

Technical Note on:

High level algorithm definition and physical and mathematical optimisations

(MIPAS Level 2 Algorithm Theoretical Baseline Document)

25 June 2001

Issue 3

(DRAFT)

(Compliant with ORM_ABC Vers. 1.2.2)

Delivery of the study:

**"Development of an Optimised Algorithm for Routine P, T and VMR Retrieval from MIPAS
Limb Emission Spectra"**

Prepared by:

Name	Institute
M. Carlotti	University of Bologna
M. Höpfner	IMK
P. Raspollini	IROE-CNR
M. Ridolfi	University of Bologna

Approved by:

Name	Institute
B. Carli	IROE-CNR

TABLE OF CONTENTS

1 - INTRODUCTION	5
<i>1.1 Changes with respect to Issue 2A of the present document</i>	5
2 - OBJECTIVES OF THE TECHNICAL NOTE	6
3 - CRITERIA FOR THE OPTIMISATION	6
4 - THE MATHEMATICS OF THE RETRIEVAL PROBLEM	8
<i>4.1 Mathematical conventions</i>	8
<i>4.2 Theoretical background</i>	8
4.2.1 The direct problem	8
4.2.2 The Gauss Newton method	9
4.2.3 The Marquardt method	11
4.2.4 Review of the possible convergence criteria	11
4.2.5 Use of external (a-priori) information in the inversion model	12
4.2.6 Use of the optimal estimation, for inclusion of LOS engineering information in p,T retrieval	13
<i>4.3 The global fit analysis</i>	16
<i>4.4 High level mathematics of the forward model</i>	17
4.4.1 The radiative transfer	17
4.4.2 Convolution with the AILS	21
4.4.3 Convolution with the FOV	21
4.4.4 Instrumental continuum	21
4.4.5 Summary of required variables	21
<i>4.5 Calculation of the VCM of the measurements</i>	22
4.5.1 Operations performed on the interferogram to obtain the apodised spectrum	22
4.5.2 Computation of the VCM relative to a single microwindow	24
4.5.3 Computation of the inverse of the VCM	26
<i>4.6 Calculation of the Jacobian matrix K of the simulations</i>	27
<i>4.7 Generalised inverse</i>	28
<i>4.8 Variance-covariance matrix of tangent heights corrections</i>	28
5 - SCIENTIFIC ASPECTS AND PHYSICAL OPTIMISATIONS	31
<i>5.1 Choice between retrieval of profiles at fixed levels and at tangent altitude levels</i>	31
5.1.1 Retrieval at tangent altitude and interpolation between retrieved values	31
5.1.2 Retrieval at fixed levels	32
5.1.3 Discussion of the problem	32
5.1.4 Conclusions	33
<i>5.2 Use of a-priori information</i>	33
5.2.1 - Accuracy improvement	33
5.2.2 - Systematic errors	34
5.2.3 - Hydrostatic equilibrium and LOS Engineering information	35

5.3 Latitude effects	36
5.3.1 Angular spread	36
5.3.2 Climatological differences	36
5.4 Earth model and gravity	37
5.4.1 Earth model	37
5.4.2 Gravity	37
5.5 Ray-tracing and refractive index	39
5.5.1 Pressure-temperature dependence of refractive index and effects on tangent heights	39
5.5.2 Method of ray tracing	42
5.6 Line shape modelling	43
5.6.1 Numerical calculation of the Voigt profile	43
5.6.2 Approximation of the Voigt profile by the Lorentz function	44
5.6.3 χ -factors in the case of CO ₂ and H ₂ O	44
5.7 Line-mixing	44
5.8 Pressure shift	45
5.9 Implementation of Non-LTE effects	45
5.10 Self broadening	46
5.11 Continuum	46
5.11.1 Instrumental continuum	46
5.11.2 Near continuum	47
5.11.3 Far continuum	47
5.12 Interpolation of the profiles in the forward / retrieval model	49
5.13 Interpolation of the “retrieved” profiles to a user-defined grid	50
5.14 Optimized algorithm for construction of initial guess profiles	52
5.14.1 Generation of initial guess continuum profiles	55
5.15 Profiles regularization	56
6 - MATHEMATICAL OPTIMISATIONS	58
6.1 Radiative Transfer integral and use of Curtis-Godson mean values	58
6.1.1 Layering of the atmosphere	61
6.2 Secant law approximation for the calculation of Curtis-Godson quantities and definition of paths	64
6.2.2 Sequence of the operations	67
6.3 Interpolation of cross sections for different geometries	68
6.4 Calculation of spectrum: exploitation of spherical symmetries	69
6.5 Use of interpolation for the calculation of Planck function	69
6.6 Finite instrument field of view.	71
6.7 Analytical derivatives	76
6.7.1 General considerations	76
6.7.2 Derivative with respect to the volume mixing ratio	78
6.7.3 Derivative with respect to temperature	79

6.7.4 Derivative with respect to the atmospheric continuum	80
6.7.5 Derivative with respect to the tangent pressure	80
6.7.6 Independence of retrieved variables	81

6.8 Convergence criteria	81
---------------------------------	-----------

6.9 Pre-calculation of line shapes	82
-------------------------------------------	-----------

6.10 Different grids during the cross-section calculation	82
------------------------------------------------------------------	-----------

6.11 Cross-section look-up tables	83
------------------------------------------	-----------

6.12 Variable frequency grids for radiative transfer computation	84
-------------------------------------------------------------------------	-----------

7 - VALIDATION OF THE CHOICES AND FUTURE WORK	85
------------------------------------------------------	-----------

REFERENCES	86
-------------------	-----------

APPENDIX A: MEMORANDUM ON DETERMINATION OF THE VCM OF ENGINEERING TANGENT HEIGHTS IN MIPAS	88
---------------------------------------------------------------------------------------------------	-----------

APPENDIX B: EVALUATION OF RETRIEVAL ERROR COMPONENTS AND TOTAL ERROR BUDGET	97
------------------------------------------------------------------------------------	-----------

APPENDIX C: GENERATION OF MW DATABASES AND OCCUPATION MATRICES	101
-----------------------------------------------------------------------	------------

APPENDIX D: GENERATION OF LUTS AND IRREGULAR FREQUENCY GRIDS (IG)	103
--------------------------------------------------------------------------	------------

APPENDIX E: GENERATION OF MW-DEDICATED SPECTRAL LINELISTS	104
------------------------------------------------------------------	------------

1 - Introduction

MIPAS (Michelson Interferometer for Passive Atmospheric Sounding) is an ESA developed instrument to be operated on board ENVISAT-1 as part of the first Polar Orbit Earth Observation Mission program (POEM-1). MIPAS will perform limb sounding observations of the atmospheric emission spectrum in the middle infrared region. Concentration profiles of numerous trace gases can be derived from MIPAS observed spectra.

According to the current baseline, from MIPAS measurements altitude profiles of atmospheric pressure and temperature (p,T), and of volume mixing ratio (VMR) of six high priority species (O_3 , H_2O , HNO_3 , CH_4 , N_2O and NO_2) will be routinely retrieved in near real-time (NRT). The retrieval of these parameters from calibrated spectra (Level 1b data) is indicated as NRT Level 2 processing.

Level 2 processing is expected to be a critical part of the Payload Data Segment (PDS) because of both the long computing time that may be required and the need for a validated algorithm capable of producing accurate and reliable results.

In the study "Development of an Optimised Algorithm for Routine P, T and VMR Retrievals from MIPAS Limb Emission Spectra" a scientific code for NRT Level 2 analysis was developed, suitable for implementation in ENVISAT PDS and optimised for the requirements of speed and accuracy. The results of the study were used by industry as an input for the development of the industrial prototype of Level 2 code.

The present Technical Note is a deliverable of the above study and provides the high level definition of the retrieval algorithm and a discussion of the physical and mathematical optimisations that have been implemented in the algorithm.

1.1 Changes with respect to Issue 2A of the present document

Changes in this document with respect to Issue 2A are highlighted by change bars. The main changes consist in the introduction of new sections regarding the description of items that previously were either described in sparse memorandums and small notes or not described at all. Namely, the following new sections have been introduced:

- Sect. 5.13: Interpolation of the retrieved profiles to a user-defined grid,
- Sect. 5.14: Optimized algorithm for construction of initial guess profiles (including generation of continuum profiles described in sub-sect 5.14.1).
- Sect. 5.15: Profiles regularization
- Appendix A: Determination of the VCM of the engineering tangent heights in MIPAS
- Appendix B: Evaluation of retrieval error components and total error budget (includes pT error propagation approach)
- Appendix C: Algorithm for generation of MW databases and occupation matrices
- Appendix D: Algorithm for generation of LUTs and irregular frequency grids
- Appendix D: Algorithm for generation MW-dedicated spectral linelists

Furthermore, Sect. 4.5 regarding the calculation of the VCM of the measurements was strongly modified in order to be consistent with recent baseline modifications. A new sub-section, 4.5.3, describing the method used to calculate the inverse of the VCM of the measurements was introduced. This sub-section replaces the old Sect. 6.13 (now removed).

Additional sparse modifications were introduced in order to remove obsolete statements and make the document in line with the current status of the study.

2 - Objectives of the technical note

Objectives of the technical note are:

- provision of a summary of the equations implemented Level 2 algorithm,
- identification of the existing options for the optimisation of the code,
- assessment of advantages and disadvantages of the individual options,
- to provide a rationale for the choice of preferred option for implementation and identification of a strategy for validation of the choice.

For a further self-consistent high-level description of the the algorithms implemented in the Level 2 scientific code of MIPAS, the reader should also refer to the paper of *Ridolfi et al. (2000)*.

3 - Criteria for the optimisation

The code must take into account requirements due to:

- characteristics of input data
- scientific requirements of output data
- correctness of the atmospheric model
- correctness of the instrument model
- numerical accuracy
- robustness in presence of erroneous observational data
- reduced computing time.

The main difficulty is due to the last requirement, which, in presence of the others, imposes the search for physical and mathematical optimisations in the implementation of the code.

The present document provides a description of the algorithm at a stage in which the code has already been developed and therefore a choice among the possible options has already been made.

We must be aware, however, that some choices have been made on a purely theoretical basis and therefore must be confirmed by tests performed with real data.

The acceptance criterion of the code, and therefore our choices, are based on the combined development of an Optimised Retrieval Model (ORM), an Optimised Forward Model (OFM) and a Reference Forward Model (RFM). Retrievals with the ORM from spectra simulated with the OFM and the RFM, with and without measurement noise, allow the identification of errors due to:

1. measurement error,
2. retrieval error,
3. approximations due to the optimisation.

Tests performed with different computing accuracy and with different profile representations determine, respectively:

4. the computing accuracy
5. and the null-space error.

The acceptance criterion requires the ORM to limit errors 2. and 3. so that the overall error budget including errors from 1. to 5. as well as systematic error, is kept below the following requirements:

- 3% error in tangent pressure retrieval,
- 2 K error in temperature retrieval,
- 5% error in VMR retrievals,

in the tangent altitude range 8 - 53 km. These were the requirements established at the beginning of the study in absence of specific indications regarding the ultimate accuracy attainable from MIPAS

measurements. However, test retrievals performed so far have shown that the above requirements can not be met in the whole altitude range explored by the MIPAS scan and for all the target species. Recently, the problem of assessing the ultimate retrieval accuracy attainable from MIPAS measurements has been tackled in the framework of the study for the selection of optimized spectral intervals (microwindows, MW) for MIPAS retrievals (see Appendix C and Bennett et al.,(1999)).

An acceptance test for the code, on the basis of actual retrieval error has been performed by using the complete OFM/ORM chain as well as the reference spectra generated by the RFM.

The strategy adopted for the operated choices is the following:

- since an altitude error is directly connected to a pressure error which in turn corresponds also to a VMR error, whenever the approximation corresponds to an altitude error the approximation is accepted if the error is less than 0.15 km (corresponding to a 2% pressure error). Actually, this is not a very conservative criterion but it is still satisfactory because it is applied only for the evaluation of FOV and self-broadening approximations
- if the approximation does not correspond to an altitude error, the approximation is accepted on the basis of the radiance error. Random error must be smaller than NESR, systematic errors must be smaller than NESR divided by the square root of the multiplicity of the effect. If individual approximations behave as either random or systematic errors can only be assessed by the full retrieval process. An educated compromise is made by using an acceptance threshold equal to NESR/4.

In Sect. 4 the mathematical problem connected to the retrieval is summarised. Sect. 5 is dedicated to the scientific aspects that affect the atmospheric and the instrument model and to the corresponding physical optimisations. Sect. 6 is dedicated to the choices related to the implementation of the calculations in the computing software and to the corresponding mathematical optimisations.

4 - The mathematics of the retrieval problem

4.1 Mathematical conventions

The mathematical conventions used in the present technical note are herewith summarised.

The functions may have the following attributes:

- **qualifiers:** qualifiers are given only as superscripts and consist of a note that helps to distinguish the different functions or the same function at different levels of the calculation. Parentheses are used to separate the qualifier from the other mathematical operations that can be confused with the qualifiers.
- **The variables of the functions** can appear either as a subscript or as arguments. In order to provide a representation consistent with the convention of matrices and vectors, whenever possible, the variables relative to which the variability of the function is explicitly sampled within the code are shown as a subscript, while variables relative to which a dependence only exists implicitly in the equations are shown as arguments.

When dealing with matrices and vectors, bold symbols are used.

4.2 Theoretical background

The problem of retrieving the altitude distribution of a physical or chemical quantity from limb-scanning observations of the atmosphere, drops within the general class of problems that require the fitting of a theoretical model, that describes the behaviour of a given system, to a set of available observations of the system itself. The theoretical model describes the system through a set of parameters so that the retrieval procedure consists in the search of the set of values of the parameters that produce the "best" simulation of the observations. The most commonly adopted criterion to accomplish the objective is the minimisation of the χ^2 function (generally defined as the weighted squared summation of the differences between observations and simulations) with respect to the value of the parameters. This criterion is generally referred as Least Squares Fit (LSF). When the theoretical model does not depend linearly on the unknown parameters the problem, called Non-linear Least Squares Fit (NLSF), cannot be solved directly by using a solution formula, and an iterative procedure must be used. Several methods exist for the NLSF; the one adopted for our purposes is the Gauss Newton (GN) method modified following the Marquardt's criterion (GNM).

In order to provide the framework of the subsequent discussion, the general mathematical formulation of the problem is herewith briefly reviewed. The mathematical formulation of the problem is described with full details in *Carlotti and Carli, (1994)*.

4.2.1 The direct problem

The signal S that reaches the spectrometer can be modelled, by means of the radiative transfer equation (described in Sect. 4.6), as a function $S = S(x, q_z)$ of the observation parameters x and of the distribution profile q_z of the atmospheric quantity which is to be retrieved. Since the radiative transfer does not represent a linear transformation, the problem of deriving the distribution q_z from the observed values of S cannot be solved through the analytical inversion of the radiative transfer equation.

A linear transformation connecting S and q_z can be obtained by operating a Taylor expansion of the radiative transfer equation, around an assumed profile \tilde{q}_z . In the hypothesis that \tilde{q}_z is near enough to the true profile to drop in a linear behaviour of the function S , the Taylor expansion can be truncated to the first term to obtain:

$$S(x, q_z) = S(x, \tilde{q}_z) + \int_0^\infty \left[\frac{\partial S(x, q_z)}{\partial q_z} \right]_{\tilde{q}_z} [q_z - \tilde{q}_z] dz \quad (4.2.1)$$

Note that the use of the integral is required in the above equation since the profile q_z is here considered as a continuous function.

Equation (4.2.1) can be written as:

$$N(x) = \int_0^\infty K(x, z) y_z dz, \quad (4.2.2)$$

where:

$$N(x) = S(x, q_z) - S(x, \tilde{q}_z) \quad (4.2.3)$$

$$K(x, z) = \left[\frac{\partial S(x, q_z)}{\partial q_z} \right]_{\tilde{q}_z} \quad (4.2.4)$$

$$y_z = [q_z - \tilde{q}_z]. \quad (4.2.5)$$

Equation (4.2.2) is an integral equation that represents a linear transformation of the unknown y_z leading to the observations $N(x)$ by way of the kernel $K(x, z)$.

4.2.2 The Gauss Newton method

In the case of practical calculations, the mathematical entities defined in Sect. 4.2.1 are represented by discrete values. Actually, we will deal with a finite number (n) of observations and a finite number (m) of values to represent, in a vector \mathbf{q}_z , the altitude distribution of the unknown quantities (these m values will be denoted as "parameters" from now on). As a consequence the integral operator of Eq. (4.2.2) becomes a summation and the equation itself can be expressed in matrix notation as:

$$\mathbf{n} = \mathbf{K} \mathbf{y} \quad (4.2.6)$$

In equation (4.2.6):

- \mathbf{n} is a vector of dimension n . The entry n_j of \mathbf{n} is the difference between observation j and the corresponding simulation calculated using the assumed profile \tilde{q}_z (Eq. 4.2.3).
- \mathbf{K} is a matrix (usually denoted as Jacobian matrix) having n rows and m columns. The entry k_{ij} of \mathbf{K} is the derivative of observation i made with respect to parameter j (Eq. 4.2.4)
- $\mathbf{y} = \mathbf{q}_z - \tilde{\mathbf{q}}_z$ is a vector of dimension m . The entry y_i of \mathbf{y} is the correction needed to the assumed value of parameter \tilde{q}_z in order to obtain its correct value q_z . The goal of the retrieval is the determination of this vector.

The problem is therefore that of the search for a "solution matrix" \mathbf{D} (having m rows and n columns) that, multiplied by vector \mathbf{n} provides \mathbf{y} .

If the vector \mathbf{n} is characterised by the variance-covariance matrix \mathbf{V}^n (square of dimension n), the χ^2 function which must be minimised is defined as:

$$\chi^2 = \mathbf{n}^T (\mathbf{V}^n)^{-1} \mathbf{n} \quad (4.2.7)$$

and matrix \mathbf{D} is equal to:

$$\mathbf{D} = (\mathbf{K}^T (\mathbf{V}^n)^{-1} \mathbf{K})^{-1} \mathbf{K}^T (\mathbf{V}^n)^{-1}. \quad (4.2.8)$$

The superscript "T" denotes the transpose and the superscript "-1" denotes the inverse of the matrix, if the inverse of \mathbf{V}^n does not exist, its generalised inverse must be used instead (see *Kalman (1976)* and Sect. 4.7). If the unknown quantities are suitably chosen, the inverse of matrix $(\mathbf{K}^T (\mathbf{V}^n)^{-1} \mathbf{K})$ always exists.

If the real minimum of the χ^2 function is found and \mathbf{V}^n is a correct estimate of the errors, the quantity defined by equation (4.2.7) has an expectation value equal to $(n - m)$ and a standard deviation equal to $\sqrt{n - m}$. The value of the quantity $\frac{\chi^2}{n - m}$ provides therefore a good estimate of

the quality of the retrieval. Values of $\frac{\chi^2}{n - m}$ which deviate too much from unity indicate the presence of incorrect assumptions in the retrieval.

The unknown vector \mathbf{y} is then computed as:

$$\mathbf{y} = \mathbf{D} \mathbf{n} \quad (4.2.9)$$

and the new estimate of the parameters as:

$$\mathbf{q}'_z = \tilde{\mathbf{q}}_z + \mathbf{y} \quad (4.2.10)$$

The errors associated with the solution to the inversion procedure can be characterised by the variance-covariance matrix (\mathbf{V}^q) of \mathbf{q}'_z given by:

$$\mathbf{V}^q = \mathbf{D} (\mathbf{V}^n) \mathbf{D}^T = (\mathbf{K}^T (\mathbf{V}^n)^{-1} \mathbf{K})^{-1} \quad (4.2.11)$$

Matrix \mathbf{V}^q permits to estimate how the experimental random errors map into the uncertainty of the values of the retrieved parameters. Actually, the square root of the diagonal elements of \mathbf{V}^q measures the root mean square (r.m.s.) error of the corresponding parameter. The off-diagonal element v_{ij} of matrix \mathbf{V}^q , normalised to the square root of the product of the two diagonal elements v_{ii} and v_{jj} , provides the correlation coefficient between parameters i and j .

If the hypothesis of linearity made in Sect. 4.2.1 about the behaviour of function S is satisfied, Eq. (4.2.10) provides the result of the retrieval process. If the hypothesis is not satisfied, the minimum of the χ^2 function has not been reached but only a step has been done toward the minimum and the vector \mathbf{q}_z computed by Eq. (4.2.10) represents a better estimate of the parameters with respect to $\tilde{\mathbf{q}}_z$. In this case the whole procedure must be reiterated starting from the new estimate of the parameters which is used to produce a new matrix \mathbf{K} . Convergence criteria are therefore needed in order to establish when the minimum of the χ^2 function has been approached enough to stop the iterations.

4.2.3 The Marquardt method

The Marquardt method introduces a modification to the procedure described in the previous subsection. This modification permits a faster convergence specially in the case of strongly non-linear problems.

Let's define the matrix \mathbf{A} :

$$\mathbf{A} = (\mathbf{K}^T (\mathbf{V}^n)^{-1} \mathbf{K}) \quad (4.2.12)$$

Following the Marquardt criterion, the diagonal elements of matrix \mathbf{A} are modified with a damping parameter λ , accordingly to:

$$A_{i,i} = A_{i,i} \cdot (1 + \lambda) \quad (4.2.13)$$

The algorithm proceeds then as follows:

1. calculate the χ^2 function and the matrix \mathbf{A} for the initial values of the parameters,
2. set λ to a initial "small" value (e.g. 0.001) and modify \mathbf{A} following Eq. (4.2.13),
3. calculate the new estimate of the parameters for the current choice of λ using equation (4.2.9).
4. calculate the new value of χ^2 using equation (4.2.7),
5. if χ^2 calculated at step 4 is greater than that calculated at step 1, then increase λ by an appropriate factor (e.g. 10) and repeat from step 3 (micro iteration),
6. if χ^2 calculated at step 4 is smaller than that calculated at step 1, then decrease λ by an appropriate factor (e.g. 10), adopt the new set of parameters to compute a new matrix \mathbf{A} and proceed to step 3 (macro iteration).

The (macro) iterations are stopped when a pre-defined convergence criterion is satisfied.

Obviously, the advantage of using Marquardt's method is that it avoids to calculate the Jacobian matrix when unnecessary.

For the development of the ORM code, however, since most operational retrievals are expected not to deal with a strongly non-linear problem and since the calculation of the Jacobian matrix is faster when performed within the forward model, ORM optimisations are performed for a Gauss Newton loop, i.e. also inside the Marquardt loop the Jacobian matrix is determined.

A more detailed description of Marquardt's method can be found in *Press et al (1992)*.

4.2.4 Review of the possible convergence criteria

We review here several conditions which can be considered for the definition of a convergence criterion.

1. The relative variation of the χ^2 function obtained in the present iteration with respect to the previous iteration is less than a fixed threshold t_1 i.e.:

$$\left| \frac{\chi^2(\mathbf{q}_z^{iter-1}) - \chi^2(\mathbf{q}_z^{iter})}{\chi^2(\mathbf{q}_z^{iter})} \right| < t_1 \quad (4.2.14)$$

where *iter* is the current iteration index.

2. The maximum correction that has to be applied to the parameters for the next iteration is below a fixed threshold t_2 i.e.:

$$\text{Max}_j \left| \frac{(\mathbf{q}_z^{iter-1})_j - (\mathbf{q}_z^{iter})_j}{(\mathbf{q}_z^{iter})_j} \right| < t_2 \quad (4.2.15)$$

different thresholds can be eventually used for the different types of parameters. The absolute variations of the parameters can also be considered instead of the relative variations, whenever an absolute accuracy requirement is present for a parameter (as for the case of temperature). Non-target parameters such as continuum and instrumental offset parameters should not be included in this check.

Since the above expression is singular whenever a parameter is equal to zero, an alternative formula which can be considered in this case is:

$$\text{Max}_j \left| \frac{(\mathbf{q}_z^{iter-1})_j - (\mathbf{q}_z^{iter})_j}{\sqrt{(\mathbf{V}_{iter}^q)_{j,j}}} \right| < t_2 \quad (4.2.15\text{bis})$$

where $\sqrt{(\mathbf{V}_{iter}^q)_{j,j}}$ represents the error associated with the parameter $(\mathbf{q}_z^{iter})_j$ at iteration $iter$. The reason which discourages for always using (4.2.15bis) instead of (4.2.15) is that $\sqrt{(\mathbf{V}_{iter}^q)_{j,j}}$ does not really represent the error on the parameter $(\mathbf{q}_z^{iter})_j$ when the retrieval is far from the convergence (i.e. far from the linear behaviour).

3. The difference between the real χ^2 and the chi-square computed in the linear approximation (χ_{LIN}^2) is less than a fixed threshold t_3 :

$$\left| \frac{\chi^2(\mathbf{q}_z^{iter}) - \chi_{LIN}^2(\mathbf{q}_z^{iter})}{\chi^2(\mathbf{q}_z^{iter})} \right| < t_3 \quad (4.2.16)$$

where χ_{LIN}^2 is computed using the expression:

$$((\mathbf{1} - \mathbf{KD})\mathbf{n})^T (\mathbf{V}^n)^{-1} ((\mathbf{1} - \mathbf{KD})\mathbf{n}) \quad (4.2.17)$$

4. The iteration index has reached a maximum allowed value (t_4):

$$iter \geq t_4 \quad (4.2.18)$$

The choice of the most appropriate logical combination of the above conditions (which provides the convergence criterion) is discussed in the section of mathematical optimisations (see Sect. 6.8).

4.2.5 Use of external (a-priori) information in the inversion model

When some a-priori information on the retrieved parameters is available from sources external to the MIPAS interferometer, the quality of retrieved parameters can be improved by including this

information in the retrieval process. Assuming the a-priori information as consisting of both an estimate \mathbf{q}^A of the vector of the retrieved parameters and of the variance covariance matrix \mathbf{V}^A related to \mathbf{q}^A , the combination of the retrieved vector with the externally provided vector \mathbf{q}^A can be made, after the convergence has been reached, by using the formula of the weighted average:

$$\mathbf{q} = (\mathbf{V}^q + \mathbf{V}^A)^{-1} (\mathbf{V}^q (\mathbf{D} \mathbf{n} + \tilde{\mathbf{q}}) + \mathbf{V}^A \mathbf{q}^A) \quad (4.2.19)$$

Introducing the explicit expressions of \mathbf{D} and \mathbf{V}^q given respectively by equations (4.2.8) and (4.2.11), equation (4.2.19) becomes:

$$\mathbf{q} = (\mathbf{K}^T \mathbf{V}^n \mathbf{K} + \mathbf{V}^A)^{-1} (\mathbf{K}^T \mathbf{V}^n \mathbf{n} + \mathbf{V}^q \tilde{\mathbf{q}} + \mathbf{V}^A \mathbf{q}^A) \quad (4.2.20)$$

This is the so called “optimal estimation” formula (see *Rodgers (1976)*). Eq. (4.2.20) can be used also at each retrieval iteration step, instead of eq. (4.2.9), for deriving the new estimate of the unknowns (at the same time also the observations \mathbf{q}^A are included in the vector of the observations for computing the χ^2). When equation (4.2.20) is used in the iteration cycles of the retrieval, the a-priori estimate of the retrieved parameters provides information on the unknown quantities also at the altitudes where the measurements may contain only poor information. In this case the retrieval process is more stable and fast (see also Sect. 5.2).

However, when using equation (4.2.20) in the retrieval iterations, the external information and the retrieval information are mixed during the minimisation process and therefore they can not be individually accessed at any time. This prevents to easily estimate the correction and the bias introduced by the a-priori information on the retrieved quantities.

The decision on whether to use equation (4.2.20) during the retrieval iterations or to use (4.2.9) during the retrieval and (4.2.20) after the convergence has been reached, chiefly depends on the type of a-priori information we are dealing with. In the cases in which the used a-priori information is expected not to polarise the results of the retrieval (e.g. in the cases in which independent a-priori estimates are available for different retrievals), equation (4.2.20) can be profitably used during the retrieval iterations.

Further advantages and disadvantages of the use of a-priori information are described as a scientific aspect in Sect. 5.2.

4.2.6 Use of the optimal estimation, for inclusion of LOS engineering information in p,T retrieval

Engineering LOS data are updated at each scan and therefore constitute an effective and independent source of information which can be routinely used in p,T retrievals and does not bias the retrieved profiles. In this case it is really worth to use formula (4.2.20) at each and iteration step and let the LOS information to help the convergence of the retrieval. In this case the a-priori information does not provide directly an estimate of the unknowns of the retrieval, but a measurement of a quantity that is related to the unknowns by way of the hydrostatic equilibrium law.

The engineering information on the pointing is expected to consist of a vector $\Delta \mathbf{z}$ containing the differences between tangent altitudes of the sweeps of the current scan and of a Variance-Covariance Matrix (VCM) \mathbf{V}^z related to the vector $\Delta \mathbf{z}$. We assume the components of the vector $\Delta \mathbf{z}$ to be defined as:

$$\begin{aligned}\Delta z_1 &= z_2 - z_1 \\ &\vdots \\ \Delta z_{N^{sw}-1} &= z_{N^{sw}} - z_{N^{sw}-1}\end{aligned}\tag{4.2.21}$$

where N^{sw} is the number of sweeps of the considered scan.

If we define the vector \mathbf{n}_1 as:

$$\mathbf{n}_1 = \Delta \mathbf{z} - \Delta \mathbf{z}_{tg}\tag{4.2.22}$$

where $\Delta \mathbf{z}_{tg}$ is the vector of the differences between the tangent altitudes at the current iteration; instead of equation (4.2.6) we have a couple of equations defining the retrieval problem:

$$\begin{aligned}\mathbf{n} &= \mathbf{K}\mathbf{y} \\ \mathbf{n}_1 &= \mathbf{K}_1\mathbf{y}\end{aligned}\tag{4.2.23}$$

where the matrix \mathbf{K}_1 is the jacobian matrix which links the differences between tangent altitudes with the vector of the unknowns. This matrix has to be re-computed at each retrieval iteration (as matrix \mathbf{K}); the recipe for the calculation of this matrix is given in Sect. 4.2.6.1. The χ^2 function to be minimised becomes:

$$\chi^2 = \mathbf{n}^T (\mathbf{V}^n)^{-1} \mathbf{n} + \mathbf{n}_1^T (\mathbf{V}^z)^{-1} \mathbf{n}_1\tag{4.2.24}$$

and the vector $\hat{\mathbf{y}}$ which minimises this χ^2 is given by:

$$\begin{aligned}\hat{\mathbf{y}} &= \left[\mathbf{K}^T (\mathbf{V}^n)^{-1} \mathbf{K} + \mathbf{K}_1^T (\mathbf{V}^z)^{-1} \mathbf{K}_1 \right]^{-1} \cdot \\ &\quad \cdot \left[\mathbf{K}^T (\mathbf{V}^n)^{-1} \mathbf{n} + \mathbf{K}_1^T (\mathbf{V}^z)^{-1} \mathbf{n}_1 \right]\end{aligned}\tag{4.2.25}$$

Therefore, if we define matrices \mathbf{A} , \mathbf{B} , and \mathbf{B}_1 as:

$$\begin{aligned}\mathbf{A} &\equiv \mathbf{K}^T (\mathbf{V}^n)^{-1} \mathbf{K} + \mathbf{K}_1^T (\mathbf{V}^z)^{-1} \mathbf{K}_1 \\ \mathbf{B} &\equiv \mathbf{K}^T (\mathbf{V}^n)^{-1} \\ \mathbf{B}_1 &\equiv \mathbf{K}_1^T (\mathbf{V}^z)^{-1}\end{aligned}\tag{4.2.26}$$

equation (4.2.25) becomes:

$$\hat{\mathbf{y}} = [\mathbf{A}]^{-1} \cdot [\mathbf{B}\mathbf{n} + \mathbf{B}_1\mathbf{n}_1]\tag{4.2.27}$$

In the linear regime, this equation provides the solution of the retrieval problem.

At each retrieval iteration the retrieval program has to compute matrices \mathbf{K} , \mathbf{K}_1 , \mathbf{A} , \mathbf{B} and \mathbf{B}_1 , then, since Marquardt's algorithm is used, matrix \mathbf{A} has to be modified accordingly to equation (4.2.13) and afterwards used in equation (4.2.27) in order to derive $\hat{\mathbf{y}}$.

In this approach, the equation which defines the linear chi-square χ_{LIN}^2 is:

$$\chi_{LIN}^2 = [\mathbf{n} - \mathbf{K}\mathbf{y}]^T [\mathbf{V}^n]^{-1} [\mathbf{n} - \mathbf{K}\mathbf{y}] + [\mathbf{n}_1 - \mathbf{K}_1\mathbf{y}]^T [\mathbf{V}^z]^{-1} [\mathbf{n}_1 - \mathbf{K}_1\mathbf{y}] \quad (4.2.28)$$

this is the equation to be used instead of equation (4.2.17).

4.2.6.1 Calculation of the jacobian matrix K_i of the engineering tangent altitudes

Let's explicitly write the second component of equation (4.2.23):

$$\Delta \mathbf{z} = \Delta \mathbf{z}_{tg} + \mathbf{K}_1 \mathbf{y} \quad (4.2.29)$$

It is clear from this relation that the component i,j of \mathbf{K}_1 is:

$$\mathbf{K}_1(i, j) = \frac{\partial \Delta z_i}{\partial y_j} \text{ with } i=1, \dots, N^{sw}-1 \text{ and } j=1, \dots, I^{top} \quad (4.2.30)$$

where I^{top} is the total number of fitted parameters in the current retrieval.

Now, being \mathbf{y} the vector of the unknowns of p,T retrieval, it is composed as follows:

- The first N^{sw} elements represent the tangent pressures,
- The elements from $N^{sw}+1$ up to $2*N^{sw}$ represent the tangent temperatures,
- The elements from $2*N^{sw}+1$ up to I^{top} represent atmospheric continuum and instrumental offset parameters.

Since engineering tangent altitudes do not depend on continuum and offset parameters $\mathbf{K}_1(i,j)=0$ for $i=1, \dots, N^{sw}-1$ and $j=2*N^{sw}+1, \dots, I^{top}$.

On the other hand the engineering tangent altitudes are connected with tangent pressures and tangent temperatures through hydrostatic equilibrium law.

The transformation which leads to $\Delta \mathbf{z}$ starting from P,T is defined by the hydrostatic equilibrium:

$$\Delta z_i = -\frac{T_{i+1} + T_i}{2\gamma} \ln\left(\frac{P_{i+1}}{P_i}\right) \quad \text{for } i = 1, \dots, N^{sw} - 1 \quad (4.2.31)$$

where P and T indicate, as usual, pressure and temperature and γ is equal to:

$$\gamma = g(\bar{z}, \varphi) \cdot \frac{M}{R} \quad (4.2.32)$$

where g is the acceleration of gravity at the mean altitude of the layer $\bar{z} = (z_{i+1} + z_i) / 2$ and latitude φ ; M is the air mass and R the gas constant. If the altitudes are measured in km and T in Kelvin, we get $M/R = 3.483676$.

