Technical Note on:

Level 2 Algorithm Characterization & Validation Plan

Issue 1
Revision A

10 September 2001

Delivery of the study:

"Development of an Optimized Algorithm for Routine p, T and VMR Retrieval from MIPAS Limb Emission Spectra"
Contract No: 11717/95/NL/CN – CCN5

Prepared by:

Name	Institute
C. Piccolo	IROE-CNR
M.Prosperi	IROE-CNR
M. Ridolfi	University of Bologna
V. Tenna	IROE-CNR

Approved by:

Name	Institute
B. Carli	IROE-CNR
P.Raspollini	IROE-CNR

Document Change Record

Issue	Revision	Date	Change description
Draft0	-	15 Jan. 01	First draft of document
Draft1	-	17 Apr. 01	Changes resulting from ORM study progress meeting # 14
Issue 1	-	04 July 01	Change bars highlight differences between Issue 1 and Draft1 Changes resulting from comments by U. of Oxford, (A.Dudhia), e-mail 4 May 2001 Included references to tests mentioned in the ESA document "Implementation of MIPAS Post-Launch Calibration and Validation Tasks" [RD8], Added Sect. 8 regarding "data exchange strategy"
Issue 1	A	10 Sept. 01	Change bars highlight differences between Issue 1 and Issue 1 Revision A. Changes resulting from comments by ESA (e-mail from H.Nett, 13 July 2001).

CONTENTS

1. Background

In the frame of the ESA contract 11717/95/NL/CN an Optimized forward / Retrieval Model (ORM) was developed, suitable for implementation in MIPAS near real-time Level 2 processor. The developed model constitutes the scientific version of the so called 'Retrieval Components Library' and includes p,T and VMR retrieval modules. In order to cope with the very demanding runtime requirements of a near real-time processor, the retrieval model was developed aiming at the optimization of the trade-off between accuracy and runtime performance. For this reason, several approximations have been implemented in the code.

So far, in absence of MIPAS measurements, the implemented approximations have been validated only on the basis of simulations and, whenever possible, on the basis of available measurements acquired by instruments similar to the ENVISAT version of MIPAS: the balloon instrument MIPAS-B2 and the Space-Shuttle instrument ATMOS.

In this framework, two issues are posed when the first MIPAS measurements will be available:

- Some processing setup parameters can not be optimized with the currently available information and therefore will require a tuning based on MIPAS measurements,
- the impact of the most critical approximations implemented in the ORM must be adequately characterized on the basis of real MIPAS measurements,

Since the time interval (commissioning phase) in which it will be possible to carry-out these characterization and validation activities is very short, the definition of a detailed plan is of primary importance.

In the present document we review all the procedures planned for the commissioning phase and, for each of them we identify: (a) input, (b) output, (c) responsible person, (d) expected required time interval for completion. Finally all the activities are arranged in a global time chart.

2. Reference documents

#	Document	Issue	Title
[RD1]	TN-IROE-GS0002	Draft	Level 2 Algorithm Characterization & Validation Strategies
[RD2]	PO-TR-DAS-MP-0143	1	MIPAS FM instrument performance verification test report
[RD3]	V.Jay and A.Dudhia		MIPAS-B Retrievals residuals analysis (23 Jan.01)
[RD4]	TN-ISM-0002		TN on MIPAS-B data analysis: flight #6 old MWs
[RD5]	M.Carlotti et al. Applied Optics paper 20 April '01		Geo-fit approach to the analysis of satellite limb-scanning measurements
[RD6]	M.Ridolfi and B.Carli, 22 July 1999		Memorandum on determination of the VCM of engineering tangent heights in MIPAS
[RD7]	TN-IROE-GS0003	Draft	Pre-flight modifications to the ORM_ABC code
[RD8]	PO-PL-ESA-GS-1124	1	Implementation of MIPAS Post-Launch Calibration and Validation Tasks
[RD9]	TN-IROE-RSA9601	3	High level algorithm definition and physical and mathematical optimisations
[RD10]	Tech. Note by A.Dudhia		Modified Radiative Transfer Calculation (11 April 1999)

3. Objective

Objective of the present technical note is the definition of a plan for the activity related to MIPAS Level 2 algorithm characterization and validation during the commissioning phase.

