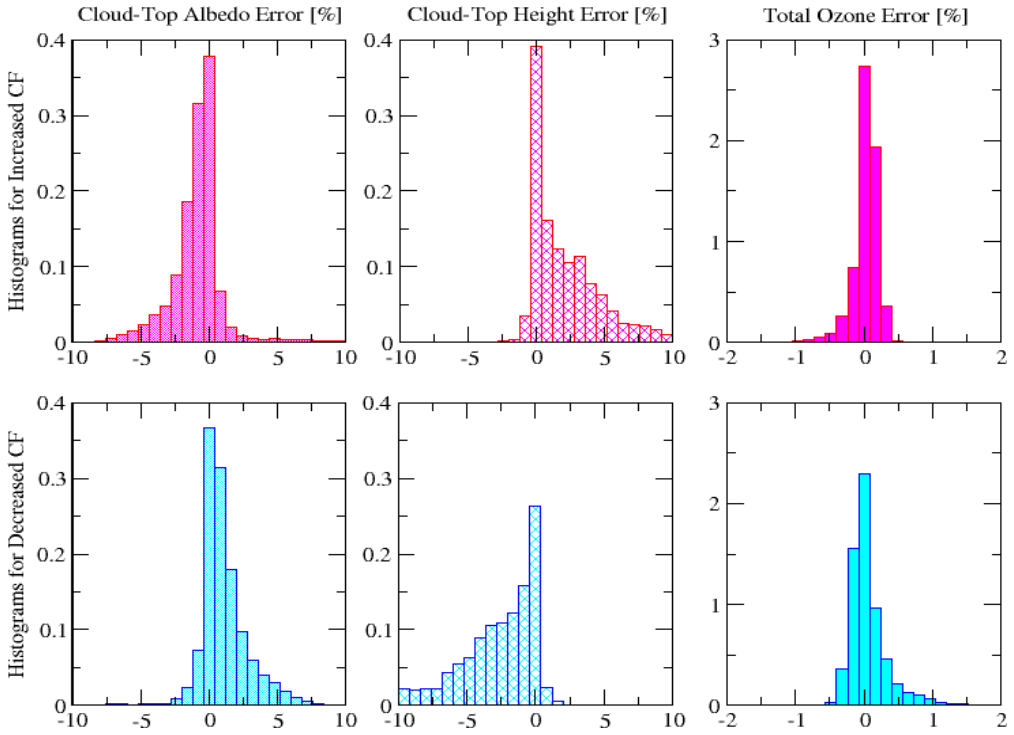


Figure 18. Normalized histograms for errors induced by a 10% overestimate of OCRA's cloud fraction (top panels) and a 10% underestimate of cloud fraction (lower panels). Relative errors for cloud-top albedo (first column), cloud-top height (middle column) and total ozone (last column) are shown. A 10% increase in cloud fraction induces a ~5% decrease of cloud-top albedo and ~5% increase of cloud-top height. The net error on total ozone stays at 0.5% level.

It has been shown that the adjustment of surface albedo in the internal closure mode of GODFIT can lead to convergence problems for pixels with cloud fraction higher than 0.85. In Phase A, such convergence problems were avoided by re-setting all perceived cloud fractions higher than a certain threshold (nominally 0.85) equal to this value. Unfortunately, this procedure did lead to unphysical surface albedos in the retrieval procedure for pixels with such high cloud fractions. For such pixels, it is better to implement a "Fully cloudy" mode in which the *cloud-top albedo* is adjusted rather than the surface albedo; this assumes a cloud fraction equal to one. In this mode, an ozone ghost column should be then added to the retrieved column.

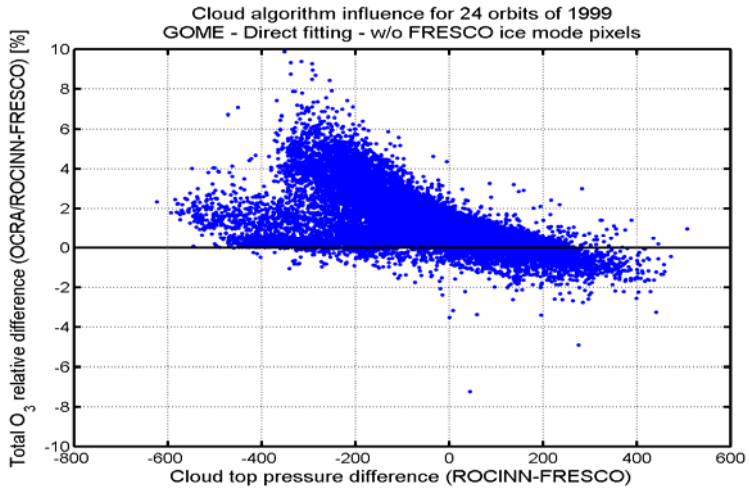


Figure 19. Correlation between total ozone relative differences derived with OCRA/ROCINN or FRESCO and the cloud top pressure differences issued from these two cloud algorithms.

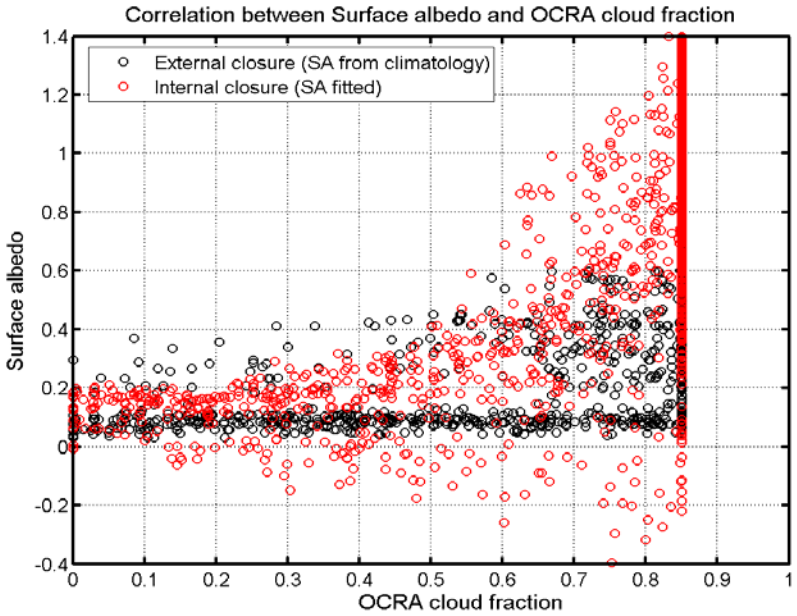


Figure 20. Illustration of the correlation between the surface albedo and the cloud fraction derived by OCRA for GOME.

It was also shown that there is a correlation between the GOME OCRA cloud fractions and the surface albedo (Fig. 20). This figure indicates that OCRA derives cloud fractions that are too large for scenes with high surface albedo, thus probably leading to overestimated O_3 columns. For clear-sky scenes, surface albedos adjusted in the internal closure mode are in very good agreement with surface albedos specified in climatologies. However, for cloud-contaminated pixels, fitted albedo values depend strongly on cloud parameters. Finally, using this internal closure mode leads to total O_3 relative differences lying between 0 and 2% compared to columns retrieved with the external closure mode. These differences are considerably reduced for high FRESCO cloud fractions but not at high OCRA cloud fractions.

It is well known that the CRB (clouds as reflecting boundaries) assumption leads to overestimation of total column ozone in DOAS algorithms. In particular, the DOAS vertical column density depends on the climatological "ghost" ozone column below cloud-top, and there is no consideration of the effect of intra-cloud scattering on ozone. This situation has been partially remedied in GDP 4.x, where an empirical intra-cloud ozone correction has been added to compensate for CRB-induced bias [*Loyola*, 2007]. Although in GODFIT the whole ozone column is fitted directly, and there is therefore no dependence on ghost column climatology, the CRB assumption is still a source of error.

