

## ENVISAT-1

### SCIAMACHY Level 1b to 2 Off-line Processing

### Algorithm Theoretical Basis Document

Doc.No.: ENV-ATB-SAO-SCI-2200-0003  
Issue: 1/B  
Date: 23. December 1998

*This document was written at the Smithsonian Astrophysical Observatory (SAO) of the Center for Astrophysics (CfA) in Cambridge, MA. It is part of the scientific documentation to be provided by SAO for the contract with DLR-DFD.*

	<i>Name, Function, Affiliation</i>	<i>Signature, Date</i>
prepared	R. Spurr, Project Scientist, SAO	
released	W. Balzer, SCIAMACHY Project Engineer, DLR-DFD	

**Additional Distribution List**

<i>Function</i>	<i>Copies</i>	<i>Name</i>	<i>Affiliation</i>
Principal Investigators	3	J. Burrows A. Goede C. Muller	IFE SRON BISA
Internal Development Team	3	S. Slijkhuis W. Thomas T. Wieland	DLR-DFD
SCIAMACHY Science Advisory Group	per WWW	K. Chance et alii	SAO various

### Document Change Log

<i>Issue</i>	<i>Rev</i>	<i>Date</i>	<i>Sheet</i>	<i>Description of Change</i>
Draft		30.05.98	all	completely new
1		30.07.98	-	numerous minor changes in several places according to comments from IFE, KNMI and RAL after the ADC Review at June 15th, 1998
1	A	27.10.98	-	numerous minor changes in several places according to a request by ESTEC to include and/or highlight the aspects related to the SCIAMACHY NRT processors
1	B	23.12.98	59 ff.	the limb VMR retrieval section was extended from 2 to 14 pages



## Table of Contents

<b>1</b>	<b>Introduction</b>	<b>9</b>
1.1	Purpose and Scope	9
1.2	Abbreviations and Acronyms	11
1.3	Applicable Documents	12
1.4	Document Overview	13
<b>2</b>	<b>SCIAMACHY Algorithms Overview</b>	<b>14</b>
2.1	Status of SCIAMACHY Data Processor	14
2.2	Algorithms Overview	16
2.3	Summary Table	20
<b>3</b>	<b>Nadir Retrieval Algorithms</b>	<b>21</b>
3.1	Cloud and Aerosol Retrieval	21
3.1.1	Introduction to Cloud Algorithms	21
3.1.2	PMD Cloud Coverage Algorithm (PCCA)	21
3.1.3	Cloud Fitting Algorithms (ICFA and RCFA)	24
3.1.4	Introduction to Aerosol Pre-processing Algorithms	28
3.1.5	Aerosol Absorbing Index Algorithm (AAIA)	28
3.1.6	Aerosol Optical Thickness Algorithm (AOTA)	29
3.2	UV-Visible Absorption Spectroscopy (UVAS)	31
3.2.1	Introduction	31
3.2.2	The DOAS algorithm	31
3.2.3	Direct Fitting UV-Visible algorithm	33
3.2.4	Choice of Fitting Windows	34
3.2.5	Reference Spectra	35
3.2.6	Air Mass Factors	36
3.2.7	Vertical Column Density	39
3.2.8	Operational Considerations	40
3.3	IR Absorption Spectroscopy (IAS)	42
3.3.1	Introduction	42
3.3.2	IAS Least-squares Fitting Algorithm	42
3.3.3	CO <sub>2</sub> height retrieval application	44
3.3.4	Trace gas retrieval application	45
3.3.5	Line-by-line cross sections	47
3.3.6	Operational Considerations	48
3.4	O <sub>3</sub> Profile Retrieval	50
3.4.1	Introduction - Optimal Estimation	50
3.4.2	Height-Resolved Fitting Algorithm (HRFA)	51
3.4.3	Operational Considerations	53
<b>4</b>	<b>Limb and Occultation Retrieval Algorithms</b>	<b>54</b>
4.1	Introduction and Retrieval Strategy	54
4.2	Limb P-T Retrieval	57
4.3	Limb Trace Gas VMR Retrieval	59
4.3.1	Introduction	59
4.3.2	Single Scatter Flux Model	60
4.3.3	Baseline Aspects of the UV Single Scatter Model	64
4.3.4	Ray Tracing and Optical Thickness Vectors	67
4.3.5	Multiple scatter corrections in the UV model	70
4.3.6	Preparation for SCIAMACHY Retrieval	72

4.4	Occultation Retrievals . . . . .	73
<b>5</b>	<b>Radiative Transfer Model Requirements . . . . .</b>	<b>75</b>
5.1	GOMETRAN and extension to SCIAMACHY (nadir only). . . . .	75
5.1.1	AMF generation. . . . .	76
5.1.2	Extension to SCIAMACHY (nadir only) . . . . .	77
5.2	MODTRAN and DISORT . . . . .	78
5.3	Limb Scattering Models. . . . .	79
<b>6</b>	<b>Input/Output Requirements. . . . .</b>	<b>80</b>
6.1	Input Data. . . . .	80
6.1.1	Level 1b input . . . . .	80
6.1.2	Initialisation File . . . . .	80
6.1.3	Ancillary input data . . . . .	81
6.2	Product Parameters . . . . .	82
6.2.1	Level 2 Cloud/Aerosol Product . . . . .	82
6.2.2	Level 2 Nadir Column Product . . . . .	82
6.2.3	Level 2 UV/Visible Nadir Height-resolved Product . . . . .	83
6.2.4	Level 2 Limb/Occultation product (suggested). . . . .	83
6.2.5	Detailed output for algorithm verification. . . . .	83
6.3	Reference Data Bases . . . . .	85
6.3.1	Climatology and Spectroscopic Databases . . . . .	85
6.3.2	RT-derived look-up tables. . . . .	85
6.3.3	Orbit pre-processed data sets. . . . .	85
<b>7</b>	<b>Appendices . . . . .</b>	<b>86</b>
7.1	Convolution . . . . .	86
<b>8</b>	<b>References . . . . .</b>	<b>88</b>

## List of Tables

Table 1:	List of proposed geophysical parameters to be retrieved . . . . .	14
Table 2:	Geophysical parameters versus algorithm including anticipated options . . . . .	20
Table 3:	Fitting Windows in the UV and visible wavelength range . . . . .	34
Table 4:	Fitting windows in the IR wavelength range . . . . .	42
Table 5:	Accuracy of CO and N <sub>2</sub> O retrievals for different slit width . . . . .	47



## 1 Introduction

### 1.1 Purpose and Scope

SCIAMACHY (SCanning Imaging Absorption SpectroMeter for Atmospheric CHartographY) is one of the earth observation research instruments to be included as part of the payload of the ESA (European Space Agency) ENVISAT-1 platform to be launched in the year 2000. The main scientific objective of SCIAMACHY is to measure distributions of a number of chemically important atmospheric trace species on a global basis. SCIAMACHY has a spectrometer and telescope system designed to observe light transmitted through and reflected and scattered from the earth's atmosphere over a spectral range of 240 - 2400 nm. It has an alternate limb and nadir viewing capability, and will be able to perform solar and lunar occultation measurements. Nadir UV/visible measurements will provide global column distributions of O<sub>3</sub>, NO<sub>2</sub> and a number of other trace species (BrO, H<sub>2</sub>CO, OClO, SO<sub>2</sub> and possibly ClO), and height-resolved profiles of O<sub>3</sub>.

Nadir infrared measurements will generate column distributions of CO<sub>2</sub>, H<sub>2</sub>O, CH<sub>4</sub>, CO and N<sub>2</sub>O. Limb observations will provide vertical profiles of many of these species, with particular emphasis on O<sub>3</sub> and NO<sub>2</sub> (UV/visible) and H<sub>2</sub>O, CH<sub>4</sub>, CO and N<sub>2</sub>O (infrared). Limb profiles of pressure and temperature will be generated from CO<sub>2</sub> absorption signatures near 2000 nm, and possibly from O<sub>2</sub> A band measurements (~760 nm). Vertical profiles will also be obtained from solar occultation measurements. Aerosol profiles will be available from limb/occultation retrievals. Cloud parameters and additional aerosol parameters will also be produced from a number of pre-processing algorithms.

This Algorithm Theoretical Basis Document (ATBD) describes all algorithms required for the operational retrieval of SCIAMACHY Level 2 data products. Both the Off-line (OL) and Near Real Time (NRT) processing elements are covered. The design, implementation and maintenance of the SCIAMACHY off-line Level 1b to 2 processor are the responsibility of the DLR (German Remote Sensing Data Centre). Operational level 1b to 2 NRT algorithms for SCIAMACHY will be implemented by an industrial consortium, with detailed product and algorithm specifications provided by DLR.

The ATBD will provide physical descriptions of the level 1b to 2 algorithms, together with essential mathematical background. This document also deals with product generation and associated issues appropriate to the implementation of the algorithms in an operational environment. This ATBD will also summarise the necessary input/output requirements for level 1b to 2 data processing, including reference databases.

Issues of this document from the present to the time of launch will be restricted to algorithms implemented for the retrieval of an established list of level 2 products to be generated on a routine basis at the time of launch. In the interim, it is expected that ongoing scientific investigations will result in changes of some of the algorithms; the development of algorithm prototypes will also generate modifications. This is particularly true for the OL algorithms; there is less flexibility with the NRT processor.

Post-launch updates of this ATBD will reflect experience gained from operational processing during and after the commissioning phase. The algorithms will be fine-tuned in response to results and feedback from verification and validation programmes, and further improved on the basis of relevant scientific research. Once the basic algorithms described in the pre-launch issues of the ATBD are functioning, post-launch updates will probably contain descriptions of synergies between nadir and limb SCIAMACHY products.

R. J. D. Spurr (Harvard-Smithsonian Center for Astrophysics) prepared the original (draft) version of this document in May 1998. This first draft was written as part of the document package produced for the SCIAMACHY Data Processing ADC Review Meeting (June 15, 1998). A number of comments received after this meeting were incorporated in the first issue (July 1998). The first issue was confined to off-line developments.

In September 1998, ESA requested a new ATBD for the level 1b to 2 NRT algorithms. Since the group of NRT algorithms is for the most part a subset of that for the OL, it is expedient to include NRT descriptions as part of a single ATBD for SGP level 1b to 2. This has been agreed, and the second issue was prepared for release in mid-December 1998, in time for a major review of selected ENVISAT ATBDs at the ESAMS conference (18-22 January 1999). In addition to the new material on NRT applications, the section on limb retrieval has been greatly expanded.

The author would like to thank W. Balzer, S. Slijkhuis and D. Loyola and colleagues at DLR for comments, and K. Chance and T. Kurosu at SAO for valuable feedback.

## 1.2 Abbreviations and Acronyms

AAIA	Absorbing Aerosol Index Algorithm
AMF	Air Mass Factor
AO	Announcement of Opportunity
AOTA	Aerosol Optical Thickness Algorithm
ATBD	Algorithm Theoretical Basis Document
ATSR	Along-Track Scanning Radiometer
AVHRR	Advanced Very High Resolution Radiometer
BIAS	Basic Infrared Absorption Spectroscopy
BUV	Backscattered ultra-violet
CRAG	Cloud Retrieval Algorithm for GOME
DFD	Deutsches Fernerkundungsdatenzentrum
DLR	Deutsches Zentrum für Luft- und Raumfahrt e.V.
DOAS	Differential Optical Absorption Spectroscopy
ECMWF	European Centre for Medium-range Weather Forecasting
ENVISAT	Environmental Satellite
ESA	European Space Agency
FD	Fast Delivery
FTS	Fourier Transform Spectrometer
GDP	GOME Data Processor
GOME	Global Ozone Monitoring Experiment
HALOE	HALogen Occultation Experiment
HRFA	Height-Resolved Fitting Algorithm
IAS	Infrared Absorption Spectroscopy
ICFA	Initial Cloud Fitting Algorithm
IR	infra-red
LBL	Line-by-line
LORE	Limb Ozone Retrieval Experiment
NRT	Near Real Time
NWP	Numerical Weather Prediction
OE	Optimal Estimation
ODIN	
OSIRIS	
PMD	Polarisation Measurement Device
SAGE	Stratospheric Aerosol and Gas Experiment
SAO	Smithsonian Astrophysical Observatory
SOLSE	Shuttle Ozone Limb Sounding Experiment
SSAG	SCIAMACHY Science Advisory Group
P-T	Pressure and Temperature
PCCA	PMD Cloud Coverage Algorithm
RCFA	Revised Cloud Fitting Algorithm
SAGE	Stratospheric Aerosol and Gas Experiment
SCIAMACHY	Scanning Imaging Absorption Spectrometer for Atmospheric Chartography
SGP_12OL	SCIAMACHY Ground Processor level 1 to 2 Off-Line
TOA	Top of Atmosphere
UV	ultra-violet
UVAS	UV/Visible Absorption Spectroscopy
VMR	Volume Mixing Ratio

### 1.3 Applicable Documents

We give here the list of applicable SCIAMACHY project documents (both general and specific). Other references arising in the main body of the ATBD will be listed separately in Chapter 8 (this includes relevant GOME project documentation).

- [A1] Burrows, J. P., K. Chance, P. Crutzen, H. van Dop, J. Geary, T. Johnson, G. Harris, I Isaksen, G. Moortgat, C. Muller, D. Perner, U. Platt, J. -P. Pommereau, H. Rodhe, E. Roeckner, W. Schneider, P. Simon, H. Sundquist, and J. Vercheval, SCIAMACHY A European proposal for atmospheric remote sensing from the ESA polar platform, Max-Planck Institut fuer Chemie, Mainz, Germany, 1988
- [A2] Burrows, J. P., K. Chance, P. Crutzen, J. Fishman, J. Fredericks, J. Geary, T. Johnson, G. Harris, I. Isaksen, H. Kelder, G. Moortgat, C. Muller, D. Perner, U. Platt, J. -P. Pommereau, H. Rodhe, E. Roeckner, W. Schneider, P. Simon, H. Sundquist, and J. Vercheval, SCIAMACHY Phase A Study Scientific Requirements Specification, 1991
- [A3] SCIAMACHY Requirements Document, SCIAMACHY Scientific Advisory Group, Issue 1A, in preparation, May 1998
- [A4] Scientific Requirements Document for Data and Algorithm Development, SCIAMACHY Algorithm Development and Data Usage Subgroup, Issue 2A, January 1998
- [A5] SCIAMACHY Instrument Requirements Document, PO-RS-DAR-EP-0001, Issue 3/1, December 1995
- [A6] SCIAMACHY Operations Concept: III. Instrument States, PO-TN-DLR-SH-0001/3, Issue2/0, July 1996
- [A7] ENVISAT Mission Conventions Document, PO-IS-ESA-GS-0561, Issue2/0, April 1997
- [A8] SCIAMACHY Level 1b to 2 NRT Processing, Input/Output Data Definition, ENV-TN-DLR-SCIA-0010, Issue 3/A, 16. July 1998
- [A9] SCIAMACHY Level 1b to 2 NRT Processing, Detailed Processing Modules / Parameter Data List, ENV-TN-DLR-SCIA-0011, Issue 1, 7. July 1998
- [A10] SCIAMACHY Level 1b to 2 Off-line Processing, Product Specification Document, ENV-PS-DLR-SCI-2200-0004, Issue 1/A, June 1998
- [A11] SCIAMACHY Level 1b to 2 Off-line Processing Interface Specification Document, ENV-IS-DLR-SCI-2200-0005, Draft, June 1998

## 1.4 Document Overview

In chapter 2, we give an overview of the algorithms required for level 1b to 2 SCIAMACHY processing in the context of the mission objectives and the list of level 2 data products. This chapter also includes a summary table; this represents the baseline on algorithms/products at the time of writing.

Chapter 3 is concerned with SCIAMACHY nadir-mode retrieval algorithms. There are two central sections on the UV/visible and infrared trace gas column retrievals, and a section on cloud and aerosol pre-processing algorithms. An additional topic deals with the optional O<sub>3</sub> profile retrieval. Apart from the latter, all algorithms in this chapter have NRT applications, and these are covered in dedicated sub-sections.

Chapter 4 (limb/occultation algorithms) deals with limb pressure/temperature retrieval, limb trace gas volume mixing ratio retrieval, and occultation profile retrievals. These algorithms are off-line only.

Requirements on radiative transfer models are discussed in Chapter 5, with applications to both NRT and OL algorithms. Chapter 6 deals with input/output and database requirements. Chapter 7 contains appendices on auxiliary mathematical operations, and is followed by the list of references.

Each algorithm exposition includes a physical description, along with the most important and relevant mathematical formulae. We discuss aspects of the retrieval strategy, and the evaluation of targeted products and their diagnostics. In addition, there are some remarks on the operational implementation of the prototype algorithms (performance aspects, testing and verification, quality control). Aspects of the forward problem (simulation of intensities and weighting functions) are noted as appropriate, with further detailed discussion of forward model simulation confined to Chapter 5.

## 2 SCIAMACHY Algorithms Overview

### 2.1 Status of SCIAMACHY Data Processor

#### Mission Objectives and Level 2 Products

In the AO proposal [A1], the following gases were targeted for measurement: O<sub>2</sub>, O<sub>3</sub>, O<sub>4</sub>, NO, NO<sub>2</sub>, NO<sub>3</sub>, CO, CO<sub>2</sub>, H<sub>2</sub>CO, CH<sub>4</sub>, H<sub>2</sub>O, SO<sub>2</sub> and possibly ClO and OCIO under ozone hole conditions. As a result of the initial sensitivity analysis, N<sub>2</sub>O, BrO and O<sub>2</sub> (<sup>1</sup>Δg) were then included. This list constitutes the set of primary scientific mission objectives. Secondary mission objectives include pressure and temperature profile generation, the determination of selected cloud and aerosol properties, and the derivation of tropospheric distributions from combined limb/nadir results. Detailed discussion of scientific objectives for SCIAMACHY may be found in the Phase A Study Report [A2], and the Scientific Requirements Document [A3] (currently in draft version). Specific scientific requirements on algorithms to retrieve trace gas constituents and other geophysical parameters have been laid down in the Report of the SCIAMACHY Algorithm Development and Data Usage Subcommittee [A4].

On the basis of this report and subsequent discussions at the Advisory Group Meetings, a table of SCIAMACHY Data Products was generated by SSAG; this is given below in Table 1, with the status as of February 1997.

	<i>Nadir Total Column Amount</i>			<i>Limb Stratospheric Profiles</i>		
	<i>UV-Vis</i>	<i>IR</i>	<i>UV-IR</i>	<i>UV-Vis</i>	<i>IR</i>	<i>UV-IR</i>
<i>NRT</i>	O <sub>3</sub>	H <sub>2</sub> O	Clouds			
	NO <sub>2</sub>	N <sub>2</sub> O	Aerosol			
	BrO	CO				
	SO <sub>2</sub>	CH <sub>4</sub>				
	OCIO					
	H <sub>2</sub> CO					
<i>OL</i>	O <sub>3</sub>	H <sub>2</sub> O	Clouds	O <sub>3</sub>	H <sub>2</sub> O	Aerosol
	NO <sub>2</sub>	N <sub>2</sub> O	Aerosol	NO <sub>2</sub>	N <sub>2</sub> O	
	BrO	CO		BrO	CO	
	SO <sub>2</sub>	CO <sub>2</sub>			CO <sub>2</sub>	
	OCIO	CH <sub>4</sub>			CH <sub>4</sub>	
	H <sub>2</sub> CO				p, T	
	UV Index					

**Table 1: List of proposed geophysical parameters to be retrieved**

The basic division here is between nadir-mode total column retrieval products and limb stratospheric profile retrieval products on the one hand, and between off-line (OL) and the fast delivery (NRT) products on the other. The OL list is the baseline that provides the Priority I products for off-line level 1b to 2 processing (see table in section 2.3). All NRT products have Priority I status.

The above table does not include occultation or nadir profile retrieval products; these have Priority II status at present; they include O<sub>3</sub> profiles (nadir) and the set of occultation stratospheric profiles (very similar to the list for limb in the above table).

With the exception of the UV Index, the present ATBD will describe all algorithms for Priority I and II off-line and near real time products as listed in the above table. The discussion of nadir/limb value-added tropospheric products is beyond the scope of the ATBD (see [A3] for a summary).

Since SCIAMACHY channels 1 to 4 are similar in scope to those for GOME, GDP algorithms have been taken over for use in the UV/visible nadir-mode column retrievals for SCIAMACHY (see below). The first major new development for SCIAMACHY was the specification and generation of an operational prototype algorithm for the NRT processor, for the total column retrieval of trace species (CO, N<sub>2</sub>O and CH<sub>4</sub>) from SCIAMACHY channels 7 and 8 infrared nadir measurements. This algorithm (chapter 3.3) will be taken over for the OL processor, and given greater flexibility and power without the constraint of NRT performance limitations. For the off-line products, limb retrieval algorithms must be developed from scratch; the first operational prototypes are now under development, and will be described in detail in this issue.

### **Relation to GOME Data Processor**

The Global Ozone Monitoring Experiment (GOME) was originally conceived as a scaled-down version of SCIAMACHY. It was given fast-track development status by ESA [G1], and was launched on 21 April 1995 on board the second European Remote Sensing Satellite (ERS-2). It has 4 spectral channels covering the range 240-790 nm, and is a nadir-only instrument. The measurement capability of GOME closely matches the UV/visible nadir capability of SCIAMACHY; the mission objectives are very similar.

The GOME Data Processor (GDP) was developed and implemented at DFD with the help of several scientific institutions [G2]. GDP became operational in July 1996, with earthshine spectra and retrieved total O<sub>3</sub> columns the main products generated on a routine basis. Total column amounts of NO<sub>2</sub>, H<sub>2</sub>CO, SO<sub>2</sub>, OCIO and BrO have been retrieved successfully from GOME back-scatter measurements. The experience gained with GOME in the implementation of operational nadir UV-visible total column retrieval will be invaluable, as there is considerable overlap with SCIAMACHY. Parts of the current ATBD will follow closely the descriptions in the GDP Technical Documentation [G3][G4].

The main algorithm in GDP is the DOAS (Differential Optical Absorption Spectroscopy) fitting for the retrieval of O<sub>3</sub> and NO<sub>2</sub> columns; this can be used directly for SCIAMACHY OL and NRT applications. It is expected that an updated form of the current cloud pre-processing algorithm in GDP will be adapted for SCIAMACHY OL, to include a new PMD cloud-clearing algorithm for cloud fractional cover (the cloud treatment in NRT will be confined to this latter development).

Substantial progress has been made in GOME studies with height-resolved O<sub>3</sub> profile retrieval algorithms, and developments in this field for SCIAMACHY will run in tandem with those for GOME. Study work done for GOME under the aegis of ESA is of relevance to several aspects of SCIAMACHY Data Processing (Cloud and aerosol studies, DOAS retrieval studies, Ring scattering study). The documents and reports from this work are noted under the section GOME Project Documentation and Study Reports in Chapter 8 (references [G5] to [G11]).

## 2.2 Algorithms Overview

### Nadir Algorithms Overview

In this and the following section, summaries of algorithms are given without references (these will be given in the main text). The summaries given include all the OL algorithms; limitations and special considerations for the NRT processor are noted in italics.

#### *Cloud/aerosol pre-processing algorithms*

These algorithms will deliver cloud information (fractional cover, cloud optical properties, cloud classification) and initial aerosol information. The term 'pre-processing' indicates that these algorithms will be the first to be executed in SGP level 1b to 2 processing chain. In particular, the cloud results will be used to correct subsequent UV/visible trace species retrievals for nadir scenes flagged as partially or totally cloudy. There are two main component cloud algorithms; they are listed separately here, but may well be used together in the (off-line) operational processor.

- PMD Cloud Coverage Algorithm (PCCA). There are two approaches here, both using sub-pixel PMD (Polarization Measurement Devices) measurements. One method is based on dynamic thresholding, the other on cloud-free composites. Prototypes have recently been developed for GOME (3 PMDs), and these will be extended to SCIAMACHY (7 PMDs). *At present an older version of PCCA is specified for the NRT, but it is expected that the upgrade planned for OL will also be taken on for the NRT.*
- Revised Cloud Fitting Algorithm (RCFA). An algorithm for the retrieval of cloud optical parameters based on least squares fitting of measured nadir back-scatter intensities with their simulated equivalents in and around the O<sub>2</sub> and O<sub>4</sub> absorption bands in the visible/near-infrared. Main retrieval products will be cloud-top pressure and optical depth; combined with the cloud-cover algorithm, a determination of cloud type will be possible. Development of this algorithm will again follow that for GOME. *Not part of the NRT processor.*
- Aerosol Absorbing Index Algorithm (AAIA) is stand-alone at present, and the results will not be used in later algorithms. *Both NRT and OL.*
- Aerosol Optical Thickness Algorithm (AOTA). Also intended to be stand-alone, this is at present an option to be considered for the OL processor. (Aerosol profiles retrieved from limb and occultation measurements are treated separately). *Not part of the NRT processor.*

#### *UV-Visible Absorption Spectroscopy (UVAS) Algorithm*

*(Applies equally to NRT and OL).*

This algorithm is the one that most closely matches the corresponding DOAS algorithm in GDP. It is based on the least squares fitting of effective slant column amounts of certain trace species in the UV and visible parts of the spectrum. The DOAS method will be one of the major options: optical densities are fitted linearly for column amounts, with accompanying non-linear fitting for spectra wavelength mismatches. The baseline will be two windows optimised for O<sub>3</sub> and NO<sub>2</sub> retrieval. Another option to be considered will be the direct non-linear least squares fitting of radiance to determine slant columns. Recent developments with GOME have demonstrated its ability to measure BrO, H<sub>2</sub>CO, OCIO, SO<sub>2</sub> and even ClO, and the SCIAMACHY OL and NRT algorithms will include fitting windows for these species. Columns of O<sub>3</sub>, NO<sub>2</sub>, BrO, SO<sub>2</sub> and H<sub>2</sub>CO will be retrieved on a global basis, with OCIO for special cases only (ozone hole scenarios).

In the first operational prototype, the conversion to vertical column amounts will be done using a single Air Mass Factor for each trace species and fitting window (the 'AMF-modified' DOAS approach will not be described in this ATBD). Experience with GOME has shown that it is preferable to use look-up tables of AMFs in an operational environment, rather than calculate them from scratch using radiative transfer code. The development for GOME of improved O<sub>3</sub> and NO<sub>2</sub> profile climatology for the generation of new AMF look-up tables for these species will be incorporated as part of the SCIAMACHY algorithms. The correction of vertical column density for cloud-contaminated scenes currently in use for GOME will be taken over as the baseline for SCIAMACHY.

### ***The IR Absorption Spectroscopy (IAS) Algorithm***

*(Applies equally to NRT and OL, with additional performance considerations for NRT).*

This algorithm is based on the direct (non-linear) least squares fitting of effective vertical column amounts of trace species in the infrared part of the back-scatter spectrum (SCIAMACHY channels 7 and 8). In its simplest form, the algorithm assumes scattering can be approximated by a multiplicative closure term on the transmittance, provided the optical depth is not too large. A small selection of micro-windows will highlight retrieval of the important species (CO and N<sub>2</sub>O are the drivers, with CH<sub>4</sub>, H<sub>2</sub>O and CO<sub>2</sub> columns also retrieved).

The same technique can be used to derive the pressure height of an equivalent Lambertian reflecting surface assuming the known absorption of CO<sub>2</sub> near 2030 nm in channel 7 (relatively low optical depth). It is more consistent to execute this CO<sub>2</sub> height fit first before attempting trace gas retrievals; the height result is used as the lower boundary in the subsequent retrieval of trace species from Channel 8 infrared measurements. It is anticipated that the infrared absorption spectroscopy results will be stand-alone (no incorporation of results from the cloud pre-processing algorithms).

The OL version will include a full line-by-line capability for cross section evaluation, and a ray-tracing formalism in a curved refracting atmosphere. *The NRT algorithm will use look-up tables of pre-computed trace gas cross-section templates at a high resolution, and an additional table of ray-traced slant path factors. In the specification documents for NRT, the algorithm is named BIAS (Basic Infrared Absorption Spectroscopy).*

### ***The height-resolved fitting algorithm (HRFA) for O<sub>3</sub> profiles***

*(Not in the NRT processor).*

This is based in part on the UV technique, where the wavelength dependence of O<sub>3</sub> scattering heights in the UV below 300 nm is used to infer stratospheric O<sub>3</sub> profiles. The algorithm also takes advantage of the temperature dependence of the Huggins bands O<sub>3</sub> absorption (300 - 360 nm) to extract height-resolved information in the troposphere; inclusion of points in the visible (Chappuis) O<sub>3</sub> absorption region may also provide an additional constraint. The algorithm uses an iterative optimal estimation technique to update the state vector of profile components; an *a priori* profile is used to constrain the problem.

The forward part of the algorithm requires repeated calculation of intensities and their parameter derivatives (weighting functions). Radiative transfer simulations must be performed with full multiple scatter, and this is (at present) prohibitively time-consuming for realistic inclusion in an operational environment. However, results derived from GOME back-scatter measurements for special scenarios and campaigns have shown clearly the feasibility of this technique, and it is

anticipated that an operational algorithm will be implemented once the performance bottleneck is overcome (perhaps with the help of extensively parameterised look-up tables).

## **Limb and Occultation Algorithms Overview**

*(Not part of the NRT processor).*

At the time of writing, there are still unresolved questions about the overall retrieval strategy for limb profiles. In order to keep options open, both onion peeling and global fit methods will be considered (and perhaps an admixture of the two). There is also a distinction between P-T retrieval and trace gas volume mixing ratio (or concentration) vertical profiles. In both cases, aerosol scattering profiles will be retrieved simultaneously. The use of optimal estimation or unconstrained non-linear least squares has also not been decided for the various limb and occultation applications.

P-T retrieval should be performed first, to allow the retrieved temperature and/or pressure profiles to be used in subsequent trace species limb retrievals. The P-T and VMR retrieval algorithms are regarded as given a stand-alone treatment in this issue of the ATBD, and it is intended to deal with the integration of these limb applications in later versions. Note that the limb and occultation products will contain both the VMR and the number density values; the algorithm is labelled ‘-VMR’ for convenience in this and following sections.

### *The P-T profile limb algorithm (Limb-PT)*

This algorithm will use non-linear least squares fitting to derive profiles of temperature and pressure from limb scatter measurements of CO<sub>2</sub> in channel 7. Fitting windows for each limb scan must be selected to include suitable combinations of pressure and temperature sensitive CO<sub>2</sub> absorption lines. The atmosphere is assumed to be in hydrostatic equilibrium; exact knowledge of the pressure and temperature at one height will enable the tangent heights to be retrieved by this method.

Simulation of limb scatter intensities and associated weighting functions will be carried out quasi-analytically in the single-scatter line-by-line approximation; multiple scattering effects (small in this part of the spectrum) will be parameterised. Aerosol scattering and extinction profiles will be retrieved in conjunction with the atmospheric variables. Pressure and temperature profiles may also be retrieved from O<sub>2</sub> A band limb back-scatter measurements around 760 nm, but this is still under investigation and will not be described here. The question of onion peeling versus global fitting has not been resolved yet for this algorithm.

### *The VMR profile limb algorithm (Limb-VMR)*

This algorithm retrieves volume mixing ratios (VMRs) of selected trace species for layers defined by the limb scan sequence (equivalently, one can retrieve concentrations, but for simplicity we stay with VMR for the remainder of the document). For SCIAMACHY channels 1 to 4, the main emphasis is on measurements from the UV (O<sub>3</sub>) and visible parts of the spectrum (NO<sub>2</sub>, O<sub>3</sub>). From two or more micro-windows in channel 8, this algorithm can also retrieve VMRs for trace species N<sub>2</sub>O, CO and CH<sub>4</sub> (and possibly H<sub>2</sub>O); CO<sub>2</sub> limb profiles can be retrieved from channel 7 measurements. In all cases, aerosol scattering and extinction profiles will be retrieved in conjunction with the main output. The algorithm will retain options on the use of optimal estimation versus non-linear least squares.