The jacobian matrix \mathbf{J}_1 associated with the transformation (4.2.31) is a $(N^{sw}-1; 2N^{sw})$ matrix containing the derivatives:

$$\begin{aligned} \mathbf{J}_1(i, j) &= \frac{\partial \Delta z_i}{\partial p_j} & \text{for } i = 1, \dots, N^{sw} - 1 \text{ and } j = 1, \dots, N^{sw} \\ \mathbf{J}_1(i, j) &= \frac{\partial \Delta z_i}{\partial T_{j-N^{sw}}} & \text{for } i = 1, \dots, N^{sw} - 1 \text{ and } j = N^{sw} + 1, \dots, 2N^{sw} \end{aligned} \quad (4.2.33)$$

and therefore, deriving equations (4.2.31) we obtain:

$$\begin{aligned} \mathbf{J}_1(i, j) &= -\frac{T_{i+1} + T_i}{2\gamma} \left[\frac{1}{P_{i+1}} \delta_{j=i+1} - \frac{1}{P_i} \delta_{j=i} \right] \\ &\quad \text{for } i = 1, \dots, N^{sw} - 1 \text{ and } j = 1, \dots, N^{sw} \\ \mathbf{J}_1(i, j) &= -\frac{1}{2\gamma} \ln \left(\frac{P_{i+1}}{P_i} \right) \left[\delta_{j-N^{sw}=i+1} + \delta_{j-N^{sw}=i} \right] \\ &\quad \text{for } i = 1, \dots, N^{sw} - 1 \text{ and } j = N^{sw} + 1, \dots, 2N^{sw} \end{aligned} \quad (4.2.34)$$

where the function δ is defined as:

$$\delta_{\text{arg}} = \begin{cases} 1 & \text{if } [\text{arg}] = \text{TRUE} \\ 0 & \text{if } [\text{arg}] = \text{FALSE} \end{cases} \quad (4.2.35)$$

Considering that the original vector of the unknowns of p,T retrieval contains also continuum and offset parameters, matrix \mathbf{K}_1 can be obtained by extending matrix \mathbf{J}_1 with as many columns as required to reach the dimension $(N^{sw} - I; I^{top})$. As aforementioned, these extra columns contain only zeroes due to the fact that the tangent altitudes do not depend on continuum and offset parameters.

For what concerns the variance-covariance matrix \mathbf{V}^z of MIPAS tangent heights required for the implementation of the equations explained in Sect. 4.2.6, this matrix is derived using a simple algorithm based on MIPAS pointing specifications. This algorithm is described in Appendix A.

4.3 The global fit analysis

In global-fit introduced by *Carlotti, (1988)*, the whole altitude profile is retrieved from simultaneous analysis of all the selected limb-scanning measurements. The retrieval is based on the least-squares criterion and looks for a solution profile that has a number p of degrees of freedom smaller than or equal to the number of the observed data points. In practice the profile is retrieved at p discrete altitudes and at intermediate altitudes an interpolated value is used.

In this approach, the vector \mathbf{n} that appears in Eq. (4.2.9) is the difference between all the selected observations and the corresponding simulations (all the spectral intervals and all the limb-scanning measurements are included in this vector, eventually also a-priori information can be included).

The unknown vector \mathbf{y} may contain a different variable depending on the retrieval we are performing, in general it is, however, an altitude dependent distribution which is sampled at a number of discrete altitudes as well as some spectroscopic and instrumental parameters (e.g. atmospheric continuum).

The use of Marquardt's method for the minimisation of the χ^2 function requires the computation of the quantities that appear in the equations (4.2.8) and (4.2.9), namely:

- simulations for all the limb-scanning measurements and all the selected microwindows,

- the variance covariance matrix \mathbf{V}^n of the observations,
- the Jacobian matrix \mathbf{K}

The simulations of the observed spectra are performed using the so called “forward model” that is described in Sect. 4.4.

The variance covariance matrix related to the apodised spectral data (observations) is derived starting from noise levels, apodisation function and zero filling information, using the algorithm described in Sect. 4.5.

The Jacobian matrix containing the derivatives of the simulated spectra with respect to the unknown parameters is computed as described in Sect. 4.6.

4.4 High level mathematics of the forward model

The task of the forward model is the simulation of the spectra measured by the instrument in the case of known atmospheric composition. Therefore, this model consists of:

1. the simulation of the radiative transfer through the Earth’s atmosphere for an optimal instrument with infinitesimal field of view (FOV), infinite instrumental spectral resolution and no distortions of the line-shape.
2. the convolution of this spectrum with the apodised instrument line shape (AILS) to obtain the apodised spectrum which includes line shape distortions.
3. the convolution of these spectra with the FOV of the instrument.

Note that while step 1. provides a model of the atmospheric signal, steps 2. and 3. simulate instrumental effects. Not all the instrumental effects are however simulated in the forward model, since the retrieval is performed from calibrated spectra, instrument responsivity and phase errors are corrected in level 1b processing. The AILS which includes the effects of finite resolution, instrument line-shape distortions and apodization is also provided by Level 1b processing.

4.4.1 The radiative transfer

In order to obtain the spectra $S(\sigma, z_g)$ (i.e. the intensity as a function of the wavenumber σ) for the different limb geometries (denoted by the tangent altitude z_g of the observation g) the following integral for the radiative transfer has to be calculated:

$$S(\sigma, z_g) = \int_{\tau^b}^1 B(\sigma, T(x_g)) d\tau(\sigma, x_g) \quad (4.4.1)$$

where

σ = wavenumber

z_g = tangent altitude of the optical path g

x_g = co-ordinate along the line of sight (LOS) belonging to
the optical path with the tangent altitude z_g

$S(\sigma, z_g)$ = spectral intensity

$T(x_g)$ = temperature

$B(\sigma, T)$ = source function

$\tau(\sigma, x_g)$ = transmission between the point x_g on the LOS and the observer located at x_0 .

This quantity depends on the atmospheric composition, pressure and temperature through the co-ordinate x .

b = indicator for the farthest point that contributes to the signal

Under the assumption of local thermodynamic (LTE) equilibrium $B(\sigma, T)$ is the Planck function:

$$B(\sigma, T) = \frac{2hc^2\sigma^3}{\exp\left[\frac{hc\sigma}{k^BT}\right] - 1} \quad (4.4.2)$$

with h = Planck's constant
 c = velocity of the light
 k^B = Boltzmann's constant

The transmission can be expressed as a function of x :

$$\tau(\sigma, x_g) = \exp\left[-\int_{x_0}^{x_g} \overline{k(\sigma, x')} \eta(x') dx'\right] \quad (4.4.3)$$

with $\eta(x_g) = \frac{p(x_g)}{k^BT(x_g)}$ = number density of the air
 $p(x_g)$ = pressure

and the weighted absorption cross section:

$$\overline{k(\sigma, x_g)} = \sum_{m=1}^{molec} k_m(\sigma, x_g) X_m(x_g) \quad (4.4.4)$$

where $molec$ = number of different molecular species that absorb in the spectral region under consideration
 $X_m(x_g)$ = volume mixing ratio (VMR) of the species m at the point x_g
 $k_m(\sigma, x_g)$ = absorption cross sections of the species m

In the retrieval model the atmospheric continuum emission is taken into account as an additional species with VMR = 1 and the corresponding cross section is fitted as a function of altitude and microwindow (see Sect. 5.11.3). For the continuum calculation in the self standing forward model the cross sections are taken from a look up table and the real VMR of the continuum species is used (see Sect. 5.11.3).

Equation (4.4.1) can now be written as:

$$\begin{aligned} S(\sigma, z_g) &= \int_{x_0}^{x_g^b} B(\sigma, T(x_g)) \frac{d\tau(\sigma, x_g)}{dx_g} dx_g = \\ &= \int_{x_0}^{x_g^b} B(\sigma, T(x_g)) \overline{k(\sigma, x_g)} \eta(x_g) \tau(\sigma, x_g) dx_g \end{aligned} \quad (4.4.5)$$

In order to determine the integral (4.4.5) two basic steps are necessary:

- the ray tracing, i.e. the determination of the optical path x_g and, consequently, the temperature $T(x_g)$, the pressure $p(x_g)$ and the volume mixing ratio $X_m(x_g)$ along the LOS and
- the calculation of the absorption cross sections $k_m(\sigma, x_g)$

Ray tracing

The line of sight in the atmosphere is given by the viewing direction of the instrument and the distribution of the refraction index in the atmosphere. The refractive index $n(p(x_g), T(x_g))$ is determined as a function of pressure and temperature by the atmospheric model (see section 5.5).

Absorption cross section calculation

The absorption cross section of one molecular species m as a function of temperature and pressure is given by the following sum over all lines of the species:

$$k_m(\sigma, T, p) = \sum_{l=1}^{lines} L_{m,l}(T) A_{m,l}(\sigma - \sigma_{m,l}, T, p) \quad (4.4.6)$$

where $L_{m,l}(T)$ = line strength of line l of species m
 $\sigma_{m,l}$ = central wavenumber of line l of species m
 $A_{m,l}(\sigma - \sigma_{m,l}, T, p)$ = line profile (line-shape)

The line strength is calculated by the formula:

$$L_{m,l}(T) = L_{m,l}(T^0) \frac{Q_m(T^0)}{Q_m(T)} \cdot \frac{\exp\left[-\frac{hcE''_{m,l}}{k^BT}\right]}{\exp\left[-\frac{hcE''_{m,l}}{k^BT^0}\right]} \cdot \frac{1 - \exp\left[-\frac{hc\sigma_{m,l}}{k^BT}\right]}{1 - \exp\left[-\frac{hc\sigma_{m,l}}{k^BT^0}\right]} \quad (4.4.7)$$

with $L_{m,l}(T^0)$ = line strength at reference temperature T^0
 $Q_m(T)$ = total internal partition function
 $E''_{m,l}$ = lower state energy of the transition

The basic line shape is the Voigt function $A_{m,l}^V(\sigma - \sigma_{m,l}, T, p)$ - the convolution of the Doppler $A_{m,l}^D(\sigma - \sigma_{m,l}, T)$ and the Lorentz profile $A_{m,l}^L(\sigma - \sigma_{m,l}, T, p)$:

$$A_{m,l}^V(\sigma - \sigma_{m,l}, T, p) = A_{m,l}^D(\sigma - \sigma_{m,l}, T) * A_{m,l}^L(\sigma - \sigma_{m,l}, T, p) \quad (4.4.8)$$

The Doppler profile is given by the formula

$$A_{m,l}^D(\sigma - \sigma_{m,l}, T) = \sqrt{\frac{\ln 2}{\pi}} \frac{1}{\alpha_{m,l}^D} \exp \left[-\ln 2 \frac{(\sigma - \sigma_{m,l})^2}{\alpha_{m,l}^{D^2}} \right] \quad (4.4.9)$$

with the half width at half maximum (HWHM) of the line:

$$\alpha_{m,l}^D = \sigma_{m,l} \sqrt{2 \ln 2 \frac{k^B T}{M_m c^2}} \quad (4.4.10)$$

where M_m = molecular mass of species m

The Lorentz function is:

$$A_{m,l}^L(\sigma - \sigma_{m,l}, T, p) = \frac{1}{\pi} \frac{\alpha_{m,l}^L}{\alpha_{m,l}^{L^2} + (\sigma - \sigma_{m,l})^2} \quad (4.4.11)$$

and the Lorentz HWHM:

$$\alpha_{m,l}^L = \alpha_{m,l}^{L_0} \frac{p}{p^0} \left[\frac{T^0}{T} \right]^{\gamma_{m,l}} \quad (4.4.12)$$

with $\alpha_{m,l}^{L_0}$ = Lorentz half width at reference temperature T^0
and reference pressure p^0
 $\gamma_{m,l}$ = coefficient of temperature dependence of the half width

Using the substitutions:

$$x_{m,l} = \sqrt{\ln 2} \frac{\sigma - \sigma_{m,l}}{\alpha_{m,l}^D} \quad (4.4.13)$$

and

$$y_{m,l} = \sqrt{\ln 2} \frac{\alpha_{m,l}^L}{\alpha_{m,l}^D} \quad (4.4.14)$$

the Voigt function can be rewritten as:

$$A_{m,l}^V(\sigma - \sigma_{m,l}, T, p) = \sqrt{\frac{\ln 2}{\pi}} \frac{1}{\alpha_{m,l}^D} K(x_{m,l}, y_{m,l}) \quad (4.4.15)$$

with

$$K(x_{m,l}, y_{m,l}) = \frac{y_{m,l}}{\pi} \int_{-\infty}^{+\infty} \frac{e^{-t^2}}{(x_{m,l} - t)^2 + y_{m,l}^2} dt \quad (4.4.16)$$

4.4.2 Convolution with the AILS

In order to take into account the

- finite spectral resolution of the instrument
- distortion of the line-shape by the instrument
- the apodisation of the observed spectra,

the spectrum $S(\sigma, z_g)$ is convoluted with $AILS(\sigma)$, giving:

$$S^I(\sigma, z_g) = S(\sigma, z_g) * AILS(\sigma) \quad (4.4.16)$$

$AILS(\sigma)$ is the Apodised Instrument Line Shape that is obtained by convoluting the measured ILS with the apodisation function used for the apodisation of the observed spectra.

4.4.3 Convolution with the FOV

The FOV of an instrument is an angular distribution. In case of satellite measurements, like MIPAS, a (nearly) linear relationship exists between viewing angle and tangent altitude, therefore the FOV can be represented using a limb-scanning-angle-invariant altitude distribution.

$FOV(z_g, z)$ describes the finite FOV of MIPAS as a function of the altitude z . In the case of MIPAS, $FOV(z_g, z)$ is represented by a piecewise linear curve tabulated in the input files. For the simulation of the spectrum affected by the finite FOV ($S^F(\sigma, z_g)$) the following convolution is calculated:

$$S^F(\sigma, z_g) = S^I(\sigma, z) * FOV(z_g, z) \quad (4.4.17)$$

4.4.4 Instrumental continuum

For the simulation of the instrumental continuum an additional (microwindow dependent and sweep independent) term is added to $S^F(\sigma, z_g)$. This term is fitted in the retrieval program.

4.4.5 Summary of required variables

For the atmospheric model:

- pressure along the line of sight
- temperature along the line of sight
- volume mixing ratio along the line of sight

$$p(x_g)$$

$$T(x_g)$$

$$X(x_g)$$

For the ray tracing:

- altitude and viewing direction of the instrument or
- tangent altitude (in case of homogeneously layered and spherical atmosphere)

$$z_g$$

For the cross section calculation:

- central wavenumber of transition l of species m
- reference line strength of transition l of species m
- lower state energy of transition l of species m
- total internal partition function of species m

$$\sigma_{m,l}$$

$$L_{m,l}(T^0)$$

$$E''_{m,l}$$

$$Q_m(T)$$

- molecular mass of species m
- reference Lorentz half width of transition l of species m
- coefficient of temperature dependence of the half width

$$M_m$$

$$\alpha_{m,l}^{L_0}$$

$$\gamma_{m,l}$$

For the AILS convolution:

- apodised instrument line shape

$$AILS(\sigma)$$

For the FOV convolution:

- field of view function

$$FOV(z_g, z)$$

4.5 Calculation of the VCM of the measurements

The variance covariance matrix (VCM) of the residuals \mathbf{V}^n , used in Eq. (4.2.24), is in principle given by the combination of the VCM of the observations \mathbf{V}^s and the VCM of the forward model \mathbf{V}^{FM} :

$$\mathbf{V}^n = \mathbf{V}^s + \mathbf{V}^{FM} \quad (4.5.1)$$

However, since:

- the amplitude of the forward model error is not accurately known;
- correlations between forward model errors are difficult to quantify

in the retrieval algorithm we choose to use $\mathbf{V}^n = \mathbf{V}^s$. The entity of the forward model errors will be evaluated by analyzing the behavior of the χ^2 - test for the different microwindows, at the different altitudes. In particular, the obtained χ^2 - test will be compared with its expected value as determined on the basis of the total error evaluated by the so called “Residuals and Error Correlation (= REC)” analysis (see Piccolo et al. (2001)).

Herewith we describe how the VCM of the observations \mathbf{V}^s can be derived.

Even if the points of the interferograms measured by MIPAS are sampled independently of each other (no correlation between the measurements), the spectral data are affected by correlation. The correlation arises from the data processing performed on the interferogram (e.g. apodisation).

For this reason the noise levels provided by Level 1B processing do not fully characterise the measurement errors and the computation of a complete Variance Covariance Matrix \mathbf{V}^s of the spectrum $S(\sigma)$ is needed.

In Section 4.5.1 we describe the operations performed on the interferogram in order to obtain the apodised spectrum. On the basis of these operations, in Sect. 4.5.2 we describe how the variance covariance matrix \mathbf{V}^s of the observations can be derived. Finally, in Sect. 4.5.3 the procedure used to invert \mathbf{V}^s is described.

4.5.1 Operations performed on the interferogram to obtain the apodised spectrum

The standard MIPAS interferogram is a double-sided interferogram obtained with a nominal maximum optical path difference of +/- 20 cm (the optical path difference is however commandable to lower values, allowing to change the trade-off between spectral resolution and acquisition time for scene data).

The final apodised spectrum $S(\sigma)$ is then obtained by subsequently performing the following operations on the interferogram:

1. Zero-filling

During Level 1B processing, in order to exploit the fast Fourier Transform algorithm, the number of points of the interferogram is made equal to a power of 2 by extending the interferogram with zeroes from the measured maximum path difference (MPD) to the zero-filled path difference (ZFPD).

The measured interferogram is therefore equal to an interferogram with maximum path difference ZFPD, multiplied by a boxcar function $\Pi(d)$ defined as:

$$\Pi^{MPD}(d) = \begin{cases} 1 & \text{when } d \in [-MPD, MPD] \\ 0 & \text{when } d \notin [-MPD, MPD] \end{cases} \quad (MPD = \text{Max. Path Difference}) \quad (4.5.2)$$

2. Fourier Transformation (FT)

Let us call $S^{NAHR}(\sigma)$ the spectrum obtained from the measured (and zero-filled) interferogram and $S^{HR}(\sigma)$ the spectrum that would have been obtained from the interferogram with maximum path difference ZFPD. $S^{NAHR}(\sigma)$, $S^{HR}(\sigma)$ and $FT[\Pi^{MPD}(d)]$ are all given in the sampling

$$\text{grid } \Delta\sigma = \frac{1}{2 \cdot ZFPD}.$$

Since the FT of the product of two functions is equal to the convolution of the FT's of the two functions, we obtain:

$$S^{NAHR}(\sigma) = S^{HR}(\sigma) * FT[\Pi^{MPD}(d)]. \quad (4.5.3)$$

3. Re-sampling at the fixed grid

Since a pre-defined and constant grid is required by the ORM for its optimisations, the spectrum is re-sampled at a fixed grid of $\Delta\sigma = 0.025 \text{ cm}^{-1} = \frac{1}{2 \cdot D}$, with D equal to 20 cm. The performed operation can be written as:

$$S^{NA}(\sigma) = S^{NAHR} * FT[\Pi^{ZFPD}(d)], \quad (4.5.4)$$

where $S^{NA}(\sigma)$ is calculated at the fixed grid $\Delta\sigma = \frac{1}{2 \cdot D}$.

In this case the operation (4.5.4) is not a classical convolution among quantities that are defined on the same grid (e.g. Eq. (4.5.3)), but is a re-sampling process which changes the grid spacing from $\frac{1}{2 \cdot ZFPD}$ of S^{NAHR} to $\frac{1}{2 \cdot D}$ of S^{NA} .

This operation does not introduce correlation between the spectral points only if $MPD \geq D$.

If $MPD \geq D$ the result of (4.5.4) is equal to the FT of a ± 20 cm interferogram.

If $MPD < D$ the result of (4.5.4) is equal to the zero-filling to 20 cm path difference of an interferogram with path difference MPD, therefore the spectral points are correlated to each other.

The NESR values given in the Level 1B product are computed after this re-sampling step.

4. Apodisation

The apodised spectrum $S(\sigma)$ is obtained by convoluting the spectrum $S^{NA}(\sigma)$ with the apodisation function $\theta(\sigma)$, sampled at $\Delta\sigma = \frac{1}{2 \cdot D}$.

$$S(\sigma) = S^{NA}(\sigma) * \theta(\sigma) \quad (4.5.5)$$

4.5.2 Computation of the VCM relative to a single microwindow

In the case of MIPAS data the microwindows are usually well separated and it is reasonable to assume that the spectral data points belonging to different microwindows are uncorrelated. As a consequence, the variance covariance matrix of the spectrum \mathbf{V}^S is a block-diagonal matrix with as many blocks as many microwindows are processed, and the dimension of each block is equal to the number of spectral points in the corresponding microwindow. We assume that different points of the microwindow are characterised by the same error, but different microwindows can have different errors. In this section we derive the relationship that applies to each block and for simplicity with \mathbf{V}^S we refer to a single block rather than the full VCM of the observations.

The correlation between different spectral points of the microwindow is due to the apodisation process and to the zero-filling that is present in the case of $MPD < D$.

If $MPD \geq D$, only the apodisation is a cause of correlation and the VCM \mathbf{V}^S of the apodised spectrum $S(\sigma)$ can be computed from the VCM \mathbf{V}^{NA} associated with $S^{NA}(\sigma)$, (\mathbf{V}^{NA} is a diagonal matrix since the spectral points of $S^{NA}(\sigma)$ are uncorrelated) and from the Jacobian \mathbf{J} of the transformation (4.5.5):

$$\mathbf{V}^S = \mathbf{J} \cdot \mathbf{V}^{NA} \mathbf{J}^T = \mathbf{V}^{NA} \mathbf{J} \mathbf{J}^T \quad (4.5.6)$$

In Eq. (4.5.6) the order of the operations has been changed because \mathbf{V}^{NA} is a diagonal matrix and all the diagonal elements are equal. The diagonal values are equal to $(NESR)^2$, where NESR is the quantity calculated after operation 3 of Sect. 4.5.1.

The calculation of matrix \mathbf{J} is straightforward. From the explicit expression of the convolution (4.5.5):

$$S(\sigma_i) = \sum_j S^{NA}(\sigma_j) \cdot \theta(\sigma_i - \sigma_j) \quad (4.5.7)$$

it follows that the entry i,k of matrix \mathbf{J} is equal to:

$$\mathbf{J}_{i,k} = \theta(\sigma_i - \sigma_k) \quad (4.5.8)$$

and the variance covariance matrix \mathbf{V}^S can be computed as:

$$\mathbf{V}_{i,j}^S = (NESR)^2 \mathbf{J}_{i,k} \mathbf{J}_{k,j}^T = (NESR)^2 \cdot \sum_k \theta(\sigma_i - \sigma_k) \cdot \theta(\sigma_j - \sigma_k). \quad (4.5.9)$$

If $MPD < D$ also the effect of zero-filling must be taken into account. Furthermore, the mathematics is more complicate because zero-filling introduces a correlation between the spectral points of $S^{NA}(\sigma)$ (Eq. (4.5.4)) and the measurement error of this spectrum is characterised by a VCM V^{NA} , with off-diagonal elements different from zero.

The apodised spectrum $S(\sigma)$ of a measurement made with $MPD < D$ is given by:

$$S(\sigma) = S^{NA}(\sigma) * \theta^{MPD}(\sigma) \quad (4.5.10)$$

where $\theta^{MPD}(\sigma)$ is the apodisation function that is used in the case of a zero-filled spectrum. In the interferogram domain the FT of $\theta^{MPD}(\sigma)$ is a function that applies the selected apodisation rule between zero and MPD and is zero-filled between MPD and D.

As in the case of Eq. (4.5.6), V^S is obtained from the following equation :

$$V^S = J \cdot V^{NA} J^T \quad (4.5.11)$$

where in this case J is the Jacobian of the apodisation with $\theta^{MPD}(\sigma)$ and V^{NA} is the VCM of the zero filled spectrum and has off-diagonal elements different from zero. Therefore, we know the diagonal elements of this matrix, which are equal to the NESR calculated in Level 1B, but further calculations must be performed in order to determine the off-diagonal elements.

A few simple considerations allow us to overcome this difficulty.

The spectrum $S^{NA}(\sigma)$ measured with maximum path different MPD can be expressed as a function of the spectrum $S^{NAD}(\sigma)$ measured with maximum path different D :

$$S^{NA}(\sigma) = S^{NAD}(\sigma) * FT[\Pi^{MPD}(d)] \quad (4.5.12)$$

and V^{NA} can be computed as a function of V^{NAD} using the following expression:

$$V^{NA} = F \cdot V^{NAD} \cdot F^T, \quad (4.5.13)$$

where F is the Jacobian matrix associated with the transformation (4.5.12).

Since V^{NAD} is a diagonal matrix

$$V^{NA} = F \cdot x \cdot I \cdot F^T = x \cdot F \cdot F^T, \quad (4.5.14)$$

where I is the identity matrix and x is the value of each diagonal element of V^{NAD} (also in this case it is assumed that the spectral data in the same microwindow are affected by the same error).

From Eq. (4.5.14), the diagonal elements of V^{NA} are equal to:

$$(V^{NA})_{ii} = |f(\sigma)|^2 \cdot x; \quad (4.5.15)$$

where $f(\sigma)$ is a “continuous representation” of the columns of the matrix F . Using the Parseval theorem we get:

$$|f(\sigma)|^2 = \frac{1}{D} \int_0^D [\Pi^{MPD}(d')] [\Pi^{MPD}(d')] dd' = \frac{MPD}{D}, \quad (4.5.16)$$

and since $(\mathbf{V}^{NA})_{ii} = (NESR)^2$, from Eq. (4.5.15) and Eq. (4.5.16) we obtain :

$$x = \frac{D}{MPD} \cdot (NESR)^2 \quad (4.5.17)$$

The result obtained with Eq. (4.5.17) implies that the measurement of the NESR, even if it does not fully determine \mathbf{V}^{NA} , can be used to fully determine \mathbf{V}^{NAD} .

Therefore, we can express $S(\sigma)$ as a function of $S^{NAD}(\sigma)$. From Eq. (4.5.10) and Eq. (4.5.12) we obtain :

$$S(\sigma) = S^{NAD}(\sigma) * FT[\Pi^{MPD}(d)] * \theta^{MPD}(\sigma) \quad (4.5.18)$$

and since from the definition of the apodisation function it follows that $FT[\Pi^{MPD}(d)] * \theta^{MPD}(\sigma) = \theta^{MPD}(\sigma)$, it results :

$$S(\sigma) = S^{NAD}(\sigma) * \theta^{MPD}(\sigma) \quad (4.5.19)$$

This expression is analogous to Eq. (4.5.5), where $S^{NA}(\sigma)$ has been replaced by $S^{NAD}(\sigma)$ and the apodisation function has been replaced with the apodisation function relative to MPD. Using this expression the calculations of the VCM relative to $S(\sigma)$ requires the following calculation:

$$\mathbf{V}^S = \mathbf{J} \cdot \mathbf{V}^{NAD} \mathbf{J}^T = \mathbf{J} \cdot x \cdot \mathbf{I} \cdot \mathbf{J}^T = x \cdot \mathbf{J} \mathbf{J}^T \quad (4.5.20)$$

and from Eq. (4.5.17) it follows that

$$\mathbf{V}^S = \frac{D}{MPD} \cdot (NESR)^2 \cdot (\mathbf{J} \cdot \mathbf{J}^T) \quad (4.5.21)$$

The result is similar to the one found for $MPD \geq D$ with the exception of the extra term $\frac{D}{MPD}$.

Besides, since the apodisation function contains 0 values between MPD and D, $\mathbf{J} \cdot \mathbf{J}^T$ and \mathbf{V}^S are singular matrices.

4.5.3 Computation of the inverse of the VCM

In the ORM the inverse of matrix \mathbf{V}^S is required. If \mathbf{V}^S is a block-diagonal matrix, also $\mathbf{V}^{S^{-1}}$ is a block-diagonal matrix and is made of blocks equal to the inverse of the blocks of \mathbf{V}^S .

As said before, when $MPD < D$ the “ranks” of the blocks are smaller than their dimension. This fact would be evident in the case in which the VCM is calculated in a sufficiently broad spectral range. However in the ORM, where microwindows are used and correlation are calculated in a limited

spectral range, the truncations make the determinant different from zero even if the number of independent pieces of information is less than the dimension of the matrix and the singularity of the matrix is not always automatically detected. The problem can be solved by modifying the routine that inverts the VCM: the VCM is inverted with the Singular Value Decomposition method (see Section. 4.7) also when the determinant is different from zero and the smallest eigenvalues are set equal to 0. The number of the eigenvalues set to 0 is given by the number of dependent points in the spectrum ($(1-MPD/D)*nl$, where nl is the dimension of the matrix (equal to the number of the sampling points of the microwindow included in the fit).

This issue is further complicated by the use of a selected set of spectral points for the retrieval (microwindows with masked points, see Appendix C). When a subset of points is used for the retrieval the blocks of \mathbf{V}^s are made by a corresponding subset of lines and columns. By reducing the dimension of the block the ratio between the rank and the dimension of the block is altered. The selection of a fraction of eigenvalues provides an useful conservative criterium.

4.6 Calculation of the Jacobian matrix \mathbf{K} of the simulations

The use of Gauss-Newton algorithm for the minimisation of the χ^2 function requires the computation of the Jacobian matrix \mathbf{K} of the simulations. The element of indexes i,j of this matrix is defined by the following relationship:

$$K_{i,j} = \left. \frac{\partial S_i(\mathbf{q})}{\partial q_j} \right|_{\mathbf{q}=\tilde{\mathbf{q}}} \quad (4.6.1)$$

where S_i is the simulated spectrum, and q_j is the j -th component of the vector \mathbf{q} containing the unknowns of the problem. The vector \mathbf{q} of the unknown parameters depends on the retrieval we are performing and a distinction between p, T and VMR retrievals is needed.

In the case of p, T retrieval the unknowns are:

- pressures at tangent altitudes,
- temperatures at tangent altitudes,
- continuum absorption cross sections at tangent altitudes, for the central frequencies of the considered microwindows
- instrumental continuum (i.e. additive term to the spectrum that is assumed as constant varying the observation geometry and function of the microwindow),

while in the case of VMR retrievals the unknowns are:

- VMR at tangent altitudes,
- continuum absorption cross sections at tangent altitudes, for the central frequencies of the microwindows that are going to be processed,
- instrumental continuum (i.e. additive term to the spectrum that is assumed as constant varying the observation geometry and function of the microwindow).

The derivatives are computed in correspondence of a vector $\tilde{\mathbf{q}}$ containing:

- the initial guess of the unknown parameters in the first iteration step,
- the new guess of the parameters in the subsequent iterations.

These derivatives can be computed either numerically or analytically. In general, the numerical approach requires an extra call to the forward model for the computation of each partial derivative,

while in the analytical approach the calculation of the derivatives can be performed in parallel to the forward model.

Tests have shown that:

- the derivatives with respect to tangent pressure, VMR, continuum cross-sections, and instrumental continuum can be performed analytically using only minor approximations.
- the derivatives with respect to tangent temperature cannot be computed analytically without introducing approximations that significantly degrade the accuracy.

The formulas to be used for the analytical calculation of the above derivatives are strictly linked to the mathematical optimisations used for the implementation of the atmospheric model into the program. These formulas are described in Sect. 6.7.

4.7 Generalised inverse

Even if a detailed description of the formulas needed for the calculation of the generalised inverse matrix can be found in *Kalman (1976)*, it is however useful to recall here a simple method that can be used for the computation of the generalised inverse of a symmetric matrix. This algorithm is used in the ORM to invert the matrices \mathbf{V}^n and $(\mathbf{K}^T(\mathbf{V}^n)^{-1}\mathbf{K})$ appearing in equation (4.2.8). In the case in which these matrices are non-singular the method provides the exact inverse matrix.

Let's call \mathbf{C} a symmetric matrix of dimension 'r'. It is possible to find a base of 'r' independent eigenvectors of \mathbf{C} . If \mathbf{V} is the matrix whose columns are the eigenvectors of \mathbf{C} , matrix \mathbf{C} can then be written in the form:

$$\mathbf{C} = \mathbf{V}\mathbf{w}\mathbf{V}^T \quad (4.7.1)$$

where \mathbf{w} is a diagonal matrix containing the eigenvalues of \mathbf{C} . The inverse matrix of \mathbf{C} is then:

$$\mathbf{C}^{-1} = \mathbf{V}\mathbf{w}^{-1}\mathbf{V}^T \quad (4.7.2)$$

The appearance of singularities in \mathbf{C} is detected by the presence of eigenvalues close to zero in \mathbf{w} . In this case the singularities can be eliminated by imposing $1/w_i = 0$ whenever $w_i \approx 0$.

This procedure corresponds to the calculation of the generalised inverse matrix of \mathbf{C} .

4.8 Variance-covariance matrix of tangent heights corrections

In p,T retrieval the retrieved quantities are tangent pressures and the temperatures corresponding to tangent pressures. In an atmosphere in hydrostatic equilibrium, after p,T retrieval is completed, it is always possible to derive from these two quantities an estimate of the differences between the tangent altitudes of two contiguous sweeps. Besides, if one of the tangent altitudes provided by engineering measurements is assumed as perfectly known, an estimate of all the tangent altitudes can be easily obtained. The differences between the tangent altitudes obtained from p,T retrieval and the corresponding engineering estimates of the tangent altitudes constitute the so called vector of 'tangent heights corrections'. Purpose of this section is to define the algorithm for the calculation of the VCM of this vector.

Let's assume that the analysed scan consists of N^{sw} sweeps and that the tangent altitude $z(N^{sw})$ of the lowest sweep is perfectly known. The corrections δz_i to the engineering tangent altitudes are defined as:

$$\delta z_i = z_i^{RET} - z_i^{ENG} \quad (4.8.1)$$

where z_i^{RET} are the tangent altitudes derived from p,T retrieval and z_i^{ENG} the engineering estimates of the tangent altitudes; the index i ranges from 1 to $N^{SW}-1$. It is important to appreciate that the error on δz_i that is given in its VCM, is not intended as the error on the difference contained in equation (4.8.1), which is complicated by the fact that both estimates of the tangent altitude are affected by an error and the two errors are not independent. The error on δz is intended as the error which should be attributed to the retrieved tangent altitude when this quantity is reconstructed by adding the correction δz to the reference levels provided by the engineering tangent altitudes.

By using hydrostatic equilibrium law and the tangent altitude of the lowest sweep, z_i^{RET} ($i = 1, \dots, N^{SW}-1$) can be expressed as:

$$z_i^{RET} = z(N^{SW}) + \sum_{j=i+1}^{N^{SW}} \frac{T_j + T_{j-1}}{2K} \cdot \log\left(\frac{P_j}{P_{j-1}}\right) \quad (4.8.2)$$

where, as usual, T_j and P_j are respectively temperature and pressure at tangent altitude z_j^{RET} and $K=M/g$ with M = air mass and g = acceleration of gravity.