An intra-cloud correction has also been developed for the GDP5 baseline. The intra cloud correction to the ozone column V_{ic} may be written as follows:

$$V_{ic} = fG(V)(1-R_c)\mu_0 . \tag{38}$$

Here, f is the cloud fraction, R_c is the cloud-top albedo, and μ_0 is the cosine of the solar zenith angle at the ground. $G(V)$ is the ozone ghost column below cloud cloud-top, depending uniquely on the final value of the retrieved total ozone V through the profile-column mapping. The corrected total ozone column is then:

$$V_{corr} = V - V_{ic} . \tag{39}$$

Recently, GDP 4.x has been given an alternative clouds-as-layers (CAL) scheme with clouds treated as plane-parallel layers of scattering particulates (water droplets or ice crystals). The CAL scheme depends upon the cloud optical depth, cloud top and bottom heights. Water droplet cloud optical properties are taken from Mie calculations with lognormal particle size distributions; AMFs in the UV Huggins bands are insensitive to the distribution parameters. Reprocessed results using the GDP 4.x CAL scheme show an improvement in total ozone validations for moderate to high solar angles in particular. Although the CRB scheme is the current baseline for GDP5, it is expected that the CAL scheme will be implemented in the near future.

7. THE GDP5 GODFIT ALGORITHM

7.1. Retrieval state vector for GODFIT in GDP5

7.1.1. Baseline sets of fitting parameters

The baseline state vector in GDP5 depends on the options chosen for the algorithm. These are

1. the choice of ozone profile climatology;
2. the choice of external or internal albedo closure;
3. the option to include a temperature shift in the fitting;
4. the use of a LUT for dealing with Ring effect interference

For a Ring correction using the empirical formulae, a fitting amplitude is required for the additive Fraunhofer Ring spectrum. This "Ring reference spectrum" is the same as that used in GDP 4.0, namely the convolution of Raman cross-sections with a high-resolution Fraunhofer spectrum [*Chance and Spurr*, 1997]. This term is not present if the LUT Ring correction option is being used. An undersampling correction was introduced in GDP 3.0 to compensate for GOME's sampling slightly below the Nyquist criterion [*Slijkhuis et al.*, 1999]. The undersampling spectrum is additive; we use the GDP 4.0 default. The implementation of a shift fitting parameter for the measured earthshine spectrum is discussed below.

Table 7.1 summarizes these options. The default number of closure parameters has been set at 3, regardless of the external or internal choice of closure method. The T-shift option applies equally to TOMS or NNORSY profiles. The first column (light green) is the GDP5 default.

	GODFIT OPTIONS			
	TOMS V8 O_3 profiles		NNORSY O_3 data	
	Ring (fitted)	Ring (LUT)	Ring (fitted)	Ring (LUT)
Total ozone DU	1	1	1	1
Closure	3	3	3	3
T-shift retrieval	1	1	1	1
Ring Fraunhofer	1	n/a	1	n/a
Earthshine Shift	1	1	1	1
Undersampling	1	1	1	1
TOTAL	8	7	8	7

Table 7.1. Summary of fitting parameters for GDP5/GODFIT direct fitting.

Total O_3	Closure 1	Closure 2	Closure 3	T-shift	Ring	λ–Shift	U-sampling
Previous	1.0 (external) R_{335} (internal)	0.0	0.0	0.0	1.0	0.0	0.0

Table 7.2. GDP5 default, initial settings

First-guess values for the GDP5 default are shown in Table 7.2. As noted in Section 5.2, the total O_3 first guess in the operational GDP5 algorithm is taken from the previous pixel value. If this previous value is not available for some reason, the initial total ozone column is taken from a zonal averaged climatology based on many years of TOMS data (J. Gleason, private communication). For external albedo closure, the first parameter is set to an initial value of 1.0; other initial albedo parameters are zero. [For internal albedo closure, an initial value for the first albedo parameter is extracted from the GOME LER database at 335 nm]. Initial values of the under-sampling amplitude, earthshine λ-shift and T-shift are zero.

NO_2 is an interfering species in the UV, but its optical thickness for absorption is small compared with O_3, and tests have shown that the inclusion of a total column of NO_2 (in [DU]) as an additional fitting parameter does not influence the O_3 result in any significant way (Phase A, Final Report, for more details).

7.1.2. Earthshine shift

Since we are dealing with sun-normalized radiances, we have opted to work on the spectral grid of the GOME solar irradiance spectrum supplied with every orbit. There is of course a wavelength registration mismatch between irradiance and radiance spectra, arising mainly from the solar spectrum Doppler shift. This mismatch (shift) varies across an orbit due to changes in the instrument temperature.

In GODFIT Phase A, the initial direct fitting was done using a *pre-set shift* of 0.08 nm (an average value for the Doppler shift) to each earthshine spectrum (re-sampling is always done by cubic-spline interpolation). This was further refined by carrying out a preliminary fit for the earthshine shift (values range between 0.07 and 0.09 nm across an orbit) before entering the main GODFIT iteration.

In GODFIT Phase B, the implementation of an earthshine spectrum shift was introduced in line with the usual procedure in DOAS: the earthshine spectrum shift is fitted as part of the retrieval procedure, and the shift value is then an element in the state vector of retrieval parameters. This allows us to manage the wavelength mismatch between the solar reference grid and the radiance spectrum. This is the GDP5 default.

7.2. Configuration settings for GDP5

7.2.1. Forward model settings

Atmospheric profiles are created on a 13-layer vertical grid with 11 pressure levels specified by the TOMS ozone climatology, and 2 additional pressure levels (0.05, 0.02 mb), with pressure zero at a pre-defined TOA level (65 km). The default choice for ozone profiles is now TOMS Version 8, with TOMS temperature climatology being used in conjunction with the T-shift procedure. The Daumont-Malicet ozone cross-sections with zero pre-shift have been selected. NO_2 is not included in the fit, and is not modeled in the radiative transfer. Internal closure has been used for the surface albedo, with 3 closure parameters to be fitted.

For LIDORT usage, we use 8 discrete ordinate streams for the multiple scatter integrals, and 2 fine-layer sub-divisions for the outgoing sphericity single scatter calculation. All calculations are done for a Rayleigh scattering atmosphere. Geometrical input to LIDORT is specified at the bottom of the atmosphere.

For every GOME footprint, local angles for the solar and viewing planes must be extracted from the Level 1b geolocation information. In an ideal world, simulated radiances should really be integrated over the footprint readout time – for a 1.5-second GOME forward scan in regular nadir-viewing mode, the line of sight zenith angle has changed by some 20°. In GOME level 1b data, viewing and solar angles are given at three points across the footprint (beginning of scan, center and end of scan). Forward calculations are most often done using only the center-scan geometry, but we have found that greater accuracy can be obtained using three scan geometries. In GDP the parabolic weighting (1:4:1) was used to approximate the scan integration. This choice has some impact on performance (see below). Studies for the GOME ozone profile algorithm have shown that this integration over scan angle can be approximated by a 2- or 3-point quadrature with little loss of accuracy.

7.2.2. Inverse model configurations

In GODFIT, the forward model delivers the necessary analytic weighting functions in an efficient manner. LIDORT generates the total ozone, T-shift and (internal) albedo closure Jacobians. The weighting function with respect to the Ring amplitude is easy, since this is just a scaling of the radiances; similarly the undersampling Jacobian is just the reference spectrum itself, since its contribution is linearly additive. Jacobians for external closure parameters are also straightforward. The earthshine shift involves a spline resampling of measurements, and this is easy to differentiate.

In GODFIT Phase B, optimal estimation (OE) software based on similar code for ozone profile retrieval [*van Oss et al.*, 2002] was implemented. Although the GODFIT inverse problem is not ill conditioned, it was found that the a little regularization helps to stabilize the retrieval – the *a priori* constraints are deliberately made very loose, so that the precision is not compromised in any serious way by *a priori* smoothing. *A priori* variances were set (by hand) at very high levels and assumed to have no cross-correlations. The *a priori* vector is taken to be the initial state vector. The level of *a priori* was chosen so that the DFS diagnostic was always close (to within 0.01) of the number of parameters in the fit.