It is likely that the first version of the prototype will use the global fit technique, especially for the UV/visible applications. Also it is probable that ‘measurements’ as defined in this algorithm will

actually be ratios of limb scan radiances at two different tangent heights. Again, single scatter simulations will be quasi-analytic in the retrieval parameters, with corrections for multiple scattering effects extracted from look-up tables. All targeted species' cross-sections are temperature-dependent. In addition it is advisable to use a retrieval heightgrid based on the foregoing P-T limb results.

### *Occultation algorithms*

In principle, the list of products defined for the first limb profile retrievals in SGP\_12OL can also be retrieved for solar occultation measurements. There is scope for only a few occultation products per orbit due to the limited time (~ 60 s) spent in this viewing mode. The forward simulation part of the retrieval algorithms is simpler, because it requires only computation of direct-beam transmittances (scattering can be neglected). Though measurement signal-to-noise and corresponding retrieval precision are high, there are radiance calibration (and possibly also wavelength calibration) problems due to the measurement strategy employed by SCIAMACHY.

### **Input/Output and Reference Database Overview**

The main input to this part of the whole SCIAMACHY Data Processor is the level 1b product, comprising measurements for nadir, limb and occultation states, and an extra-terrestrial sun reference spectrum taken by the instrument. Each atmospheric measurement contains a number of observation records (geolocation information and spectra). Also required as input is a list of parameters controlling the execution of the component algorithms in level 1b to 2 processing; this comprises the 'Initialisation File'. Additional real-time ancillary data may also be required (assimilated meteorological forecast/analysis fields, for example).

Main output will be level 2 products, as specified in [A10]. Each component algorithm as summarised above will generate its own component output data set, in addition to the housekeeping and geolocation data required for completeness. In addition to retrieved parameter values, solution covariance output from the fitting algorithms will be specified where possible, though no detailed spectral information will be included in the products. During the testing and commissioning phases, detailed results and scenarios will also be generated as part of the output.

Reference data may be divided into three classes. The first class contains fixed global and climatological data sets (topography, surface properties, atmospheric profiles including *a priori* values) and reference spectra (spectroscopic line parameters, absorption and Ring effect cross sections, aerosol and cloud optical properties). The second class comprises fixed pre-calculated look-up tables of radiative transfer simulated data (AMFs for the UV/visible column retrievals, multiple scatter correction factors for limb retrievals, line-by-line transmittances). The third class contains data sets that are pre-calculated for each orbit (in particular, the preparation of instrument response functions required for the convolution of high-resolution templates to instrument wavelength grids).

### 2.3 Summary Table

So far as possible, this table complies with table 1 in section 2.1 on page 14 above. Priority I has been subdivided. An asterisk denotes NRT product. For acronyms, see section 1.2.

<i>Parameter</i>	<i>Priority</i>	<i>Algorithm</i>	<i>Remarks and Baseline</i>
Cloud cover	I-a *	PCCA	PCCA also in level 0 to 1b processing
Cloud top pressure	I-a	RCFA	Used perhaps in conjunction with PCCA
Cloud optical depth	I-a	RCFA	Used perhaps in conjunction with PCCA
Cloud type	I-b	RCFA/PCCA	Status uncertain
Aerosol index	I-b *	AAIA	Absorbing aerosol index
O <sub>3</sub> column	I-a *	UVAS	AMF lookup-table
NO <sub>2</sub> column	I-a *	UVAS	AMF lookup-table
SO <sub>2</sub> column	I-a *	UVAS	Special scenario, may be globally
BrO column	I-a *	UVAS	Global coverage
H <sub>2</sub> CO column	I-a *	UVAS	Special scenario
OCIO column	I-a *	UVAS	Special scenario
CO column	I-a *	IAS	Channel 8 infra-red
N <sub>2</sub> O column	I-a *	IAS	Channel 8 infra-red
CH <sub>4</sub> column	I-a *	IAS	Channel 8 infra-red
H <sub>2</sub> O column	I-b *	IAS	Channel 7, also channel 5 possibility
Reflecting height	I-a *	IAS	Channel 7, CO <sub>2</sub> infra-red
O <sub>3</sub> profile	II	HRFA	Nadir, developed with GOME, UV/visible
P-T profile	I-a	Limb-PT	Global fit preferred, CO <sub>2</sub> infra-red
O <sub>3</sub> profile	I-a	Limb-VMR	Global fit preferred, UV/visible
NO <sub>2</sub> profile	I-a	Limb-VMR	Global fit preferred, visible
BrO profile	II	Limb-VMR	Status TBD
CO profile	I-b	Limb-VMR	Global fit or onion peeling, infra-red
N <sub>2</sub> O profile	I-b	Limb-VMR	Global fit or onion peeling, infra-red
H <sub>2</sub> O profile	I-b	Limb-VMR	Global fit or onion peeling, infra-red
CH <sub>4</sub> profile	I-b	Limb-VMR	Global fit or onion peeling, infra-red
Aerosol profile	II	Limb-VMR	Global fit or onion peeling, infra-red
P-T profile	II	Occul-PT	Global fit or onion peeling, CO <sub>2</sub> , infra-red
O <sub>3</sub> profile	II	Occul-VMR	Global fit or onion peeling, UV/visible
NO <sub>2</sub> profile	II	Occul-VMR	Global fit or onion peeling, visible
Minor profiles	II	Occul-VMR	UV/visible: BrO, SO <sub>2</sub> , H <sub>2</sub> CO, OCIO
Minor profile	II	Occul-VMR	Infra-red: CO, CH <sub>4</sub> , N <sub>2</sub> O
Aerosol profile	II	Occul-VMR	Status TBD

**Table 2: Geophysical parameters versus algorithm including anticipated options**

### 3 Nadir Retrieval Algorithms

#### 3.1 Cloud and Aerosol Retrieval

##### 3.1.1 Introduction to Cloud Algorithms

First, we outline the historical development of cloud parameter retrieval for GOME. To date there have been no specific developments in this field for SCIAMACHY, but the commonality of the two instruments ensures that all the GOME work done so far is relevant. There have been two approaches to cloud retrieval - the use of sub-pixel PMD data to derive fractional cloud cover, and the use of a least-squares fitting algorithm comparing reflectances in and around the O<sub>2</sub> A band for fractional cover and more recently, cloud-top pressure and optical depth.

The first algorithm using PMD data was developed in 1993 as part of a scattering study sponsored by ESA [G8]. Using a combination of reflectance and colour (ratio) thresholds, sub-pixel scenes were assigned 'clear', 'cloudy', or 'undetermined' status. The algorithm used static thresholds and simulated data, and distinguished between land and ocean surfaces. This algorithm was adapted for use in GOME but never actually used in the operational context. *It has also been specified for SCIAMACHY NRT as the cloud clearing algorithm, but recent developments (next paragraph) have superseded it, and it is anticipated that an upgraded cloud clearing algorithm will be incorporated in NRT.*

A more sophisticated algorithm has been constructed as part of the CADAPA project for GOME [G10]. This is designed to work with real GOME PMD data; it contains sets of maximum and minimum thresholds for the PMD signals compiled from two years of GOME data. The algorithm returns fractional coverage; it also possesses a dynamic threshold update capability. Another clearing algorithm for cloud coverage has recently been developed for the GOME PMD data [C2]. The key component here is a global cloud-free composite, again based on GOME historical data. These newer algorithms will form part of a new operational cloud retrieval package (CRAG) to be installed in GDP in 1999 [G11].

An O<sub>2</sub> A band algorithm (ICFA - Initial Cloud Fitting Algorithm) was developed originally for GOME to determine cloud fraction [C1]. In its current operational version for GOME, ICFA uses linear least-squares fitting, matching a combination of transmittances to ground and cloud-top in and around the O<sub>2</sub> A band. The cloud-top height is taken from climatology, the optical depth is assumed constant, and cloud-top reflectances are pre-computed from forward model computations. ICFA is part of GDP and has received some validation [C3]. More recent work [G9] has focused on non-linear least squares fitting of the O<sub>2</sub> A band reflectance values, with optical depth and cloud-top height the free parameters. This work will eventually form the basis of RCFA (Revised Cloud Fitting Algorithm). RCFA will also form part of CRAG.

In the next two sections we review these GOME developments in more detail, up to the present state of knowledge. Extensions to SCIAMACHY will be indicated where appropriate, and we review some developments at the end.

##### 3.1.2 PMD Cloud Coverage Algorithm (PCCA)

*(To be considered as part of NRT).*

For now we consider only the three broad band PMD devices covering GOME channels 2, 3 and 4 (approximate ranges 300 - 400 nm, 400 - 600 nm, and 600 - 800 nm); SCIAMACHY has comparable devices with similar ranges. For each diode array short integration read-out (1.5 seconds),

the PMDs are read out 16 times (integration time 98 ms). For the regular 960 km GOME scanning swath, this translates to 3 x 16 sub-pixel footprints, each of size 40 x 20 km<sup>2</sup> (sizes will be comparable for SCIAMACHY). For each nadir scan, the GOME level 1 product contains 3 x 16 PMD reflectances (sun-normalised intensities); these are converted to nadir values (albedos) through division by  $\mu_0$  (solar zenith angle cosine), division by  $\mu_1$  (line-of-sight zenith angle cosine) and multiplication by  $\pi$ .

### *Thresholding technique*

Computation of PMD sub-pixel cloud cover is performed as follows. Denote the PMD albedo as  $P^{(1)}$ ,  $P^{(2)}$  and  $P^{(3)}$  and suppose that the minimum value (clear sky) thresholds for the PMDs are  $P_{min}^{(j)}$ ,  $j = (1, 2, 3)$ , and the maximum value (cloudy) thresholds  $P_{max}^{(j)}$ . Suppose also that margins  $\delta_{min}^{(j)}$  and  $\delta_{max}^{(j)}$  for these thresholds have been defined. An actual PMD albedo  $P^{(j)}$  will then pass the clear or cloudy threshold test if  $P^{(j)} < P_{min}^{(j)} + \delta_{min}^{(j)}$  or  $P^{(j)} > P_{max}^{(j)} - \delta_{max}^{(j)}$  respectively. These are absolute tests; another test involves the colour ratio  $R = P^{(3)}/P^{(2)}$  of the ‘red’ and ‘green’ PMD albedos. Thresholds  $\{R_{min}, R_{max}\}$  and margins  $\{\delta_{min}^{(R)}, \delta_{max}^{(R)}\}$  are defined in a similar fashion. When the four clear and cloudy threshold tests have been completed with no determination of a totally clear or completely cloudy sub-pixel footprint, then a fractional cloud cover is assigned as follows. When we encounter the first of the  $P^{(j)}$  to lie between their threshold values  $P_{min}^{(j)}$  and  $P_{max}^{(j)}$  (the order of the tests is discussed below), then we assign fraction cloud cover (in %) from:

$$C = \frac{|P - (P_{min} + \delta_{min})|}{(P_{max} - \delta_{max}) - (P_{min} + \delta_{min})} \times 100 \quad (1)$$

If this assignment fails, then we look to an adjacent pixel, and if the latter has been classified we assign a partial cloud fractional percentage by interpolation; failing this, the pixel is genuinely indeterminate. The absolute tests are always the same for each PMD albedo, but the order in which they are performed depends on the surface type. In addition, the colour threshold test operates differently according to ocean or land surface types. For cloudy conditions, the colour ratio is close to 1, because the scattered light from the cloud has a flat signature. For cloud-free conditions, the colour ratio depends on the underlying surface over oceans, green is reflected more than red so  $R < 1$ , while over land the reverse is true ( $R > 1$ ). For a land surface, the algorithm performs first the cloud-free testing using minimum absolute thresholds  $P_{min}^{(j)}$  and the maximum colour threshold  $R_{max}$ , in the order  $P^{(2)}$ ,  $R$ ,  $P^{(3)}$ ,  $P^{(1)}$ . It then tests for cloudy sub-pixels using maximum absolute thresholds and a minimum colour threshold; the test order is the same. The partial assignment test is then applied if required. For ocean surfaces, the threshold test structure is the same, except that the cloud-free colour test uses  $R_{min}$  while the cloudy colour test uses  $R_{max}$ , and the order is  $P^{(3)}$ ,  $R$ ,  $P^{(2)}$ ,  $P^{(1)}$ . Flow diagrams of all tests can be found in [G10]. The algorithm is designed to update thresholds when sub-pixel PMD assignments have been classified as cloudy or cloud-free (this is the dynamic threshold update). Thresholds are assigned to a 1° x 1° or 1/2° x 1/2° latitude/longitude grid. A representative set of monthly minimum and maximum

values was created by running the algorithm in dynamic update mode through two years of GOME data (January 1996 to January 1998). This set will be updated dynamically as the GOME mission continues.

The classification of surface type follows the scheme adopted for GDP, which is based on earlier work from NASA [C4]. In the GOME PMD algorithm, no distinction is made between 'snow' and 'water' surface types; all land surface types are treated equally. It should be possible with SCIAMACHY to distinguish easily between ice and liquid water. Since SCIAMACHY has additional PMD detectors out in the infrared, an additional threshold test could be added for this distinction by comparing the PMD output from the channel 6 PMD (range ~ 1000-1700 nm) to that of the channel 5 PMD (range ~ 800-1000 nm). For the GOME algorithm, a set of static thresholds has been generated for test simulations using spectral albedos for five different surface types (snow, water, sand, soil and vegetation; data compiled by R. Guzzi (private communication), see also [C5] and [G4]).

Geolocation information is important in cloud detection using this algorithm. The sub-pixel latitude and longitude assignments may be found by interpolating the available geolocation information in the level 1 product to the centre of the footprint. The solar zenith angle should really be specified at the top of the atmosphere (TOA) - this value has not so far been part of the level 1 geolocation for GOME. A patch was written to provide this information so that the algorithm could run with level 1 input alone. An alternative solution uses the orbit propagator supplied by ESA as part of level 0 to 1 processing to return exact geolocation values; this has now been grafted on to the algorithm (currently undergoing tests). Initial investigations as part of the CRAG project indicate that the PMD albedos are not very sensitive to TOA height (T. Kurosu, private communication); a baseline TOA of 60 km has been proposed.

For 'ocean' surfaces, a simple sun-glint condition has been incorporated. If the viewing geometry indicates forward scatter, and if the difference between the solar and line-of-sight zenith angles is less than 0.5, then threshold testing is not done. However, post-processing using extrapolation or interpolation from adjacent successfully classified sub-pixels can often overcome the indeterminacy resulting from a sun-glint flag.

Qualitative validation of the thresholding algorithm has been performed with a selection of coincident AVHRR images [G10]. It is intended to make a more comprehensive validation of the algorithm using the Apollo cloud coverage algorithm [C6] on coincident ATSR images (a sister instrument to GOME on ERS-2, with common data stream).

The algorithm as it stands can be directly adapted for use in SCIAMACHY. The only differences are the number of sub-pixel PMD measurements (10 in SCIAMACHY) and the read-out times. The same applies to the cloud-free composite clearing algorithm described next.

#### *Cloud-free composite technique*

Another approach to cloud cover determination has been demonstrated using GOME PMD data; this work is also part of the CRAG study for GOME [G9]. It is a generic cloud recognition algorithm that works for optical remote sensors. It relies on the generation of an off-line global cloud-free composite from historical data [C2], which is then compared linearly with actual measurements to determine a fractional cover.

A simplified model of the earth-atmosphere reflectance is used for the image formation. The PMD reflectance  $\rho(x, y, \lambda)$  at location  $(x, y)$  and wavelength  $\lambda$  is defined as the ratio of the PMD radiance value  $I(x, y, \lambda)$  to the solar irradiance  $I_0(\lambda)$ , divided by the cosine of the incident

solar zenith angle, and the cosine of the viewing zenith angle. The use of this geometry-invariant corrected reflectance ensures that comparisons can be made between measurements. (This definition is in line with that for PMD albedos in the previous algorithm). Zenith angles should be calculated at the top of the atmosphere and the above remarks concerning the satellite-to-atmosphere correction of these zenith angles are also appropriate here. Also, selection of the TOA height is required.

For each location, we then find the minimum reflectance (brightness) level over an extended period of time as follows. From GOME PMD data, we generate 3 images [“Red (R)”, “Green (G)”, and “Blue (B)”]. We represent colour characteristics in RG-space by defining co-ordinates  $r = N\rho(\lambda_R)$  and  $g = N\rho(\lambda_G)$ , where  $(1/N) = \rho(\lambda_R) + \rho(\lambda_G) + \rho(\lambda_B)$  defines the normalisation factor (the location  $(x, y)$  is assumed). If there are  $M$  red-green pairs  $\{rg_i\}$  at this location, then the one that designates the cloud free (CF in the equation below) value is defined by:

$$\|rg_{CF} - W\| \geq \|rg_i - W\| \quad i = 1, 2, \dots, M \quad (2)$$

where  $W$  is the white point in the RGB chromaticity diagram [C2], and the modulus denotes distance in RG-space.

For GOME, global monthly cloud-free composites are created using this method, and these are then merged to form an annual composite. Monthly composites are not totally cloud-free (3-day cover in the tropics translates to 10 measurements per month at the equator). Combining monthly composites removes ephemeral snow cover, but permanent snow and ice fields are still present (further work is needed to separate these). GOME historical data from 4 months in 1996-1997 were combined in this way to generate the first cloud-free composite.

The fractional cover at a given location is then a measure of the difference between the actual measurement  $rg$  and the cloud-free value  $rg_{CF}$  at that location, that is,  $f = S\|rg - rg_{CF}\|$ , where  $S$  is a proportionality factor. The algorithm can process whole orbits quickly, and is well suited to an operational environment. The algorithm has been compared with ICFA results taken from GDP level 2, and preliminary results show favourable qualitative comparisons with selected Meteosat images [C2]. Comparisons with the thresholding technique will be carried out as part of the CRAG study.

#### *Concluding Remark*

It has now been agreed that the PMD cloud coverage algorithm should be installed as a stand-alone part of the GOME level 0 to 1 processing (by user-community demand), as well as constituting part of the overall cloud retrieval algorithm in GDP level 1 to 2. This may also be considered for SCIAMACHY. A requirement on such a level 0 to 1b implementation is the use of a fixed TOA height, which must be agreed on by all parties at the outset. In level 1 to 2 processing, the top-of-the-atmosphere height is a generic input (allowed to vary).

### **3.1.3 Cloud Fitting Algorithms (ICFA and RCFA)**

*(Not part of NRT).*

The use of back-scatter spectra in and around the O<sub>2</sub> A band to determine cloud-top pressure has a long history (see for example [C7] and [C8]). The depth of the main A band absorption in the centre of the R branch (~ 760 nm) below the adjacent continuum level of the reflectance spectrum gives an indication of the penetration depth of the atmosphere (for fully cloudy conditions). We now describe the main features of the fitting problem, with its applications to cloud fractional

cover (ICFA) and the application to cloud-top pressure and optical depth retrieval (RCFA). It is assumed in both algorithms that the cloud-top can be treated as a reflecting boundary; later we discuss the validity of this assumption.

Least-squares fitting involves the chi-square merit function minimisation (see equation (10) on page 32). Suppose  $R_{meas}(\lambda_i)$  are measured reflectances (sun-normalised radiance values multiplied by  $\pi$ ) in a fitting window covering all or part of the  $A$  band. Let  $F_c$  be the cloud fraction of the pixel scene ( $F_c$  may be known from some other source or (for ICFA) it may be a free parameter in the fitting). Then the simulated reflectance may be written as a linear combination of terms to cloud-top and to ground level:

$$Y_{sim}(\lambda_i) = F_c + Y_{cloud}(\tau_c, p_c, \lambda_i) + (1 - F_c) \cdot Y_{ground}(p_g, \lambda_i) \quad (3)$$

Here  $p_c$  and  $p_g$  are the pressures at cloud-top and ground levels respectively, and  $\tau_c$  is the optical thickness of the cloud. In the fitting studies carried out for GOME, the main focus has been on the parameters  $\{F_c, p_c, \tau_c\}$ . For non-linear fitting, the repeated computation of full multiple scatter intensities is too time-consuming for operational use. This is largely because of the need to calculate simulated intensities at a high spectral resolution prior to a slit-function convolution to the wavelength grid of the GOME measurements.

#### ICFA description and status

First we shall describe the Initial Cloud Fitting Algorithm (ICFA) currently in operational use in GDP [G3]. ICFA fits only the cloud fraction  $F_c$ ;  $p_c$  and  $\tau_c$  are assumed known. ICFA assumes also that the simulated reflectance is a linear combination of the slant path transmittances  $T_{cloud}(\lambda_j, p_c, \mu, \mu_0)$  and  $T_{ground}(\lambda_j, p_g, \mu, \mu_0)$  plus an additive closure term to deal with scattering contributions:

$$Y_{sim}(\lambda_i) = F_c \mu_0 \alpha(\tau_c, \mu_0, \mu, \phi) T_{cloud}(p_c, \lambda_i, \mu_0, \mu) + (1 - F_c) \mu \beta T_{ground}(p_g, \lambda_i, \mu_0, \mu) + \gamma \left(1 - \frac{\lambda_i}{\lambda^*}\right) \quad (4)$$

( $\lambda^*$  is a reference wavelength, for example the mid-point of a fitting window covering all or part of the  $A$  band).  $\mu$  and  $\mu_0$  are the line-of-sight and solar zenith cosines respectively.  $\alpha(\tau_c, \mu, \mu_0, \phi)$  is the bi-directional cloud-top reflectance function which depends on the relative azimuth  $\phi$  as well as the zenith geometry,  $\beta$  is the ground albedo (assumed Lambertian), and  $\gamma$  is a closure coefficient (a linear term in wavelength has been assumed in the above equation).

If we define parameters  $A_1 \equiv F_c \cdot \mu_0 \cdot \alpha(\tau_c, \mu, \mu_0, \phi)$ ,  $A_2 \equiv (1 - F_c) \cdot \mu_0 \cdot \beta$  and  $A_3 \equiv \gamma$ , then equation (4) is linear in the parameters  $A_1$ ,  $A_2$  and  $A_3$ , and the fitting is then linear least squares. The result  $F_c$  comes from the fitted parameter  $A_1$  and knowledge of  $\alpha$ ; apart from a requirement to examine the fit diagnostics, the other two parameters are discarded.

The bi-directional reflectance  $\alpha$  is an important quantity in this algorithm. It consists of two terms - a reflectance  $\alpha_0(\mu, \mu_0, \phi)$  due to an infinitely optically thick cloud, and a second term  $C_{esc}(\tau, \mu, \mu_0)$  representing transmission loss through the cloud.  $\alpha_0(\mu, \mu_0, \phi)$  must be calculated using an invariance principle controlling the diffuse reflection from a semi-infinite optically

thick plane-parallel atmosphere [R1][R13]. The correction factor  $C_{esc}(\tau, \mu, \mu_0)$  deals with the finite optical thickness and is calculated using asymptotic theory [R14]. For water droplet clouds above a critical optical thickness  $\tau_{crit}$ , the correction factor is:

$$C_{esc}(\tau_c, \mu, \mu_0) = \frac{4(1 - A_g)K(\mu)K(\mu_0)}{3(1 - A_g)(1 - g)(\tau_c + 2q_0) + 4A_g} \quad (5)$$

This expression contains constants  $g$  and  $q_0$ , the escape function  $K(\mu)$  which gives the relative angular distribution of the radiation transmitted through the (optically thick) cloud, and the desired dependence on optical depth (see [C9] for detailed explanations).  $A_g$  is the albedo of the underlying ground surface (assumed Lambertian); the effect of  $A_g$  can be very considerable for thinner clouds over highly reflecting surfaces like snow and ice. The above definition assumes that the clouds have single scatter albedo equal to unity (conservative scattering); for the water droplet clouds defined for GOME [C10], the Mie scattering theory shows this assumption is valid at wavelengths around 760 nm in the near infrared. The approximation in equation (5) is sufficiently accurate for  $\tau > \tau_{crit} = 9$ .

A database of cloud-top reflectance quantities has been created for GOME [G9]. 27 quadrature cosines from 0 to 1 cover the zenith angle cosines  $\mu$  and  $\mu_0$ . The azimuth dependence in  $\alpha_0$  is expressed through a Fourier cosine series, for which 35 terms are generally sufficient using the ‘Delta-m’ approximation [R12]. 10 water-droplet clouds have been classified thus, for 3 wavelengths 750 nm, 775 nm and 800 nm.

The validity of the clouds-as-reflecting-boundaries approximation has been extensively investigated as part of the cloud study work for GOME [G9]. A full layer treatment of clouds has been incorporated into the radiative transfer model GOMETRAN [R4], and this has been used in conjunction with the DISORT forward model [R11] to compare back-scatter intensities. The results tend to show a persistent underestimation of the intensity in the reflecting boundary approximation, but it is difficult to establish any sort of consistent pattern. One suggestion for further work is the generation of a parameterised ‘reflectance correction’ data set to be used in tandem with the above database for reflectance  $\alpha$ .

The ICFA as implemented in GDP does not contain any line-by-line computations of transmittance. Instead, pre-calculated layer vertical transmittances were computed at 11001 points in wavenumber space (12,780 to 13240  $\text{cm}^{-1}$  at 0.04  $\text{cm}^{-1}$  resolution) covering the  $\text{O}_2$  A band. A 16-layer atmosphere was distilled from the US Standard atmosphere, with effective layer temperatures and pressures used in the line-by-line (LBL) cross section computation. [For an explanation of the LBL computation, see section 3.3.4 on page 45].

For solar zenith angles less than 75 degrees, the secant factor  $(1/\mu) + (1/\mu_0)$  is accurate enough for the layer slant path factors (plane parallel approximation); for higher solar zenith angles, a small look-up table of ray-traced path factors has been computed. For a given geometry, cumulative transmittances are given as the product of vertical layer transmittances with layer slant path exponents. The results are then convoluted with the GOME Channel 4 instrument response function to generate quantities used as templates in the fitting problem.

Cloud-top pressures (assumed known in ICFA) are taken from the ISCCP database [C11]; values are given on a 1x1 latitude/longitude grid, and the level 1b geolocation footprint co-ordinates

(centre of pixel) select the entry. Convoluted transmittances in equation (4) must be interpolated from those computed at the levels specified in the reference atmosphere.

It is important to have an accurate line spectroscopic database for O<sub>2</sub> A band absorption in the calculation of transmittances. This been the subject of some investigation in GOME, and an improved database has been created for ICFA [C12]. The line intensities and broadening uncertainties cause errors of up to 15%. Recent laboratory work is likely to provide more accurate data [C13]. This will be incorporated into the upgraded CRAG algorithm for GOME and by extension to the corresponding SCIAMACHY algorithm.

ICFA has a tendency to overestimate cloud fraction for scenes with inhomogeneous cloud fields, and there are some conceptual problems associated with cross-correlation of parameters  $A_1$  and  $A_2$  (the convoluted transmittance shapes to cloud-top and ground are very similar). There are additional question marks over the simple closure treatment of scattering, and also over shape distortions of the A band signature perhaps due to interference spectra not considered so far in the spectral fitting (under-sampling correction for GOME, Ring filling of absorption features by Raman scattering). These latter issues are currently under investigation.

#### *RCFA developments*

The non-linear fitting of  $\tau_c$  and/or  $p_c$  has also been the subject of some investigation [G9]. Preliminary results have shown that the cloud-top pressure is a robust parameter that may be fitted with some degree of confidence, but the optical thickness is less certain, particularly if the cloud fraction is also a free parameter to be fitted. It seems likely that one of the two abovementioned PMD cloud-clearing algorithms will be the method of choice for cloud fraction, with  $\tau_c$  and/or  $p_c$  derived from a revised cloud-fitting algorithm; a combination of methods may yield a more accurate approach.

With the cloud fraction assumed known, we may use an iterative technique (non-linear least squares, optimal estimation) to fit for  $\tau_c$  and  $p_c$ , using the combination of quantities in equation (3) for the simulated back-scatter. Separate intensities to ground and to cloud-top may be simulated using the GOMETRAN radiative transfer code in a formalism with clouds [R5]. An initial study has been carried out for one scenario containing a number of scenes over and around Hurricane Fran (4 September 1996). A look-up table of simulated intensities was created for the appropriate viewing conditions, and a chi-square grid-search was used to find a pair of values  $\tau_c$  and  $p_c$  corresponding to a minimum value of chi-square. The results are very promising, in particular for the cloud-top height, as verified by available synoptic information (T. Kurosu, private communication).

Two additional methods have been suggested for cloud fraction retrieval. The first uses a simple linear regression on continuum intensities (GOME or SCIAMACHY channel 4, away from H<sub>2</sub>O or O<sub>2</sub> absorption bands), using a look-up table. The second is also a linear regression on intensities in and around the prominent Ca II Fraunhofer lines (~ 393 nm). This method utilises the difference in the levels of Raman scattered light for cloudy and clear scenes.

#### *Further developments for SCIAMACHY*

The database of cloud-top reflectance quantities has recently been extended to cover SCIAMACHY wavelengths. Mie theory computations show that single scattering albedos are significantly

different from unity for SCIAMACHY infrared wavelengths, and in consequence the expression for the transmission loss term in the asymptotic approximation contains extra terms. With the notation as in equation (5), and with constants  $m$ ,  $n$ ,  $l$ ,  $k$  and  $A^*$  calculated under the assumption of non-conservative scattering [C9], the loss term is:

$$C_{esc}(\tau_c, \mu, \mu_0) = \frac{m[(1 - A_g A^*)l - A_g m n^2]K(\mu)K(\mu_0)e^{-2k\tau_c}}{[(1 - A_g A^*)(1 - l^2 e^{-2k\tau_c}) + A_g m n^2 l e^{-2k\tau_c}]} \quad (6)$$

As far as O<sub>2</sub> A band fitting algorithms are concerned, GOME and SCIAMACHY developments are concomitant. However, SCIAMACHY does have infrared channels, and there are additional possibilities for cloud information retrieval using related spectral fitting algorithms in the water bands around 930 nm (channel 5), and the CO<sub>2</sub> bands around 2000 nm (channel 7). In particular, the retrieval of cloud-top pressure from CO<sub>2</sub> absorption is in principle very similar to the IAS application for the retrieval of an equivalent reflecting height (see section 3.3 on page 42). Here, one could assume that the cloud fraction was known, and then compute transmittances to fit for the cloud-top pressure and optical depth. Light scattering by clouds is non-conservative in this part of the infrared and expression (6) would be required for the reflectance escape function correction.

Finally a word on cloud type. It should be possible with the PMD information available from SCIAMACHY to distinguish between water and ice clouds, with a further classification of ice clouds into optically thin and optically thick types (examination of brightness values in infrared?). The cloud fitting algorithm will provide additional information about the cloud-top height and the optical depth, but this is not in itself enough to make an unambiguous assignment of water cloud type. An assignment of cloud type should probably be made according to the WMO scheme. It is also intended to look at fractional polarisation values (as delivered by the level 0 to 1 algorithm) for the identification of cirrus.

### 3.1.4 Introduction to Aerosol Pre-processing Algorithms

The last section deals with the retrieval of aerosol from SCIAMACHY nadir data. This field is still under development; we confine our attention to the generation of an absorbing aerosol index value (the AAIA algorithm), and the generation of an optical thickness value (AOTA algorithm). The AAIA has been considered for GOME as well, and the treatment is the same. The AOTA has been developed for GOME, and the treatment is also applicable directly to SCIAMACHY.

