THE GODFIT (GOME DIRECT FITTING) ALGORITHM: A NEW APPROACH FOR TOTAL COLUMN RETRIEVAL

R. Spurr[(1)], M. Van Roozendael[(2)], C. Fayt[(2)], J.-C. Lambert[(2)]

[(1)]Harvard-Smithsonian Center for Astrophysics,
60 Garden Street, Cambridge, MA 02138, USA, rspurr@cfa.harvard.edu
[(2)]Belgian Institute for Space Aeronomy (BIRA-IASB),
3 Av. Circulaire, 1180 Brussels, Belgium

ABSTRACT

We present a new direct fitting algorithm for retrieval of total column amounts from remote sensing UV/visible spectrometers. The GODFIT algorithm for GOME total ozone makes a single fit for the vertical column; there is no DOAS slant-column fitting and AMF conversion. GODFIT relies on a 1-1 column-profile map for ozone; all forward model simulations are done from scratch using the LIDORT linearized radiative transfer model. The algorithm includes a new molecular Ring correction implementation. A 2300-orbit validation showed closer agreement with ground-based results (excepting some polar regions), and much reduced scatter and seasonal variation. These findings are in line with results from the new generation of DOAS algorithms.

1. INTRODUCTION

1.1 Background

The program for long-term global monitoring of total ozone from remote-sensing UV/visible spectrometers is in a transitional phase. GOME (launched on board the ERS-2 satellite in April 1995) is still operational, but in the last two years, data has been sporadic (problems with tape storage) and the quality has dropped because of instrument degradation. SCIAMACHY was launched in March 2002 on the ENVISAT platform, and GOME-2 is scheduled for launch in early 2006 on board the first METOP satellite. OMI (Ozone Monitoring Instrument) was launched in July 2004 on NASA's Aura platform, and will replace the older TOMS instruments.

ESA-ESRIN funded three studies in 2003 to investigate improved algorithms for GOME total ozone. This paper reports work done by the BIRA/SAO consortium. The main outcome of this work is the development of a new algorithm GODFIT (GOme Direct FITting). A second outcome is the development of GDOAS, an improved DOAS algorithm that has been implemented in 2004 as part of GDP 4.0 at DLR, for the purpose of reprocessing the entire 9-year GOME data record. This paper summarizes the GODFIT part of the study; a separate paper in these Proceedings deals with GDOAS.

1.2 Rationale

The main operational Level 2 product from GDP (GOME Data Processor [1]) is the global distribution of total vertical column amounts of ozone. Version 3.0 of the GDP algorithm is the latest to be validated [2]. GDP 3.0 is a traditional DOAS-style algorithm, although it uses an iterative scheme for AMF determination based on profile-column adjustment, and a neural-network training scheme for fast extraction of AMF results. For details see [3]. An error study on GDP 3.0 [4] has helped to characterize many error sources in the algorithm; most noteworthy is the need for molecular Ring correction. Both GODFIT and GDOAS contain a new implementation for this correction.

GODFIT makes a direct comparison between simulated radiances and GOME measurements. There is a single-step direct fit of the state vector comprising total ozone and a number of ancillary parameters. The inversion is non-linear least-squares minimization. Simulated radiances and weighting functions are calculated "on-the-fly" using the linearized radiative transfer (RT) model LIDORT in the forward model. There is a 1-1 correspondence between fitted total columns and ozone profiles required as input for RT. Each forward model step is followed by the molecular Ring correction. Ancillary fitting parameters are the Fraunhofer Ring scaling and several albedo closure coefficients. The main Level 2 product is total ozone; there are no intermediate DOAS products (slant columns, AMFs).

The algorithm description in section 2 comprises an overview and discussion of key ideas (section 2.1), a discussion of RT and forward model aspects (2.2), a summary of the molecular Ring correction (2.3), a section on the inverse model (2.4), and some notes on the algorithm configuration (2.5). Section 3 contains a digest of the validation exercise carried out as part of the ESA-ESRIN study. For more details, refer to the GODFIT ATBD [5] and Validation Report [6].

2. ALGORITHM DESCRIPTION

2.1 Overview and Key Ideas

The flowchart in Fig. 1 gives an overview of GODFIT for the GOME total ozone application.

Proc. of the 2004 Envisat & ERS Symposium, Salzburg, Austria
6-10 September 2004 (ESA SP-572, April 2005)

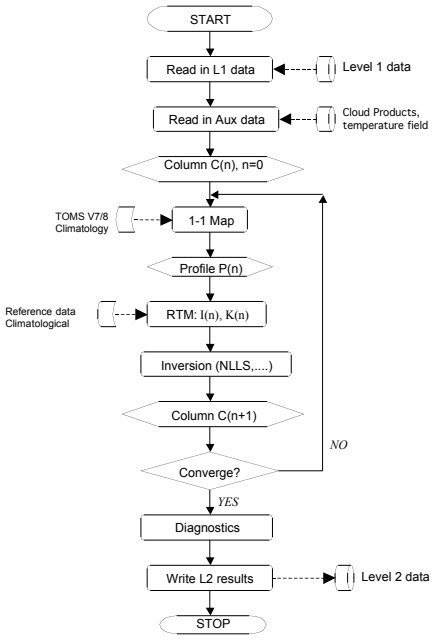

Fig 1. Flow Diagram for GODFIT

The algorithm is a standard non-linear iterative minimization of a chi-square merit functional, with a linearization of the forward model at each iteration step. Following the initial read of GOME Level 1b data (radiances and irradiances), and auxiliary data (cloud information, temperature profiles), an initial guess is made for the total column. A unique ozone profile is then constructed using a 1-1 column-profile map based on column-classified ozone profile climatology (e.g. TOMS V7 or V8 data). This profile is a basic input to the forward model RT simulation of top-of-atmosphere (TOA) radiances and weighting functions (Jacobians) for ozone column and albedo. The iteration stops when suitable convergence criteria have been satisfied. Fitting parameter errors come from the solution variance-covariance matrix; the results are then written to file.

