Technical Note on:

High level algorithm definition and physical and mathematical optimisations

8 August 1997

Issue 2
Revision -
(Compliant with ORM Vers. 2.3)

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	FMA
M. Ridolfi	IROE-CNR

Approved by:

Name	Institute
B. Carli	IROE-CNR

TABLE OF CONTENTS

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 five high priority species (O_3, H_2O, HNO_3, CH_4 and N_2O) will be routinely retrieved in near real time (NRT). The retrieval of these parameters from calibrated spectra (Level 1b data products) 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.

The study "Development of an Optimised Algorithm for Routine P, T and VMR Retrievals from MIPAS Limb Emission Spectra" is meant to develop a scientific code for NRT Level 2 analysis, suitable for implementation in ENVISAT PDS and optimised for the requirements of speed and accuracy. The results of the study will be 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 on the physical and mathematical optimisations that are relevant for the implementation of the algorithm.

2 - Objectives of the technical note

Objectives of the technical note are:
- summary of the equations of Level 2 code that define the output data as a function of the input data,
- identification of the options which exist for the optimisation of the code,
- assessment of advantages and disadvantages of the different options,
- choice of preferred option for implementation and validation of the choice.

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 a certain level of maturity of the choices has been reached. We must be aware, however, that some choices have been made on the basis of incomplete information and some feedback between testing and code optimization may still be possible.
The acceptance criterion of the code is 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 of spectra simulated with the OFM and the RFM, with and without measurement noise, will 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 will 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.
However, we expect that scientists may wish to review these requirements in the light of the overall error budget (not yet determined in absence of a retrieval study).
An acceptance test on the basis of actual retrieval error will be performed by using the complete OFM / ORM chain as well as the reference spectra generated by the RFM.
The strategy adopted for the preliminary choices of this document 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 once, for the evaluation of the FOV approximation (see later).
- 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/5.

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. This rule is not applied in the case of exponents, inverse (-1) and transpose (T) operations when they can not be confused with 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. In many cases either incomplete or preliminary representation are used since the final distinction between implicit and explicit variables depends on the implementation.

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.

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. 4.4.1), as a function $S = S(x, q_Z)$ of the observation variables 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, the Taylor expansion can be truncated to the first term to obtain:

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

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:

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

where:

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

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

$$y_Z = [q_Z - \tilde{q}_Z]\ . \qquad (4.2.5)$$

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 are the "fitted parameters" and 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:

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

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 $\tilde{\mathbf{q}}_Z$ in order to obtain its correct value $\mathbf{q}_Z$. 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:

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

and matrix $\mathbf{D}$ is equal to:

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

Apex T denotes the transpose and the apex -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:

$$\mathbf{y} = \mathbf{D}\,\mathbf{n} \tag{4.2.9}$$

and the new estimate of the parameters as:

$$\mathbf{q}'_z = \tilde{\mathbf{q}}_z + \mathbf{y} \tag{4.2.10}$$

The errors associated with the solution are characterised by the variance-covariance matrix $(\mathbf{V}^q)$ of $\mathbf{q}'_z$ given by:

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

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 q_Z computed by Eq. (4.2.10) represents a better estimate of the parameters with respect to

$\tilde{q}_z$. 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 **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 sub-section. This modification permits a faster convergence specially in the case of non-linear problems.
Let's define the matrix **A**:

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

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

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

The algorithm proceeds then as follows:
1. the χ^2 function and the matrix **A** are calculated for the initial values of the parameters,
2. λ is initialised to an initial "small" value (e.g. 0.001) and **A** is modified with Eq. (4.2.13)
3. the new estimate of the parameters is calculated for the current choice of λ using equation (4.2.9).
4. the new value of χ^2 is calculated using equation (4.2.7),
5. if χ^2 calculated at step 4 is greater than that calculated at step 1, then λ is increased by a factor 10 and the above steps are repeated starting from step 3 (micro iteration),
6. if χ^2 calculated at step 4 is smaller than that calculated at step 1, then λ is decreased by a factor 10, the new set of parameters is adopted for computing a new matrix **A** and the above steps are repeated starting from 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 to deal with a 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 some conditions which may 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_j i.e.:

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

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.:

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

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, a similar but alternative formula which could be considered is:

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

where $\sqrt{\left(\mathbf{V}_{iter}^{q}\right)_{j,j}}$ represents the error associated with the parameter $\left(\mathbf{q}_z^{iter}\right)_j$ at iteration *iter*. The reason which discourages the use of (4.2.15bis) instead of (4.2.15) is that $\sqrt{\left(\mathbf{V}_{iter}^{q}\right)_{j,j}}$ does not represent the error on the parameter $\left(\mathbf{q}_z^{iter}\right)_j$ 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^2_{LIN}) is less than a fixed threshold t_3:

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

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

$$\left((\mathbf{1}-\mathbf{KD})\mathbf{n}\right)^T\left(\mathbf{V}^n\right)^{-1}\left((\mathbf{1}-\mathbf{KD})\mathbf{n}\right) \qquad (4.2.17)$$

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

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

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:

$$\mathbf{q} = (\mathbf{V}^{-1}_{q} + \mathbf{V}^{-1}_{a})^{-1}(\mathbf{V}^{-1}_{q}(\mathbf{D}\,\mathbf{n} + \tilde{\mathbf{q}}) + \mathbf{V}^{-1}_{a}\mathbf{q}^a) \qquad (4.2.19)$$

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:

$$\mathbf{q} = (\mathbf{K}^T\mathbf{V}^{-1}_{n}\mathbf{K} + \mathbf{V}^{-1}_{a})^{-1}(\mathbf{K}^T\mathbf{V}^{-1}_{n}\mathbf{n} + \mathbf{V}^{-1}_{q}\tilde{\mathbf{q}} + \mathbf{V}^{-1}_{a}\mathbf{q}^a) \qquad (4.2.20)$$

This is the so called "optimal estimation" formula.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 $\mathbf{q}^a$ should be included in the vector of the observations for the computation of χ^2Eq. 4.2.7). 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 faster and more stable.

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 Line Of Sight (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 contiguous tangent altitudes 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:

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

where $\Delta\mathbf{z}_{ig}$ 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 **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:

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

Therefore, if we define matrices **A**, **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 **K**, $\mathbf{K}_1$, **A**, **B** and $\mathbf{B}_1$, then, since Marquardt's algorithm is used, matrix **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^2_{LIN} is:

$$\chi^2_{LIN} = \left[\mathbf{n} - \mathbf{K}\mathbf{y}\right]^T \left[\mathbf{V}^n\right]^{-1} \left[\mathbf{n} - \mathbf{K}\mathbf{y}\right] + \\ + \left[\mathbf{n}_1 - \mathbf{K}_1\mathbf{y}\right]^T \left[\mathbf{V}^z\right]^{-1} \left[\mathbf{n}_1 - \mathbf{K}_1\mathbf{y}\right] \tag{4.2.28}$$

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

4.2.6.1 Calculation of the jacobian matrix K_1 of the engineering tangent altitudes

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

$$\Delta\mathbf{z} = \Delta\mathbf{z}_{tg} + \mathbf{K}_1\mathbf{y} \tag{4.2.29}$$

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

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

where I^{top} is the total number of fitted parameters in the current retrieval.
Now, being **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 $i=1, .., N^{sw}-1$ and $j=2*N^{sw}+1, \ldots, I^{top}$.
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:

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

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

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

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 $M/R = 3483.676$ [$s^2 * K / m^2$].
The jacobian matrix $\mathbf{J}_1$ associated with the transformation (4.2.31) is a $(N^{sw}-1;\ 2N^{sw})$ matrix containing the derivatives:

$$\mathbf{J}_1(i,j) = \frac{\partial \Delta z_i}{\partial P_j} \qquad \text{for } i = 1, \ldots, N^{sw} - 1 \text{ and } j = 1, \ldots, N^{sw}$$

$$\mathbf{J}_1(i,j) = \frac{\partial \Delta z_i}{\partial T_{j-N^{sw}}} \qquad \text{for } i = 1, \ldots, N^{sw} - 1 \text{ and } j = N^{sw} + 1, \ldots, 2N^{sw} \tag{4.2.33}$$

and therefore, deriving equations (4.2.31) we obtain:

$$\mathbf{J}_1(i,j) = -\frac{\mathrm{T}_{i+1} + \mathrm{T}_i}{2\gamma}\left[\frac{1}{\mathrm{P}_{i+1}}\delta_{j=i+1} - \frac{1}{\mathrm{P}_i}\delta_{j=i}\right]$$

$$\text{for } i = 1, \ldots N^{sw}\text{-}1 \text{ and } j = 1, \ldots N^{sw} \tag{4.2.34}$$

$$\mathbf{J}_1(i,j) = -\frac{1}{2\gamma}\ln\left(\frac{\mathrm{P}_{i+1}}{\mathrm{P}_i}\right)\left[\delta_{j-N^{sw}=i+1} + \delta_{j-N^{sw}=i}\right]$$

$$\text{for } i = 1, \ldots N^{sw}\text{-}1 \text{ and } j = N^{sw}+1, \ldots 2N^{sw}$$

where the function δ is defined as:

$$\delta_{\text{arg}} = \begin{cases} 1 \text{ if [arg] = TRUE} \\ 0 \text{ if [arg] = FALSE} \end{cases} \tag{4.2.35}$$

Considering that the original vector of the unknowns of p,T retrieval contains also continuum and offset parameters, matrix $\mathbf{K}_i$ can be obtained by extending matrix $\mathbf{J}_i$ with as many columns as required to reach the dimension (N^{sw}-1; 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.

4.3 The global fit analysis

In global fit the whole altitude profile is retrieved from simultaneous analysis of all the limb-scanning measurements without any partitioning of either the measurements or the unknowns. 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 **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 ideal instrument with infinitesimal field of view (FOV), infinitesimal spectral resolution.
2. the convolution of this spectrum with the apodised instrument line shape (AILS) to obtain the apodised spectrum.
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 assumed to be 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:

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

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(σ,T) is the Planck function:

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

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:

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

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

and the weighted absorption cross section:

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

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:

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

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 line of sight (LOS) and
- the calculation of the absorption cross sections $k_m(\sigma,x_g)$

Ray tracing

The LOS in the atmosphere is determined by the viewing direction of the instrument and the refractive index of the atmosphere. The refractive index $n(p(x_g),T(x_g))$ is a function of pressure and temperature (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:

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

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:

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

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 $A^V_{m,l}(\sigma-\sigma_{m,l},T,p)$ equal to the convolution of the Doppler $A^D_{m,l}(\sigma-\sigma_{m,l},T)$) and the Lorentz profile $A^L_{m,l}(\sigma-\sigma_{m,l},T,p)$:

$$A^V_{m,l}(\sigma-\sigma_{m,l},T,p)=A^D_{m,l}(\sigma-\sigma_{m,l},T)*A^L_{m,l}(\sigma-\sigma_{m,l},T,p) \qquad (4.4.8)$$

The Doppler profile is given by the formula

$$A^D_{m,l}(\sigma-\sigma_{m,l},T)=\sqrt{\frac{\ln 2}{\pi}}\frac{1}{\alpha^D_{m,l}}\exp\left[-\ln 2\frac{(\sigma-\sigma_{m,l})^2}{\alpha^{D\ 2}_{m,l}}\right] \qquad (4.4.9)$$

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

$$\alpha_{m,l}^{D} = \sigma_{m,l}\sqrt{2\ln 2\frac{k^{B}T}{M_{m}c^{2}}} \tag{4.4.10}$$

where M_m = molecular mass of species m

The Lorentz function is:

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

and the Lorentz HWHM:

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

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:

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

and

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

the Voigt function can be rewritten as:

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

with

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

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:

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

$AILS(\sigma)$ 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 the case of satellite measurements, like MIPAS, a linear relationship exists between viewing angle and tangent altitude and the FOV can be defined by a distribution as a function of altitude.
$FOV(z_g,z)$ describes the finite FOV of MIPAS as a function of the altitude z. In the case of MIPAS, $FOV(z_g,z)$ is a trapezium defined by the length of its greater base and the half difference between the two bases (parameters). It is centred around the tangent altitude of the geometry. For the simulation of the spectrum affected by the finite FOV ($S^F(\sigma,z_g)$) the following convolution is calculated:

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

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, but not simulated in the self standing forward model.