Therefore, the tangent altitude corrections can be expressed as a function of T_j and P_j by substituting equation (4.8.2) in (4.8.1):

$$\delta z_i = \sum_{j=i+1}^{N^{SW}} \frac{T_j + T_{j-1}}{2K} \cdot \log\left(\frac{P_j}{P_{j-1}}\right) - \Delta z_j^{ENG} \quad i = 1, \dots, N^{SW} - 1 \quad (4.8.3)$$

where Δz_i^{ENG} are defined as:

$$\Delta z_i^{ENG} = z_{j-1}^{ENG} - z_j^{ENG} \quad j = 2, \dots, N^{SW} \quad (4.8.4)$$

Now, equation (4.8.3) allows the evaluation of the variance-covariance matrix \mathbf{V}^{HC} of the heights corrections δz_i starting from the variance-covariance matrix (\mathbf{V}^q) of the retrieved values of pressure and temperature (see Eq. (4.2.11) in Sect. 4.2.2, this matrix is directly provided by the retrieval algorithm). The transformation which links \mathbf{V}^{HC} and \mathbf{V}^q is:

$$\mathbf{V}^{HC} = \mathbf{K}_D \mathbf{V}^q \mathbf{K}_D^T \quad (4.8.5)$$

where \mathbf{K}_D is the jacobian matrix connecting δz_i with T_i and P_i . The elements of \mathbf{K}_D are the derivatives:

$$\mathbf{K}_D(i, j) = \frac{\partial(\delta z_i)}{\partial P_j} \quad \text{with } i = 1, \dots, N^{SW}-1 \text{ and } j = 1, \dots, N^{SW}$$

(4.8.6)

and

$$\mathbf{K}_D(i, N^{SW} + j) = \frac{\partial(\delta z_i)}{\partial T_j} \quad \text{with } i = 1, \dots, N^{SW}-1 \text{ and } j = 1, \dots, N^{SW}$$

these expressions can be easily evaluated by deriving Eq. (4.8.3) with respect to pressure and temperature. We obtain:

$$\frac{\partial(\delta z_i)}{\partial p_h} = \sum_{j=i+1}^{N^{SW}} \frac{T_j + T_{j-1}}{2K} \cdot \left(\frac{\delta_{h,j}}{P_j} + \frac{\delta_{h,j-1}}{P_{j-1}} \right) \text{ with } i = 1, \dots, N^{SW}-1 \text{ and } h = 1, \dots, N^{SW}$$

and

(4.8.7)

$$\frac{\partial(\delta z_i)}{\partial T_h} = \sum_{j=i+1}^{N^{SW}} \frac{1}{2K} \cdot (\delta_{j,h} + \delta_{j-1,h}) \cdot \log\left(\frac{P_j}{P_{j-1}}\right) \text{ with } i = 1, \dots, N^{SW}-1 \text{ and } h = 1, \dots, N^{SW}$$

Summarising, the steps to be carried-out for the calculation of the variance-covariance matrix \mathbf{V}^{HC} are:

- calculation of the jacobian matrix \mathbf{K}_D by using equations (4.8.7),
- transformation of the variance-covariance matrix \mathbf{V}^q of the retrieved pressures and temperatures, by using equation (4.8.5).

5 - Scientific aspects and physical optimisations

In this section the baselines for the choice of the implementation of specific physical aspects and optimisations into the code are discussed. This discussion includes the following items:

- explanation of the physical effects
- possible physical models for the description of these effects
- options for the implementation into the code
- optimisations: improved algorithms, simplifications, items to be neglected
- choices for the implementation
- accuracy of baselines

The aspects which will be discussed here are:

1. Retrieval grid: retrieval of the unknowns at fixed levels or at tangent altitudes
2. Use of a-priori information
3. Latitudinal gradients of atmospheric parameters
4. Model of the earth and calculation of the gravity
5. Ray tracing: equation of the refractive index and determination of the optical path
6. Line shape
7. Line mixing
8. Pressure shift
9. Non local thermodynamic equilibrium (NLTE)
10. Self broadening
11. Continuum: instrumental, nearby, far
12. Interpolation of the profiles

5.1 Choice between retrieval of profiles at fixed levels and at tangent altitude levels

In the case of the onion peeling method, the retrieved values of VMR can only be determined at the pressure levels that correspond to the tangent altitude of the limb scanning sequence. In the case of global fit this constraint does not exist and other discrete levels can be used. Since in Level 3 data processing global maps on pressure surfaces are produced, an interesting possibility offered by global fit is that of using fixed pressure levels which will in general be different from the tangent altitude levels.

5.1.1 Retrieval at tangent altitude and interpolation between retrieved values

If the pressure levels at which the retrieval is performed are the ones identified by the observation geometries of the limb scanning sequence they may not correspond to those needed by the user. In this case an interpolation can be applied and, as it is shown in *Carli (1995)*, the equations that fully characterise the interpolated values of the profiles assess that, even if a reduced statistic error applies to the profile at the interpolated altitude levels, the vertical resolution of the measurement is degraded.

Numerical tests have shown that between two retrieved values the measurement error has a minimum while the width of the averaging kernel (see e.g. *Rodgers (1976)*) has a maximum (i.e. the interpolation changes the trade-off between vertical resolution and accuracy in favour of the latter). Interpolation of retrieved values provides therefore a variable trade-off between accuracy and

vertical resolution of the measurement. Alternative interpolation schemes to be applied to retrieved profiles have been studied in *Carli et al., (2001)*. One of the methods proposed in this reference is discussed in Sect. 5.13 of the present document.

5.1.2 Retrieval at fixed levels

If the data utilisation requires VMR at fixed levels (for instance in level 3 data processing global maps on pressure surfaces are produced) an alternative could be that of retrieving the VMR directly at the required fixed levels.

The quality of the retrieval performed at fixed levels should be assessed with appropriate tests, but it is reasonable to expect that in general it provides larger noise and a constant vertical resolution.

5.1.3 Discussion of the problem

The basic problem that is behind the different trade-off between accuracy and vertical resolution of the measurement measured with the two approaches, can be explained by the Nyquist theorem. This theorem states that in order to measure a periodic variation of a distribution, the distribution must be sampled with intervals equal to one half the period of the variation as long as the maxima and the minima of the variation coincide with the sampling points (detection of the cosine variation). A sampling equal to a quarter of the period is needed in order to detect a variation with any phase (detection of both the cosine and the sine components).

According to the Nyquist theorem, we have that:

- the vertical resolution of the measurement coincides (within the limits of the retrieval problem) with the vertical resolution of the sounding if the retrieval is performed at tangent altitudes levels
- the vertical resolution of the measurement cannot be equal to the vertical resolution of the sounding if the retrieval is performed at intermediate levels.

Therefore, if the offset introduced by the pointing system, between wanted and implemented tangent altitudes causes a sounding of the atmosphere at tangent altitudes located in between the fixed levels required by the user, it is impossible to obtain at the fixed levels the maximum vertical resolution. This result, which is based on the implicit assumption that the weighting functions of limb sounding observations peak at the tangent altitudes, may have a partial exception if the difference between weighting functions at different frequencies and in different microwindows provides some information at intermediate altitudes. However, the exploitation of this second order information is bound to cause a major increase of the measurement error.

The choice is therefore between:

1. retrieval at tangent altitude levels followed by interpolation for determination of VMR at required levels: this procedure makes the best use of the data when no interpolation is used. If interpolation is used, the vertical resolution of the measurement depends upon the offset between retrieved and interpolated points. Up to a factor two loss in vertical resolution can be encountered.
2. retrieval at fixed pressure levels: this option has not been adequately tested, it is expected to provide retrievals at roughly (T dependence) constant vertical resolution, but the noise depends upon the offset between wanted and implemented tangent altitudes. A very large increase of noise can be encountered and it is not easy to quantify this increase.

The first option is the one with less risks. The second option is simpler from the conceptual point of view.

The two options imply a different implementation of level 2 data analysis and a significant compromise in scientific requirements, but from the code point of view there is no strong reasons in favour of any of the two.

5.1.4 Conclusions

The following strategy has been assumed for the development of Level 2 scientific code:

- for the level 2 retrieval algorithm we have adopted the option of the retrieval at tangent altitudes levels eventually followed by interpolation. The reason for this choice is that at the time of the code development we could not afford the unknowns of the other option.
- Some flexibility is maintained in the code in order not to prevent the implementation of the alternative approach in subsequent versions, in case it is recommended by retrieval studies.

The trade-off between vertical resolution and accuracy and the alternative of retrieving the profiles either at tangent altitude levels or at fixed (user-defined) levels have been fully addressed in a quantitative study carried-out in a parallel ESA contract (12055/96/NL/CN). We therefore refer to the final report of that study for a quantitative assessment of the topics described in the present section. In any case no explicit recommendation for a retrieval at fixed levels emerged from the mentioned study.

5.2 Use of a-priori information

The use of information provided by sources different from the spectroscopic measurements can increase the overall information content (equal to retrieval information plus extra information) and improve the quality of the retrieved profiles. This possibility is source of both, improvements and concern because, if on one hand it can lead to a positive result in the case of marginal accuracy in the retrieval, on the other hand it can become a cosmetic exercise which hides serious systematic errors. These two aspects of external information will be discussed in the next two sections.

5.2.1 - Accuracy improvement

The exploitation of external information is worthwhile only if it leads to a significant accuracy improvement. In order to understand the entity of the improvement, the mathematics of the combination of information is herewith briefly recalled.

It is well known that if two independent measurements q_1 and q_2 exist of a scalar quantity q , the two measurements can be combined by way of their r.m.s. errors σ_1 and σ_2 leading to the new estimate:

$$q_t = (\sigma_1^{-2} + \sigma_2^{-2})^{-1} (\sigma_1^{-2} q_1 + \sigma_2^{-2} q_2) \quad (5.2.1)$$

with an error:

$$\sigma_t = (\sigma_1^{-2} + \sigma_2^{-2})^{-1/2} \quad (5.2.2)$$

We know that in this case the error of the new estimate:

- is reduced by a factor $1/\sqrt{2}$ when the two measurements have the same error,
- is practically equal to the error of the best measurement when a large difference exists between the two errors.

In practice, combining information from different measurements brings no advantage when the measurements have different quality.

A similar error combination can be made in the case in which the measured quantity is a vector \mathbf{x} . The weighted combination \mathbf{x}^c of the two measurements \mathbf{x}_1 and \mathbf{x}_2 having respectively the variance-covariance matrices \mathbf{V}_1 and \mathbf{V}_2 is equal to:

$$\mathbf{x}^c = (\mathbf{V}_1^{-1} + \mathbf{V}_2^{-1})^{-1} (\mathbf{V}_1^{-1} \mathbf{x}_1 + \mathbf{V}_2^{-1} \mathbf{x}_2) \quad (5.2.3)$$

and has a variance-covariance matrix equal to

$$\mathbf{V}^c = (\mathbf{V}_1^{-1} + \mathbf{V}_2^{-1})^{-1} \quad (5.2.4)$$

The similarity of respectively expressions (5.2.1), (5.2.2) and (5.2.3), (5.2.4) may suggest that also similar properties apply, and it is not worthwhile to combine two measurements if their errors are very different. In our case this would imply that if the external information is better than the retrieved information we do not need the limb scanning measurements and if the external information is worse than the retrieved information we do not need to waste efforts combining the two. However, the situation is not so simple in the case of measurements of vectors and the considerations made for scalar quantities do not apply anymore.

Limb scanning observations often provide very good measurements with low errors at some altitudes, and undetermined measurements with large errors at other altitudes. An approximate estimate, which may be available from either statistical studies or models, of course does not directly add information where good measurements have been retrieved, but can reduce significantly the errors where the retrieved measurements are undetermined. Since a correlation exists between measurements at different altitudes, the reduction of the errors at some altitudes may lead to a reduction of all the errors. This explains why in the case of combination of vectors the errors may be reduced more than what is expected on the basis the quadratic combination.

In *Carlotti et al.* (1995) some numerical tests were performed showing that a significant improvement is possible by using a-priori knowledge such as that which can be obtained from either the previous measurement or statistical seasonal and geographical maps.

The use of external information can, therefore, be very profitable and should be seriously considered as part of the retrieval strategy.

5.2.2 - Systematic errors

A concern about the combination of the retrieved information with an a-priori estimate of the profile is that whenever the same a-priori information is used for several profiles, the error budget of each profile contains both a random and a systematic component (the first due to the measurement and the second due to the constant a-priori information).

Usually it is a good rule to list separately random and systematic errors in order to avoid mistakes in the subsequent operations. In fact in the case of averages random errors are reduced and systematic errors remain constant, while in the case of differences systematic errors cancel and random errors increase. If we want to maintain this separation between random and systematic errors it is necessary, therefore, to make retrievals without a-priori information (our primary output) and combine externally provided profiles with the retrieved profiles only optionally.

The different approach of using a-priori information routinely, during the retrieval iterations, can be adopted when the a-priori information has a random character.

5.2.3 - Hydrostatic equilibrium and LOS Engineering information

In the case of p, T retrieval an external information is also provided by the hydrostatic equilibrium. This information is a relationship between the unknowns and another measurement (provided by the engineering data) rather than a constant a-priori estimate and implies therefore different choices.

Hydrostatic equilibrium is a condition that applies to an ideal atmosphere which is perfectly stratified. The integrated form of the equation that describes this condition is:

$$p_n = p_0 \exp \left(- \frac{M}{R} g \sum_{i=0}^n \frac{1}{T_i} \Delta z_i \right) \quad (5.2.5)$$

where the notations are:

p_n pressure at a given altitude,

p_0 pressure at the reference altitude,

M average molecular weight of the atmosphere,

R universal gas constant,

T_i temperature of the atmospheric layer i ,

Δz_i thickness of the atmospheric layer i .

Now, since equation (5.2.5) is a relationship between T , p_n/p_0 and Δz_i increments, the measurements of pressure and temperature at tangent altitude obtained from the spectroscopic observations, can be used to get an estimate of the differences Δz_i between the tangent altitudes of the sweeps in the same limb-scanning sequence. Another estimate of the differences between tangent altitudes is provided by the engineering measurements. The two estimates are then combined using equation (5.2.3) and the variance covariance matrix of the new estimate is computed using (5.2.4). The variance covariance matrix related to the engineering pointings that is needed for the above operations is an input of p, T retrieval program. The mathematics required by this operation is discussed with further details in Sect. 4.2.6.

Note that in this approach, only the differences between tangent altitudes and not the absolute pointing altitudes are improved by the retrieval process. In fact, when pressure is a fitted quantity, the sensitivity of radiative transfer to the tangent altitude of the measurement is very weak and it is not possible to retrieve any of the tangent altitudes of the limb-scanning sequence.

Baselines:

Engineering LOS data will be routinely used inside p,T retrieval by exploiting the constraint provided by hydrostatic equilibrium law as explained in the above section.

The externally provided profiles of temperature, VMR, continuum cross-sections and instrumental offsets could be combined with the retrieved profiles only optionally after the retrievals are completed (this operation is not performed by the Level 2 algorithm).

5.3 Latitude effects

5.3.1 Angular spread

Here we will examine the extent of the angular spread of the beam of a limb scanning observation with respect to the earth centre. Exact computation gives for this value

$$\vartheta = 2 \arccos\left(\frac{R + H}{R + h}\right) \quad (5.3.1)$$

where R is the earth radius (supposed spherical for simplicity), H is the tangent height and h is the height for which the angle should be calculated. For a tangent height of 10 km and an atmospheric boundary of 100 km the result is 19° .

However, most of the emission originates from the lower layers. The integrated columns as a function of the angular spread are shown in Figure 5.1 for tangent altitudes 10 km and 17 km. This figure shows that the 90% of the emitting gas is concentrated within 4° , for a species with constant VMR.

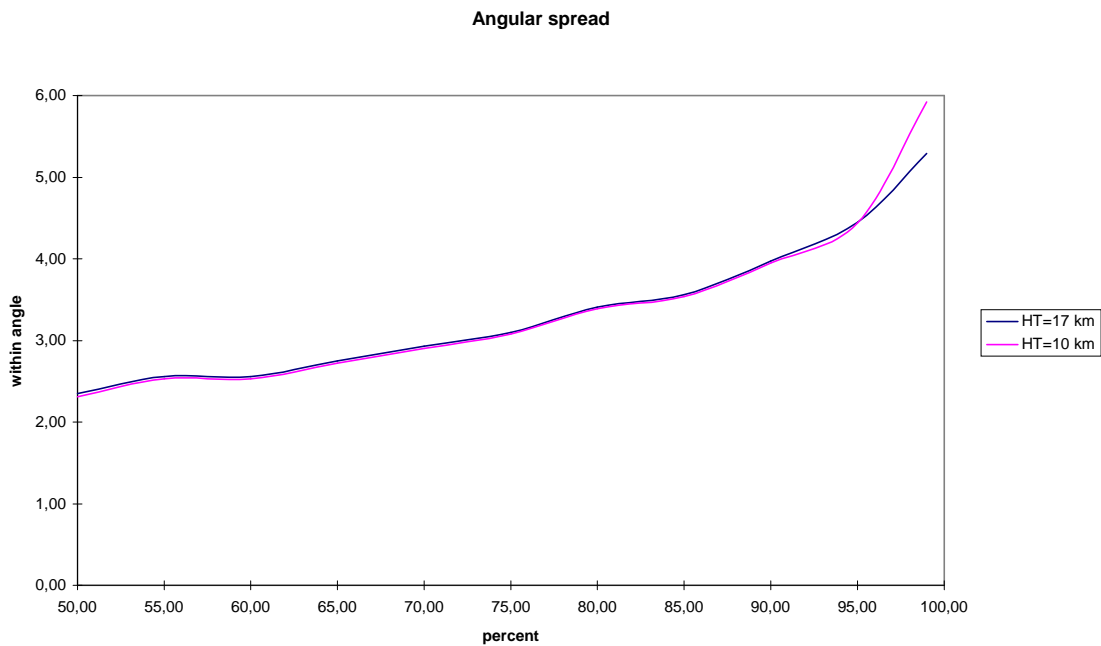


Figure 5.1: Angle [$^\circ$] at the earth centre over the percentage of the integrated column for two different tangent altitudes HT.

5.3.2 Climatological differences

As a sample on how much latitudinal variations may influence the final computation we show in the following figure the differences between temperatures (over the Northern hemisphere) at corresponding heights over different latitude ranges (spaced by 15°). (Data from the COSPAR International Atmosphere Reference (1986)).

Temperature changes

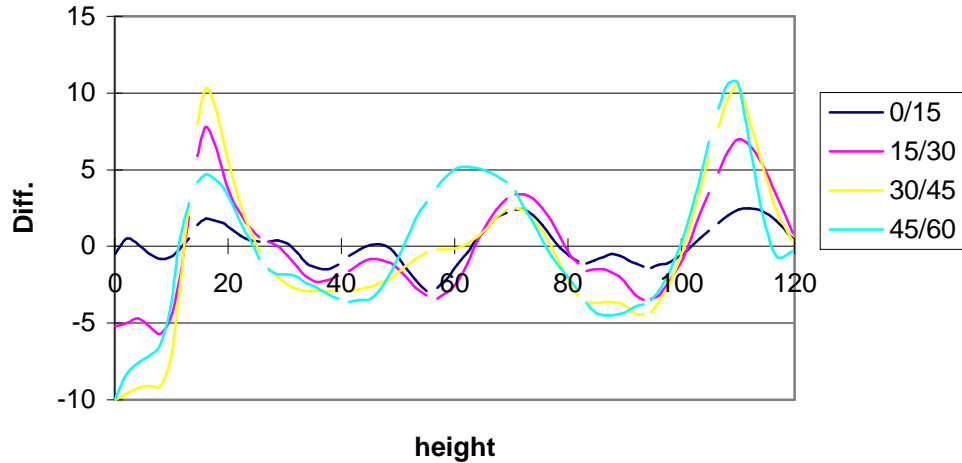


Figure 5.2: Temperature differences [K] as a function of the altitude [km] between latitudes separated by 15°.

Figure 5.2 shows that maximal temperature gradients are about 0.6-0.7 K/°latitude. This results in differences of about 2-3 K along the LOS where the main contribution to the emitted intensity originates.

Instantaneous latitudinal variations can be larger than these climatological differences, but taking into consideration the large amount of memory and computation time which are necessary to simulate horizontal gradients, the baseline of the NRT code is to perform retrievals of individual limb-scan sequences for which horizontal homogeneity is assumed.

A more quantitative evaluation of the effect of latitudinal gradients on the accuracy of the retrieved parameters is included in the final report of the ESA contract 12055/96/NL/CN. A further evaluation of the impact of horizontal homogeneity assumption in the forward model is included in *Carlotti et al. (2001)*. In that paper the authors also propose an innovative method that avoids the horizontal homogeneity assumption in the retrieval algorithm.

5.4 Earth model and gravity

5.4.1 Earth model

For the earth shape the simplified WGS84 model has been adopted. This model simulates the earth as an ellipsoid with the half-major axis $a = 6378.137$ km and the half-minor axis $b = 6356.752$ km. From the WGS84 model, also the local radius of curvature is derived.

5.4.2 Gravity

For the calculation of the gravity as a function of altitude and latitude a formula taken from *Clarmann (1986)* has been adopted.

The acceleration of gravity at the sea level (identified by the WGS84 ellipsoid), as a function of geodetic latitude Φ can be computed using the following empirical formula which includes the centrifugal effect:

$$g_0 = 9.80616 \cdot [1 - 0.0026373 \cdot \cos(2\Phi) + 0.0000059 \cdot \cos^2(2\Phi)] \quad (5.4.1)$$

Since the centrifugal component of g_0 has a different dependence on the altitude compared to the gravitational component, it is then necessary to separate the two components in order to properly insert in g_0 the dependence on the altitude. Let us define:

$$\tilde{g} = g_0 + \Omega^2 \frac{f^2}{R} \cos^2 \Phi \quad (5.4.2)$$

where \tilde{g} is the gravitational component and does not contain the centrifugal effect; Ω is the angular speed of the earth and is equal to:

$$\Omega = \frac{2\pi}{86400 \text{ sec./day}} / 1.002737904 \quad (5.4.3)$$

the factor 1.002737904 is the star time rotation factor which takes into account the motion of the earth along the orbit. Furthermore, the meaning of R and f is explained in Fig. 5.3 and they are computed using the following formulas:

$$f = \frac{a}{\sqrt{1 - \left(1 - \frac{b^2}{a^2}\right) \sin^2 \Phi}} \quad (5.4.4)$$

$$R = \sqrt{f^2 \cos^2 \Phi + \left(\frac{b^2}{a^2} f \cdot \sin \Phi\right)^2} \quad (5.4.5)$$

where a and b are respectively the equatorial and the polar radius of the earth.

The dependence on the altitude is then included in g :

$$g = \tilde{g} \left(\frac{R}{R+z} \right)^2 - \Omega^2 f \left(\frac{f+z}{R} \right) \cos^2 \Phi \quad (5.4.4)$$

In the above equation, z is the altitude of the considered point with respect to the sea level.

The results provided by equation (5.4.4) have been compared with the values of the gravity provided by other models (see e.g. *List (1963)*, *Stern (1960)*, *Defense (1987)*): the observed relative differences are always smaller than $2 \cdot 10^{-4}$.

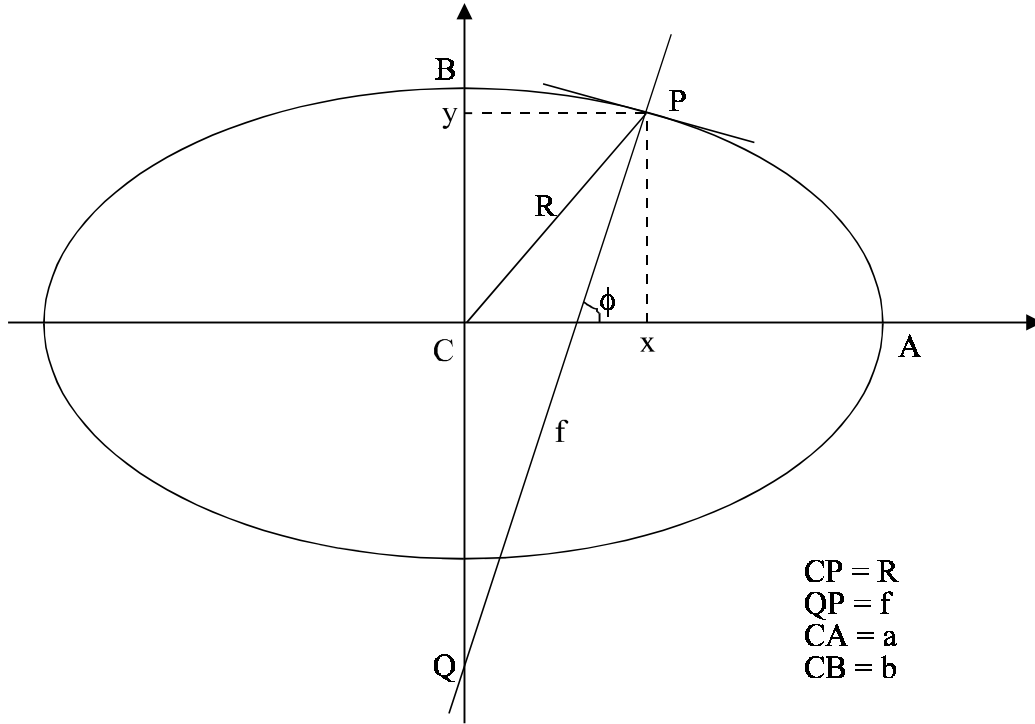


Fig. 5.3: Earth model

5.5 Ray-tracing and refractive index

Due to refraction, the ray-path bends towards the earth. Therefore the radiative transfer integral is a curvilinear integral along the line of sight, determined by the viewing direction of the instrument and by the refractive index of the air.

5.5.1 Pressure-temperature dependence of refractive index and effects on tangent heights

Refractive index is a function of pressure, temperature and water vapour content: this dependence can be defined using the Clausius-Mossotti (or Lorentz-Lorenz) formula (see e.g. *Born and Wolf*, 1975):

$$\frac{n^2 - 1}{n^2 + 2} = \text{const} \cdot \rho \quad (5.5.1)$$

where n is the refractive index, ρ the air density.

Since $n - 1 \ll 1$, we can write:

$$n = 1 + \alpha \cdot \rho(p, T). \quad (5.5.2)$$

Where α is a factor dependent on the empirical model chosen for the refractive index. If we consider the Barrel - Sears' s empirical formula:

$$(n - 1) \cdot 10^{-6} \approx \left(77.48 + \frac{0.44}{\lambda^2} + \frac{0.007}{\lambda^4} \right) \cdot \frac{p}{T} - \left(12.79 - \frac{0.14}{\lambda^2} \right) \cdot \frac{e}{T}, \quad (5.5.3)$$

where ‘ e ’ is the water vapour pressure, we see that the dependence on the water vapour content can be neglected because its maximum contribution to the value of n is about 0.02% (this value refers to height 5 km, where water VMR is equal to $1.4 \cdot 10^{-3}$ for the standard atmosphere).

Besides also the frequency dependence of the refractive index can be neglected in the mid-infrared. Different models can be used for the determination of refractive index, for instance Edlen (1966) or SAO (*priv. com.*), each of which finds a different value of α .

The effect of refraction in the evaluation of tangent height is shown below.

Assuming a spherical earth and spherical atmospheric layers, the following relation is valid for all the points on the line of sight:

$$n(r) \cdot r \cdot \sin(\theta(r)) = \tilde{c} = n(r^t) \cdot r^t, \quad (5.5.4)$$

where r is the summation of the local radius of curvature of the earth (which is an input of the ORM) and the altitude of the considered point referred to the surface of the earth; r^t is the value of r at the tangent point, θ is the local zenith angle. r^t is an input of the ray tracing module and $n(r^t)$ is computed using the actual p,T profiles.

Namely, the previous expression represents the Snell’s law in the case of spherical geometry. Considering an atmospheric spherical layer as drawn in fig. 5.4, Snell’s law can be written in point P as:

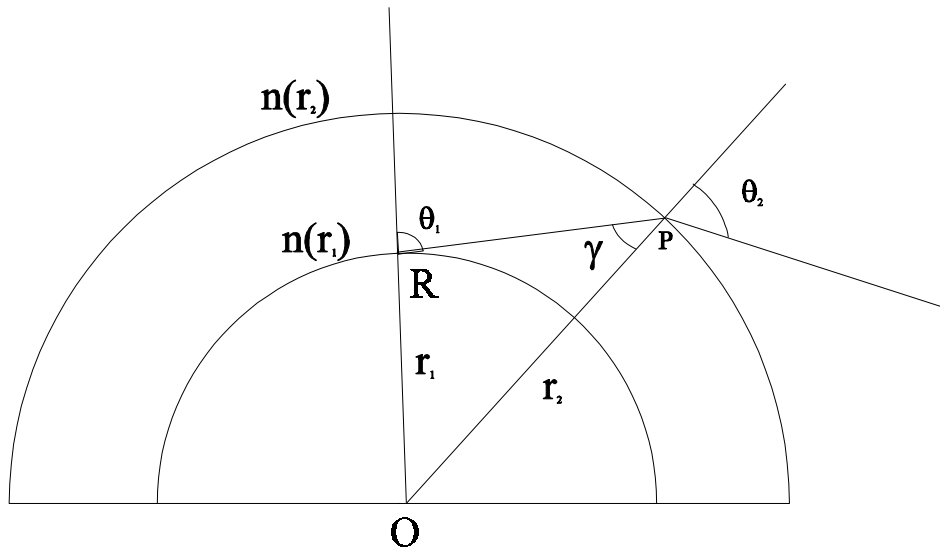


Figure 5.4: Deduction of Snell’s law for media with spherical symmetry.

$$n(r_2) \cdot \sin \theta_2 = n(r_1) \cdot \sin \gamma \quad (5.5.5)$$

From the ‘sine’ theorem applied to triangle OPR we obtain:

$$\frac{r_1}{\sin \gamma} = \frac{r_2}{\sin(\pi - \theta_1)} \quad (5.5.6)$$

Combining the two equations we get:

$$r_2 \cdot n(r_2) \cdot \sin \theta_2 = r_1 \cdot n(r_1) \cdot \sin \theta_1 \quad (5.5.7)$$

from which eq (5.5.4) is demonstrated.

From eq. (5.5.4) we find that, keeping the same flight altitude and the same limb angle, the tangent height (referred to the centre of the earth) r^r for a refracted beam is related to the height of a non-refracted r^0 beam by:

$$r^r = r^0/n \approx r^0 \cdot (1 - \alpha\rho) \quad (5.5.8)$$

Inserting the numerical values ($\alpha\rho \approx 1.17 \cdot 10^{-4}$ at 8 km) one finds shifts in the expected tangent heights ($h^r - h^0 \approx -\alpha\rho \cdot R$, with R equal to earth radius) as large as 0.5 - 1.0 km (error increasing towards lower tangent heights).

On the other hand, if the refraction has a large effect on the tangent height, the choice of the model, e.g. the definition of α value, is not significant: the difference between the tangent altitudes determined using two different models (SAO and Edlen) can be obtained from the following relation

$$r_1^r - r_2^r = r^0 \cdot (\alpha_1 - \alpha_2) \cdot \rho \quad (5.5.9)$$

to be:

$$h_1^r - h_2^r = (h^0 + R) \cdot (\alpha_1 - \alpha_2) \cdot \rho. \quad (5.5.10)$$

Since the percentage difference between the two considered models is about 0.2 %, we can say that both models are equivalent for our aims.

Discrepancy SAO-EDLEN

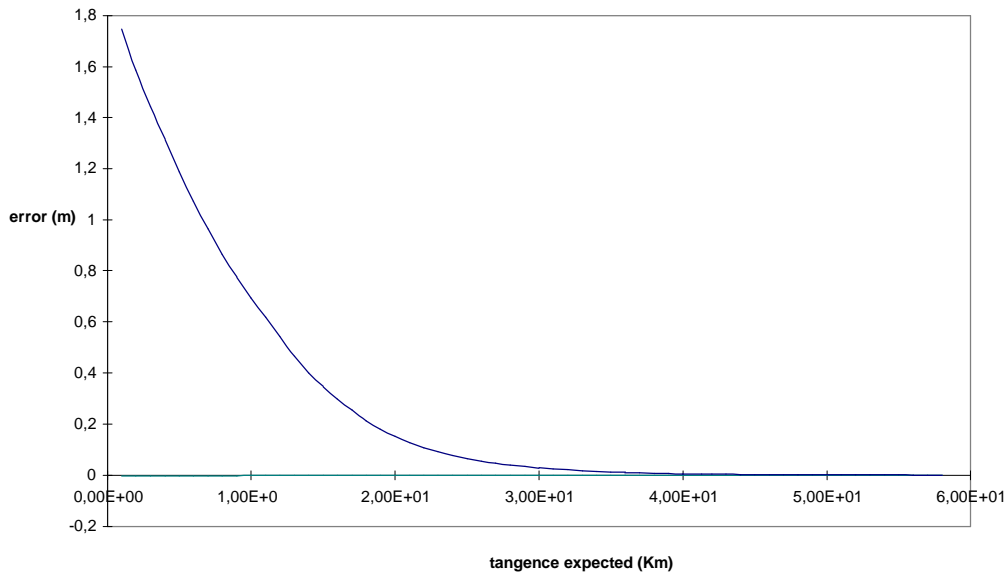


Figure 5.5: Difference in the tangent height between different models for the refractive index (SAO-Edlen).

The differences on the tangent heights found using two models of refractive index is shown in Fig. 5.5 to be about 1 m. This error may therefore be neglected.

The model used in the scientific code, for refractive index is the simplified version of Edlen:

$$n = 1 + 0.000272632 \cdot \frac{\rho}{\rho^0}, \quad (5.5.11)$$

with $\frac{\rho}{\rho^0} = \frac{p}{T} \frac{T^0}{p^0}$, $p^0 = 1013.25$ hPa and $T^0 = 288.16$ °K.

The refractive index is calculated for each pressure and temperature, therefore no interpolation is necessary.

5.5.2 Method of ray tracing

As discussed in Sect. 6.1, the preferred integration variable of the radiative transfer equation is the altitude. In this case the expression for dx / dr must be calculated. It is therefore necessary to determine the expression of optical path x as a function of the altitude and the refractive index. In order to define the ray path in each position it is necessary to know the local zenith angle $\theta(r)$. Since the layers are assumed as spherical, eq.(5.5.4) can be used to calculate θ .

Since the following relation is valid:

$$dr = dx \cdot \cos(\theta), \quad (5.5.12)$$

from eq.(5.5.4) and eq.(5.5.8) we obtain:

$$dx = \frac{1}{\sqrt{1 - \frac{\tilde{c}^2}{n^2(r) \cdot r^2}}} \cdot dr \quad (5.5.13)$$

This formula has a singularity at tangent altitude, however the singularity can be removed by changing the variable of integration from r to y where:

$$y = \sqrt{r^2 - (r^t)^2} \quad (5.5.14)$$

is the geometrical distance from the tangent point.

In this case we obtain:

$$dx = \frac{1}{\sqrt{1 + \frac{(r^t)^2}{n^2(r)} \cdot \frac{n^2(r) - n^2(r^t)}{y^2}}} \cdot dy \quad (5.5.15)$$

wherein the singularity is not present.