All RT simulations are done from scratch using the linearized scattering code LIDORT [7]. Inputs depend on reference and climatological data (trace gas cross-sections, temperature and pressure profiles, albedo and topography) and on ingested cloud information (fractional cover, cloud-top pressure). There are no multidimensional look-up tables; these are cumbersome to manipulate and take time to create. GODFIT is more modular and flexible; it is easy to replace reference data and climatologies. With modern computing power, the use of "on-the-fly" RT simulations of radiances and Jacobians is feasible from a data turnover standpoint.

Direct fitting: column-profile mapping

In GODFIT, total ozone determines the profile shape uniquely, and we now define a correspondence between profiles and columns. Column-classified ozone profile climatologies were created for TOMS Version 7 and 8 retrieval algorithms [8-9]. If the profile is represented as a set $[U_j]$ of partial columns, the total column is $C = \Sigma U_j$. For two 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 the linear relation:

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

Column weighting functions are related to profile Jacobians through the derivatives of Eq. (1):

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

This map allows us to interpolate smoothly between profile entries in the climatology; we are drawing on an ensemble of possible profiles of which the climatology is a sample. In Fig. 2, we show this map for a set of 10 high-latitude TOMS profiles. TOMS Version 7 profiles were used as the baseline in GODFIT work reported here. Pressure levels defined in the TOMS data were used to set the vertical layering for RT simulations.

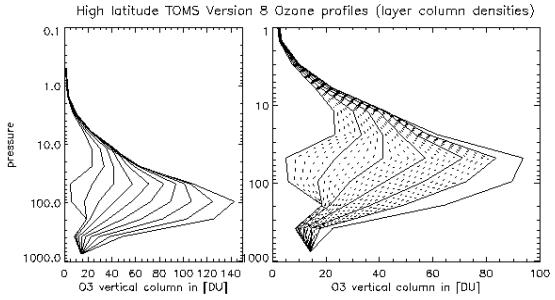

Fig. 2. (Left) TOMS high latitude profiles. (Right) Intermediate profiles (dotted) for columns at 10 Dobson Unit (DU) intervals between climatology values.

Direct fitting: internal albedo closure

Scattering from tropospheric aerosols is very difficult to decouple from surface reflectivity in the UV ozone range. In DOAS slant column fitting, all broadband radiative effects (due to clouds and aerosols, surface reflection and molecular scattering) are filtered out with a low-order polynomial. This filtering is an *external* closure; given the highly simplified forward model assumption (Beer-Lambert extinction law), DOAS closure parameters have no physical interpretation.

In GODFIT, aerosols and surface reflectivity radiative effects are brought together in an *albedo closure term* that is fitted internally: coupling between surface and tropospheric scatterers is treated properly in a multiple scattering context. Thus GODFIT determines (along with total ozone) an *effective* wavelength-dependent

albedo in a *molecular* atmosphere. This avoids introduction of extraneous (and highly uncertain) aerosol information into the forward model. Sensitivity results have shown that this albedo closure treatment makes the algorithm insensitive to the presence of aerosols (even absorbing aerosols in the troposphere). GODFIT is simplified enormously by performing all RT calculations in a Rayleigh-only atmosphere, with the effective surface albedo (assumed Lambertian) treated as a low order polynomial in wavelength.

2.2 The GODFIT Forward Model

Radiances and Jacobians from LIDORT

In an iterative inversion, the forward model must deliver simulated radiances and Jacobians with respect to fitted parameters. Multiple scattering is important in the UV and we use the linearized scattering code LIDORT [7]. The atmosphere is assumed stratified, with a number of optically uniform layers. For each layer, the radiative transfer equation (RTE) is solved by separating the radiation field azimuth dependence, then approximating the multiple scatter integral by means of a quadrature scheme that allows solutions to be found at discrete ordinates (quadrature polar directions). The resulting set of coupled linear differential equations is solved for the homogeneous solutions via eigenmethods, and for the particular solutions (due to scattering of the solar beam) by substitution. LIDORT uses the pseudo-spherical approximation: solar beam attenuation before scattering is calculated in a curved spherical shell atmosphere.

Layer solutions are combined by means of boundary conditions: (1) no downwelling diffuse radiation at TOA; (2) surface reflection at the lower boundary; and (3) continuity of the radiation field at intermediate levels. Solution of the resulting boundary value problem (a sparse linear algebra system) generates the complete radiance field at discrete ordinates. To obtain the field at arbitrary stream angle, the discrete ordinate solution is substituted in the original RTE, which is then integrated over optical thickness ("post processing" via source function integration). Azimuth components are summed until convergence is reached.

LIDORT inputs for layer p are the extinction optical thickness τ_p, the total single scatter albedos ω_p, and the total phase function scattering coefficients β_{pl}. These quantities are constructed from atmospheric profiles of temperature, pressure, trace gas amounts and aerosol loading, plus details of molecular scattering, absorption cross-sections, and aerosol optical properties (see below, section 2.3). We assume a Lambertian surface characterized by a total albedo.

The basic input set $\{\tau_p, \omega_p, \beta_{pl}\}$ will depend on the total ozone column C, and in order to derive a Jacobian we require the basic *derivative* inputs $\partial\tau_p/\partial\xi$, $\partial\omega_p/\partial\xi$ and $\partial\beta_{pl}/\partial\xi$. with respect to variable ξ in layer p. *Profile* Jacobians are then computed by explicit analytic differentiation of each RTE solution step. This involves repeated applications of the chain rule to each component part of the radiance field, always expressing the answers in terms of this basic derivative set. Finally we compute the *column* Jacobian using the chain rule:

$$K_{col}(0,\Omega) \equiv \frac{\partial I(0,\Omega)}{\partial C} = \sum_{p=1}^{n} \frac{\partial I(0,\Omega)}{\partial U_p} \frac{\partial U_p}{\partial C} \qquad (3)$$

Here, $(0,\Omega)$ denotes output at TOA in direction Ω. Note the use of profile-column mapping derivatives (Eq. (2)). Version 2.3E of LIDORT outputs a complete set of profile Jacobians; a column weighting function would then be summed as in Eq. (3). Version 2.5 of LIDORT (written specifically for GODFIT) performs the sum in Eq. (3) *internally*, and is much more efficient.