*The AAIA appears in the NRT specifications; AOTA is not part of NRT.*

### 3.1.5 Aerosol Absorbing Index Algorithm (AAIA)

*(Applies equally to SCIAMACHY OL and NRT).*

It has been shown that an *absorbing* aerosol index (AAI) may be obtained for remote sensing observations by looking at the logarithmic difference between two earthshine reflectance that have been corrected for Rayleigh scattering. The usefulness of this index as a tracer for dust and smoke aerosols was first demonstrated for the TOMS instrument for measurements of biomass smoke burning over the Amazon basin [C14]. Similar results have been obtained using GOME level 1 data (J. Gleason, private communication, 1997). For the specification of the SCIAMACHY NRT algorithm it was decided to adopt this definition of an absorbing aerosol index [A8].

The AAI in itself is not a physical quantity, but recent work has established a relation between this index and aerosol optical depth [C15].

Reflectances at 340 nm and 360 nm have been selected for the TOMS work. Trace gas absorption is low at these wavelengths, and the spectral signature of non-absorbing aerosols is flat in this part of spectrum. However, absorbing aerosols such as soot have low single scattering albedo in the UV, and the absorption shows a marked fall-off with wavelength. Thus once the Rayleigh  $\lambda^{-4}$  signature has been accounted for, the corrected reflectance ratio should mark the presence of the absorbing aerosols. The definition of AAI is:

$$AAI = -100 \cdot \log \left( \frac{I_{meas}(340)}{I_{calc}(340)} \right) \quad (7)$$

The calculation is facilitated by using the measurement albedo at 360 nm in the theoretical simulation at 340 nm. High values of the logarithm of the reflectance pair ratio will indicate the presence of absorbing aerosols. It has been noticed that even without the Rayleigh correction, the ratio of reflectances at 340 and 360 nm shows a marked deviation from normal values in the presence of smoke aerosols [U9]. Such absorbing aerosols are found principally in the atmospheric boundary layer and their detection is compromised for partially and totally cloudy scenes.

SCIAMACHY earthshine and solar measurements are smoothed and binned in small buffers around each wavelength, and reflectances calculated for these buffers. Mean values of reflectance are then taken across the buffers. A simulated reflectance for a pure Rayleigh atmosphere must then be calculated at 340 nm, using the reflectance at 360 nm as the surface albedo. A set of pre-calculated look-up table coefficients was derived for the extraction of appropriate Rayleigh correction reflectances.

### 3.1.6 Aerosol Optical Thickness Algorithm (AOTA).

We confine our attention here to summarising what has been done so far for GOME; the adaptation to SCIAMACHY will be clear. Some of the GOME work has been done as part of the CADAPA study [G10], and the following summary is drawn from that source. In conjunction with this study, further work was done as part of a project to develop operational cloud and aerosol products at I-PAF (see for example, the presentation at GOME and SCIAMACHY Working Sessions 7, April 1998), and this operational implementation is now under way.

Aerosol optical thickness is best retrieved from nadir reflectance measurements in parts of the spectrum where trace gas absorption is minimal. For GOME, there are 3 suitable regions; one from 357-430 nm, plus two other regions adjacent to the O<sub>2</sub> A band (750.8 - 758.4 nm, and 774.5 - 787.0 nm). Wavelength windows were chosen by examining a number of simulated reflectances (produced by MODTRAN [R10]) for which atmospheric trace gas transmittances were greater than 0.996 over ranges greater than 3 nm. A total of 28 wavelengths were selected from the above three regions (3 nm resolution from 361 - 427 nm, plus 5 others from channel 4). Some measurements around 425-430 nm may require correction for O<sub>3</sub> Chappuis absorption.

These GOME measurements span channels 2, 3 and 4, and it is important to take care of inter-channel discrepancies at the channel 2/3 border (cross-over gradient, jump mismatch of radiances). Box averaging was used to render the spectra Ring-free [C16]. A new and flexible data base of aerosol optical properties has been assembled for this work [C17]. Properties for a large number of aerosol classes can be derived from a data set of basic component properties using

internal or external (WMO) mixing [C18]. 15 aerosol classes were considered for the present algorithm.

The simulation of nadir back-scatter measurements was done using a modified form of the DIS-ORT model. In the absence of trace gas absorption, multiple scattering can be approximated using a limited number of zenith stream angles, so long as the single scatter contributions are modelled with full accuracy [C19]. A direct inversion retrieval for this problem has not been attempted. For each pixel scene, simulations of reflectance values at the chosen wavelengths are made for a number of values of aerosol optical thickness and for a number of aerosol classes. These results are then compared with measurements to find the best fit; simulated values are not analytic functions of aerosol optical thickness, so the fitting is non-linear least squares. For each aerosol class, the fitting is performed for the single varying parameter (the aerosol optical thickness at 500 nm); the desired result is obtained for that class which has the lowest chi-square merit function [G10].

Sensitivity analyses have been carried out for two effects. The first is cloud contamination and the results clearly indicate that the AOT retrieval is seriously compromised in the presence of clouds. The second study focussed on the effect of heterogeneous footprint surfaces; results are unreliable for a predominantly water surface with some land contamination, less so for mixed land-only footprints. Therefore only cloud-free pixels have been considered in the test cases considered so far for the GOME application. A dust aerosol test case over the Sahara and Atlantic Ocean was selected. Contemporaneous METEOSAT images were used for scene selection, with some validation measurements from a ground-based station [G10]. Preliminary results show good semi-quantitative estimates of AOT over the ocean and desert, though difficulties were found in the definition of suitable dust aerosol classes [C20].

Derivations of aerosol optical thickness values are possible from SCIAMACHY near infrared nadir measurements (channels 5 - 8), but there are no algorithms on the table at the time of writing.

## 3.2 UV-Visible Absorption Spectroscopy (UVAS)

### 3.2.1 Introduction

Long-path differential absorption is a well-known method for measuring amounts of atmospheric trace gases. The technique has been used widely in the last two decades for the derivation of trace species abundance from a variety of ground-based, balloon and air-borne, and ship-based instruments measuring the zenith sky in the near UV and visible ([U1] [U2], [U3] and references therein). The use of DOAS (Differential Optical Absorption Spectroscopy) in GOME total column retrieval represents the first successful application of this technique to measurements from space-borne passive remote sensing instruments [G3]. Besides O<sub>3</sub>, abundances of a number of trace species have now been detected with GOME (see table below).

Since SCIAMACHY has a similar nadir-scan measurement capability to GOME, the DOAS retrieval technique is clearly applicable to its UV and visible back-scatter spectra. Much of the following algorithm description applies equally well to either instrument.

We start with a description of DOAS, following for the most part the exegesis in [G3]. This is followed by a summary of a new variant on the slant column fitting of trace species which may be considered as an option for SCIAMACHY OL. The discussions on fitting window choice and reference spectra are based on the text in [G3], with special applications to SCIAMACHY. The following two sections are concerned with Air Mass Factor (AMF) computation and the final evaluation of vertical column densities respectively. The last section deals with operational implementation issues.

*For the NRT, the DOAS algorithm as used in GDP has been taken over. AMF tables will be created especially for the NRT, based on O<sub>3</sub> and NO<sub>2</sub> profile climatology used in GOME. Because of the need to freeze specifications at an early stage, recent developments on these AMFs as described below will not be considered for NRT. Similarly the ‘direct fitting’ slant column algorithm will only be considered for the OL processor.*

### 3.2.2 The DOAS algorithm

Trace gas absorption is based on the Beer-Lambert law - the optical density due to trace gas absorption is the coefficient of absorption times the absorber column amount. If  $I(\lambda)$  is the attenuated intensity and  $I_0(\lambda)$  the incoming intensity at the top of the atmosphere, then we write:

$$\ln\left(\frac{I(\lambda)}{I_0(\lambda)}\right) = -\sigma_1(\lambda) \cdot A_1(s) - \sigma_2(\lambda) \cdot A_2(s) - \dots \quad (8)$$

Here,  $A_j(s)$  is the effective slant column density of the  $j^{\text{th}}$  absorber over path length  $s$ , and  $\sigma_j(\lambda)$  is the coefficient of absorption.

Incident solar light is further attenuated by molecular (Rayleigh) scattering, by absorption and scattering due to particulate matter (clouds and aerosols), and by reflection from the earth’s surface. In contrast with the differential (sharp with moderate resolution) spectral features characteristic of some trace gas absorption in the UV/visible, these other attenuation effects all contribute broad-scale features to the back-scatter spectrum. The broad features are approximated by a low order additive polynomial in wavelength, and the simulated optical density  $Y_{sim}(\lambda)$  is given by:

$$Y_{sim}(\lambda) \equiv \ln\left(\frac{I(\lambda)}{I_0(\lambda)}\right) = -\sigma_1(\lambda) \cdot A_1(s) - \sigma_2(\lambda) \cdot A_2(s) - \dots - B_1 - B_2(\lambda - \lambda_0) - B_3(\lambda - \lambda_0)^2 \quad (9)$$

The measured optical density may be expressed in terms of the SCIAMACHY-measured nadir back-scatter spectrum  $I_{nadir}(\lambda)$  and the extra-terrestrial reference spectrum  $I_{sun}(\lambda)$  as the logarithm of the reflectance ratio:  $Y_{meas}(\lambda) = \ln[I_{nadir}(\lambda)/I_{sun}(\lambda)]$ .

Least-squares fitting involves minimisation of the chi-square merit functional for N observations included in a given fitting window:

$$\chi^2 = \sum_{i=1}^N \left( \frac{Y_{meas}(\lambda_i) - Y_{sim}(\lambda_i)}{\varepsilon_{meas}(\lambda_i)} \right)^2 \quad (10)$$

The simulated optical density in equation (9) has a linear dependence on the set of fitting amplitudes  $\{A_i, B_j\}$ ; the fitting for these quantities is a straightforward multi-linear regression. Inclusion of the measurement errors  $\varepsilon_{meas}(\lambda_i)$  makes the fitting weighted; an assumption of constant error produces an un-weighted fit.

The trace gas cross sections are assumed known, and are usually taken from laboratory measurements (see below) - they constitute reference data for the fitting. Experience with DOAS in GOME has shown that an additional reference spectrum should be added to the list of fitted spectra; this is the ‘‘Ring’’ spectrum, due to atmospheric inelastic Raman scattering of the Fraunhofer spectrum. In the fitting, the Ring spectrum is treated as a linear pseudo-absorber.

In practice, there will be some uncertainties in the wavelength matching between the various spectra, and the fitting can be improved by compensating for these registration errors. Each reference spectrum can be translated in wavelength by a single value (shift), and stretched or compressed by a single value (squeeze) about some reference point. The shift/squeeze adjustment is:

$$\lambda^* = \lambda + \langle shift \rangle + (\langle squeeze \rangle - 1) \cdot (\lambda - \lambda_{ref}) \quad (11)$$

$Y_{sim}(\lambda)$  depends non-linearly on the shift/squeeze parameter pairs, which must be determined iteratively using a non-linear least squares method [U4]. The linear regression for  $\{A_i, B_j\}$  is now embedded in the non-linear shift/squeeze fitting, and must be performed at each iteration. Once the non-linear fit has converged, another linear regression is performed to establish the final values of the amplitudes  $\{A_i, B_j\}$ .

In this DOAS approach, wavelength calibration of measured spectra is already performed in level 0 to 1b processing, and there is no attempt to further reference them. Another assumption is that the form of the SCIAMACHY instrument response function (sometimes called the ‘‘slit function’’) is assumed known, so that convolution of reference spectra to the wavelength resolution of the instrument (if required) is in principle known for all wavelength calibrations. Convolution is not necessary when reference spectra have been derived from flight model laboratory measurements (as part of the pre-launch calibration program); this has been the practice in GOME with  $O_3$  and  $NO_2$  cross sections. The advantage of using such cross sections is that some instrument-derived distortions can be accounted for in the fitting [U5]. For ‘‘minor’’ absorbers, it has often proved expedient to run a low-pass smoothing filter over the measurement data to improve the fit,

at the expense of some precision. (If GOME spectra are smoothed, then so must the reference spectra be smoothed).

The most important diagnostics from the least-squares fitting are the solution covariance matrices  $S^{(L)}$  and  $S^{(N)}$  for the linear and non-linear problems respectively. For weighted fitting, the fitting amplitude errors are the square roots of the diagonal entries of these matrices; thus the error on effective slant column  $A_i$  is  $\sqrt{S_{ii}^{(L)}}$ . For un-weighted fitting, the fitting amplitude errors must be multiplied by  $\sqrt{(\chi^2/\nu)}$ , where  $\nu$  is the number of degrees of freedom (number of observations less number of fitted parameters). Off-diagonal entries in the solution covariance will be more conveniently expressed in terms of the cross-correlation matrix, given by  $C_{ab} \equiv S_{ab} / [\sqrt{S_{aa}} \sqrt{S_{bb}}]$ .

Chi-square is also a diagnostic. Another single-number diagnostic is the RMS value:

$$RMS = \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^N [Y_{meas}(\lambda_i) - Y_{sim}(\lambda_i)]^2 \quad (12)$$

The goodness-of-fit statistic is sometimes useful for weighted fitting; this is the probability that the final value of chi-square obtained in the fit should occur by chance, and is given by the value of the incomplete gamma function  $Q\left(\frac{1}{2}\nu, \frac{1}{2}\chi^2\right)$ , where  $\nu$  is again the number of degrees of freedom. (Un-weighted fitting is equivalent to the assumption of a good fit).

For SCIAMACHY retrieval, trace gas slant column amplitudes constitute the main results to be included in the product, which will also contain some diagnostic information (see chapter 6 for product content). Before we discuss fitting windows and targeted trace species, we mention a recent improvement in spectral fitting.

### 3.2.3 Direct Fitting UV-Visible algorithm

An alternative approach involves the direct fitting of measured back-scatter spectra to their simulated equivalents [U6]. There are no logarithms; in the above equations, we use  $I_{nadir}(\lambda)$  in place of  $Y_{meas}(\lambda)$  as the basic measurement quantity of the fit. The amplitudes  $\{A_i, B_j\}$  plus polynomial closure terms are fitted directly by non-linear least squares iteration. The fit includes a Ring correction.

Shifts and squeezes are not fitted in this algorithm; instead all radiance and irradiance spectra are wavelength-registered against a very accurate external Fraunhofer reference spectrum [U7] before the fit takes place. Using least squares fitting for this registration also allows an independent determination of the (assumed Gaussian) instrument response half-width. Reference spectra are also pre-calibrated against more accurate standards.

The fitting of GOME spectra for trace species retrieval includes an additional reference spectrum to account for unwanted structures in the measurement spectra caused by incomplete sampling of the slit function (FWHM for GOME in channel 2 is typically 1.7 pixels, below the Nyquist limit of 2 pixels). A synthetic *under-sampling* spectrum is constructed from a convolution of the high-resolution Fraunhofer reference spectrum with the fitted Gaussian slit function by contrasting sampled representations of this convoluted spectrum. Inclusion of this under-sampling in the trace

gas retrieval improves the fitting dramatically; the results are more precise, and there is no need for additional smoothing.

The technique has been tried for whole orbits of GOME data to produce global distributions of BrO [U6]. The enhanced accuracy of the method (in particular the inclusion of the under-sampling correction) allows absorption features of several parts in  $10^{-4}$  to be seen by GOME; it is now possible to measure global background SO<sub>2</sub> and ozone-hole ClO with this method (K. Chance, private communication, 1998). Work is currently underway to extend this technique to NO<sub>2</sub> and H<sub>2</sub>CO retrieval.

### 3.2.4 Choice of Fitting Windows

There are several considerations to bear in mind when choosing fitting windows. In general, each window should contain a set of SCIAMACHY observations from a single cluster readout, with no cross-channel or cross-cluster mixing of measurements. (Measurements are usually, but not necessarily, contiguous). The number of points in a window should not be too low (poor fit quality) or too high;  $50 < N < 500$  is suitable.

Windows are selected to highlight particular trace gases. Interfering species should if possible be avoided, along with regions containing strong Fraunhofer lines (though the Ring effect can be partially compensated for, as noted above). One should also avoid regions of low signal-to-noise (large measurement uncertainty) and wavelengths at which unwanted instrumental residues present in the measurement spectra (stray light corruption, dichroic effects).

In the 'classical' DOAS approach, a single Air Mass Factor (AMF) is used to convert slant to vertical column density. For a given trace species, the window should be chosen so that the corresponding AMF is representative; for minor absorbers showing optical depths  $10^{-2}$  to  $10^{-4}$ , this is not a problem, and one normally selects the window mid-point wavelength for the AMF computation. However, for strong non-linear absorption in the O<sub>3</sub> Huggins bands, this AMF issue is an important driver in the choice of window limits (see below in section 3.2.6).

<i>Window (nm)</i>	<i>Featured Absorber</i>	<i>Interfering Species</i>	<i>Usage to date</i>	<i>Source</i>
325 - 335	O <sub>3</sub> (Huggins)	NO <sub>2</sub> , Ring	GDP	[U8]
425 - 450	NO <sub>2</sub>	O <sub>3</sub> , Ring	GDP	[U8][U10]
510 - 560	O <sub>3</sub> (Chappuis)	NO <sub>2</sub> , Ring, H <sub>2</sub> O, O <sub>4</sub>	GDP proposed	[G3]
336 - 354	H <sub>2</sub> CO	O <sub>3</sub> , NO <sub>2</sub> , Ring, BrO	GOME off-line	[U9][U10]
314 - 327	SO <sub>2</sub>	O <sub>3</sub> , Ring	GOME off-line	[U10]
340 - 360	BrO	O <sub>3</sub> , NO <sub>2</sub> , Ring	GOME off-line	[U11][U6]
350 - 380	OCIO	O <sub>3</sub> , NO <sub>2</sub> , Ring, BrO	GOME off-line	[U11]
305 - 312	ClO	O <sub>3</sub> , NO <sub>2</sub> , Ring	GOME off-line	KC

**Table 3: Fitting Windows in the UV and visible wavelength range**

In the table above, the first two windows for O<sub>3</sub> and NO<sub>2</sub> were selected for GDP, and it is proposed that they be also used for SCIAMACHY. Work done on minor trace species DOAS retrieval from GOME measurements has helped to define windows for H<sub>2</sub>CO, BrO, SO<sub>2</sub>, OCIO

and ClO. For minor species in the UV, O<sub>3</sub> is often a serious interfering species and must be fitted simultaneously; the ClO signatures are particularly hard to distinguish from the very substantial overlying O<sub>3</sub> absorption. In the visible, NO<sub>2</sub> and O<sub>3</sub> (Chappuis) are fitted together; the main issue here is to avoid wherever possible regions of H<sub>2</sub>O and O<sub>2</sub>-O<sub>2</sub> absorption. [The fitting of H<sub>2</sub>O requires a different approach and has not so far been considered for SCIAMACHY].

As far as the products are concerned, research on BrO and SO<sub>2</sub> retrieval using GOME data indicates that the fitting is mature enough to consider these two gases as retrievable on a global basis. This will be the default for SCIAMACHY. For HCHO the fitting is still most effective for special scenarios such as biomass burning, but despite this, a decision has been made to retrieve this species on a global basis for SCIAMACHY. OCIO is restricted to special circumstances, and a baseline has to be determined for the scenarios to be considered for fitting this gas. The ClO status is still too undeveloped for a determination yet as a SCIAMACHY product.

### 3.2.5 Reference Spectra

#### Ozone

The benchmark UV data of Bass and Paur [U12] was used in the original GDP and is still a candidate for SCIAMACHY UV fitting. The well-known temperature dependence in the Hartley-Huggins bands is expressed through the quadratic parameterisation:

$$\sigma_{\lambda}(T) = \sigma_{\lambda}(T_0) \cdot (1 + a_{\lambda}(T - T_0) + b_{\lambda}(T - T_0)^2) \quad (13)$$

$\sigma_{\lambda}(T_0)$  is the cross section at reference temperature  $T_0$  (273.15K);  $\{a_{\lambda}, b_{\lambda}\}$  are parameterisation coefficients. For the visible, the unpublished data of Johnson [U13] was used originally for GOME (no temperature dependence); this is probably out of date now, and a more suitable compilation should be considered (for example, see [U14]).

As part of the calibration programme, laboratory measurements were made with a reaction cell and the GOME flight model during the pre-launch calibration phase. O<sub>3</sub> UV and visible cross sections at 5 temperatures were derived from these measurements; the results have recently been recalibrated [U15], and these 'flight model' cross-sections are now the default for GDP. The quadratic parameterisation in the previous equation has been applied to the data. Unlike the older literature data, they do not require convolution or smoothing before use in the fitting. It is intended to derive cross sections from pre-launch laboratory measurements with SCIAMACHY at the end of the calibration phase (October 1998). A further option for the future will be the use of high-resolution cross sections derived from laboratory Fourier Transform spectrometer (FTS) measurements (work in progress at a number of institutions).

#### NO<sub>2</sub>

The original data [U16] used for GOME have now been supplanted by GOME 'flight model' cross sections also derived from pre-launch laboratory measurements taken during the calibration phase [U17]. This molecule has in recent years been the subject to a number of new laboratory studies, and there is wealth of new data available [U18][U19]. It was agreed last year (SSAG 14, Brussels, October 1997) that a new compilation should be made, and this work has now started. The compilation will distil the work of several FTS teams active in the field. Temperature dependence is turning out to be important for this species, but only preliminary attempts have been made so far at a parameterisation scheme [U20]. It is also intended to derive SCIAMACHY flight-model cross sections for this gas.

## Ring

A differential Ring cross section was derived from GOME pre-launch zenith sky observations using an established measurement technique [U21]. Theoretical Ring spectra have been derived from model studies of Raman inelastic scattering [U7][R6][G12]. It now seems clear that the model-produced Ring reference spectra are to be preferred, and the theoretical results obtained in GOME studies can certainly be used again for SCIAMACHY. Some zenith sky measurements are planned with the flight model to determine an instrumental Ring absorption spectrum.

## Other Species

For BrO, the literature reference spectrum used for GOME [U22] is still extant, though new FTS work is expected to provide more accurate data (P. Wennberg, private communication, 1998). For SO<sub>2</sub>, the situation is much the same: for the GOME standard, see [U23]. For ClO [U24], OClO [U25], H<sub>2</sub>CO [U26] and O<sub>2</sub>-O<sub>2</sub> [U27], improvements over the spectra currently used for GOME are also needed.

## Temperature Dependence

For trace species with temperature-sensitive cross-sections, the choice of a representative temperature to use in the DOAS fitting is important. The current default for GOME GDP O<sub>3</sub> is to choose the temperature at an altitude corresponding to the peak number density of an appropriate ozone profile (temperature and ozone profiles are part of the same climatology or model output). An improvement here might be the use of analysed ECMWF temperature fields rather than climatological profiles.

Studies on GOME data have shown that O<sub>3</sub> columns are sensitive to the choice of cross section temperature; in this connection, a data set of effective temperatures for DOAS-type retrievals of UV ozone has been derived for use on a global basis (A. Piters, GOME and SCIAMACHY Working Sessions 5, February 1997).

Another alternative is to treat the temperature as a variable to be extracted from the fitting. To this end, one employs two O<sub>3</sub> cross sections at different (but known) temperatures in the fitting, and the actual ozone slant column content and effective temperature are then derived from the two fitting parameters associated with the two reference spectra (A. Richter, GOME and SCIAMACHY Working Sessions 7, April 1998).

For GOME GDP NO<sub>2</sub>, the flight model cross sections at 221 K are used (nearest representative temperature corresponding to the peak concentration of NO<sub>2</sub> in the stratosphere). The temperature dependence for this species is not so marked as that for ozone, and it is not clear yet whether this value is really optimal.

### 3.2.6 Air Mass Factors

Irrespective of the form of the fitting algorithm, trace gas slant columns must be converted to vertical column amounts (which are independent of viewing geometry). This conversion is accomplished by dividing with suitable Air Mass Factors (AMFs); these represent the enhancement of trace gas absorption due to slant paths of incident light through the atmosphere. For a given gas *g*, the AMF is the ratio of slant path to vertical optical depths, and is defined by:

$$AMF_{\lambda}(g) = \frac{\ln(I_{nogas}(\lambda)) - \ln(I_{total}(\lambda))}{\tau_{vert}(\lambda)} \quad (14)$$

$I_{nogas}$  is the back-scattered intensity including all absorbers except the trace gas of interest, while  $I_{total}$  is the intensity including all absorbers. The vertical optical depth is determined from the extinction integration of profile concentration  $C_g(z)$  multiplied by the cross section  $\sigma_g(z, \lambda)$  at the wavelength  $\lambda$  of interest:

$$\tau_{vert}(\lambda) = \int_0^{z_0} \sigma_g(z, \lambda) \cdot C_g(z) dz \quad (15)$$

For a trace species with small optical depth and for a plane-parallel atmosphere, the AMF is well approximated by the geometrical secant factor  $AMF_{geom} = \sec\theta_0 + \sec\theta_1$ , where the angles are the solar and nadir line-of-sight zenith angles respectively. For long paths and for regimes where the trace species absorption is deeper and non-linear, a full radiative transfer (RT) simulation in a curved refracting atmosphere is required.

For an accurate computation of the AMF, one must run the RT forward model twice to compute  $I_{nogas}$  and  $I_{total}$ . In some circumstances, a good approximation to the AMF may be obtained from a single computation of  $I_{total}$  and its associated weighing function output [R7]. Note that a top of the atmosphere (TOA) height must be defined for the RT transfer simulations.

The UV/visible *vertical* column retrieval as described so far is conveniently separated into a fitting algorithm for the slant column and a pure simulation of the AMF. However, there is a flaw in this logic. It is assumed that the trace species profile is known in advance in order to compute the AMF to achieve the retrieved column amount for that species. In most cases (and in general for low optical depths) the column amount is insensitive to choice of assumed profile in the AMF computation. There are circumstances in which AMFs are sensitive to profile shape, and choosing a profile which bears little resemblance to the actual atmospheric profile can lead to an AMF that is incorrect and hence to a vertical column result that is inaccurate or wrong. In the next section on vertical column calculation, a simple quality control method for AMF selection is outlined.

Details of AMF simulation depend on the nature of the RT model used for the computation. The AMF will depend not only on viewing geometry, but also on the chosen trace species climatology, on the assumed molecular density and aerosol loading, and on the lower boundary surface reflecting property. For cloud-contaminated scenes, AMFs to the ground and to cloud-top are required. The assumption adopted for GOME is that clouds may be treated as bi-directional reflecting surfaces. This is good enough for water droplet clouds above a certain optical thickness (see above in section 3.1 on page 21, and for more details, see [G9] for example).

A full multiple scattering treatment cannot be avoided in the UV and visible regions. For SCIAMACHY (as was the case for GOME), viewing geometry information is available several times over one pixel, and assuming horizontal homogeneity in all other atmospheric variables, it is possible to calculate several AMFs and form a weighted sum that will be representative for that footprint. This is known as the “extended field-of-view (FOV)” weighting option [G3]. For GOME the 3-sum weighting was done with intensity ratios and parabolic weights ( $W_1 = W_3 = 1, W_2 = 4$ ) in the following manner (the assumption of small optical depth is a prerequisite of this result):

$$AMF_{fov}(g) = \sum_{i=1}^3 \frac{I_{nogas}(i)}{I_{total}(i)} \cdot W_i \cdot AMF_i(g) \quad (16)$$

For species with small optical depths, AMFs do not vary much with wavelength over the relevant fitting windows, and a value computed at the mid-point of the window is representative ( $\text{NO}_2$  AMFs were calculated at 437.5 nm for example). For  $\text{O}_3$  in the Huggins bands, the absorption is stronger and non-linear, and the AMFs show considerable variation with wavelength between 325 and 335 nm, and variation also with solar zenith angle. Studies carried out for GOME have shown that the choice of 325 nm for the AMF gives a value closest to the actual enhancement of slant path absorption [G6]. Ideally one should account for the wavelength dependence of the AMF in the  $\text{O}_3$  Huggins window, and perhaps try to fit for the vertical column density by using an AMF-modified cross section in the fitting (see remarks below).

It was intended originally with GOME to calculate the AMFs from scratch, but due to unforeseen circumstances, an interim solution was adopted for the first operational GDP. In this, the single scatter AMFs are computed *ab initio* using a model developed at the University of Heidelberg [G6], and multiple scatter correction factors are then used to scale up the AMF to its full value. These factors were parameterised in a look-up table, and pre-calculated using the GOMETRAN model [R4]. No attempt was made to estimate errors on the AMFs due to model input assumptions.

The updated GDP planned for 1999 will use look-up tables only, with no radiative transfer simulations from scratch; this update will also use improved climatology for  $\text{NO}_2$  and  $\text{O}_3$  profiles in the AMF selection. The SAGE-II/Balloon Climatology for  $\text{O}_3$  [U28] contains 26 profiles expressed as column Umkehr layers (units [DU]), and a single error covariance matrix (in units of  $[\text{DU}]^2$ ). By evaluating the AMF derivatives with respect to these Umkehr amounts, one can then use the error covariance to compute variances for the AMFs, and ultimately to examine errors on the vertical column result due to the  $\text{O}_3$  profile choice uncertainties (see next sub-section).

GOMETRAN is a suitable RT model for generating look-up tables of  $\text{O}_3$  AMF components, and the weighting function output from this RT can be used directly to compute AMF Umkehr column derivatives (see section 5.1 on page 75 for details). The work undertaken for GOME can be taken over directly for SCIAMACHY. It is proposed to use new look-up tables based on the improved climatology for  $\text{O}_3$  profiles in both the NRT and OL processors for SCIAMACHY.

For  $\text{NO}_2$ , a suitable profile climatology for AMF computation does not exist, and the model output profiles used for GDP [G4] have so far proved unreliable. Work is currently underway to generate a new climatology. In the stratosphere down to the tropopause, this is based on HALOE  $\text{NO}_2$  occultation profile data [U29]. Stratospheric profiles show little variation on a zonal basis, and stratospheric  $\text{NO}_2$  AMFs are fairly insensitive to the column content above the tropopause. Tropospheric  $\text{NO}_2$  is very variable, and large boundary layer concentrations can greatly affect AMF values. To allow for this, the ground level concentration is regarded as a variable parameter, and a number of analytic tropospheric profiles are matched to the stratospheric data to generate the climatology. AMFs constructed from this basis may be used iteratively to generate a vertical column result that supports the fitted slant column amount (see next section).

#### *Remark on Modified DOAS*

This is a form of DOAS wherein reference cross-sections are multiplied by pre-computed AMFs in the DOAS fitting environment; the fitted parameters are then vertical column amounts. So far this has been applied only to Huggins  $\text{O}_3$  in the UV [G6]; the advantage here is that the wavelength dependence of the AMF can be accounted for in the fitting window 325 - 335 nm. Off-line case studies have shown that the fit quality is improved by this method, but there is some way to

go before the method could be implemented globally in the operational environment (at the very least, a comprehensive parameterisation scheme would be required). Operational implementation would have to rely on pre-calculated look-up tables; AMFs would have to be computed at many wavelengths, adding an extra dimension to the size of the tables already being developed for regular DOAS. Status TBD.

### 3.2.7 Vertical Column Density

For a clear sky footprint, the vertical column  $V$  to the ground is simply the slant column  $E$  divided by the Air Mass Factor  $AMF_{clear}$ , the latter computed or extracted for the appropriate viewing geometry and climatological profile (and also ground height, surface reflectance and particulate loading). In order to deal with cloud-contaminated scenes, we assume that clouds are equally distributed over the footprint scene with the same cloud-top height. We then take the fractional cloud cover result  $F_c$  from the cloud pre-processing algorithm, and compute a total AMF for the scene:

$$AMF_{total} = F_c \cdot AMF_{cloudy} + (1 - F_c) \cdot AMF_{clear} \quad (17)$$

(If the option has been set, the extended-FOV weighting is assumed for these AMFs). For the cloudy parts of the pixel, the AMF is known only to the cloud-top, and we have no knowledge of the vertical column between ground and cloud-top. This latter column is the “ghost vertical column (GVC)”, required for the final computation of the total vertical column  $V_{total}$  to the ground.