4.4.5 Summary of required variables

For the atmospheric model:

- pressure along the line of sight $p(x_g)$
- temperature along the line of sight $T(x_g)$
- volume mixing ratio along the line of sight $X(x_g)$

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 $\sigma_{m,l}$
- reference line strength of transition l of species m $L_{m,l}(T^0)$
- lower state energy of transition l of species m $E''_{m,l}$
- total internal partition function of species m $Q_m(T)$
- molecular mass of species m M_m
- reference Lorentz half width of transition l of species m $\alpha^{L_0}_{m,l}$
- coefficient of temperature dependence of the half width $\gamma_{m,l}$

For the AILS convolution:
- apodised instrumental 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

Even if the points of the interferograms measured by MIPAS are sampled independently on each other (no correlation between the measurements), the spectral data are affected by correlation. The correlation arises essentially from apodisation process performed on the spectrum.
For this reason the noise levels provided by Level 1b processing do not fully characterise the error on the measurements and the computation of the complete Variance Covariance Matrix $\mathbf{V}^S$ of the spectrum $S(\sigma)$ is needed.
The inputs required for the calculation of $\mathbf{V}^S$ are:
1. the frequency spacing between spectral data points $\Delta\sigma$,
2. the spectral interval of each microwindow,
3. the Noise Equivalent Spectral Radiance for each microwindow $\delta(\mu W)$,
4. the apodisation Function $\theta(\sigma)$,
5. Maximum path difference.

Below we describe how the variance covariance matrix $\mathbf{V}^S$ of the observations can be derived starting from the above input quantities.
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-filling path difference (ZFPD).
The measured interferogram is therefore equal to the interferogram with path difference ZFPD, multiplied by a boxcar function $\Pi(d)$ defined as:

$$\Pi^{MPD}(d) = \begin{cases} 1 & \text{when } d \in [-\text{MPD}, \text{MPD}] \\ 0 & \text{when } d \notin [-\text{MPD}, \text{MPD}] \end{cases} \quad (\text{MPD} = \text{Max. Path Difference}) \qquad (4.5.1)$$

2. **Fourier Transformation (FT)**: Let us call $\mathbf{S}^{HR}(\sigma)$ the spectrum obtained from the interferogram with path difference ZFPD and $\mathbf{S}^{NAHR}(\sigma)$ the spectrum obtained from the Fourier transformation of the measured (and zero-filled) interferogram. Since the FT of the product of two functions is equal to the convolution of the FT's of the two functions, we obtain:

$$S^{NAHR}(\sigma) = S^{HR}(\sigma) * \text{FT}\,[\Pi^{MPD}(d)]\,. \qquad (4.5.2)$$

$S^{NAHR}(\sigma)$, $\mathbf{S}^{HR}(\sigma)$ and FT $[\Pi^{MPD}(d)]$ are all given in the sampling grid

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

3. **Interpolation to the fixed grid**: since a constant grid is required, the spectrum is interpolated to a fixed grid of $\Delta\sigma = 0.025$ cm^{-1} $= \frac{1}{2 \cdot D}$, with D equal to 20 cm. The performed operation is:

$$S^{NA}(\sigma) = S^{NAHR} * FT\left[\Pi^{D}(d)\right], \tag{4.5.3}$$

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

In this case the convolution provides also the interpolation process which changes the grid spacing from $\frac{1}{2 \cdot ZFPD}$ to $\frac{1}{2 \cdot D}$.

This operation does not introduce correlation between the spectral points as long as MPD ≥ D.

If MPD ≥ D the result of (4.5.3) is equal to the FT of a ±20 cm interferogram.

If MPD < D the result of (4.5.3) 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 interpolation step.

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

$$S(\sigma) = S^{NA}(\sigma) * \theta(\sigma) \tag{4.5.4}$$

If MPD ≥ D, the variance covariance matrix $\mathbf{V}^{S}$ of the apodised spectrum $\mathbf{S}(\sigma)$ can be computed from the VCM $\mathbf{V}^{NA}$ associated with $S^{NA}(\sigma)$, (this is a diagonal matrix since the spectral points of $S^{NA}(\sigma)$ are uncorrelated) and the Jacobian of the convolution.

The convolution process introduces correlation and the variance covariance matrix $\mathbf{V}^{NA}$ is modified accordingly to:

$$\mathbf{V}^{S} = \mathbf{J} \cdot \mathbf{V}^{NA} \mathbf{J}^{T} = \mathbf{V}^{NA} \mathbf{J}\mathbf{J}^{T} \tag{4.5.5}$$

where $\mathbf{J}$ is the Jacobian matrix associated with the convolution (4.5.4). The order of the operations can be changed because $\mathbf{V}^{NA}$ is a diagonal matrix and all the diagonal elements are equal (see later). Now, since convolution (4.5.4) is a linear operation equal to:

$$S(\sigma_i) = \sum_j S^{NA}(\sigma_j) \cdot \theta(\sigma_i - \sigma_j) \tag{4.5.6}$$

the entry i,k of matrix $\mathbf{J}$ is given by:

$$\mathbf{J}_{i,k} = \theta(\sigma_i - \sigma_k) \tag{4.5.7}$$

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

$$\mathbf{V}^S_{i,j} = \delta_i \, \delta_j \cdot \sum_k \theta(\sigma_i - \sigma_k) \cdot \theta(\sigma_j - \sigma_k) \tag{4.5.8}$$

δ_i is the NESR calculated in Level 1b after step 3 (see above).

If MPD < D , the problem is more complicated and also the effect of zero-filling must be taken into account. Furthermore, since zero-filling introduces a correlation between the spectral points of $S^{NA}(\sigma)$ (eq. 4.5.3), its error is characterised by a VCM $\boldsymbol{V}^{NA}$ (with off-diagonal elements different from zero) of which the NESR provides only the measurement of the diagonal.
This problem can be overcome considering that, when MPD < D, the apodised spectrum is given by:

$$S(\sigma) = S^{NAD}(\sigma) * \theta'(\sigma), \tag{4.5.9}$$

where $S^{NAD}(\sigma)$ is the non-apodised spectrum which would be obtained if MPD was equal to D, $\theta'(\sigma)$ is the Fourier Transform of the apodisation function modified according to:

$$\theta'(d) = \begin{cases} \theta(d) & \text{when d} \in [-\text{MPD}, \text{MPD}] \\ 0 & \text{when d} \notin [-\text{MPD}, \text{MPD}] \end{cases}. \tag{4.5.10}$$

In order to exploit Eqn. (4.5.9) it is necessary to calculate the VCM associated with $S^{NAD}(\sigma)$ (characterised by having off-diagonal elements equal to zero), knowing the diagonal elements of the VCM associated with $S^{NA}(\sigma)$ (equal to NESR). The relationship between $S^{NAD}(\sigma)$ and $S^{NA}(\sigma)$ is:

$$S^{NA}(\sigma) = S^{NAD}(\sigma) * FT\left[\Pi^{MPD}(d)\right] \tag{4.5.11}$$

The VCM of the observations related to $S^{NA}(\sigma)$, $\mathbf{V}^{NA}$, is connected with the VCM related to $S^{NAD}(\sigma)$, $\mathbf{V}^{NAD}$, by the following equation:

$$\boldsymbol{V}^{NA} = \boldsymbol{F} \cdot \boldsymbol{V}^{NAD} \cdot \boldsymbol{F}^T = \boldsymbol{F} \cdot x \cdot \boldsymbol{I} \cdot \boldsymbol{F}^T = x \cdot \boldsymbol{F} \cdot \boldsymbol{F}^T, \tag{4.5.12}$$

where $\boldsymbol{F}$ is the Jacobian matrix associated with the convolution by $f(\sigma) = FT\left[\Pi^{MPD}(d)\right]$, $\boldsymbol{I}$ is the identity matrix and x the diagonal element of $\boldsymbol{V}^{NAD}$ which has to be determined (it is assumed that the spectral data in the same microwindow are affected by the same error, even if different microwindows can have different errors).
The diagonal elements of $\mathbf{V}^{NA}$ which are equal to $(\text{NESR})^2$ from Eqn. (4.5.12), are given by:

$$\left(\mathbf{V}^{NA}\right)_{ii} = \left|f(\sigma)\right|^2 \cdot x; \tag{4.5.13}$$

using the Parseval theorem we get:

$$\left|f(\sigma)\right|^2 = \frac{1}{D}\int_0^D \left[\Pi^{MPD}(d')\right]\cdot\left[\Pi^{MPD}(d')\right]dd' = \frac{MPD}{D}, \tag{4.5.14}$$

and hence:

$$x = \frac{D}{MPD}\cdot(NESR)^2 \tag{4.5.15}$$

Finally the matrix $\mathbf{V}^S$ is given by the following expression:

$$\mathbf{V}^S = \frac{D}{MPD}\cdot(NESR)^2\cdot(\mathbf{J}'\cdot\mathbf{J}') \tag{4.5.16}$$

The result is that, when MPD < D, the value of NESR is scaled by the factor $\frac{D}{MPD}$ and the product of $(\boldsymbol{J}'\cdot\boldsymbol{J}')$, instead of $(\boldsymbol{J}\cdot\boldsymbol{J})$, is used in order to calculate $\boldsymbol{V}^S$. Besides, in this case $\mathbf{V}^{S^{-1}}$ is a singular matrix.