For the albedo Jacobian, we note that the layer inputs for LIDORT have no dependence on the albedo, which only appears through the boundary value problem. Thus for this Jacobian, we start with the linearized boundary problem and then work through differentiation of the post-processed solution.

Physical aspects of the RT computation

<u>*Pressure and temperature*</u>. We use TOMS V7 and V8 pressure levels to set atmospheric layering; pressures are halved at successive atmospheric boundaries. TOA is set at 0.03 hPa, and surface pressure is taken from auxiliary input. Temperature profiles are required for hydrostatic balance and determination of ozone cross sections. Height levels are determined from hydrostatic equilibrium based on reference surface topography. Layering is set up for clear-sky and full-cloud scenes.

TOMS V7 ozone climatology comes with a set of temperature profiles and these were used without interpolation. For TOMS V8, we used separate latitude-zone monthly temperature climatology. We also used historical analysis fields from the ECMWF NWP models (ERA-40). In GODFIT, temperature selections for ozone cross-sections are done on a layer-by-layer basis; we do not ingest or retrieve an effective "Bass-Paur" temperature as is done for DOAS fitting.

<u>*Ozone profiles*</u>. There are 26 TOMS V7 O_3 profiles, in 3 latitude zones, with 10 profiles from 125 DU to 575 DU at intervals of 50 DU in high and mid-latitudes, and 6 profiles from 225 DU to 475 DU at low latitudes [8]. Combining two sets of profiles with a cosine-latitude weighting incorporates latitude variation. Adjustments are made to tropospheric columns to account for lower boundary pressure.

<u>*Ozone cross-sections*</u>. Ozone cross sections come from the GOME FM 98 data set [10]; the baseline uses re-

sampled data including a solar I_0 correction [11], and the optimized pre-shift is 0.017 nm.

Molecular Scattering. Rayleigh scattering $\sigma^{Ray}(\lambda)$ will be determined from standard results [12]. In hydrostatic equilibrium, air column density D_p in layer p depends only on the pressure drop across the layer. In terms of ozone partial columns U_p, extinction optical thickness and single scatter albedo and their derivatives with respect to U_p are:

$$\delta_p = \sigma^{Ray}(\lambda)D_p + \sigma_p^{O3}(\lambda)U_p \;\; ; \;\; \omega_p = \frac{\sigma^{Ray}(\lambda)D_p}{\delta_p} \qquad (4)$$

$$\frac{\partial \delta_p}{U_p} = \sigma_p^{O3}(\lambda) \qquad \frac{\partial \omega_p}{\partial U_p} = -\frac{\omega_p}{\delta_p}\frac{\partial \delta_p}{\partial U_p} \qquad (5)$$

The Rayleigh phase function has a $cos^2\Theta$ dependence on scatter angle Θ, with depolarization ratio from [12].

Nitrogen Dioxide. Number density profiles in the stratosphere were taken from a new climatology compiled from HALOE and balloon data. NO_2 in the troposphere was not used (future work will include an NO_2 GODFIT application in the visible spectrum). GOME FM 98 NO_2 cross-sections [13] are appropriate. Although the GODFIT algorithm baseline for O_3 validation testing does not include NO_2 as an interfering species, some tests have been performed including this minor species in the fit (see [5]).

Clouds: treatment and input data. In the independent pixel approximation (IPA), TOA radiance in a partially cloudy scene is simulated as a linear combination of backscatter from clear-sky and full-cloud scenarios, weighted by the effective cloud fractional cover f_c. Clouds are treated as highly reflecting Lambertian surfaces. A separate "ghost column" calculation is not needed in GODFIT, since the profile-column map ensures that the tropospheric column below cloud-top (the "ghost") is automatically adjusted in the fit.

GODFIT ingests cloud information from the FRESCO O_2 *A*-band algorithm [14] which supplies retrieved values of f_c and effective cloud-top pressure q_c, plus errors on these two parameters. Supplementary data in the FRESCO input files include the surface albedo and pressure; the latter is taken as reference for hydrostatic balance. (GDP 3.0 uses the ICFA IPA algorithm for cloud pre-processing, but takes q_c from climatology).

Surface data. In GODFIT, the albedo closure term $R(\lambda)$ is a 3-parameter quadratic function of wavelength λ:

$$R(\lambda) = \gamma_0 + \gamma_1(1-\lambda/\lambda_0) + \gamma_2(1-\lambda/\lambda_0)^2 \qquad (6)$$

From the LIDORT albedo weighting function $K_R(\lambda)$, it is easy to derive Jacobians for closure coefficients γ_0, γ_1 and γ_2. We assume first guess values $\gamma_1 = \gamma_2 = 0$, and an initial value for γ_0 is taken from the database of Lambertian equivalent reflectivities (LER) prepared from 5.5 years of GOME reflectivity data [15]. Monthly data is on a 1°x1° grid; we use data at λ_0=335 nm.

Aerosols. The baseline for GODFIT is to dispense with explicit knowledge of aerosols in RT simulations. 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 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.

Polarization issues. In ozone profile retrievals, the polarization correction applied to Level 1 data is an important source of error; the inclusion of polarization in RT simulations is also relevant. However, in DOAS fitting with narrow windows, the polarization signature is buried in the low-filter polynomial; polarization is neglected in AMF RT calculations. The polarization default for GODFIT is the same: any polarization effects are subsumed by the internal albedo closure. We should note that GODFIT is applicable anywhere in the UV (it is possible to use windows with wavelengths as low as 312 nm at the beginning of GOME Band 2a), and we intend to look at polarization effects with the new vector LIDORT model.

2.3 Ring correction method

The Ring effect - Fraunhofer and telluric filling-in due to rotational Raman scattering (RRS) - induces a small-amplitude distortion in earthshine spectra. In DOAS-type algorithms, the Fraunhofer Ring effect is treated as a pseudo-absorber, with an additive scaling parameter fitted using a Ring reference spectrum computed from a convolution of rotational Raman (RR) cross-sections with a reference solar spectrum [16]. The 2002 study on GDP [4] pointed to the need for a second correction to deal with filling-in of ozone absorption features, and new molecular Ring correction methods were developed for the GODFIT and GDOAS algorithms.