In GDP5, the optimal estimation was replaced by the LVMR_VR (Levenberg-Marquardt with variable regularization) and line-search algorithms, as noted in section 5.2. Here no special consideration applies to the level of regularization, and the optimal choice of three wavelengths for the line-search algorithm was noted above in section 2.5. For operational orbit processing, the default is to use the previous pixel result as the first guess for the ozone total column.

The major diagnostic is the solution covariance matrix, with diagonal entries indicating the individual fitting parameter variances. RMS and chi-square are useful single number diagnostics. For iteration convergence, we looked at the relative change (between iterations) in χ^2 and the relative change in the total ozone column (the first of the fitting parameters). We adopted a 0.1% criterion for the change in total ozone (the same value is used for the convergence of iterative AMFs). Convergence is rapid in the vast majority of cases (2 to 4 iterations).

7.2.3. Re-sampling issues

The GOME earthshine and solar spectra are produced by the GDP level 0-to-1 algorithm by means of the Level 0-to-1b extractor [*Aberle et al.*, 2002; *Slijkhuis et al.*, 2004]. Wavelength calibration of the level-1 spectra was improved selectively through application of window-dependent pre-shifts to parts of the solar spectrum before each orbit of data is processed. These pre-shifts are established by cross-correlation with a high-resolution solar spectrum [*Chance and Spurr*, 1997] over limited wavelength ranges covering the main fitting windows (325-335 nm for O_3, 425-450 nm for NO_2 in the visible, and 758-772 nm covering the O_2 *A* band as used in the ROCINN algorithm). This procedure was used in GDP 4.0 [*van Roozendael et al.*, 2006]. There have been other attempts to improve wavelength calibration [*van Geffen and van Oss*, 2003]; see also [*van der A et al.*, 2002] for a discussion of the effects of using recalibrated GOME data on ozone profile retrieval accuracy.

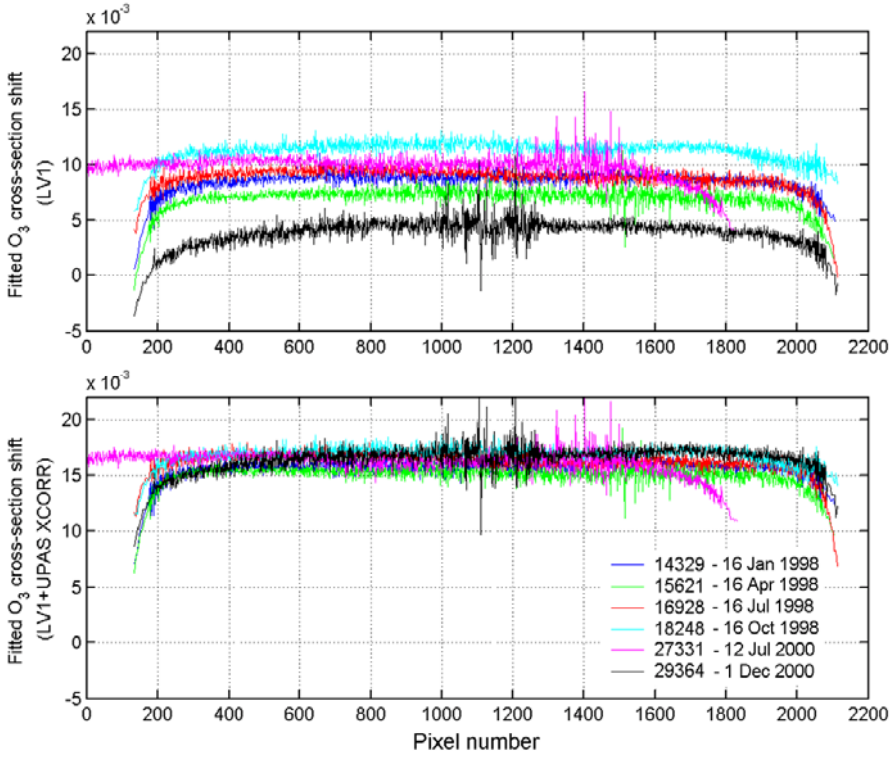


Figure 21. Improvement in fitted ozone cross-section wavelength shifts resulting from an additional post Level 1 wavelength cross-correlation. (Upper panel) fitted shifts derived using the original extracted Level 1 data; (lower panel) fitted shifts determined after the additional wavelength cross-correlation ("UPAS XCORR") applied to Level 1 data. Results are shown for 6 orbits as indicated.

Pre-shifting of ozone cross-sections is required to compensate for inaccuracies in the wavelength calibration of the GOME FM 98 data. Following a recommendation from the GDP 3.0 geophysical validation campaign [*Van Roozendael et al.*, 2002], O_3 cross-sections are now corrected for the so-called solar I_0 effect [*Aliwell et al.*, 2002], and have been implemented in GDP 4.0 and GODFIT/GDP5 with a pre-shift of +0.016 nm. The same value applies to NO_2 cross-sections. This value was determined from a series of test retrievals in which the ozone cross-section shift was optimized as part of the DOAS procedure itself. This exercise removed a systematic positive bias of 1.5 % in GDP 3.0 total ozone, and also a bias

in the DOAS effective temperature [*Van Roozendael et al*., 2006]. Figure 21 (taken from this paper) shows the improved stability in the choice of ozone cross-section shift due to this calibration enhancement.

In GODFIT, there has been no pre-convolution of reference spectra, at least outside of any wavelength registration using the high-resolution Kitt Peak spectrum [*Chance and Spurr*, 1997]; the one exception is the high-resolution Daumont-Malicet ozone cross-section data set. For GODFIT Phase B, undersampling and Fraunhofer Ring spectra were taken from GDP 4.0 data sets.

7.3. The GODFIT software package

7.3.1. Overall capability

Phase-A GODFIT algorithm had two modes – the direct fitting and GDOAS options. Direct fitting was confined to total ozone, and DOAS-style retrievals were restricted to total ozone and nitrogen dioxide. The Phase A algorithm was designed to deal with only GOME measurements. GDOAS was incorporated in GDP 4.0 and higher variants. Members of the present GODFIT team have been involved with the use of GDOAS as part of the development of operational algorithms for the GOME-2 project. For this and other reasons, the GODFIT Phase B tool has the ability to ingest and process data from the GOME-1, GOME-2 and SCIAMACHY instruments. It can also generate synthetic data for closed-loop retrieval testing. A contract change notice from ESA allowed the GODFIT team to participate in the SAUNA campaign in March-April 2006, and for this, GODFIT was given the capability to process ground-based measurements from instruments involved in this campaign (DOAS and direct fitting options are available). The ground-based facility has proven useful in other validation contexts.

The initial GODFIT Phase A algorithm was written at SAO (Smithsonian Astrophysical Observatory) in 2003 in FORTRAN 77 on a Sun Unix workstation, and was transferred to PCs with Linux and Windows-based operating systems at BIRA in June 2003. The Phase A algorithms were then tested on whole orbits of GOME data, and linked to the validation software at BIRA. With the exception of the OCRA/ROCINN Version 2.0 development, all phase B work was done at BIRA. This includes the implementation of the T-shift scheme, the Ring correction LUTs, the earthshine shift fitting, new reference data sets, and the integration of the OCRA/ROCINN package. The whole package was re-organized and given a strict exception handling procedure. In particular, all input configurations, whether Level 1b GOME data or auxiliary FRESCO or ICFA data or variables to be assigned from the input configuration file, are initialized, then read from file as necessary, and then checked before the algorithm can proceed. LIDORT has its own error handling structure, which is incorporated in the larger whole.