In order to properly organize this activity (see Sect. 6), first all planned char./val. operations must be identified (see Sect. 4) and their logical sequence defined (see Sects. 5 and 6). The description of the various tests and operations reported in Sect. 4 sometime contains more details than the description reported in RD1. In fact, the goal of RD1 was only to establish the ‘priority’ of the individual tests while in the present document their time-sequence must be identified. To this purpose all the required details of the operations are identified in Sect. 4.

The description of the operations as reported in Sect. 4 is also a baseline to establish the requirements of the software tools that will be used in the cal./val. activity.

4. Planned characterization validation operations

The sequence of operations aiming at the characterization and validation of MIPAS Level 2 processor can be summarized in the following steps:

1. Tuning of processor setup parameters:
   - ➢ starting from a “reference” set of processor setup parameters and a set of MIPAS measurements (input)
   - ⇒ the processing setup parameters will be tuned (see Sect. 4.1)
   - ➢ the output of this procedure is an optimized set of processing setup parameters and the generation of a “reference” retrieval

2. Perform tests for critical-baseline verification
   - ➢ Using the outputs of the reference retrieval generated above (input)
   - ⇒ the tests for critical-baseline verification will be carried-out (see Sect. 4.2)
   - ➢ Outputs are:
     - Recommendations (arising from the verification of Level 1b assumptions) to Level 1b experts
     - Identification of ORM approximations that cause unacceptable errors
     - Characterization of systematic errors affecting Level 2 input data (e.g. MWs)
     - Identification of specific issues

3. Contingency activity required to understand specific issues identified under step 2.
   - ➢ Input: specific issues identified under step 2.
   - ⇒ Study of the specific issues identified under step 2.
   - ➢ Recommendations for further improvements of Level 2 processor and auxiliary data

4. Preparation of a report that summarizes results and recommendations of findings. Possibly the report will be discussed in a meeting
   - ➢ Uses the results of the above analysis (under steps 2 and 3.)
   - ➢ Output: recommendations for actions concerning Level 1b, auxiliary data and ORM.

5. Correction of unacceptable approximations in the ORM code and in auxiliary data
   - ➢ Using the results of the above meeting (under step 4.)
   - ⇒ the identified corrections for ORM and auxiliary data will be implemented
   - ➢ Output: corrected version of the ORM and of auxiliary data

- Possible re-iteration of above activity depending on type of modifications implemented under step 5.

In the present section we describe the procedures required to optimize all processor input parameters (Sect. 4.1) and the tests (Sect. 4.2) required to verify critical Level 2 baselines (see also RD1). Tests reported in RD1 with “C” priority are not mentioned here, as they will not be done. For each parameter subject to tuning and for each test aiming at the verification of a critical Level 2 baseline we also raise the requirements for the software tools to be developed in support to the mentioned procedures. Namely, the following tools will be used:

1. Statistical Tool (ST, to be developed). This tool shall include several functionalities of both mathematical and graphical type, supporting the evaluation of retrieval results on a statistical basis.
2. ORM_SDC (to be developed using the ORM_ABC as a starting basis). This code will have the capability of fitting additional (instrumental and calibration) parameters compared to the ORM_ABC.
3. OFM (already available). Is the self-standing forward model.
4. Tool for REC (Residual and Error Correlation) analysis (see [RD3]).

The individual procedures reported in Sect.s 4.1 and 4.2 will be terminated when the termination criterion reported for each procedure is fulfilled.