The end result is:

$$V_{total} = \frac{E + F_c \cdot GVC \cdot AMF_{cloudy}}{AMF_{total}} \quad (18)$$

The ground height (topography data base) and cloud-top height (retrieved from RCFA cloud pre-processing) are assumed known. One way of estimating GVC is to integrate the appropriate part of the trace gas climatological profile chosen for the AMF evaluations (this is the current GOME default). Another way (so far untried) is to use results from two different fitting windows. If  $O_3$  has been fitted with slant column results  $E_1$  from the UV (Huggins) window and  $E_2$  from a visible (Chappuis) window, and the corresponding clear sky AMF values are  $A_1$  and  $A_2$ , then an estimate for GVC is  $(E_2/A_2) - (E_1/A_1)$ .

So far with GDP, there has been no simple way to exercise any sort of quality control on the vertical column result. In the classical DOAS/AMF approach, only one set of AMFs can be computed for a given scene; there is no way to improve upon the VCD result if the AMFs are not correct. The most serious situation where this could occur is in a depleted  $O_3$  scenario (associated with the ozone hole phenomenon). Here, AMF values computed with a ‘normal’ (non-hole)  $O_3$  climatological profile will be in error by as much as 10%; they are not representative for the corresponding low slant column result obtained from the DOAS fit.

The use of look-up tables based on the SAGE-II/balloon column-classified  $O_3$  climatology allows us to adjust the AMF and the vertical column result, so that the computed AMFs for that footprint reflect the true content of the atmosphere. To do this, we take an initial guess  $V_0$  for the vertical column and use it in the look-up table to return our first calculations of the AMFs to ground and cloud-top. A new estimate of the vertical column  $V_1$  is then obtained using equation (18). This process is repeated until the iteration converges. The result is then  $V_n$  if  $|1 - (V_n/V_{n-1})| < \epsilon$

after the  $n^{\text{th}}$  iteration ( $\epsilon$  is a pre-set convergence criterion). Note that the ghost column will be iterated automatically with this climatology (the ghost column height is the difference between the ground and cloud-top heights).

$V_0$  may be taken from a nearest neighbour Level 2 result, or from climatology; the TOMS zonal mean  $O_3$  total column climatology [U30] is a suitable candidate for the latter option. In most cases the first guess  $V_1$  is sufficient, since for regular  $O_3$  scenarios, AMFs are insensitive to profile shape and column amount. In these cases iteration is not necessary. However for long solar paths and/or depleted  $O_3$  scenarios, where initial guess AMF values may be incorrect, the iteration will act as a quality control on the AMFs. The look-up table enables AMFs to be repeatedly extracted quickly from database, without the need for lengthy RT simulations. It has been found in practice that a few iterations are sufficient for a convergence criterion of  $\epsilon = 10^{-4}$ . Despite the repeated extraction from look-up tables implicit in this iteration, the procedure is many times faster than the original GDP scheme for AMF and VCD computation. Work is now underway to validate this approach.

A similar iteration is possible for the  $NO_2$  vertical column, using the suggested climatology mentioned above. Local uncertainties in the tropospheric  $NO_2$  burden could be accounted for in this method (work under way).

Finally for  $O_3$  vertical columns, we derive the total error on the VCD including AMF uncertainties from the SAGE II/balloon climatology. In equation (18), we assume that the errors on fitted parameters  $E$  (slant column) and  $F_c$  (fractional cloud cover) are independent of each other, and also independent from errors due to  $O_3$  profile choice. From the Gaussian error propagation law, we have:

$$\epsilon_{total} = \sqrt{\left(\frac{\partial V}{\partial E}\right)^2 (\delta E)^2 + \left(\frac{\partial V}{\partial F_c}\right)^2 (\delta F_c)^2 + \sum_{k,m} \left(\frac{\partial V}{\partial U_k}\right) C_{km} \left(\frac{\partial V}{\partial U_m}\right)} \quad (19)$$

The matrix  $C_{ij}$  is the error covariance matrix present in the SAGE-II/balloon climatology. The climatology is parameterised in terms of a number of Umkehr column amounts  $U_i$ . The partial derivatives may be found from the definition in equation (18). In particular the derivatives  $\partial V / \partial U_i$  depend on the AMF and ghost column derivatives  $\partial A / \partial U_i$  and  $\partial(GVC) / \partial U_i$ . The ghost column is a linear combination of the database Umkehr columns, so its derivative is easy to compute. A note on the AMF derivatives is given in section 5.1 in the summary of the AMF look-up table construction.

So far in GDP, AMF uncertainties have been ignored in computing the error on the vertical column density result (that is, only the first two terms in the right hand side of equation (19) are computed in GDP).

### 3.2.8 Operational Considerations

Because of the similar measurement capability in the UV/visible, and the similarity of the algorithms, operational experience gained with GDP using the DOAS/AMF algorithm is invaluable in the implementation and testing of the SCIAMACHY UVAS algorithms. For example, the use of an *ab initio* single scatter model and a look-up table for the multiple scatter correction in GDP has

proved hard to interpret. From the start with SCIAMACHY, it is intended to use updated climatologies and pre-computed look-up tables to return AMF values.

Some optimisation is possible with the spectral fitting. By re-mapping the spectra wavelength grids onto diode numbers, interpolation over integer diode numbers can proceed much more quickly using B-spline interpolation (C. Leue, GOME and SCIAMACHY Working Sessions 6, October 1997). The situation with regard to under-sampling is not yet clear for SCIAMACHY, as measurements of the channel instrument response functions are pending (preliminary high values of the channel 8 slit function half-widths have been shown to cause retrieval degradation, see section 3.3 below).

### 3.3 IR Absorption Spectroscopy (IAS)

#### 3.3.1 Introduction

The infrared absorption spectroscopy (IAS) algorithm, as applied to SCIAMACHY nadir-mode earthshine data in channels 7 and 8, is a straightforward application of non-linear least squares fitting for the retrieval of trace gas total column amounts. The wavelength ranges covered are 1940 to 2040 nm (channel 7) and 2265 to 2380 nm (channel 8), and the 5 trace species of interest are H<sub>2</sub>O, CO<sub>2</sub>, N<sub>2</sub>O, CO, and CH<sub>4</sub>. It is also possible to retrieve an equivalent reflecting height with this algorithm. The fitting depends on the selection of points from certain micro-windows highlighting specific trace gas absorbers. In channel 8, CO and N<sub>2</sub>O are drivers for the window selection, as both species possess limited numbers of absorption lines that can be usefully differentiated from the interfering species signatures (H<sub>2</sub>O and CH<sub>4</sub>). A separate window will also be allocated for CH<sub>4</sub>, even though it can be retrieved as an adjunct species from CO or N<sub>2</sub>O fitting. In the initial definition, micro-windows will be made up of (possibly non-contiguous) segments within the ranges given below in the table.

<i>SCIA Channel</i>	<i>Approximate Window (nm)</i>	<i>Absorbing Trace Species</i>	<i>Retrieved Parameters</i>
7	2030 - 2040	CO <sub>2</sub> and H <sub>2</sub> O	Equivalent height, H <sub>2</sub> O
8	2265 - 2276	H <sub>2</sub> O, N <sub>2</sub> O and CH <sub>4</sub>	N <sub>2</sub> O (driver), CH <sub>4</sub>
8	2358 - 2365	H <sub>2</sub> O, CO and CH <sub>4</sub>	CO (driver), H <sub>2</sub> O, CH <sub>4</sub>

**Table 4: Fitting windows in the IR wavelength range**

Provided the optical depth is relatively small, the simulation of back-scattered radiance can be approximated as the product of two terms. These are the transmittance term, due solely to the direct attenuation of the solar beam by trace gas absorbers; and the closure term, which includes all slowly varying spectral signatures (reflection from ground surfaces and cloud-top, aerosol extinction and scattering, Rayleigh scattering). This separation is basic to the algorithm; it is akin to the use of an additive polynomial in the UVAS algorithm. In the version of IAS described here, non-linear fitting will be used to determine the vertical column amounts (there is no separate AMF conversion of the slant column).

*An initial version of the IAS algorithm was developed for the SCIAMACHY NRT prototype [A8][A9]. The OL algorithm (described here) is essentially the same. The main difference lies with the use of look-up tables of slant path factors and LBL-computed cross sections in the NRT algorithm (OL has options to calculate these from scratch, plus a larger set of reference data).*

Following the basic theory of the algorithm (next section), we then discuss the channel 7 height application and the channel 8 trace gas applications respectively. The penultimate section is a recapitulation of the standard line-by-line calculation of trace gas cross sections required for this algorithm. The final section deals with the operational requirements for implementing this algorithm in SGP\_12, and includes a discussion on NRT aspects.

#### 3.3.2 IAS Least-squares Fitting Algorithm

Least-squares fitting involves the minimisation of the chi-square merit function (see equation (10) on page 32). As before,  $Y_{meas}(\lambda_i)$  are the SCIAMACHY earthshine reflectance values (sun-nor-

malised radiance), and the errors  $\varepsilon_{meas}(\lambda_i)$  can be taken directly from the measurement uncertainties (weighted fitting), or all set to a constant input value (un-weighted fitting). Following the closure/transmittance separation for BIAS, the simulated earthshine reflectance will be written:

$$Y_{sim}(\lambda_i) = P(\lambda_i, \underline{c}) \cdot T(\lambda_i, \underline{b}) \quad (20)$$

where  $P(\lambda_i, \underline{c})$  is the closure term, and  $T(\lambda_i, \underline{b})$  is the trace species atmospheric transmittance. The symbol  $\underline{c}$  denotes the vector of closure parameters to be fitted, while  $\underline{b}$  is the vector of trace gas parameters to be fitted. Together these constitute the set of fitting amplitudes for the IAS problem. To a good approximation, the closure term can be treated as a low order polynomial in wavelength with  $c_j$  the polynomial coefficients:

$$P(\lambda, \underline{c}) = \sum c_j (1 - \lambda/\lambda^*)^{j-1} \quad (21)$$

( $\lambda^*$  is a reference wavelength, for example the mid-point of the fitting window). Three closure parameters are normally sufficient (quadratic polynomial).

The closure term includes the effects of Rayleigh scattering (small in the infrared), aerosol scattering and extinction, surface reflectance property and the far-wing H<sub>2</sub>O continuum absorption. If the surface reflectance property (assumed Lambertian) is the only contributing factor (other properties being neglected) then the closure term is simply the constant value  $A\mu_0$  (albedo times solar zenith cosine).

$T(\lambda_i, \underline{b})$  is the convoluted atmospheric transmittance for a set of trace gas absorbers. The parameters  $\{b_m\}$ ,  $m = 1, \dots, N_{gases}$ , are simply the total column amounts normalised to the total columns in a reference atmosphere. We are actually retrieving scale factors rather than column amounts in absolute physical units [mol.cm<sup>-2</sup>]. The transmittance may be written:

$$T(\lambda_i, \underline{b}) = \left\langle \exp \left[ - \sum_{m=1}^M b_m \tau_m(\lambda_k) \right] \right\rangle_i \quad (22)$$

where the symbol  $\langle \dots \rangle_i$  denotes the instrument slit function convolution to wavelength  $\lambda_i$  from the high-resolution wavelength grid  $\{\lambda_k\}$ . In this equation,  $\tau_m(\lambda_k)$  is the slant path atmospheric optical depth for the m<sup>th</sup> trace species at wavelength  $\lambda_k$ .

The computation of high-resolution slant optical depths requires knowledge of the slant path factors computed by means of ray tracing. For this latter, a variant of the code developed for the retrieval of FIRS [N1] was used, with only the direct solar path receiving a refractive spherical-shell treatment in the nadir mode. A single height grid is selected for the transmittance computations. Path factors and effective layer temperatures and pressures (Curtis-Godson equivalent values; for more, see section 4.2 on page 57 below) will be returned by the ray-tracing package. For low solar zenith angles, the path factor is well approximated by the geometrical secant factor  $\sec \theta_0 + \sec \theta_1$  for solar and nadir line-of-sight zenith angles  $\theta_0$  and  $\theta_1$  respectively.

In the first version of IAS, the set of reference atmospheres will be taken from the well-known AFGL database [N2], which has climatological profiles of pressure, temperature and number mixing ratio for the 5 main absorbing species.

The cumulative slant path optical depth to height  $z_L$  is given by:

$$\tau_m(\lambda_k, z_L) = \sum_{l=1}^L \frac{1}{2} [\sigma_m(z_l, \lambda_k) C_m(z_l) + \sigma_m(z_{l+1}, \lambda_k) C_m(z_{l+1})] \cdot \delta z_l \cdot S_l \quad (23)$$

Here, the layer between levels  $z_l$  and  $z_{l+1}$  has thickness  $\delta z_l$ ,  $C_m(z_l)$  is the concentration profile for trace species  $m$ , the species effective slant path factor for this layer is  $S_{l,m}$ , and  $\sigma_m(z_l, \lambda_k)$  is the corresponding absorption cross section at the specified height and wavelength. The concentration profiles are derived from the climatology; the cross sections require line-by-line computations (see below). Equation (23) employs average extinction values computed from quantities expressed at the layer boundaries - an alternative is to use vertical extinction calculated at the effective (Curtis-Godson equivalent) heights of the layers.

*In the NRT prototype, all cross sections and concentrations were pre-calculated on a single reference atmosphere; this atmosphere is a 17-level distillation of the US Standard atmosphere..*

The quantity  $Y_{sim}(\lambda_i)$  is the fitting function. In non-linear least squares, we also require the kernel matrix of parameter derivatives of the fitting function (these are sometimes called the weighting functions). From the above definitions, we have:

$$\frac{\partial Y_{sim}(\lambda_i)}{\partial b_m} = P(\lambda_i, \underline{c}) \cdot \langle -\tau_m(\lambda_k) \cdot \exp \left[ - \sum_{m=1}^M b_m \tau_m(\lambda_k) \right] \rangle_i \quad (24)$$

$$\frac{\partial Y_{sim}(\lambda_i)}{\partial c_j} = \left( 1 - \frac{\lambda_i}{\lambda^*} \right)^{j-1} \cdot T(\lambda_i, \underline{b}) \quad (25)$$

for the trace gas parameter and closure parameter derivatives respectively. The first of these contains an additional convolution, to be performed after the derivative of the high-resolution transmittance had been calculated analytically.

As with the UVAS algorithm applications, a standard set of fitting diagnostics will be generated from each IAS fit. This will include the fitting amplitude variances (diagonal elements of the solution covariance matrix), the cross-correlation matrix, the root-mean-square and chi-square values, plus the goodness-of-fit statistic (weighted fits only). The number of iterations required in the non-linear process will also be retained in the output. Chi-square for this algorithm is defined in the usual way, while the other diagnostic definitions have been given already in the section on UVAS. Signal-to-noise is critical with these infrared measurements, and the fitting will always be weighted with the estimated measurement errors.

### 3.3.3 CO<sub>2</sub> height retrieval application

In channel 7 in the region 2030-2040 nm, there is relatively weak CO<sub>2</sub> and H<sub>2</sub>O absorption (other regions of channel 7 are too optically thick with respect to CO<sub>2</sub> absorption for the IAS approximation to work). We assume that the level of CO<sub>2</sub> in the atmosphere is known; this is simply a scale factor  $f$  times the concentration of the appropriate reference atmosphere. In the AFGL climatol-

ogy, the CO<sub>2</sub> reference atmospheres are uniformly mixed (at least to 70 km) with a volume-mixing ratio of 330 ppmv. The CO<sub>2</sub> scale factor would depend on geographical location and season, and an auxiliary data set should be created for its generation (based perhaps on the Mauna Loa record, see [N3]).

The main physical assumption in this IAS application is that the atmosphere has an equivalent Lambertian reflecting surface at some lower boundary height, regardless of the state of cloudiness. While this is to some extent not physical, it is the most expedient method available at the present time for dealing with cloud contamination. The unknown (assumed Lambertian) albedo is subsumed in the first coefficient of the closure polynomial, while the unknown height will be retrieved separately along with the H<sub>2</sub>O column. This process is analogous to the retrieval of a reflecting height in the revised cloud fitting algorithm from O<sub>2</sub> A band back-scatter spectra.

The equivalent Lambertian assumption makes the IAS applications stand-alone, and it precludes the use of retrieved cloud parameters from other algorithms in SDP. One could envisage using the results of the cloud pre-processing algorithms PCCA and RCFA (fractional cover and cloud-top pressure) as input to this IAS application, with the known CO<sub>2</sub> absorption then used to generate a temperature at the cloud-top. In the next section we show how the equivalent height result of this IAS application may be used in the channel 8 trace gas applications.

The free parameters to be fitted are the reflecting height  $h$  and the normalised water column  $w$ , plus the closure parameters. The fitting function is:

$$Y_{sim}(\lambda_i) = P(\lambda_i, \underline{c}) \cdot \langle \exp[-f\tau_{CO_2}(\lambda_k, h) - w\tau_{H_2O}(\lambda_k)] \rangle_i \quad (26)$$

The notation is as before; the transmittance term has been written explicitly. In the off-line version, the layering of the atmosphere is repeatedly changed to reflect the fact that the lower boundary height is a free parameter; for each guess of this height, the ray-tracing and line-by-line cumulative slant path optical depth calculations are executed. Finally we note that the transmittance is not an analytic function of lower boundary height, and the kernel derivative with respect to this height will be evaluated by finite differences.

*In the NRT prototype, the cumulative transmittances and their parameter derivatives are calculated on the lowest 7 levels of the 17-level BIAS reference grid mentioned above. For any given value of  $h$  in the fitting iteration, actual transmittances must be interpolated to  $h$  (interpolation is over the lowest 7 levels; convolution takes place before interpolation). The NRT solution is convenient and faster, but there is a loss of accuracy.*

### 3.3.4 Trace gas retrieval application

Using the same window in channel 7, one can retrieve a total CO<sub>2</sub> column if the cloud fraction and cloud-top height are given (perhaps already retrieved from the cloud pre-processing algorithms). In this case the parameter  $f$  in the above section is now to be fitted, and  $h$  in equation (26) is either the cloud-top height or the ground topography height. A combined simulated reflectance is constructed as a linear cloud-fraction-weighted combination of simulations from cloud-top and ground.

A similar approach could be adopted for channel 8 IAS retrievals, but we here describe an approach using the equivalent reflecting height  $h$ . This assumes that the equivalent Lambertian reflecting surface approximation is valid for all IAS applications; the CO<sub>2</sub> height retrieval must be performed first. The error  $\epsilon(h)$  (a diagnostic result from the channel 7 IAS application) on the fit-

ted parameter  $h$  can be treated as a ‘process’ error (sometimes called a ‘model’ or ‘system’ error) in the IAS trace gas retrievals. This combination of IAS applications allows for a degree of consistency in the near infrared nadir column retrievals from SCIAMACHY.

The basic mathematical exegesis has already been given above. In the definition of chi-square, the weighting is given by the inverse of the measurement error covariance matrix  $S_m$ . We assume that channel 8 measurement errors are independent, so that  $S_m$  has diagonal entries equal to  $\varepsilon(\lambda_j)^2$ , for  $j = 1, \dots, N$  (number of points in fitting window). When  $\varepsilon(h)$  is included as a process error, the measurement error covariance matrix is replaced by:

$$(S_m)'_{ij} = (S_m)_{ij} + \frac{\partial Y_i}{\partial h} \cdot \frac{\partial Y_j}{\partial h} \cdot \varepsilon(h)^2 \quad (27)$$

The derivatives with respect to height are defined as in the previous section, and will be done using finite differences.

It has been noted that CO and N<sub>2</sub>O are important drivers for window selection. In the Phase A sensitivity study, it was shown that for CO, only three lines (P7, P8 and P9 from the first vibrational overtone  $2 \leftarrow 0$ ) would produce a signature detectable by SCIAMACHY [A2]. The IAS algorithm requires that regions of significant optical depths from interfering species other than CO must be masked in the selection of the fitting points. One must be careful to mask out strong H<sub>2</sub>O and CH<sub>4</sub> lines from the fitting windows - hence the requirement for non-contiguous segments. For N<sub>2</sub>O, the P19, P15, P13 and P12 lines of the  $00^0_2 \leftarrow 00^0_0$  transition are recommended for micro-window selection.

There are a number of possibilities for selecting a window for CH<sub>4</sub> retrieval; a determination has not yet been made, but it should be straightforward to select a one-segment window in channel 8 for this absorber (large number of lines spread over the channel).

In all applications, it has been assumed that the shape of the slit function is known, and that the convolutions would be carried out with a known instrument response half-width. One can adapt the IAS algorithm to fit this half-width as a free parameter in a manner similar to the direct fitting method described above as an option in the UVAS algorithm; this has not yet been attempted.

However, in response to concern about the effect of instrument slit function width on the accuracy of CO and N<sub>2</sub>O retrieval, sensitivity studies were carried out to test this effect [N4][N5]. These studies were performed with SCIAMACHY measurements simulated using LBL code [see next section] and the forward model GOMETRAN, plus the SCIAMACHY Instrument Simulation Software Package [N6]. IAS fits were made for a number of solar zenith angles, with convolution using Gaussian response functions of various half-widths. Results are shown in the table below for three values of the slit function full width half maximum (in pixel numbers). It is clear that, in general, the CO precision is reduced by a factor of 2, but the N<sub>2</sub>O retrieval is seriously compromised by wide slit widths (poorly resolved optics). The N<sub>2</sub>O result for FWHM 4.7 pixels is particularly poor, but is not surprising given that the useable P branch lines (mentioned above) are typically separated by  $\sim 0.52$  nm, which is about 4.7 pixels at the nominal SCIAMACHY spectral resolution for this channel.

SZA	CO in Window			N <sub>2</sub> O in Window		
	% error for FWHM 2.0	% error for FWHM 3.3	% error for FWHM 4.7	% error for FWHM 2.0	% error for FWHM 3.3	% error for FWHM 4.7
16	7.39	10.28	13.95	2.21	5.39	11.94
30	7.85	10.92	14.82	2.33	5.69	12.57
45	8.80	12.25	16.59	2.60	6.31	13.88
60	10.63	14.77	19.94	3.12	7.47	16.21
70	13.01	18.01	24.26	3.81	8.97	19.02
75	17.77	23.66	30.62	4.45	10.32	21.47
80	22.46	29.81	38.39	5.71	12.94	26.13

Table 5: Accuracy of CO and N<sub>2</sub>O retrievals for different slit width

### 3.3.5 Line-by-line cross sections

The standard result for the computation of the line absorption coefficient for a single species at a given temperature  $T$ , pressure  $P$  and wavenumber  $\nu$  is:

$$\alpha(\nu, P, T) = \sum_i S_i \cdot \frac{Q_{rot}(T_0)}{Q_{rot}(T)} \cdot \frac{Q_{vib}(T_0)}{Q_{vib}(T)} \cdot \exp\left(\frac{E_i}{kT_0} - \frac{E_i}{kT}\right) \cdot \frac{1 - \exp(-hc\nu_i/kT)}{1 - \exp(-hc\nu_i/kT_0)} \cdot V_i(\nu, P, T) \quad (28)$$

Here the sum is over all contributing line transitions.  $h$  is Planck's constant,  $k$  is Boltzmann's constant,  $c$  the speed of light. The ratio of rotational partition functions is to a good approximation given by  $Q_{rot}(T)/Q_{rot}(T_0) \sim (T_0/T)^n$ . For CO, CO<sub>2</sub> and N<sub>2</sub>O, the exponent  $n$  is 1.0; for CH<sub>4</sub> and H<sub>2</sub>O,  $n = 1.5$ . The vibration partition functions require explicit computation using the wavenumbers  $\omega_m$  and degeneracy values  $d_m$  of the vibrational modes [L6]:

$$Q_{vib}(T) = \prod_{i=1}^N \frac{1}{[1 - \exp(-hc\omega_i/kT)]^{d_i}} \quad (29)$$

In equation (28),  $E_i$  are the lower state energies,  $\nu_i$  the line positions and  $S_i$  the line intensities measured at a reference temperature  $T_0$ . The profile shape function  $V_i$  depends on the distance  $\nu - \nu_i$  from the line position, the Doppler half-width  $\gamma_{i,D}(P, T)$  and the air-broadened half-width  $\gamma_{i,L}(P, T)$ .  $\gamma_{i,L}$  is specified in the database at a reference temperature and pressure  $P_0, T_0$ , and a temperature-dependence exponent  $\beta_i$  gives values at other temperatures via the following:

$$\gamma_{i,L}(P, T) = \gamma_{i,L}(P_0, T_0) \cdot (P/P_0) \cdot (T/T_0)^{\beta_i} \quad (30)$$

The set  $\{E_i, \nu_i, S_i, \gamma_{i,L}(P_0, T_0), \beta_i\}$  are taken from the HITRAN96 data base [N7], with  $P_0 = 1$  atmosphere and  $T_0 = 296K$ . The shape function is the Voigt profile, defined by:

$$K(x, y) = \frac{y}{\pi} \int_{-\infty}^{\infty} \frac{e^{-t^2}}{(x-t)^2 + y^2} dt \quad \text{where } x = \sqrt{\ln 2} \cdot \frac{v - v_i}{\gamma_{i,D}} \quad \text{and } y = \sqrt{\ln 2} \cdot \frac{\gamma_{i,L}}{\gamma_{i,D}} \quad (31)$$

A number of standard packages are available for the numerical computation of this integral. So far, the IAS prototype algorithms have tested three implementations of the Voigt function [N8][N9]. Recent work [N10][N11] has shown how line profile calculations may be optimised for speed. In particular the modules prepared for the MIPAS project (also on ENVISAT-1) should be available for use in the SCIAMACHY off-line operational algorithms. Finally, we note that the LBL computation of cross sections also applies to transmittance calculations in the limb stratospheric profile retrievals, and the formulae mentioned above are valid also in section 4 on page 54.

*The NRT algorithm will use only look-up tables of calculated according to the above, so these performance considerations do not apply.*

### 3.3.6 Operational Considerations

This algorithm is less complicated to implement and maintain in an operational environment than the UVAS package. The range of databases is smaller, and the calculation of simulated back-scatter somewhat simpler. However there are a number of crucial performance considerations.

The most important is the line-shape profile computation. As noted above, there has been recent progress in optimising this task; the main idea is to minimise the number of calls to a Voigt function module by using interpolation wherever possible, and by pre-computing as many quantities as possible to avoid repetition. Though this complicates the line-shape determination, the saving in run time is impressive [N12].

Another possible time saver in these algorithms involving LBL computations lies in the pre-computation of convoluted quantities. If the wavelength calibration is known beforehand, then all normalised slit function values can be pre-computed, thus reducing the convolution integrals to simple summations. This consideration also applies to the cloud fitting algorithms and the limb retrieval algorithms. Such a procedure assumes that the slit function shape and half-width are fixed values known in advance. *This time-saving applies also to the NRT algorithm but has not so far been considered for the specifications.*

Further improvements in speed can be obtained by optimising the ray-tracing (a) only to operate at sufficiently high solar zenith angles where the plane parallel approximation is invalid; and (b) to have all climatological pressure, temperature and refractive indices pre-calculated on optimised fine-level height grids. *This does not apply to the NRT algorithm, where slant path factors are pre-computed in a look-up table.*

Another improvement can be gained by reducing the number of iterations in the non-linear least-squares fitting loop. This depends on the first guess of the set of fitting parameters; a very significant saving can be achieved if the first guess is closer to the final fitted set of values in parameter space. This would suggest that those nearest neighbour results (in space and time) from previous IAS applications should be used as first guess values.

Complete system tests on the IAS applications will require the simulation of SCIAMACHY level 1b data. For this task, the radiative transfer model GOMETRAN (see section 5 on page 75) will be used in conjunction with LBL code to generate back-scatter reflectance values, and the SCIAMACHY Instrument Simulation Software Package will generate measurement error values. One

other option that will be attempted is to merge all applications of IAS in a kind of global fit with three or four windows, for all trace species amounts and the equivalent reflecting height.

### 3.4 O<sub>3</sub> Profile Retrieval

#### 3.4.1 Introduction - Optimal Estimation

This first section is a general introduction to the mathematical technique of optimal estimation. The application of this technique to the particular problem of height-resolved retrieval from SCIAMACHY (and GOME) nadir measurements is the subject of the rest of this chapter. Optimal estimation will be used again in the limb retrievals. Suitable general reviews can be found in [P1][P2], and in the context of GOME studies, the SERCO/ESA Trace Gas Study Final Report [G5] and its follow-up [G6].

Given a vector of measurements  $y$  with noise covariance  $S_y$ , how do we find the best estimate of the true state vector  $x$  (of atmospheric constituents), knowing that  $x$  and  $y$  are related by a forward model  $F$ , that is  $y = F(x)$ ? We assume that the solution of the inverse problem is constrained by the existence of an independent *a priori* state vector  $x_a$  with error covariance  $S_a$ . The best estimate for the inverse  $x = F^{-1}(y)$  is the one that maximises the conditional probability  $P(x|y, x_a)$ ; if the statistics are Gaussian, then this is equivalent to minimising the following functional:

$$\|Y_{meas} - F(x)\|_{S_y} + \|(x - x_a)\|_{S_a^{-1}} \quad (32)$$

If we linearise the forward model about the state  $x_n$ , and write  $K_n = (\partial F / \partial x)$  for the matrix of forward model weighting functions evaluated at  $x = x_n$ , then the minimisation yields the following estimate for the next guess  $x_{n+1}$ :

$$x_{n+1} = x_a + G^{-1} \cdot K_n^T \cdot S_y^{-1} \cdot [(Y - F(x_n)) - K_n(x_a - x_n)] \quad (33)$$

where:

$$G = K_n^T \cdot S_y^{-1} \cdot K_n + S_a^{-1} \quad (34)$$

Here, the symbol  $(\dots)^T$  denotes the matrix transpose. A test for convergence may be made by examining the root mean square of the deviation of the state vector from one iteration to the next. Alternatively one can examine the convergence of the so-called cost function (sum of weighted deviations of state vector  $x$  from a priori  $x_a$  and measurement vector  $y$  from simulation  $F(x)$ ) [P3][P4].

Upon convergence, the matrix  $G$  in equation (34) is recalculated using the last estimate of the state vector. The resulting inverse  $G^{-1}$  is the solution variance-covariance matrix, from which parameter errors and the cross-correlation matrix can be derived. Other diagnostics from the fitting are the RMS values of the residuals, the cost function itself, and the averaging kernel and contribution functions. The averaging kernel represents the response of the system to a delta-function impulse of the true state [P2]; widths of the peaked rows and columns of the averaging kernel are a measure of the resolution of the retrieval.

Besides the *a priori* and measurement noise, errors due to uncertainties in the model parameters can also be worked into the formulation. If the vector of model parameters is  $\underline{b}$  with error covari-

ance  $S_b$ , then it is necessary to replace  $S_y$  by  $S_m = S_y + K_b S_b K_b^T$  in the above formulation to include this additional source of uncertainty [ $K_b = (\partial F / \partial b)$  in this definition].

### 3.4.2 Height-Resolved Fitting Algorithm (HRFA)

The retrieval of ozone in the stratosphere has been established with the SBUV instrument [P5]. The method is based on the variation of effective back-scatter height with wavelength in the Hartley and short-wavelength Huggins absorption bands. This is the so-called UV technique [P6], which allows profiles to be established down to peak ozone concentration levels. SBUV has only 11 wavelengths; with higher resolution and wider spectral coverage, GOME and SCIAMACHY should be able to extend the retrieval, most importantly to the troposphere, where the temperature dependency of the Huggins bands absorption should yield additional information. This is indeed the case, as was demonstrated in the initial study for GOME. This study used a non-linear least squares fitting algorithm to examine the profile retrieval, and also proposed optimal estimation for operational use [P7].