7.3.2. Performance issues

In GODFIT, forward model simulations of basic radiances and weighting functions are done from scratch. For GDP 4.0, the entire 10-year record of GOME total ozone was reprocessed without the use look-up tables for fundamental RT calculations of air mass factors. Doing RT simulations from scratch is the time-consuming step, and this depends on the two discretizations used in LIDORT – the number of vertical layers and the number of discrete

ordinate streams. For a 13-layer Rayleigh-only atmosphere with 8 discrete ordinates in total (this is the baseline), a single LIDORT call will return the backscatter radiance and column/albedo weighting functions in a small fraction of a second.

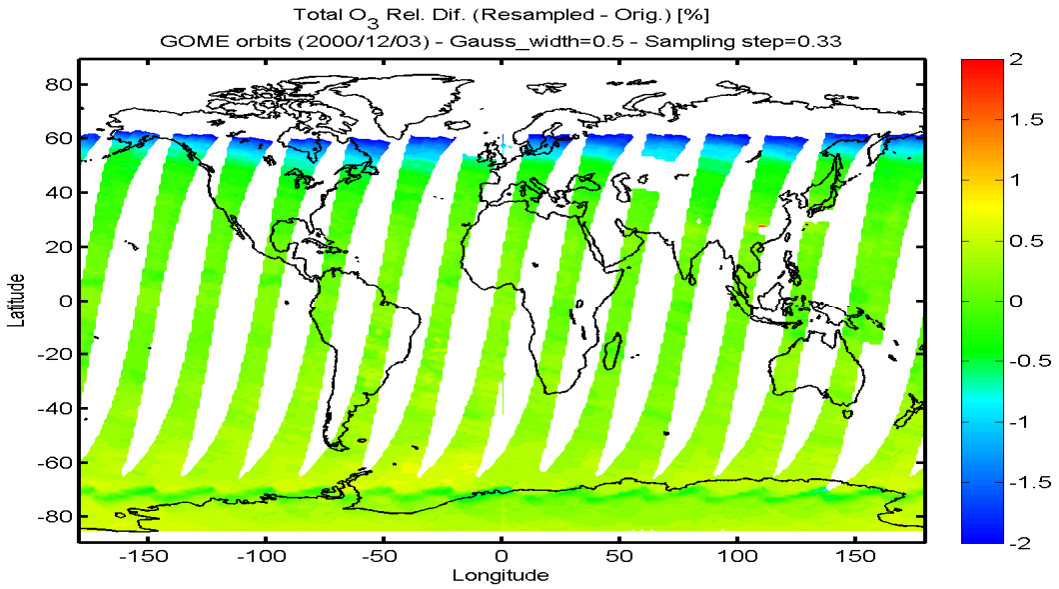


Figure 22. Resampling for a day of GOME orbits. Percentage differences between total ozone with and without optimal resampling as indicated.

Based on these considerations, we make the following estimates based on the requirement that the data granularity is one orbit. We assume an orbit has ~2000 footprints, 90% of which are partially cloudy (requiring dual cloud/clear sky computations); there is a 10 nm fitting window containing ~90 spectral points. Further, we take an average of 3.5 iterations for the fitting, and assume the use of all three footprint geometries. Then we find that there are almost 4 million forward model calls to LIDORT during the processing of an orbit granule.

Based on these numbers and on timing tests done at BIRS-IASB in September 2008, it takes about 3 minutes to process an orbit using the GDOAS algorithm, and about 40 times longer to process using direct fitting GODFIT. This is a reflection of the greatly increased number of LIDORT calculations. This is an order of magnitude slower.

For GOME forward scan pixels, forward calculations are done using only center-scan geometry in order to save computing time. The associated small loss of accuracy for these pixels is acceptable. On the contrary, in GODFIT, all three geometries are needed to maintain sufficient accuracy for backscan pixels.

At the 6th progress meeting in Brussels (17 December 2009), it was decided to omit regular backscan pixels in the forthcoming GDP5 reprocessing in early 2010; this represents about a 40% saving in computer time. [In the past, there has been no significant use of backscan results]. However, larger integration time pixels over Polar Regions will be processed in addition to the usual 3 forward scans, The GDP5 default will use single center-scan geometry throughout.

One way to improve the performance is to reduce the sampling by means of spectral smoothing. GODFIT has software to resample radiance and irradiance spectra, but this

requires a degradation of the resolution of radiance and irradiance spectra, and reference cross-sections as well. Smoothing is convolution with a Gaussian function (the half width is the controlling parameter), and for the sampling, a sampling step size is the other control parameter. The half-width and the sampling step must satisfy the Nyquist/Shannon theorem.

An initial choice of optimal resampling has a half-width of 0.5 nm for the Gaussian, and a sampling step of 0.33 nm. Using these values, Figure 22 compares ozone results for a day of GOME orbits, without and with resampling using these two parameters. In general, we find that the impact is negligible at low and intermediate SZA, but clearly the resampling should not be used at higher SZA. For this example, processing time was reduced by not more than a factor of 3. Further work is required to establish limits here, and the exercise needs to be repeated with the T-shift option turned on. This issue has also been discussed in a recent paper from the 2008 QOS Symposium [*Lerot et al*., 2010].

Performance has been an issue in the choice of the ozone profile climatology. We have already seen an aspect of this in section 2.1.3 (Figure 2.3). There, it was noted that use of the "Mode 1" NNORSY profiles led to consistently larger iterations required for least-squares convergence in the inverse model. This is due to the lack of variability in the NNORSY derivative profiles (Figure 2.3, lower right) as delivered by the NNORSY extraction software. In September 2008, the calculation of NNORSY derivatives was changed to a finite difference method, with results now much more closely resembling the TOMS profile derivatives (Figure 2.3, lower left), and with concomitant performance now becoming very similar to that obtained using TOMS Version 8 profiles.

7.3.3. Algorithm verification

There are two main algorithm checks. The first is to make sure that LIDORT is delivering analytic Jacobians in the correct manner. We use finite difference estimates for this task, in which one parameter (the column or albedo) is changed by a small amount and the resulting change in simulated intensity is divided by the small parameter change to get a finite difference estimate of the derivative. A dedicated "weighting function check" module has been written for this task.

The second verification is the "quasi-perfect" or "closed-loop" retrieval test. We first run the algorithm in non-retrieval mode to produce simulated radiances that are then converted to synthetic level 1b GOME data through the addition of measurement noise. Then we run the algorithm again, this time in retrieval mode, using these synthetic data as measurement input. Measurement noise may be added by hand or by means of an instrument model. If this is the only source of error, then the retrieval should return the "truth parameters" that were used in the first place to generate the synthetic data. The algorithm has a full capacity for this kind of closed-loop testing. [Closed-loop testing on the molecular Ring correction was done off-line].

7.4. The UPAS Installation

7.4.1. UPAS implementation issues

A schematic of the GDP UPAS environment is shown in Figure 23. The GDP has now been given the GDP5 capability in addition to existing features. The transition to LIDORT Version 3.3 was completed in UPAS in autumn 2008, and was first tested on the GDP4 algorithm. Excellent agreements for ozone AMF values were obtained between results calculated using

Version 3.3 and the older Version 2.5. As part of the GDP5 implementation, a C++ interface for LIDORT was developed in 2009, and has now been completed and tested in the GDP4 setting. GDP also has configurations for several inverse methods, and one of these has been chosen for GDP5 (section 5.2).

Currently, UPAS runs on 5 eight core machines in parallel without essential slowdown, reading/writing products via NFS. GODFIT performance on this system should be the same as UPAS single processor performance (factor 40). The current single UPAS GDP4 performance: 70 seconds for one GOME-1 orbit (including extraction); the estimated performance for GDP5 ~0.85 orbits per minute (~ two months for 65,000 orbits).

The UPAS nightly build chain ensures continuous code quality and is completed at the time of writing. Basically, if there is a new SVN revision of UPAS, the chain should check it out, compile it, and process some products. It is implemented on the build host of the operational version, and it checks for errors, and will keep statistics about processing time and memory consumption. The chain can also be used to build branch revisions for comparison purposes, and to build past revisions for verification of specific code parts.