4.1 Tuning of critical processing setup parameters

The processing setup parameters are the Level 2 processor input parameters that characterize and identify options and criteria used by the processor during the calculations.
The processing setup parameters will be optimized by performing a set of test retrievals and / or simulations for one selected "reference" orbit. This "reference" orbit will be selected with the help of the software tools available to ESA, allowing to inspect the Product Confidence Data (PCD) derived by the Level 1b processor. After the tuning (optimization) phase, the processing setup parameters will be consolidated by systematically processing a significant number of orbits within the M_RP_1 sample [a statistically significant set of orbits (≈ 43) of nominal measurements]. The MIPAS Level 2 pre-processor (ML2PP) could be used to support this exercise (TBC). Presently we do not make of the consolidation task a firm requirement for our post-launch activities because of the large uncertainty existing on the computing workload that we will have to face during the commissioning phase. For this reason the "consolidation" activities are not described in the present document.
The processing setup parameters will be optimized considering all the retrievals of the orbit (which means all types of retrievals, i.e. p,T, H_2O O_3, etc. and all the scans of the orbit). The setup parameters may be personalized for type of retrieval, therefore for each parameter, 7 optimum values (one value for each type of retrieval) must be determined which best cope with the needs of the various types of retrievals along the whole orbit (the setup parameters are not personalized for the individual scans).
The general approach that will be used to tune an individual setup parameter consists in:

1. perform a retrieval with a “reference” set of processing setup parameters. This “reference” set is obtained from the experience on both MIPAS-B data analysis and simulated retrievals (see Task 7 of the CCN5 of ESA contract 11717/95/NL/CN). In principle this “reference” set of parameters can be different from the set that will be used for the initial Level 2 processor operations.
2. vary within a pre-defined range, and with a pre-defined sampling, the parameter to be optimized and perform test retrievals with the modified parameter.

3. the “optimum” value for the tuned parameter is the value (within the range explored in step 2.) that provides the best performance with minimum cost.

The proposed optimization procedure is based on heuristic choices made on the first real atmospheric measurements that will become available. For efficient results the quantitative definitions of "performance" and “cost” of the retrieval do not require a too rigorous quantifier, also considering that educated choices must be made in order to have a margin that accounts for the fact that the tests are not performed on an exhaustive sample of possible atmospheres.

4.1.1 Atmospheric continuum

The atmospheric continuum is assumed to be constant within the microwindows and to be both altitude and microwindow dependent. In order to limit the number of retrieved parameters, the atmospheric continuum is fitted only below a user-defined altitude z_{ucl}. Above this altitude the continuum is obtained by scaling the initial guess profile to match the highest fitted continuum parameter. Below z_{ucl} the continuum is assumed to vary linearly with frequency within user defined frequency ranges (“umbrella” radii) that depend on both microwindow and tangent altitude. Furthermore, the atmospheric continuum is forced to be zero above a user-defined altitude z_{c0}.

Umbrella radii will be set equal to an arbitrarily small value (e.g. 0.1 cm^{-1}, so that MWs are not grouped for continuum retrieval) and will not be tuned. This is because in its present version, the algorithm selecting the optimized MWs for MIPAS retrievals (see RD9) uses the fitted continuum to compensate for systematic errors having a continuum-like impact on the spectra. Therefore the fitted "continuum" has lost its original physical meaning and its spectral behavior may easily contain sharp features that cannot be represented with linear interpolation.

Apart umbrella radii, the other continuum-related settings will be tuned with the following strategy: for each parameter independently, we start from a retrieval with very weak constrains (example: continuum set to zero above 60 km, fitted below 36 km) and we gradually increase their strength so that the retrieval stability (see definition in appendix A) improves. At the same time we expect that as the constrains are increased, the chi-square will also increase. The strength of the constrains should not be further increased when they prevent the retrieval from reaching the chi-square value that is reached when weak constraints are applied.

Required Tools:

- Statistical tool

Tool requirements arising from this procedure:

- The statistical tool should be able to evaluate the quantifier of retrieval stability as defined in appendix A.