The measurement vector  $Y_{meas}$  of back-scatter reflectances is constructed by taking the ratio of the level 1b earthshine radiance to the extra-terrestrial solar data. Although both level 1 spectra are instrument-calibrated in the level 0 to 1b processing step, it has been found necessary to improve the wavelength calibration before this ratio is taken. One possibility is to fit for a shift between the two spectra. Another way is to match the spectra against an accurate Fraunhofer standard [U7], allowing for an independent determination of the instrument response function across the chosen fitting window (this is analogous to the procedure in section 3.2 for direct UV/visible fitting).

The most important elements of the state vector are the ozone profile concentration values at selected heights on the retrieval grid; it is customary to retrieve volume mixing ratios. Studies carried out for GOME have shown that it is necessary to include a number of additional parameters in the state vector [G5]. These include a parameter for the Ring effect, wavelength registration shifts of the ozone cross section reference and the (re-calibrated) solar level 1 spectrum against the (re-calibrated) radiance spectrum, total column amount of  $\text{NO}_2$ , a surface albedo, and (sometimes) total aerosol and molecular optical depths.

One of the forward models of choice is GOMETRAN [R4]; this has the multiple scatter formalism necessary for intensity calculations in the UV and visible, and it has the additional capability to generate all required weighting function information simultaneously with the simulated back-scatter intensity [R7]. MODTRAN version 3.5 and above [R10] is also suitable for forward model simulation (it contains the DISORT multiple scatter formulation [R11]). More information can be found in section 5 on page 75. In all studies so far, Ring reference spectra have been derived by model simulations [U7][R5]. A temperature shift parameter has been included in the list of parameters to be fitted in some of the investigative work [G6], as it is now possible to define weighting functions in GOMETRAN with respect to this parameter (R. de Beek, GOME and SCIAMACHY Working Session 6, Brussels, October 1997).

The choice of an *a priori* state vector and associated error covariance is important. For GOME, the MPI model-output database has been used with a specially constructed error covariance [G6]. Another source is the stratospheric ozone profile climatology based on SAGE II data [P8]; this is a true climatology based on real measurements and their uncertainties. For the troposphere levels, it is only necessary to impose a loose constraint on the *a priori* values. *A priori* and starting values for temperatures,  $\text{NO}_2$  columns, aerosol and air density profiles are usually taken from climatol-

ogy, with suitable *a priori* covariance chosen on an *ad hoc* basis. A starting value for the albedo can be taken from a global data set, but it is better to use a value from the measurement data - the earthshine albedo at 400 nm is a suitable candidate (lower Rayleigh scattering, very low trace gas absorption). This value can also be used as the albedo *a priori* though this is not strictly necessary.

The retrieval strategy depends not only on the choice of state vector parameters but also on the fitting window definition. For GOME, 3 separate channels cover the range of ozone absorption, with band 1A (237 - 307 nm) covering the Hartley band (240 - 300 nm), and band 2B (312 - 405 nm) the Huggins bands (310 - 340 nm). Bands 3 and 4 cover the Chappuis region. Because of low signal-to-noise, band 1A is normally read every 12 seconds, as opposed to 1.5 seconds (or 0.375 seconds before co-adding) for bands 2B, 3 and 4. Band 1A footprints cover the whole swath (normally 960 km width), and include many individual band 2B scenes. It is apparent that under normal scanning operations, band 1A and 2B data cannot readily be included in the same measurement vector. Not only do the reflectances not cover coincident scenes, but also there are some real problems associated with the representative and accurate simulation of back-scattered light from such mixed-scene footprints.

One strategy is to adopt a two-step approach [P9][P10][P3]. The first stage fits an ozone profile in the stratosphere using 1A measurement data only from the Hartley bands. For this, a window between 260 and 300 nm is suitable; there is a mask to take out NO  $\gamma$ -band emission lines [P11] corrupting the spectra. Rayleigh single scattering is dominant at these wavelengths, and it is possible to greatly reduce the computation time taken over simulations by turning off the multiple scatter formalism in the forward model. An average intensity for the large footprint scene requires a number of simulations for various viewing geometries. However, the single scatter assumption is not really valid for  $\lambda > 295$  nm, and valuable spectral information may be lost if these wavelengths are omitted.

The second step involves an optimal estimation fit in the Huggins bands (band 2B data) using the 1A retrieved profile as *a priori*. This places a reasonably tight constraint on the *a priori* for the second fit; the tropospheric *a priori* on the other hand is loosely constrained. GOME band 2 earthshine spectra at wavelengths below 320 nm are not considered reliable for inclusion in the fit (straylight corruption). Besides a contiguous region inside the prominent part of the Huggins bands ( $< 340$  nm), it is also possible to include points from band 3 in the Chappuis - this may provide a little more information for the troposphere. Unlike the band 1A retrieval, the surface albedo and NO<sub>2</sub> column parameters should be included in the state vector.

Results for selected cloud-free scenarios with O<sub>3</sub> sonde validation data have shown that the tropospheric ozone profile retrieval is reliable and reasonably accurate. A typical averaging kernel for a combined 1A/2B two-stage GOME-retrieved profile has narrow peaks in the stratosphere (half-widths less than 5 km), and broader but still well-defined tropospheric peaks. There is a clear advantage in using a single fitting step, particularly if one can use a contiguous and spatially consistent set of measurements covering all required UV wavelengths.

Recently a change was made to the virtual pixel boundary in GOME channel 1A; this boundary has been moved from 307 nm to 282 nm. This allows 1B data from 282 - 307 nm to be read out alongside the other (short-readout) bands. One could then define a single window with points covering both the Hartley and Huggins bands, thereby obviating the need for the two-step retrieval. Despite the loss of signal concomitant with the reduced band 1B integration time, it is thought that the rich information present in the spectrum below 307 nm is enough to secure the accurate retrieval of the stratospheric part of the profile. The spatial coverage would be the same.

The effect on retrieval accuracy is currently being investigated. A similar proposal is now on the table for SCIAMACHY.

Investigation has started on the treatment of clouds in this retrieval. The simplest idea for partially cloudy scenes is to assume a linear combination of cloud-top and surface intensities in the simulation of the total reflectance. This is similar to the treatment of cloud-contaminated scenes in the total column differential absorption fitting (see section 3.2 on page 31); it assumes atmospheric homogeneity across the footprint and all clouds at the same height. This method is also relatively simple to implement in an operational environment. It is also possible to carry out the retrieval assuming the clouds as layers, using a version of GOMETRAN with a proper parameterisation of clouds [R5]; this investigation has just started.

All four groups working on this type of profile retrieval have shown that partial orbits (or even whole orbits) of GOME data may be processed to generate substantial amounts of profile information. However, such tasks demand many hours of computer run time (10 - 20 minutes per pixel is typical), and we now address the performance issue for this algorithm.

### 3.4.3 Operational Considerations

For SCIAMACHY and for GOME, the major bottleneck for the operational implementation of this algorithm has been the performance time. The basic problem is the time required for numerous multiple scatter simulations of the back-scatter reflectance and their weighting functions, for a considerable number of wavelengths, for many viewing geometries, and for a number of iterations in the optimal estimation process. To date, no suitable parameterisation scheme has been devised for the construction of reliable and comprehensive look-up tables of simulated variables. The performance improvement issue is now being addressed for GOME. The intention is to reach a retrieval throughput of 10% of the data acquisition rate in 1998, this figure hopefully rising to 100% within a 3-year time frame with further increases in computing power (B. Kerridge, private communication).

There are some other issues regarding operational usage. Though it is possible to shift the temperature by a single number at all levels, studies have shown that the O<sub>3</sub> profile precision is sensitive to temperature uncertainty [G6]. Model output climatology is probably not accurate enough; ideally one should use assimilated data from an NWP program such as that run by ECMWF [P12]. This would require an additional input to the retrieval to be generated by an independent processor. ECMWF data files will be ingested into the Envisat PDS ground segment on a regular basis, and will thus be available for SCIAMACHY data processing. The data will contain 4 analysis files and 4 forecast files, each file of size 8.7 Mbytes, ingested every 6 hours; the analysis fields will be used for the off-line algorithm data processing as described in the present document (NRT will use forecast fields).

Another suggestion has been the use of nearest neighbour previously-retrieved profiles. While this would in principle greatly reduce *a priori* errors and thereby cut down on the number of iterations in the optimal estimation, there is a risk of a "forward time bias" creeping into the results.

A lot of work has already been done to show the feasibility of this technique, and the results obtained so far mark a significant advance in our ability to measure tropospheric ozone profiles. More work is required to ensure that the algorithm will be capable of dealing with all geophysical scenarios likely to be encountered by these instruments.

## 4 Limb and Occultation Retrieval Algorithms

### 4.1 Introduction and Retrieval Strategy

Before going into details on SCIAMACHY limb and occultation retrievals, we give a general discussion on retrieval strategy. This will set the scene for the detailed retrieval algorithms to follow. Much of the nadir retrieval for SCIAMACHY is based on the experience gained with GOME; limb/occultation retrieval is a new departure. At the time of writing the present issue, algorithm development for limb is still in an early phase and some of the descriptions here are therefore not as comprehensive as those in section 3.

Limb stratospheric profiles for pressure, temperature and trace gas volume mixing ratios have been retrieved routinely from atmospheric emission spectra in the infrared (and beyond to the microwave regions). A typical example is the MIPAS instrument [L1], which will fly alongside SCIAMACHY as part of the ENVISAT-1 payload; the list of target species includes O<sub>3</sub>, H<sub>2</sub>O, CH<sub>4</sub>, N<sub>2</sub>O and HNO<sub>3</sub>. In contrast, SCIAMACHY limb profiles will be derived chiefly from backscatter spectra in the UV/visible and near-infrared. Although there have been a number of previous limb scatter experiments (see for example the literature on SME [L17] and the balloon-borne NO<sub>2</sub> measurements [L18]), there are a number of new passive remote sensing applications. Besides SCIAMACHY, NASAs SOLSE/LORE experiments (demonstration instruments) have flown recently on a Shuttle mission [L2], and the OSIRIS instrument on ODIN [L3] is scheduled for launch in the near future.

It has been pointed out that at sufficiently high altitudes, fluorescent and chemi-luminescent molecular emission will be the main light source for limb radiances. In this document, we shall be concerned with limb measurements in the troposphere and low/middle stratosphere, for which scattering is the dominant source, and emission effects (including thermal) can be neglected.

As with other remote sensing applications for SCIAMACHY, the limb retrieval will involve least-squares minimisation of chi-square merit functions. Normally the simulated quantities are not linear in the retrieval parameters; the fitting then uses either a standard iterative non-linear least-squares algorithm [U4] or the commonly used optimal estimation method [P1]. These fitting algorithms will be standard for SCIAMACHY. The modified Chahine inversion technique [L4] has been used particularly in occultation retrieval, but so far is not being considered for SCIAMACHY.

Limb retrieval has traditionally followed the onion peeling strategy [L5]. Here, each limb scan is fitted individually, starting with the top-most scan, and working down to the lowest tangent height in the sequence of limb scans. At any given level, retrieved profile values from all the previously-fitted higher levels must be included in the fit; it is important to include parameter error estimates propagating downwards from the higher levels. Onion peeling is convenient and relatively fast. As it is normally used, layers below the current scan do not contribute, so the weighting functions are truncated below this level. This method has been used successfully for a number of balloon-borne instruments [L6][L7].

The more sophisticated global fit strategy [L8] requires the simultaneous fit of all limb scan measurements in a sequence; entire atmospheric profiles are fitted in one operation. This is a more consistent approach, allowing for a more accurate rendition of the simulated spectra and weighting function matrix. However it is more cumbersome and harder to interpret, and often imposes an excessive computational burden. One variant that has been suggested is to carry out a post-processing global fit on a set of onion peeling results (K. Chance, private communication, 1997).

The global fit approach has been taken as standard for MIPAS on ENVISAT [L9]. It is the preferred method for simultaneous P-T retrieval (see for example SAGE III algorithm description [L10]), and has been used for SOLSE/LORE and OSIRIS O<sub>3</sub> limb profile retrieval from UV/visible measurements.

For SCIAMACHY, limb back-scatter intensities will contain diffuse contributions from multiply scattered atmospheric and surface-reflected light. This mitigates against the onion peeling technique (unless climatological contributions from below the tangent layer are included). Multiply scattered light contributes more strongly in the UV/visible than in the infrared. It is possible also to compensate for surface reflected light by using ratios of two limb scans as the measurements instead of the actual limb radiance.

If the limb instrument has a P-T retrieval capability, it is important to ensure consistency in the use of pressure and temperature information in the trace gas retrievals. This is certainly the case for SCIAMACHY, which has two P-T capabilities (O<sub>2</sub> and CO<sub>2</sub> absorption). Thus the requirement for SCIAMACHY will be to perform P-T retrieval alongside the various targeted trace gas retrievals. There are some concerns in SCIAMACHY regarding the pointing accuracy in limb [A5], and it is thought that tangent height (pressure) levels derived from limb fitting will be more accurate than the ephemeris-derived values.

All limb trace gas products will be given in VMR and number density units, the latter derived using temperature and pressure values retrieved concomitantly with the trace gas VMR retrievals. Regarding the retrieval of aerosol parameters, at this stage it is not clear how much information on aerosols can be extracted. For an initial baseline, we will assume that aerosol scattering and extinction coefficients can be retrieved for each limb scan, and for all limb applications. These variables are the simplest to use in the retrieval; an alternative might be layer optical thickness and single scatter albedo. It has been decided not to derive additional aerosol properties dependent on particle size distribution (for example, the mode radius).

A word on geolocation is in order. It is important to have geolocation quantities clearly defined for limb retrieval. The top-of-the-atmosphere height (TOA) will be defined as the atmospheric height above the surface at the point of entry of the line of sight beam. The actual value of this TOA will depend on the instruments capability, but a baseline needs to be agreed on by members of the SSAG (need not be the same as that defined for the nadir algorithms). The solar zenith angle should also be defined at the TOA height. A recommendation was made to include two values of the tangent height in the geolocation, one calculated by pure geometry, the other with an additional refraction calculation, and this will be adopted. Derivation of geolocation quantities must follow the methods described in the Mission Conventions Document [A7].

The above discussion refers to limb strategy. In occultation viewing, the instrument detects direct sunlight attenuated along an oblique atmospheric track; atmospheric scatter is not a significant source of light. Simulation is restricted to absorption (transmittance) along the line of sight. Space-borne occultation instruments have looked at a number of spectral regions, for example the FTS ATMOS instrument in the middle infrared [L11], HALOE in the visible [U29], and the various SAGE missions (visible, near infrared) [L12][L10]. For trace species emission retrieval in the infrared, the onion peeling approach has been satisfactory. P-T retrieval has been attempted using both retrieval strategies [L13]. A different approach is used in the SAGE II and III inversion algorithms [L12] whereby an initial fitting for slant column amounts or slant optical depths is performed using differential absorption and multi-linear regression, and this is followed by a non-linear inversion for the vertical profile values. The list of products retrieved from occultation is similar to that from limb retrievals (see table in section 2.3).

---

We now discuss the main applications for SCIAMACHY. First, the pressure and temperature retrieval from limb scans, second, the trace species limb profile retrievals, and finally a description of the occultation algorithms.

## 4.2 Limb P-T Retrieval

The motivation is to assign a pressure-temperature profile and a tangent height to each limb scan measurement, so that trace gas profiles can be retrieved with a minimum of uncertainty in the atmospheric information (in this regard, climatological atmospheric data are often insufficiently accurate). The derivation of P and T profiles brings an important degree of self-consistency to the whole limb retrieval process.

The distributions of O<sub>2</sub> and CO<sub>2</sub> are assumed known; the dependence of their absorption features on pressure and temperature provides the key to the retrieval. We confine ourselves here to a discussion of P-T retrieval from CO<sub>2</sub> absorption (channel 7), but most aspects of the discussion also hold for P-T retrieval using channel 4 data around the O<sub>2</sub> A band.

There are a number of pressure- and temperature-sensitive CO<sub>2</sub> absorption lines in channel 7. Once a P-T profile has been obtained, a height assignment can be made by assuming one absolute pressure-height correspondence and hydrostatic equilibrium for the rest of the atmosphere. If such a correspondence is absent, then the whole retrieval could be repeated with an improved hydrostatic reference point until convergence is reached.

Assuming for a single P-T retrieval that a global fit analysis is in force, the state vector of fitting parameters will consist of P-T pairs, one for each limb scan included in the fit, plus a number of additional parameters. The latter may include H<sub>2</sub>O mixing ratio values (if this gas is an interfering species), plus closure parameters for the continuum level of the spectra. For O<sub>2</sub> A band P-T retrieval, O<sub>3</sub> and (possibly) NO<sub>2</sub> are interfering species.

The vector of observations will be made up of a number of micro-window segments, each segment containing a small number of observations around a prominent CO<sub>2</sub> line or a small set of lines selected for either pressure sensitivity or temperature sensitivity. Interference from H<sub>2</sub>O lines cannot be ruled out with the sort of resolution SCIAMACHY possesses (it will probably be necessary to fold in some H<sub>2</sub>O line optical depths in the instrument response function convolution). A study needs to be carried out to optimise the choice of lines.

The assumed profile of CO<sub>2</sub> mixing ratios is also important here. It is enough to assume that CO<sub>2</sub> is uniformly well mixed up to the mesopause, and that there are no non-LTE effects below this altitude. In any case, signal-to-noise ratios for scattered light from limb scans above 35 km may be too small for useful information retrieval above this altitude (see [A5] for instrument performance simulations). The value of the volume mixing ratio might be derived from measurements in the stratosphere (e.g. [L14]) updated to the current year, and with some allowance for the seasonal level (photosynthesis cycle). The O<sub>2</sub> mixing ratio is usually taken as 20.95%.

The forward model part of the algorithm will require the simulation of limb-scattered radiance values. Formulae for the line-by-line computation of absorption cross sections have already been specified (see section 3.3). Limb scatter requires ray tracing for both the solar and line-of-sight paths, and we shall use the following as the definition of transmittance for a ray with path coordinate  $s$ :

$$\tau(\nu, s) = \exp \left[ - \int_{s_0}^s \bar{k}(\nu, s') \eta(s') ds' \right] \quad (35)$$

Here  $\eta(s)$  is the number density of air, and the weighted cross section  $\bar{k}(v, s)$  is the sum of individual cross sections  $\alpha_m(v, s)$  multiplied by trace species mixing ratios  $\chi_m(s)$ . In this application  $\chi_{CO_2}(s)$  is assumed known and constant, and the transmittance will depend on  $P$  and  $T$  through the number density and the cross section dependence.

With onion peeling, we would be restricted to using the pressure heights as retrieved successively by the algorithm, and this would define the vertical structure of the atmosphere to be used in the simulations. With global fitting, we can choose any height grid we want for the retrieval, though it is more convenient to use a grid suggested by the data; this issue has not been resolved yet. It is customary to use a finer grid for the simulation, to account for variations of temperature and pressure across the various layers. Rather than use mean values (requiring a finer structure), one can use a coarser height grid and Curtis-Godson (CG) equivalent pressure and temperature values. The CG temperature is defined for each trace species  $m$  and for each layer  $L$  as follows:

$$T_{m,L}^{eff} = \frac{\int_{z_{L-1}}^{z_L} T(z) \cdot \chi_m(z) \cdot \eta[p(z), T(z)] \cdot \frac{ds}{dz} \cdot dz}{\int_{z_{L-1}}^{z_L} \chi_m(z) \cdot \eta[p(z), T(z)] \cdot \frac{ds}{dz} \cdot dz} \quad (36)$$

Here  $\eta$  and  $\chi$  are defined as above, and  $z$  is the local vertical co-ordinate; the effective pressure is defined in a similar fashion. For P-T retrieval we assume also that the atmosphere is horizontally stratified, and obeys the hydrostatic equation  $dp/dz = -K\eta(z)g(z)$ , where  $K$  is a constant and  $g(z)$  is the acceleration due to gravity. The variation of  $g(z)$  with height may be modelled along the lines suggested in [L9], and this is explained in more detail in one of the Appendices.

The retrieval requires a set of initial values, and for P and T these may come from climatology (e.g. [G4]) or from an additional data set of assimilated P-T fields generated by a numerical weather prediction model [P12]. Initial values of tangent heights are taken from the geolocation information ephemeris results. It is of course possible to constrain the retrieval with these starting values used as *a priori*, but experience has shown that this is not advisable (at least as far as the tangent height information is concerned). As a consequence, the preferred fitting for P-T retrieval from limb measurements has been non-linear least squares rather than optimal estimation.

Elements of the Jacobean matrix (weighting functions) should be derived analytically (as far as this is possible). For the single scatter formalism, analytic expressions derived for the transmittances and scattered fluxes may be differentiated with respect to P and T values for each layer, to generate parameter derivatives. This means that the T and P dependence must be followed through the definitions of air density and cross section, the latter including P- and T-dependence in the Voigt profile shape, and T-dependence in the line strength. Voigt profile derivatives may be found from the properties of the complex error function (of which  $K(x, y)$  in equation (31) is the real part; see [N9] for definitions).

## 4.3 Limb Trace Gas VMR Retrieval

### 4.3.1 Introduction

A primary application is the derivation of  $O_3$  profiles from limb measurements in the UV and visible parts of the spectrum. A number of other trace gas limb profile applications have been considered for the UV/visible, with the possibility to retrieve  $NO_2$  and  $BrO$  profiles on a global basis, and  $SO_2$ ,  $OCIO$  and  $H_2CO$  in special scenarios. In all cases, Rayleigh scattering and aerosol scattering and extinction must be accounted for, and in particular the retrieval of aerosol parameters will be undertaken concurrently. In the infrared,  $H_2O$  in Channel 5,  $CO_2$  in channel 7, and  $CH_4$ ,  $CO$  and  $N_2O$  in Channel 8 are the target species. For these species, the line-by-line computation of cross sections is required, but this need only be done once in the fitting, once the temperature and pressure are assumed known; convolution to SCIAMACHY wavelengths is necessary. Rayleigh scattering is much reduced in the infrared, but aerosol parameters will still play an important role.

The choice of fitting strategy and fitting windows depends on the target species and the spectral region. In limb viewing of scattered light, it is difficult to account for diffuse multiply scattered contributions in the atmosphere and from the lowest surface layers. This is especially the case in the UV and visible, where  $O_3$  absorption is optically thick and multiple scattering effects are large. Multiple scattering is not so significant in the infrared, especially in Channel 8. Another problem with Channel 8 measurements is the low level of light; under most circumstances, there will only be enough signal-to-noise to distinguish the  $CO$  and  $N_2O$  features in the lowest views in a limb scan sequence.

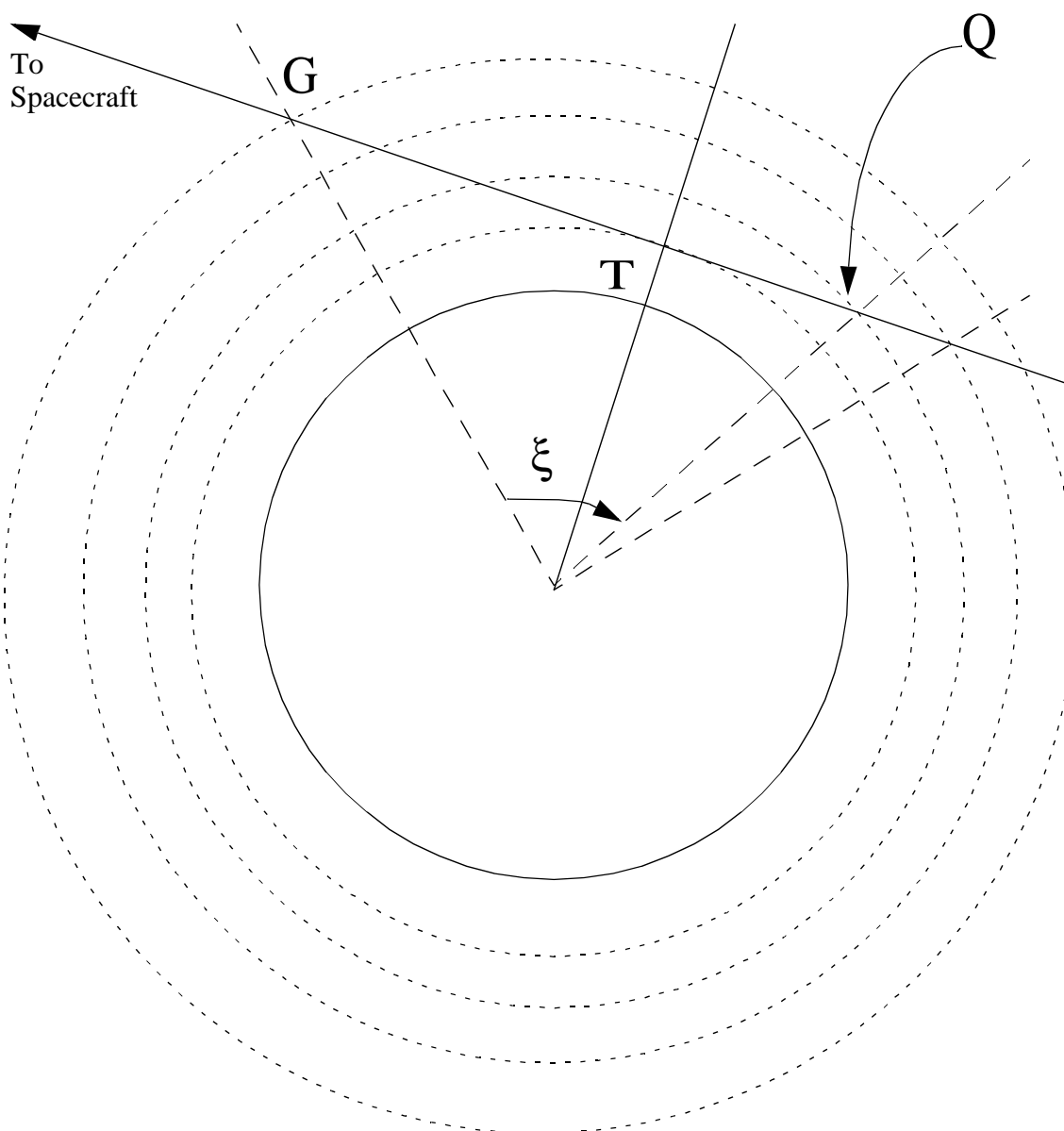
This issue of retrieval strategy has received some attention in the Requirements Phase Documentation. In all scattering applications in the UV and visible, global fitting has been the preferred method. Onion peeling has been successfully used in many limb applications, and options will be kept open to use this strategy alongside and maybe in conjunction with global fitting. Questions of retrieval strategy will not be considered in this issue of the ATBD.

It will not be possible in the initial operational data processor to perform simulations from scratch with a full multiple scattering RT model. In common with other new limb scattering developments, a single scattering model with analytic weighting function derivatives will be implemented, along with a look-up table of correction factors to handle the multiple scattering contributions (see [L12] and also the discussion below). At the time of writing, limb radiative transfer modelling has concentrated on the  $O_3$  UV problem, and the following sections deal with this application. However many aspects of the modelling (such as ray tracing, for example) are relevant for the infrared applications.

In the next 3 sections, we give a description of a single scatter forward model. The first section deals with the overall scattering formalism and the derivation of fluxes and weighting functions. This is followed by a section on the baseline optical properties of the atmospheric constituents used in the radiative transfer. The third section deals with ray tracing and initial radiative transfer computations. The fourth section has some material on multiple scatter corrections and the use of ratioed limb intensities in the retrieval. The last part deals with the preparation of the retrieval, describing the current state-of-the-algorithm.

### 4.3.2 Single Scatter Flux Model

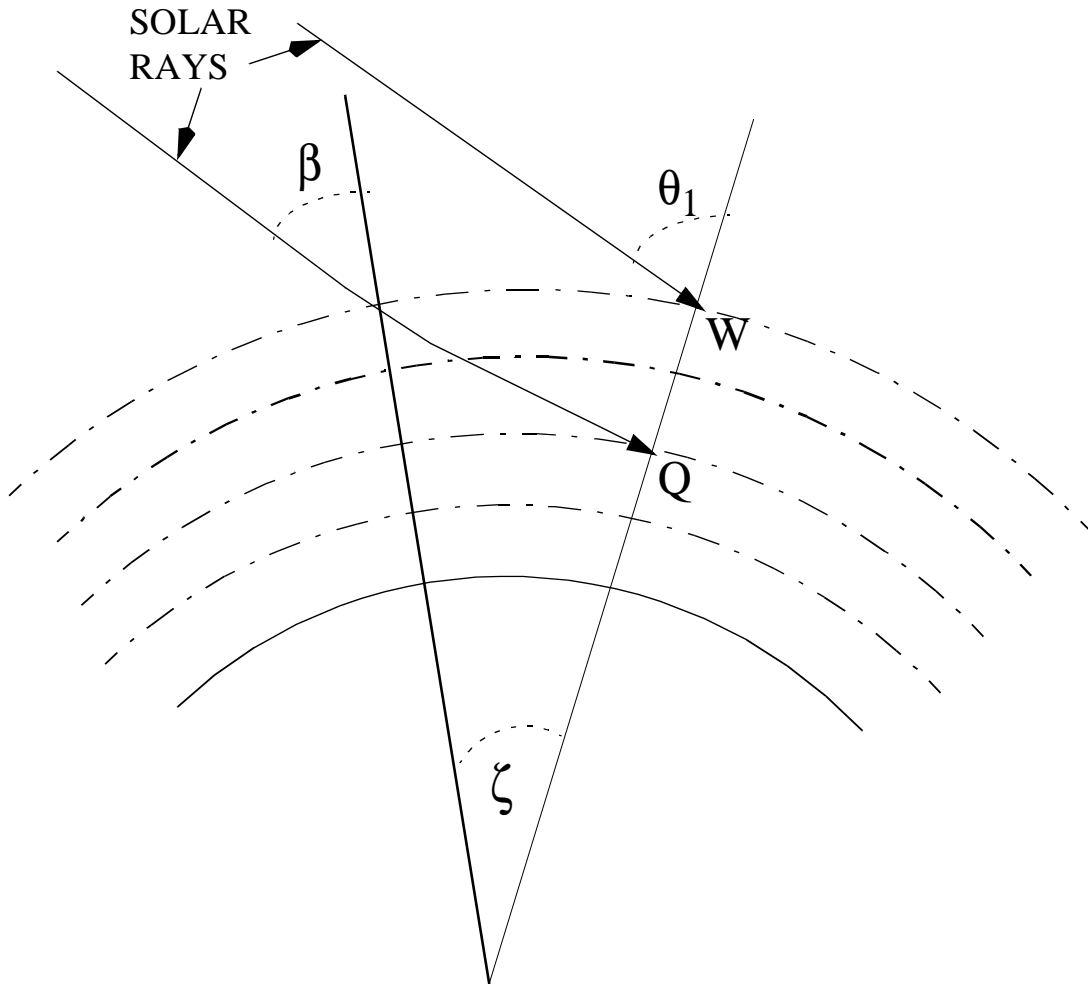
We shall assume that direct scattering of the solar beam into the line-of-sight is the only source of light; we omit for now any multiple scattering processes. A sequence of limb views defines a layer structure for the atmosphere through the computation of tangent heights, together with an assumption about the hydrostatic nature of the atmosphere. The TOA (top of the atmosphere) level above the first tangent height is defined by the user, and all input viewing geometries should be specified at this level. It is assumed that atmospheric refraction is neglected above this level, and that solar radiation reaching the atmosphere at this level has not been attenuated. Figure 1 illustrates schematically a limb view through 3 layers of the atmosphere.



**Figure 1: Limb-path geometry**

for a view through 3 layers of the atmosphere, with tangent point T. Note cumulative earth-centered angles, in particular  $\xi$  at point Q on the far side of tangent point T. Geolocation for this beam is specified at TOA point G, where the solar zenith angle is  $\theta_0$ .