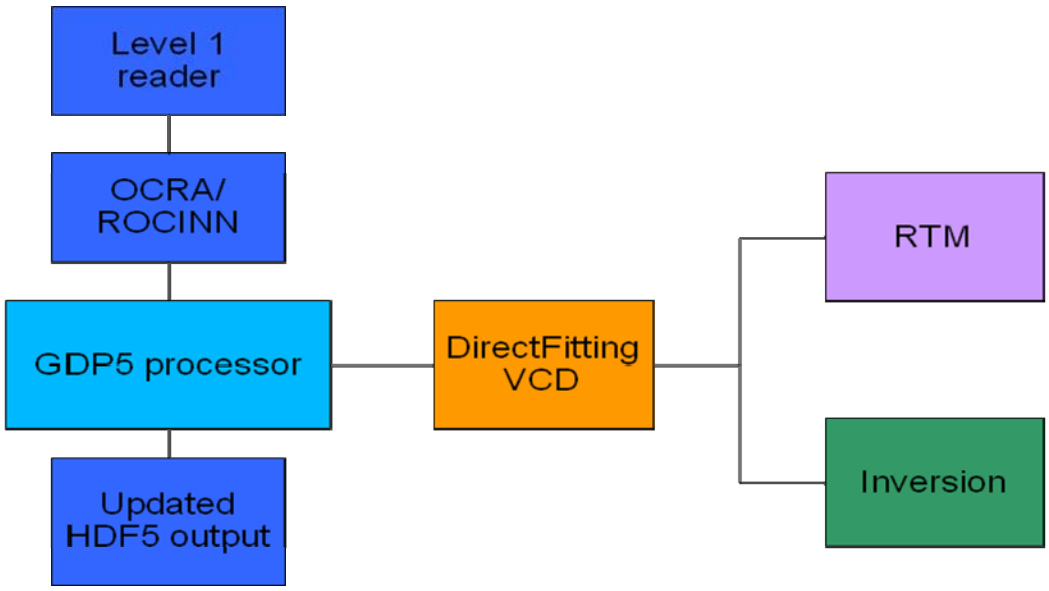


Figure 23. Outline of GDP5 in UPAS.

A number of possible performance enhancements for LIDORT are under consideration for the reprocessing in 2010. It is estimated that a 45% improvement can be obtained by optimizing the single scattering correction (SSCORR_OUTGOING) module inside LIDORT. A further 33% gain comes from optimizing LAPACK BLAS routines (by threading for multi-core architectures and using CUDA/OpenCL acceleration). Other smaller levels of optimization can be applied to the eigenvalue routine ASYMTX, and to the initialization in BVP_MATRIX_INIT. In the longer term, LIDORT should be re-designed with accelerators and multi-core processing in mind.

UPAS now has the ability to handle clouds treated as scattering layers (CAL) as well as the more traditional clouds as reflecting boundaries (CRB) method. CAL and CRB results have been compared extensively for GDP4. GDP5 will use CRB as the default.

7.4.2. Level 1 issues

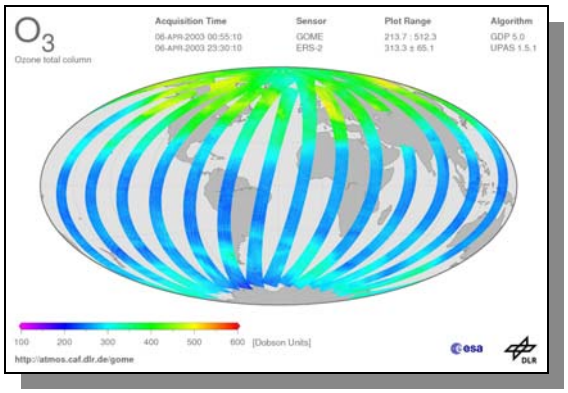
Although the Level 1 product has been standard for GOME for many years, there have been some changes in the last years which are worth noting here. In particular, there has been a change-over to angle definitions at the bottom of the atmosphere (BOA). In addition to the usual trios of solar zenith angles, viewing zenith angles and relative azimuth angles defined at the spacecraft and at TOA, these angles are now specified at BOA in the extracted Level 1b product. Also newly specified is the topographic height. Secondly, additional cloud information from OCRA/ROCINN has been added: in addition to the cloud fraction, cloud-top albedo and cloud-top height (and errors in %), we now have cloud-top pressure and cloud optical thickness, the latter derived from a look-up table inversion [*Loyola et al*., 2009], and an indication of cloud type (based on the ISCCP classification).

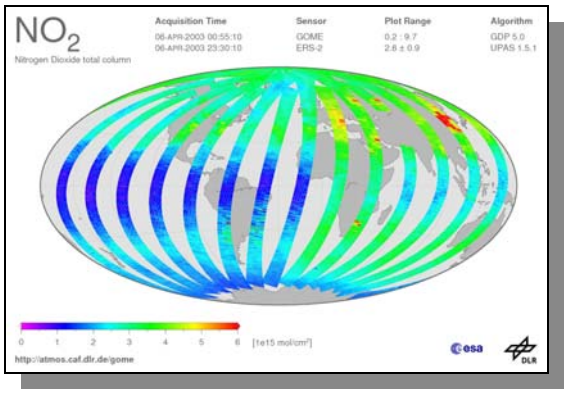
Thirdly, the degradation correction has been improved. Instead of an interpolating polynomial is time, the new correction is based on a look-up table of polynomial coefficients versus time, and uses the Savitsky-Golay algorithm for interpolation. Fourthly some unstable calibration lines have been removed. For details, see [*Coldewey-Egbers et al*., 2008].

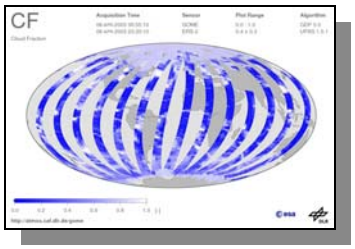
The latest Level 1 product is an input to the GDP5 GODFIT algorithm: GDP L1 version 4.00 from 1995 to 2003 and GDP L1 version 4.03 from 2003 to 2011.

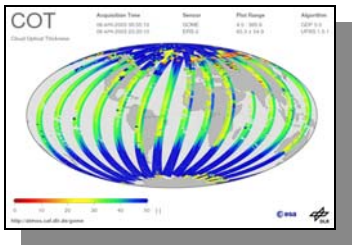
7.4.2. Documentation and Imagery

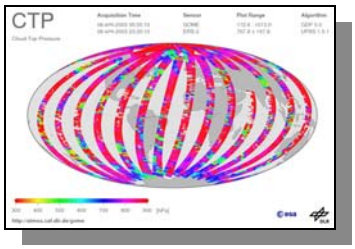
GOME documentation (including this report) and imagery can be found at http://atmos.eoc.dlr.de/gome











REFERENCES

Project Documentation

[R1] ATBD, GOME Direct Fitting Phase A, ESA AO/1-4235/02/I-LG, November 2003.

[R2] ATBD, GOME Direct Fitting Phase B, ESA 11817/04/I-OL, May 2007.

[R3] ATBD, GOME Total Column densities of O_3 and NO_2, UPAS/GDOAS 4.0, ESA ERSE-DTEX-EOPG-TN-04-0007, December 2004.

[R4] J-C. Lambert, M. Koukouli, D. Balis, J. Granville, C. Lerot, and M. Van Roozendael, GDP 5.0 Upgrade of the GOME Data Processor for Improved Total Ozone Columns — Validation Report, TN-IASB-GOME-GDP5-VR, Iss./Rev. 1B, August 2012.