4.1.2 Marquardt damping factor

The Marquardt method involves the introduction of a damping factor that reduces the amplitude of the parameter correction vector. This method is intended to induce smoother convergence especially in case of non-linear problems. The damping factor is initialized to a user-defined value and during the retrieval iterations it is increased or decreased depending on whether the chi-square function increases or decreases. The initial value of this damping factor and the factors used to increase and decrease it during the iterations are subject to tuning. Furthermore, the use of Marquardt algorithm requires that also the maximum allowed number of micro-iterations must be established by the user.

The strategy for the choice of the parameters that control the behavior of the damping factor is based on the minimization of the number of iterations needed to reach the convergence and on the maximization of the retrieval stability (see appendix A). Furthermore we must consider that the forward model internal to the ORM calculates also the Jacobian at each run, therefore a micro-iteration costs as much as a macro-iteration in terms of computing time. For this reason the

Marquardt-related parameters should be optimized trying to avoid the occurrence of micro-iterations.
The trade-off between the Marquardt-related parameters and the parameters driving the regularization is of course very strong. The approach is therefore first to optimize the Marquart-related parameters without any regularization and subsequently optimize the strength of the regularization (Sect. 4.1.3).

Required Tools:
- Statistical tool

Tool requirements arising from this procedure:
- The statistical tool shall be able to display, for each retrieval, the chi-square, the value of Marquardt damping factor and the micro- and macro- iteration indices. The statistical tool shall be able to build the plot reported in Fig.A2 of appendix A.

4.1.3 Regularization parameters

In some cases the retrieved profiles are oscillating more than what can be reasonably expected from the physical point of view. This oscillation is intrinsic with the retrieval problem, because the solution is represented in a base of functions different from the base of observations. The techniques intended to reduce these instabilities are called 'regularization' techniques. Tikhonov-Phillips regularization is adopted in MIPAS retrievals.
The user-defined parameters controlling profile regularization are subject to tuning. In particular, these parameters are the elements of the regularization operator for the individual groups of retrieved parameters and a parameter which establishes the global strength of the applied constraint. The optimum strength of the regularization is given by the regularization that produces the smoothest retrieved profiles without significant impact on the final value of the chi-square. The optimum regularization strength is identified using a plot similar to Fig. A2 (see appendix A) with the merit figure M is defined as:

$$\mathrm{M} = rms\left\{ \frac{x_i - \dfrac{x_{i-1} + x_{i+1}}{2}}{\sigma_i} \right\} \qquad (1)$$

where x_i are the elements of the retrieved profile, σ_i their error and *rms* represents the root mean square of the quantity reported in brackets for $i = 2, \ldots, npoints - 1$, with $npoints$ = number of retrieved points. The smaller is the quantity M, the more regular is the retrieved profile, therefore the strength of the regularization constraint is tuned by finding a compromise between accuracy (attained chi-square) and smoothness (M) of the retrieved profile.
If comparable performances are attained with different sets of regularization parameters, preference will be given to the choice of using no (or weaker) regularization for the target parameters (p,T and VMR) and to choose an optimum regularization strength for continuum parameters.

Required Tools:
- Statistical tool

Tool requirements arising from this procedure:
- The statistical tool shall be able to display, for each retrieval, the chi-square, the value of Marquardt damping factor and the micro- and macro- iteration indices. Furthermore it shall be

possible to create a plot of the eigenvalues of the inverse VCM of the retrieved parameters. The statistical tool shall also be capable of building the plot of Fig. A2 reported in appendix A, with M defined in Eq. (1).

4.1.4 Threshold for the eigenvalues in the inversion of the VCM of the retrieved parameters

A procedure for the optimization of this threshold is still being worked-out and may involve some change in the definition of the retrieved continuum parameters (TBD). In case a procedure for the optimization of this parameter is still pending during the commissioning phase, the presently used conservative value (10^{-20}) will be maintained.

Required Tools:
- Statistical tool

Tool requirements arising from this procedure:
- The statistical tool shall be able to plot the eigenvalues relating to the various retrieval parameters.