In conclusion we use:

1. Exact model for ray tracing. No effort has been made to introduce simplifications in these calculations because only a small fraction of the total forward model computing time is spent for ray tracing (0.1% when the 20 microwindows of p,T retrieval are simulated).
2. exact model of optical refraction for spherical symmetry.
3. negligible approximation in refractive index value

5.6 Line shape modelling

The line shape function which has to be modelled appropriately inside the forward model is the Voigt profile (equation 4.4.15), equal to the convolution of the Doppler profile caused by the velocity of the molecules and the Lorentz profile induced by collisions. While the Doppler function is the correct description of the physical effect of Doppler shift, the Lorentz profile is an approximation that is valid if:

- the spectral region under consideration is not far from the line centre and
- the distance from other lines of the molecule is so large that the line can be considered as isolated

The first assumption fails if we try to model lines far away from the centre. Two important examples for this are the sub-Lorentzian shape of the line wings of CO₂, or the super- and sub-Lorentzian behaviour of H₂O. These effects are normally modelled by introducing a (experimentally determined) χ -factor into the line-wing description (see 5.6.3).

The second assumption causes major problems in the modelling of Q-branches since the lines are very close to each other and molecular collisions lead to an intensity transfer between the transitions. This effect is called line-mixing (see also 5.7).

5.6.1 Numerical calculation of the Voigt profile

Since the convolution integral of the Voigt profile cannot be evaluated in an analytical form it has to be calculated numerically. In order to fulfil this task several algorithms were developed and compared (e.g. Schreier, 1992) with regard to speed and accuracy, especially for the application in line-by-line models. Due to the recommendations of the latter intercomparison and our own tests regarding computation time of different approaches (see Table 5.1) it was decided to use the algorithm described by Humlicek (1982). This routine calculates a rational approximation (with a relative accuracy of 10^{-4}) of the complex probability function:

$$w(z) = e^{-z^2} \left(1 + \frac{2i}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \right) = K(x, y) + iL(x, y) \quad (5.6.1)$$

with $z = x + iy$

$K(x, y)$ is the convolution integral of the Voigt function from equation (4.4.15).

Method	Relative run time
Humlicek (Humlicek, 1982)	1
AFGL (Clough et al., 1981)	1.2
Drayson (Drayson, 1976)	3.0

Table 5.1: Run time comparison between different approaches for the calculation of the Voigt lineshape.

As a baseline Humlicek (1982) has been implemented in the scientific code. A more sophisticated and optimised version of the Humlicek algorithm has recently been developed at IMK by M.Kuntz (priv. com.). However, the algorithm currently implemented in the scientific code will not be upgraded because, as it will be shown in Sect. 6.11, the final baseline of MIPAS Level 2 processor is to compute cross-sections using compressed Look-Up Tables (LUTs).

5.6.2 Approximation of the Voigt profile by the Lorentz function

Since the calculation of the Voigt function is much more time consuming than the Lorentz function it is reasonable to use the Voigt shape only near the line centre where the differences between both are relatively large. Outside this region the Lorentz line shape can be used. The criterion that can be used for the application of the different functions is the relative error as a function of the distance from the centre in multiples of the Doppler half width α^D . The maximum relative errors are:

Distance from the line centre	Relative error (Lorentz-Voigt)/ Voigt
$10 \alpha^D$	-1.5%
$20 \alpha^D$	-0.37%
$30 \alpha^D$	-0.17%
$40 \alpha^D$	-0.10%
$50 \alpha^D$	-0.06%

Table 5.2: The maximum ($y \rightarrow 0$ in eq. 4.4.15) relative error between the Voigt and the Lorentz function as a function of the difference from the line centre.

Besides, test calculations using the $30\alpha^D$ boundary for 6 selected p-T microwindows showed maximum differences in the order of NESR/120. In these cases the time saving was 60% with respect to the calculation with the Voigt-profile.

The baseline is to use the Voigt profile only within 30 Doppler half-widths from the line centre. A less conservative transition can be considered if further computing time saving is found to be necessary.

5.6.3 χ -factors in the case of CO_2 and H_2O

In order to describe the sub-Lorentzian behaviour of the CO_2 lines χ -factors were experimentally determined e.g. by Cousin et al. (1985) for the CO_2 ν_3 band head at $4.3 \mu\text{m}$. These factors start from unity at the line centre and remain 1 until $0.5\text{-}5 \text{ cm}^{-1}$ distance (temperature dependent). Afterwards they decay exponentially.

The χ -factor for H_2O represents the super-Lorentzian behaviour of water vapour until some 100 cm^{-1} from the line centre and the sub-Lorentzian shape beyond (Clough et al., 1989).

The behaviour of the χ -factors of being equal 1 up to some wavenumbers from the line centre allows us to disregard them inside one microwindow. On the other hand the factors are considered for those lines which contribute as near continuum (see Sect. 5.11.2) to the radiation inside the microwindow.

5.7 Line-mixing

Line mixing, known also as line interference, line coupling, collision narrowing, Q-branch collapse, corresponds to the deviation of the measured line shape from the Lorentzian function (generally in regions with dense rotational structures, but effects in microwindows of transparency in vibration-rotation bands have been observed as well).

For practical calculations the suggested line shape (Rosenkranz, 1975) is:

$$A(\sigma, \sigma_l) = \frac{1}{\pi} \frac{\alpha_l^L + (\sigma - \sigma_l) p Y_l}{(\sigma - \sigma_l + \text{Im}\{\Gamma_{ll}\})^2 + \alpha_l^{L^2}} \quad (5.7.1)$$

with the first order coupling coefficient:

$$Y_l = 2 \sum_{l' (l' \neq l)} \frac{d_{l'}}{d_l} \cdot \frac{\Gamma_{l'l}}{\sigma_l - \sigma_{l'}} \quad (5.7.2)$$

Γ is the (frequency dependent) relaxation matrix, with diagonal elements determining the shape of uncoupled lines ($\text{Re}\{\Gamma_{ll}\} = \alpha_l^L$, the Lorentz half widths) and the lineshifts ($-\text{Im}\{\Gamma_{ll}\}$), and off-diagonal elements responsible for non additive effects (line mixing) when the lines overlap, and d_l is the reduced matrix of the dipole moment.

The Rosenkranz expression for the line shape is easily convoluted with the Doppler function. The modified Voigt function resulting from this convolution may be written in terms of the real ($K(x,y)$) and imaginary parts ($L(x,y)$) of the complex error function $w(z)$ which are calculated by the Humlicek algorithm (equation 5.6.1):

$$A_l^{V_{LM}}(\sigma - \sigma_l) = \sqrt{\frac{\ln 2}{\pi}} \frac{1}{\alpha_l^D} [K(x_l, y_l) + p Y_l L(x_l, y_l)] \quad (5.7.3)$$

The baseline for Level 2 processor is not to use Q-branches where line-mixing is known to have strong effects. These spectral regions will be avoided with an appropriate choice of the microwindows, so that line mixing does not need to be simulated.

5.8 Pressure shift

Beside the line mixing effect, equation (5.7.1) contains the pressure shift in the form of the term $-\text{Im}\{\Gamma_{ll}\}$ which is proportional to the atmospheric pressure p . Since pressure shift data is only given for CH_4 and for CO_2 above 2300 cm^{-1} in the HITRAN96 data base we will not implement it into the forward code. However, if complete data on the shift will become available it can easily be inserted into the code.

5.9 Implementation of Non-LTE effects

The recommendation arising from the final report of Non-LTE study (Bologna, January 23, 1996) is that no inclusion of externally provided vibrational temperatures for target transitions is needed in operational processing. The impact of Non-LTE effects in p , T and VMR retrievals is reduced by using an appropriate selection of microwindows.

The capability of handling NLTE is therefore not implemented in the OFM / ORM.

5.10 Self broadening

The Lorentz half width from equation (4.6.12) includes both, foreign α^{L_f} and self broadened α^{L_s} components:

$$\alpha^{L_0} = (1 - X)\alpha^{L_f} + X\alpha^{L_s} \quad (5.10.1)$$

With the volume mixing ratio X .

The relative error which is done when neglecting the self broadening is $X(\alpha^{L_s}/\alpha^{L_f} - 1)$. Table 5.3 gives the maximum errors of the half width when assuming maximum values for X and maximum values for the quotient of self and foreign broadened half widths.

Gas	Max. $\alpha^{L_s}/\alpha^{L_f}$	Max. X (8-50 km) [ppmv]	Max. error [%]
CO ₂	1.3	360	0.011
O ₃	1.3	7	0.0002
H ₂ O	10 5 (average)	700 (tropical) 200 (mid-latitude)	0.7 0.08 (average/mid-lat.)
CH ₄	1.3	1.7	0.00005
N ₂ O	not in HITRAN92	-	-
HNO ₃	not in HITRAN92	-	-

Table 5.3: Maximum errors in the Lorentz half width when neglecting the self-broadening. Data was taken from HITRAN92 considering the entire range 685-2410 cm⁻¹

From Table 5.3 it is evident, that the self-broadening is negligible for all target species except water vapour in the tropical troposphere where maximum errors of about 0.7% can occur. However, this maximum error corresponds (through the Lorentz line shape formulation) directly to a pressure error of 0.7% which is one third of our acceptance criteria for approximations (2%, see chapter 3). Therefore our baseline is to disregard the self-broadening in all cases.

5.11 Continuum

We have to distinguish between three different kinds of effects which contribute to the spectral intensity of a microwindow as continuum:

- the instrumental continuum
- the near continuum
- the far continuum

These effects and their sources will be discussed first. It is described how they can be simulated in the light of the objectives of this study. For their simulation by the forward model it will be necessary to distinguish between the self standing forward model and the model included in the retrieval code.

5.11.1 Instrumental continuum

This continuum contribution is caused by the instrument itself. Its effect on the spectrum is a pure additive offset. The reasons for an instrumental continuum are manifold - e.g. self emission of the

instrument, scattering of light into the instrument, or 3rd order non-linearity of the detectors. These distortions are corrected during the calibration of the level 1b data processing but present specifications indicate that the residual instrumental continuum averaged over the spectral range of the microwindow can be larger than the measurement error. Therefore we plan to model the remaining radiometric errors.

The simulation of the instrumental continuum can be performed by adding a wavenumber dependent offset to the spectrum.

Our baselines are:

- for the self-standing forward model to disregard this effect since this program should only simulate atmospheric contributions.
- for the retrieval to assume that the instrumental offset does not vary with changing limb scan angles and fit for each microwindow only one instrumental continuum offset value.

5.11.2 Near continuum

This contribution to the intensity inside a microwindow is caused by nearby atmospheric lines. Therefore, the simulation of this effect has to be performed during the calculation of the absorption cross sections. The different possibilities for its calculation are:

1. explicit calculation of the wings of the lines at each fine-grid point inside the microwindow (see 5.6).
2. calculation only at three grid points inside the microwindow and parabolic interpolation in between

We plan to simulate the effect of nearby lines using a switch provided by the spectroscopic database to decide if the line has to be calculated explicitly (option 1) or can be interpolated inside the microwindow (option 2). Baseline is to use option 2.

The same option is used for both self standing and retrieval forward model.

5.11.3 Far continuum

This term includes all continuum-like contributions which are not included in the previous two definitions. These are e.g. the line wings of far lines (the most important contribution here is from H₂O), the pressure broadened bands of O₂ at 1550 cm⁻¹ and N₂ at 2350 cm⁻¹ and the absorption by aerosols.

For this continuum we have to distinguish between the self standing forward model and the one implemented into the retrieval code.

Self standing forward model:

The self standing forward model must produce realistic simulations of the atmospheric spectra that include the continuum.

We decided to use the same continuum as in FASCOD (Clough et al., 1989). The water continuum is described in Clough et al. (1989). Herein the continuum cross section $k^{cont_{H_2O}}(\sigma)$ is given by the sum of the far line wings using the Van Vleck and Huber line shape function which is modified in order to fit the experimental data of Burch and Alt (1984):

$$k^{cont_{H_2O}}(\sigma) = \sigma \tanh\left(\frac{hc\sigma}{2k^BT}\right) \sum_{l(ines)} \frac{S_l}{\sigma_l \tanh(hc\sigma_l / 2k^BT)} [f(\sigma, \sigma_l) \chi(\sigma, \sigma_l) + f(-\sigma, \sigma_l) \chi(-\sigma, \sigma_l)] \quad (5.11.1)$$

where

$$f(\pm\sigma, \sigma_l) = \frac{1}{\pi} \frac{\alpha_l^L}{\alpha_l^{L^2} + (\sigma \pm \sigma_l)^2} \quad \text{if } |\sigma \pm \sigma_l| \geq 25 \text{ cm}^{-1} \quad (5.11.2)$$

$$f(\pm\sigma, \sigma_l) = \frac{1}{\pi} \frac{\alpha_l^L}{\alpha_l^{L^2} + 25^2} \quad \text{if } |\sigma \pm \sigma_l| \leq 25 \text{ cm}^{-1} \quad (5.11.3)$$

and $\chi(\sigma, \sigma_l)$ is the χ -factors of water.

For the implementation into the forward model the following equation is used:

$$k^{cont_{H_2O}}(\sigma) = \sigma \tanh\left(\frac{hc\sigma}{2k^BT}\right) \frac{\eta}{\eta^0} \left(X^{H_2O} C^{0s}(\sigma, T) + (1 - X^{H_2O}) C^{0f}(\sigma, T) \right) \quad (5.11.4)$$

$C^{0s}(\sigma, T)$ and $C^{0f}(\sigma, T)$ are the continuum absorption parameters for the self and the foreign broadening at the reference number density η_0 . η is the actual air density. To determine the temperature dependence of the self broadening values, exponential interpolation between the tabulated parameters $C^{0s}(\sigma, 260 \text{ K})$ and $C^{0s}(\sigma, 296 \text{ K})$ is performed. For the foreign broadening $C^{0f}(\sigma, 296 \text{ K})$ is used for all temperatures. The interpolation in frequency is performed linearly.

This continuum description includes only contributions of lines farther than 25 cm^{-1} from the line centre. Since the self standing forward model will use the spectroscopic database which uses various selection criteria for the lines, not all water lines within 25 cm^{-1} may be considered. However, it has been estimated that the error due to this fact is less than $0.175 \times \text{NESR}$ (and in most cases much less than this). Because this is presumably smaller than the absolute accuracy of the continuum model and because we don't have to describe the continuum with very high precision our baseline is to use the spectroscopic database. If higher accuracy is needed, the missing water lines can be added to the data-base.

The N_2 continuum is parameterised, temperature independently, every 5 cm^{-1} between 2020 and 2800 cm^{-1} for a reference number density η_0 . In order to calculate it for the actual pressure and temperature it has to be multiplied by the ratio η/η_0 , where η is the actual number density, and linearly interpolated to the wavenumber. For a details on the implemented N_2 continuum model, please refer to Lafferty et al., (1996).

The O_2 continuum is given in the form of three parameters (one strength and two for the temperature correction) from 1395 - 1760 cm^{-1} . For a detailed description of the adopted model of O_2 continuum, please refer to the paper of Thibault et al., (1997).

Forward model in the retrieval:

The forward model implemented into the retrieval code does not have to simulate any far-continuum effects. These are included into a single atmospheric continuum for each microwindow which is fitted like an additional absorption cross section. This leads to one cross section at each atmospheric layer for each microwindow. In order to limit the number of retrieved parameters we have implemented the following constraint.

For each microwindow at each altitude the frequency range is defined in which the continuum is expected to vary linearly with frequency. This frequency range is used by the retrieval routine to establish the constraints between retrieved continua.

Therefore:

- an additional field is required for each microwindow (in the microwindow-definition file) for a real constant that specifies the size of the frequency interval in which atmospheric continuum can be assumed to have a linear variation.
- this parameter is used to establish a rule of ‘tight contiguity’ and ‘loose contiguity’ between the selected microwindows. The expression ‘tight contiguity’ means that the same constant continuum value can be assumed for the microwindows. Two microwindows have tight contiguity when their separation is less than a suitable fraction of their intervals of linearity. Microwindows are assumed to have ‘loose contiguity’ when their separation is less than their interval of linearity.
- at each altitude the microwindow will be classified as:
 1. single (no tight contiguity, or loose contiguity with only one mW)
 2. double (tight contiguity with one mW)
 3. multiple at the edge of a group (multiple mWs are those that have a loose contiguity with at least two mWs; the group is defined as all the elements contiguous to the first multiple element as encountered during a search that starts from a frequency end; double mWs count as one mW in this search; elements at the edge of a group are the first and last element of the group; the last element of one group can be the first element of another group)
 4. multiple in between a group
- at each altitude, one continuum value will be retrieved for each class 1) microwindow, for each pair of class 2) microwindows, and for each class 3) microwindow. The interpolated values will also contribute to the fit.

Only recently, the algorithm for automated selection of spectral microwindows (MWMAKE, see Appendix C and Bennett et al., 1999) started using the fitted atmospheric continuum parameters to compensate for various systematic error sources that introduce in the spectrum a continuum-like effect. For this reason, the fitted “continuum” parameters have lost their physical meaning and no constraints can be applied to their spectral behavior.

5.12 Interpolation of the profiles in the forward / retrieval model

Since both initial guess and retrieved profiles are often represented on a grid much coarser ($\approx 3\text{km}$) than the grid used for the discretization of the atmosphere ($\approx 1\text{km}$, for the radiative transfer calculation), a set of interpolation rules must be established for the various types of profiles.

In the scientific code, pressure and temperature profiles are always constrained to obey the hydrostatic equilibrium equation. Tests have shown that, using any of the atmospheric standard models, neglecting the temperature gradient leads to negligible errors when layers thinner than 6 km are adopted and the temperature in the middle of the layer is used as the representative temperature of the layer.

When the altitude is the independent variable, it is common to consider the temperature and the VMR profiles as varying linearly with altitude, and pressure as varying accordingly to the hydrostatic equilibrium law (i.e. exponentially with the altitude if the temperature is assumed locally constant for the calculation of pressure).

Now, since in our case the independent variable is pressure, the most appropriate interpolation rule is logarithmic interpolation for both temperature and VMR profiles.

The baseline is therefore to use logarithmic interpolation of temperature and VMR profiles whenever an interpolation is required.

Regarding continuum cross-section profiles: whenever an interpolation is required, since the continuum emission behaves as the square of pressure (Clough et al., 1989), and since the fitted continuum cross-section profiles are multiplied by the air column (proportional to pressure) in order to obtain continuum emission, it is then reasonable to interpolate continuum cross-section profiles linearly with pressure.

5.13 Interpolation of the “retrieved” profiles to a user-defined grid

5.13.1 Description of the problem

The baseline of the Level 2 scientific processor is to retrieve the target profiles in correspondence of the tangent pressures. However the users of MIPAS products may need to resample the retrieved profiles to a user-defined set of pressures. The problem arises therefore of providing a recommended interpolation scheme consistent with the Level 2 algorithm assumptions.

The first possibility is to recommend the use of the same set of interpolation rules used in the retrieval processor (see Sect. 5.12). However, considering that Level 2 profiles have been derived from limb-scanning measurements, the a-posteriori interpolation process should generate profiles that, when provided in input to the forward model, reproduce the simulated spectra calculated in the retrieval at convergence. Therefore, considering that the simulated limb-radiances strongly depend on the total column amounts above the tangent point, the interpolation scheme should maintain the total column. In other words the column calculated for the interpolated points should be equal, within some pre-defined tolerance, to the column obtained with the original data profile.

We recall that the vertical column at a particular altitude r_{z1} is defined as

$$Col = Cost \int_{r_{z1}}^{r_{latm}} X_{gas} \frac{P(z)}{T(z)} dz \quad (5.13.1)$$

where r_{latm} is the altitude of the upper boundary of the atmosphere and X_{gas} , $P(z)$ and $T(z)$ are the gas VMR, pressure and temperature respectively.

The quantities X_{gas} , $P(z)$ and $T(z)$ are measured for a discrete altitude grid, and hence in order to have a realistic result we must calculate the total column as the sum of partial columns between the altitudes of this grid. In turn these partial columns can be calculated interpolating the pressure, temperature and VMR profiles between the two edge altitudes: linear interpolation is used for Temperature and VMR, and exponential interpolation for the pressure.

Therefore the problem of the interpolation of the VMR profiles becomes one of finding an appropriate transformation of the VMR profile that satisfies the constraint of an unchanged vertical column.

5.13.2 Strategy

The strategy adopted consists of constructing a new grid of VMR points and to impose a one-to-one correspondence of partial columns with the original profile. By finding the VMR values that satisfy this condition, a profile is identified at the new grid points. The latter will be different of the one obtained with a classical interpolation rule.

A ‘classical’ interpolation law is defined in order to identify from the discrete output of the ORM a continuous representation of the atmosphere. For each value of the independent variable pressure, the corresponding altitude and temperature can be obtained by means of hydrostatic equilibrium equation, assuming a linear dependence of temperature on altitude.

In turn, the VMR values at each user-defined altitude can be obtained by linear interpolation between the two nearest retrieved values.

On the basis of this continuous representation partial columns and hence total columns can be defined with respect to both the retrieval grid and the user defined grid in the following way:

1. First partial columns are calculated on the retrieval-grid
2. Then the VMR values are calculated at the user-defined grid using linear interpolation with the altitude.
3. Then the partial columns at the retrieval-grid are calculated following a spline that joins the user-defined grid points only.
4. We then vary the VMR values at the user-defined grid until the difference between the partial columns of the new profile and of the original profile are minimum. The resulting profile by definition has the property of maintaining the vertical column unmodified.

The variation of the VMR values on the user-defined grid are calculated using a non-linear least-squares fitting procedure. The quantities to be fitted are the partial columns on the retrieval grid. The parameters to be fitted are the VMR values at the user-grid pressure levels. The fitting method is the Gauss-Newton method: if \mathbf{n}_v is a vector containing the difference between the observed partial columns and the calculated partial columns, then the correction \mathbf{y}_v to be applied to the values of the VMR at the user-grid to minimise this difference is given by:

$$\mathbf{y}_v = \mathbf{D}_v \mathbf{n}_v \quad (5.13.2)$$

where $\mathbf{D}_v = \left(\mathbf{K}_v^T \mathbf{V}_{col}^{-1} \mathbf{K}_v \right)^{-1} \cdot \mathbf{K}_v^T \mathbf{V}_{col}^{-1}$ and \mathbf{K}_v is the Jacobian matrix and \mathbf{V}_{col} is the Variance Covariance Matrix (VCM) of the measured columns.

Since the values of the partial columns are numbers that change by several orders of magnitude from the highest altitude to the lowest, instead of fitting their values we fit the logarithm of their values. So \mathbf{V}_{col} becomes the VCM of the logarithm of the partial columns $\mathbf{V}_{\log(col)}$.

An output of the retrieval code is the VCM of the measured columns $col(i)$ at the various tangent altitudes i . To get the VCM of the logarithm of the columns we apply the following transformation:

$$\left(\mathbf{V}_{\log(col)} \right)_{i,j} = \frac{\left(\mathbf{V}_{col} \right)_{i,j}}{col(i) * col(j)} \quad (5.13.3)$$

The errors associated with the solution of the fit are given by the squared root of the diagonal elements of the VCM of the solution (\mathbf{V}) given by:

$$\mathbf{V} = \left(\mathbf{K}_v^T \mathbf{V}_{\log(col)}^{-1} \mathbf{K}_v \right)^{-1} \quad (5.13.4)$$

After solving equation (5.13.2) the VMR profile on the user-defined grid will be given by the new vector \mathbf{VMR}_{int} :

$$\mathbf{VMR}_{int} = \mathbf{VMR}_{linint} + \mathbf{y}_v \quad (5.13.5)$$

The fitting procedure is repeated until the sum of the squares of the differences between the new columns and the measured ones doesn't change of more than 1%. In the ORM code the retrieved profile above the highest fitted altitude is obtained by scaling the initial-guess profile of the same quantity used for the highest fitted point. The same procedure has to be applied to the interpolated

points above the highest fitted altitude or to the highest point of the user-defined grid if it happens to be below the highest fitted altitude.

The errors associated with the new VMR values are characterised by the variance-covariance matrix V provided by equation (5.13.4).

5.14 Optimized algorithm for construction of initial guess profiles

The scientific version of this function has not been implemented because the scientific code has only the capability of handling individual scans, whereas the elaboration of a suitable strategy for handling initial guess profiles requires consideration of the whole orbit. Data analysis at the ‘orbit’ level is handled only by the industrial Level 2 processor.

For the analysis of a given scan, the ORM code needs the following atmospheric profiles as input:

1. Pressure and temperature profiles
2. Continuum profiles for microwindows used in p,T retrieval
3. H₂O VMR profile
4. Continuum profiles for microwindows used in H₂O retrieval
5. O₃ VMR profile
6. Continuum profiles for microwindows used in O₃ retrieval
7. HNO₃ VMR profile
8. Continuum profiles for microwindows used in HNO₃ retrieval
9. CH₄ VMR profile
10. Continuum profiles for microwindows used in CH₄ retrieval
11. N₂O VMR profile
12. Continuum profiles for microwindows used in N₂O retrieval
13. VMR profiles for other contaminants,

which can be used in the different retrievals either as a first guess of the profiles that are going to be retrieved or as assumed profiles of the atmospheric model (profiles of interfering species and p,T profiles in the case of VMR retrievals).

For each of these profiles both the a-priori estimate (given by either the IG2 profiles or, preferably, if available, ECMWF profiles extended with IG2 profiles in the altitude ranges not covered by ECMWF data) and the result of the most recent measurement (obtained either from the retrieval of the previous scans or from a previous retrieval of the same scan) are available.

Up to now the strategy has been to use whenever possible the most recent measurement (in practice the a-priori estimate was used only for the retrieval of the first sequence or when previous retrievals were unsuccessful).

On the light of the fact that in some cases the errors of the retrieved profiles may be very large (VMR retrieval at very low altitudes, continuum retrieval), we now propose to change the concept of ‘most recent measurement’ into the concept of ‘best estimate’. As it will be discussed later, the ‘best estimate’ coincides with the most recent measurement for the input files of type 1 (pressure and temperature profiles used as assumed profiles in VMR retrievals), with the a-priori profiles for the continuum profiles associated to the microwindows used in the different retrievals, and for all the others is equal to the optimal estimation between the most recently measured profiles and the a-priori ones.

The optimal estimation method consists in weighting the retrieved profile, with its VCM, with the pre-stored profile, which will be characterized by a VCM with big diagonal values.

The optimal estimations of the profiles have to be determined not only at the beginning of each scan analysis, but also after each VMR retrieval, because the retrieved VMR profile is used as assumed profile for subsequent retrievals.

As discussed in Sect. 5.2, the Optimal Estimation is not adopted for the determination of the results of Level 2 retrievals because it may introduce biases when averaging different results. Furthermore MIPAS limb scanning retrievals have redundancy of measurements and do not depend on the aid of a-priori information, therefore the choice of what a-priori information is added to the retrieved quantity can be left to the user. However, when selecting the assumed profile of interfering species and the initial guess profiles of the retrieved species, we are users of the results of previous retrievals and in this case we are authorized to apply optimal estimation without entering in contradiction with our stated principles.

Below the description of the algorithm which computes the Optimal Estimation between a-priori and retrieved profiles is provided.

Inputs:

- ⇒ Last retrieved profile x^r (profiles from retrievals up to n scans backwards can be used, where n is provided by the users)
- ⇒ simplified VCM of the retrieved profile V^r , whose diagonal elements and the first off-diagonal ones are set equal to the corresponding elements of the VCM of the retrieved profile, the other elements are set to 0.
- ⇒ Interval of time elapsed since the scan in which the profile x^r was retrieved.
- ⇒ First guess profile x^f of the previous retrieval (used for extending the VCM V^r for the values of the profile above the highest tangent altitude)
- ⇒ ECMWF profile x^e (if available)
- ⇒ climatological profile IG2 x^i , selected according to the actual latitude/longitude position of a scan.
- ⇒ VCM V^c of the a-priori profile (i.e. an appropriate merging of the ECMWF and IG2 profiles), characterized by equal elements along the diagonal and equal values (which can be different from the diagonal ones) for the first off-diagonal elements, the other elements being 0.

Outputs:

- ⇒ the best estimation of the profile to be used as first guess or as assumed profile of the retrieval

Procedure:

1. Definition of the a-priori profile x^c .

if x^e is not available, then:

$$x^c = x^i$$

else

$$x^c = x^e \quad (\text{in the altitude range covered by the ECMWF profiles})$$

$$x^c = x^i * \alpha \quad (\text{in the altitude range not covered by the ECMWF profiles})$$

end if

The grid of x^c is defined by the grid of x^e in the altitude range covered by the ECMWF profiles and by the grid of x^i in the altitude range not covered by the ECMWF profiles; the number of points of this new grid is named *igrid*.

In order to avoid discontinuities, the merging of the ECMWF and IG2 profiles is matched scaling the values of the IG2 profile in correspondence of the highest point of the ECMWF profile and the points above by a factor α equal to the ratio between the value of the highest point of the ECMWF

profile and the corresponding value of the IG2 profile (that may be obtained interpolating in the IG2 profile).

2. Correction of the VCM of the retrieved profile V^r due to latitude variation.

If the profile x^r has been retrieved in a previous scan (profiles from retrievals up to n scans backwards can be used), it means that it has been measured at a different latitude, and hence the estimated total retrieval error associated with this profile may underestimate its real error which is given by the total estimated retrieval error plus the error due to the latitudinal variation. In this case, the elements of the matrix V^r will be multiplied by a factor $\left(1 + \frac{t}{t_0}\right)$ proportional to the time interval elapsed since the measurement of the profile x^r (the constant t_0 is a user input).

3. Interpolation of the retrieved profile x^r on the grid of the a-priori profile x^c (the new profile is named x^{ri})

The interpolation rules are the same as used in the ORM code, i.e. logarithmic interpolation with pressure between the nearest values for temperature and VMR, linear interpolation with pressure for continuum; altitude is re-built using hydrostatic equilibrium starting from the lowest altitude of the a-priori profile x^c .

The interpolated profile x^{ri} is related to the retrieved profile x^r by the following equation:

$$x^{ri} = J x^r,$$

where J is the matrix (of dimensions $igrid \times ibase$) which transforms the retrieved profile (on a grid of $ibase$ points) to the interpolated one (on a grid of $igrid$ points): the element (i, k) of matrix J is given by:

$$J_{ik} = \frac{\partial x_i^{ri}}{\partial x_k^r}$$

4. Extension of the VCM of the retrieved profile above the highest retrieved tangent altitude.

V^r provides the variance and covariance of the retrieved profile only for the retrieved 0 tangent altitudes. The values of the retrieved profile above the highest retrieved tangent altitude are obtained by scaling the first-guess profile x^f with a factor equal to the ratio between the value of the retrieved profile at the highest tangent point and the corresponding value of the first-guess profile. Therefore the quadratic errors of these points are obtained by scaling the quadratic error at the highest tangent altitude (V_{11}^r) with the square of the same factor:

$$V_{ii}^{r-ext} = V_{11}^r \cdot \left(\frac{x_i^f}{x_n^f} \right)^2,$$

n is the index of the highest tangent altitude in the retrieved grid (dimension $ibase$) and $i < n$.

Correlations of the points of the profile above the highest tangent altitude, which, strictly speaking, should be set to 1, can be set to 0.5, in order to reduce a possible source of instability in the determination of the profile.

5. Determination of the VCM of the profile x^{ri} , V^{ri} .

Matrix V^{ri} is obtained performing the following matrix product:

$$V^{ri} = J V^{r-ext} J^T$$

6. Optimal estimation between profiles x^{ri} and x^c .

The optimal estimation (OE) of the profile to be used as first guess or as assumed profile of the retrieval is obtained as follows:

$$x_{OE} = (V^{c^{-1}} + V^{ri^{-1}})^{-1} (V^{c^{-1}} x^c + V^{ri^{-1}} x^{ri})$$

In the case of temperature and pressure profiles, OE method is used only for the determination of the first-guess profiles for p, T retrieval, while the assumed p, T profiles in VMR retrievals are the p, T values retrieved in the current scan.

The reason of the different treatment of p, T profiles according to the different use as first-guess or as assumed profiles is that the retrieval of minor constituents from MIPAS measurements is based on the simultaneous and reliable measurement of the temperature profile. For this reason it is not possible to rely on any a-priori information. Indeed, if pressure and temperature retrieval fails or if these profiles are measured with too big uncertainties, VMR retrievals will not be performed.

On the contrary, in the case of VMR retrievals, where effects of interference should be limited by the optimized choice of microwindows, if the previous VMR retrieval fails or returns a profile with big uncertainties, the subsequent retrievals will be performed anyway. In this case, the use of a-priori information for the definition of the VMR assumed profiles of the interfering key species allows to minimize the effects of interference. Therefore, after each VMR retrieval the OE between the retrieved profile and the a-priori profile of that VMR is computed again and the obtained profile is used in the subsequent retrievals of the actual scan.

Continuum profiles do not have to be treated with the optimal estimation method, because these profiles are strongly undetermined and their contribution to the spectrum is very low (even if the fit of these profiles ‘helps’ the retrieval procedure). Therefore, the a-priori continuum profiles, and not the retrieved continuum profiles, are used as first-guess for the subsequent scans.

5.14.1 Generation of initial guess continuum profiles

As explained in Sect. 5.14, the initial guess continuum profiles used by the retrieval algorithm are purely climatological profiles. Retrieved continuum profiles are not considered for the generation of the initial guess as they suffer of very large uncertainties (and also because retrieved continuum parameters, being used also to compensate for continuum-like error sources, may be very different from scan to scan).

Climatological continuum profiles are generated on the basis of a simple algorithm that, starting from given profiles of pressure, temperature and water vapor, calculates profiles of water vapor cross-sections in correspondence of the central frequencies of the considered microwindows. Water vapor continuum cross-sections are calculated using the CKD v.2.4 model (Clough, (1989)). The continuum of other species like CO₂, N₂O, aerosols etc. is not considered in this algorithm because we have found that in presence of large uncertainties on the continuum emission, it is by far preferred to use initial guess profiles that underestimate the real continuum. In fact, an overestimated initial guess continuum easily leads to a very opaque atmosphere in which the line of sight may not be able to penetrate. In these conditions the fitting procedure may be unable to recover the correct continuum emission.