Since in the case of MIPAS data it is reasonable to assume that the spectral data points belonging to different microwindows are uncorrelated, the matrix $\mathbf{V}^S$ can be decomposed in as many submatrices as many microwindows are processed, and the dimension of the submatrices is equal to the number of spectral data in each microwindow.
This assumption allows to save both computing time (the summation which appears in equation (4.5.9) can be computed only once, for the largest considered microwindow) and memory (it is not necessary to store the complete matrix $\mathbf{V}^S$).
The variance covariance matrix of the spectra can be combined with the variance covariance matrix of the forward model $\mathbf{V}^{FM}$ for the calculation of the variance covariance matrix $\mathbf{V}^n$ of the quantities (4.2.3):

$$\mathbf{V}^n = \mathbf{V}^S + \mathbf{V}^{FM} \tag{4.5.17}$$

We assume for the moment that $\mathbf{V}^{FM}$ can be neglected and $\mathbf{V}^S$ provides an adequate estimate of $\mathbf{V}^n$. This assumption must be validated as part of subsequent sensitivity tests on systematic errors. The use of an incorrect estimate of $\mathbf{V}^n$ can be easily identified from the deviation of the quantity:

$$\frac{\chi^2}{n-m} \tag{4.5.18}$$

from unity (χ^2- test).

4.6 Calculation of the Jacobian matrix **K** *of the simulations*

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

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

where S_i is the simulated spectrum, and q_j is the *j-th* component of the vector **q** containing the unknowns of the problem. The vector **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 to be constant varying the observation geometry and a 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 to be constant varying the observation geometry and a 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 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 can be used 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 **C** a symmetric matrix of dimension 'r'. It is possible to find a base of 'r' independent eigenvectors of **C**. If **V** is the matrix whose columns are the eigenvectors of **C**, matrix **C** can then be written in the form:

$$\mathbf{C} = \mathbf{V}\mathbf{w}\mathbf{V}^T \tag{4.7.1}$$

where **w** is a diagonal matrix containing the eigenvalues of **C**. The inverse matrix of **C** is then:

$$\mathbf{C}^{-1} = \mathbf{V}\mathbf{w}^{-1}\mathbf{V}^T \tag{4.7.2}$$

The appearance of singularities in **C** is detected by the presence of eigenvalues close to zero in **w**. In this case the singularities can be eliminated by imposing $1/w_i = 0$ whenever $w_i \approx 0$.
It can be demonstrated that this procedure corresponds to the calculation of the generalised inverse matrix of **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:

$$\delta z_i = z_i^{RET} - z_i^{ENG} \tag{4.8.1}$$

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 $\delta \mathbf{z}$ is intended as the error which should be attributed to the retrieved tangent altitude when this quantity is reconstructed by adding the correction $\delta \mathbf{z}$ 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,..., N^{SW}-1) can be expressed as:

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

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 altitudes corrections can be expressed as a function of T_j and P_j by substituting equation (4.8.2) in (4.8.1):

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

where Δz_j^{ENG} are defined as:

$$\Delta z_j^{ENG} = z_{j-1}^{ENG} - z_j^{ENG} \qquad j=2, ...,N^{SW} \tag{4.8.4}$$

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:

$$\mathbf{V}^{HC} = \mathbf{K}_D \mathbf{V}^q \mathbf{K}_D^T \tag{4.8.5}$$

where $\mathbf{K}_D$ is the jacobian matrix which connects δz_i with T_j and P_j. The elements of $\mathbf{K}_D$ are the derivatives:

$$K_D(i,j) = \frac{\partial(\delta z_i)}{\partial P_j} \qquad \text{with } i=1, ...,N^{SW}-1 \text{ and } j=1, ..., N^{SW}$$

and (4.8.6)

$$K_D(i,N^{SW}+j) = \frac{\partial(\delta z_i)}{\partial T_j} \qquad \text{with } i=1, ...,N^{SW}-1 \text{ and } j=1, ..., N^{SW}$$

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

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

and (4.8.7)

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

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

1. Calculation of the jacobian matrix $\mathbf{K}_D$ by using equations (4.8.7),
2. 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 will be 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 refraction 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.

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 directly VMR 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 sampling if the representation is made at the sampling points
- the vertical resolution of the measurement cannot be equal to the vertical resolution of the sampling if the representation is made at intermediate levels.

Therefore, if the offset 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 conclusion, 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 an 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, but the vertical resolution of the measurement depends upon the offset between wanted and implemented tangent altitudes. 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 constant vertical resolution, but the noise depends upon the offset between wanted and implemented tangent altitudes. A very large increase of noise could 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 is presently assumed for the development of Level 2 scientific code:

- we adopt as a baseline for the code that is being developed for level 2 data analysis the option of the retrieval at tangent altitudes levels followed by interpolation. The reason for this choice is that we cannot afford the unknowns of the other option.
- we notice that the scientific requirement for a 3 km vertical resolution of the measurement has been used to choose a vertical resolution of the sounding equal to 3 km. Since these two vertical resolutions can differ significantly when a retrieval at fixed levels is required, either a reconsideration of the limb scanning sequences or a new definition of the vertical resolution requirement is necessary.
- Some flexibility is maintained in order not to prevent the implementation of the alternative approach in subsequent versions of the code in case it is recommended by retrieval studies.

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 described.
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:

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

with an error:

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

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 **x**. The weighted combination $\boldsymbol{x}^c$ of the two measurements $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ having respectively the variance-covariance matrices $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ is equal to:

$$\boldsymbol{x}^c = \left(\boldsymbol{V}_1^{-1} + \boldsymbol{V}_2^{-1}\right)^{-1}\left(\boldsymbol{V}_1^{-1}\boldsymbol{x}_1 + \boldsymbol{V}_2^{-1}\boldsymbol{x}_2\right) \qquad (5.2.3)$$

and has a variance-covariance matrix equal to

$$V^c = \left(V_1^{-1} + V_2^{-1}\right)^{-1} \qquad (5.2.4)$$

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. Consequently, 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:

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

where the notations are:
p_n pressure at a given altitude,
p_o 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:

1. 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.
2. 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.

5.3 Latitude effects

5.3.1 Latitude spread

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

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

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. In order to quantify this statement, the integrated columns as a function of the latitude spread are shown in Figure 5.1 for tangent altitudes 10 km and 17 km. This figure shows that According to that the 90% of the emitting gas is concentrated within 4°, for a species with constant VMR.

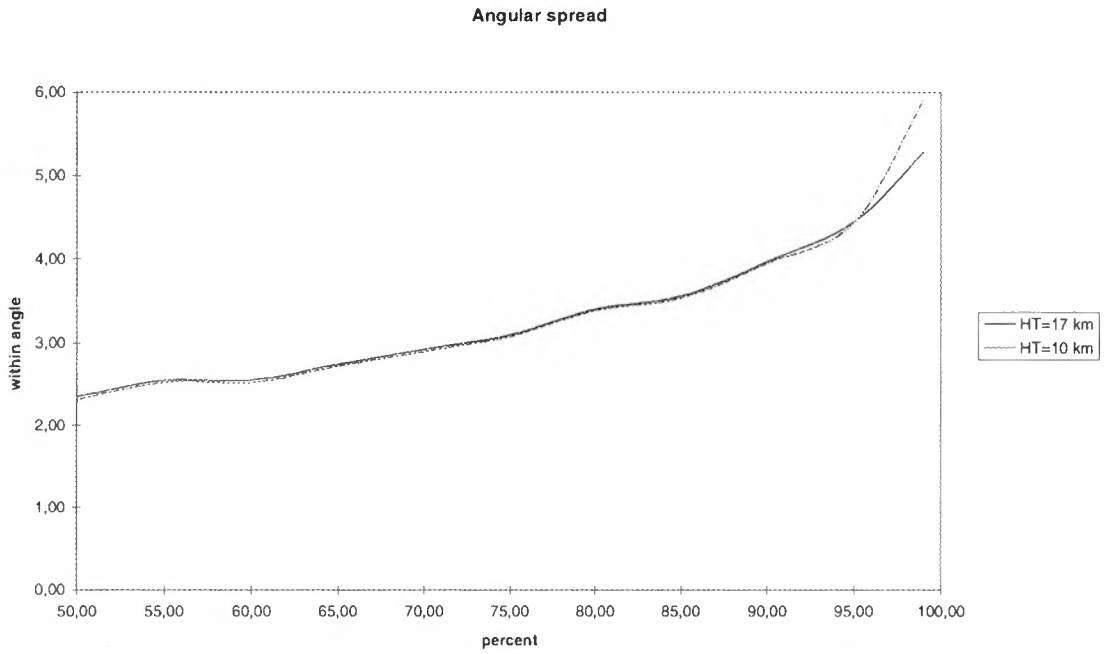


Figure 5.1: Angle [°] at the earth centre over the percentage of the integrated column for two different tangent altitudes HT.

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).)

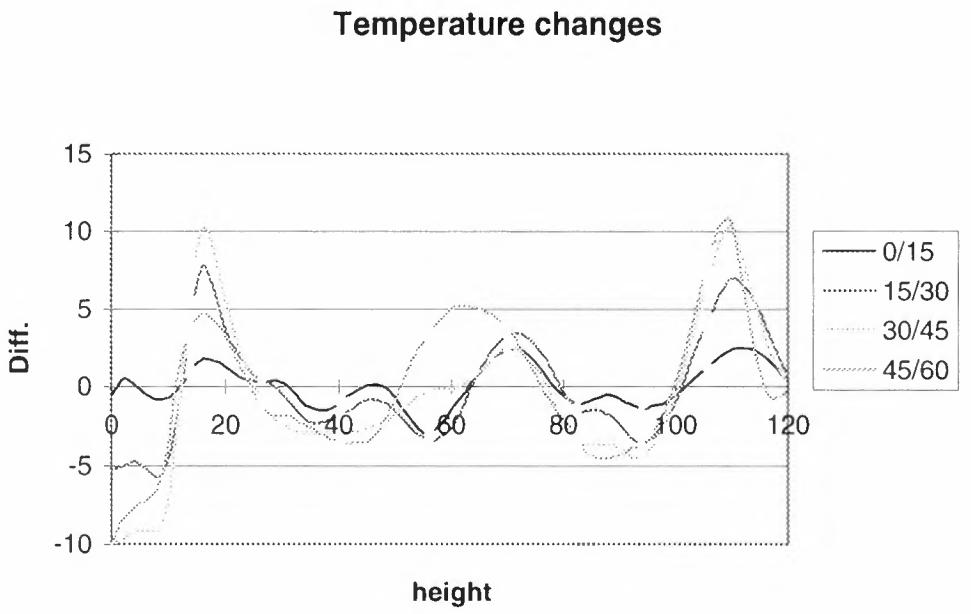


Figure 5.2: Temperature differences [K] as a function of the altitude [km] between latitudes separated by 15°.

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 for a latitude spread of 4 degrees (See Sect. 5.3.1).
Instantaneous latitudinal variations can be larger than these climatological differences, but taking into consideration the large amount of memory and computation time which would be 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. Horizontal gradients will be considered in the off-line data processor.

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 surface of the WGS84 ellipsoid), as a function of geodetic latitude Φ can be computed using the following empirical formula which includes the centrifugal effect:

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

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:

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

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:

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

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:

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

$$R = \sqrt{f^2 \cos^2\Phi + \left(\frac{b^2}{a^2} f \cdot \text{sen}\,\Phi\right)^2} \qquad (5.4.5)$$

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:

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

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 $2*10^{-4}$.