The correction is based on a simplified forward model of intensity $I(\lambda)$ at the satellite; this includes an explicit contribution due to RRS, and for GODFIT we write:

$$\frac{I(\lambda)}{I^0(\lambda)} = \left[\frac{I(\lambda)}{F(\lambda)}\right]_{EL} + E_{Ring} I_0^{RRS}(\lambda) e^{\left[-\tau_{O3}^{RRS}\right]} \qquad (7)$$

The first term on the right is the sun-normalized LIDORT *elastic* scattering radiance, and $I_0^{RRS}(\lambda)$ is the usual zero-order Fraunhofer-RRS source spectrum, with fitting amplitude E_{ring}. τ^{RRS} is the effective optical density of ozone in the Raman light along the line-of-

sight path from surface (where RRS is assumed to occur) to satellite. To a first approximation, this can be evaluated from simple geometry:

$$\tau_{O3}^{RRS} = \tau_{O3}^{vertical} \cdot \sec(\theta_0) \quad (8)$$

Ozone absorption in Raman light can be represented with sufficient accuracy by means of a geometrical enhancement factor; θ_0 is the solar zenith angle (SZA). We neglect the impact of ozone absorption in the incident beam before generation of Raman photons, since O_3 absorption structures will be mostly scrambled in the RRS process. Tests have shown that Eq. (8) is valid for most observations, but for large SZA > 85°, we use a more accurate definition:

$$\tau_{O3}^{RRS} = \tau_{O3}^{vertical} \cdot \left[\sec(\theta_0) + \Phi(\theta_0) \cdot \frac{\sigma_{O3}^{RRS}}{\sigma_{O3}} \right] \quad (9)$$

$\Phi(\theta)$ is a geometrical enhancement factor of the incident beam (allowing for the Earth's sphericity), and σ^{RRS} an O_3 absorption cross-section for Raman scattering after smoothing by the Raman process. The two cross-sections in Eq. (9) are illustrated in Fig. 3.

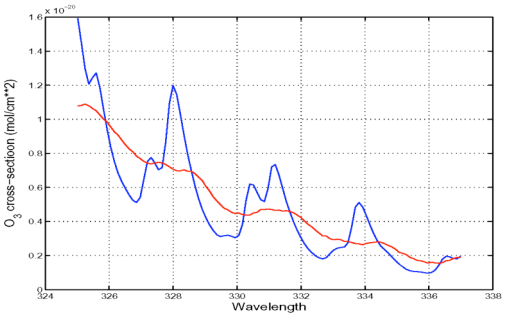

Fig. 3. O_3 absorption cross-sections at 241 K. (Blue) regular, (red) after smoothing by RRS.

The accuracy of this correction has been checked with closed-loop tests using look-up tables of synthetic radiances calculated with and without inelastic RRS.

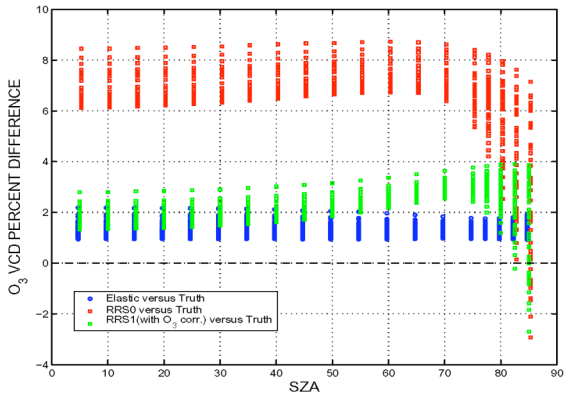

Fig. 4. (See text for explanation)

For GODFIT, the LIDORT-RRS model [17] was used to create an extensive LUT with 400 geometries, 7 surface albedos, 26 O_3 profiles and 11 surface pressures. All calculations were for a Rayleigh atmosphere. Although SZA and albedo are the major dependencies, lower boundary pressure is important because most inelastic scattering takes place in the lower troposphere. Corrections were calculated for 20 wavelengths between 315 and 335 nm, using a GOME solar spectrum at resolution ~0.105 nm. Fig. 4 compares O_3 column results with and without the new correction, displaying errors as a function of SZA. Blue dots denote errors on pure elastic retrievals (there is a small offset due to atmospheric set-up issues), and errors with no molecular Ring correction are shown in red. Green dots show residual errors when applying the new molecular Ring correction. The new correction significantly reduces the bias on O_3 retrievals.

2.4 Inverse model

Least squares minimization

The least squares fitting procedure minimizes the chi-square merit functional:

$$\lfloor \chi^2 \rfloor_{oise} = (\mathbf{Y}_m - F(\mathbf{X}))^{\mathrm{T}} \mathbf{S}_y^{-1} (\mathbf{Y}_m - F(\mathbf{X})) \quad (10)$$

Here, the measurement vector of TOA radiances is $\mathbf{Y}_m$, the state vector $\mathbf{X}$, the forward model radiance simulations $F(\mathbf{X})$, and the error covariance matrix $\mathbf{S}_y$. For $F(\mathbf{X})$ with non-linear dependence on $\mathbf{X}$, the fitting is non-linear least squares (NLLS), proceeding iteratively via a series of linearizations about $\mathbf{X}$ at each iteration. When the number of fitting parameters exceeds the number of pieces of independent information, it is usual to add a second regularization term to χ^2:

$$\chi^2 = \lfloor \chi^2 \rfloor_{oise} + (\mathbf{X} - \mathbf{X}_a)^{\mathrm{T}} \mathbf{S}_a^{-1} (\mathbf{X} - \mathbf{X}_a) \quad (11)$$