5.15 Profiles regularization

In some cases the retrieved profiles oscillate more than what can be reasonably expected from the physical point of view. As outlined in the ESA study 12055/NL/CN, the oscillations are confirmed by the shape of the Modulation Transfer Functions (MTF) which show values greater than 1 (see e.g. the final report of the mentioned study). Tikhonov regularization (see e.g. Rodgers, (2000)) can be adopted in MIPAS retrievals with the objective of constraining the MTF to values equal or smaller than 1. Test calculations are however needed in order to tune the strength of the Tikhonov constraint (analysis of the MTF shape as a function of the strength of Tikhonov constraint). The implementation of this algorithm in the ORM required limited changes and it is useful to have this option accessible. The inversion formula used at each retrieval iteration for deriving the new parameters estimate, in the Levenberg-Marquardt (LM) approach is:

$$\hat{\mathbf{y}} = \mathbf{A}^{-1} \cdot \mathbf{K}^T (\mathbf{V}^n)^{-1} \mathbf{n} \quad (5.15.1)$$

with

$$\begin{aligned} \mathbf{A}_{i,j} &= \left(\mathbf{K}^T (\mathbf{V}^n)^{-1} \mathbf{K} \right)_{i,j} & \text{for } i \neq j \text{ and} \\ \mathbf{A}_{i,i} &= \left(\mathbf{K}^T (\mathbf{V}^n)^{-1} \mathbf{K} \right)_{i,i} \cdot (1 + \lambda) \end{aligned} \quad (5.15.2)$$

where λ is the LM damping factor and is adjusted during the iterations.

When using Tikhonov regularization the inversion formula (5.15.1) must be modified accordingly to:

$$\hat{\mathbf{y}} = \left(\mathbf{A} + \gamma \mathbf{L}^T \mathbf{L} \right)^{-1} \cdot \mathbf{K}^T (\mathbf{V}^n)^{-1} \mathbf{n} \quad (5.15.3)$$

where \mathbf{L} is a ‘regularization operator and γ is a parameter that can be used for tuning the strength of Tikhonov constraint.

$\mathbf{L}^T \mathbf{L}$ is a square matrix whose dimension is equal to the number of fitted parameters, each diagonal element of this matrix corresponding to a fitted parameter. In the ORM implementation matrix $\mathbf{L}^T \mathbf{L}$ consists of three non-zero square blocks referring to the three different groups of parameters that are being fitted: VMR, atmospheric continuum and instrumental offset in VMR retrievals. In the case of p,T retrieval matrix $\mathbf{L}^T \mathbf{L}$ consists of four non-zero square blocks which refer to p,T, atmospheric continuum and instrumental offset. In both cases, matrix $\mathbf{L}^T \mathbf{L}$ is a strip-diagonal matrix whose elements different from zero are only the diagonal elements and the first off-diagonal ones.

$$\mathbf{L}^T \mathbf{L} = \begin{pmatrix} 1 & -1 & 0 & \dots & \dots & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 \\ 0 & -1 & 2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -1 & 0 \\ \vdots & \dots & 0 & -1 & 2 & -1 \\ 0 & \dots & \dots & 0 & -1 & 1 \\ & & & 1 & -1 & 0 & \dots & \dots & 0 \\ & & & -1 & 2 & -1 & 0 & \dots & 0 \\ & & & 0 & -1 & 2 & \ddots & \ddots & \vdots \\ & & 0 & \vdots & \ddots & \ddots & \ddots & -1 & 0 \\ & & & \vdots & \dots & 0 & -1 & 2 & -1 \\ & & & 0 & \dots & \dots & 0 & -1 & 1 \\ & & & & & & 1 & -1 & 0 & \dots & \dots & 0 \\ & & & & & & -1 & 2 & -1 & 0 & \dots & 0 \\ & & & & & & 0 & -1 & 2 & \ddots & \ddots & \vdots \\ & & & & & & \vdots & \ddots & \ddots & \ddots & -1 & 0 \\ & & & & & & \vdots & \dots & 0 & -1 & 2 & -1 \\ & & & & & & 0 & \dots & \dots & 0 & -1 & 1 \end{pmatrix}$$

Both diagonal and first off-diagonal elements of each individual block can be scaled by user-defined input parameters.

The ORM first builds matrix $\mathbf{L}^T \mathbf{L}$ taking into account the actual number of fitted parameters and the input scaling factors, subsequently this matrix is normalized to the initial guess value of the unknown parameters $par(i)$ and tuned according to an input parameter η used for setting the overall strength of the constraint:

$$\gamma(\mathbf{L}^T \mathbf{L})_{i,j} \Rightarrow \frac{(\mathbf{L}^T \mathbf{L})_{i,j}}{\eta^2 par(i) \cdot par(j)}. \quad (5.15.4)$$

6 - Mathematical Optimisations

In this section we describe the different mathematical optimisations implemented in the optimised forward-retrieval model (OFM/ORM): in particular, the advantages and disadvantages of the different options are reviewed and the preferred option is identified.

The parameters defining the different optimisations, like parameters dealing with layering, have been determined on the basis of tests performed with the RFM.

These are the guidelines used in the search for possible mathematical optimisations:

- Research of feasibility of performing analytically integrals and derivatives.
- When this is not possible, in case of numerical integrals, minimisation of the number of the intervals over which complex expression are evaluated.
- In case of very time consuming calculations, study of feasibility of using pre-tabulated data and interpolation schemes.
- Research and exploitation of the symmetries that can reduce the number of calculations.
- Study of the possibility of storing quantities that are used more than once.

When it is possible, for each of the implemented optimisations we identify:

- physical problem
- different options for modelling this physical problem
- selected choice
- reason of choice
- how optimisation is implemented into the program
- performed validations
- validations to be performed

The optimisations concern mainly the forward model and its interfaces with retrieval model, since calculation of the synthetic spectra and of the Jacobian matrix used in the Marquardt algorithm are the time consuming parts of the retrieval program.

6.1 Radiative Transfer integral and use of Curtis-Godson mean values

Forward model consists essentially of the calculation of Radiative Transfer integral (eq. (4.4.5)), a curvilinear integral along the line of sight.

Optimising the forward model means to optimise the calculation of this integral.

An analytical expression of absorption cross-sections as a function of pressure and temperature is not available, so the integral can be solved only by using a discretisation, i.e. calculating the single contribution of shorter paths and then summing them.

This implies that atmosphere has to be segmented.

Two main optimisations can be implemented:

- the first refers to the kind of segmentation: since spectra corresponding to different lines of sight have to be calculated, a segmentation that can be used for all geometries avoids to repeat several times the same calculations;
- the second refers to the length of the segments and consequently to the number of segments to be considered: the coarser is the segmentation and the faster is the forward model.

In this section we describe two different methods for building the segmentation and then different possibilities to calculate the contribution of single segments to the whole integral, highlighting the chosen options.

Integration variable

If x , the co-ordinate along the line of sight, is used as integration variable, the optical path is subdivided in a number of intervals of equal length, so that the altitude thickness of the layers in which atmosphere turns out to be stratified decreases when approaching the tangent altitude: in this way the atmosphere is sampled with greater detail near the tangent altitude, from where most of the signal originates. The disadvantage of this method is that layering changes for each viewing geometry, so that results from one geometry cannot be used for the others.

On the contrary, if the altitude z , is used as integration variable, the atmosphere can be divided in a pre-defined number of layers, independent on the geometry. This layering can be critical near the tangent altitude, if layering is not sufficiently fine, but it has an important advantage when several different limb geometries have to be simulated at the same time: in this case layering does not change from one geometry to the other.

Since a whole limb-scanning sequence is analysed at a time, the atmosphere is subdivided in a set of fixed layers, whose thickness is defined by criteria that will be discussed in section 6.1.1.

Layers versus levels in the discretisation of the unknowns

The unknown profiles (temperature and VMR) must be transformed into discrete values corresponding to a finite set of altitudes. The altitude distribution that corresponds to these discrete values can be obtained either with the layer approach (the unknown is constant within contiguous layers of finite thickness) or with the level approach (the unknown is linearly interpolated in between the altitudes where the discrete altitudes are defined). The level approach has been chosen as baseline for OFM /ORM algorithms.

Calculation of the transmission of each layer

In order to perform the radiative transfer integral in the single layer, different options have been analysed. The first option consists in using interpolated cross-sections from the values computed at the boundaries of the layer: this method requires layers not thicker than 1km, because of the critical dependence of cross-sections on temperature and VMR. The possibility of performing the analytical calculation of the integral, that could be possible with this method and should reduce the number of computations, fails when refractive index has to be included correctly in the calculation, because the expression becomes, in this case, very complicate (see section 5.5).

Another option consists in the calculation of layer transmission by computing the cross-sections at some values of pressure and temperature representative of the layer.

These quantities can be either the mean pressure and temperature of the layer, or the Curtis-Godson equivalent values, calculated for each gas, by performing an integral along the ray-path of temperature and pressure, taking into account the variation of pressure and temperature along the ray-path inside the layer.

Equivalent pressure and temperature are given respectively by:

$$p_{m,l,g}^e = \frac{\int_{z_{l-1}}^{z_l} p(z) \cdot X_m(z) \cdot \eta(p(z), T(z)) \cdot \frac{dx^g}{dz} \cdot dz}{\int_{z_{l-1}}^{z_l} X_m(z) \cdot \eta(p(z), T(z)) \cdot \frac{dx^g}{dz} \cdot dz}, \quad (6.1.1)$$

$$T_{m,l,g}^e = \frac{\int_{z_{l-1}}^{z_l} T(z) \cdot X_m(z) \cdot \eta(p(z), T(z)) \cdot \frac{dx^g}{dz} \cdot dz}{\int_{z_{l-1}}^{z_l} X_m(z) \cdot \eta(p(z), T(z)) \cdot \frac{dx^g}{dz} \cdot dz} : \quad (6.1.2)$$

z is the altitude, z_l and z_{l-1} are the heights on the boundaries of the layer, $X_m(z)$ represents the Volume Mixing Ratio of the m -th gas, x^g is the line of sight dependent on geometry, $\eta(p(z), T(z))$ is the air number density.

The normalisation factor of these expressions represents the column of the considered gas, layer and geometry, $C_{m,l,g}$:

$$C_{m,l,g} = \int_{z_{l-1}}^{z_l} X_m(z) \cdot \eta(p(z), T(z)) \cdot \frac{dx^g}{dz} \cdot dz. \quad (6.1.3)$$

Integrals (6.1.1), (6.1.2), (6.1.3) are solved taking into account refraction after an appropriate change of variable (see section 5.5).

We have verified that the use of Curtis-Godson pressure and temperature, instead of mean values, allows the use of a coarser stratification of the atmosphere.

A complication of this method is that, in principle, the equivalent pressure and temperature are personalised for each gas, each layer and each geometry.

In particular, their dependence on the gas, that is useful when performing the analytical derivatives with respect to VMR, has as consequence that cross-sections for all the gases have to be stored (see Sect. 6.2.2), and a big amount of memory is required.

On the contrary, the use of mean temperature and pressure, that do not depend neither on the molecules nor on the geometry, requires a finer layering and consequently more computing time.

We have to underline that the calculation of equivalent pressure and temperature, as well as the columns, is not a time consuming part of forward model calculation, and no optimisation effort is worthwhile.

Using equivalent pressure and temperature, the transmission τ of layer l for the geometry g due to all the gases is given by:

$$\begin{aligned}
 \tau_{\sigma,l,g} &= \exp \left(- \sum_m \int_{z_{l-1}}^{z_l} k_{\sigma}(p(z), T(z)) \cdot X_m(z) \cdot \eta(p(z), T(z)) \cdot \frac{dx^g}{dz} \cdot dz \right) \approx \\
 &\approx \exp \left(- \sum_m k_{\sigma,l,g}(p_{l,g}^e, T_{l,g}^e) \cdot \int_{z_{l-1}}^{z_l} X_m(z) \cdot \eta(p(z), T(z)) \cdot \frac{dx^g}{dz} \cdot dz \right) = \\
 &= \exp \left(- \sum_m k_{\sigma,l,g}(p_{l,g}^e, T_{l,g}^e) \cdot C_{m,l,g} \right)
 \end{aligned} \tag{6.1.4}$$

Using this result, expression (4.4.5) can be written as:

$$S_{\sigma,g} = \sum_{l=1}^N B_{\sigma} \left(T_{l,g}^{e \text{ main}}(z) \right) \cdot (1 - \tau_{\sigma,l,g}) \cdot \prod_{j=1}^{l-1} \tau_{\sigma,j,g} \tag{6.1.5}$$

N represents the number of paths obtained from the intersection of the line of sight with the levels used for atmospheric layering, that are equal to twice the number of layers minus one.

The source function B_{σ} , that has to be calculated for each frequency of the considered microwindow and each layer, is independent on the molecule, while the equivalent temperature of the layer is molecule dependent.

Since we can assume that each selected microwindow is characterised mainly by one gas, while the others molecules have very small influences on the total spectrum, we calculate the source function at the Curtis-Godson temperature of the main (retrieved) molecule of the considered microwindow.

6.1.1 Layering of the atmosphere

On the basis of the above choices, the atmosphere is modelled using layers whose boundaries are marked by levels at fixed pressure. Within the levels the temperature and the VMR profiles are assumed to vary linearly with the altitude, while the behaviour of pressure profile is assumed exponential with altitude. Pressure and temperature profiles obey to the hydrostatic equilibrium law. Either altitude or pressure can be considered the independent variable at this stage, provided that we use the correct interpolation rules for dependent variables.

While setting-up the layering of the atmosphere the chosen independent variable is the altitude; this is because the visual inspection of the generated levels is easier.

The algorithm which builds the levels proceeds as follows:

- Step 1:

A set of levels corresponding to the tangent altitudes of the spectra we want to simulate is set-up; radiative transfer calculation is indeed simpler if the tangent altitudes are at the boundary of one layer. Since we want to take into account the FOV effect using interpolation of the spectra in the altitude domain (see Sect. 6.6), not only the spectra whose tangent altitudes correspond to measurements have to be simulated, but ‘extra’ spectra are needed as well. In the case of p,T retrieval, the simulated spectra are the ones corresponding to the measurements, plus one extra spectrum located below the lowest measurement and one extra spectrum located above the highest measurement. The distance between the tangent altitude of each extra spectrum and the tangent altitude of the nearby measurement is kept equal to half of the FOV width.

Tests have shown that the explained set of simulations does not allow to perform an accurate interpolation of the spectra when the VMR of the main gas of the retrieval has a large gradient as at low altitudes in the case of H_2O , (see Sect. 6.6). In these cases further intermediate simulations are needed.

- Step 2:

Each couple of adjacent levels generated at step 1 is considered. We check whether, moving from one level to the other, the following two conditions are satisfied:

1. the variation of the temperature is below a fixed threshold. Two different thresholds are used depending on the altitude of the first considered level, a more conservative threshold is used at low altitudes.
2. The variation of the Voigt half-width of a reference line is below a fixed threshold.

If both these conditions are satisfied then we consider the next couple of levels generated at step 1 and redo checks 1 and 2.

If one or both the above conditions are not satisfied, then we insert new evenly-spaced levels within the couple of considered levels, until conditions 1 and 2 are satisfied for all the new sub-levels.

After this step is completed, it turns out that the altitude range in which the tangent altitudes of the simulated geometries lie, is sub-divided into layers of suitable thickness, whose boundaries are marked by the levels.

- Step 3:

Above the tangent altitude of the highest simulated spectrum a set of levels is determined which divide the atmosphere into layers for the radiative transfer calculation.

Starting from the tangent altitude of the highest simulated spectrum,

^(*) a user-defined guess increment Δz is used to build next level, then, conditions 1. and 2. (used at the previous step) are checked and:

⇒ if the two conditions are both satisfied then the guess level is accepted and the algorithm proceeds to ^(*)

⇒ else, the guess increment Δz is reduced using an appropriate factor and the conditions 1. and 2. are checked again.

It is clear that after this procedure the maximum allowed thickness of the layers is equal to the initial value of Δz that is controlled by the user.

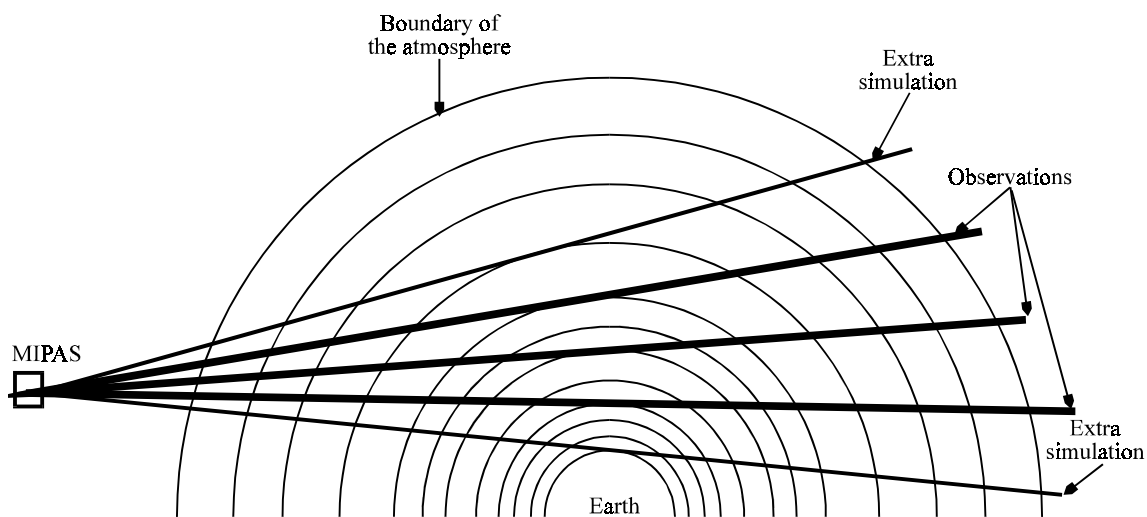
The new levels are added to those obtained in step 2.

The user-defined parameters that control the layering of the atmosphere are subject of tuning: because of the speed requirements, in operational conditions, the parameters that allow a more coarse layering without significantly affecting the accuracy of the computed spectra have to be adopted. Tests have been carried-out using some microwindows involved in p,T retrieval. The results are shown in Table 6.1; the spectra computed using layerings 2 and 3 have been compared with spectra obtained using a really conservative layering (reference layering 1). It turns out that layering 2 represents a suitable compromise between accuracy and number of levels.

	Layering 1 (reference)	Layering 2	Layering 3
Low altitude T threshold (K)	1.5 K	5 K	25
Hig altitude T threshold (K)	5 K	15 K	35
Altitude where the threshold is changed (km)	56	56	56
Max. HW-variation	1.05	1.5	2.5
Max. thickness of the layers (km)	10	10	10
N. of obtained levels	146	42	22
Max. difference		NESR / 5	NESR/0.6
Average difference		NESR / 60	NESR/17

Table 6.1: Tuning of the parameters used for the layering of the atmosphere. The tests have been performed considering 7 microwindows of p,T retrieval. The upper limit of the atmosphere has been set equal to 100 km.

Figure 6.1: Sketch of the layering of the atmosphere.



6.2 Secant law approximation for the calculation of Curtis-Godson quantities and definition of paths

Secant law approximation consists in the calculation of Curtis-Godson quantities as if the layer was flat and the line of sight a straight line. In this case the secant law applies and the same values of p^e and T^e are obtained independently of the angle θ between the line of sight and the vertical direction. In Fig. 6.2 the percent deviation of p^e and T^e from the values calculated in the case of vertical penetration are reported.

These tests were done using ozone VMR profile and standard atmosphere, with layers 3 km and 1 km thick at tangent altitude of 8 km. No significant changes occur at different tangent altitudes.

Secant law approximation causes only very small errors at all altitudes, except at the tangent layer and to the layer above.

Exploiting this result, it is sufficient to calculate the values of p^e and T^e for all the layers of the lowest geometry, and only for the lowest layers for all the other geometries.

If we associate a *path* with each combination of layer and geometry of the complete limb-scanning sequence, we can say that the values of p^e and T^e do not have to be calculated for all the different paths. In Fig. 6.3 the table of the paths is reported: in horizontal position the levels used for the simulations are represented, in vertical position all the different geometries that have to be simulated are shown.

The grey boxes represent all the possible paths.

The grey boxes with either 'X' or 'x' are the paths for which a customised calculation of equivalent pressure and temperature have to be calculated. In the following these particular $p^e - T^e$ couples will be called 'Implemented Atmospheric Pressures and Temperatures', IAPTs.

In the grey box without either 'X' or 'x', values of the top box can be used.

The number of extra-paths to be calculated for each limb view is an input parameter of the retrieval program, but the current baseline is to re-compute only the IAPTs relative to the tangent layer. Tests have shown that this is a good approximation, since the tangent layer is significantly thinner than 3 km.

The use of IAPTs is a crucial optimisation, not only because less equivalent pressures and temperatures have to be calculated (calculation of Curtis-Godson quantities is not very time consuming), but mainly because less cross-sections have to be calculated (see Table 6.2) and stored. The saving of number of calculations is significant: without using secant law approximation, the number of intervals where cross-sections have to be calculated, that is the number of total paths, is given by one half the product of the number of the layers (about 40) in each geometry times the number of geometries used for the simulations (18), that is 360 paths; on the contrary, the number of the IAPTs is given by the number of paths for the lowest geometry (about 40) plus the number of extra-paths (about 2) times the number of remaining geometries (17), that is about 74.

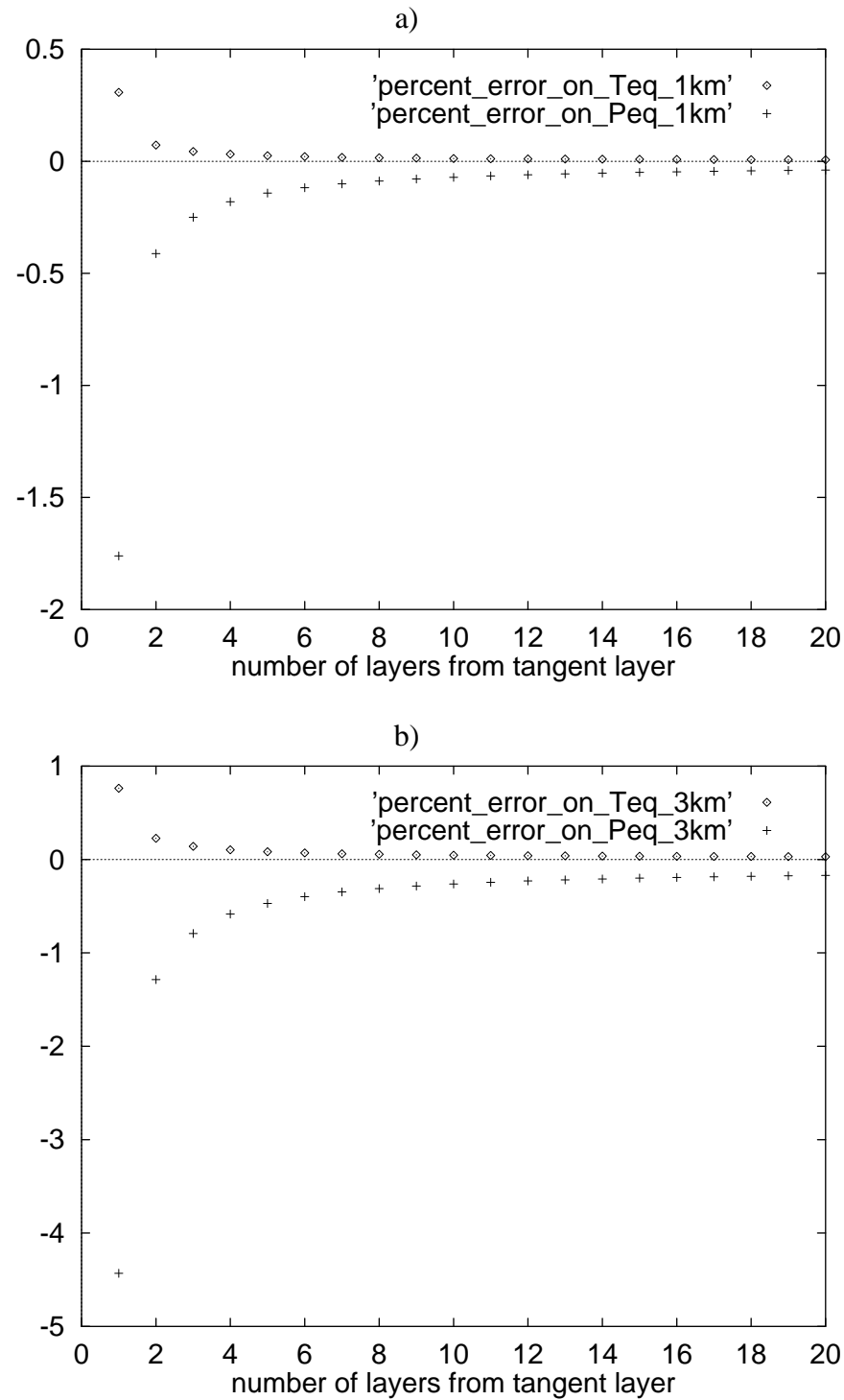


Fig. 6.2 The percentage deviation of equivalent temperature (squares) and equivalent pressure (crosses) from that in the case of vertical penetration are reported for the different layers, starting from the tangent layer. a) Layers thickness is about 1 km; b) Layers thickness is about 3 km. Test performed for a tangent altitude of 8 km.

Fig.6.3_TABLE of PATHS

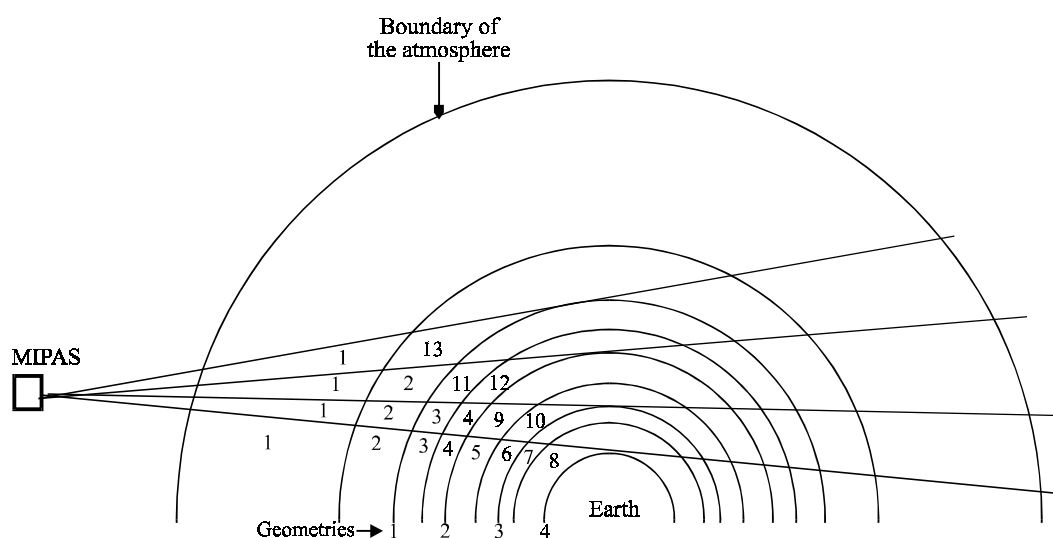
[illegible]

6.2.2 Sequence of the operations

The fact that only a limited number of layers for each geometry needs a customised calculation of equivalent pressure and temperature is the basis of the structure of the optimised forward model.

After the setting-up of the layering of the atmosphere (see section 6.1) the matrix of the IAPT numbers is built. This matrix associates with each path a number that refers to the corresponding IAPT.

Fig. 6.4 Example of the association of the appropriate IAPT to each path.



We start from the line of sight to which the lowest tangent altitude corresponds and a progressive IAPT number is associated to each layer.

For all the other geometries a new IAPT number is assigned to the tangent layer and in some cases to others layers above it, all the other layers have the same IAPT number as the lowest geometry.

The matrix below refers to the example of figure 6.4:

$$\begin{pmatrix} 13 \\ 1 & 2 & 11 & 12 \\ 1 & 2 & 3 & 4 & 9 & 10 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{pmatrix}$$

At this point the calculation of ray-tracing is performed and all the IAPTs are computed, while the columns are calculated for all the paths.

For each of the selected microwindows the computation of the cross-sections is performed for the different IAPTs (some possible optimisations are discussed in section 6.3, 6.8, 6.9, 6.10).

The spectra of all the limb-scanning sequence are calculated summing the contribution of all the layers (see Sect. 6.4).

After the convolution of the spectrum with the apodised instrument line-shape (AILS), convolution with the function that describes FOV is performed (see Sect. 6.6).

6.3 Interpolation of cross sections for different geometries

We have explained in the previous section that the equivalent temperature and pressure of each layer remain nearly constant for the different geometries of a limb scan. As was shown in Figure 6.2 the differences become larger the closer we are to the tangent layer and are the largest for the tangent layer itself. The absorption cross sections don't have to be calculated for all paths, but only for the IAPTs (see Sect. 6.2).

After calculating the absorption cross sections for all the IAPTs of the lowest geometry of the limb scan, in order to calculate the cross sections for the different IAPTs of the other geometries we have to distinguish between two kinds of IAPTs, either the IAPTs corresponding to the paths indicated with 'X' in Figure 6.3, or the IAPTs corresponding to the paths indicated with 'x'.

1. IAPTs corresponding to paths with 'x': the cross sections can be interpolated (in pressure) between the cross sections of the lowest geometry.
2. IAPTs corresponding to paths with 'X' in geometries different from the lowest one: new calculation of the absorption cross sections for the equivalent temperature and pressure of the new path is done.

We performed calculations in order to test the feasibility of case 1. Since the Lorentz line wings are in first approximation proportional to the pressure we used this parameter for the interpolation value (i.e. we interpolated the cross sections of the lowest layer to the equivalent pressure of the new path). We decided to use linear interpolation. (Tests of higher order interpolation gave often better results but failed in those cases where there was an inversion of the absorption cross section profile with altitude.)

In these tests the maximum differences of the cross sections between recalculation and linear interpolation was 2% for the tangent layer, 0.3% for the layer above the tangent layer and 0.1% for the second layer above the tangent layer (layer thickness 3 km). The reason for this decreasing errors results obviously from the fact that the secant law approximation becomes more and more valid when moving away from the tangent layer.

Table 6.2 shows the results of test calculations which were performed using 6 microwindows for p-T retrieval. It is obvious that no recalculations of the cross sections for the tangent layer (1st column) or interpolation for the tangent layer (2nd column) results in maximum errors larger than NESR/5. Recalculation of the tangent layer (3rd column) leads to acceptable maximum errors of NESR/21 and recalculation of the tangent and interpolation of the layer above the tangent layer to NESR/70. As a baseline we will recalculate the absorption cross sections only for the tangent layer. Since the code is structured in order to be very flexible in handling these three different cases higher accuracy can be obtained only by changing one input parameter.

no. of recalculations 'X'	0	0	1	1	2
no. of interpolations 'x'	0	1	0	1	0
max. difference	NESR/0.9	NESR/4.4	NESR/21	NESR/70	NESR/70
average difference	NESR/3.7	NESR/20	NESR/79	NESR/ 114	NESR/ 330

Table 6.2: Maximum and average difference measured in NESR between a reference simulation and simulations made with different methods of cross-sections calculation for the geometries above the lowest one for 6 selected p-T microwindows.

6.4 Calculation of spectrum: exploitation of spherical symmetries

The use of the altitude as the integration variable and the layering of the atmosphere that results from this choice, together with the hypothesis of homogeneity of the atmosphere with latitude, allows to exploit some symmetries and reduce the number of computations.

In fact, the line of sight crosses each layer twice, in a symmetrical position with respect to tangent layer. The symmetry derives from the fact that the atmospheric layers are spherical, and dependence on latitude is neglected (section 5.3).

The two contributions of the same layer to the total intensity reaching the observer are characterised by the same emission, but different transmissions.

Since the cross-sections for all the layers have been previously calculated, while the first contribution is calculated, also the second one is taken in account.

So, instead of calculating the integral for all the altitude intervals in which the line of sight intersects the different layers, that are twice the number of layers, it is possible to calculate the integral only for all the layers.

Expression (6.1.5) is modified into the following expression:

$$S_{\sigma,g} = \sum_{l=1}^L B_{\sigma}(T_l^e) \cdot (1 - \tau_{\sigma,l,g}) \cdot \left(1 + \tau_{\sigma,l,g} \cdot \prod_{j=l+1}^L \tau_{\sigma,j,g}^2 \right) \cdot \prod_{j=1}^{l-1} \tau_{\sigma,j,g}, \quad (6.4.1)$$

with $\prod_j^{j-1} \tau_{\sigma,j,g} \equiv 1$.

L is the total number of the layers (in order to maintain the symmetry, the tangent layer is also divided into two parts, symmetrical with respect to tangent point).

6.5 Use of interpolation for the calculation of Planck function

In order to save computing time for the calculation of the Planck function $B(\sigma,T)$, a linear interpolation is used between the values assumed by the function at the edges of the microwindow.

The expression for $B(\sigma, T)$ was given in eq. (4.4.2).

The plot of the absolute differences between the theoretical value of B and its linearly interpolated value is shown in Fig. 6.5. It has been calculated, every 100 cm^{-1} in the centre of the microwindow, for a microwindow width of 2 cm^{-1} , from 600 cm^{-1} to 2000 cm^{-1} . On the plot a scale factor of 10^{12} has been applied. The extreme values of the temperature 180K and 270K are considered; we can see that the absolute error decreases when T decreases and when the frequency of the microwindow increases. This effect can be explained by considering that, for $T=250\text{K}$, the Planck function has its peak value around 500 cm^{-1} , so that the linear approximation becomes better when the derivative of the B function tends to become constant, i.e. at high frequencies (far from the peak).

While the precision of the computation is not affected by this approximation, there is a saving of the CPU time due to the elimination of the computation of some thousands of exponential functions (one for each point in fine grid) for each microwindow.

A run time test on the simulation of a 0.25 cm^{-1} wide microwindow containing 74 transitions led to a time saving of about 10%.