4.1.5 Convergence Criteria

The adopted convergence criteria are based on three conditions:
1. Linearity of the inversion problem. The maximum relative difference (in two subsequent iterations) between linear and real chi-square must be less than a pre-defined threshold T1.
2. Attained accuracy. The maximum relative variation (in two subsequent iterations) of the fitted parameters must be less than a pre-defined threshold T2.
3. Computing time. Due to general computing time constraints in MIPAS Level 2 processor, there is a max. number of iterations beyond which the retrieval must be stopped. However, the present runtime requirements of the Level 2 processor are not very stringent and therefore the max. number of allowed iterations can be set on the basis of the ORM team experience based on both simulated and MIPAS-B2 retrievals. From the experience we learned that if a retrieval is not converging after 10 iterations then also after 30 iterations it will not converge because of some contingent problem. Therefore we will set the max. number of both micro- and macro- iterations equal to 10 for all retrieval types.

The retrieval is stopped if one of the above 3 conditions is fulfilled, the convergence is reached if one of the first two conditions is fulfilled. The task is therefore to tune the thresholds T1 and T2 related to conditions 1. and 2.
Procedure:
The threshold T1 is not subject to tuning because from the physical point of view it is directly connected with the accuracy of the retrieved parameters. Therefore T1 must be a fraction (e.g. $1/10^{th}$) of the expected total error affecting the target parameter to which it refers.
The threshold T2 will be tuned using this approach: T1 and T2 will be initially set to an arbitrarily small value (e.g. = 0), all the retrievals relating to the selected orbit will be carried-out and, for each retrieval the following plot will be built:

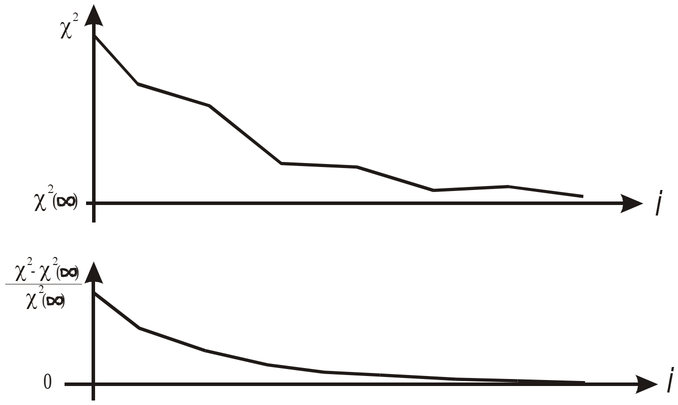

Fig.1: chi-square and relative difference between value of the chi-square at the current iteration and asymptotic value of the chi-square.

Then, T2 will be chosen in such a way that at convergence:

- in 99 % of the retrievals $\frac{\chi^2-\chi^2(\infty)}{\chi^2(\infty)}<0.10$,
- in 95 % of the retrievals $\frac{\chi^2-\chi^2(\infty)}{\chi^2(\infty)}<0.05$
- in 85 % of the retrievals $\frac{\chi^2-\chi^2(\infty)}{\chi^2(\infty)}<0.01$

these conditions will be checked by making use of 3 plots. Each plot refers to one of the 3 conditions specified above and reports the number of the retrievals for which the condition is fulfilled, as a function of the value of T2.
In order to gain confidence on the selected value of T2, the above analysis should be carried-out for a statistically significant number of orbits. However we consider the above conditions more "guidelines" rather than "stringent requirements", and we do not consider a "stringent requirement" the task of processing large sets of orbits. This is because of the large uncertainty existing on the computing workload that we will have to face during the commissioning phase (see also comment in Sect. 4.1).

Required Tools:
- Statistical tool

Tool requirements arising from this procedure:
- The statistical tool shall be able to build the plot of Fig.1 for each retrieval.