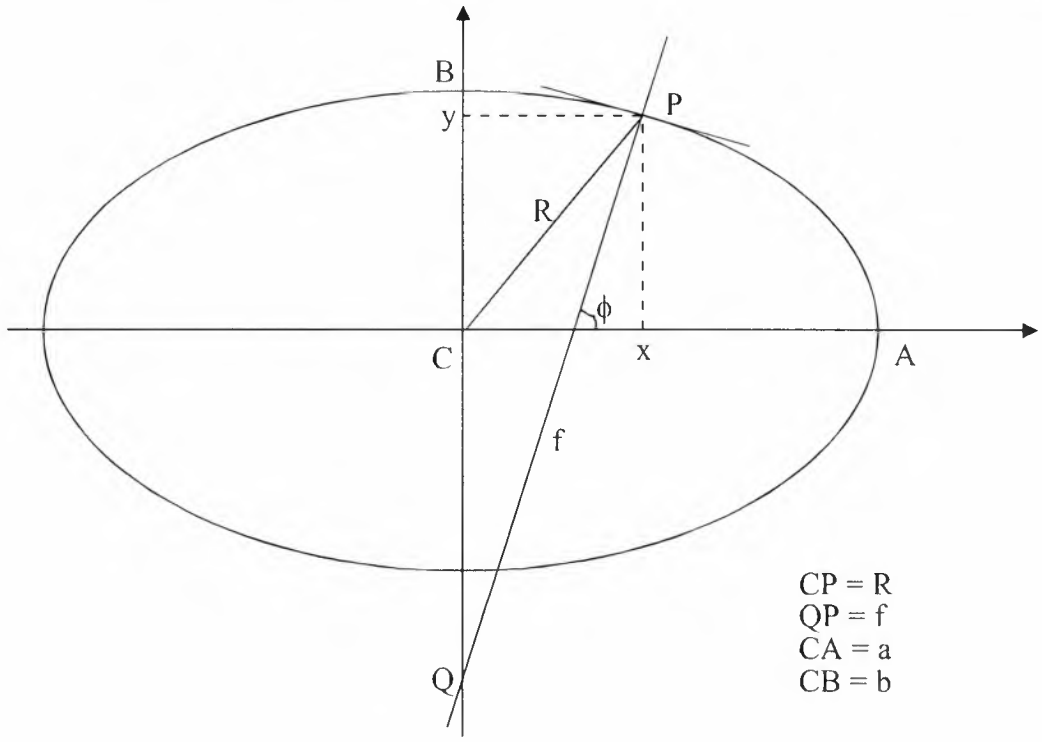


Fig. 5.3: Earth model

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 refraction index of the air.

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

Refraction 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):

$$\frac{n^2-1}{n^2+2} = const \cdot \rho \tag{5.5.1}$$

where n is the refractive index, ρ the air density.
Since $n-1 << 1$, we can write:

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

If we consider the Barrel - Sears' s empirical formula:

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

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 the height of 5 km, where water VMR is equal to $1.4 \cdot 10^{-3}$ for the standard atmosphere).
Besides also the frequency dependence of the refraction index can be neglected in the mid-infrared.
Different models can be used for the determination of refraction index, for instance Edlen (1966) or SAO (*priv. com.*), each of which finds a different value of α.
The effect of refraction on the evaluation of tangent height will be shown below.
Assuming a spherical earth and spherical atmospheric layers, the following relation is valid for all the points on the line of sight:

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

where r is the summation of the local radius of curvature of the earth (which is an input of the scientific code) 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 is the Snell's law in the case of spherical geometry. Considering an atmospheric spherical layer as drawn in the picture, Snell's law can be written in point P as:

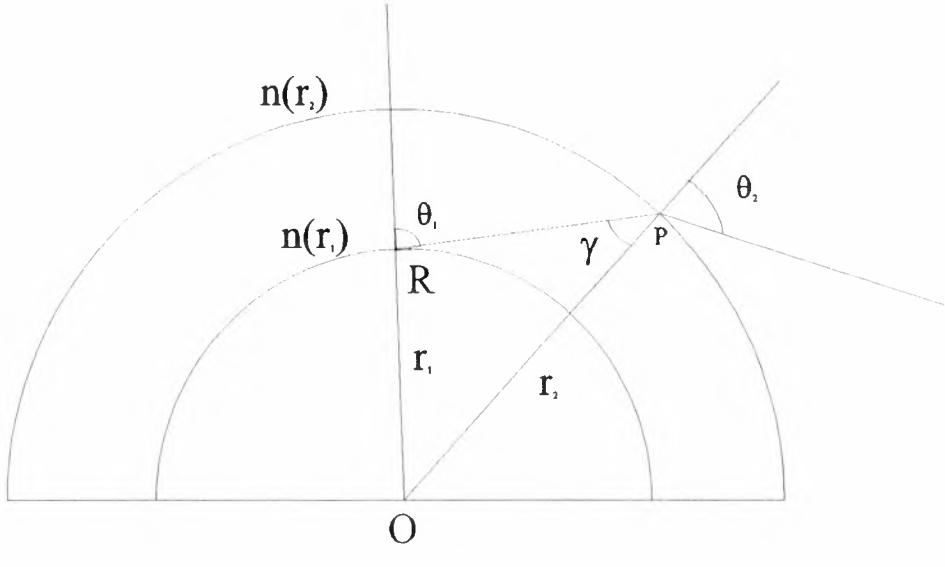


Figure 5.4: Deduction of Snell's law for media with spherical symmetry.

$$n(r_2) \cdot \sin\theta_2 = n(r_1) \cdot \sin\gamma \tag{5.5.5}$$

From the 'sine' theorem applied to triangle OPR we obtain:

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

Combining the two equations we get:

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

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:

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

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, even 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

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

to be:

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

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

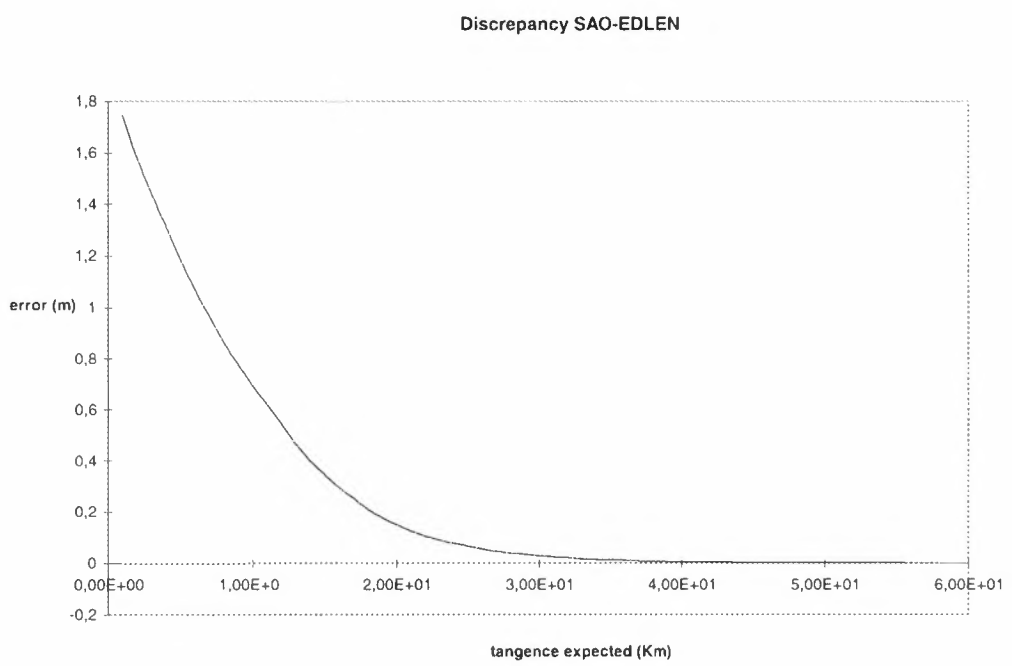


Figure 5.5: Difference in the tangent height between different models for the refraction index (SAO-Edlen).

The differences on the tangent heights found using two models of refraction 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 refraction index is the simplified version of Edlen:

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

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 refraction 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 n. 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:

$$dr = dx \cdot \cos(\theta), \tag{5.5.12}$$

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

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

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

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

is the geometrical distance from the tangent point.
In this case we obtain:

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

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 refraction value

5.6 Line shape modelling

The line shape function which has to be modelled 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 CO_2, or the super- and sub-Lorentzian behaviour of H_2O. 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 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:

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

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 line-shape.

As a baseline Humlicek (1982) is implemented. In future versions of the code further run time improvements can be achieved by implementation of optimised versions of the Humlicek algorithm (e.g. in development at the IMK, M. Kuntz, priv. com.).

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 the two 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 α^D	-1.5%
20 α^D	-0.37%
30 α^D	-0.17%
40 α^D	-0.10%
50 α^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 distance 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 μm. These factors start from unity at the line centre and remain 1 until 0.5-5 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:

$$A(\sigma,\sigma_l)=\frac{1}{\pi}\frac{\alpha_l^L+(\sigma-\sigma_l)pY_l}{\left(\sigma-\sigma_l+\mathrm{Im}\{\Gamma_{ll}\}\right)^2+\alpha_l^{L\,2}} \qquad (5.7.1)$$

with the first order coupling coefficient:

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

Γ is the (frequency dependent) relaxation matrix, with diagonal elements determining the shape of uncoupled lines ($\mathrm{Re}\{\Gamma_{ll}\}=\alpha_l^L$, the Lorentz half widths) and the lineshifts ($-\mathrm{Im}\{\Gamma_{ll}\}$), 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):

$$A_l^{V_{LM}}(\sigma-\sigma_l)=\sqrt{\frac{\ln 2}{\pi}}\frac{1}{\alpha_l^D}\left[K(x_l,y_l)+pY_lL(x_l,y_l)\right] \qquad (5.7.3)$$

The baseline for our code development is not to use Q-branches where line-mixing is known to have strong effects with an appropriate choice of microwindows, so that line mixing does not need to be simulated.
This baseline can be verified during comparisons with the RFM-code.

5.8 Pressure shift

Beside the line mixing effect equation (5.7.1) contains the pressure shift in the form of the term $-\mathrm{Im}\{\Gamma_{ll}\}$ 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 HITRAN92 data base we will not implement it into the forward code.
This baseline can be verified during comparisons with the RFM-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 can be reduced by using an appropriate selection of microwindows.
For the moment the capability of handling non-LTE effects is not implemented in the OFM, but some auxiliary data that could be used for this purpose are maintained in the list of inputs.
This baseline can be verified during comparisons with the RFM-code.

5.10 Self broadening

The Lorentz half width from equation (4.6.12) includes both, foreign $\alpha^{L_0^f}$ and self broadened $\alpha^{L_0^s}$ components:

$$\alpha^{L_0} = (1 - X)\alpha^{L_0^f} + X\alpha^{L_0^s} \tag{5.10.1}$$

With the volume mixing ratio X.