This is regularization for optimal estimation (OE): $\mathbf{X}_a$ and $\mathbf{S}_a$ are the *a priori* state vector and covariance matrix. Optimal estimates are found by iteration [18]:

$$\mathbf{X}_{i+1} = \mathbf{X}_a + \mathbf{D}_y \left[\mathbf{Y}_m - F(\mathbf{X}_i) - \mathbf{K}_i (\mathbf{X}_i - \mathbf{X}_a) \right] \quad (12)$$

$$\mathbf{D}_y = \mathbf{S}_{i+1} \mathbf{K}_i^{\mathrm{T}} \mathbf{S}_y^{-1} \quad ; \quad \mathbf{S}_{i+1} = \left(\mathbf{K}_i^{\mathrm{T}} \mathbf{S}_y^{-1} \mathbf{K}_i + \mathbf{S}_a^{-1} \right)^{-1} \quad (13)$$

$\mathbf{K}_i = dF(\mathbf{X}_i)/d\mathbf{X}_i$ is the matrix of Jacobians, $\mathbf{D}_y$ is the matrix of contribution functions, and $\mathbf{S}_{i+1}$ is the next-guess solution covariance matrix. [Results for the NLLS case are obtained by dropping the *a priori* terms]. The computation proceeds efficiently when one uses SVD (singular value decomposition) on the (scaled) matrix of Jacobians [18]. In terms of the singular values Λ_n resulting from SVD, the DFS (degrees of freedom for signal) indicator is:

$$DFS = \sum_{n=1}^{NP} \frac{\Lambda_n^2}{1+\Lambda_n^2} \qquad (14)$$

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 $\mathbf{X}$, then DFS = N_P when the measurement completely determines the state. 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. From Rodgers [18], we distinguish three error sources: (1) *measurement errors* due to random noise contributions are given by $\boldsymbol{\varepsilon}_{me} = \mathbf{D}_y\,\boldsymbol{\varepsilon}_y$ with solution covariance $\mathbf{S}_{\text{noise}} = \mathbf{D}_y\mathbf{S}_y\mathbf{D}_y^{\mathbf{T}}$; (2) *model parameter errors* $\boldsymbol{\varepsilon}_{\text{par}} = \mathbf{D}_y\mathbf{K}_b\boldsymbol{\Delta}\mathbf{b}$, where $\mathbf{K}_b$ is the sensitivity of the forward model to parameter $\mathbf{b}$, and $\boldsymbol{\Delta}\mathbf{b}$ is the error on $\mathbf{b}$, and if the error is random with covariance $\mathbf{S}_b$, the solution covariance contribution is $\mathbf{S}_{\text{par}} = \mathbf{D}_y\mathbf{K}_b\mathbf{S}_b\mathbf{K}_b^{\mathbf{T}}\mathbf{D}_y^{\mathbf{T}}$; and (3) *forward model errors* $\boldsymbol{\varepsilon}_{\text{fwd}} = \mathbf{D}_y\boldsymbol{\Delta}\mathbf{F}$, where $\boldsymbol{\Delta}\mathbf{F}$ is the forward model error (non-random) due to incorrect physical or mathematical assumptions in the model.

The inverse model in GODFIT

The baseline state vector in GODFIT has 6 parameters:

- Total column of O_3 (in [DU]) 1
- Albedo closure parameters 3
- Ring spectrum scaling factor 1
- Undersampling scaling factor 1

The Ring spectrum is the usual convolution of Raman cross-sections with a high-resolution Fraunhofer spectrum (additive term). The undersampling spectrum is also additive (default as for GDP). First guess values for total O_3 are taken from an external database (TOMS zonal mean climatology). An initial value for the first albedo parameter is taken from the LER database at 335 nm. The total column of NO_2 has been considered, but tests have shown that its inclusion does not influence the O_3 result in any significant way. With an internal albedo closure term, it has not been necessary to apply other closure (multiplicative or additive polynomials) in the fitting. Weighting functions for the above two scaling factors are trivial, since their radiance contributions are additively linear.

In GODFIT, we implemented both an OE package [19] and SLATEC NLLS code [20]. For OE, we set *a priori* variances at very high levels (loose constraints, so that DFS was always greater than 5.99). The *a priori* vector is the initial state vector. Although the GODFIT inverse problem is not ill conditioned, we found that OE is much faster and more transparent: SLATEC spends time searching in parameter space for convergence, and it also performs unnecessary finite difference checks on Jacobians (our Jacobians are analytic). We found that loosely constrained OE code converges quickly (3-4 iterations); it was also completely robust over one data granule (orbit). For iteration convergence, we looked at relative changes (between iterations) in χ^2 and O_3 column (0.1% criterion for the latter). The baseline window for most of the validation was 325-335 nm. Some tests were done using a window 331.6-336.6 nm.

2.5 Algorithm implementation

Processing issues

In an operational algorithm, some pre-processing needs to be done on Level 1b data before ingestion into GODFIT. The main issue is wavelength misregistration, often handled by doing preliminary fitting for shift (and sometimes squeeze) parameters. GODFIT uses sun-normalized radiances, so the spectral grid of the GOME solar irradiance spectrum (supplied with every orbit or backscatter data) is the wavelength baseline. We perform a preliminary fit for earthshine shift (magnitude ~0.08). Geolocation pre-processing requires that zenith and azimuth angles for line-of-sight and solar viewing directions be converted from values at spacecraft (level 1b data) to values at TOA.

Forward model results should be integrated over scan angle: for GOME forward scans, the line of sight zenith angle has changed by ~20°. For forward scans, the default is to perform a single calculation using only the center-scan geometry. Studies for the GOME ozone profile algorithm have shown that scan-angle integration can be approximated by a 2-point quadrature with little loss of accuracy (R. van Oss, private communication); we use this option for back-scans.

Algorithm verification is the closed-loop test ("quasi-perfect" retrieval): we generate synthetic Level 1b GOME data and add some noise, then run the algorithm again in retrieval mode using this synthetic data. If added noise is random, then (subject to noise uncertainty), the retrieval should return the "truth parameters" used to generate synthetic data.

Performance and portability

Portability is important, and "makefiles" were written for several platforms. The initial algorithm was written in FORTRAN 77 on Sun Unix, and was transferred to PCs with Linux and Windows-based operating systems. The algorithm was then tested on orbits of GOME data, and linked to validation software at BIRA-IASB. All whole-orbit runs were done on BIRA platforms.