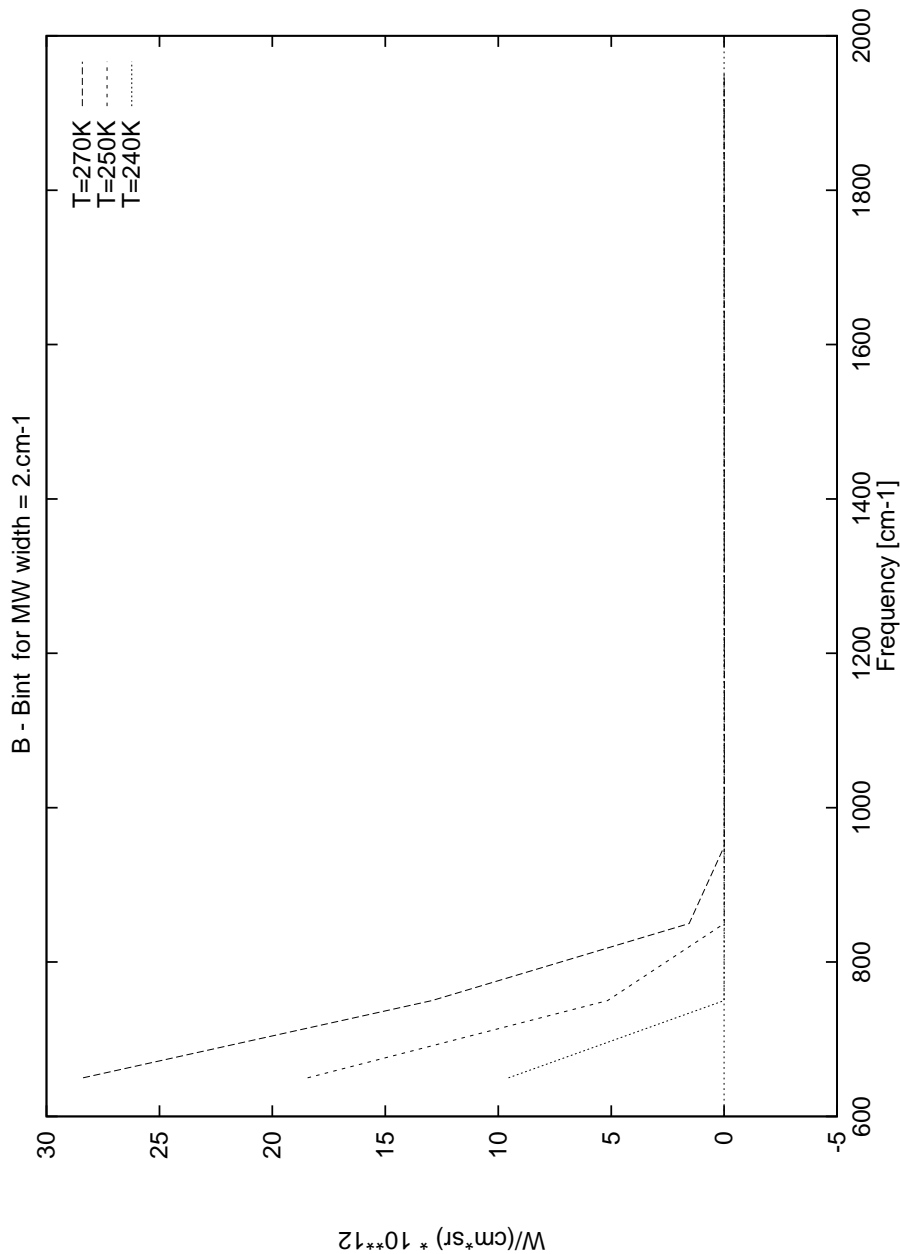


Fig. 6.5: Absolute difference between the theoretical value of Planck function and its linearly interpolated value. A scale factor of 10^{12} is applied. These differences are completely negligible if compared with the NESR, whose values range from $50 \text{ nW}/(\text{cm}^2 \cdot \text{sr} \cdot \text{cm}^{-1})$ in the wave-number region around 800 cm^{-1} to $4.2 \text{ nW}/(\text{cm}^2 \cdot \text{sr} \cdot \text{cm}^{-1})$ around 2000 cm^{-1} .

6.6 Finite instrument field of view.

The problem of finite field of view is a consequence of the fact that:

- the input diaphragm of the interferometer has non-zero angular size;
- light from an extended source crosses it;
- this source is characterised by a vertical exponential energy distribution.

These factors have two main effects: a modification in the ILS and a modification in the ‘effective’ tangent altitude of the spectrum.

The change of the ILS in the specific case of MIPAS rectangular aperture, with a vertical exponential energy distribution across it, have been analysed by L. Debouille and G. Roland (1995): they found that the use of a rectangular aperture creates a small asymmetry in the ILS, but even the strongest exponential energy distribution across the field of view does not significantly affect the ILS, calculated in the case of homogenous source.

The other effect is that the exponential distribution introduces a non-negligible difference between the geometrical tangent altitude, defined by the centre of the input diaphragm, and the ‘spectral tangent height’, that is the position, along the vertical scale, of the mean emitted signal.

Delbouille and Roland (1995) found that corrections (dependent on the molecule) up to nearly 1 km have to be applied to tangent altitude when the rate of change of the emission is of the order of a factor three per kilometre.

We have verified that, at least at low tangent altitudes, neglecting field of view effect brings an error in the spectrum of H₂O and CO₂ bigger than NESR.

The antenna pattern of the field of view provided by ESA was initially represented by a spread in the altitude domain $FOV(z)$ constant as a function of altitude and of shape equal to a trapezium with the greater base of about 4 km and the smaller base of about 3 km. Recent measurements on the MIPAS flight module indicated however that the MIPAS FOV can be more adequately represented using a piecewise linear shape. A tabulated piecewise distribution is in fact the FOV representation presently adopted in the ORM/OFM.

According to these arguments, the effect of field of view can be taken into account in two different ways:

- by using an equivalent observation geometry,
- by performing, for each spectral frequency, the convolution between the spectrum and the antenna pattern (Sect. 4.4.3):

$$S^F(\sigma, z) = S(\sigma, p(z)) * FOV(z). \quad (6.6.1)$$

Since the equivalent observation geometry is strongly dependent on the molecule, the second option has been chosen.

The standard method used in this case is to perform a numerical convolution with the FOV function after repeating forward model calculation for a number of lines of sight that span the a user-defined vertical range around tangent altitude.

In order to reduce the number of computations, the following optimisations have been implemented:

- convolution of the high resolution spectra with the apodised instrument line shape, before taking into account FOV effects, in order to operate FOV convolution in the coarse frequency grid, instead of the fine grid.
- interpolation of the spectra calculated at the tangent pressures to determine the dependence of the spectra as a function of altitude; the result is used to perform an analytical convolution.

This interpolation is critical: it doesn’t seem reasonable to use a high order interpolation extended to remote tangent altitudes, because the spectrum corresponding to a particular layer depends on the value of temperature and VMR profiles at that layer, and these are not necessarily related with those at layers above and below.

For this reason, an improvement in the approximation can not be obtained increasing the order of interpolation by including spectra at remote tangent altitudes, but must be obtained increasing the number of simulated spectra used for the interpolation between two contiguous tangent altitudes.

Some tests have been performed for determining the minimum number of spectra necessary for performing a correct interpolation.

The critical aspect is given by discontinuities in the rate of change of temperature and of molecule density with the altitude.

The most critical molecules are H₂O and CO₂. H₂O has a strong rate of change in VMR profile, and hence in density profile, in the troposphere, but also CO₂, which is characterised by a constant VMR, is strongly affected by change in temperature near the tropopause.

At higher altitudes the profiles don't show significant change in the VMR slope, with the only possible exception of O₃ and HNO₃, and FOV effect are expected to be less significant.

Tests on CO₂ were performed by comparing the analytical convolution, made using interpolated spectra between three spectra at three contiguous tangent pressures, with a numerical convolution between spectra corresponding to tangent altitudes distant 200 metres and the FOV function.

The results on tests on CO₂ are reported in table 6.3, where the error in tangent altitude is written for different microwindows.

MW ↓	TA→	8 km	11 km	14 km	17 km
12PT37		33	70		
13PT38		20	85	15	
14PT41			85	5	20
15PT44		20	85	5	15

Table 6.3 Results of comparison between reference numerical convolution of field of view and analytical convolution using interpolation with 3 contiguous spectra, for some of the microwindows selected for p-T retrieval. The error in altitude are expressed in metres.

These errors are acceptable, according to the acceptance criteria reported in section 3.

Therefore, for CO₂ and, consequently, for all the other molecules, except water, interpolation can be built from spectra calculated at three contiguous tangent pressures.

In this case, the interpolated spectrum is represented by:

$$S^I(\sigma, z(p)) = cof_1(\sigma) + cof_2(\sigma) \cdot z + cof_3(\sigma) \cdot z^2, \quad (6.6.2)$$

cof_1 , cof_2 , cof_3 are the coefficients of the interpolation calculated, for each frequency, from the values of spectra at the considered tangent pressures.

The spectrum with FOV is given by:

$$S^F(\sigma, z(p)) = \int S^I(\sigma, z(p)) \cdot FOV(z) \cdot dz, \quad (6.6.3)$$

The integral can be easily calculated analytically.

Tests on H₂O have shown that at low altitudes, up to the boundary between troposphere and stratosphere, the interpolation with three spectra at three contiguous tangent pressures produces discrepancies between the analytical and numerical ('exact') convolution.

These discrepancies can be reduced calculating an additional spectrum at a tangent altitude intermediate between two contiguous tangent altitudes, and hence drawing a quartic order polynomial through five spectra (see Table 6.4).

Because of these results, the retrieval program has been made flexible for the computation of additional spectra in some specific cases. This does not represent a big increase in computing time, because additional spectra have to be calculated only in the troposphere, and only for H₂O.

μW_s ↓ TA →	8 km	11 km	14 km
1H ₂ O2B	26	17	21
2H ₂ O3	24	74	21
3H ₂ O4B	20	42	24
4H ₂ O5B	30	17	24
5H ₂ O6B			47
6H ₂ O33	23	116	22

Table 6.4 Results of comparison between reference numerical convolution of field of view and analytical convolution using 5 spectra with tangent heights 1.5 km distant, for some microwindows selected for retrieval of H₂O VMR. The equivalent error in tangent height is expressed in metres.

It has to be underlined that, using this approach, the error due to the interpolation is very small when the mean tangent altitude of S^f coincides with that of one of the simulated spectra. The error increases when an offset is introduced.

The final validation of the model for taking into account FOV has been done using RFM spectra. Work is still in progress to improve the model in the cases where significant discrepancies with RFM spectra have been found.

In Fig. 6.6 the values of the reference spectrum with FOV at a significant frequency at different altitudes is plotted as a function of the corresponding spectrum obtained by analytical convolution. The deviation of the curve from a straight line indicates the presence of a variable error. This variation as a function of the tangent altitude offset indicates the presence of a potential error in the computation of the analytical derivatives with respect to tangent pressure (see below).

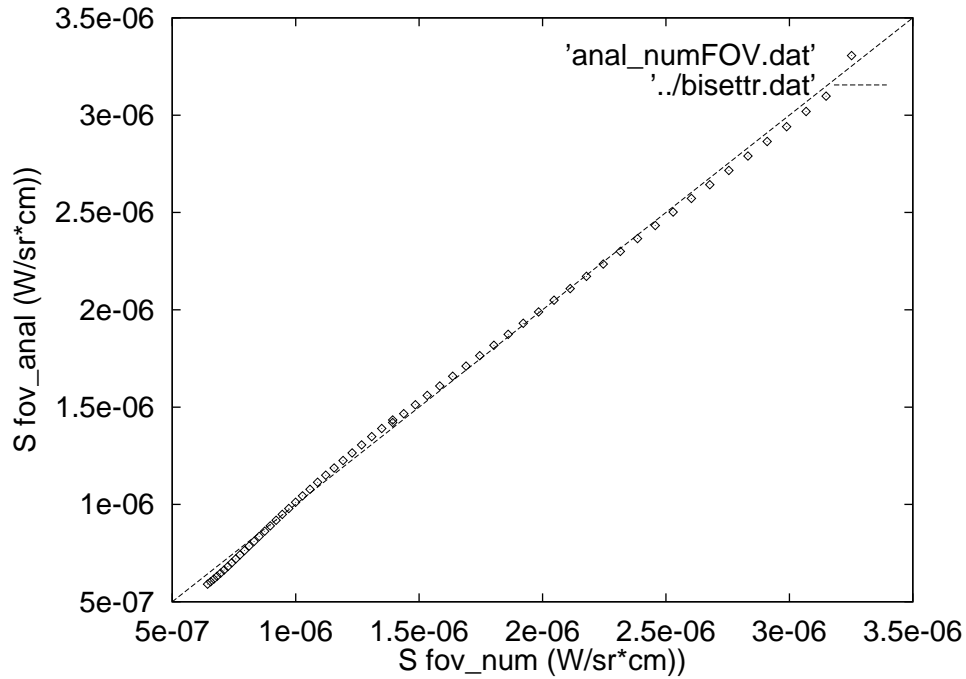


Fig. 6.6 In this plot the values of the spectrum at a significant frequency calculated with a analytical convolution at altitudes between 6.5 km and 12.5 km are plotted as a function of the corresponding values of the spectrum calculated with reference numerical convolution. A microwindow selected for p-T retrieval has been used.

The analytical derivative with respect to tangent pressure, obtained by the following calculation:

$$\frac{dS^F}{dp^{\tan g}} = \frac{dS^F}{dz} \cdot \frac{dz}{dp^{\tan g}}, \quad (6.6.4)$$

(the relation between z and p is derived by Hydrostatic equilibrium equation) has been compared with the numerical derivative, calculated using two spectra that take into account field of view and are characterised by a difference in tangent altitude of 100 m.

A percentage difference of the order of 10-15 % is obtained (see Fig. 6.7).

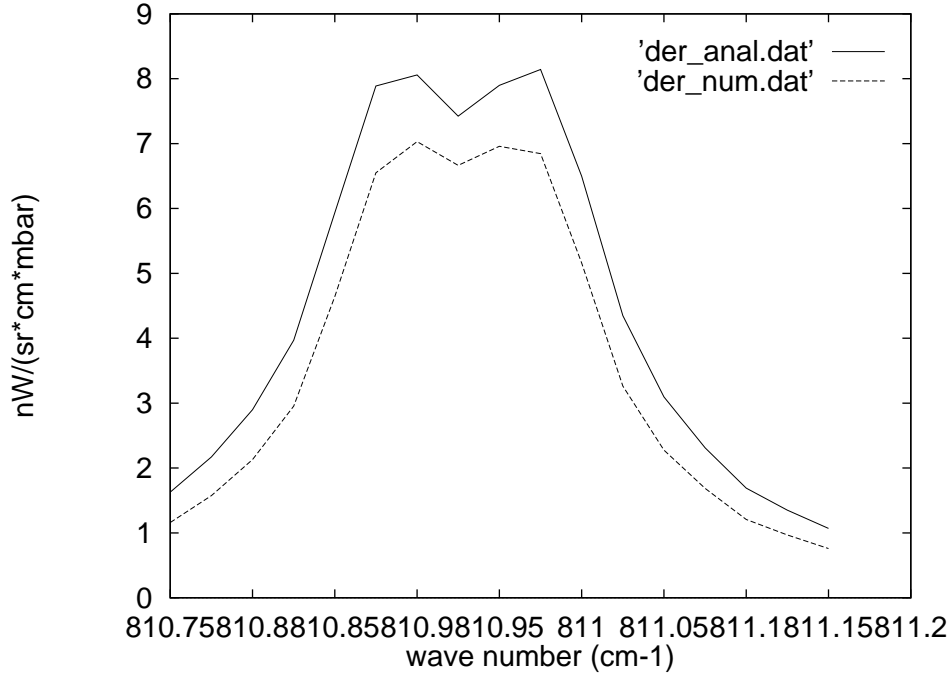


Fig. 6.7 Numerical and analytical derivatives with respect to tangent pressure for a microwindow selected for p-T retrieval, tangent altitude equal to 11 km.

Tests performed so far have shown that this error on tangent pressure derivatives does not increase the number of iterations required for reaching the convergence; however, in case this error should turn-out to be too large when dealing with real data, further efforts must be made for the improvement of the interpolation of the spectra.

6.7 Analytical derivatives

6.7.1 General considerations

In contrast to numeric derivatives for which many reruns of the forward model are necessary, analytical derivatives can be calculated from parameters determined during the forward model calculation. Obviously the use of analytical derivatives makes only sense if the time consumption of their calculation is considerably smaller than the recalculation of the forward model. As a rule of thumb this is the case if the determination of the analytical derivatives avoids the recalculation of the absorption cross sections and if they are sufficiently precise that no extra iteration steps are necessary.

The basic equation of the derivative of the spectrum S with respect to an unknown variable q_r^{ret} (temperature, pressure or volume mixing ratio) on the levels to be retrieved is (for clearness of the equations we omit here the dependence of S on the wavenumber and the tangent altitude and consider only one absorber species, i.e. omit index m on the gases):

$$\frac{dS}{dq_r^{ret}} = \frac{d}{dq_r^{ret}} \sum_{l=1}^N B_l(T_l^e) \left[e^{-\sum_{j=1}^{l-1} k_j(T_j^e, p_j^e) C_j} - e^{-\sum_{j=1}^l k_j(T_j^e, p_j^e) C_j} \right] \quad (6.7.1)$$

where N = total number of optical paths used to calculate the radiative transfer

T_l^e = equivalent temperature of the layers
 p_l^e = equivalent pressure of the layers
 C_l = column amount of the absorber species in each layer

$$\begin{aligned}
 \frac{dS}{dq_r^{ret}} = & \sum_{l=1}^N \frac{dB_l(T_l^e)}{dq_r^{ret}} \left[e^{-\sum_{j=1}^{l-1} k_j(T_j^e, p_j^e) C_j} - e^{-\sum_{j=1}^l k_j(T_j^e, p_j^e) C_j} \right] \\
 & + \sum_{l=1}^N B_l(T_l^e) \left[e^{-\sum_{j=1}^{l-1} k_j(T_j^e, p_j^e) C_j} \sum_{j=1}^{l-1} \left(\frac{-dk_j(T_j^e, p_j^e)}{dq_r^{ret}} C_j + k_j(T_j^e, p_j^e) \frac{-dC_j}{dq_r^{ret}} \right) \right] \\
 & - \sum_{l=1}^N B_l(T_l^e) \left[e^{-\sum_{j=1}^l k_j(T_j^e, p_j^e) C_j} \sum_{j=1}^l \left(\frac{-dk_j(T_j^e, p_j^e)}{dq_r^{ret}} C_j + k_j(T_j^e, p_j^e) \frac{-dC_j}{dq_r^{ret}} \right) \right]
 \end{aligned} \quad (6.7.2)$$

To evaluate this expression, the values that have to be additionally calculated inside the forward model are the derivatives of the Planck function $\frac{dB_l(T_l^e)}{dq_r^{ret}}$, the derivative of the absorption cross sections $\frac{dk_l(T_l^e, p_l^e)}{dq_r^{ret}}$ and the derivatives of the absorber columns $\frac{dC_l}{dq_r^{ret}}$. In order to write these derivatives in a more explicit form, we have to regard that the Curtis-Godson layer values T_l^e, p_l^e and C_l are dependent on the values of temperature, pressure and volume mixing ratio at the levels which are used for the radiative transfer ($T_n^{mod}, p_n^{mod}, X_n^{mod}$). These are themselves dependent on the levels where the unknowns are retrieved ($T_r^{ret}, p_r^{ret}, X_r^{ret}$).

$$\begin{aligned}
 T_l^e &= T_l^e \left[T_n^{mod}(T_r^{ret}, p_r^{ret}), p_n^{mod}(T_r^{ret}, p_r^{ret}), X_n^{mod}(T_r^{ret}, p_r^{ret}, X_r^{ret}) \right] \\
 p_l^e &= p_l^e \left[T_n^{mod}(T_r^{ret}, p_r^{ret}), p_n^{mod}(T_r^{ret}, p_r^{ret}), X_n^{mod}(T_r^{ret}, p_r^{ret}, X_r^{ret}) \right] \\
 C_l &= C_l \left[T_n^{mod}(T_r^{ret}, p_r^{ret}), p_n^{mod}(T_r^{ret}, p_r^{ret}), X_n^{mod}(T_r^{ret}, p_r^{ret}, X_r^{ret}) \right]
 \end{aligned} \quad (6.7.3)$$

Finally the derivatives can be written as:

$$\begin{aligned}
 \frac{dB_l(T_l^e)}{dq_r^{ret}} &= \frac{\partial B_l(T_l^e)}{\partial T_l^e} \frac{dT_l^e}{dq_r^{ret}} \\
 \frac{dk_l(T_l^e, p_l^e)}{dq_r^{ret}} &= \frac{\partial k_l(T_l^e, p_l^e)}{\partial T_l^e} \frac{dT_l^e}{dq_r^{ret}} + \frac{\partial k_l(T_l^e, p_l^e)}{\partial p_l^e} \frac{dp_l^e}{dq_r^{ret}} \\
 \frac{dC_l}{dq_r^{ret}} &= \frac{dC_l}{dq_r^{ret}}
 \end{aligned} \quad (6.7.4)$$

with

$$\begin{aligned}\frac{dT_l^e}{dq_r^{ret}} &= \frac{\partial T_l^e}{\partial T_n^{mod}} \frac{dT_n^{mod}}{dq_r^{ret}} + \frac{\partial T_l^e}{\partial p_n^{mod}} \frac{dp_n^{mod}}{dq_r^{ret}} + \frac{\partial T_l^e}{\partial X_n^{mod}} \frac{dX_n^{mod}}{dq_r^{ret}} \\ \frac{dp_l^e}{dq_r^{ret}} &= \frac{\partial p_l^e}{\partial T_n^{mod}} \frac{dT_n^{mod}}{dq_r^{ret}} + \frac{\partial p_l^e}{\partial p_n^{mod}} \frac{dp_n^{mod}}{dq_r^{ret}} + \frac{\partial p_l^e}{\partial X_n^{mod}} \frac{dX_n^{mod}}{dq_r^{ret}} \\ \frac{dC_l}{dq_r^{ret}} &= \frac{\partial C_l}{\partial T_n^{mod}} \frac{dT_n^{mod}}{dq_r^{ret}} + \frac{\partial C_l}{\partial p_n^{mod}} \frac{dp_n^{mod}}{dq_r^{ret}} + \frac{\partial C_l}{\partial X_n^{mod}} \frac{dX_n^{mod}}{dq_r^{ret}}\end{aligned}\quad (6.7.5)$$

and

$$\begin{aligned}\frac{dT_n^{mod}}{dq_r^{ret}} &= \frac{\partial T_n^{mod}}{\partial T_r^{ret}} \frac{\partial T_r^{ret}}{\partial q_r^{ret}} + \frac{\partial T_n^{mod}}{\partial p_r^{ret}} \frac{\partial p_r^{ret}}{\partial q_r^{ret}} \\ \frac{dp_n^{mod}}{dq_r^{ret}} &= \frac{\partial p_n^{mod}}{\partial T_r^{ret}} \frac{\partial T_r^{ret}}{\partial q_r^{ret}} + \frac{\partial p_n^{mod}}{\partial p_r^{ret}} \frac{\partial p_r^{ret}}{\partial q_r^{ret}} \\ \frac{dX_n^{mod}}{dq_r^{ret}} &= \frac{\partial X_n^{mod}}{\partial T_r^{ret}} \frac{\partial T_r^{ret}}{\partial q_r^{ret}} + \frac{\partial X_n^{mod}}{\partial p_r^{ret}} \frac{\partial p_r^{ret}}{\partial q_r^{ret}} + \frac{\partial X_n^{mod}}{\partial X_r^{ret}} \frac{\partial X_r^{ret}}{\partial q_r^{ret}}\end{aligned}\quad (6.7.6)$$

In (6.7.5) and (6.7.6) implicit summations are assumed where a repeated index is present.

6.7.2 Derivative with respect to the volume mixing ratio

The different contributions of the volume mixing ratio derivatives are investigated.

When changing the volume mixing ratio the major effect to the derivative (equation (6.7.2)) is the change of the gas columns in each layer: $\frac{dC_l}{dX_r^{ret}}$. Test calculations have shown that using only this

term the residual errors with respect to the total derivatives are about 1-10%, with the largest errors near the tangent level.

These errors are due to neglecting terms $\frac{dB_l(T_l^e)}{dX_r^{ret}}$ and $\frac{dk_l(T_l^e, p_l^e)}{dX_r^{ret}}$ mainly through the dependence

of the Curtis-Godson value T_l^e on the volume mixing ratio. Adding the effect of $\frac{dB_l(T_l^e)}{dX_r^{ret}}$ to our

previous calculations reduces the errors to about 1-5%.

In order to further improve this derivative a lot of effort is needed: it is necessary to determine

$\frac{dk_l(T_l^e, p_l^e)}{dX_r^{ret}}$, for which (equation 6.7.4) $\frac{\partial k_l(T_l^e, p_l^e)}{\partial T_l^e}$ and $\frac{\partial k_l(T_l^e, p_l^e)}{\partial p_l^e}$ must be calculated:

$$\begin{aligned} \frac{\partial k_l(T_l^e, p_l^e)}{\partial T_l^e} &= \frac{\partial \left[\sum_{li} L_{l,li}(T_l^e) A_{l,li}(T_l^e, p_l^e) \right]}{\partial T_l^e} \\ &= \sum_{li} \left[A_{l,li}(T_l^e, p_l^e) \frac{\partial L_{l,li}(T_l^e)}{\partial T_l^e} + L_{l,m}(T_l^e) \frac{\partial A_{l,li}(T_l^e, p_l^e)}{\partial T_l^e} \right] \end{aligned} \quad (6.7.7)$$

and

$$\frac{\partial k_l(T_l^e, p_l^e)}{\partial p_l^e} = \frac{\partial \sum_{li} [L_{l,li}(T_l^e) A_{l,li}(T_l^e, p_l^e)]}{\partial p_l^e} = \sum_{li} L_{l,li}(T_l^e) \frac{\partial A_{l,li}(T_l^e, p_l^e)}{\partial p_l^e} \quad (6.7.8)$$

where li = index for the different lines
 $A_{l,li}(T_l^e, p_l^e)$ = line shape of line li
 $L_{l,li}(T_l^e)$ = line intensity of line li

It is easy to calculate the derivative of the line intensity with respect to T_l^e but the derivative of the line shape is more problematic. Also if there would be a possibility to calculate it this would not save computation time, since the formula would be more complicated and would need more calculation time than a recalculation of the whole spectrum.

In any case, these calculations have to be done for each single line during the calculation of the absorption cross sections. This needs much more time than the calculation of the quantity $\frac{dC_l}{dX_r^{ret}}$.

Therefore the baseline for our program is to compute analytical derivatives which only contain $\frac{dC_l}{dX_r^{ret}}$. In case in future it turns-out that the accuracy of these VMR derivatives is not satisfactory, improvements can be obtained by reducing the layers thickness since this minimises the influence of the other contributions to the whole derivative.

6.7.3 Derivative with respect to temperature

The main contributions to this derivative result from the derivative of the Planck function $\frac{\partial B_l(T_l^e)}{\partial T_l^e}$

(equation 6.7.4) and the derivative of the line strength $\frac{\partial A_{l,m}(T_l^e)}{\partial T_l^e}$. The larger is the temperature dependence of the line strength, i.e. the bigger E'' and the more important is the latter derivative. Therefore, in our tests, the errors when only taking into account $\frac{\partial B_l(T_l^e)}{\partial T_l^e}$ range from 5-40%. Since,

as we said above, the calculation of $\frac{\partial \mathcal{L}_{l,m}(T_l^e)}{\partial T_l^e}$ has to be done during the calculation of the absorption cross sections, it requires much more time than $\frac{\partial \mathcal{B}_l(T_l^e)}{\partial T_l^e}$. Our baseline is not to calculate the analytical derivatives with respect to the temperature and use numerical derivative instead. However, the numerical derivatives are implemented in an optimised form (i.e. the calculation of spectra with the ‘T-perturbed’ profiles is parallel to the one with the ‘original’ profile), and not by just recalling the forward model.

6.7.4 Derivative with respect to the atmospheric continuum

The derivative with respect to the atmospheric continuum can easily be performed since the continuum is taken into account as an absorption cross section which is multiplied by the air column of each path. The Curtis-Godson temperatures, pressures or the total air column do not change when varying the continuum cross sections. Equation 6.7.1 can be written as:

$$\frac{dS}{dk_r^{cont,ret}} = \frac{d}{dk_r^{cont,ret}} \sum_{l=1}^N B_l(T_l^e) \left[e^{-\sum_{j=1}^{l-1} k_j^{cont,e} C_j^{air}} - e^{-\sum_{j=1}^l k_j^{cont,e} C_j^{air}} \right] \quad (6.7.9)$$

and for 6.7.2 follows:

$$\frac{dS}{dk_r^{cont,ret}} = \sum_{l=1}^N B_l(T_l^e) \left[e^{-\sum_{j=1}^{l-1} k_j^{cont,e} C_j^{air}} \sum_{j=1}^{l-1} \frac{-dk_j^{cont,e}}{dk_r^{cont,ret}} C_j^{air} - e^{-\sum_{j=1}^l k_j^{cont,e} C_j^{air}} \sum_{j=1}^l \frac{-dk_j^{cont,e}}{dk_r^{cont,ret}} C_j^{air} \right] \quad (6.7.10)$$

$k_l^{cont,e}$ are the continuum absorption cross sections of the forward model layers and are therefore dependent on the forward model levels ($k_n^{cont,mod}$) which are themselves dependent on the levels where the continuum has to be retrieved ($k_r^{cont,ret}$):

$$\frac{dk_l^{cont,e}}{dk_r^{cont,ret}} = \frac{\partial k_l^{cont,e}}{\partial k_n^{cont,mod}} \frac{\partial k_n^{cont,mod}}{\partial k_r^{cont,ret}} \quad (6.7.11)$$

The baseline is to use these analytical derivatives.

6.7.5 Derivative with respect to the tangent pressure

The main effects of this derivative result from the change of the column, the change of the line shape and the change of the temperature with the tangent pressure. For the modelling of the change of the line shape it is necessary to do calculations in the domain of the absorption cross sections which are very time consuming. Therefore, we decided not to calculate this derivative as given by the formulas above, but to determine it during the convolution with the FOV-function. The

spectrum $S^F(\sigma, p)$ is calculated for the field of view function FOV centred at the tangent pressure. Since the convolution with the FOV is an analytical expression, its derivative with respect to the central pressure of the FOV can be easily obtained.

Our baseline is to use this kind of analytical derivatives for the calculation of the derivatives with respect to the tangent pressure.

6.7.6 Independence of retrieved variables

The Jacobian matrix (equation 4.2.6) contains the partial derivatives of the spectrum with respect to the parameters. Therefore, while performing the calculation of the derivatives, one has to take care that either they are partial (if parameters are dependent) or that the parameters are independent ($\partial \text{parameter}_i / \partial \text{parameter}_j = 0, \forall i \neq j$), so that the partial derivatives are equal to the total ones.

This assumption is clearly fulfilled for the volume mixing ratio retrieval where the parameters are the volume mixing ratios and the continuum cross sections at the tangent levels and the instrumental offset.

During the p-T retrieval one has to be more careful since (due to the optimised calculation of the tangent pressure derivatives, cf. 6.7.5) varying the tangent pressure implies also a change of the tangent temperature and of the atmospheric continuum on the tangent levels. This difficulty disappears when we define for each iteration cycle as the fitted parameters the new tangent pressure and the temperature and continuum cross sections at the tangent pressures of the previous iteration. In a subsequent step the new temperatures and continuum cross sections at the previous tangent pressures are interpolated to the new tangent pressures.

6.8 Convergence criteria

In Sect. 4.2.4 four possible conditions have been considered for the definition of reached convergence. These are:

1. A small relative variation of χ^2 with respect to the previous iteration (Eq.(4.2.14)).
2. A small maximum correction applied to the parameters vector (Eq.(4.2.15)).
3. A small relative difference between the χ^2 value calculated at present iteration and the value estimated with linear extrapolation at the previous iteration (Eq.(4.2.16)).
4. A maximum allowed number of iterations (Eq.(4.2.18)).

While condition 4 is only used to avoid loops in case of convergence difficulties, each of the other three conditions has important physical implications.

Condition 1 indicates that a minimum has been reached and the desired level of exploitation of the available information has been obtained. This does not guarantee, however, that in retrieved parameters the random errors of the observations are mapped to the desired level of accuracy so that systematic errors do not dominate in case of averaging. Furthermore, in case of non-linearities a false convergence may be obtained.

Condition 2 indicates that the desired level of accuracy has been reached in the determination of the unknowns, however, if a conservative threshold is applied at all the altitudes, it may be very

difficult to meet this requirement. The alternative condition 2 bis (Eq.(4.2.15 bis)) may be easier to meet, but can provide false convergence at the beginning of the iterative process.

Condition 3 indicates that a linear region has been reached by the iteration process (the reliability of the test is limited by the accuracy of the derivatives calculations).

Both conditions 1 and 2 are needed to assess the quality of the solution with respect of both measurement error and retrieval accuracy requirements. Condition 3 is needed to assess the validity of condition 1.

The scientific code baseline is to use as convergence criterion the result of the OR operation applied to conditions 2 and 3. The thresholds used for conditions 2 and 3 must be optimised experimentally with suitable tests.

6.9 Pre-calculation of line shapes

This item concerns the line shape calculation in the case of HNO_3 lines. Since a lot of lines have to be taken into consideration in HNO_3 microwindows this is a very time consuming part of the VMR-retrieval process. In order to optimise this calculation we can use the fact that in the HITRAN data base the HNO_3 lines have the same Lorentz half width and coefficient of the temperature dependence (equation 4.4.12). It is therefore not necessary to recalculate the line shape (equation 4.4.15), but it can be calculated once at the beginning of the cross section calculation for a new path ($p_{\text{eq}}-T_{\text{eq}}$ pair) and used for all other HNO_3 lines of the microwindow. For each transition the pre-calculated line shape is then centred at the central frequency and interpolated to the wavenumber grid of the microwindow. This interpolation is performed linearly. The resulting error was determined by test calculations. The maximum difference between exact calculation and use of the interpolated pre-calculated line shape was NESR/88. Due to this small value of the introduced error and to the time saving of 66%, our baseline is to use the pre-calculated HNO_3 line-shapes.

6.10 Different grids during the cross-section calculation

Up to now we assumed, that the cross-sections are calculated at each grid point of the fine grid, in which also the radiative transfer calculations are performed. The interval between two grid points is of the order of $5 \times 10^{-4} \text{ cm}^{-1}$. Tests indicated that a reasonable and conservative value for the interval between two fine grid points can be $5 \times 10^{-4} \text{ cm}^{-1}$. This value, which can in principle be optimised in terms of faster computations, as been adopted as baseline. This results in 4000 points for a 2 cm^{-1} microwindow where cross sections for each transition have to be calculated. Therefore, the run time is directly proportional to the number of grid points. In order to reduce the number of grid points during the calculation of cross sections two methods are used by recent line-by-line codes (e.g. Edwards, 1991; Gordley, 1994):

1. the grid can be coarser far off the line centre than near the centre.
2. the grid can be proportional to the half-width of the line, i.e. it can be dependent on the pressure of the layer for which the cross-sections have to be calculated.

These two methods have been implemented into the subroutine for the line-by-line calculation of cross section in the following way:

In addition to the constant general fine grid ($\Delta^{gf} = 5 \times 10^{-4} \text{ cm}^{-1}$) two grids, the local coarse (Δ^{lc}) and the local fine grid (Δ^{lf}), are defined for each path where the cross sections are calculated. The grid distances are multiple integers of each other:

$$\Delta^{lf} = n\Delta^{gf}, \Delta^{lc} = m\Delta^{lf} \quad (6.11.1)$$

with $n, m \in \mathbb{N}$, and $1 \leq n, 1 \leq m$

- m , which determines the local coarse grid, is a test parameter which can be varied during the optimisation phase.
- n is the nearest positive integer value so that $\Delta^{lf} \approx \xi(\alpha^{L_m} + \alpha^{D_m})$, where $\alpha^{L_m}, \alpha^{D_m}$ are the Lorentz and the Doppler half width of the target gas transition with the largest intensity, and ξ is a second optimisation parameter that determines the period of the local fine grid.
- The third parameter ζ defines the distance $\zeta(\alpha^{L_m} + \alpha^{D_m})$ of the transition between local fine and local coarse grid from the line centre.