The relative error which is done when neglecting the self broadening is $X\left(\alpha^{L_0^s}/\alpha^{L_0^f} - 1\right)$. 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. $\alpha^{L_0^s}/\alpha^{L_0^f}$	Max. X (8-50 km) [ppmv]	Max. error [%]
CO_2	1.3	360	0.011
O_3	1.3	7	0.0002
H_2O	10 5 (average)	700 (tropical) 200 (mid-latitude)	0.7 0.08 (average/mid-lat.)
CH_4	1.3	1.7	0.00005
N_2O	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. It's 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. This 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 it's 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 microwindow data-base to decide if the line has to be calculated explicitly (option 1) or can be interpolated inside the microwindow (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 H_2O), the pressure broadened bands of O_2 at 1550 cm^{-1} and N_2 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):

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

where

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

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

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

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

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

$C^{0s}(\sigma,T)$ 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 microwindow database which uses different 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×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 microwindow 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.

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 Orlando et al., 1996.

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 implement 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 mW.s; 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.

A more sophisticated continuum constraint is being developed in a parallel retrieval study and could be considered in future versions of the code.

5.12 Interpolation of the profiles

Since the baseline for the Level 2 scientific code is to retrieve the profiles at tangent pressure, an interpolation of the retrieved profiles to user-defined pressure levels may be required. The choice of the most appropriate interpolation rules is herewith briefly discussed.
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 the pressure, the most appropriate interpolation rule in the pressure domain seems to be logarithmic interpolation for both temperature and VMR profiles.
The baseline is therefore to use logarithmic interpolation of temperature and VMR profiles whenever an interpolation to the user-defined pressure grid 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.

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 preliminary tests performed with the RFM. However, the final 'tuning' of these parameters is not analysed in the present document.

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 sections 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. by calculating the single contributions of shorter paths and then summing them.
This implies that atmosphere has to be segmented.
Two main optimisations can be realised:
- the first refers to the kind of segmentation: since spectra corresponding to different lines of sights have to be calculated, a segmentation that uses discontinuities at the same altitudes 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 the faster is the forward model.

In this section we describe two different methods to build 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 values are defined). The level approach is chosen as a baseline. The fact that the segmentation of the radiative transfer integral is made in the altitude domain makes it possible to have some overlap between the altitudes of the radiative transfer layering and those of the discretisation of the unknowns.

_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 thinner 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 refraction 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:

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

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

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, $\eta\big(p(z),T(z)\big)$ is the air number density.

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

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

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. 6.2.1), and a big amount of memory is required.

On the contrary, the use of mean temperature and pressure, that do not depend either on the molecules, or 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 a small 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:

$$\tau_{\sigma,l,g} = \exp\left(-\sum_m \int_{z_{l-1}}^{z_l} k_\sigma\left(p(z),T(z)\right)\cdot X_m(z)\cdot\eta\left(p(z),T(z)\right)\cdot\frac{dx^g}{dz}\cdot dz \right) \approx$$

$$\approx \exp\left(-\sum_m k_{\sigma,l,g}\left(p^e_{l,g},T^e_{l,g}\right)\cdot \int_{z_{l-1}}^{z_l} X_m(z)\cdot\eta\left(p(z),T(z)\right)\cdot\frac{dx^g}{dz}\cdot dz \right) = \qquad (6.1.4)$$

$$= \exp\left(-\sum_m k_{\sigma,l,g}\left(p^e_{l,g},T^e_{l,g}\right)\cdot C_{m,l,g} \right)$$

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

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

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 the 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 occurs 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:

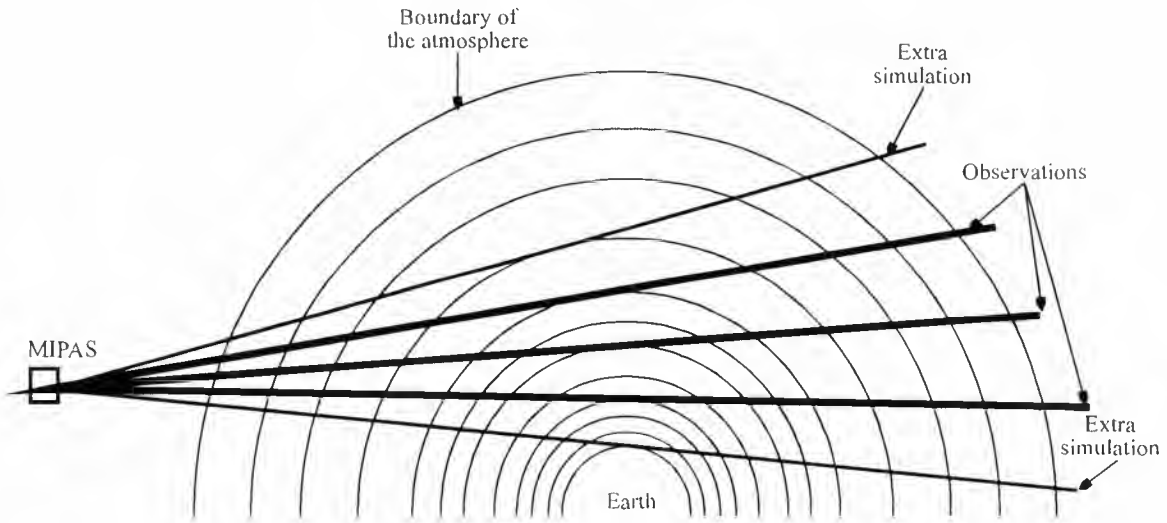
Above the tangent altitude of the highest simulated spectrum a set of levels is determined which divide the atmosphere into layers for the radiative transfer calculation.
Starting from the tangent altitude of the highest simulated spectrum,
(*) a user-defined guess increment Δz is used to build next level, then, conditions 1. and 2. (used at the previous step) are checked and:
$\Rightarrow$ if the two conditions are both satisfied then the guess level is accepted and the algorithm proceeds to (*)
$\Rightarrow$ 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 to 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.
Further tests have to be performed using the microwindows of the five VMR retrievals. In the case in which insufficient accuracy is achieved using a reasonable number of levels, a further constraint, on the maximum variation of the VMR between the levels, can be considered.

	Layering 1 (reference)	Layering 2	Layering 3
Lower threshold (K)	1.5 K	5 K	25
Higher 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^c and T^c 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^c and T^c 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^c and T^c 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^c and T^c have not 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^c - T^c 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 revealed 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 isn't 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.

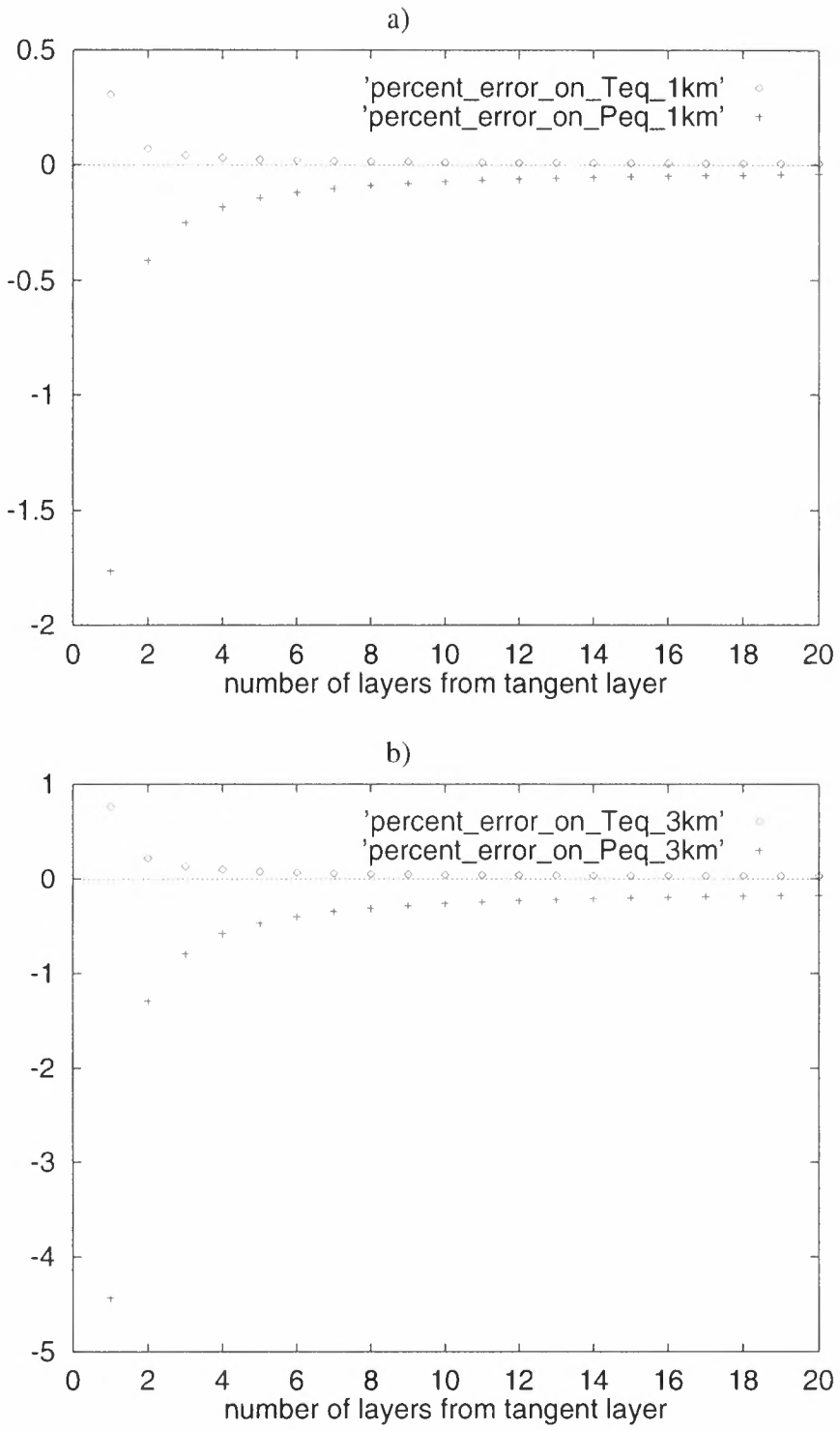


Fig. 6.2 The percentage deviation of equivalent temperature (squares) and equivalent pressure (crosses) from that in the case of vertical penetration are reported for the different layers, starting from the tangent layer. a) Layers thickness is about 1 km; b) Layers thickness is about 3 km. Test performed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