[R5] D. Loyola, W. Zimmer, S. Kiemle, P. Valks, M. Pedergnana, Product User Manual for GOME Total Columns of Ozone, NO2, tropospheric NO2, BrO, SO2, H2O, HCHO, OClO, and Cloud Properties, DLR/GOME/PUM/01, Iss./Rev. 2E, August 2012.

Other References

Aberle, B., W. Balzer, A. von Bargen, E. Hegels, D. Loyola and R. Spurr, GOME Level 0 to 1 Algorithms Description, Tech. Note ER-TN-DLR-GO-0022 (Issue 5/B),Deutsches Zentrum für Luft und Raumfahrt, Oberpfaffenhofen, Germany, 2002.

Aliwell, S. R., M. Van Roozendael, P. V. Johnston, A. Richter, T. Wagner, D. W. Arlander, J. P. Burrows, D. J. Fish, R. L. Jones, K. K. Tørnkvist, J.-C. Lambert, K. Pfeilsticker, and I. Pundt, Analysis for BrO in zenith-sky spectra: An intercomparison exercise for analysis improvement, *J. Geophys. Res.*, **107**, D14, doi: 10.1029/2001JD000329, 2002.

Balis, D., J-C. Lambert, M. van Roozendael, R. Spurr, D. Loyola, Y. Livschitz, P. Valks, V. Amiridis, P. Gerard, J. Granville, and C. Zehner, Ten years of GOME/ERS2 total ozone data: the new GOME Data Processor (GDP) Version 4: II. Ground-based validation and comparisons with TOMS V7/V8, *J. Geophys Res.*, doi: 10.1029/2005JD006375, 2006.

Bass, A., and R. Paur, The Ultraviolet Cross Sections of Ozone: I. The Measurements, in *Atmospheric Ozone, Proceedings of the Quadrennial Ozone Symposium*, edited by C. S. Zerefos and A. Ghazi, D. Reidel Publishing Company, 1984.

Bhartia, P., Algorithm Theoretical Baseline Document, TOMS v8 Total ozone algorithm (http://toms.gsfc.nasa.gov/version8/version8_update.html), 2003.

Bodhaine, B., N. Wood, E. Dutton, and J. Slusser, On Rayleigh optical depth calculations, *J. Atmos. Ocean. Tech.*, **16**, 1854-1861, 1999.

Boersma, K., H. Eskes, and E. Brinksma, Error analysis for tropospheric NO_2 retrieval from space, *J. Geophys. Res.*, **109**, D04311, doi:10.1029/2003JD003962, 2004.

Bogumil, K., J. Orphal, T. Homann, S. Voigt, P. Spietz, O. Fleischmann, A. Vogel, M. Hartmann, H. Bovensmann, J. Frerick, and J. Burrows: Measurements of molecular absorption spectra with the SCIAMACHY Pre-Flight Model: instrument characterization and reference data for atmospheric remote-sensing in the 230--2380 nm region, *Journal of Photochemistry and Photobiology A,* **157**, 167-184, 2003.

Bovensmann, H., J. Burrows, M. Buchwitz, J. Frerick, S. Noel, V. Rozanov, K. Chance, and A. Goede, SCIAMACHY: Mission Objectives and Measurement Modes, *J. Atmos. Sci.,* **56**, 127–150, 1999.

Bowker, D., R. Davies, D. Myrick, K. Stacy, and W. Jones, Spectral Reflectances of Natural Targets for Use in Remote Sensing Studies, NASA Reference Publication, 1139, 1985.

Burrows, J., 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 231-794 nm range, *J. Quant. Spectrosc. Radiat. Transfer.*, **60**, 1025-1031, 1998.

Burrows, J., 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 231-794 nm range, *J. Quant. Spectrosc. Radiat. Transfer*, **61**, 509-517, 1999.

Casadio, S., D. Loyola, and C. Zehner, GOME-MERIS cloud products inter-comparison on global scale, Atmospheric Science Conference, Frascati, 2006.

Caudill, T., D. Flittner, B. Herman, O. Torres, and R. McPeters, Evaluation of the pseudo-spherical approximation for backscattered ultraviolet radiances and ozone retrieval, *J. Geophys. Res.*, **102**, 3881-3890, 1997.

Chance, K. and R. Spurr, Ring effect studies: Rayleigh scattering including molecular parameters for rotational Raman scattering, and the Fraunhofer spectrum, *Applied Optics*, **36**, 5224-5230, 1997.

Chandrasekhar, S., Radiative Transfer, Dover Publications Inc., New York, 1960.

Coldewey-Egbers, M., M. Weber, L. Lamsal, R. de Beek, M. Buchwitz, and J. Burrows, Total ozone retrieval from GOME UV spectral data using the weighting function DOAS approach, *Atmos. Chem. Phys.,* **5**, 1015–1025, 2005.

Coldewey-Egbers, M., S. Slijkhuis, B. Aberle, and D. Loyola, Long-term analysis of GOME in-flight calibration parameters and instrument degradation, *Appied Optics*, **47**, 4749-4761, 2008.

Dahlback, A., and K. Stamnes, A new spherical model for computing the radiation field available for photolysis and heating at twilight, *Planet. Space Sci.*, **39**, 671, 1991.

Daumont, D., J. Brion, J. Charbonnier, and J. Malicet, Ozone UV spectroscopy. I. Absorption cross-sections at room temperature, *J. Atmos. Chem.*, **15**, 145-155, 1992.

Doicu, A., F. Schreier, and M. Hess, Iterative regularization methods for atmospheric remote sensing, *J. Quant. Spectrosc. Radiat. Transfer*, **83**, 47-61, 2004.

Doicu, A., F. Schreier, S. Hilgers, A. von Bargen, S. Slijkhuis, M. Hess, and B. Aberle, An efficient inversion algorithm for atmospheric remote sensing with application to UV limb observations, *J. Quant. Spectrosc. Radiat. Transfer*, **103**, 193-208, 2007.

Eskes, H, and K. F. Boersma, Averaging kernels for DOAS total-column satellite retrievals, *Atmos. Chem. Phys*., **3**, 1285-1291, doi:10.5194/acp-3-1285-2003, 2003.

Eskes, H., R. van der A, E. Brinksma, J. Veefkind, J. de Haan, and P. Valks, Retrieval and validation of ozone columns derived from measurements of SCIAMACHY on Envisat, *Atmos. Chem. Phys. Discuss.,* **5**, 4429–4475, 2005.

ESA, GOME Global Ozone Monitoring Experiment Users Manual, ed. F. Bednarz, ESA SP-1182, 1995.

ETOP05, Data Announcement 88-MGG-02, Digital relief of the Surface of the Earth. NOAA, National Geophysical Data Center, Boulder, Colorado, 1988.

Grainger, J., and J. Ring, Anomalous Fraunhofer Line Profiles, *Nature*, **193**, 762, 1962.

Hassler, B., G. Bodeker, and M. Dameris, Technical Note: A new global database of trace gases and aerosols from multiple sources of high vertical resolution measurements, *Atmos. Chem. Phys*., **8**, 5403-5421, 2008.

Hasekamp, O., J. Landgraf, and R. van Oss, The need of polarization monitoring for ozone profile retrieval from backscattered sunlight, *J. Geophys. Res*., **107**, 4692, 2002.

Herman, J., and E. Celarier, Earth surface reflectivity climatology at 340 nm to 380 nm from TOMS data, *J. Geophys. Res*., **102**, 28003-28011, 1997.

Joiner, J., and P. Bhartia, Accurate determination of total ozone using SBUV continuous spectral scan measurements, *J. Geophys. Res.*, **102**, 12957-12969, 1997.

Kaifel, A., J. Kaptur, and M. Felder, NNORSY-GOME Tropospheric Ozone Products, The Remote Sensing of Atmospheric Constituents from Space, Accent Troposat-2, Annual Report, 2008.

Kneizys, F., E. Shettle, L. Abreu, J. Chetwynd, G. Anderson, W. Gallery, J. Selby, and S. Clough, Users Guide to LOWTRAN 7, Air Force Geophysics Laboratory, Environmental Research Papers, No. 1010, AFGL-TR-88-0177, 1998.