Following the calculation of the cross sections for all lines of the microwindow both grids are linearly interpolated to the general fine grid.

Run time tests with this implementation showed a considerable time saving for the calculation of the absorption cross sections of more than 50%.

6.11 Cross-section look-up tables

The use of pre-computed look-up tables (LUTs) is an alternative method to the explicit line-by-line calculation of absorption cross sections (equation 4.4.6). The basic idea of this method is to pre-calculate for each frequency grid point the absorption cross sections of each gas for a set of different pressures and temperatures within the range of the atmospheric variability. These data are stored in files which are read at the beginning of each retrieval. Then they are interpolated to the equivalent pressure and temperature of the atmospheric paths (p_{eq}, T_{eq} pairs). Since the frequency grid in which the cross sections have to be calculated is rather fine ($5 \times 10^{-4} \text{ cm}^{-1}$, see Sect. 6.10) the amount of data is large.

In order to reduce the amount of data contained in the look-up tables and their reading time, a compression procedure has been studied at University of Oxford (see Morris (1997)). The algorithm used to build compressed look-up tables and procedure for the decompression are based on the matrix singular value decomposition applied to the cross-sections and is described in Strow et al., (1998).

Since version 2.3, the ORM code is able to handle these compressed look-up tables, including cases in which the look-up tables are available only for a sub-set of the operational microwindows and / or for a sub-set of the gases contributing to the emission in each microwindow.

The importance of time savings obtained using LUTs instead of line-by-line calculations depends very much on the considered retrieval. Largest savings are obtained in the case of HNO_3 retrieval (38% reduction of forward model runtime), less significant savings (15% reduction of forward model runtime) are obtained in the cases of O_3 and CH_4 . In other cases the use of compressed LUTs provides no time savings (N_2O) or an increase of computing time like in p,T and H_2O retrievals.

LUTs represent an optimisation of certain efficiency only when irregular frequency grids (see Sect. 6.12) are used in combination.

6.12 Variable frequency grids for radiative transfer computation

Limb radiance spectra contain spectral features on a range of scales varying from the narrow, isolated, Doppler-broadened line centers at high altitudes, to wide, overlapping, Lorentz-broadened line wings from low altitudes.

Therefore, a minimum subset of spectral grid points (irregular grid) can be determined that are sufficient to reconstruct full radiance spectra, applicable over a range of tangent altitudes and atmospheres. Full radiative transfer calculations are then only required for this subset of points, the remaining fine grid points are obtained using a pre-determined interpolation scheme. The irregular grid is a function of the microwindow boundaries, the chosen interpolation scheme and the spectral convolution represented by the Instrument Line Shape.

Typically it is found that only 5-10 % of the full resolution grid is required for reconstruction of the spectra. When combined with the Look-Up Tables, this also means that absorption cross-section only has to be reconstructed at the same fraction of grid points resulting in even more efficient optimization. On the other hand, the use of irregular grids does not reduce the computation time of the algorithm for line-by-line evaluation of the cross-section. This algorithm in fact already uses its own internal, line-specific, irregular grid (see Sect. 6.10) and takes little advantage from the externally provided optimized grids.

Since version 2.3, the ORM code is able to exploit externally provided, microwindow-specific irregular grids. Due to the very significant time savings obtained with the combined use of LUTs and irregular grids (a factor 10 on the computing time of the full retrieval chain) and to the smallness of the introduced inaccuracies ($< \text{NESR} / 10$ in spectral radiances), the ORM baseline is to use both LUTs and irregular grids in operational retrievals.

The adoption of irregular grids also takes care of the frequency dependent optimisation of the fine grid baseline.

Both LUTs (see Sect. 6.11) and irregular grids optimized for MIPAS microwindows are currently calculated by a dedicated algorithm developed at Oxford University (see Morris, (1997) and Wells, (1997)). A summary description of this algorithm is reported in Appendix D.

7 - Validation of the choices and future work

Several physical and mathematical choices made for the design of the scientific code require either a “tuning” for the optimisation of some variables or a validation for an experimental assessment of the critical assumptions that have been used to progress in the study.

Presently, several input parameters have already been tuned on the basis of comparisons between spectra simulated with different levels of accuracy. Also, most of the adopted choices have already been validated by results of intercomparisons with the RFM. However, several assumptions regarding both the instrument and the atmospheric model still require a final validation on the basis of real measurements. Some work has already been done in this direction by analysing both MIPAS-B2 (Dinelli et al., (2000)) and ATMOS (see e.g Flaud and Piccolo, (2001)) data sets with the developed retrieval algorithm. However these instruments are significantly different compared to MIPAS / ENVISAT in particular, MIPAS-B2 operates from a stratospheric balloon, while ATMOS measures absorption. For this reason we expect that very important outcomes may still emerge from the calibration and validation activity planned for the commissioning phase (Piccolo et al., 2001).

References

- Bennett, V. L., A. Dudhia and C. D. Rodgers, *Microwindow Selection for MIPAS using Information Content*, ESAMS'99 Proceedings, 18-22 January 1999, Noordwijk, NL.
- Born, M., E. Wolf, '*Principles of Optics*', Oxford, Pergamon Press, 5th edition, p. 87, 1975.
- Carli, B., '*Document on the choice between retrieval of profile at fixed levels and at tangent altitude levels*', Action item from 24th MIPAS SAG Meeting of 15 November 1995, 4 December 1995.
- Carli, B., B.M.Dinelli, P.Raspollini, M.Ridolfi, '*Discrete representation and resampling in the case of limb-sounding measurements*', Appl. Optics, Vol. 40, No. 8, p. 1261-1268, (2001).
- Carlotti M., '*Global-fit approach to the analysis of limb-scanning atmospheric measurements*', Appl. Optics, 27, 3250-3254, (1988).
- Carlotti, M., B.Carli, '*Approach to the design and data analysis of limb-scanning atmospheric measurements*', Appl. Optics, 33, 15, 3237-3249, (1994)
- Carlotti, M., G. Paruolo, F. Alboni, B. Carli, F. Mencaraglia, M. Ridolfi, '*Determination of atmospheric pressure and temperature distributions from MIPAS limb emission line spectra*', ESA contract report No. 142956, 1995.
- Carlotti, M., B.M.Dinelli, P.Raspollini, M.Ridolfi, '*Geo-fit approach to the analysis of satellite limb-scanning measurements*', Appl. Optics, Vol. 40, No. 12, p. 1872 – 1875, (April 2001).
- Clarmann T., '*Untersuchungen zur Strahldichteberechnung mit Linie - für - Linie - Computerprogrammen*', Diplomarbeit für Meteorologie, Universität München, Meteorologisches Institut, 1986.
- Clough, S. A., F. X. Kneizys, L. S. Rothman, and W. O. Gallery, '*Atmospheric spectral transmission and radiance: FASCOD1B*', SPIE 277, 152-166, 1981.
- Clough, S. A., F. X. Kneizys and R. W. Davies, '*Line shape and the water vapor continuum*', Atmospheric Research, 23 229-241, 1989.
- Cousin, C., R. Le Doucen, C. Boulet, and A. Henry, '*Temperature dependence of the absorption in the region beyond the 4.3 μm band head of CO_2 . 2: N_2 and O_2 broadening*', Appl. Opt., 24, 3899-3907, 1985.
- Defense Mapping Agency, '*WGS84 Ellipsoidal Gravity Formula and Gravity Anomaly Conversion Equations*': DMA Aerospace Center, St. Louis p. 10.
- Delbouille, L., G. Roland, '*Assesment of finite field-of-view effects on MIPAS ILS & Review of resolution enhancement techniques*', Answer to ESA NTO/ME/1573.
- Dinelli, B.M., E.Battistini, B.Carli, '*MIPAS-B2 data analysis*' TN-ISM-0002, (2000).
- Drayson, S. R., '*Rapid computation of the Voigt profile*', J Quant. Spectr. Radiat. Transfer, 16, 611-614, 1976.
- Dudhia, P. E. Morris and R. J. Wells, *Acceleration of Radiative Transfer Calculations* ESAMS'99 Proceedings, 18-22 January 1999, Noordwijk, NL.
- Dudhia, A., '*MIPAS ultimate retrieval accuracy*', final report, 19 June 2000 (a).
- Dudhia, A., '*MIPAS ultimate retrieval accuracy*' appendix to the final report, 7 December 2000 (b).
- Dudhia, A., '*Latitude-dependence of MIPAS microwindow selection*', report of task 5.7 of the CCN5 of the ESA contract 11717/95/NL/CN, 12 February 2001.
- Edlen, B. Metrologia, 2:12, 1966.
- Edwards, D. P., '*GENLN2 a general line-by-line atmospheric transmittance and radiance model*' Version 3.0 Description and Users Guide, 1992.
- Flaud, J.-M., C.Piccolo, B.Carli, '*Update A of the spectroscopic Database for MIPAS*', TN-LPM-IROE-01, draft, (May 2001).
- Gordley, L. L., B. T. Marshall, D. A. Chu, *LINEPAK: algorithms for modelling spectral transmittance and radiance*, J Quant. Spectr. Radiat. Transfer, 52, 563-580, 1994.
- Hollweg, H.D., V.S. Kostsov, G. Schlusser, P. Schlusser, Y.M. Timofeyev, *Interaction at MM and Optical Frequencies*, ESA Interim Report Contract 10603/93/NL/NB, 1994.
- Humlicek, J., '*Optimized computation of the Voigt and complex probability functions*', J. Quant. Spectr. Radiat. Transfer, 27, 437-444, 1982.

Lafferty, W..J., A.M.Solodov, A.Weber, W.B.Olson, J.-M.Hartmann: '*Infrared collision-induced absorption by N₂ near 4.3 μ m for atmospheric applications: measurements and empirical modelling*', Appl. Opt., **35**, 5911-5917, (1996).

List R. J. ed.: *Smithsonian Meteorological Tables*, Sixth Revised Edition, Washington, D.C., 1963.

Morris P., '*Generation of compressed cross-section look-up tables for NRT MIPAS retrievals*', ESA report PO-TN-OXF-GS-0011, (1997).

C. Piccolo, M. Prosperi, M. Ridolfi, V. Tenna, '*Level 2 Algorithm Characterization and Validation Plan*', prog. doc. n. TN-IROE-GS0101, (2001).

Press, W. H., S. A. Teukolsky, W. T. Wetterling, B. P. Flannerly: '*Numerical Recipes in Fortran*', Second Edition Cambridge University Press 1992.

P.Raspollini and M.Ridolfi, '*Mapping of Temperature and Line-of-Sight Errors in Constituent Retrievals for MIPAS / ENVISAT Measurements*', Atmospheric Environment, Vol. **34**, Nos. 29 – 30, p. 5329 – 5336 (2000).

Ridolfi M., B.Carli, M.Carloti, T.v.Clarmann, B.M.Dinelli, A.Dudhia, J.-M.Flaud, M.Hoepfner, P.E.Morris, P.Raspollini, G.Stiller, R.J.Wells, '*Optimized forward model and retrieval scheme for MIPAS near-real-time data processing*' Appl. Optics, Vol. **39**, No. 8, p. 1323 – 1340 (10 March 2000).

Rodgers C. D. '*Retrieval of Atmospheric Temperature and Composition From Remote Measurements of Thermal Radiation*', Reviews of Geophysics and Space Physics, Vol. 14 N. 4, p.609, 1976.

Rodgers C. D., '*Inverse Methods for Atmospheric Sounding: Theory and Practice*', Series on Atmospheric, Oceanic and Planetary Physics – Vol. 2, World Scientific, (2000).

Rosenkranz, P. W., '*Shape of the 5 mm oxygen band in the Atmosphere*', IEEE Trans. Antennas Propag. 23, 498-506, 1975.

Schreier, F., '*The Voigt and complex error function: a comparison of computational methods*', J Quant. Spectr. Radiat. Transfer, 48, 743-762, 1992.

Sterne T. E. '*An Introduction to Celestial Mechanics*', Interscience publishers Ltd, London, 1960.

Strow, L.L., Motteler H.E., Benson R.G., Hannon S.E. and De Souza-Machado S. '*Fast Computation of monochromatic infrared atmospheric transmittances using compressed look-up tables*'. J. Quant. Spectrosc. Radiat. Transfer, 59, p481-493, (1998).

Thibault, F., V.Menoux, R.Le Doucen, L.Rosenmann, J.-M.Hartmann, Ch.Boulet: '*Infrared collision-induced absorption by O₂ near 6.4 μ m for atmospheric applications: measurements and empirical modelling*', Appl. Opt., **36**, 563-567, (1997).

Turner, D. S., '*Absorption coefficient estimation using a two-dimensional interpolation procedure*', J. Quant. Spectrosc. Radiat. Transfer, 55, 633-637, 1995.

Wells, R., '*Generation of optimized spectral grids*', Technical report for ESA study 11886/96/NL/GS, document: PO-TN-OXF-GS-0010, (14th July 1997).

Appendix A: Memorandum on determination of the VCM of engineering tangent heights in MIPAS

Prepared by M.Ridolfi and B.Carli, (22 July 1999)

A1. Introduction

The ORM code uses engineering pointing information in p,T retrieval. This information is included in the retrieval using the optimal estimation method, therefore also a variance-covariance matrix (VCM) is needed for a proper weighting of engineering data by the inversion algorithm (see Sect. 4.2.6). Since MIPAS pointing system is presently characterized only by very general engineering specifications, some assumptions must be made and an algorithm must be set up to build a realistic VCM of the engineering tangent altitudes, starting from the specified performances. Based on the results of this algorithm, a suitable strategy must be studied for an efficient storage of MIPAS pointings VCMs in Level 2 framework.

In the present memorandum we elaborate an algorithm for deriving a realistic VCM of pointings to be used by p,T retrieval and subsequently propose a strategy for storing these VCMs in Level 2 framework.

A2. MIPAS pointing performance

The VCM of pointings is built on the basis of some pieces of information provided by British Aerospace (BAe) which is responsible for the platform and for compiling the pointing budgets. BAe reports MIPAS pointing stability for 4.0 and 75 s time intervals for the three satellite axes (x -axis being the most critical for MIPAS pointing accuracy). x -axis stability, in terms of tangent altitude, is:

- 230 m for 4 s stability
- 660 m for 75 s stability

BAe provides also the total pointing accuracy:

- 2000 m is the total accuracy

The reported values have a confidence level of 95.4%, meaning that the above values are not exceeded in 95.4% of the cases. The errors are not purely gaussian because they include e.g. linear drifts due to temporary unavailability of the stars used by the satellite star sensors. However, in order to exploit the formalism of the statistics, we will consider these errors as gaussian with standard deviation equal to **half** of the above reported figures (Note: we are assuming that the stability provided by BAe is an excursion from an average value, a quarter should be used if the provided value is a peak-to-peak excursion).

BAe tells also that for time intervals between 4 and 75 s no analyses have been made, however in these cases, the best approximation is to linearly interpolate between the above reported figures. This approximation will not be exploited in the proposed algorithm because it does not provide realistic stability figures for time intervals much shorter than 4s and much longer than 75s. In Sect.A3 a more sophisticated interpolation scheme is proposed.

Another assumption we will use in the following is about the speed of MIPAS interferometer. We assume that MIPAS will be always operated at a 5 cm/s speed independently of the adopted spectral resolution. Furthermore we will assume the 'turn-around' time, i.e. the time required for speed inversion and positioning of the limb-scanning mirror, to be equal to 0.45 s. Scans with altitude step greater than 10 km characterized by a turn-around time greater than 0.45 s will not be considered

here. In this hypothesis the time Δt required for measuring a sweep with resolution identified by MPD is given by:

- $$\Delta t(s) = \frac{MPD(cm)}{5cm/s} + 0.45s$$

A3. Algorithm

From the above figures, the total pointing error ($\sigma_{tot} = 1000$ m) can be intended as absolute error of the individual tangent heights, while the stability specifications can be exploited (as it will be explained) to derive the correlations between tangent heights.

Let's calculate explicitly the correlation $c_{i,k}$ between two generic tangent heights z_i and z_k , assuming that they have been measured at times t_i and t_k . The general expression of the correlation provides:

$$c_{i,k} = \frac{\sum_{j=1}^N (z_i(j) - \bar{z}_i)(z_k(j) - \bar{z}_k)}{\left[\left(\sum_{j=1}^N (z_i(j) - \bar{z}_i)^2 \right) \cdot \left(\sum_{j=1}^N (z_k(j) - \bar{z}_k)^2 \right) \right]^{\frac{1}{2}}} \quad (A1)$$

where \bar{z}_i is given by:

$$\bar{z}_i = \sum_{j=1}^N \frac{z_i(j)}{N} \quad (A2)$$

and the index j ranges over an hypothetical set of N measurements of the tangent heights z_i and z_k with $i, k = 1, 2, \dots, N_{LS}$ (N_{LS} = number of sweeps in the considered limb-scanning (LS) sequence).

Let's indicate:

$$\varepsilon_i(j) = (z_i(j) - \bar{z}_i) \quad (A3)$$

$\varepsilon_i(j)$ is the error on $z_i(j)$ in the sense that it is the deviation of $z_i(j)$ from its 'true' value which is represented by the average of equation (A2).

If the two tangent heights z_i and z_k have been measured at times t_i and t_k such that $\Delta t = |t_i - t_k|$, their errors cannot differ too much due to the stability specifications of the pointing. In particular we will have:

$$\varepsilon_k(j) = \varepsilon_i(j) + \delta_{\Delta t}(j) \quad (A4)$$

where $\delta_{\Delta t}(j)$ is a random term with standard deviation $\sigma_{\Delta t}$. In order to calculate $\sigma_{\Delta t}$ from the specified short- and long- term stability we will use the following function:

$$\sigma_{\Delta t} = \sigma_{tot} \cdot \left(1 - \exp(-\alpha \cdot \Delta t^\beta)\right) \quad (A5)$$

with α and β constants determined imposing $\sigma_{\Delta t=4s} = \sigma_{4s}$ and $\sigma_{\Delta t=75s} = \sigma_{75s}$, where $\sigma_{4s} = 115$ m and $\sigma_{75s} = 330$ m are the standard deviations associated respectively to the 4s and to the 75 s specified stability. Please note that, as it is logically required, expression (A5) provides $\sigma_{\Delta t} \Rightarrow 0$ for $\Delta t \Rightarrow 0$ and $\sigma_{\Delta t} \Rightarrow \sigma_{tot}$ for $\Delta t \Rightarrow \infty$. The behavior of $\sigma_{\Delta t}$ as a function of Δt is plotted in Fig.A1.

The standard deviation of ε_k (equation (A3)) can be expressed as:

$$\begin{aligned} \frac{1}{N} \sum_{j=1}^N \varepsilon_k^2(j) &= \frac{1}{N} \sum_{j=1}^N [\varepsilon_i(j) + \delta_{\Delta t}(j)]^2 = \\ \frac{1}{N} \sum_{j=1}^N \varepsilon_i^2(j) &+ \frac{2}{N} \sum_{j=1}^N \varepsilon_i(j) \delta_{\Delta t}(j) + \frac{1}{N} \sum_{j=1}^N \delta_{\Delta t}^2(j) = \\ \sigma_{tot}^2 &+ \frac{2}{N} \sum_{j=1}^N \varepsilon_i(j) \delta_{\Delta t}(j) + \frac{1}{N} \sum_{j=1}^N \delta_{\Delta t}^2(j) \end{aligned} \quad (A6)$$

Now, since ε_k must have standard deviation equal to σ_{tot} , from equation (A6) we get:

$$\sum_{j=1}^N \varepsilon_i(j) \delta_{\Delta t}(j) = -\frac{1}{2} \sum_{j=1}^N \delta_{\Delta t}^2(j) \quad (A7)$$

Substituting expressions (A3) and (A7) in (A1) we obtain:

$$c_{i,k} = \frac{\sum_{j=1}^N \varepsilon_i(j) [\varepsilon_i(j) + \delta_{\Delta t}(j)]}{N \sigma_{tot}^2} = 1 - \frac{\sum_{j=1}^N \delta_{\Delta t}^2(j)}{2N \sigma_{tot}^2} \equiv 1 - \frac{\overline{\delta_{\Delta t}^2}}{2\sigma_{tot}^2} \quad (A8)$$

Where in the second step we have used:

$$\sigma_{tot}^2 = \frac{\sum_{j=1}^N \varepsilon_i^2(j)}{N} \quad (A9)$$

Considering that:

$$\overline{\delta_{\Delta t}^2} = \sigma_{\Delta t}^2 + \overline{\delta}^2_{\Delta t} \quad (\text{A10})$$

and that, from heuristic considerations, it should be:

$$\overline{\delta}^2 = 2\sigma_{tot}^2 - \sigma_{\Delta t}^2 - 2\sigma_{tot}\sqrt{\sigma_{tot}^2 - \sigma_{\Delta t}^2} \quad (\text{A11})$$

equation (A8) becomes:

$$c_{i,k} = \sqrt{1 - \frac{\sigma_{\Delta t}^2}{\sigma_{tot}^2}} \quad (\text{A12})$$

From a more qualitative point of view, this very simple result can also be justified as follows. The tangent height z_k has two error components: the first component (σ_1) is linked to the measurement of the neighboring tangent height z_i , the second component $\sigma_{\Delta t}$ does not depend on previous measurements. The two components must satisfy:

$$\sqrt{\sigma_{\Delta t}^2 + \sigma_1^2} = \sigma_{tot} \quad (\text{A13})$$

since the error associated to z_i is σ_{tot} , the correlation between tangent heights z_i and z_k is by definition:

$$c_{i,k} = \frac{\sigma_1}{\sigma_{tot}} = \frac{\sqrt{\sigma_{tot}^2 - \sigma_{\Delta t}^2}}{\sigma_{tot}} = \sqrt{1 - \frac{\sigma_{\Delta t}^2}{\sigma_{tot}^2}} \quad (\text{A14})$$

where the value for σ_1 has been extracted from equation (A13).

Equation (A12), together with expression (A5) provides the tool for calculating the correlation between two generic tangent heights z_i and z_k . This tool can be exploited for computing the VCM \mathbf{V} of the tangent heights whose elements $V_{i,k}$ are given by:

$$V_{i,k} = \sigma_{tot}^2 \cdot c_{i,k} \quad (\text{A15})$$

The VCM \mathbf{V}^z relating to the differences between tangent heights (whose inverse is used by the ORM) can be obtained through the transformation:

$$\mathbf{V}^z = \mathbf{J}\mathbf{V}\mathbf{J}^t \quad (\text{A16})$$

where \mathbf{J} is the jacobian matrix that represents the linear transformation leading from tangent heights to differences between tangent heights. If we indicate with $\Delta z_i \equiv z_{i+1} - z_i$, the jacobian \mathbf{J} contains the derivatives:

$$J_{i,k} = \frac{\partial(\Delta z_i)}{\partial z_k} = \begin{cases} -1 & \text{if } i = k \\ 1 & \text{if } i = k - 1 \\ 0 & \text{in the other cases} \end{cases} \quad (\text{A17})$$

with $i = 1, \dots, N_{LS}-1$ and $k = 1, \dots, N_{LS}$.

A4. Software tool

A very simple software tool has been implemented that computes the VCMs \mathbf{V} and \mathbf{V}^z of MIPAS pointing system using the explained algorithm. Besides the parameters defining pointing performances described in Sect.A2, the only inputs of this tool are the max. path difference and the number of sweeps of the limb-scanning sequence for which we want to calculate the VCM of pointings. The outputs of this program are:

- ✓ correlation matrix of tangent heights
- ✓ VCM of tangent heights
- ✓ correlation matrix of differences between tangent heights
- ✓ errors on differences between tangent heights
- ✓ VCM of differences between tangent heights
- ✓ inverse of VCM of differences between tangent heights

A5. Results

In Fig. A2 we report the correlations between different tangent heights for a scan of 16 sweeps and MPD = 20 cm as a function of the sweep index. In Fig. A3 we report correlations of differences between tangent heights for the same scan of Fig.A2 as a function of the index. The same quantities are reported in Fig's A4 and A5 respectively, for a scan of 16 sweeps and MPD = 5 cm (reduced resolution).

In the adopted approach, the absolute errors of both tangent heights and differences between tangent heights are constant with altitude. The errors on differences between tangent heights depend however on the selected MPD. The dependence of these errors on the MPD is shown in Fig. A6.

General comments are:

- the absolute error on tangent heights is a constant (does not depend on MPD)
- the correlation between tangent heights increases when decreasing the resolution (i.e. decreasing MPD)
- decreasing the resolution, in consequence of the increased correlations, the errors on the differences between tangent heights decrease (see Fig. A6).

A6. Adopted strategy for handling pointings VCM in Level 2 framework

The quantity required in input to the ORM is the inverse of the VCM of the differences between tangent heights. Given the invariance of the obtained results with respect to the sweep index we the following approach has been adopted for storage / handling of pointing VCM in Level 2 framework. The file 'PI_VCM.DAT' (defined in the ICD document, PO-IF-DOG-GS-0002, Issue 1c) contains VCMs of the tangent heights tabulated as a function of max. path difference. The tabulated VCMs refer to a scan with a maximal number N_{\max} of sweeps (e.g. $N_{\max} = 30$ sweeps). Given a scan

with N_{sw} (with $N_{sw} < N_{max}$) sweeps to be analyzed and max. path difference $MPD = xx$, a block matrix of dimension $N_{sw} \times N_{sw}$ is extracted from the VCM relating to $OPD = xx$. Rows and columns relating to corrupted sweeps will be then removed from this block. The remaining rows and columns are transformed according to equation (A16) and the resulting matrix is inverted and provided in input to Level 2 processor.

Despite the differences existing between the present algorithm and the algorithm developed during the first MIPAS pT retrieval study (ESTEC Purchase Order No: 142956 terminated in Sept.'95) for the calculation of pointing VCMs, the results of the two algorithms are consistent in the case of $MPD = 20$ cm. Presently it is not possible to use any longer the old algorithm due to the fact that in the old algorithm the spectral resolution was assumed constant. The analytical expression presented in this memorandum provides a more simple and versatile calculation tool.