Fig.6.3_**TABLE of PATHS**

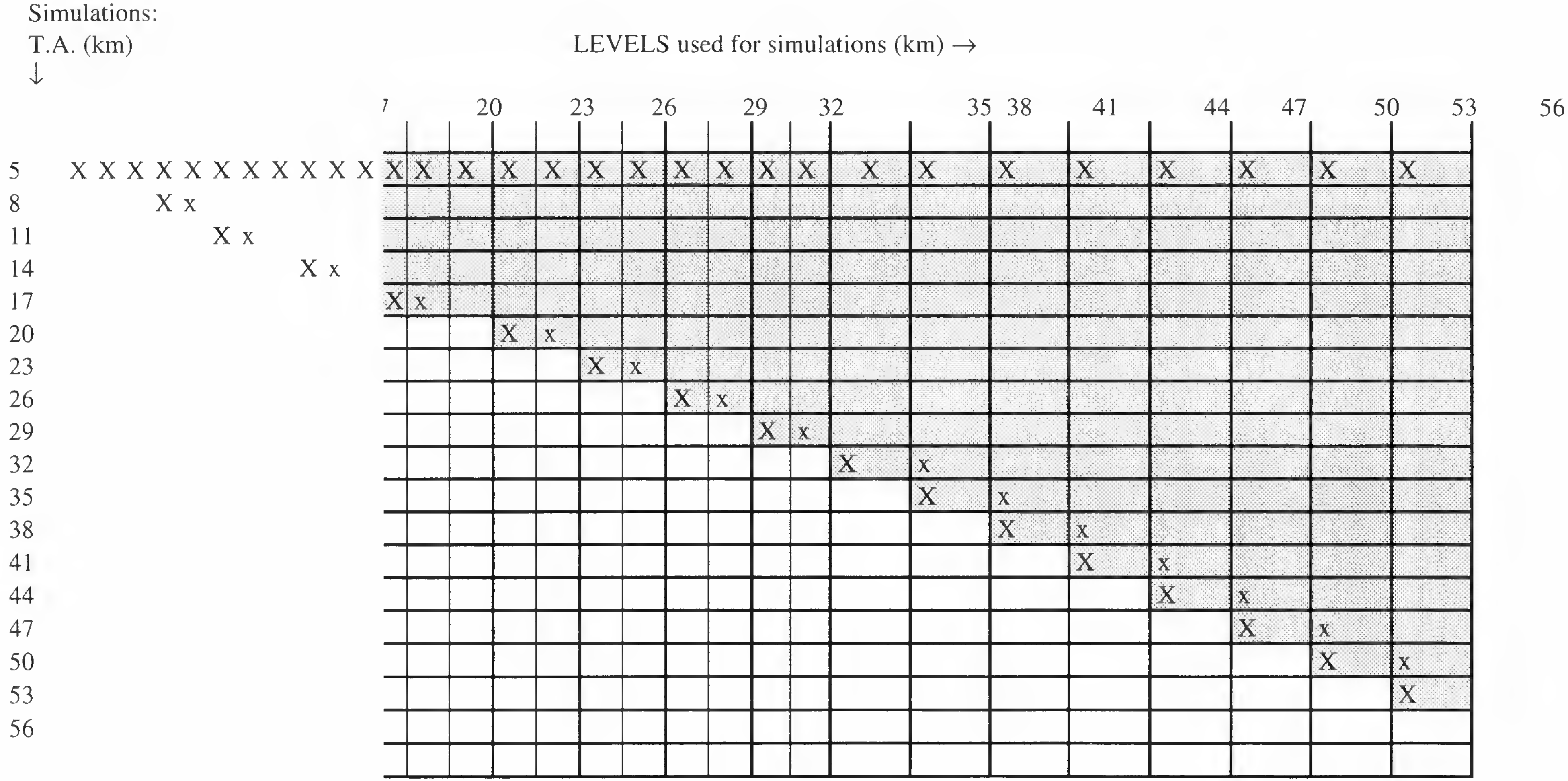


Fig.6.3_**TABLE of PATHS**

Simulations:
T.A. (km)
↓

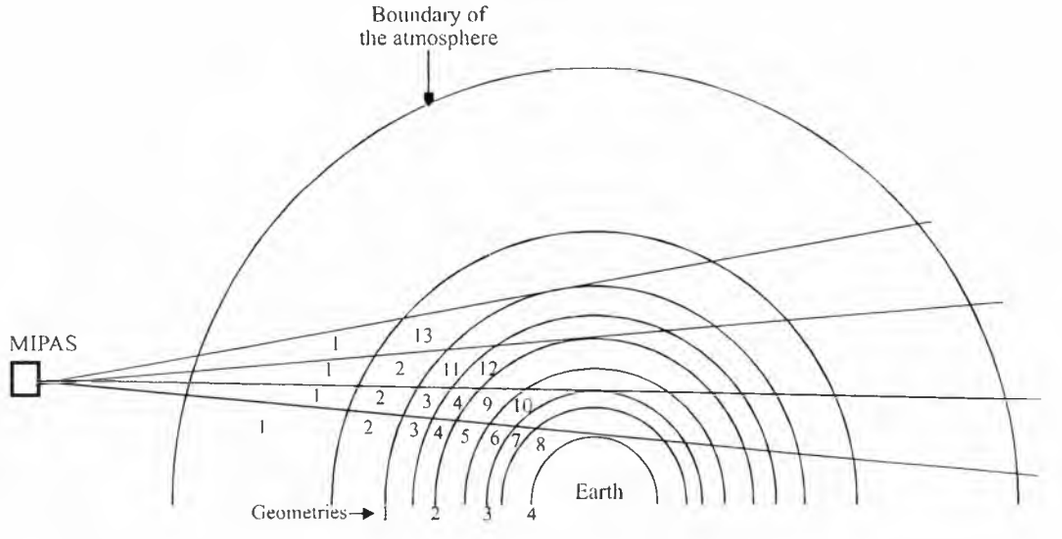
LEVELS used for simulations (km) →

	5			8		11			14			17			20		23		26		29		32		35 38	41	44	47	50	53	56
5	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
8				X	x																										
11						X	x																								
14									X	x																					
17												X	x																		
20															X	x															
23																	X	x													
26																			X	x											
29																					X	x									
32																							X	x							
35																								X	x						
38																									X	x					
41																										X	x				
44																											X	x			
47																												X	x		
50																													X	x	
53																														X	
56																															

6.2.1 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:

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

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 remains nearly constant for the different geometries of a limb scan. As 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 latitudes, 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:

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

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=250K, 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%.

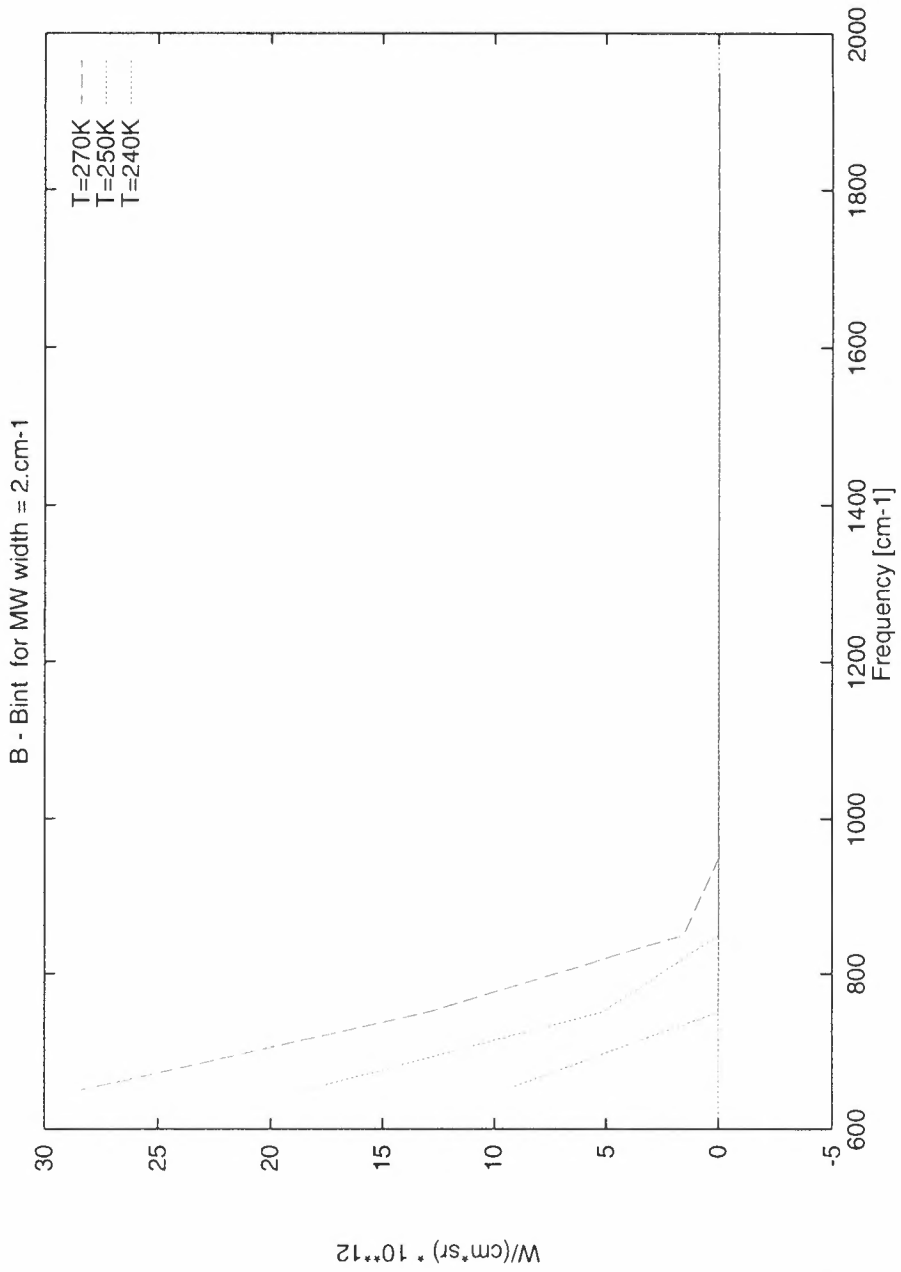


Fig. 6.5: Absolute difference between the theoretical value of Planck function and its linearly interpolated value. A scale factor of 10^{12} is applied.
These differences are completely negligible if compared with the NESR, whose values range from 50 nW/(cm^2*sr*cm^{-1}) in the wave-number region around 800 cm^{-1} to 4.2 nW/(cm^2*sr*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 Delbouille et all (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 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 H_2O and CO_2 bigger than NESR.

The antenna pattern of the field of view provided by ESA is 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.

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):

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

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 line of sight that span the vertical range of 4 km 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 work in the frequency coarse 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 seem to be H_2O and CO_2. H_2O has a strong rate of change in VMR profile, and hence in density profile, in the troposphere, but also CO_2, 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 O_3 and HNO_3, and FOV effect are expected to be less significant.
Tests on CO_2 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 CO_2 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 CO_2 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:

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

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 F.O.V. is given by:

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

The integral can easily be performed in analytical way.

Tests on H_2O have revealed that at low altitudes, up to the boundary between troposphere and stratosphere, the interpolation using 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 H_2O.

μWs ↓ TA →	8 km	11 km	14 km
1H_2O2B	26	17	21
2H_2O3	24	74	21
3H_2O4B	20	42	24
4H_2O5B	30	17	24
5H_2O6B			47
6H_2O33	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 H_2O 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.
However, the final validation of the model for taking into account FOV has been doing using RFM spectra. Work is 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.