Koelemeijer, R., and P. Stammes, A fast method for retrieval of cloud parameters using oxygen *A* band measurements from the Global Ozone Monitoring Experiment, *J. Geophys. Res*., **106**, 3475-3490, 2001.

Koelemeijer, R., J. de Haan, J. Hovenier, and P. Stammes, A database of spectral surface reflectivity in the range 335-772 nm derived from 5.5 years of GOME observations, *J. Geophys. Res.*, **108,** D4070, doi:10.1029/2002JD0024, 2003.

Kuze, A., and K. Chance, Analysis of Cloud-Top Height and Cloud Coverage from Satellites Using the O_2 A and B Bands, *J. Geophys. Res.*, **99**, 14481-14491, 1994.

Lacis, A., J. Chowdhary, M. Mishchenko, and B. Cairns, Modeling errors in diffuse sky radiance: vector vs. scalar treatment, *Geophys. Res. Lett*., **25**, 135-138, 1994.

Lambert, J.-C., et al., A climatology of NO_2 profile for improved Air Mass Factors for ground-based vertical column measurements, Proc. 5th Europ. Symp. on Stratosph. Ozone, Air Pollution Research Report (CEC DG XII), **EUR 19340**, 703-706, 2000.

Lambert, J.-C., G. Hansen, V. Soebijanta, W. Thomas, M. Van Roozendael, D. Balis, C. Fayt, P. Gerard, J. Gleason, J. Granville, G. Labow, D. Loyola, J. van Geffen, R. van Oss, C. Zehner, and C. Zerefos, "ERS-2 GOME GDP3.0 Implementation and Validation, ESA Technical Note ERSE-DTEX-EOAD-TN-02-0006", 138 pp., Ed. by J.-C. Lambert (IASB), 2002.

Lerot, C., M. Van Roozendael, J. van Geffen, J. van Gent, C. Fayt, R. Spurr, G. Lichtenberg, and A. von Bargen, Six years of total ozone column measurements from SCIAMACHY nadir observations, *Atmos. Meas. Tech.,* **2**, 87–98, 2009.

Lerot, C., M. van Roozendael, J. van Gent, D. Loyola, and R. Spurr, The GODFIT algorithm: a direct fitting approach to improve the accuracy of total ozone measurements from GOME, *Int. J. Remote Sensing*, **31**(2), 543-550, doi: 10-1080/01431160902893576, 2010.

Levelt, P., G. van den Oord, M. Dobber, A. Malkki, H. Visser, J. de Vries, P. Stammes, J. Lundell, and H. Saari, The Ozone Monitoring Instrument, IEEE *Trans. Geosc. Rem. Sens*., **44**(5), 1093-1101, 2006.

Liu, X., K. Chance, C. Sioris, R. Spurr, T. Kurosu, R. Martin, and M. Newchurch, Ozone Profile and Tropospheric Ozone Retrievals from Global Ozone Monitoring Experiment: Algorithm Description and Validation, *J. Geophys. Res.*, **110**, D20307, doi:10.1029/2005JD006240, 2005.

Loyola, D., B. Aberle, W. Balzer, K. Kretschel, E. Mikusch, H. Muehle, T. Ruppert, C. Schmid, S. Slijkhuis, R. Spurr, W. Thomas, T. Wieland, and M. Wolfmüller, Ground Segment for ERS-2 GOME Data Processor, 3rd Symposium on Space in the Service of our Environment, Florence, Italy, *ESA SP-414*, 591-597, 1997.

Loyola, D., and T. Ruppert, A new PMD cloud-recognition algorithm for GOME, *ESA Earth Observation Quarterly*, **58**, 45-47, 1998.

Loyola, D., Cloud retrieval for SCIAMACHY, http//:wdc.dlr.de/sensors/gome/gdp4/loyola_2000.pdf, 2000

Loyola, D., Automatic Cloud Analysis from Polar-Orbiting Satellites using Neural Network and Data Fusion Techniques, *IEEE International Geoscience and Remote Sensing Symposium*, **4**, 2530-2534, Alaska, 2004.

Loyola, D., W. Thomas, Y. Livschitz, T. Ruppert, P. Albert, and R. Hollmann, Cloud properties derived from GOME/ERS-2 backscatter data for trace gas retrieval, IEEE *Trans. Geosc. Rem. Sens*., **45**(9), 2747-2758, 2006.

Loyola, D., A semi-transparent Lambertian cloud model for ozone retrieval, DLR presentation, September 2007.

Loyola D., Thomas W., Spurr, R., B. Mayer, Global patterns in daytime cloud properties derived from GOME backscatter UV-VIS measurements, *Int. J. Remote Sensing*, **31**(16), 4295-4318, doi: 10-1080/01431160903246741, 2010.

Loyola D., Koukouli M. E., Valks P., Balis D. S., Hao N., Van Roozendael M., Spurr R. J. D., Zimmer W., Kiemle S., Lerot C., Lambert J.-C. The GOME-2 total column ozone product: Retrieval algorithm and ground-based validation, Journal of Geophysical Research, vol. 116, D07302, 2011.

Malicet, J., D. Daumont, J. Charbonnier, C. Parisse, A. Chakir, and J. Brion, Ozone UV Spectroscopy. II. Absorption Cross-Sections and Temperature Dependence, *J. Atmos. Chem*., **21**, 263-273, 1995.

Matthews, E., Global Vegetation and Land Use: New High-Resolution Data Bases for Climate Studies, *J. Clim. Appl. Met.*, ***22***, 474-487, 1983.

Mishchenko M., A. Lacis, and L. Travis, Errors induced by the neglect of polarization in radiance calculations for Rayleigh scattering atmospheres, *J. Quant. Spectrosc. Radiat. Transfer*, **51**, 491-510, 1998.

Müller, M. D., A. Kaifel, M. Weber, S. Tellmann, J. Burrows, and D. Loyola, Ozone profile retrieval from Global Ozone Monitoring Experiment (GOME) data using a neural network approach (Neural Network Ozone Retrieval System (NNORSY)), *J. Geophys. Res.*, **108(**D16), 4497, doi:10.1029/2002JD002784, 2003.

Munro, R., M. Eisinger, C. Anderson, J. Callies, E. Corpaccioli, R. Lang, A. Lefebvre, Y. Livschitz, and A. Perez Albinana, GOME-2 on METOP: From in-orbit verification to routine operations, in: Proceedings of EUMETSAT Meteorological Satellite Conference, Helsinki, Finland, 12-16 June 2006.

Orphal, J., A critical review of the absorption cross-sections of O_3 and NO_2 in the 240--790 nm region, *Journal of Photochemistry and Photobiology A,* **157**, 185-209, 2003.

Richter, A., and J. Burrows, Tropospheric NO_2 from GOME measurements, *Adv. Space Res.*, **29**, 1673-1683, 2002.

Rodgers, C., Inverse Methods for Atmospheres: Theory and Practice, *World Scientific Press*, 2001.

Rothman, L., et al., The HITRAN molecular spectroscopic database: edition of 2000 including updates through 2001, *J. Quant. Spectrosc. Radiat. Transfer*, **82**, 5-44, 2003.

Rozanov, A., V. Rozanov, and J. Burrows, Combined differential-integral approach for the radiation field computation in a spherical shell atmosphere: Nonlimb geometry, *J. Geophys. Res.*, **105**, 22937-22942, 2000.

Rozanov, A., A. Kokhanovsky, D. Loyola, R. Siddans, B. Latter, A. Stevens, and J. Burrows, Intercomparison of cloud top altitudes as derived using GOME and ATSR-2 instruments onboard ERS-2, *Remote Sensing of Environment*, **102**, 186-193, 2006.