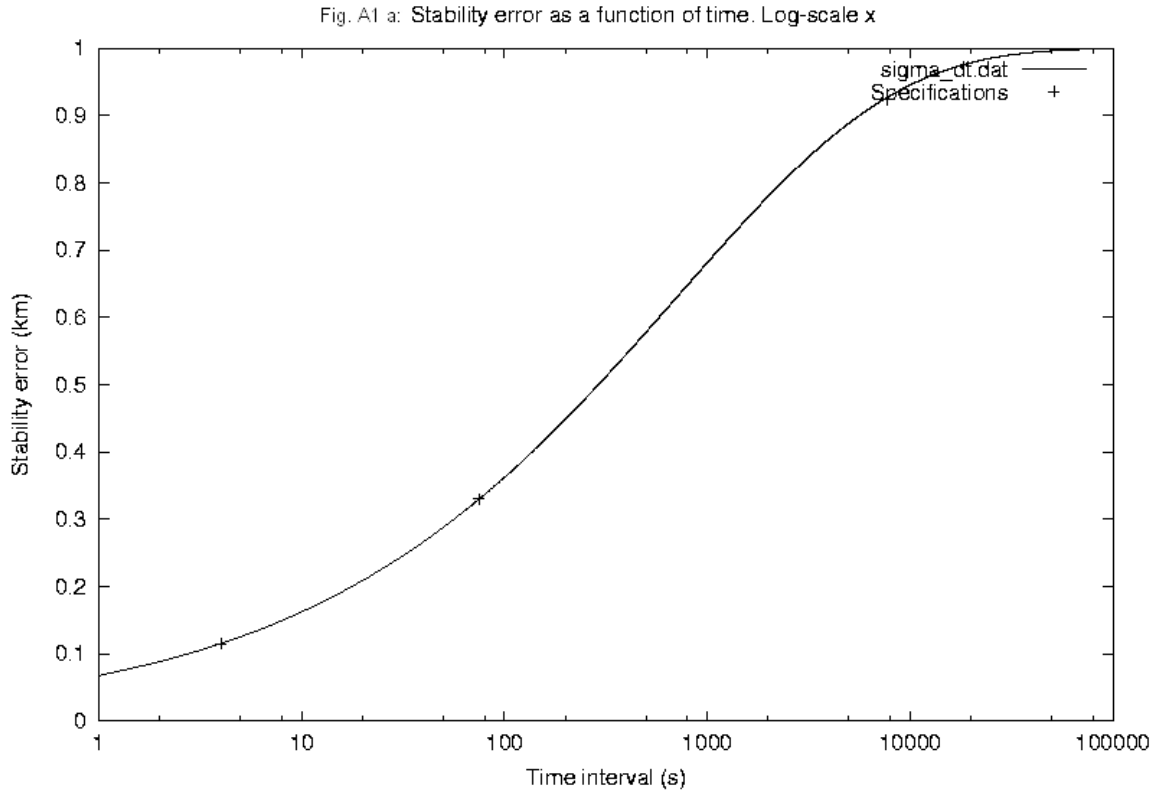


Fig. A1 b: Stability error as a function of time. Linear scale x

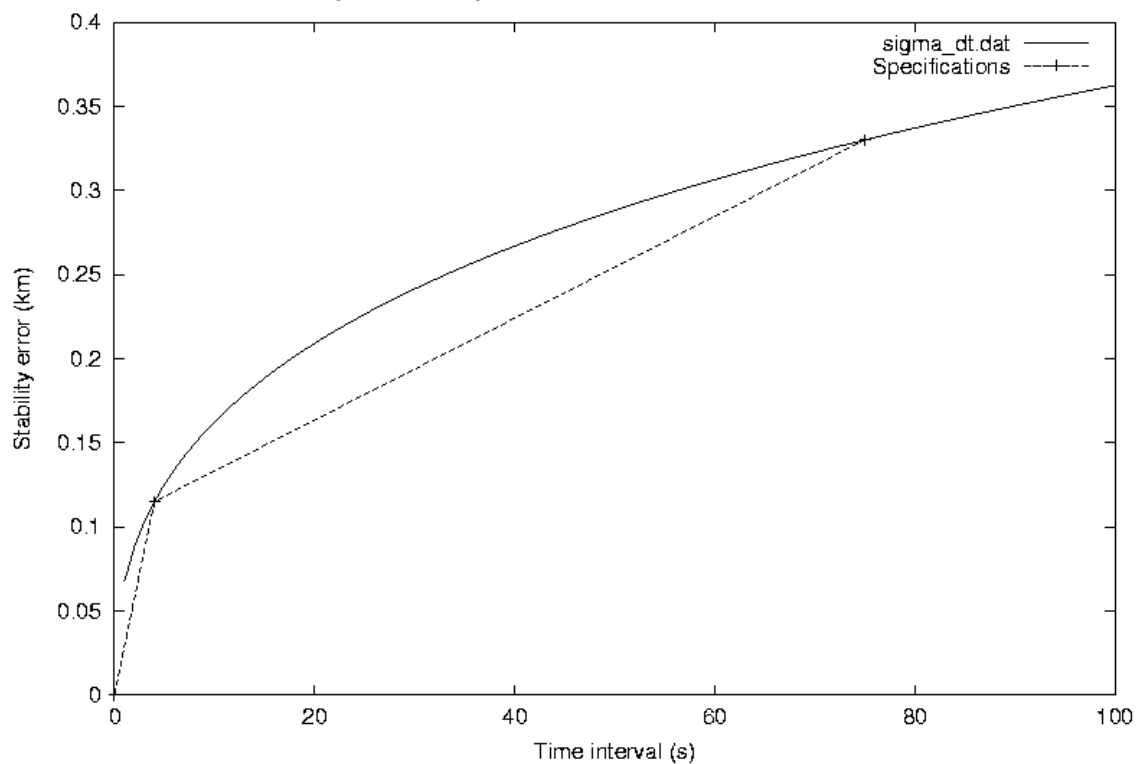


Fig. A2: Correlations between Z, OPD = 20 cm, N. of sweeps = 16

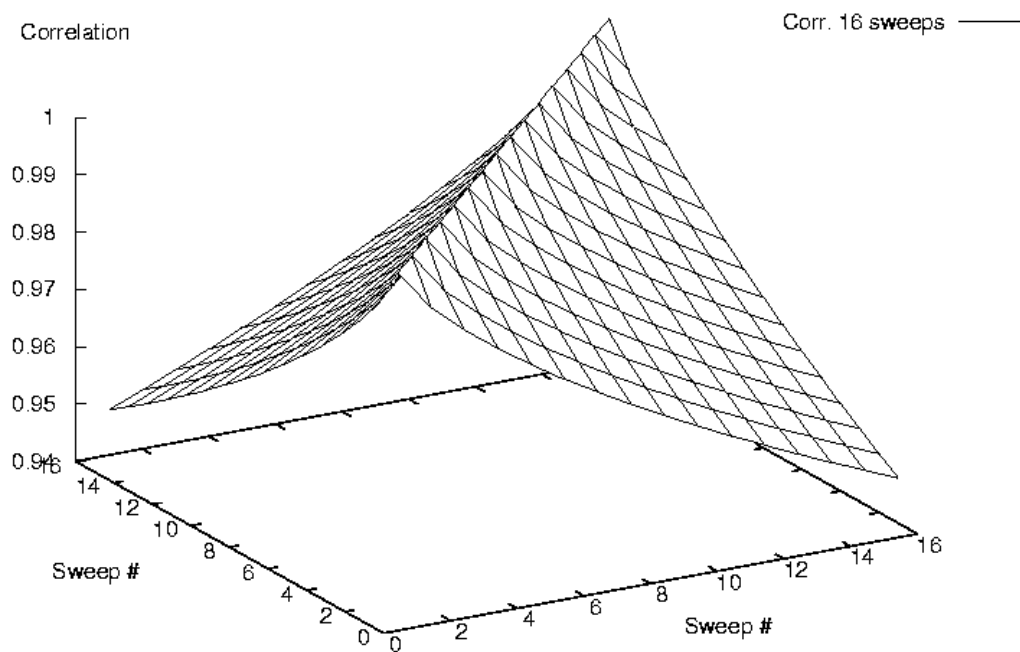


Fig. A3: Correlations between dZ, MPD = 20 cm, N. of sweeps = 16

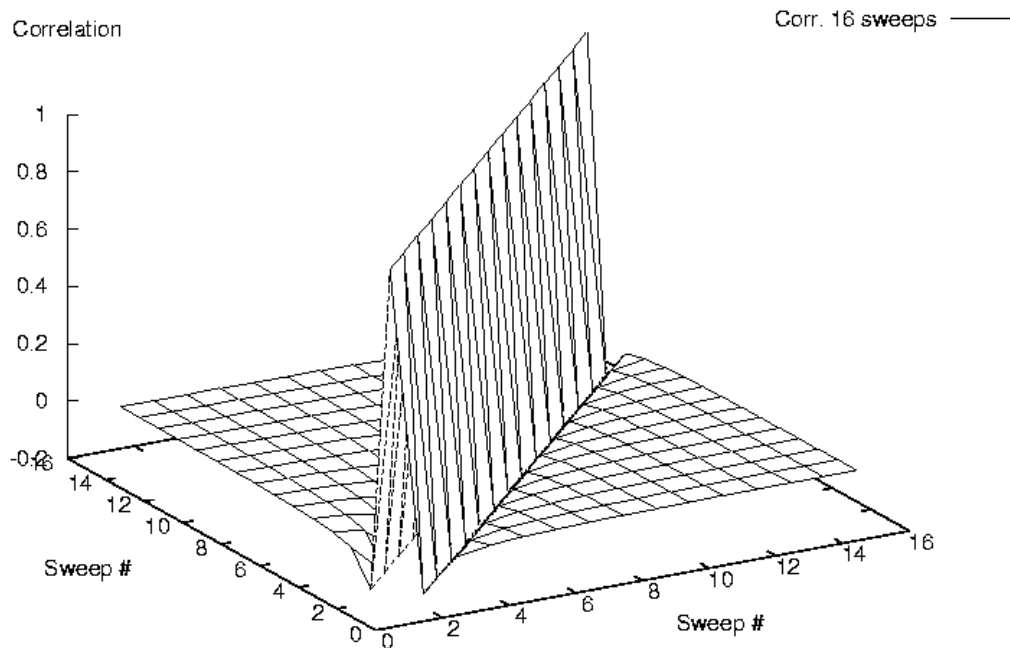


Fig. A4: Correlations between Z, OPD = 5 cm, N. of sweeps = 16

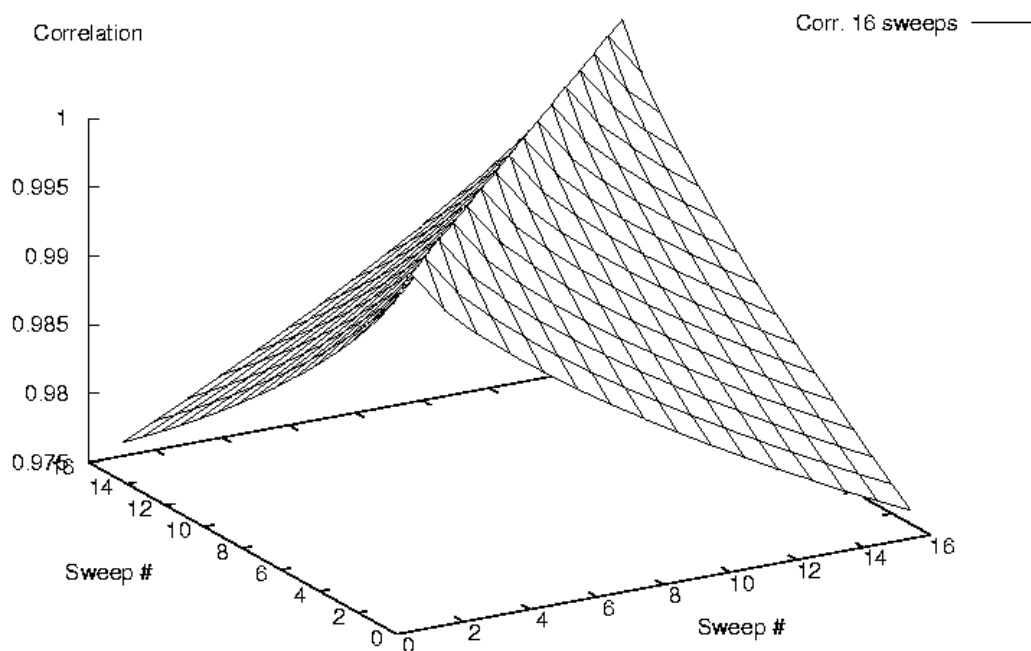


Fig. A5: Correlations between dZ , MPD = 5 cm, N. of sweeps = 16

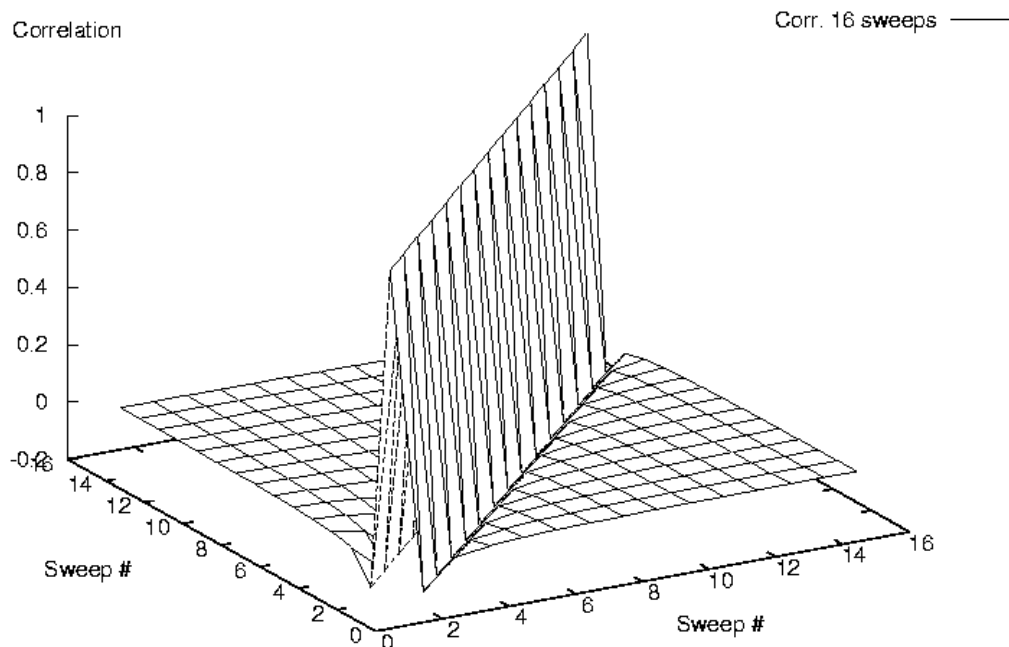
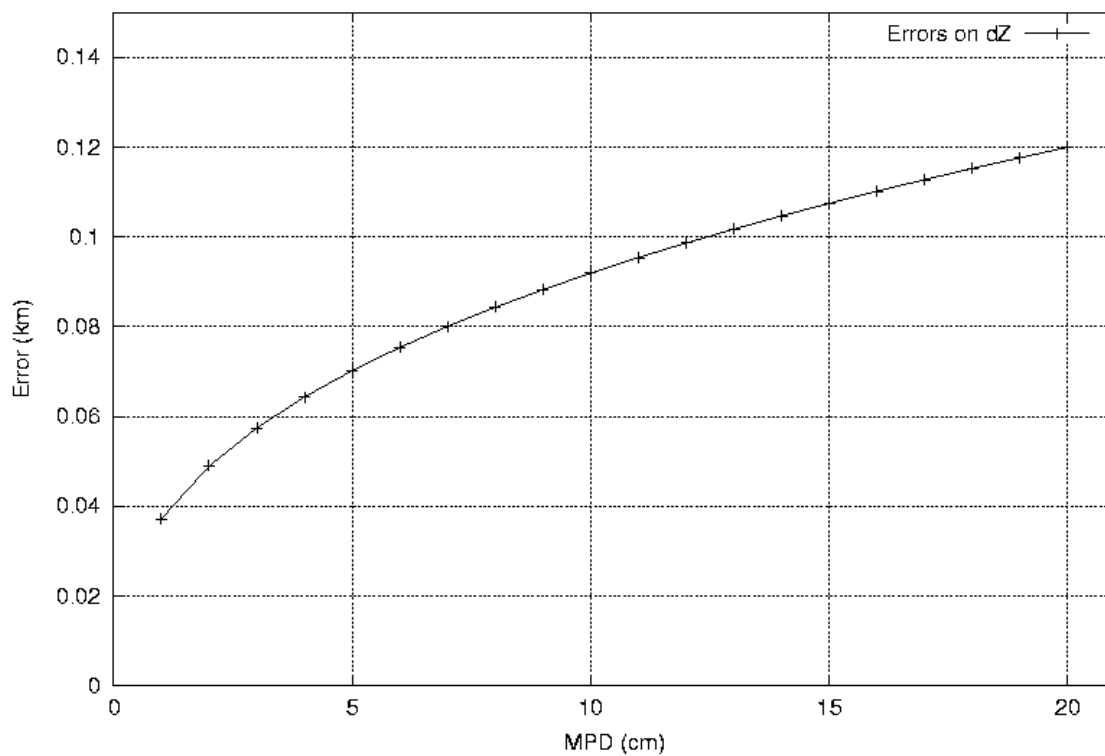


Fig. A6: Errors on dZ for different MPDs. Errors do not depend on sweep.



Appendix B: Evaluation of retrieval error components and total error budget

The recommendations arising from the 30th MIPAS Scientific Advisory Group meeting (ESTEC, 25th and 26th June 1998) regarding total retrieval error (see SAG minutes MIN-MIPAS-0030), can be summarized by the following statements:

- the random error affecting the retrieved profiles has two components:
 - ⇒ The first random component is due to measurement noise. This component is evaluated as part of the inversion procedure (Gauss-Newton method) by the retrieval program. The ORM output files already contain the retrieved profile(s) variance-covariance matrix obtained from measurement noise (derived in Level 1b processing) without scaling by the actual χ^2 of the fit. This variance-covariance matrix will be included in Level 2 products.
 - ⇒ The second random component of profiles error is due to pressure and temperature error propagation in VMR retrievals. This component will be evaluated, on-line, by a function implemented in Level 2 processor (Framework) and included in Level 2 products.
- Systematic error on the retrieved profiles has several components:
 - ⇒ **Contaminants:** this error component is due to imperfect assumption of the contaminants profiles in the forward model.
 - ⇒ **Horizontal homogeneity error:** due to horizontal homogeneity assumption in the forward model (only T gradient is considered)
 - ⇒ **Spectroscopic errors:** are due to errors in spectroscopic data used by the forward model.
 - ⇒ **Instrumental errors.** Due to imperfect (frequency and intensity) calibration of the observed spectra, imperfect knowledge of the ILS (and of the FOV).
 - ⇒ **Model errors.** These are purely systematic errors biasing the simulated observations. For the moment only errors arising from neglecting Non-LTE and using LUTs and IGs in the forward model are expected to belong to this group.

Systematic errors in contaminant profiles, spectroscopic data, horizontal homogeneity, and instrument have some 'random' character in the sense that, in principle, their value may vary from retrieval to retrieval. For this reason the rule with which they combine may vary as a function of the application. This prevents the calculation of a general purpose error budget.

The evaluation of the total systematic errors affecting the retrieved profiles is not a task of the on-line Level 2 processor. The development of a database is planned containing systematic error profiles relating to VMR profiles retrieved using different (pre-stored) occupation matrices. Level 2 products will contain a reference to that database. Systematic errors will not be available for generic occupation matrices which eventually could be constructed on-line by a function implemented in the Level 2 processor.

A scientific software tool was developed for evaluating the different error contributions explained above and the resulting total error budget. Purpose of the present section is to describe the algorithms implemented in this software tool. The tool evaluates separately:

1. Temperature / pressure induced errors in VMR retrievals
2. Systematic error components
3. Total error budget

as a function of:

- error quantifiers reported in the microwindow database,
- current occupation matrix, p-,T-error propagation matrix
- current accuracy of temperature / pressure retrieved profiles

- user input parameter defining the error components to be taken into account for the evaluation of total error.

The part of this tool calculating temperature and pressure induced errors in VMR retrieval constitutes a scheme for the implementation of the corresponding function in Level 2 processor framework. If necessary, this tool may be used for the generation of a database of systematic errors to be attached to profiles retrieved with different (pre-stored) occupation matrices.

B1. Temperature and pressure induced errors in VMR retrievals

A generic retrieved VMR profile \mathbf{y} is obtained through the inversion formula:

$$\mathbf{y} = \left(\mathbf{K}^T \mathbf{V}_{obs}^{-1} \mathbf{K} \right)^{-1} \mathbf{K}^T \mathbf{V}_{obs}^{-1} \mathbf{n} \equiv \mathbf{Dn} \quad (\text{B1.1})$$

where \mathbf{n} is the residuals vector, \mathbf{V}_{obs} is the VCM of the observed spectra and \mathbf{K} is the jacobian of the VMR retrieval.

An uncertainty $\Delta(\mathbf{p}, \mathbf{T})$ on the assumed tangent pressures and temperatures, translates into an error $\Delta\mathbf{n}$ on the simulated spectra and therefore into an error $\Delta\mathbf{y}$ on the retrieved profile equal to:

$$\Delta\mathbf{y} = \mathbf{D} \cdot \Delta\mathbf{n} = \mathbf{D} \cdot \mathbf{C} \cdot \Delta(\mathbf{p}, \mathbf{T}) \quad (\text{B1.2})$$

where \mathbf{C} is the matrix accounting for p,T error propagation in the simulated spectra of VMR retrieval and contains the derivatives:

$$C_{i,j} = \frac{\partial \mathcal{S}_i}{\partial \left(\begin{matrix} P \\ T \end{matrix} \right)_j} \quad (\text{B1.3})$$

The index 'i' identifies the fitted spectral points (as a function of frequency for all the microwindows and all the tangent altitudes) and the index 'j' identifies the retrieved tangent altitudes.

In equation (B1.2) we have assumed \mathbf{D} as locally independent of p,T (always true for small errors $\Delta(\mathbf{p}, \mathbf{T})$).

As the error on the retrieved p,T is in our case described by a VCM \mathbf{V}_{PT} , the corresponding VCM \mathbf{V}_y relating to \mathbf{y} and due to p,T error is given by:

$$\mathbf{V}_y = \mathbf{D} \mathbf{C} \mathbf{V}_{PT} (\mathbf{D} \mathbf{C})^T \equiv \mathbf{E} \mathbf{V}_{PT} \mathbf{E}^T \quad (\text{B1.4})$$

where we have defined $\mathbf{E} = \mathbf{D} \mathbf{C}$. \mathbf{E} is the matrix transforming p,T error into VMR error. The order of magnitude of the dimensions of this matrix are:

\mathbf{E} : (16 VMR retrieved points) x (32 p,T retrieved points) x (4 bytes/datum) = 2 Kb

In principle, matrix \mathbf{E} depends on:

- current atmospheric status (p,T and VMR)
- set of adopted MWs in VMR retrieval (Occupation Matrix)

These dependencies have been studied with full details in Raspollini and Ridolfi, (2000).

The strategy adopted in the Level 2 framework to calculate pT error propagation in VMR retrievals consists in:

- creating in Level 2 framework a database of matrices \mathbf{E} : since matrix \mathbf{E} depends on the chosen set of MWs (occupation matrix), as many \mathbf{E} matrices as many pre-defined occupation matrices must be stored in Level 2 framework. Given the relatively small size of matrix \mathbf{E} , this is not cause of concern. Matrices \mathbf{E} are calculated by means of sensitivity tests.
- A first order dependence of \mathbf{E} on the atmospheric status is allowed. The coefficients that establish the first order dependence are calculated with numerical tests and optionally attached to the \mathbf{E} matrices included in the database.
- \mathbf{V}_y' is evaluated using equation (B1.4).

B2. Systematic error components

The microwindow database contains quantifiers for the evaluation of systematic errors affecting the retrieved profiles. All the quantifiers have the form:

$$d_{\varepsilon_k} = \frac{1}{VMR(z)} \cdot \frac{\partial VMR}{\partial \varepsilon_k} \cdot \Delta \varepsilon_k \quad (B2.1)$$

where ε_k identifies the systematic error source. In case of a multi-MW retrieval in which N_{MW} microwindows are used at altitude z , the VMR error $\sigma_{\varepsilon_k}(z)$ induced by source ε_k at altitude z is given by:

$$\frac{\sigma_{\varepsilon_k}(z)}{VMR(z)} = \frac{\sum_{j=1}^{N_{MW}} \left(d_{\varepsilon_k}(z) \right)_j}{\sum_{j=1}^{N_{MW}} \left(\sigma_N^2(z) \right)_j} \quad (B2.2)$$

In the case of **model errors**, which have known sign, $\left(d_{\varepsilon_k}(z) \right)_j$ is the summation of all model error components relating to MW j at altitude z .

B3. Total error budget

Assuming as ‘independent’ the various sources of error affecting the retrieved profiles, the total error is obtained from the summation of the variance covariance matrices (VCM) connected with the different error sources. As for the construction of the MW database the correlation among different altitudes is neglected, the VCM connected with error sources different from measurement noise and p,T error are diagonal matrices.

In conclusion, the total retrieved profile error is represented by the VCM \mathbf{V}_{TOT} given by:

$$\mathbf{V}_{TOT} = \mathbf{V}_N + \mathbf{V}_y' + \sum_{k=1}^{N_{SYST}} \mathbf{V}_{\varepsilon_k} \quad (\text{B3.1})$$

where:

the summation extends over all the (N_{SYST}) systematic errors, \mathbf{V}_N is the VCM due to measurement noise and is directly provided by the retrieval algorithm (Gauss-Newton method), \mathbf{V}_y' is the VCM connected with p,T error propagation in VMR retrieval (provided by the algorithm described in Sect. B1 of the present appendix) and $\mathbf{V}_{\varepsilon_k}$ is the VCM due to error source ε_k . $\mathbf{V}_{\varepsilon_k}$ are diagonal matrices, with diagonal elements equal to:

$$\mathbf{V}_{\varepsilon_k}(i, i) = \sigma_{\varepsilon_k}^2(z_i) \quad (\text{B3.2})$$

where σ_{ε_k} is provided by equation (B2.2).

In equation (B3.1), the first two terms represent the random component of the profile error, of course, in the case of p,T retrieval \mathbf{V}_y' does not exist and will not be included in the total error evaluation. The summation appearing in the last term of expression (B3.1) represents the systematic component of the profile error.

Appendix C: Generation of MW Databases and Occupation matrices

Microwindow selection is performed by an algorithm which simulates the propagation of random and systematic errors through a retrieval and attempts to maximise the information content (Bennett et al, 1999). The information content of a microwindow increases as the "log" of the determinant of the total covariance decreases, total covariance being the sum of the random and various systematic error covariances. Broadly speaking, 1 'bit' of information is equivalent to a factor 2 reduction in the uncertainty at one profile altitude.

Microwindows are created by first selecting a number of single measurements, identified by location in the spectral and tangent altitude grids, as starting points. Adjacent measurements are added to each until the information content no longer improves or the maximum width of 3 cm^{-1} is reached. The best of these trial microwindows is selected, the retrieval covariance modified, and the process repeated for a new set of measurements as starting points. The procedure of growing microwindows also allows for measurements within microwindows to be 'masked', i.e., excluded from the retrieval. This usually applies to measurements where the associated systematic errors such as the uncertainty in modelling a contaminant, outweigh any benefit in the reduction of the random error when considering the total covariance.

Initially, a set of typically 10 microwindows, or 10000 measurements (whichever occurs first) is selected based on the assumption that spectra for all MIPAS bands are available. Further microwindows are then selected to maximise information retrieved in situations where data from different bands may be unavailable. This set of 20-30 microwindows constitutes the database.

Occupation matrices represent subsets of microwindows to be used under different retrieval circumstances, and these are constructed using the same approach: selecting the microwindows from the database (rather than growing new microwindows) in the sequence which maximises the retrieved information. A number of these OMs are precomputed, corresponding to different band-availabilities, and associated with each of these is a single figure-of-merit representing the information content.

Further details of the algorithm used to grow optimized spectral microwindows are included in Dudhia, (2000 a), Dudhia, (2000 b) and in Dudhia, (2001).

From the above description it is clear that a side product of this algorithm is the total error affecting the retrieved profiles as a function of altitude. By operating the algorithm without constraints limiting the maximum number of observations included in the analysis, it is possible to obtain ultimate accuracy figures for the parameters retrieved from MIPAS measurements.

Fig. A1 reports a summary of the ultimate retrieval accuracy as evaluated in Dudhia, (2000 b).

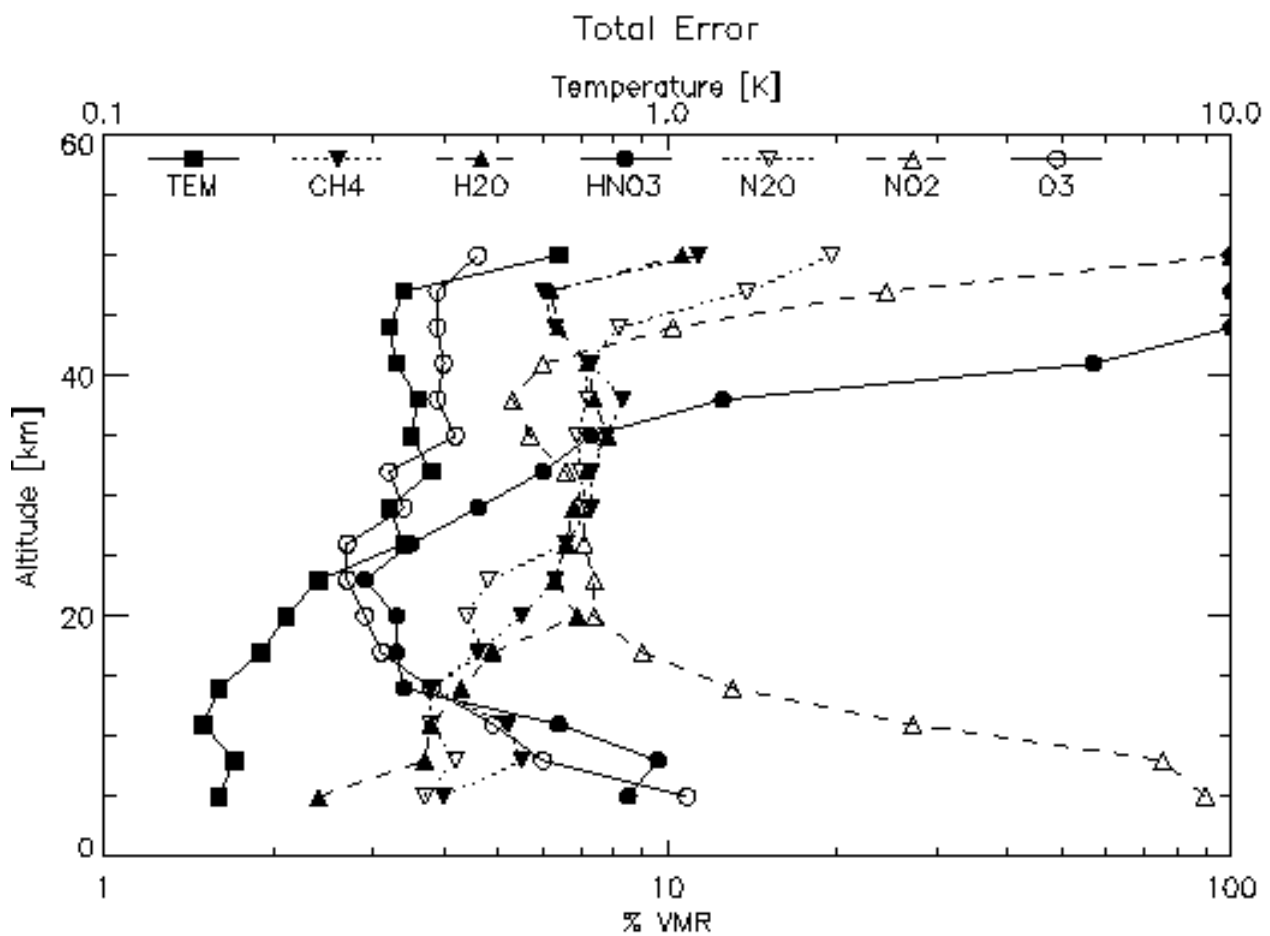


Fig. C1: Summary of ultimate accuracy results (from Dudhia, (2000 b))

Appendix D: Generation of LUTs and Irregular frequency Grids (IG)

The creation of lookup tables and the irregular grids for spectral calculations is a process consisting of four stages (Dudhia et al, 1999).

The first stage is to determine the significant absorbers for each microwindow. The maximum radiance for all potential absorbers is assessed considering atmospheres containing just that absorber in isolation with its maximum concentration. Any absorber contributing more than 10% of the NESR is considered significant and added to the absorber list for the microwindow.

The second stage is to establish the tabulation axis increments in temperature and $\ln[\text{pressure}]$ for the absorber look-up tables. Large tables (i.e., small axis increments) of monochromatic absorption coefficient are initially created for all microwindow absorbers. MIPAS radiances are modelled for three different atmospheres (representing nominal, minimum and maximum absorber concentrations) using these tables and compared with line-by-line calculations for the same cases. The tabulation axis increments are iteratively increased for each absorber until the maximum discrepancy reaches NESR/30.

The third stage is to perform the singular value decomposition (SVD) of these tables. Initially, each absorber table is SVD-compressed, retaining 30 singular values. MIPAS radiances are modelled using the SVD-compressed tables for the three atmospheres as in the previous stage, and compared with the same line-by-line calculations. The number of singular values for each absorber table is iteratively reduced until the maximum discrepancy reaches NESR/15.

The fourth stage is to determine the irregular grid for spectral integration. This starts with the MIPAS radiances calculated with the SVD-compressed LUTs in the previous stage, which are determined on a regular fine grid (0.0005 cm^{-1} spacing) prior to the AILS convolution. Grid points are then selectively removed until the maximum discrepancy compared to a line-by-line on the full spectral grid reaches NESR/10.

Appendix E: Generation of MW-dedicated spectral linelists

By feeding a given line-by-line radiative transfer forward model only with the spectral lines actually needed for the microwindows and measurement geometry under consideration, a lot of computation time can be saved. This lines selection can be performed either online, i.e. during the run of radiative transfer codes or offline. Off-line, spectroscopic data preselection offers the advantage, that the importance of a certain transition can be more easily assessed on the basis of the related radiance signal rather than by means of absorption cross sections. This is particularly important for nonlinear radiative transfer in the case of a non-isothermal atmosphere.

Each transition in the HITRAN96 database within 25 cm^{-1} of the microwindow boundaries is examined, while lines outside this margin are supposed to be parametrized by a continuum model. The effect of neglecting a line in radiance calculations is estimated for each nominal MIPAS tangent altitude between 5 and 83 km and for each predefined microwindow and forms the basis of the decision whether a line should be included or excluded. For reasons discussed below, the contribution of a spectral line has to be assessed within an iterative loop. Therefore, a simple and quick approach to estimate the contribution of a transition is needed, in order to reduce computation time to tolerable limits. A common and simple approach for this purpose is to neglect convolution by any lineshape function and to approximate the contribution of a spectral line, L as

$$L = B(T) \times (1 - \exp(-S(T) \times M)) \quad (\text{E1})$$

for a homogeneous atmosphere where

L = radiance

T = Curtis-Godson mean value of temperature over whole path,

B = Planck function

S = line intensity

M = integrated mass of the gas along the optical path.

This approach is equivalent to the so-called $S \times M$ spectra which are commonly used to estimate the radiance contribution of a transition. This approximation, however, which avoids both layer-by-layer calculation and evaluation of the line shape, while being favourable with respect to computational expenses, turned out not to be satisfactory in terms of accuracy. We found, that for tangent altitudes below the stratopause, the homogeneous and optically thin approximation is not sufficient, because higher and thus warmer layers in the atmosphere significantly contribute to the spectrum. Furthermore, due to pressure broadening and spectral apodization, signal from other lines is imported to the signal at a particular line center. Therefore, we had to account for more realistic line shapes.

A more sophisticated approach describes the monochromatic radiative transfer through the atmosphere by a two-layer model which supports temperature inhomogeneities:

$$L = B(T_1) \times (1 - \exp(-k_1(T_1, p_1) \times M_1)) \times \exp(-k_2(T_2, p_2) \times M_2) \\ + B(T_2) \times (1 - \exp(-k_2(T_2, p_2) \times M_2)) \quad (\text{E2})$$

with

p_1, p_2 = Curtis-Godson mean values of pressure in layers 1 and 2, respectively,
 T_1, T_2 = Curtis-Godson mean values of temperature in layers 1 and 2, respectively,
 M_1, M_2 = slant path column amounts in layers 1 and 2, respectively,
 k_1, k_2 = absorption cross sections in layers 1 and 2, respectively.

The optimal layering turned out to be such that the slant path absorber amount in the tangent layer M_1 is about 96% of the total slant path column amount $M_1 + M_2$.

The line shape is modeled by the Voigt function. The radiance spectra are convolved with a triangular (in frequency space) apodization function leading to a spectral resolution of 0.05 cm^{-1} in terms of FWHM (= full width at half maximum). The frequency grid is set up using 5 gridpoints within each HWHM (= half width at half maximum) of each line, and a coarser grid in the gaps between the lines.

By this model, the contribution of other lines are explicitly considered at the center of one particular line under consideration. However, continua (aerosols etc.) and no gases for which only absorption cross-sections but no spectroscopic line data are available (CFCs etc.) are not taken into account. This simplified model provides reasonably accurate radiance spectra within a tolerable amount of computing time. For all cases checked, it was within 10 - 20% of the "true" radiance usually providing an overestimate of the true radiance. It should be mentioned again that the purpose of this simplified radiance modeling is only to assess the importance of a certain transition within a line rejection loop.

Due to non-linearity of radiative transfer and the contribution of overlapping lines to the radiance at a particular line position, and since the contribution of each line considered for exclusion depends on the signal of lines still to be included, this approach requires an iterative processing. Starting with the spectrum of all lines, the radiance spectrum of lines considered for exclusion is compared to an error threshold. By checking the spectrum of neglected lines only rather than the difference between all-lines and selected-lines spectra, it is guaranteed not to underestimate the contribution of any line in a potentially saturated spectrum. Within each iteration step, the lines contributing most to this 'error spectrum' are re-included in the list of transitions to be considered.

Selection criteria are based on the expected apodized noise equivalent spectral radiance (NESR) of the MIPAS instrument. The error threshold to fall below in any case is set to NESR/10. In a first step, lines from outside the microwindows are selected according to their radiance contribution within the microwindow. Those lines remaining in the linelist for a microwindow will therefore provide a sufficiently accurate estimate of the overall contribution from outside lines. Due to non-linearity in radiative transfer, this continuum-like contribution has to be taken into account when selecting the lines lying inside the microwindow.

The treatment of lines outside the microwindow is performed as follows: Only HITRAN lines less than 25 cm^{-1} from the microwindow boundaries are considered for further investigation. First, lines are sorted according to the criterion:

$$\alpha \times S \times M_1^2 \times \left(\frac{\partial k}{\partial \nu} \Big|_{\nu_1} - \frac{\partial k}{\partial \nu} \Big|_{\nu_2} \right) \quad (\text{E3})$$

where

S = line intensity of candidate transition

α = Lorentzian half width of the candidate transition

M_l = mass in tangent layer

$\frac{\partial k}{\partial \nu}|_{\nu_1}$ = wavenumber derivative of absorption coefficient at microwindow boundary ν_1

$\frac{\partial k}{\partial \nu}|_{\nu_2}$ = wavenumber derivative of absorption coefficient at microwindow boundary ν_2

This expression weights the contribution of the considered line to the radiance of a far wing continuum brought in by lines from outside the microwindow. Furthermore, it gives less weight to weakly frequency-dependent contributions, reflecting the fact that in the OFM/ORM the continuum-like signal is not modeled on a line-by-line basis.

Beginning with the weakest ones, according to this criterion, lines are excluded as long as the apodized radiances of the excluded lines at the microwindow boundaries do not exceed a pre-defined threshold which depends on the apodized noise equivalent spectral radiance and the radiance of all lines at the microwindow boundary under consideration:

- a) $T_a = 0.07 \text{ NESR}$ if $L_b \geq 0.7 \text{ NESR}$
- b) $T_a = 0.1 L_b$ if $0.7 \text{ NESR} > L_b \geq 0.07 \text{ NESR}$
- c) $T_a = 0.007 \text{ NESR}$ if $L_b \leq 0.07 \text{ NESR}$.

For exclusion of lines inside the microwindow, the threshold for the maximum allowed radiance signal of removed lines is calculated from the maximum radiance of all lines.

- a) $T_i = \text{NESR}/10$ if $L_{\max} \geq \text{NESR}$,
- b) $T_i = L_{\max}/10$ if $L_{\max} < \text{NESR}$.

If at any spectral gridpoint ν inside the microwindow the radiance signal of excluded lines exceeded the threshold, T_i , the line which contributed most at this spectral gridpoint and is centered within a predefined interval $\nu \pm \Delta\nu/2$ where $\Delta\nu = \text{apodized spectral resolution} = 0.05 \text{ cm}^{-1}$, is re-included to the linelist. This step is repeated iteratively as long as the threshold is no longer exceeded at any spectral gridpoint. In order to avoid endless loops, lines once re-included are never again excluded. Typically after 5 to 10 iterations the error threshold is no longer reached at any spectral gridpoint within the microwindow. By checking the radiance spectrum of the potentially negligible lines instead of checking the residual spectrum between full calculation and calculation with a reduced linelist, a conservative estimate of the contribution of the neglected lines is achieved.

The remaining lines are compiled in a database which is organized microwindow by microwindow and contains, beside spectroscopic data, the following table entries:

- tangent altitude range where a line has to be considered
- a flag indicating whether the line shape has to be evaluated on a fine grid or if a coarse grid is sufficient

The suitability and appropriateness of the produced linelists is checked by FASCOD2 reference calculations. FASCOD2 radiance spectra were calculated with the full HITRAN96 database and the selected linelist for more than 70 cases. These test calculations proved that the radiance difference due to neglecting lines is always significantly below the MIPAS NESR, thus justifying the two-layer approach for the purpose of lines selection.

The efficiency of line reduction is very high due to the optimized iterative selection scheme. The number of lines outside a microwindow (corresponding to a 25 cm^{-1} interval around the microwindow) that need to be considered is reduced to 0.15 - 1.8% of the full HITRAN line number

for low tangent heights (< 20 km) while for tangent heights above 40 km it is less than 0.03%. Inside the microwindows, 17 - 77% of the full HITRAN lines contribute considerably below 20 km, and 1.3 - 18% have to be used above 40 km (see Table 1). The resulting reduction of computational time for the line-by-line absorption coefficient calculation is directly proportional to the line number reduction.

Target gas	Number of mw's	Average number of HITRAN lines	Reduced number of lines (target and contaminant gases) in the linelist [% of full HITRAN]					
O ₃	121	Out: 23331 In: 255	1.1 53	0.73 42	0.26 31	0.032 18	0.003 8.7	0 5.1
p-T	130	Out: 23309 In: 111	1.8 44	0.67 40	0.17 30	0.034 15	0.001 6	0.001 2.5
H ₂ O	101	Out: 9107 In: 152	0.302 24	0.145 18	0.034 11	0.001 4.4	0 2.1	0 1.4
CH ₄	57	Out: 22562 In: 230	0.66 53	0.31 41	0.048 20	0.001 2.9	0 2.2	0 1.8
N ₂ O	45	Out: 15209 In: 116	0.63 34	0.23 24	0.033 12	0 5	0 4.9	0 4.2
HNO ₃	47	Out: 27349 In: 729	0.9 77	0.78 72	0.16 36	0.001 15	0 9.5	0 4.6

Table 1: Averaged line reduction efficiency for the 6 target gases.