If the limb path travels through  $N$  tangent layers, there are  $2N+1$  different solar rays that need to be accounted for. The first of these is the TOA ray on the far side of the tangent point - the scattering point is at TOA level, so no ray tracing is necessary. Similarly the last solar ray is scattered at TOA at the point where the limb path line-of-sight exits the atmosphere. All intermediate rays have some passage through the atmosphere to lower tangent levels, so these must be ray-traced (see Figure 2). Note that each solar ray scattered into the limb path has a unique angle of entry at the top of the atmosphere.



**Figure 2: Sun path geometry**

to point  $Q$  on the limb path Solar zenith at point  $W$  is  $\theta_1 = \theta_0 - \xi$  (See Figure 1). In the diagram,  $\zeta = \theta_1 - \beta = \zeta_{\text{Refraction}}(\beta)$ .

Denote the aerosol scattering coefficient at level  $n$  by  $\sigma_n^{\text{Aer}}$  and the molecular (Rayleigh) scattering coefficient is  $\sigma_n^{\text{Mol}}$ . Referring to Figure 1, solar beams on the far side of the tangent point are scattered “downwards” into the limb path on its way to the tangent point, and “upwards” into the limb path on the near side of the tangent point. In terms of the molecular (Rayleigh) and aerosol phase functions  $P_{\text{Mol}}$  and  $P_{\text{Aer}}$ , and the cumulative atmospheric transmittance  $C_n$  of the solar beam that is to be scattered at the  $n^{\text{th}}$  level, the scattering source terms at this level are given by:

$$S_n^R = \frac{\sigma_n^{Mol} P_{Mol}(\Theta_n^R) + \sigma_n^{Aer} P_{Aer}(\Theta_n^R)}{4\pi} \cdot C_n^R$$

The index R denotes the solar ray and runs from 1 (first solar ray at TOA on far side of tangent point) to 2N+1 (final solar ray at point of exit of limb path). The level index n is related to R:  $n = R$  for “downward” scatter points, and  $n = 2N+2-R$  for “upward” scatter points. In addition to the limb path ray tracing, each of the 2N+1 solar rays has a unique path through the atmosphere which must be determined from ray tracing. At TOA, we have no attenuation of the solar ray:

$$C_1^1 = 1 \quad \text{and} \quad C_1^{2N+1} = 1$$

Now consider scattering on the far side of the tangent point. If the limb path transmittance through layer m is  $T_m$ , then the “downward” flux  $F_{m+1}$  emerging at the bottom of this layer is expressed in terms of incident flux  $F_m$  at the top of the layer, and the downward scatter source terms as follows:

$$F_{m+1} = T_m F_m + (S_{m+1}^{m+1} + T_m S_m^m) D_m$$

Here,  $D_m$  is the slant distance through this layer along the limb path. This relation can be applied repeatedly from  $m = 1$  to  $m = N$  (the tangent layer), with  $F_1 \equiv 0$ . A similar recurrence relation exists for the “upward” fluxes  $G_m$  emerging from the top of the  $m^{\text{th}}$  layer:

$$G_m = T_m G_{m+1} + (S_m^{2N+2-m} + T_m S_{m+1}^{2N+1-m}) D_m$$

At the tangent level, we have  $F_{N+1} \equiv G_{N+1}$ , and this provides the link between the two recursion relations. The end result is  $G_1$ , which constitutes the single-scatter simulated intensity of backscattered light from a single limb view. This procedure must be repeated for each of the views in a scan sequence, and all values of  $G_1$  lumped together in a vector of ‘simulated intensities’ for use in the global fit algorithm. The above recursion relations can be used to derive weighting functions (derivatives of the outgoing intensity with respect to certain atmospheric variables), but first we must set up a state vector of such variables.

Let  $\underline{X}_m$  be the state vector of atmospheric variables to be retrieved or modelled at level m. For the UV application, this will include (as a minimum) two trace gas volume mixing ratios (for  $O_3$  and  $NO_2$ ), the Rayleigh scattering coefficient, and the aerosol extinction and scattering coefficients. [We will also include a variable for the inelastic component of molecular scattering, but this will be parameterised after the RT model simulation; this is discussed later in this chapter]. These 5 quantities (2 trace gas VMRs, 1 Rayleigh coefficient, 2 aerosol coefficients) will constitute the default baseline of atmospheric constituent variables to be included in the single scatter model simulation.

The aerosol coefficients are the values at a fixed wavelength (for example, the midpoint of a fitting window). One can parameterize the wavelength dependence of the aerosol optical properties, and introduce additional parameters characterizing this dependence. In the discussion of aerosol properties below, we suggest a linear-with-wavelength parameterization of aerosol extinction and scattering - this would introduce two more variables to be scaled and retrieved, bringing the total to 7 quantities in the state vector. It is not clear yet whether these is enough information in the spectra to retrieve these additional aerosol quantities. The molecular (Rayleigh) scattering coeffi-

cient has a wavelength dependence that is known to a high degree of accuracy (see next section) and this is not parameterised. We assume also that phase functions for molecular and aerosol scattering are known in advance (no retrieval).

It is convenient to use *normalised* (dimensionless) quantities for the vector  $\underline{X}_m$ , where the entries in the vector are just scale factors with respect to a set of baseline variables. Baseline Rayleigh and aerosol quantities are discussed separately below. Trace gas VMRs may be taken direct from a reference climatology, and cross-sections from a suitable source - see section below for details. The use of normalised atmospheric parameters is helpful in an iterative retrieval, as most of the radiative transfer can be done using baseline values before the iteration starts. In addition the 'first guess' in an iterative retrieval is usually  $\underline{X}_m = \underline{1}$  (unity) for all levels m (though this first guess could be derived dynamically in an operational context).

To compute backscattered intensities and weighting function kernels quickly and expeditiously, we express optical thicknesses and scattering source terms as linear combinations of these state vectors of (normalised) atmospheric variables. For the transmittance  $T_m$  along the limb path for layer m, we can write:

$$T_m = \exp(-\Phi_m \vec{X}_m - \Phi_{m+1} \vec{X}_{m+1})$$

$T_m$  has been expressed in terms of scaled atmospheric variables  $\underline{X}_m$  and  $\underline{X}_{m+1}$  at the two layer boundaries. The optical thickness baseline vectors  $\Phi_m$  are computed from baseline optical properties expressed in the reference hydrostatic atmosphere, taking into account the slant path variation across the layer due to changes in temperature and pressure and refractive index. In section 3.2 below, we illustrate the computation of the components of  $\Phi_m$ .

The above expression is convenient for computing parameter derivatives. With the notation

$\nabla_m T \equiv \frac{\partial T}{\partial \vec{X}_{m\alpha}}$  where vector  $\underline{X}_m$  has components  $X_{m\alpha}$ , we may write:

$$\nabla_m T_m = -T_m \bar{\Phi}_m \quad \text{and} \quad \nabla_{m+1} T_m = -T_m \bar{\Phi}_{m+1}$$

$$\nabla_k T_m = 0 \quad \text{for } k \neq m+1 \quad \text{and} \quad k \neq m$$

Similar considerations apply to the solar ray transmittance  $C_{m+1}^R$  for which the solar ray has passed through m layers down to a scatter point at level m+1 (ray index R = m+1 ("down") or R = 2N+1-m ("up")). We write:

$$C_{m+1}^R \equiv \exp\left(-\sum_{j=1}^{m+1} \vec{\Omega}_j^R \vec{X}_j\right) \quad \text{with derivatives } \nabla_k C_{m+1}^R = -C_{m+1}^R \vec{\Omega}_k^R \quad \text{for } k \leq m+1$$

Again the quantities  $\Omega$  will be calculated using baseline state vector variables and reference atmospheric pressures and temperatures - they must be defined for all solar rays appropriate to a given limb view, and for all limb views in a scan sequence. Finally we write the scatter source terms in the form:

$$S_n^R = \Lambda_n^R \vec{X}_n \cdot C_n^R$$

where  $\Lambda_n$  are baseline scattering contributions at level  $n$  for solar beam  $R$ . For the derivatives, we have then:

$$\nabla_m S_n^R = \vec{\Lambda}_n^R \vec{X}_n \cdot \nabla_m C_n^R \quad \text{for } m < n, \text{ and}$$

$$\nabla_n S_n^R = \vec{\Lambda}_n^R \vec{X}_n \cdot \nabla_n C_n^R + C_n^R \vec{\Lambda}_n^R \quad \text{at the scattering level } n.$$

Derivatives of the flux equations may now be written down using the definitions above. The flux computations and parameter derivatives can be done easily using the above rules, once the baseline vectors  $\Phi$ ,  $\Omega$  and  $\Lambda$  are evaluated. The first two vectors are baseline optical thickness quantities, the last is a scattering quantity. All these vectors need only be evaluated once at the start of a retrieval - to update intensities and kernel derivatives, one simply changes the state vector parameter values in the flux and derivative equations. Obviously this is useful in an iterative retrieval algorithm where simulated quantities require repeated updates. We now turn our attention to the baseline constituent optical properties selected for the model. Evaluation of the vectors  $\Phi$ ,  $\Omega$  and  $\Lambda$  is outlined in the following section.

### 4.3.3 Baseline Aspects of the UV Single Scatter Model

This section outlines the optical properties of the constituents used in the UV single scatter model. First we describe the reference atmosphere (assumed in hydrostatic equilibrium), then discuss baselines for Rayleigh scattering, aerosol extinction and scattering, and trace gas absorption.

#### 4.3.3.1 Reference Atmosphere

The trace gas retrieval in the UV will concentrate on VMRs for  $O_3$  and  $NO_2$ . The temperature and pressure are assumed known, either from a concomitant retrieval using channel 7 data, or from another source (assimilated meteorological data, for example). In what follows we assume that the temperature is specified on a given reference height grid, and that the pressure is known at one level (this is the hydrostatic reference level). For this, we take pressure and temperature at a user-defined reference level below the lowest tangent height. Pressures at all other points are then calculated using the hydrostatic equation:

$$Ln\left(\frac{p(z)}{p(z_0)}\right) = -\int_{z_0}^z \frac{g(\phi, z') dz'}{R_{gas} T(z')}$$

$R_{gas}$  is the gas constant for dry air. The acceleration due to gravity  $g(\phi, z)$  shows significant dependence on height and this must be accounted for in the integral. It also depends weakly on latitude  $\phi$ . We follow the gravity model found in the MIPAS technical documentation [L1]. This has the following height dependence:

$$g(\phi, z) = A + Bz + \frac{C}{(R + z)^2}$$

where parameters  $A$ ,  $B$ ,  $C$  and  $R$  depend on latitude. We take the latitude of the tangent point on the earth's surface as being representative (the limb path is symmetric about the tangent point), and ignore any variations of latitude over the limb path. If the temperature dependence on height is taken to be linear, then the hydrostatic integral may be solved in analytic form (no approximation).

In practice, reference values of temperature are first linearly interpolated to the retrieval or model grid (here defined by the tangent heights of the limb views), and then the hydrostatic equation is used to assign the pressure at these retrieval heights.

#### 4.3.3.2 Rayleigh Scattering Baseline

The Rayleigh phase function depends on the square of the cosine of the scattering angle, and is given by:

$$P_{\lambda}(\Theta) = A_{\lambda} + B_{\lambda} \cos^2 \Theta \quad \text{with} \quad A_{\lambda} = \frac{3(1 + \delta_{\lambda})}{2(2 + \delta_{\lambda})} \quad \text{and} \quad B_{\lambda} = \frac{3(1 - \delta_{\lambda})}{2(2 + \delta_{\lambda})}$$

The depolarization ratio  $\delta_{\lambda}$  is expressed in terms of the well-known King factor  $F_K$ , which in turn has a wavelength dependence expressed through a Sellmeier-type empirical equation [L20] [U7]:

$$\delta_{\lambda} = \frac{6F_K - 6}{7F_K + 3} \quad \text{where} \quad F_K = 1.0469541 + 3.2502153 \times 10^{-4} \lambda^{-2} + 3.8622851 \times 10^{-5} \lambda^{-4}$$

[Wavelength  $\lambda$  is in microns]. To get the molecular scattering coefficient at a given height, the cross section is multiplied by the air density at that height:

$$\sigma_{Mol}(\lambda, z) = \alpha_{Mol}(\lambda) \cdot \rho(z)$$

where the cross-section  $\alpha(\lambda)$  has a wavelength dependence close to the familiar  $\lambda^{-4}$  law, in the following expression derived recently from the data of Bates [L20] [U7] (units are  $\text{cm}^2/\text{mol}$ ):

$$\alpha_{Mol}(\lambda) = \frac{3.9992662 \times 10^{-28} \lambda^{-2}}{1.0 - 1.0689770 \times 10^{-2} \lambda^{-2} - 6.6814090 \times 10^{-5} \lambda^{-4}}$$

The air density is given in terms of pressure and temperature through the relation:

$$\rho(z) = \frac{P(z)}{1013.25} \cdot \frac{273.15}{T(z)} \cdot \rho_{zero}$$

where  $\rho_{zero} = 2.68676 \times 10^{24} \text{ mol.cm}^{-2}.\text{km}^{-1}$  is Loschmidt's number (density of air at standard temperature and pressure).

#### 4.3.3.3 Trace Gas Absorption Baseline

Cross-sections for  $\text{O}_3$  and  $\text{NO}_2$  can be taken from data derived from laboratory measurements taken with the flight model during the pre-launch calibration phase, or from literature sources. For  $\text{O}_3$ , the cross-sections  $\alpha(T)$  will have the usual Bass-Paur quadratic parameterization [U12] in temperature in the Hartley-Huggins bands:  $\alpha(T) = \alpha_0 + \alpha_1 T + \alpha_2 T^2$  ( $T$  in degrees Kelvin). This temperature dependence is important for limb retrieval (most of the ozone burden lies in the stratosphere).

Baseline VMR values for  $\text{O}_3$  and  $\text{NO}_2$  values may be taken from a variety of sources. The default has been to use values referenced on the set of USA standard atmospheres (some latitude/seasonal dependency is included in the  $\text{O}_3$  profiles). These reference VMRs will be linearly interpolated to

the model height grid as defined by the tangent heights. For *a priori* O<sub>3</sub> values, we take ozone profiles from SAGE II data (status TBD).

#### 4.3.3.4 Aerosol Extinction and Scattering Baseline

Most GOME and SCIAMACHY studies are now using the WMO aerosol scheme [L21], which has an explicit mixing of a number of basic aerosol components. This scheme to a certain extent supersedes the older LOWTRAN aerosol classification [L22], which relies on a “pre-mixed” classification of aerosols into certain broad types (“continental”, “maritime”, etc.). In the WMO scheme, the database consists of aerosol optical properties (scattering and extinction coefficients per particle, phase function moments) specified for a number of fixed components over a suitable wavelength range.

Suitable WMO-style component data sets have now been compiled for GOME and SCIAMACHY. Aerosol particles are assumed spherical and Mie computations for the optical properties are performed for some 10-12 components, some of which have a hygroscopic variation (humidity dependence). For the limb problem, we really require the external mixing of just two or three components - these are the sulphate aerosol component (in dry air), a volcanic ash component, and a meteoric dust component (mesosphere).

The atmosphere is divided into a number of regimes specified by height boundaries. For each regime, a number of WMO component types are listed together with their proportions  $f_k$  ( $\sum f_k=1$ ). If the aerosol loading is specified as a number density  $N(z)$ , then the extinction coefficients for the mixture are given by:

$$\epsilon_{\lambda}(z) = N(z) \sum_k \epsilon_{\lambda, k}(R_z) f_k$$

where  $R_z$  denotes the regime corresponding to height  $z$ , and the extinction coefficients inside the sum are the WMO data base values. To get values away from the data base wavelengths, linear interpolation is sufficient. A similar expression holds for the scattering coefficient.

The suggested default parameterization for the aerosol extinction coefficients has the following form for the wavelength dependence:

$$\epsilon(\lambda, z) = \epsilon_0(z) + \left(1 - \frac{\lambda}{\lambda^*}\right) \epsilon_1(z)$$

Here,  $\lambda^*$  is a reference wavelength (midpoint of fitting window). We express the pair of values  $\epsilon_0(z)$  and  $\epsilon_1(z)$  in terms of the database-derived extinction coefficients, interpolated to wavelength  $\lambda = \lambda^*$ . The results (together with the wavelength polynomial itself) constitute the baseline for aerosol extinction. The same parameterization pertains to the scattering coefficient.

The aerosol phase functions may be retrieved from the expansion:

$$P_{Aer}(\Theta) = \sum_{l=0} \beta_l P_l(\cos \Theta)$$

in terms of the Legendre polynomials  $P_l(\mu)$ , where  $\mu = \cos \Theta$ . The Legendre coefficients  $\beta_l$  are specified for a number of wavelengths for each WMO component; an overall phase function moment is defined by using an extinction-weighted sum over component moments. The total number of Legendre coefficients should be specified as an input variable. The moments will have wavelength dependence expressed linearly as functions of the database entries. The default will

be to assume that the phase functions are known in advance, with no attempt made to retrieve any phase function information.

It is possible to use the delta-M approximation [R12], in which the original phase function is replaced by a delta-function peak in the forward scattering direction plus a smoother phase function with a reduced number of Legendre moments. This approximation is useful for handling the strong preferential scattering in the forward direction that is often a feature of particulates scattering. This approximation has the effect of scaling the extinction and scattering coefficients. This approximation could be an important adjunct to aerosol phase function determination in the stratosphere if there is a large amount of volcanic ash in the external WMO mixture - phase functions for this aerosol component are strongly peaked in the forward direction (this is not the case for the regular sulphate aerosols, however).

Another option that has been considered is the smooth Henyey-Greenstein phase function which can be derived explicitly from knowledge of the “asymmetry parameter”  $g$ :

$$P(\Theta) = \frac{1 - g^2}{(1 - 2g\mu + g^2)^{1.5}} \quad \text{where } g = \frac{\beta_1}{3}$$

This is convenient but not always representative of aerosols in the atmosphere.

#### 4.3.4 Ray Tracing and Optical Thickness Vectors

##### 4.3.4.1 Ray Tracing

This is one of the most important aspects of the limb-scatter problem. In limb emission and occultation retrieval it is usually enough to consider ray tracing only along the line of sight, but for limb scatter, the solar rays must also be ray traced to the scatter points on the line-of-sight path. This greatly increases the number of ray tracing computations, and we want a quick and convenient way to handle this part of the algorithm. Ray tracing in a refractive atmosphere follows Snell’s law for the layer incident and refracted angles at height  $z$ :

$$\frac{\sin I(z)}{\sin R(z)} = n(z)$$

where the dependence of the refractive index  $n(z)$  on height comes through the atmospheric temperature and pressure  $T$  and  $P$ . An improved Edlen-type formula has recently been derived for the refractive index of air at STP for  $\lambda$  in microns [U7]:

$$(n_{Air} - 1) \times 10^4 = 0.7041 + \frac{315.90}{157.39 - \lambda^{-2}} + \frac{8.4127}{50.429 - \lambda^{-2}}$$

In limb viewing, it is sufficiently accurate to use the Born-Wolf (ideal gas law) approximation to this equation:

$$n(z) = 1 + \kappa \frac{P(z) T_0}{T(z) P_0}$$

where  $1 + \kappa$  is the value of  $n$  at reference values  $P_0$  and  $T_0$ . In the UV at 330 nm, and assuming the value of  $n_{Air}$  given above at STP, we get  $\kappa = 0.00030395$ . This is significantly different from the value 0.0002874 used for MIPAS infrared applications.

The first task of the ray-tracing is to determine the tangent heights. This is done by iteration, starting from a first guess equal to the geometric tangent height value:

$$H_0 = (R_e + H_{TOA}) \cos \gamma$$

where  $R_e$  is the earth's radius and  $\gamma$  is the limb view elevation angle at TOA. [This angle is part of the geolocation record input to level 1b to 2 processing, along with the TOA value for limb retrieval]. We then find the values of temperature and pressure in the reference atmosphere corresponding to this height, and construct the refractive index of air for these values of P and T. One then uses Snell's law to find a new tangent height reached by the incident limb ray with this degree of refraction. This process converges quickly after a small number of iterations. Tangent heights are determined for each of the limb views in a scan sequence, and these heights constitute the levels at which the retrieval and RT modelling will be carried out.

We then subdivide these "tangent" layers into a number of fine layers with a height resolution of typically 0.1 km in the lower stratosphere and anything up to 5 km for the topmost layer (this depends on TOA height above first tangent level). Temperatures and (hydrostatic) pressures are then computed on these fine levels, along with air density and refractive index values. An optimised fine layering resolution will be worked out for the operational prototype.

Next we perform (for each limb view) ray tracing for the limb path, computing the slant path distances (the quantities  $D_m$  in the flux recurrence relations) and associated quantities required for the optical thickness computations (see next section for details). For each limb view, we also compute the earth-centred angles subtended by the passage of the ray from TOA to all scatter points on intermediate levels (see Figure 1). Each line-of-sight path is symmetric about its tangent point.

The input requirement on limb geolocation for level 1b to 2 processing states that the solar angles must be specified at the TOA height for each limb view at the point of exit of the line-of-sight path. Figure 2 shows the situation for one solar ray to be scattered at point Q on the limb path. Let  $\xi$  be the earth centred angle generated along the limb path from Q to the exit point. If  $\theta_0$  is the solar zenith at the exit point for this line-of-sight, then  $\theta_1 = \theta_0 - \xi$  is the solar zenith at the TOA level above Q. To find the entry angle  $\beta$  for this ray, we must use the ray tracing to compute the earth-centered angle  $\zeta$  (see figure 2). If  $\zeta_{Ref}(\beta)$  is the result of computing  $\zeta$  by ray tracing, given an entry zenith angle  $\beta$ , then the solution for  $\beta$  is found from the equation  $\zeta_{Ref}(\beta) = \theta_1 - \beta$ . This is solved iteratively, with the first guess  $\beta_0$  given by the geometric approximation:

$$\sin \beta_0 = \left[ \frac{(R_e + H_Q)}{(R_e + H_{TOA})} \sin \theta_1 \right]$$

$R_e$  is the earth radius, and  $H$  denotes height above the earth's surface. If  $\beta_n$  is the value after the  $n^{\text{th}}$  iteration, the next guess is given by:

$$\beta_{n+1} = \beta_n + F |\zeta_{Ref}(\beta_n) - \theta_1 + \beta_n|$$

F is a factor inserted for convenience to speed up the iteration. For most cases, it is sufficient to take  $F=1$ , but for large angles it is better to use smaller values. It is possible to pre-calculate a data set of F-values for a reference atmosphere, for a number of height levels and a number of solar zenith angles, and then use these pre-computed factors to speed up the iteration. This procedure ensures that accurate results can be found quickly for a small number of iterations for all possible values of the solar zenith angles.

In order to calculate scattering angle cosines  $\cos \Theta_n^R$  for solar ray R at level n, we use the following standard result in terms of solar ray and limb path zenith angles  $\theta_S$  and  $\theta_L$ ; and the relative azimuth  $\varphi_S - \varphi_L$  between the planes containing the two paths:

$$\cos \Theta = -\cos \theta_S \cos \theta_L + \sin \theta_S \sin \theta_L \cos(\varphi_S - \varphi_L)$$

The zenith angles must be found by ray tracing for each scatter point.

#### 4.3.4.2 Baseline Radiative Transfer

This section deals with the computation of baseline vectors  $\Phi$ ,  $\Omega$  and  $\Lambda$  as defined above. These quantities may be expressed in terms of scaled state-vector variables (VMRs, Rayleigh scattering coefficient, aerosol coefficients) at the tangent levels. In the computation of optical thickness vectors  $\Phi$  and  $\Omega$ , we can allow for differential attenuation along the ray traced slant paths, by using the fine layer structure of the atmosphere already set up for the ray tracing, and by making certain interpolation assumptions. For now we will assume that all variables vary linearly across a layer. This is fine for “extrinsic” variables (VMRs, temperature), but exponential interpolation may be more accurate for the “intrinsic” variables (Rayleigh and aerosol coefficients). However the latter assumption sacrifices simplicity, and for now we take linearity.

Consider the vectors  $\Phi_n$  and  $\Phi_{n+1}$  which appear in the definition of limb path transmittance  $T_n$  in a single layer between adjacent tangent levels; first we look at the components of these vectors due to ozone absorption. Let  $X_n$  and  $X_{n+1}$  be the upper and lower boundary values of the  $O_3$  VMR scaling parameters, with actual baseline values  $V_n$  and  $V_{n+1}$  (the latter are interpolated values from a reference database for example). Suppose that the layer is divided into M fine layers, each such fine layer having a midpoint height  $z_j$  where  $j = 1, \dots, M$ . We assume that VMR values may be linearly interpolated to these points:

$$V_j = \frac{(z_{n+1} - z_j)V_n X_n + (z_j - z_n)V_{n+1} X_{n+1}}{(z_{n+1} - z_n)}$$

Then if the  $O_3$  cross sections have the form indicated in the discussion above about trace gas absorption (quadratic in temperature), and if the fine layering slant path distances are  $ds_j$ , the air densities  $\rho_j$ , and the temperatures  $T_j$  in the sub-layers, then the optical thickness across the whole layer is:

$$\tau_n^{O3} = \sum_i (\alpha_0 + \alpha_1 T_j + \alpha_2 T_j^2) V_j \rho_j ds_j \equiv \Phi_n^{O3} X_n + \Phi_{n+1}^{O3} X_{n+1}$$

from which it follows that:

$$\Phi_n^{O3} = \frac{(\alpha_0 \langle 1 \rangle + \alpha_1 \langle T \rangle + \alpha_2 \langle T^2 \rangle) z_{n+1} V_n - (\alpha_0 \langle z \rangle + \alpha_1 \langle zT \rangle + \alpha_2 \langle zT^2 \rangle) V_n}{(z_{n+1} - z_n)}$$

Here the brackets indicate a slant path summation  $\langle x \rangle = \sum_j x_j \rho_j ds_j$  over the fine layer division.

$\langle 1 \rangle$  is just the slant column amount in this notation. A similar expression can be written down for  $\Phi_{n+1}$ . If there is no temperature variation of the cross-section, the computation is simpler.

A straightforward alternative to this kind of fine-layer summation is to take an average value of the trace gas absorption at the layer boundaries, and multiply by the slant path distance to get the optical thickness. This gives the following:

$$\Phi_n^{O_3} = \frac{D_n V_n \rho(z_n) \alpha(T_n)}{2} \text{ and similarly for } \Phi_{n+1}^{O_3}$$

$D_n$  is the slant path distance (the only quantity with fine-layering evaluation in this approximation). However this method does not account for fine-layer variations in temperature and trace gas concentration.

The polynomial dependence on temperature allows us to compute the fine-layer variation in a quick way. To date, no convenient temperature parameterization scheme has been established for the NO<sub>2</sub> cross-sections, but a polynomial-based interpolation scheme over a number of fixed temperatures would be convenient for the above treatment.

For line absorption in the infrared, the temperature dependence of the cross-sections is complicated and not explicit, and the computational burden of cross-section evaluation dictates the use of layer effective temperatures as defined by the Curtis-Godson sums (q.v.).

For the molecular (Rayleigh) scattering contribution to limb path optical thickness, we write:

$$\tau_n^{Mol} = \Phi_n^{Mol} Y_n + \Phi_{n+1}^{Mol} Y_{n+1}$$

where  $Y_n$  is the scaling parameter for the Rayleigh scattering coefficient  $\sigma_{Mol}(\lambda, z_n)$  at height  $z_n$ . From the definition of this coefficient, and assuming again a linear variation with height across the layer, we get:

$$\Phi_n^{Mol} = \frac{\alpha_{Mol}(\lambda) [\langle 1 \rangle z_{n+1} - \langle z \rangle]}{(z_{n+1} - z_n)}$$

A similar consideration applies to the aerosol optical thickness contributions.

Elements of vectors  $\Omega_n^R$  are evaluated in the same manner. The scatter contributions (vectors  $\Lambda_n$  in the model) are specified at the level heights only, so there is no averaging across the layer. It is simply then a matter of multiplying the baseline scattering coefficients with their respective phase functions.

These sorts of calculations must be repeated for every layer traversed by the limb path in a given view, and for all layers traversed by the solar rays on their way to the scatter points.

#### 4.3.5 Multiple scatter corrections in the UV model

The proper RT treatment requires a fully spherical multiple scattering model. For past and existing limb scatter remote sensing experiments, the approach has been to use a quasi-analytic single scattering model, and deal with multiple scattering in the form of correction factors. For the SOLSE/LORE experiment, the target species is O<sub>3</sub> in the UV/visible, and simulations and weighting functions are calculated according to:

$$I_{ms}(\lambda, s) = I_{ss}(\lambda, s) \cdot (1 + C(\lambda, s))$$

$$W_{ms}(\lambda, s, \underline{a}) = W_{ss}(\lambda, s, \underline{a}) \cdot (1 + C(\lambda, s))$$

These relations are valid for wavelength  $\lambda$ , limb view number  $s$  of a scan sequence, and for state vector  $\underline{a}$  of parameters to be retrieved. The single scatter results  $I_{ss}$  and  $W_{ss}$  come from the quasi-exact single scattering model.  $I_{ms}$  values are tabulated as corrections to the single scattering results, and they are computed using a state-of-the-art fully spherical model [R15]. This model is too slow for retrieval purposes, but it has been used to correct simulations for nadir-mode total column retrievals where the line-of-sight is significantly away from nadir [R16].

This will be the approach adopted for SCIAMACHY, at least in the UV application. The acquisition of a multiple scattering model for table computation is currently under negotiation. The above formulation assumes that the correction factors are independent of the vector of atmospheric parameters to be varied.

Correction factors will cover diffuse multiple scatter contributions, including light reflected and scattered from the ground and tropospheric parts of the atmosphere. The problem remains to characterise such contributions. In a stand-alone limb-scatter retrieval, little is known about the ground and cloud cover and the tropospheric constituent burdens, and simply basing correction factors on climatology and assumed reference quantities could lead to large errors in the forward simulation. One solution is to use a set of correction factors based on a single equivalent atmospheric parameter which accounts for all the lower atmosphere effects. We could assume that the atmosphere has a fixed composition below the lowest tangent height down to a fixed lower boundary with an assumed Lambertian reflecting surface. The free parameter is then the Lambertian albedo  $A$  of this surface, and we set up a table of “equivalent correction factors” by matching values of  $A$  to a wide variety of scenarios. The albedo could then be an additional parameter in the limb retrieval, one which is orthogonal to the state vector of atmospheric parameters as outlined for the single scatter model. That is, the single scatter contribution does not vary with  $A$ , and the correction factor gradient with respect to  $A$  would be done by finite differencing. This approach requires extensive off-line studies with the full RT models.

Another option is to use a ratio of limb view radiances. This has been the approach adopted by the SOLSE/LORE team [L2]. In order to ameliorate the lowest-layer and surface reflection contamination, a ratio of two limb radiance spectra is used as the basic measurement quantity; radiances at any given tangent height are divided by their values for the 55 km tangent height. To first order, the surface reflection multiple scatter contributions are factored out in the simulation of these ratios.

If the simulation for limb view  $N$  yields single scatter quantities  $I_n(\lambda)$  and  $W_n(\lambda, \underline{a})$ , and correction factors  $C_n(\lambda)$ , and we want to ratio the intensities with values from view  $M$ , then we have:

$$R_n(\lambda) = \frac{I_n(\lambda)}{I_m(\lambda)} \cdot D_n(\lambda) \text{ where } D_n(\lambda) = \frac{1 + C_n(\lambda)}{1 + C_m(\lambda)}, \text{ and}$$

$$Z_n(\lambda, \underline{a}) = \left[ \frac{W_n(\lambda, \underline{a})}{I_n(\lambda)} - \frac{W_m(\lambda, \underline{a})}{I_m(\lambda)} \right] \cdot R_n(\lambda)$$

Typically one divides by radiances from a limb view with high tangent height (this will be one of the last in a SCIAMACHY limb scan sequence). One must take care with the derivatives in the second of these two equations, because  $W_m(\lambda, \underline{a})$  will not be defined for levels below  $m$ . Experience has shown that the ratio of correction factors  $D_n(\lambda)$  is much less strongly dependent on lower atmosphere variations, and it is probably expedient to construct a look-up table of these quantities, rather than their unratioed equivalents. Status TBD.