GODFIT does all forward model simulations "on-the-fly", and it is expected that the algorithm is slower than DOAS algorithms with AMF look-ups. We use 13-layer and 8-stream discretizations in LIDORT – accurate enough for Rayleigh atmospheres, yet fast enough to achieve data turnover. A data granule of 1 orbit (~2000 pixels, a 10 nm window of ~100 spectral points, ~3.5

iterations for each fit) is processed in well under 15 minutes on the fastest PC at BIRA.

3. ALGORITHM VALIDATION

Here we examine the overall behavior of the GODFIT total ozone product through comparisons with a large subset of the NDSC network. Results are discussed in relation to both GDP 3.0 and TOMS data products. Validation methodology and selected data sets are noted below (section 3.1), and in section 3.2 we summarize the performance of GDP 3.0 total O_3 and outline some of its algorithmic problems, to set the scene for the main GODFIT validation summary in section 3.3.

3.1 Validation Methodology and Tools

We use comparisons of GOME data with ground-based GAW/NDSC measurements; these are reliable ground-based stations in all latitudinal bands, measuring total ozone using Dobson/Brewer instruments [21].

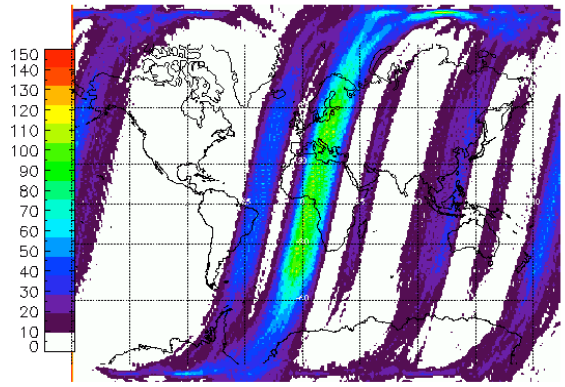

Fig 5. Validation orbits (courtesy, NASA)

The set of 2257 GOME orbits (Fig. 5) used for GODFIT validation is the same as for GDP 3.0 Δ-validation [2]. This is a good compromise between processing time and optimizing station coverage. An orbit is selected whenever the average value of relative GOME/NDSC difference is close to the median value. Most data records were taken from the 1996-97 period before the impact of significant GOME instrument degradation, but for long-term verification, data records were extended through 2001 for selected stations.

The extraction algorithm is applied in a consistent way to all data sets (same number of pixels): the same statistical analyses can be used on different data products. Relative differences between GOME ozone values and ground-based correlative measurements are then calculated, with further analysis and discussion of the results focusing on the most representative features (cyclic signatures, long-term stability, SZA and total column dependencies).

3.2 GDP 3.0 issues

GDP 3.0 results are the basis for the present study, so we note the main findings (see [2] for details). At Northern middle latitudes, the average agreement of GDP 3.0 with ground-based correlative measurements is within ±2-3%. At higher latitudes, a solar zenith angle (SZA) dependency appears. In addition, there is a persistent dependence of GDP data on ozone columns., For SZA < 70°, the average deviation of GOME from ground-based data does not exceed ±2-4%. At higher SZA, the average error ranges from -8% to +5% depending on season. Lowest total ozone values are overestimated by 5-10% during ozone hole conditions. A general view of the comparison between GDP 3.0 total ozone and NDCS measurements is seen in Fig. 6 for the 23 most significant NDSC stations. These are monthly averages (1996-2001) of relative differences between GOME and ground-based total ozone.

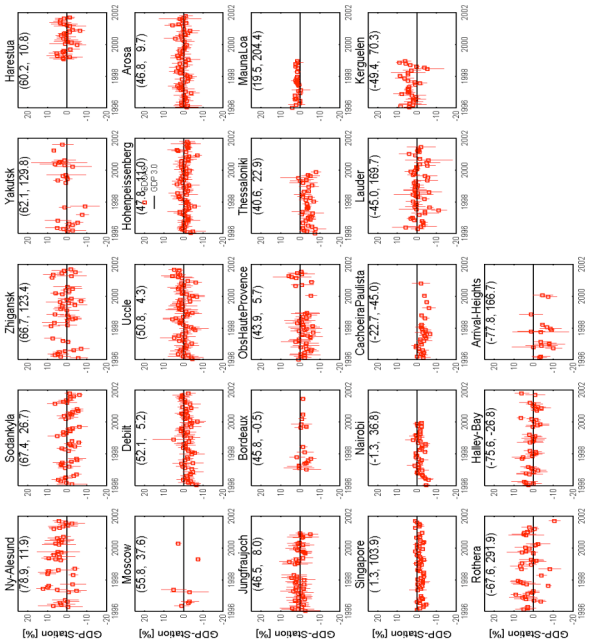

Fig. 6. Monthly averaged % differences in total O_3 GDP 3.0 vs. NDSC measurements, 23 stations 1996-01.

The main conclusions from the study on DOAS-related error sources [4] are: (1) retrievals in the 325-335 nm window using literature O_3 cross-sections [22] give slant columns ~2% larger than those using GOME FM data; (2) the slit function line shape is strongly asymmetric on both edges of channel 2; (3) an incorrect pre-shift for O_3 cross-sections in GDP 3.0 gives a +5 K bias in retrieved effective temperature (+1.5% bias in O_3 slant columns); (4) the lack of a solar-I_0 effect leads to an effective temperature bias of 2.5 K; (5) the lack of a molecular Ring correction leads to a negative bias in slant columns (5-10% range), with SZA dependence; (6) AMFs at 325.0 nm lead to a positive bias in retrieved columns at high SZA (> 80°), reduced by using AMFS at 325.5 nm These are issues for a DOAS algorithm, and all of them have been addressed in the GDOAS algorithm now part

of the GDP 4.0. A number of them (items 1-2, 3-5) are relevant to GODFIT; the most important is item 5.