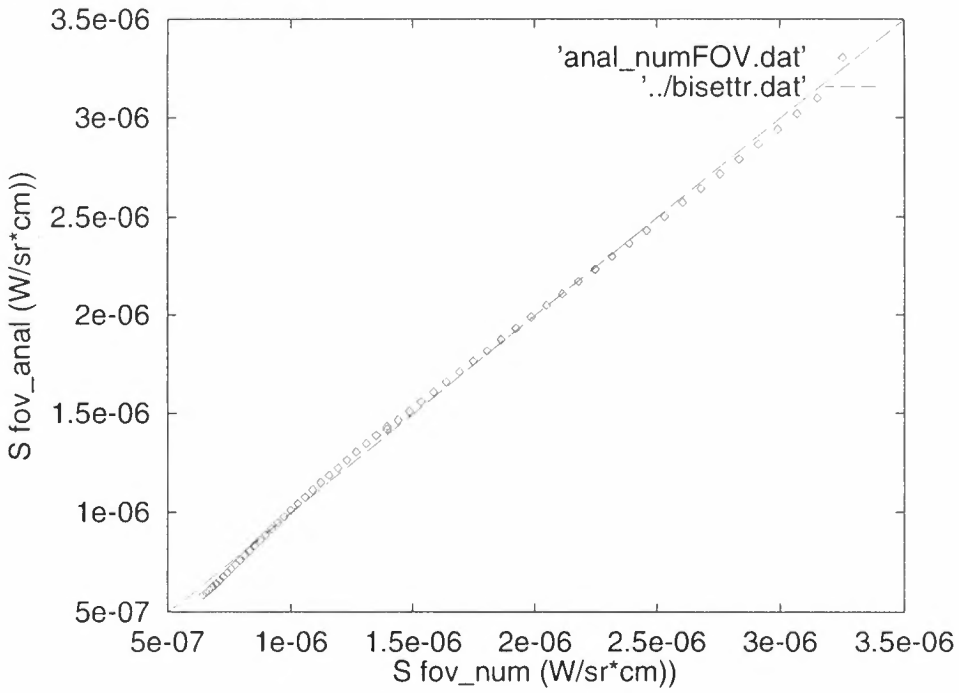


Fig. 6.6 In this plot the values of the spectrum at a significant frequency calculated with a analytical convolution at altitudes between 6.5 km and 12.5 km are plotted as a function of the corresponding values of the spectrum calculated with reference numerical convolution.
A microwindow selected for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

The analytical derivative, obtained by the following calculation:

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

(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 in 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).

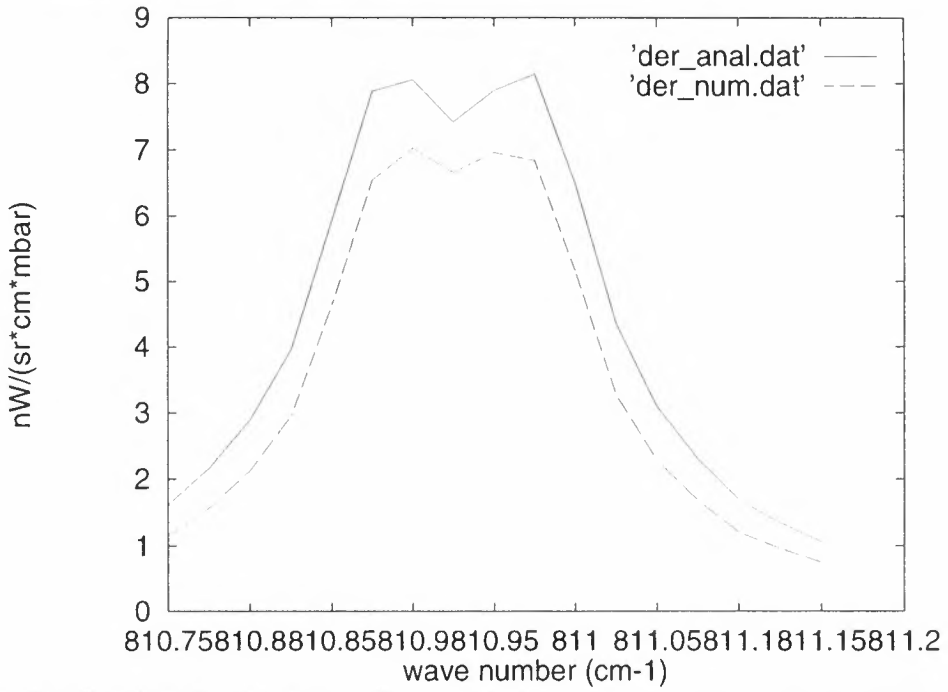


Fig. 6.7 Numerical and analytical derivatives with respect to tangent pressure for a microwindow selected for p-T retrieval, tangent altitude equal to 11 km.

It is not yet clear how an error in the derivatives affects the retrieval results (see discussion on convergence criteria, Sect. 6.8).
Our hope is that this kind of error on the derivatives does not increase substantially the number of iterations necessary for reaching the convergence: however, in the case in which the error should result too big, a greater 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 is necessary.
The basic equation of the derivative of the spectrum S with respect to a 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):

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

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

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

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

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

To solve this equation 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 ($T_r^{ret}, p_r^{ret}, X_r^{ret}$).

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

Finally the derivatives can be written as:

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

with

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

and

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

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 improve this further 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:

$$\frac{\partial k_l(T_l^e, p_l^e)}{\partial T_l^e} = \frac{\partial\left[\sum_{li} L_{l,li}(T_l^e) A_{l,li}(T_l^e, p_l^e)\right]}{\partial T_l^e} \qquad (6.7.7)$$

$$= \sum_{li}\left[A_{l,li}(T_l^e, p_l^e)\frac{\partial L_{l,li}(T_l^e)}{\partial T_l^e} + L_{l,m}(T_l^e)\frac{\partial A_{l,li}(T_l^e, p_l^e)}{\partial T_l^e}\right]$$

and

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

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 is to implement into the program analytical derivatives which only contain $\frac{dC_l}{dX_r^{ret}}$. If retrieval tests show that the use of $\frac{dC_l}{dX_r^{ret}}$ is not sufficient we intend to reduce the layer thickness since this minimises the influence of the other contributions to the whole derivative.

6.7.3 Derivative with respect to the temperature

The main contributions to this derivative results 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 $\frac{\partial L_{l,m}(T_l^e)}{\partial T_l^e}$. The larger 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 $\frac{\partial L_{l,m}(T_l^e)}{\partial T_l^e}$ has to be done during the calculation of

the absorption cross sections and therefore needs much more time than $\frac{\partial B_l(T_l^e)}{\partial T_l^e}$. 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. This allows us to test different approximations without loosing computation time with respect to the implementation of analytical T-derivatives.

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 whole gas column of each path. The Curtis-Godson temperatures, pressures or the total air column don't change when varying the continuum cross sections. Equation 6.7.1 can be written as:

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

and for 6.7.2 follows:

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

$k_l^{cont,e}$ 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 ($k_r^{cont,ret}$):

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

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 function it's 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) consists of the partial derivatives of the spectrum with respect to the parameters. Therefore, 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 ($\partial parameter_i / \partial parameter_j = 0, \forall i \neq j$), 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

2. A small maximum correction applied to the parameter

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.

4. A maximum allowed number of iterations.

While condition 4 is only used to avoid waste of time 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 unknown, however, if a conservative value is applied at all altitudes, it may be very difficult to meet this requirement. The alternative condition 2 bis may be easier to meet, but can provide false convergence at the beginning of the iteration 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 derivative calculations).
Both condition 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 an AND of conditions 1, 2 or 3. The parameters used to set the values that each condition should reach must be optimised experimentally with suitable tests. The optimisation of the convergence criteria must take into account the effect that also other features, such as the accuracy of the calculation of the derivatives and the number of retrieved parameters have on the convergence process.

6.9 Cross-section look-up tables

The use of pre-computed look-up tables is an alternative method to the explicit 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 which represent 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 where the cross sections have to be calculated is rather fine (in the order of 5×10^{-4} cm^{-1}) the amount of data is large.
In order to reduce the amount of data contained in the look-up tables and the time necessary for reading them, a procedure that allows to compress them has been studied at University of Oxford (see Morris et all. (1997)).
The algorithm used to build the compressed look-up tables and the procedure for the decompression is described in the paper of Larrabee et all., (1997).
Version 2.3 of the ORM code is able to handle these compressed look-up tables, and also 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.

6.10 Pre-calculation of line shapes

This item concerns the line shape calculation in the case of HNO_3 volume mixing ratio retrieval. 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_{eq}-T_{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 assessed with test calculations. The maximum difference between exact calculation and use of the interpolated pre-calculated line shape was NESR/88. Regarding this small value and the time saving of 66% our baseline is to use the pre-calculated line shape during the HNO_3 retrieval.

6.11 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 in the order of 5×10^{-4} cm^{-1}. 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 at some half-widths away from the line-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 cross section calculation in the following way:
In addition to the constant general fine grid ($\Delta^{gf}=5\times10^{-4}$ 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:

$$\Delta^{lf} = n\Delta^{gf}, \Delta^{lc} = m\Delta^{lf} \quad \text{with } n,m \in \mathrm{N}, \text{ and } 1 \le n,\ 1 \le m \tag{6.11.1}$$

- 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 $\Delta^{lf} \approx \xi\left(\alpha^{L_m} + \alpha^{D_m}\right)$, 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 $\zeta\left(\alpha^{L_m} + \alpha^{D_m}\right)$ 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%.

6.12 Variable frequency grids for radiative transfer computation

Usually, also before the convolution by the apodised instrument line-shape (AILS), the high resolution spectra in the microwindows used in MIPAS retrievals have a rather smooth shape, especially at low tangent altitudes. This implies that it is not strictly necessary to compute the radiative transfer for all the fine frequency grid points of the spectrum: it is enough to calculate the radiative transfer integral for only a set of selected spectral points and derive the remaining points by just using linear interpolation between the points computed with the radiative transfer. With this approximation, valuable time savings are expected at both low and high tangent altitudes.

At low tangent altitudes, the spectral features are quite broad and have a smooth shape so that several points can be derived with linear interpolation without degrading the accuracy of the simulated spectrum. On the other hand, at high tangent altitudes the spectral lines are narrow but often separated by large flat regions which can be computed through interpolation with high accuracy.
From the implementation point of view, this optimisation does not involve serious complications in the code; however, in order to use it operationally, some effort has to be spent for defining, as a function of tangent altitude, the frequency grid points for which the radiative transfer is to be computed.
The current baselines are the following:

- An ESA study is being carried-out at University of Oxford for defining logical vectors which identify, as a function of microwindow and tangent altitude, the fine frequency-grid points for which the radiative transfer is to be computed (see Wells (1997)).
- The ORM code (Version 2.3) has the capability of handling the logical vectors defined by Oxford University. The spectral data points which are not computed with the radiative transfer are obtained by using linear interpolation.
- The time savings and accuracy performances obtained with this optimisation will be assessed as soon as the aforementioned logical vectors will be available for a standard set of microwindows and tangent altitudes used in MIPAS retrievals.