Another advantage of this approach is that the use of ratioed spectra will obviate the requirement on a solar spectrum as an absolute radiometric standard. We also avoid problems of wavelength misrepresentation between the solar spectrum and the limb backscattered spectra.

#### 4.3.6 Preparation for SCIAMACHY Retrieval

The level 1b input data required for this part of SGP comprises limb spectra and geolocation. The limb sequence will contain a number of limb views, and we will extract spectra from each view to be used in the fit. Quality checks should be made to ensure no cross-clustered data or error-flagged pixels are present within the fitting windows. We will assume for now that the wavelength calibration is the same for all limb backscattered spectra in the scan sequence - this is a proper requirement for the use of ratioed measurements. The limb sequence will also come with a solar extraterrestrial spectrum.

The baseline geolocation definition requires values of the solar zenith angles, the limb path elevation angles and the relative azimuths between the planes of view to be specified at the TOA level, at a point where the limb line-of-sight exits the atmosphere. Thus the TOA level must be given along with the viewing geometry. The default value will be 80 km.

Measurement spectra are then buffered into fitting windows. These windows could in theory be defined uniquely for every view, but in practice we will use the same window for all views in the scan sequence (in contrast with the infrared retrieval, where one might tailor window choices to specific trace gas absorption lines). Not all views in a sequence might be used; we could for example take a block of typically 12-15 limb views from a lowest tangent height of the order of 16-20 km, up to a highest tangent height of 50-60 km. [The limb view elevation angle steps translate into a vertical height separation of about 3.1 km between tangent levels in the stratosphere].

For the global fit retrieval, we now prepare the simulated single-scatter spectra and weighting functions as described above, plus the extraction of correction factors from look-up tables. The use of a Ring scaling parameter has proved to be a vital ingredient in the ozone profile retrieval carried out for GOME, and we propose to do this here. We can take over the Ring spectrum prepared for GOME studies in the UV, and convolute this down to wavelength grid of the SCIAMACHY limb measurements, using the instrument response function in the folding. We then add a multiple of this convolved spectrum to each of the simulated limb view spectra - a different scaling factor will be retrieved for each view.

Finally we will allow for some closure parameters in the fitting. For each view, we will define additional continuum and tilt factors. The continuum will be an additive polynomial of low order in wavelength, while the tilt will be a multiplicative polynomial, also of low order in wavelength. These coefficients round up the set of 'parameters-to-be-retrieved' which will comprise the state vector to be entered into the retrieval package. The final set of simulated spectra and kernel derivatives for the selected scan sequence will be assembled as a single vector and matrix to be passed to the retrieval package.

If we are using optimal estimation, we require some *a priori* input. For the most important constituent, we use SAGE II O<sub>3</sub> climatology which comes with a proper covariance matrix (ozone amounts should be normalised to the baseline VMR values as defined above, since the retrieval is designed to work with normalised values). For all other constituents we take the *a priori* values to be the baseline values used in the model setup, and simply assign covariance values on an ad-hoc basis. Thus we might assume that the *a priori* value for the scaled NO<sub>2</sub> VMR at each level is simply 1, with a 50% error to be used.

#### 4.4 Occultation Retrievals

With the sun as the direct source of light, occultation measurements have high signal-to-noise (low measurement errors); additional sun-follower options will also allow for finer vertical resolution in the scanning [A5]. In principle, transmission along the direct line-of-sight to the sun is the only significant factor in the simulation, and this greatly simplifies the radiative transfer aspects of the retrieval. Despite the limited availability of measurements (~ 60 s before sunrise in Arctic latitudes), occultation measurements could have great potential for accurate profile retrievals.

However, there are problems to be solved in association with the measurement strategy employed by SCIAMACHY. In sun occultation, SCIAMACHY is first pointed to the approximate location at which the sun is expected to be situated some 17 km above the horizon (this avoids problems with atmospheric refraction). Because the position of the sun is not known exactly (spacecraft pointing inaccuracies), the elevation mirror scans continuously up and down. When the sun enters the field-of-view, a sun-follower locks the azimuth mirror on the sun, but the elevation mirror continues to scan the slit up and down over the solar image. We thus get a sequence of spectra for which the height in the atmosphere and the position on the solar surface are not accurately known. Since different horizontal slices of the solar image cover different surface areas, there is no radiance calibration of the spectra.

Further uncertainties in radiance level come from solar limb darkening and sunspots. Limb darkening may also slightly change the depths of the Fraunhofer lines, while Doppler shifts due to the rotation of the sun may yield slightly wider or narrower Fraunhofer lines depending on which sun-slice is in view. Thus, the self-calibration (as used in e.g. the SAGE instruments) of the occultation method is compromised. The self-calibration may be regained by summing all slices which make up a complete solar image, but this reduces the effective vertical resolution of the measurement to ~25 km. It is uncertain that sequences of solar images at different heights in the atmosphere will contain enough information to retrieve profiles with higher vertical resolution than this value.

Additional problems may be caused by the small diaphragm inserted in the beam to reduce light levels when viewing the sun. It may be expected that this reduces the width of the slit function, and also the wavelength calibration may be slightly changed. In summary, the solution of these problems requires a large degree of further scientific work, and it is unrealistic to expect that an operational processing algorithm for this viewing mode can be implemented with complete success by the time of launch.

Some of the earliest studies for SCIAMACHY focussed on the high precision to be expected from occultation retrieval of trace gas profiles [L15]. Precision estimates were found by examining small changes from a simulated set of transmittances, subject to an *a priori* constraint. These results represent the initial linearisation step in the non-linear estimation procedure, and an optimal estimation retrieval algorithm has now been developed for occultation studies. The fitting involves the direct comparison of simulated and measured transmission values (ratios of attenuated to vacuum irradiance observations). Rayleigh and aerosol closure coefficients are retrieved simultaneously with the trace gas VMRs and the P-T values.

As with the limb applications, the P-T profile should be derived concurrently from the occultation measurements to ensure that a consistent set of trace species profiles is achieved. The list of target species is essentially the same as that for the limb problem, though the greater precision may mean that additional species (such as NO<sub>3</sub>) could be seen in the occultation mode.

The SAGE II and III instruments have used another approach to occultation retrieval [L12]. The initial step here is to exploit the linearity of the transmission functions to derive slant column densities of aerosol and trace species. This is a straightforward multi-linear regression along the lines of the DOAS fitting of section 3.2 on page 31 (it is assumed that wavelength registration is not a problem). In this approach, the Rayleigh scattering contribution to the slant optical depth is subtracted before the differential absorption fitting is performed (this assumes knowledge of the P-T profile). Trace species optical depths may be derived from a variety of fitting windows (especially O<sub>3</sub>, NO<sub>2</sub> and aerosol). The final step involves an inversion from slant to vertical optical depths using a standard Twomey-Chahine non-linear inversion method (see [L12]). This technique has not yet been considered for SCIAMACHY. It should be noted that the P-T retrieval cannot be done this way.

## 5 Radiative Transfer Model Requirements

### 5.1 GOMETRAN and extension to SCIAMACHY (nadir only)

The forward model GOMETRAN was developed especially for simulation of back-scatter intensities and weighting functions associated with the geophysical retrieval of atmospheric parameters from GOME measurements. It is based on a finite difference approach, and in the next paragraphs we give a short resume to highlight the main points.

For the nadir back-scatter of incident sunlight, we separate the direct-beam and diffuse radiation fields. The direct beam term requires a transmittance calculation through the atmosphere, plus a description of the surface reflecting property. For the diffuse field, the integro-differential equation of radiative transfer for the intensity field  $I(z, \Omega)$  at height  $z$  and solid angle  $\Omega$  can be written in the general form [R2]:

$$\mu \frac{d}{dz} I(z, \Omega) = -\epsilon(z) I(z, \Omega) + \frac{1}{4\pi} \sigma(z) \int_{\Omega'} P(z, \Omega, \Omega') I(z, \Omega') d\Omega' + \sigma(z) \zeta(z, \Omega) \quad (37)$$

Here,  $\epsilon(z)$  and  $\sigma(z)$  are the extinction and scattering coefficients, and the integral represents the multiple scatter contribution with phase function  $P(z, \Omega, \Omega')$ .  $\zeta(z, \Omega)$  is the source function, which contains the direct single scatter contribution and the singly scattered light from surface reflections of the direct beam (and emission terms if required).

The angular dependence of the radiation field is separated by means of an expansion of the phase function in Legendre polynomials in the scattering angle cosine. Using the definition of scatter angle in terms of zenith and relative azimuth angles  $\{\mu, \phi\}$ , and by means of the completeness relation between associated and full Legendre polynomials, it is possible to determine the azimuth dependence in a Fourier cosine series:

$$I(z, \mu, \phi) = \sum_{m=0}^{\infty} (2 - \delta_{m0}) I_m(z, \mu) \cos m\phi \quad (38)$$

Working with Fourier components  $I_m(z, \mu)$ , we then discretise  $\mu$  with Gaussian quadratures, replacing  $\int \mu d\mu$  by the sum  $\sum \mu_j w_j$ . The height derivative in equation (37) is computed as a finite difference [R2][R3]. These two assumptions allow the RT equation to be expressed as a matrix algebra equation  $Q_m I_m = S_m$  for each Fourier component  $I_m$ , source term  $S_m$ , and matrix operator  $Q_m$ . The formal solution may be found by matrix substitution, and this is one of the major tasks in the numerical integration. Fourier components are calculated until a convergence criterion applied to equation (38) is satisfied.

The advantage of the matrix formulation is that the equation  $Q_m I_m = S_m$  can be linearised about the solution [R7]. The linearisation is able to deliver a whole field of weighting functions (that is, responses of the radiation field at different heights to small changes in selected atmospheric constituent variables) with little additional effort. This is in contrast to all other RT solution methods, which require extensive perturbation calculations for weighting function evaluation. Output of weighting functions depends on which atmospheric parameters are allowed to vary. To date, the set of weighting functions for surface albedo variation, trace species concentration variation, molecular scattering, aerosol extinction and scattering, and temperature shift.

From databases of pressure, temperature and trace species concentrations, aerosol loading and optical properties, surface data sets and trace gas absorption properties, simulations across the GOME spectrum from 240 to 790 nm have been performed. For the O<sub>2</sub> A band, both line-by-line calculations and exponential sum-fitted approximations (also known as correlated-k) have been done [M. Buchwitz, GOME and SCIAMACHY Working Session 7, April 1998]. Rotational Raman scattering has been introduced into the model to investigate the Ring effect [R6]. Clouds have been incorporated into the model both in a quasi-exact layer treatment, and in an approximation assuming a bi-directionally reflecting cloud-top as the lower boundary of the atmosphere [R5]. GOMETRAN has also been used as the forward model in the optimal estimation retrieval of O<sub>3</sub> profiles for a selection of scenarios [P4][P10]. The original formulation [R4] was for plane-parallel atmospheres; later versions have incorporated a ray-tracing calculation for the attenuation of the direct solar beam to the lower boundary (the pseudo-spherical approach).

GOMETRAN has been validated against DISORT [R11] in a number of situations, with excellent agreement. In the context of GOME study work [G6], some comparisons have also been made with other RT models, including the Monte-Carlo code used for AMF calculations at Heidelberg (see [R9] for an exposition) and the Doubling-Adding model [R8].

### 5.1.1 AMF generation

In GDP, the model was used off-line to generate tables of multiple scattering correction factors to be applied to the single scattering AMF values calculated from scratch as part of GDP DOAS code. In fact the quantities stored were fitted coefficients for the correction factors, assuming polynomial dependence on the solar and line-of-sight zenith angles [G3]. For O<sub>3</sub>, the look-up classification allowed for some 6 surface albedo values from 0.02 to 0.9, 9 lower boundary heights from 0 to 10 km, some 8 latitude zones, 2 aerosol types (land/sea), and 24 coefficients covering the two zenith angle dependencies. For a given viewing geometry ( $\mu, \mu_0, \phi$ ), the AMFs are recovered as polynomials in  $\mu_0$  and  $\mu$ , and interpolation takes care of the  $\phi$  values and all the other quantities.

In the updated GDP (and this will also apply to SCIAMACHY), AMFs for O<sub>3</sub> and NO<sub>2</sub> will be generated completely from GOMETRAN-calculated look-up tables. Rather than catalogue the actual AMF values, it is more convenient to store Fourier components of the intensities (as in equation (38) above); this takes care of the azimuth dependence explicitly when recovering the AMFs. (This is important in distinguishing the east, centre and west forward pixel scan geometry). One would need also to store the number of Fourier azimuth harmonics required for an accurate result from the Fourier series (an accuracy parameter governs the series convergence). AMFs may then be restored from the definition  $AMF = \ln(I_{nogas}/I_{total})/\tau_{vert}$  and interpolated to the appropriate viewing zenith angles and other climatological variables.

It is also possible to estimate the error on an AMF due to climatology uncertainty. We suppose that the climatology for trace species  $g$  depends on a set of parameters  $U_j, j = 1, M$ , which have error covariance matrix  $C_{ij}$ . Then the variance is:

$$var(A) = \sum_{i,j} \frac{\partial A}{\partial U_j} C_{ij} \frac{\partial A}{\partial U_i} \quad (39)$$

where

$$\frac{\partial A}{\partial U_j} = -\frac{1}{\tau} \left[ A \frac{\partial \tau}{\partial U_i} + \frac{1}{I_{total}} \frac{\partial I_{total}}{\partial U_i} \right] \quad (40)$$

There is no dependence of  $I_{nogas}$  on  $U_j$ , because this intensity was computed with the trace gas  $g$  excluded. If GOMETRAN takes a concentration profile  $p(z_i)$  as input, for height levels  $z_i, i = 1, \dots, N$ , then the intensity derivatives have a direct expression in terms of GOMETRAN weighting function output  $W_i \equiv p_i \partial I / \partial p_i$  :

$$\frac{\partial I_{total}}{\partial U_j} = \sum_{i=1}^N \frac{W_i}{p_i} \frac{\partial p_i}{\partial U_j} \quad (41)$$

*The NRT look-up table of AMFs will not be broken down into Fourier components of Intensity. In the interests of simplicity and to provide continuity with GDP, the NRT table will contain actual AMF values computed for a number of solar zenith angles, line-of-sight-zenith angles, azimuth angles, a number of lower boundary heights, a number of surface albedos, and for all profiles present in the climatology. Due to the need for early specification, there will be no incorporation of AMF errors in these NRT tables.*

### 5.1.2 Extension to SCIAMACHY (nadir only)

The nadir mode solution to the radiative transfer equation embodied in the GOMETRAN model can be taken over for SCIAMACHY applications. The extension to SCIAMACHY is more a question of ensuring that the radiative transfer code can take suitable inputs for wavelengths out to 2400 nm. To this end, the GDP database of aerosol optical properties has been extended in wavelength [A9]. Additional trace species number mixing ratio profiles are required for those absorbing only in the infrared channels (CO, CH<sub>4</sub>, N<sub>2</sub>O and CO<sub>2</sub> have been prepared so far). The original spectral database of spectral albedos was prepared for GOME and SCIAMACHY [G3]. The databases of cloud optical properties and cloud-top reflectances have also been extended (T. Kurosu, private communication, 1998) to 2.5 microns.

The largest extension to the model has been the incorporation of line-by-line absorption coefficients in the GOMETRAN code. It is more convenient to maintain the line-by-line code separately from the GOMETRAN scattering code. Required sets of absorption coefficients pre-calculated using LBL at temperatures and pressures are then regarded as input to the RT model. An alternative approach using the correlated-k (exponential sum-fitted) method for transmittance involves the creation of a comprehensive data set of pre-calculated sum-fitted coefficients that could be used directly with GOMETRAN. This has been accomplished with the O<sub>2</sub> A band (M. Buchwitz, GOME and SCIAMACHY Working Sessions 7, April 1998) and there are plans to extend this work to some of the infrared molecules.

*The level 1b to 2 NRT processor will not contain explicit RT code. However, Look-up tables of Air Mass Factors (UVAS algorithm) and Rayleigh Correction Factors (AAIA) will be generated off-line by GOMETRAN. These tables and their derivations are described elsewhere in this document.*

The description of a forward simulation radiative transfer model for SCIAMACHY limb scatter scenarios is a major new development and is treated separately in this document in section 5.3 below.

## 5.2 MODTRAN and DISORT

DISORT [R11] is discrete-ordinate monochromatic multiple scattering code. It contains no reference to any particular trace species, or to types of scattering particulate (molecular/aerosol); there are no transmittance calculations. It requires the following generic inputs for the optical properties of the reflecting atmosphere: layer optical thickness values, layer single scatter albedos, layer phase function Legendre expansion coefficients and coefficients for the surface reflection property. It is up to the user to create these inputs according to the problem under consideration. Like GOMETRAN, DISORT uses the Legendre polynomial decomposition for the scatter angular dependence, with the Fourier cosine decomposition in azimuth, and a zenith cosine quadrature scheme. However, the vertical atmospheric co-ordinate is optical depth (not altitude), and the multi-layer RT equation requires the simultaneous solution of a number of first-order differential equations.

The model has a long history and has been widely used and tested; it is straightforward to interface. It has been used frequently in comparison with GOMETRAN (see [G6]). The major disadvantages are (i) the lack of a pseudo-spherical version, and (ii) the treatment of bi-directional reflecting surfaces needs an upgrade. The use of optical depth as vertical co-ordinate is advantageous in the treatment of clouds. So far in this model, weighting functions must be evaluated by perturbation methods.

DISORT has been assimilated to MODTRAN (versions 3.0 and higher), to ensure that the latter now has an adequate multiple scatter formalism. MODTRAN is a moderate resolution radiative transfer model with an extensive band model capability [R10]. It has replaced LOWTRAN in recent years, and is a widely-used standard. A major development for MODTRAN in 1998 will be the advent of correlated-k multiple scattering formalism (versions 4.0 and higher).

### 5.3 Limb Scattering Models

SCIAMACHY limb retrieval requires the forward model RT treatment of scattering for this viewing mode. Trace gas transmittances must be computed in a curved refracting atmosphere, both for the line-of-sight path and also for the solar path. Aerosol extinction must also be computed along these paths. The forward model must account for molecular (Rayleigh) and aerosol scattering, both single and multiple, including scattered light from the lowest layers of the atmosphere. This latter term will have a component from the surface reflection of direct and diffuse sunlight.

The proper RT treatment requires a fully spherical multiple scattering model. For all past and existing limb scatter remote sensing experiments, the approach has been to use a quasi-analytic single scattering model, and deal with multiple scattering in the form of correction factors. For the SOLSE/LORE experiment, the target species is O<sub>3</sub> in the UV/visible, and weighting functions are calculated according to:

$$W_{ms} = W_{ss} \cdot \left( 1 + \frac{I_{ms}}{I_{ss}} \right) \quad (42)$$

The single scatter results  $I_{ss}$  and  $W_{ss}$  come from the quasi-exact single scattering model (see for example [L18]).  $I_{ms}$  values are tabulated as corrections to the single scattering results, and they are computed using a state-of-the-art fully spherical model [R15]. This model is too slow for retrieval purposes, but it has been used to correct simulations for nadir-mode total column retrievals where the line-of-sight is significantly away from nadir [R16].

For SCIAMACHY, single scatter models have been developed at two institutions. The single scattering atmospheric source terms at level  $n$  in the atmosphere may be written schematically as:

$$S_n^\pm = \frac{1}{4\pi} \cdot I_0 \cdot \Omega_n^\pm \cdot \xi_n \quad (43)$$

Here,  $\pm$  refers to upwards and downward scattered light,  $I_0$  is the solar flux at the top of the atmosphere,  $\xi_n$  is the cumulative transmittance along the solar path, and  $\Omega_n^\pm$  are the single scatter albedos. In terms of the total limb path extinction  $\beta_n$ , the molecular and aerosol scattering coefficients  $\sigma_{Ray,n}$  and  $\sigma_{Aer,n}$ , and the phase functions  $P_{Ray}(\Theta_n^\pm)$  and  $P_{Aer}(\Theta_n^\pm)$  dependent on the scatter angles  $\Theta_n^\pm$ ,  $\Omega_n^\pm$  may be written:

$$\Omega_n^\pm = \frac{P_{Ray}(\Theta_n^\pm) \sigma_{Ray,n} + P_{Aer}(\Theta_n^\pm) \sigma_{Aer,n}}{\beta_n} \quad (44)$$

Work has now started on the production of a limb model for SCIAMACHY with a multiple scattering formalism.

## 6 Input/Output Requirements

This section is intended as a reference for the I/O requirements of the SGP<sub>12</sub> algorithms. Detailed definitions and product formatting can be found in the appropriate documents ([A10] for the off-line, and [A9] for the NRT). Here we shall list the main physical content of the input and output files.

### 6.1 Input Data

#### 6.1.1 Level 1b input

The following quantities in the extracted level 1b file will be used in level 1 to 2 processing. Detector measurements are always accompanied by PMD information (there are 7 PMDs for SCIAMACHY), and geolocation. Each detector spectrum contains an index to a set of spectral calibration coefficients, which are required to generate the pixel wavelength assignments.

- Mean sun reference spectrum (8 channels), plus radiometric precision
- Average sun mirror azimuth and elevation position
- Mean sun PMD reference measurements
- Scan spectrum arranged in clusters according to states [A6], including radiometric precision
- PMD sub-pixel measurements corresponding to these scan spectra
- Viewing geometry information (solar and line-of-sight zenith/azimuth angles at start/middle/end of integration time, earth radius, satellite height, sub-satellite position, scan mirror positions)
- Footprint location (nadir only; 4 corner and centre latitude/longitude co-ordinates)
- Tangent height information (limb/occultation only; tangent height and ground point co-ordinates at start/middle/end of integration time)
- Fractional cloud cover from PCCA cloud coverage algorithm (nadir only)

#### 6.1.2 Initialisation File

The following remarks give an indication of the parameters required for control of the algorithms in SCIAMACHY level 1 to 2 parameters. At the time of writing, a detailed parameter list for the Initialisation File has been drawn up for the NRT level 1 to 2 processor, but not so far for the off-line algorithms. Many of the parameter choices are for test purposes only, but until we have optimised the operational algorithms, it is diplomatic to include all options.

For the fitting algorithms, the selection of fitting window limits for whole or segmented windows (IR application will use non-contiguous segments) is of paramount importance. In limb, it will be necessary to select different window limits for each scan in a sequence, because the set of observations to be fitted will not necessarily be the same through the atmosphere. Other options might include the choice of slit function and whether the slit function half-width should be fitted or not, and so on.

For each fit, we must specify the geophysical parameters to be derived from the fitting, whether they are volume mixing ratios (profile retrieval), total column amounts (nadir column fitting), cloud variables (PCCA, RCFA), closure parameters (polynomial coefficients in UVAS and IAS), shift and squeeze parameters (DOAS), etc. For retrieval algorithms involving non-linear fitting we need to specify starting values for these geophysical parameters (initial guesses).

The use of reference data will require control. For instance in UVAS, we need to consider which reference spectra will be included in the fit, and what will be the options on smoothing, convolution, shift/squeeze limits for these reference spectra. Will we be using multiple scattering look-up corrections in the UV/visible limb profile algorithms, or will we run that algorithm in single scattering mode only? Will we include the AMF error computation in the UV/visible computation? There are many decisions to be made on algorithm control and it is vitally important to keep the options open during the development phase.

### 6.1.3 Ancillary input data

The requirement here is for additional databases that contain data pertinent to the actual geophysical scenario being processed. In order to use them in SGP level 1 to 2 processing, it is necessary to have dedicated auxiliary processing exterior to the main processing. The most obvious example is the generation of up-to-date atmospheric profiles to be used as first guess and/or *a priori* in limb temperature/pressure profile retrieval, and as accurate and representative reference values in height-resolved nadir profile retrieval. Such data would come normally from numerical weather prediction model output (ECMWF analysis fields, for example). An ancillary processor would have to assimilate this output on a regular basis and generate the necessary inputs for level 1 to 2 on an operational basis.

Another important source is CO<sub>2</sub> volume mixing ratio data. This is required for a number of applications in the infrared using CO<sub>2</sub> absorption to retrieve atmospheric information. Though the gas is uniformly mixed up to mesospheric levels, and the mean annual VMR is accurately known, there are seasonal and hemispheric variations. These variations can be as large as 5%, and are linked to the cycle of photosynthesis. The timing of the well-known CO<sub>2</sub> uptake varies considerably from year to year in the northern hemisphere for example. Clearly it is expedient to use assimilated data from an established source to ensure that those SCIAMACHY retrieval algorithms based on known CO<sub>2</sub> levels receive input information that is accurate in space and time. A second ancillary processor should be written for this purpose.

## 6.2 Product Parameters

### 6.2.1 Level 2 Cloud/Aerosol Product

The following is an initial list of suggested cloud/aerosol product output associated with the algorithms described in this version of the ATBD. Most of the geophysical parameters listed below come from the cloud and aerosol algorithms described in section 3.1 on page 21. The equivalent Lambertian reflecting height from the CO<sub>2</sub> nadir infrared fit (section 3.3 on page 42) is included here. Some climatology (ground pressure and albedo) is helpful for referencing the retrieved products.

- Cloud fractional coverage and error
- Number of PMD pixels clear/cloudy/partially cloudy or undetermined
- Cloud-top pressure and error
- Cloud optical depth and error
- Cloud type (bit-wise assignment on WMO scheme?)
- Ground pressure and surface albedo
- Cloud-top bi-directional reflectance and error (reference values)
- Height of CO<sub>2</sub> equivalent Lambertian reflecting surface, plus error
- Flag describing cloud output (which way through the algorithms!)
- Absorbing aerosol index (AAI)
- Diagnostic for AAI
- Flag describing aerosol output

### 6.2.2 Level 2 Nadir Column Product

The recommended policy here is to define one data set for each trace species retrieved. Given that UVAS and IAS algorithms are both variants on the least-squares fitting, it is possible to define a generic nadir column product comprising one data set per retrieved trace gas species. Each such data set will have a main product value, which will be the total vertical column amount of the trace gas, plus the retrieved error on this quantity.

The fitting may be linear or non-linear or both (for example, DOAS); selected diagnostic output is generated for each fit regardless of the linearity. We include all fitted parameters regardless of their physical usage, plus all fitted parameter errors, plus all cross-correlation entries, plus chi-square, RMS, goodness-of-fit, and number of iterations (non-linear least squares). If there are  $M$  fitted parameters, then there are  $M(M-1)/2$  independent entries in the cross-correlation matrix (by definition symmetric, this is just the total number of off-diagonal elements). For DOAS, which has a linear fit embedded in a non-linear algorithm, the values of chi-square and RMS are the same for both algorithms. No detailed spectral output will be given (see below). For the UVAS algorithm, additional AMF values will be specified to ground and cloud-top, with errors. The product may look like this:

- Main product value (total column) and error
- Number of linear and non-linear fitted parameters
- Linear fitted parameters and their errors
- Linear fit cross-correlation entries
- Non-linear fitted parameters and their errors
- Non-linear fit cross-correlation entries

- Chi-square, RMS, Goodness-of-fit, number of iterations (non-linear)
- (UVAS only) centre pixel AMF values to ground and cloud-top, plus errors
- Flag controlling output for this data set

### 6.2.3 Level 2 UV/Visible Nadir Height-resolved Product

The main geophysical retrieved parameter here will be a vector containing the profile of O<sub>3</sub> concentrations on a specified retrieval grid. Besides the usual solution variance/covariance and single-number diagnostic output, it is useful to generate two additional quantities for this product. These are (i) the averaging kernel for the O<sub>3</sub> profile entries (discounting the other fitted parameters), and (ii) the improvement in the profile entries over the *a priori*. With the additional constraint imposed by the use of *a priori*, the fitting functional has two parts (see section 3.4 on page 50, equation (32), and it is often useful to output partial RMS and chi-square values for these parts.

- Main product values (ozone concentrations) and errors, plus retrieval height grid
- Number of fitted parameters in state vector
- Fitted parameter values and their errors
- cross-correlation entries
- measurement Chi-square, RMS, Goodness-of-fit, number of iterations
- *a priori* Chi-square, RMS, Goodness-of-fit
- Flag controlling output for this data set

### 6.2.4 Level 2 Limb/Occultation product (suggested)

It is best to include complete profiles from a limb/occultation scan sequence. At the present stage of development, we confine the output to profiles of VMR, profiles of pressure and temperature on a fixed grid. The precision (relative error in %) for each retrieved quantity will be generated; cross-correlation values are not considered in the first version. The flag controlling output will give some indication of the retrieval strategy. Occultation products will be similar, but they will be specified on a different vertical grid. The issue of classification by window against molecule could be avoided by defining the limb product to include all VMR profiles and the associated limb-derived self-consistent P-T profiles. This has been done here in the following generic list.

- Retrieval height grid
- Number of separate fitting applications (including P-T)
- Number and list of trace species retrieved
- Fitted P-T values and their precisions
- For each gas, fitted VMR values and their precisions
- For each fitting application, flag controlling output for this data set

### 6.2.5 Detailed output for algorithm verification

In GDP, three additional output files may be created when the processor is executed in 'debug' or 'test' mode (this is in addition to the regular error log-file). These are the debug log file, the detailed scenario file and the detailed results file. The scenario file contains all pertinent information used in the set-up stages of the retrieval application. It includes all atmospheric profiles used in the calculation, all (convoluted) cross sections, all additional climatological information extracted from data sets (topography, surface albedo at various wavelengths, etc.). The scenario file also contains a complete statement of all the parameters input from the Initialisation File, plus

output of the level 1b geolocation and (window-buffered) measurement data used in the retrieval. Scenario files should be defined for each algorithm. Algorithms using optimal estimation should also include all *a priori* information in the detailed scenario file output.

The detailed results file should contain all pertinent *intermediate* output from an algorithm. In particular it is here that spectral output should be made available for verification purposes - this would include the vector of measured backscatter intensities, the vector of simulated intensities, the rest spectrum, and various fitted and reference optical densities, etc. Additional AMF output is also valuable (UVAS only). For algorithms using optimal estimation it is possible to output a number of additional diagnostics (contribution functions, averaging kernels, etc.). There should also be an option to include detailed output at each iteration of any iterative fitting process (this would not only include the spectral data, but also ongoing values of the state vector of parameters to be fitted, plus the current solution variance/covariance matrix).

### 6.3 Reference Data Bases

We confine ourselves here to a list of the main data sets. As mentioned in the overview, there are three classifications for reference data climatology, look-up tables, and orbit-processed data. This list is the current baseline.

#### 6.3.1 Climatology and Spectroscopic Databases

- (UV/visible) Cross sections for O<sub>3</sub>, NO<sub>2</sub>, BrO, OClO, SO<sub>2</sub>, ClO, H<sub>2</sub>CO, NO<sub>3</sub>, O<sub>2</sub>-O<sub>2</sub>
- (Near infrared) Line spectroscopic data for O<sub>2</sub>, H<sub>2</sub>O, CO, N<sub>2</sub>O, CH<sub>4</sub>, CO<sub>2</sub>
- Ring spectra, Fraunhofer reference spectrum, slit function parameters
- Topography, constant albedo, land surface types (global)
- Spectral surface albedo (complete SCIAMACHY range)
- Atmospheric profile databases for pressure, temperature, VMRs for O<sub>3</sub>, NO<sub>2</sub>, BrO, OClO, SO<sub>2</sub>, ClO, H<sub>2</sub>CO, NO<sub>3</sub>, H<sub>2</sub>O, CO, N<sub>2</sub>O, CH<sub>4</sub> and CO<sub>2</sub>. *A priori* profiles O<sub>3</sub>, aerosol and maybe other trace gases, plus error covariance
- Aerosol loading and optical properties
- Cloud optical and reflectance properties, ISCCP climatology
- PMD threshold data sets, cloud-free composite

#### 6.3.2 RT-derived look-up tables

- Fourier components of intensities for O<sub>3</sub> AMF at 325 nm and for NO<sub>2</sub> AMF at 437.5 nm, plus climatology derivatives. Alternatively - complete AMF databases for these two species (NRT).
- Ratios of multiple scatter intensities to single scatter intensities for use in limb O<sub>3</sub> VMR profile retrieval (TBD).
- Intensity and weighting function look-up table for nadir O<sub>3</sub> profile retrieval (suggested).