3.3 GODFIT comparisons and validations

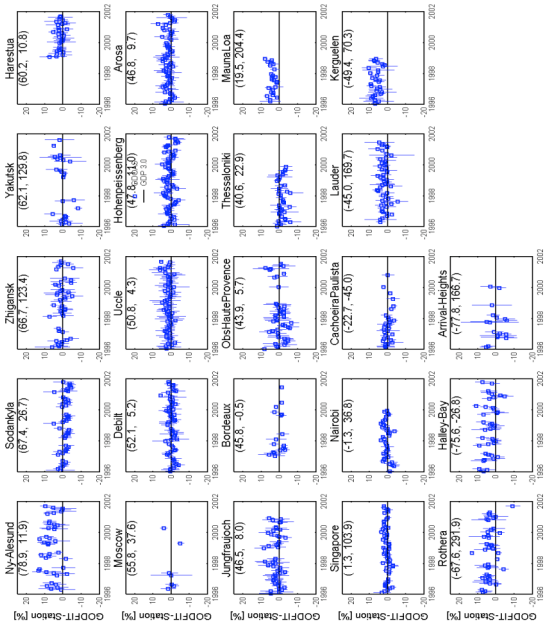

Fig. 7. GODFIT vs. NDSC.

Fig.7 gives a comparison overview for GODFIT, similar to that for GDP 3.0 in Fig. 6; sampling is the same for both GOME data sets. Both GODFIT and GDP 3.0 show no indication of significant drift. In Northern mid-latitudes, there are differences in behavior compared with GOME; this points to limitations in calibrations of ground-based instruments (stations have cyclic variations and dependencies).

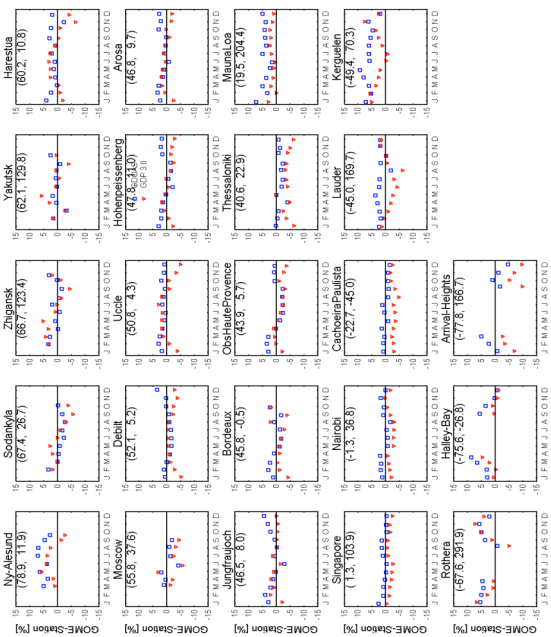

Fig 9. Monthly averaged % differences GODFIT retrievals (blue) or GDP 3.0 (red) and NDSC at 23 stations. Results are plotted as a function of month.

Average seasonal behavior is compared in Fig. 8. Mean seasonal variations of relative differences have different phase; GODFIT variations are generally smaller. For high latitudes (>70°), the ground-based comparisons still display significant seasonal dependencies similar to those for GDP 3.0; GODFIT values are larger by 2-5%.

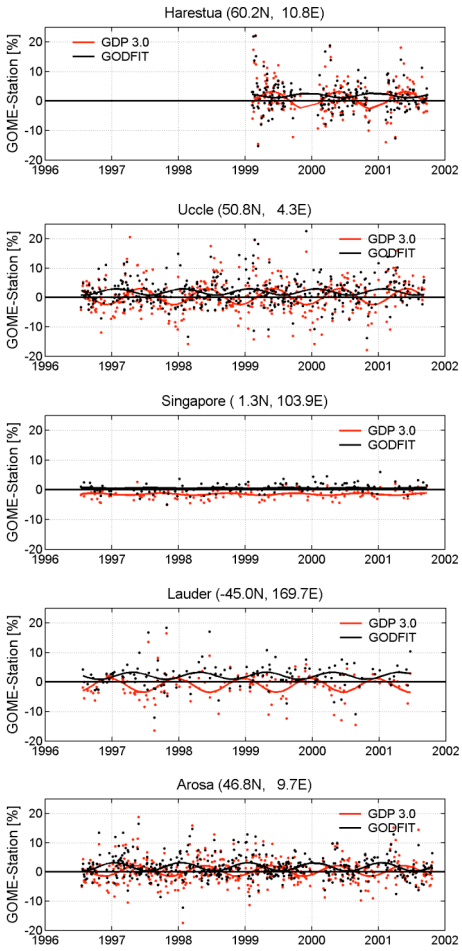

Fig. 9. Time series of total ozone relative differences between GODFIT (black dots) or GDP 3.0 (red dots) and ground measurements. Lines are one-year period cosine fits. Harestua, Uccle, Arosa, Singapore, Lauder.

Time-series of comparison data points were fitted to the following equation:

$$R = A + B \cdot \cos\left(2\pi\left[t - \Phi\right]\right) \qquad (15)$$

R is % relative difference between GOME and ground measurements; A is the offset parameter (mean relative difference); B is the amplitude of the annual wave variation; t is time in decimal year; Φ is the phase. Data points and fitted cosine curves are displayed graphically in Fig. 9 for 5 stations. For data-sparse high latitude

stations, simple averages are given. In Table 1, mean relative differences and wave amplitudes are averaged in zonal bands (high- and mid-latitudes and tropics).

From this analysis, GODFIT total ozone values are on average 2-3% larger than those from GDP 3.0, in better agreement with ground-based measurements in tropical regions and Northern mid-latitudes. Most significant is the clear reduction of cyclic variations. Annual wave amplitudes are reduced by ~40% at mid-latitudes in both hemispheres. Cyclic signatures characteristic of GODFIT retrievals are in most cases opposite in phase to those of GDP 3.0.

Table 1. Total O_3 mean % differences and seasonal wave amplitudes from cosine function fits to time-series of relative differences between GODFIT and GDP 3.0 and measurements at 18 ground-based stations. Averaged according to latitudinal bands.