Schiffer, R., and W. Rossow, The international satellite cloud climatology project ISCCP: The first project of the world climate research programme, *Bull. Am. Meteorol. Soc.*, **54**, 779-784, 1983.

Siddans, R., B. Kerridge, W. Reburn, A. Stevens, and R. Munro, Height-resolved ozone retrievals spanning the troposphere and stratosphere from GOME, ESAMS '99 - European Symposium on Atmospheric Measurements from Space, Noordwijk, The Netherlands, ESA WPP-161, 299-305, 1999.

Siewert, C., A concise and accurate solution to Chandrasekhar's basic problem in radiative transfer, *J. Quant. Spectrosc. Radiat. Transfer*, **64**, 109-130, 2000.

Sioris, C. and W. Evans, Filling-in of Fraunhofer and gas-absorption lines as caused by rotational Raman scattering, *Applied Optics*, **38**, 2706-2713, 1999.

Slijkhuis S., A. von Bargen, W. Thomas, and K. Chance, Calculation of Under-sampling correction spectra for DOAS spectral fitting, ESAMS'99 - European Symposium on Atmospheric Measurements from Space, Noordwijk, The Netherlands, *ESA WPP-161*, 563-569, 1999.

Slijkhuis, S., B. Aberle, and D. Loyola, GOME Data Processor Extraction Software User's Manual, ER-SUM-DLR-GO-0045, 2004.

Spurr, R., Simultaneous derivation of intensities and weighting functions in a general pseudo-spherical discrete ordinate radiative transfer treatment, *J. Quant. Spectrosc. Radiat. Transfer*, **75**, 129-175, 2002.

Spurr, R., LIDORT V2PLUS, a comprehensive radiative transfer package for nadir viewing spectrometers, Proceedings SPIE conference, Barcelona, 2003.

Spurr, R., A New Approach to the Retrieval of Surface Properties from Earthshine Measurements, *J. Quant. Spectrosc. Radiat. Transfer*, **83**, 15-46, 2004.

Spurr, R., D. Loyola, W. Thomas, W. Balzer, E. Mikusch, B. Aberle, S. Slijkhuis, T. Ruppert, M. Van Roozendael, J.-C. Lambert, and T. V. Soebijanta, GOME Level 1-to-2 Data Processor

Version 3.0 (2005): A Major Upgrade of the GOME/ERS-2 Total Ozone Retrieval Algorithm, *Applied Optics*, **44**, 7196-7209, 2005.

Spurr, R. J. D., VLIDORT: A linearized pseudo-spherical vector discrete ordinate radiative transfer code for forward model and retrieval studies in multilayer multiple scattering media, *J. Quant. Spectrosc. Radiat. Transfer*, **102**(2), 316-342, doi:10.1016/j/jqsrt.2006.05.005, 2006.

Spurr, R., J. de Haan, R. van Oss, and A. Vasilkov, Discrete Ordinate Theory in a Stratified Medium with First Order Rotational Raman Scattering; a General Quasi-Analytic Solution, *J. Quant. Spectrosc. Radiat. Transfer*, **109**, 404-425, doi: 10.1016/j.jqsrt.2007.08.011, 2008.

Spurr, R., LIDORT and VLIDORT: Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems. *Light Scattering Reviews*, Volume 3, ed. A. Kokhanovsky, Springer, 2008.

Stammes, P., P. Levelt, J. de Vries, H. Visser, B. Kruizinga, C. Smorenburg, G. Leppelmeier, and E. Hilsenrath, Scientific requirements and optical design of the Ozone Monitoring Instrument on EOS-CHEM, Proceedings of the SPIE Conference on Earth Observing Systems IV, July 1999, Denver, Colorado, USA, vol. SPIE 3750, 221-232, 1999.

Stamnes, K., S.-C.Tsay, W. Wiscombe, K. Jayaweera, Numerically stable algorithm for discrete ordinate method radiative transfer in multiple scattering and emitting layered media, *Applied Optics*, **27**, 2502-2509, 1988.

Tanskanen, A., A. Arola, and J. Kujanpää, Use of the moving time-window technique to determine surface albedo from the TOMS reflectivity data, In: Proc. SPIE Vol. 4896, p. 239--250, 2003.

Tanskanen, A. Lambertian Surface Albedo Climatology at 360 nm from TOMS Data Using Moving Time-Window Technique. In: Proceedings of the XX Quadrennial Ozone Symposium, 1-8 June 2004, Kos, Greece.

Thomas, G., and K. Stamnes, Radiative Transfer in the Atmosphere and Ocean, 1st ed., Cambridge University Press, 1999.

Thomas W., T. Erbertseder, T. Ruppert, M. van Roozendael, J. Verdebout, D. Balis, C. Meleti, and C. Zerefos**,** On the Retrieval of Volcanic Sulfur Dioxide Emissions from GOME Backscatter Measurements, *J. Atmos. Chem*., **50**, 295-320, 2005.

Tuinder, O., R. de Winter-Sorkina, and P. Builtjes , Retrieval methods of effective cloud cover for the GOME instrument: an intercomparison, *Atmos. Chem. Phys*., **4**, 255-273, 2004.

Van der A, R., R. van Oss, A. Piters, J. Fortuin, Y. Meijer, and H. Kelder, Ozone profile retrieval from recalibrated GOME data, *J. Geophys. Res.*, **107**(D15), 10.1029/2001JD000696, 2002.

Van Geffen, J., and R. van Oss, Wavelength calibration of spectra measured by the Global Ozone Monitoring Experiment by use of a high-resolution reference spectrum, *Applied Optics*, **42**, 2739-2753, 2003.

Van Oss, R., and R. Spurr, Fast and accurate 4 and 6 stream linearized discrete ordinate radiative transfer models for ozone profile retrieval, *J. Quant. Spectrosc. Radiat. Transfer*, **75**, 177-220, 2002.

Van Oss, R., R. Voors, and R. Spurr, Ozone Profile Algorithm, OMI Algorithm Theoretical Basis Document, Volume II, OMI Ozone products (P.K. Bhartia ed.), ATBD-OMI-02, Version 2.0, 2002.

Van Roozendael M., V. Soebijanta, C. Fayt, and J.-C. Lambert, Investigation of DOAS Issues Affecting the Accuracy of the GDP Version 3.0 Total Ozone Product, in ERS-2 GOME GDP

3.0 Implementation and Delta Validation, Ed. J-C. Lambert, ERSE-DTEX-EOAD-TN-02-0006, ESA/ESRIN, Frascati, Italy, Chap.6, pp.97-129, 2002.

Van Roozendael, M., D. Loyola, R. Spurr, D. Balis, J-C. Lambert, Y. Livschitz, P. Valks, T. Ruppert, P. Kenter, C. Fayt, and C. Zehner, Ten years of GOME/ERS2 total ozone data: the new GOME Data Processor (GDP) Version 4: I. Algorithm Description, *J. Geophys Res.*, doi: 10.1029/2005JD006375, 2006.

Vountas, M., V. Rozanov, and J. Burrows, Ring effect: Impact of rotational Raman scattering on radiative transfer in earth's atmosphere, *J. Quant. Spectrosc. Radiat. Transfer*, **60**, 943-961, 1998.

Wang, P., P. Stammes, R. van der A, G. Pinardi, and M. van Roozendael, FRESCO+: an improved O_2 A-band cloud retrieval algorithm for tropospheric trace gas retrievals, *Atmos. Chem. Phys. Discuss*., **8**, 9697-9729, 2008.

Weber, M., L. Lamsal, M. Coldewey-Egbers, K. Bramstedt, and J. Burrows, Pole-to-pole validation of GOME WFDOAS total ozone with ground based data, *Atmos. Chem. Phys. Discuss.,* **4,** 6909–6941, 2005.

Wellemeyer, C., S. Taylor, C. Seftor, R. McPeters and P. Bhartia, A correction for total ozone mapping spectrometer profile shape errors at high latitude, *J. Geophys. Res.,* **102**, 9029-9038, 1997.