#### 6.3.3 Orbit pre-processed data sets

- For each wavelength registration, pre-computed slit function quantities to be used in all convolutions anticipated for all retrieval applications.
- Others TBD.

## 7 Appendices

### 7.1 Convolution

The formal definition of the convolution of  $Y(x)$  with slit function  $S(x)$  is defined as:

$$Y_c(x) = \frac{1}{2\Delta} \int_{x-\Delta}^{x+\Delta} Y(x')S(x-x')dx' \quad (45)$$

The slit function is assumed symmetric about the central wavelength, and the band-pass is  $2\Delta$ . Although the slit function is parameterised as an analytic function of wavelength, in general the spectrum  $Y(x)$  is not analytic, and is specified on a wavelength ( $x_k$ ). The integral is then replaced by the trapezoidal sum:

$$Y_c(x_i) = \frac{\sum_{k=k_1}^{k_2} Y(x'_k) \cdot S_{\underline{a}}(x_i - x'_k) \cdot dx'_k}{\sum_{k=k_1}^{k_2} S_{\underline{a}}(x_i - x'_k) \cdot dx'_k} \quad (46)$$

$\underline{a}$  is the vector of parameters determining the analytic shape of the slit function.  $dx_k$  are the intervals in wavelength space given by:

$$dx_{k_1} = x_{k_1+1} - x_{k_1} \quad dx_{k_2} = x_{k_2} - x_{k_2-1} \quad dx_k = x_{k+1} - x_{k-1} \quad (47)$$

The number of points runs from  $k_1$  to  $k_2$ ; these depend on the wavelength  $x_i$ . We denote  $\{x'_k\}$  as the incoming wavelength grid. We will be convoluting to a group of wavelengths  $\{x_i\}$  (the outgoing wavelength grid), and we use symbols  $k_1(i)$  and  $k_2(i)$  to indicate this dependence on the outgoing wavelengths. In general we will not have regularly spaced wavelength grids, and this means the summation integrals must be done from first principles. The slit function shape is also dependent on  $x_i$ , so we write  $\underline{a}(i)$  to indicate the wavelength dependence of the slit function parameters. The slit function expressions are given in terms of pixel distance  $p_{ik}$ , rather than wavelength distance. The relation between these two is the pixel spacing  $R(x_i)$ , which also changes with wavelength across any channel. Thus  $x_i - x'_k = p_{ik}R(x_i)$ .

The range  $k_1(i)$  to  $k_2(i)$  depends on  $\Delta$ , the half-width slit band-pass (in pixel numbers). In fact, the range of  $k$ -values must lie between  $x_{lower} = x_i - \Delta R(x_i)$  and  $x_{upper} = x_i + \Delta R(x_i)$ . In general the spectrum to be convoluted is a large array with many points and a fine resolution in wavelength space. The set  $\{x'_k\}$  must include and extend beyond the set  $\{x_i\}$ . It is clear that the extreme wavelengths  $x_{lower, ext} = x_1 - \Delta R(x_1)$  and  $x_{upper, ext} = x_N + \Delta R(x_N)$  must lie inside the set  $\{x'_k\}$ . Here,  $x_1$  and  $x_N$  are the first and last wavelengths to which the convolution will be done. For each  $x_i$  we must not only trawl through the set  $\{x'_k\}$  to find the summation limits

$k_1(i)$  and  $k_2(i)$ , but also extract the correct pixel resolution  $R(x_i)$  and the correct vector of slit parameters  $\mathbf{a}(i)$  from the slit function database.

We write the convolution as  $Y_c(x_i) = \sum Y(x'_k)Q_i(x'_k)$  where the sum is between  $k_1(i)$  and  $k_2(i)$ . The quantities  $Q_i(x'_k)$  are then just normalised slit function values, and can be pre-computed independently of any spectrum  $Y$  that needs convolution. This slit function pre-computation should be done at the beginning of any orbit for each of the wavelength calibrations to be encountered during that orbit. The convolution itself is then just a simple summation with a minimum number of operations. Considerable time can be saved by pre-computing the variables  $Q_i(x'_k)$ , and knowing the range of values over which the summation can be done. More time can be saved in the set-up by starting the search for the offset  $k_1(i+1)$  with the already-found value  $k_1(i)$  instead of starting at  $k_1 = 1$  for each point.

At the time of writing, there are some preliminary measurements of SCIAMACHY slit functions. We must wait until the calibration is completed before a proper data base of slit function parameters can be compiled. In any case, it may be sufficient to assume the shape is known, and try to fit for one or more of the slit function parameters (main candidate for this would be the half-width). In the IAS prototype development, a Gaussian form  $S(x) = \exp[-0.693(x/w)^2]$  has been adopted. Here,  $w$  is the full width half maximum. Another possibility is the simple hyperbolic function used in GDP, namely  $S(x) = a_0^2/(x^4 + a_1^2)$ . All slit functions are expressed in terms of pixel numbers from the centre; all should have unit area.

## 8 References

### Cloud/aerosol pre-processing references (section 3.1)

- [C1] Kuze, A., and K. V. Chance, Analysis of Cloud-Top Height and Cloud Coverage from Satellites Using the O<sub>2</sub> A and B Bands, *J. Geophys. Res.*, **99**, 14481-14491 (1994)
- [C2] Loyola, D., A New Cloud Recognition Algorithm for Optical Sensors, Proceedings of the IGARSS Conference, Seattle, July 1998
- [C3] Koelemeijer, R. B. A., P. Stammes and J. A. Konings, Validation of GOME cloud cover fraction relevant for accurate ozone retrieval, Proc. 3220 EOS/SPIE Satellite remote sensing of clouds and the atmosphere, September 1997, in press
- [C4] Mathews, E., Vegetation, land-use and albedo data sets, Technical Memorandum #86107, NASA (1984)
- [C5] Bowker, D. et. al., Spectral reflectances of natural targets for use in remote sensing studies, NASA report NASA-RP-1139 (1985)
- [C6] Saunders, R. W., and K. T. Kriebel, An improved method for detecting clear sky and cloudy radiances from AVHRR data, *Int. J. Remote Sens.*, **9**, 123-150 (1988)
- [C7] Saiedy, F., H. Jacobowitz, and D. Q. Wark, On cloud-top determination from Gemini-5, *J. Atmos. Sciences*, **24**, 63-69 (1967)
- [C8] Wu, M. L. C., Remote Sensing of Cloud-top Pressure Using Reflected Solar Radiation in the Oxygen A-band, *J. Clim. & Appl. Met.*, **24**, 539-546 (1985)
- [C9] King, M. D., Determination of the Scaled Optical Thickness of Clouds from Reflected Solar Radiation Measurements, *J. Atmos. Sciences*, **44**, 1734-1751 (1987)
- [C10] Stephens, G. L., Optical Properties of 8 Water Cloud Types, Division of Atmospheric Physics, CSIRO, Technical Paper No. 36 (1979)
- [C11] Rossow, W., A. Walker, D. Beuschel, and M. Roiter, International satellite cloud climatology project (ISCCP) documentation of new cloud data sets, Technical Report WMO/TD-No.737, World Meteorological Organization (1996)
- [C12] Chance, K., Improvement of the O<sub>2</sub> A band spectroscopic database for satellite-based cloud detection, *J. Quant. Spectrosc. Radiat. Transfer*, **58**, 375-378 (1997)
- [C13] Newnham, D., and J. Ballard, Visible absorption cross-sections and integrated absorption intensities of molecular oxygen (O<sub>2</sub> and O<sub>4</sub>), submitted to *J. Geophys. Res.* (March 1998)
- [C14] Hsu N. C., J. R. Herman, P. K. Bhartia, C. J. Seftor, O. Torres, A. M. Thompson, J. F. Gleason, T. F. Eck and B. N. Holben, Detection of biomass burning smoke from TOMS measurements, *Geophys. Res. Lett.*, **23**, 745-748 (1996)
- [C15] Torres, O., P. K. Bhartia, J. R. Herman, Z. Ahmad, Derivation of aerosol properties from satellite measurements of back-scattered ultraviolet radiation: theoretical basis, Proc. IRS'96: Current problems in atmospheric radiation, A. Deepak Publishing, Hampton, VA, 534 (1997)
- [C16] Fish, D., and R. Jones, Rotational Raman scattering and the Ring effect in zenith-sky spectra, *Geophys. Res. Lett.*, **22**, 711-814 (1995)
- [C17] Levoni, C., M. Cervino, R. Guzzi and F. Torricella, Atmospheric aerosol optical properties: a database of radiate characteristics for different components and classes, *Applied Optics*, **36**, 8031-8041 (1997)
- [C18] World Climate Research Program (WCP), A preliminary cloudless standard atmosphere for radiation computation, WCP-112, Boulder, Co., USA (1986)

- [C19] Nakajima, T. and M. Tanaka, Algorithms for radiate intensity calculations in moderately thick atmospheres using a truncation approximation, *J. Quant. Spectrosc. Radiat. Transfer*, **40**, 51-69 (1988)
- [C20] C. Levoni et. al., presented at IGARSS 98, Seattle, July 1998

#### **GOME project documentation and study reports**

- [G1] GOME Interim Science Report, edited by T. D. Guyenne and C. J. Readings, SP-1151, ESA publications Division, ESTEC, Noordwijk, The Netherlands, ISBN 92-9092-041-6 (1993)
- [G2] GOME Users Manual, ESA SP-1182, ESA/ESTEC, Noordwijk, The Netherlands (1996)
- [G3] GOME Level 1 to 2 Algorithm Description, ER-TN-DLR-GO-0025, Issue 2, August 1996.
- [G4] GOME Software Databases for Level 1 to 2 Processing, ER-TN-IFE-GO-0018, Issue 2, October 1993
- [G5] A Study of the Method for Retrieval of Atmospheric Constituents, ESA/SERCO Final Report, 9687/91/NL/BI, July 1994
- [G6] Detailed Analysis of the Retrieval Algorithms Selected for the Level 1-2 Processing of GOME data, Final Report, ESA/SERCO 10728/94/NL/CN, July 1995
- [G7] Design, development and upgrading of prototype software tools for use with the ESRIN atmospheric reference processor, Final Report, ESRIN/Bremen study 12030/96/I-HE, January 1998
- [G8] A Study on the effects of Scattering on the Monitoring of Atmospheric Constituents, volume 2: Albedo and Cloud Studies, ESA/ESTEC Final Report, 9740/91/NL/BI, 1994.
- [G9] A Study of Cloud Detection, ESA/SERCO Final Report, 10997/94/NL/CN, December 1995.
- [G10] GOME Cloud and Aerosol Data Products Algorithms Development (CADAPA) Study, Final Report, 11572/95/NL/CN, February 1998.
- [G11] Proposal to the Remote Sensing Data Centre for Cloud Retrieval for GOME Development, P4198-12-97, December 1997
- [G12] Study of the Ring effect, Final Report, ESA/SERCO 10996/94/NL/CN, 1996

#### **Limb/occultation references (section 4)**

- [L1] MIPAS Instrument specification, PO-RS-DOR-MP-001, Issue 2A
- [L2] SOLSE/LORE, Shuttle Ozone limb Sounding Experiment, Limb Ozone Retrieval Experiment, NASA/GSFC, launched November 1997
- [L3] K. Strong, Retrieval of Vertical concentration profiles from ODIN/OSIRIS UV-visible limb spectra: current status, GOME and SCIAMACHY Working Session 6, Brussels, September 1997
- [L4] Chahine, M., A general relaxation method for inverse solution of the full radiative transfer equation, *J. Atmos. Science*, **29**, 741-747 (1972)
- [L5] Gille, J. C., and F. B. House, On the inversion of limb radiance measurements: I. Temperature and thickness, *J. Atmos. Science*, **28**, 1427-1442 (1971)
- [L6] Norton, R. H., and C. P. Rinsland, ATMOS data processing and science analysis methods, *Applied Optics*, **30**, 389-400 (1991)
- [L7] Wetzal, G., H. Oelhaf, T. von Clarmann, H. Fischer, F. Friedl-Vallon, G. Maucher, M.

- Seefeldner, and O. Trieschmann, Vertical profiles of  $N_2O_5$ ,  $HO_2NO_2$  and  $NO_2$  inside the arctic vortex, retrieved from nocturnal MIPAS-B2 infrared limb emission measurements in February 1995, *J. Geophys. Res.*, **102**, 19377-19386 (1997)
- [L8] Carlotti, M., Global-fit approach to the analysis of limb-scanning atmospheric measurements, *Applied Optics*, **27**, 3250-3254 (1988)
- [L9] Development of an optimized algorithm for routine P, T and VMR retrieval from MIPAS limb emission spectra, Technical Note on algorithm definition, and physical and mathematical optimizations, TN-IROE-RSA9601, Issue 1, June 1996
- [L10] SAGE III ATBD Ozone data products, Version 1.1, November 1997
- [L11] Rinsland, C. P., M. R. Gunson, R. Zander, and M. Lopez-Puertas, Middle and upper atmosphere pressure-temperature profiles and the abundances of  $CO_2$  and CO in the upper atmosphere from ATMOS/Spacelab 3 observations, *J. Geophys. Res.*, **97**, 20479- 20495 (1992)
- [L12] Chu, W. P., M. McCormick, J. Lenoble, C. Brogniez, and P. Pruvost, SAGE II inversion algorithm, *J. Geophys. Res.*, **94**, 8339-8351 (1989)
- [L13] Stiller, G. P., M. Gunson, L. Lowes, M. Abrams, O. Raper, C. Farmer, R. Zander, and C. Rinsland, Stratospheric and mesospheric pressure-temperature profiles from rotational analysis of  $CO_2$  lines in atmospheric trace molecule spectroscopy/ATLAS 1 infrared solar occultation spectra, *J. Geophys. Res.*, **100**, 3307-3117 (1995)
- [L14] Schmidt, U., and A. Khedim, In situ measurements of carbon dioxide in the winter arctic vortex and at mid-latitudes: An indicator of the age of stratospheric air, *Geophys. Res. Lett.*, **18**, 763-766 (1991)
- [L15] Burrows, J., V. Rozanov, Y. Timofeyev, A. Polyakov, R. Spurr, and K. Chance, A study of the accuracy of atmospheric trace gas vertical profile retrieval from satellite-based occultation measurements, IRS conference on current problems in atmospheric radiation, Keevalik & Kaerner (eds.), DEEPAK publishing, ISBN 0-937194-28-X, 398-400 (1992)
- [L16] tbd1
- [L17] tbd2
- [L18] Aruga and Heath
- [L19] Bates, D. R., Rayleigh Scattering by Air, *Planet. Space. Sci.* **32**, 785 (1984)
- [L20] A preliminary cloudless standard atmosphere for radiation computation. Tech. Rep. WCP-112, WMO/TD-NO.24, World Meteorological Organization, March 1986
- [L21] Shettle, E. P. and R. W. Fenn, Models of the aerosol of the lower atmosphere and the effects of humidity variations on their optical properties, Tech. Rep. AFGL-TR-790214, Air Force Geophysical Laboratory, Hanscom Air Force Base, Massachusetts, USA (1979)

### IAS references (section 3.3)

- [N1] Johnson, D. G., K. W. Jucks, W. A. Traub, and K. V. Chance, The Smithsonian Stratospheric Far-Infrared Spectrometer and Data Reduction System, *J. Geophys. Res.*, **100**, 3091-3106 (1995)
- [N2] Anderson, G. P., S. A. Clough, F. X. Kneizys, J. H. Chtewynd, and E. P. Shettle, AFGL atmospheric constituents profiles (0-120 km), AFGL, Hanscom, MA, Report AFGL-TR-86-1001, AD175173 (1986)
- [N3] Thoning, K. W., P. P. Tans, and W. D. Komhyr, Atmospheric carbon dioxide at Mauna Loa observatory. 2. Analysis of the NOAA GMCC data, 1974-1985, *J. Geophys. Res.*, **94**,

- 8549-8565 (1986)
- [N4] Spurr, R. J. D., and K. Chance, BIAS: An algorithm for the retrieval of trace gas vertical column amounts from near-infrared earthshine measurements by the SCIAMACHY spectrometer, Proceedings of the IGARSS conference, Seattle, July 1998
  - [N5] Schrijver, H., S. Slijkhuis, M. G. Roemer, and A. P. H. Goede, Noise-related limits on the detectability of concentration variations of CH<sub>4</sub> and CO with SCIAMACHY, Atmospheric Sensing and Modeling, ed. R. P. Santer, Proc SPIE 2311, 39 (1995). Note: Runs have recently been updated by H. Schrijver, private communication (1997)
  - [N6] SCIAMACHY Instrument Simulation Software, SRON, Utrecht (1996)
  - [N7] L. Rothman et. al., The HITRAN molecular spectroscopic database and HAWKS (HITRAN Atmospheric Work-station, J. Quant. Spectrosc. Radiat. Transfer, in press (1998)
  - [N8] Humlicek, J., Optimized computation of the Voigt and complex probability functions, J. Quant. Spectrosc. Radiat. Transfer, **27**, 437-444 (1982)
  - [N9] Schreier, F., The Voigt and complex error function: a comparison of computational methods, J. Quant. Spectrosc. Radiat. Transfer, **48**, 743-762 (1992)
  - [N10] Sparks, L., Efficient line-by-line calculation of absorption coefficients to high numerical accuracy, J. Quant. Spectrosc. Radiat. Transfer, **57**, 631-650 (1997)
  - [N11] Kuntz, M., A new implementation of the Humlicek algorithm for the calculation of the Voigt profile function, J. Quant. Spectrosc. Radiat. Transfer, **57**, 819-824 (1997)
  - [N12] Kuntz, M., 20th Annual Review Conference on Atmospheric Transmission Models, Phillips Laboratory, Bedford, MA, June 1997

#### **HRFA (nadir O3 profile) references (section 3.4)**

- [P1] Rodgers, C. D., Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation, Rev. Geophys. Space Phys., **14**, 609-624 (1976)
- [P2] Rodgers, C. D., Characterization and Error Analysis of Profiles Retrieved from Remote Sounding Experiments, J. Geophys. Res., **95**, 5587 (1990)
- [P3] Spurr, R. J. D., Development of a prototype algorithm for the operational retrieval of height-resolved products from GOME, Proceedings of the 3rd ERS Symposium Space at the Service of our Environment, SP-414, 621-628, Florence, Italy, March 1997
- [P4] deBeek, R., R. Hoogen, V. Rozanov, and J. Burrows., Ozone profile retrieval from GOME satellite data 1. Algorithm Description, Proceedings of the 3rd ERS Symposium Space at the Service of our Environment, SP-414, 749-754, Florence, Italy, March 1997
- [P5] Heath, D., A. J. Krueger, H. A. Roeder, and B. D. Henderson, The Solar Backscatter Ultraviolet and Total Ozone Mapping Spectrometer (SBUV/TOMS) for Nimbus G, Opt. Eng., **14**, 323-331 (1975)
- [P6] Bhartia, P. K., R. D. McPeters, C. L. Mateer, L. E. Flynn, C. Wellemeyer, Algorithm for the Estimation of Vertical Ozone Profiles from the Backscattered Ultraviolet Technique, J. Geophys. Res., **101**, 18793-18806 (1996)
- [P7] Chance, K., J. P. Burrows, D. Perner, and W. Schneider, Satellite measurements of atmospheric ozone profiles, including tropospheric ozone, from UV/visible measurements in the nadir geometry: a potential method to retrieve tropospheric ozone, J. Quant. Spectrosc. Radiat. Transfer, **57**, 467-476 (1997)
- [P8] McPeters, R. D., T. Miles, L. E. Flynn, C. G. Wellemeyer, and J. M. Zawodny, Compari-

son Of SBUV And SAGE II Ozone Profiles: Implications For Ozone Trends, *J. Geophys. Res.*, **99**, 20513-20524 (1994)

- [P9] Siddans, R., W. J. Reburn, B. J. Kerridge, and R. Munro, Height-resolved ozone information in the troposphere and lower stratosphere from GOME, Proceedings of the 3<sup>rd</sup> ERS Symposium Space at the Service of our Environment, SP-414, 615-620, Florence, Italy, March 1997
- [P10] Munro, R., R. Siddans, W. J. Reburn, and B. J. Kerridge, Direct measurement of tropospheric ozone distributions from space, *Nature*, **392**, 168-171 (1998)
- [P11] McPeters, R. D., Climatology of Nitric Oxide in the Upper Stratosphere, Mesosphere and Thermosphere: 1979-1986, *J. Geophys. Res.*, **94**, 3461 (1989)
- [P12] ECMWF reference

#### **Radiative transfer references (section 5)**

- [R1] Chandrasekhar, S., Radiative Transfer, Dover, New York, 1960
- [R2] Lenoble, J., Radiative Transfer in Scattering and Absorbing Atmospheres, Deepak publishing, Hampton, VA, USA (1985)
- [R3] Barkstrom, B., A finite difference method of solving anisotropic scattering problems, *J. Quant. Spectrosc. Radiat. Transfer*, **16**, 725-739 (1976)
- [R4] Rozanov, V., D. Diebel, R. Spurr and J. Burrows, GOME-TRAN: A Radiative Transfer Model for the Satellite Project GOME - the Plane Parallel Version, *J. Geophys. Res.*, **102**, 16683 (1997)
- [R5] Kurosu, T., V. Rozanov, and J. Burrows, 'Parameterization schemes for terrestrial water clouds in the radiative model GOMETRAN, *J. Geophys. Res.*, **102**, 21809-21823 (1997)
- [R6] Vountas, M., V. Rozanov and J. P. Burrows, Ring effect: impact of rotational Raman scattering on radiate transfer in earth atmosphere, *J. Quant. Spectrosc. Radiat. Transfer*, in press (1998)
- [R7] Rozanov, V., T. Kurosu and J. Burrows, Retrieval of atmospheric constituents in the UV/visible: A new quasi-analytical approach for the calculation of weighting functions, *J. Quant. Spectrosc. Radiat. Transfer*, in press (1998)
- [R8] de Haan, J. F., P. B. Bosma and J. W. Hovenier, The adding method for multiple scattering calculations of polarized light, *Astron. & Astrophys.*, **183**, 371-391 (1987)
- [R9] Perliski, L. M., and S. Solomon, On the evaluation of air mass factors for atmospheric near-ultraviolet and visible absorption spectroscopy, *J. Geophys. Res.*, **98**, 10363-10374 (1993)
- [R10] Bernstein, L. S., A. Berk, D. C. Robertson, P. K. Acharya, G. P. Anderson, J. H. Chetwynd, Addition of a Correlated-k Capability to MODTRAN, Proceedings of the 19th Annual Conference on Atmospheric Transmission Models, Phillips Laboratory, Geophysics Directorate, June 1996.
- [R11] Stamnes, K., S.-C. Tsay, W. Wiscombe, and K. Jayaweera, Numerically stable algorithm for discrete ordinate method radiative transfer in multiple scattering and emitting layered media, *Applied Optics*, **27**, 2502-2509 (1988)
- [R12] Wiscombe, W., The delta-M method: Rapid yet accurate radiate flux calculations for strongly asymmetric phase functions, *J. Atmos. Sciences*, **34**, 1408-1422 (1977)
- [R13] Dlugach, J. M., and E. G. Yanovitskij, The Optical Properties of Venus and the Jovian Planets. II. Methods and results of Calculations of the Intensity of Radiation Diffusely

Reflected from Semi-infinite homogeneous Atmospheres, *Icarus*, **22**, 66-81 (1974)

- [R14] van de Hulst, H. C., Asymptotic Fitting. A method for solving anisotropic transfer problems in thick layers, *J. Comp. Phys.*, **3**, 291-306 (1968)
- [R15] Herman, B. M., A. Ben-David, and K. J. Thome, Numerical technique for solving the radiative transfer equation in a spherical-shell atmosphere, *Applied Optics*, **33**, 1760-1770 (1994)
- [R16] Caudill, T. R., D. E. Flittner, B. M. Herman, O. Torres, and R. D. McPeters, Evaluation of the pseudo-spherical approximation for back-scattered ultraviolet radiances and retrieval, *J. Geophys. Res.*, **102**, 3881-3890 (1997)

### UVAS references (section 3.2)

- [U1] Noxon J. F., Nitrogen Dioxide in the stratosphere and troposphere as measured by ground-bases absorption spectroscopy, *Science*, **189**, 547 (1975)
- [U2] Platt U., D. Perner and H.W. Paetz, Simultaneous Measurement of atmospheric H<sub>2</sub>CO, O<sub>3</sub>, and NO<sub>2</sub> by differential optical absorption, *J. Geophys. Res.*, **84**, 6329-6335 (1979)
- [U3] Platt, U., Differential optical absorption spectroscopy (DOAS), in *Air Monitoring by Spectroscopic techniques*, ed. M. Sigrist, *Chem. Anal. Ser.*, **127**, 27-84 (1994)
- [U4] Marquardt, D. L., An algorithm for least squares estimation of nonlinear parameters, *J. Soc. Indust., App. Math.*, **2**, 431-441 (1963)
- [U5] Harder, J. W., R. O. Jakoubek, and G. H. Mount, Measurement of tropospheric trace gases by long-path differential absorption spectroscopy during the 1993 OH Photochemistry Experiment, *J. Geophys. Res.*, **102**, 6215-6226 (1997)
- [U6] Chance K., BrO Measurements from GOME. I. Method of Spectroscopic Analysis, submitted to *Geophys. Res. Lett.*, May 1998
- [U7] Chance K., and R. J. D. Spurr, Ring Effect Studies: Rayleigh Scattering, Including molecular Parameters for Rotational Raman Scattering, and the Fraunhofer Spectrum, *Applied Optics*, **36**, 5224-5230 (1997)
- [U8] 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. Wolfmueller, Ground Segment for ERS-2 GOME Sensor at the GERMAN D-PAF, *Proceedings of the 3<sup>rd</sup> ERS Symposium Space at the Service of our Environment*, SP-414, 591-596, Florence, Italy, March 1997
- [U9] Thomas, W., E. Hegels, S. Slijkhuis, R. Spurr, and K. Chance, Detection of trace species in the troposphere using back-scatter spectra obtained by the GOME spectrometer, *Geophys. Res. Lett.*, **25**, 1317-1320 (1998)
- [U10] Perner, D., T. Klüpfel, E. Hegels, P. J. Crutzen, and J. P. Burrows, First results on tropospheric observations by GOME, *Proceedings of the 3<sup>rd</sup> ERS Symposium Space at the Service of our Environment*, SP-414, 647-652, Florence, Italy, March 1997
- [U11] Hegels, E., P. J. Crutzen, T. Klüpfel, D. Perner, J. P. Burrows, A. Ladstätter-Weissenmayer, M. Eisinger, J. Callies, A. Hahne, K. Chance, U. Platt and W. Balzer, Satellite measurements of halogen oxides by GOME: Distribution of OClO and BrO and comparison with ground-based observations, *XVIII Quadrennial Ozone Symposium*, Vol. Abstracts, p 229, L'Aquila, Italy, September 1996
- [U12] Bass, A. M., and R. J. Paur, The UV cross-sections of ozone: 1. Measurements in atmospheric ozone, *proceedings of the quadrennial ozone symposium*, Halikidi, Greece, ed. C.

- Zeferos and A. Ghaz, pp. 606-616 (1985)
- [U13] Johnston, H. S., Ozone absorption cross-sections in the Chappuis bands, unpublished results, Lawrence Livermore Laboratory (1990)
- [U14] Shettle, E. P., and S. M. Anderson, New visible and near IR ozone cross sections for MODTRAN, 17<sup>th</sup> Annual Review Conference on Atmospheric Transmission Models, Phillips Laboratory, Bedford, MA, June 1994
- [U15] J. P. Burrows, 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<sub>3</sub> in the 231-794 nm range, *J. Quant. Spectrosc. Radiat. Transfer*, in press (1998)
- [U16] Schneider, W., G. K. Moortgat, G. S. Tyndall, and J. P. Burrows, Absorption cross-sections of NO<sub>2</sub> in the UV and visible region (200-700 nm) at 298K, *J. Photochem. Photobiol.*, **40A**, 195-217 (1987)
- [U17] J. P. Burrows, 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<sub>2</sub> in the 231-794 nm range, *J. Quant. Spec. Rad. Transfer*, in press (1998)
- [U18] Harder, J. W., J. W. Brault, P. V. Johnston, and G. H. Mount, Temperature dependent NO<sub>2</sub> cross-sections at high spectral resolution, *J. Geophys. Res.*, **102**, 3861-3879 (1997)
- [U19] Yoshino, K., J. R. Esmond, and W. H. Parkinson, High resolution absorption cross-section measurements of NO<sub>2</sub> in the UV and visible region, *J. Chem. Phys.*, in press, (1998)
- [U20] Kirmse, B., A. Delon, and R. Just, NO<sub>2</sub> absorption cross-section and its temperature dependence, *J. Geophys. Res.*, **102**, 16089-16098 (1997)
- [U21] Solomon, S., A. L. Schmeltekopf, and R. W. Sanders, On the interpretation of zenith sky absorption measurements, *J. Geophys. Res.*, **92**, 8311-8319 (1987)
- [U22] Wahner, A., A. R. Ravishankara, S. P. Sander, and R. R. Friedl, Absorption cross-section of BrO between 312 and 385 nm at 298 and 223 K, *Chem. Phys. Lett.*, **152**, 507-512 (1988)
- [U23] Hearn, C. H., and J. A. Joens, The near UV absorption spectrum of CS<sub>2</sub> and SO<sub>2</sub> at 300K, *J. Quant. Spectrosc. Radiat. Transfer*, **45**, 69-75 (1991)
- [U24] Simon, F. G., W. Schneider, G. K. Moortgat, and J. P. Burrows, A study of the ClO absorption cross-section between 240 and 310 nm and the kinetics of the self-reaction at 300K, *J. Photochem. Photobiol.*, **55A**, 1-23 (1990)
- [U25] Wahner, A., G. S. Tyndall, and A. R. Ravishankara, Absorption cross-sections of OClO as a function of temperature in the wavelength range 240-480 nm, *J. Phys. Chem.*, **91**, 2734-2738 (1987)
- [U26] Meller, R., Absorption cross-sections of formaldehyde, Max Planck Institute for Chemistry, Mainz, private communication (1991)
- [U27] Greenblatt, G. D., J. J. Orlando, J. B. Burkholder, and A. R. Ravishankara, Absorption measurements of oxygen between 330 and 1140nm, *J. Geophys. Res.*, **95**, 18577-18582 (1990)
- [U28] Wellemayer, C. G., S. L. Taylor, C. J. Seftor, R. D McPeters and P. K. Bhartia, A Correction for TOMS Profile Shape Errors at High Latitude, *J. Geophys. Res.*, **102**, 9029- 9038 (1997)

- [U29] Russell, J. M., L. Gordley, J. H. Park, S. R. Drayson, W. D. Hesketh, R. J. Cicerone, A. F. Tuck, J. E. Frederick, J. E. Harries, and P. J. Crutzen, The Halogen Occultation Experiment, *J. Geophys. Res.*, **98**, 10777-10797 (1993)
- [U30] Nimbus-7 Total Ozone Mapping Spectrometer (TOMS) Data Products User's Guide, NASA Reference Publication (1996)