6.13 Calculation of inverse of Variance Covariance Matrix of the observations.

In section 4.5 we discussed how the VCM of the observations can be derived and it has been shown that it is a block-diagonal matrix constituted of as many blocks as many microwindows are considered. Each block is a sub-matrix of the widest block.

Looking at equation (4.2.7) we see that the inverse of $\mathbf{V}^S$, $\left(\mathbf{V}^S\right)^{-1}$, is required.

If $\mathbf{V}^S$ is a block-diagonal matrix, $\left(\mathbf{V}^S\right)^{-1}$ is a block-diagonal matrix too.

We recall that each block of $\mathbf{V}^S$ can be calculated using equation (4.5.4):

$$V^S = \delta_i^{\,2} \cdot \boldsymbol{J} \cdot \boldsymbol{J}^T \qquad (6.13.1)$$

where $\boldsymbol{J}$ is given by eq.(4.5.7). Then the blocks of $\left(\mathbf{V}^S\right)^{-1}$ can be computed with routines that invert matrices.

If MPD ≥ D, an optimised procedure can be used for the calculation of $\left(\mathbf{V}^S\right)^{-1}$: it consists in calculating directly the block of $\left(\mathbf{V}^S\right)^{-1}$ relative to the widest used microwindow $\left(\mathbf{V}_{opt}{}^{S^{-1}}\right)$ not using matrix algebra, but using properties of the space of the functions.
The optimised procedure, based on the assumption that the matrix $\boldsymbol{J}$, as well as each microwindow, is unlimited, results very useful even if this assumption is not true.

In the following, the optimised procedure used to calculate $\left(\mathbf{V}^S\right)^{-1}$ when MPD ≥ D, is described.

Each element of the matrix $\boldsymbol{J}$, which is the transformation matrix of a convolution with the symmetrical function $\theta(\sigma)$, can be expressed as:

$$J_{ij} = \theta\left(\left|\sigma_i - \sigma_j\right|\right) = \theta\left(|i-j|\cdot\Delta\right) = \theta(n\cdot\Delta), \tag{6.13.2}$$

where $n = |i-j|$ assumes all the values from 0 to the maximum number of points in the widest microwindow minus 1, Δ is the frequency step.
Therefore, ***J*** has the property that each column and each row are obtained shifting by one position the previous column and row:

$$\mathbf{J} = \begin{pmatrix} \theta(0) & \theta(1) & \theta(2) & .... \\ \theta(1) & \theta(0) & \theta(1) & .... \\ \theta(2) & \theta(1) & \theta(0) & .... \\ .... & .... & .... & .... \end{pmatrix} \tag{6.13.3}$$

The different blocks $\mathbf{V}^S$ are obtained from matrices ***J*** of different size leading to different results. However, in the case of unlimited matrices, because of property (6.13.3), each row of the sub-matrix $\mathbf{V}^S$ is equal to the convolution of the function $\theta(\sigma)$ with itself; if we define the function:

$$G(\sigma) = \theta(\sigma) * \theta(\sigma), \tag{6.13.4}$$

each element of the sub-matrix V^S is equal to:

$$V_{ij}^S = \delta^2 \cdot G\left(|i-j|\cdot\Delta\right), \tag{6.13.5}$$

so that $\mathbf{V}^S$ has the same property (6.13.3) as the matrix **J**.
In order to calculate $\left(\mathbf{V}^S\right)^{-1}$ we can assume that also this matrix has the structure of **J** and $\mathbf{V}^S$:

$$\left(V^S\right)^{-1}_{i,j} = \frac{1}{\delta^2}\cdot H\left(|i-j|\cdot\Delta\right). \tag{6.13.6}$$

$\left(\mathbf{V}^S\right)^{-1}$ must satisfy the equation:

$$\mathbf{V}^S\cdot\left(\mathbf{V}^S\right)^{-1} = \left(\mathbf{V}^S\right)^{-1}\cdot\mathbf{V}^S = \mathbf{1} \tag{6.13.7}$$

Because of the property (6.13.3) of matrices $\mathbf{V}^S$ and $\left(\mathbf{V}^S\right)^{-1}$, this matrix equation corresponds to the convolution between the functions:

$$G(\sigma) * H(\sigma) = \hat{\delta}(\sigma), \tag{6.13.8}$$

$\hat{\delta}(\sigma)$ is the Dirac-function.
Fourier transforming this equation we obtain:

$$FT^{-1}(H(\sigma)) = \frac{1}{FT^{-1}(G(\sigma))} = \frac{1}{\Theta(d)^2} \qquad (6.13.9)$$

$\Theta(d)$ is the apodisation function in OPD domain.
Finally we obtain:

$$H(\sigma) = FT\left(\frac{1}{\Theta(d)^2}\right). \qquad (6.13.10)$$

An expression for $H(n \cdot \Delta)$ has been determined, hence, since if the inverse of a matrix exists, it is unique, $\left(\mathbf{V}^S\right)^{-1}$ can be calculated using equation (6.13.6) and (6.13.10).

The optimised procedure described above is valid in the case of unlimited matrices. However, the block of $\left(\mathbf{V}^S\right)^{-1}$ resulting from this procedure can be used, with some modifications, also if the matrices are limited, and this is the case we are interested in.

When the apodisation function, which has to be convoluted with the spectrum, is made of a finite number of points, the only effect of considering limited microwindows is that the spectral points near the borders of the microwindow result less correlated than they would be if the microwindow was unlimited, while the other points are not affected by boundary effects. This means that matrix $\left(\mathbf{V}^S\right)^{-1}$, calculated with the optimised procedure, can be used as the core of the sub-matrices related to the different microwindows, while the first and last rows and columns of these sub-matrices have to be substituted with more adequate values.

Moreover, if the number of spectral points contained in a microwindow is greater than the number of points of the convoluting function, the modifications to be made on the blocks of the matrix $\left(\mathbf{V}^S\right)^{-1}$ related to a particular microwindow do not depend significantly on the number of spectral points of the microwindow itself: therefore the corrected rows and columns of each block of $\left(\mathbf{V}^S\right)^{-1}$ can be computed at the beginning and used for correcting all the blocks.

The above considerations are not true for the microwindows constituted of a number of spectral points less than the number of points of the apodisation function, because these microwindows are more affected by boundary effects. In this case the blocks of $\left(\mathbf{V}^S\right)^{-1}$ related to these particular microwindows have to be computed with the exact procedure.

In conclusion, if MPD ≥ D, the sub-matrix $\left(\mathbf{V}^S\right)^{-1}$ related to the widest microwindow is computed with the optimised procedure at the beginning of each retrieval (but it could be computed also in a pre-processor program), then the blocks of the matrix $\left(\mathbf{V}^S\right)^{-1}$ related to the microwindows extended more than the apodisation function can be derived by using as core the core of the sub-matrix $\left(\mathbf{V}^S\right)^{-1}$ previously determined, and as external rows and columns the ones calculated using the exact procedure.

For the small microwindows the exact procedure must be used.

If MPD < D, the optimised procedure cannot be used since from equ. (6.13.10) we see that if $\Theta(n \cdot \Delta')$ is equal to zero (and this is the case for MPD < D), $H(n \cdot \Delta)$ does not exist. In fact in this case $\mathbf{V}^S$ is a singular matrix.

Therefore, each block of $\left(\mathbf{V}^S\right)^{-1}$ has to be derived by computing first the blocks of $\mathbf{V}^S$ related to the different microwindows and then inverting them. The inversion of these singular matrices is performed with the method of singular value decomposition described in paragraph 4.7.

7 - Validation of the choices

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 assumptions and the anticipations that have been used to progress in the study.
At the present stage of the study several parameters have already been tuned and the most part of the choices has already been validated by the first results of the tests with the RFM. In the following list we identify the most important issues which in our opinion require a further validation or still have to be addressed.

- Validation of the criteria used for building the layering of the atmosphere
- Verification of baseline of neglecting line mixing
- Verification of baseline of neglecting pressure shifts
- Verification of baseline of neglecting NLTE
- Verification of use of analytical convolution of FOV
- Verification of impact of continuum constraints on the retrieval, i.e. tuning of the parameters which define the strength of continuum constraints
- Optimisation of convergence criteria
- Optimisation of the parameters which control the evolution of Marquardt damping factor during the iterations
- Assessment of the total error budget affecting the retrieved VMR profiles. In particular, the propagation of p,T retrieval errors into VMR retrievals should be investigated.
- Verification of the rules used for the extrapolation of the profiles outside the altitude range covered by the measurements
- Assessment of differences between retrieval at tangent altitude levels and at fixed altitude levels.
- Tuning of frequency grid parameters
- Choice between regular or irregular grids for radiative transfer calculations
- Choice between stored and calculated cross-sections
- Interval for Voigt line-shape calculation.

Some of these open issues can be fixed by just performing dedicated tests against the RFM. However, in some cases the choices would be better validated by exercising the ORM program with real data.

References

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.

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.

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, *'Atmopspheric 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, *'Assessement of finite field-of-view effects on MIPAS ILS & Review of resolution enhancement techniques'*, Answer to ESA NTO/ME/1573 (1995).

Morris P. .. 'Generation of compressed cross-section look-up tables for NRT MIPAS retrievals', ESA report PO-TN-OXF-GS-0011 (still to be issued) (1997).

Drayson, S. R., *'Rapid computation of the Voigt profile'*, J Quant. Spectr. Radiat. Transfer, 16, 611-614, 1976.

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.

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. Schlussel, P. Schlussel, 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.

Larrabee Strow, Robert G. Benson, Scott E. Hannon, Howard E. Motteler: "Fast Computation of Monochromatic Infrared Atmospheric Transmittances Using Compressed Look-Up Tables", still to be published (1997).

List R. J. ed.: *Smithsonian Meteorological Tables,* Sixth Revised Edition, Washington, D.C., 1963.

Orlando J.J.,G.S.Tyndall,K.E.Nickerson,J.G.Calvert, 'The temperature dependence of collision induced absorption by Oxygen near 6 μm', JGR,96,D11,20755-20760

Press, W. H., S. A. Teukolsky, W. T. Wetterling, B. P. Flannerly: *'Numerical Recipes in Fortran'*, Second Edition Cambridge University Press 1992.

Rodgers C. D. *'Retrieval of Atmospheric Temperatire and Composition From Remote Measurements of Thermal Radiation'*, Reviews of Geophysics and Space Physics, Vol. 14 N. 4, p.609, 1976.

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.

Turner, D. S., *'Absorption coefficient estimation using a two-dimensional interpolation procedure'*, J. Quant. Spectrosc. Radiat. Transfer, 55, 633-637, 1995.

Sterne T. E. *'An Introduction to Celestial Mechanics'*, Interscience publishers Ltd, London, 1960.

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).