Latitude band	GODFIT		GDP 3.0	
	A [%]	B [%]	A [%]	B [%]
> 70°N	10	-	6.7	-
60 – 67°N	0.9	2.1	0.12	3.4
30 – 60°N	0.45	1.4	-1.4	2.2
Tropics	0.4	0.7	-2.0	0.8
30° – 50°S	3.9	1.1	0.9	2.9
> 70°S	1.8	-	-2.1	-

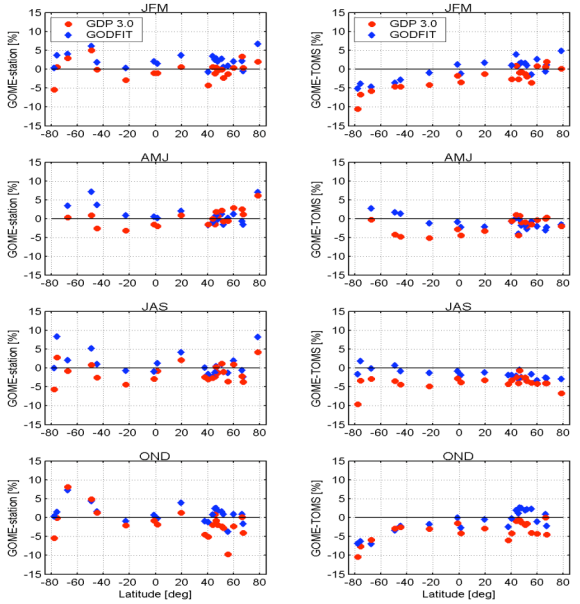

Fig. 10. Meridian and seasonal variation of percentage differences in total ozone, comparing GODFIT (blue) or GDP 3.0 (red) with NDSC (left panel), and collocated TOMS measurements (right panel).

In Fig. 10, we display meridian behavior of differences between GOME total O_3 retrievals (GODFIT and GDP 3.0) and ground-based correlative measurements (left panel), as well as collocated EP-TOMS satellite measurements (right). Each data point corresponds to one NDSC station. In tropical and Northern mid-latitude regions, meridian fluctuations are generally reduced in comparison with GDP 3.0. It is instructive to contrast the GOME/stations and GOME/TOMS comparisons. Excepting Northern mid-latitudes, there are less meridian fluctuations in comparisons between the two satellite products. This points to limitations in the internal consistency of the ground-based network. Comparing GOME retrievals and the TOMS Version 7 data product, overall consistency between the satellite instruments is significantly improved with GODFIT, with the exception of Southern high latitudes during the spring-summer period. However, it is known that TOMS shows a pronounced tendency to overestimate total ozone under these conditions.

Acknowledgements

We would like to acknowledge the contribution of Vincent Trisnanto Soebijanta, who died in tragic circumstances as this work was nearing completion.

Contributing NDSC PIs and ground-based instrument operators, DLR-IMF, science and Operation Teams are acknowledged for data provision. GOME level-1 and level-2 data products were processed at DLR-IMF on behalf of ESA.

4. REFERENCES

1. Loyola D., et al. Ground Segment for ERS-2 GOME Data Processor, 3rd Symposium on Space in the Service of our Environment, Florence, Italy, 1997.

2. Lambert J.-C., et al. ERS-2 GOME GDP 3.0 Implementation and Delta Validation, Validation Report for GOME Level-1-to-2 Data Processor Upgrade to Version 3.0, ERSE-DTEX-EOAD-TN-02-0006, ed. J.-C. Lambert, 2002.

3. Spurr R., et al. GOME Level 1-to-2 Data Processor Version 3.0: A Major Upgrade of the GOME/ERS-2 Total Ozone Retrieval Algorithm, *Applied Optics* (submitted), 2004.

4. Van Roozendael M., Soebijanta V., Fayt C. and Lambert J.-C., 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.

5. Spurr R. and van Roozendael M., Algorithm Theoretical Basis Document, GOME Direct Fitting (GODFIT), ESA ITT, AO/1-4235/02/I-LG, 2004.

6. Van Roozendael M. and Spurr R., Validation Report, GOME Direct Fitting (GODFIT), ESA ITT, AO/1-4235/02/I-LG, 2004.

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

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

9. Bhartia P., Algorithm Theoretical Baseline Document, TOMS V8 Total ozone algorithm, 2003. (http://toms.gsfc.nasa.gov/version8/version8_update.html)

10. Aliwell S., et al. Analysis for BrO in zenith-sky spectra: An intercomparison exercise for analysis improvement, *J. Geophys. Res.,* Vol. 107, D14, doi: 10.1029/2001JD000329, 2002.

11. Burrows J., et al. 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*, Vol. 61, 509-517, 1999.

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

13. Burrows J., et al. Atmospheric Remote-Sensing Reference Data from GOME: Part 1. Temperature-dependent Absorption cross-sections of NO_2 in the 231-794nm range, *J. Quant. Spectrosc. Radiat. Transfer*, Vol. 60, 1025-1031, 1998.

14. Koelemeijer R. and Stammes P., A fast method for retrieval of cloud parameters using oxygen *A* band measurements from GOME, *J. Geophys. Res.*, Vol. 106, 3475-3490, 2001.

15. Koelemeijer R., de Haan J. and Stammes P., A database of spectral surface reflectivity in the range 335--772 nm derived from 5.5 years of GOME observations*, J. Geophys. Res.*, Vol. 108, 4070, doi: 10.1029/2002JD0024, 2003.

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

17. Spurr R., Discrete Ordinate Theory in a Stratified Medium with First Order Rotational Raman Scattering; a General Quasi-Analytic Solution, 2004.

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

19. Van Oss R., Voors R. and Spurr R., Ozone Profile Algorithm, OMI Algorithm Theoretical Basis Document, Volume II, OMI Ozone products (ed. P. Bhartia), ATBD-OMI-02, Version 1.0, September 2001.

20. Fong K., et al. Guide to the SLATEC mathematical library, 1993. http://www.netlib.org/slatec/guide

21. Lambert J.-C., et al. Combined characterization of GOME and TOMS total ozone measurements from space using ground-based observations from the NDSC, *Adv. Space. Res*., Vol. 26, 1931-1940, 2000.

22. Bass A. and Paur R., The ultraviolet cross-sections of ozone: I. The measurements in Atmospheric ozone (Ed. C. S. Zerefos and A. Ghazi), Reidel, Dordrecht, Boston, Lancaster, pp. 606-610, 1985.