4.1.6 Retrieval grid

The vertical resolution and the accuracy with which the retrieved profiles are determined are strongly dependent on the altitude grid where the retrieved points are represented (retrieval grid). The ORM choice is to retrieve vertical profiles at an altitude grid defined by the tangent altitude levels, since this provides the most stable results. It is however possible to limit (via input parameters) the retrieval to a subset of the tangent altitudes. This option allows therefore to choose for each tangent altitude whether the profile is fitted or obtained using interpolation/extrapolation

from adjacent values. Large retrieval errors are caused by lack of information on the target profile at the retrieval altitude. In this case, retrieving the investigated parameter for each measured altitude could be harmful to the overall retrieval because the large error associated with a single retrieved parameter could propagate to the other altitudes as well. Therefore, for some gases it may be desirable to retrieve the VMR profile only at those tangent altitudes carrying sufficient information on the profile itself.
Sensitivity tests have shown that all target gas profiles but HNO_3, N_2O and NO_2 can be retrieved at all tangent altitudes of the measured scan. The retrieval grid must be optimized for HNO_3, N_2O and NO_2. This optimization involves two steps: the definition of the altitude range (already made in the OM selected by Oxford) and the definition of the vertical resolution (assumed to be equal to the measurement resolution by Oxford). No further tuning is possible and only some validation tests with different retrieval grids will be made.

Required Tools:
- ORM_SDC

Tool requirements arising from this procedure:
- None.

4.1.7 “Linear-in-tau” method

In the radiative transfer of the ORM the Planck function is calculated in for a constant (frequency independent) temperature equal to the Curtis-Godson temperature of the target gas of the retrieval. This approximation may turn out to be too rough for tangent layers in an opaque atmosphere. In some codes this problem is solved by assuming that the Planck function varies linearly with optical depth in the layer. This approach is usually referred to as the ‘linear-in-tau’ method and takes into account the fact that only a spectral element associated with small optical depth radiates with temperature determined in the Curtis-Godson approximation, while a spectral element with large optical depth radiates with a temperature characteristic of the layer boundary.
The ‘linear-in-tau’ method has been tested at Oxford University, the results are reported in [RD10].
Due to the mentioned approximation, the parameters that control the automatic layering of the atmosphere (temperature and half-width variation thresholds) must be re-tuned using the ‘real’ atmosphere as retrieved from MIPAS measurements. If the real atmosphere turns-out to be more opaque than that used for pre-flight tuning of the layering, the tuning operation may lead to the selection of a layering finer than the optimized one determined (before the flight) with model atmospheres.

Required Tools:
- OFM

Tool requirements arising from this procedure:
- None.

4.1.8 Tropopause altitude

An input parameter that characterizes the tangent altitude grid at which the simulations are made (for FOV convolution) is the altitude level of the tropopause. Above and below this altitude the maximum separation between contiguous tangent altitudes of two simulated spectra is determined by a different user-defined parameter: the separation below the tropopause level being smaller than above in order to properly account for temperature and water vapor gradients.

The tropopause altitude depends on latitude and is determined on the basis of two user-defined parameters (tropopause height at poles and height increment) subject to tuning.
Procedure: initially the two parameters will be arbitrarily set to get a conservatively “high” tropopause. All the scans relating to the selected orbit will be then processed and the global maps relating to the retrieved temperature an water vapor profiles will be visualized. The visual inspection of these maps will allow to identify experimentally the behavior of tropopause altitude along the considered orbit. The two user-defined settings will be then set to best reproduce the tropopause height behavior along the orbit.
The comment reported in Sect. 4.1 regarding the computing workload during the commissioning phase applies also to the work to be done for the consolidation of these two parameters.

Required Tools:
- Statistical tool

Tool requirements arising from this procedure:
- The statistical tool should be able to visualize color maps of all profiles retrieved along the orbit.

4.1.9 Field of view

The effect of finite FOV is taken into account in the ORM by convolving the tangent altitude dependent spectrum with the FOV pattern. The FOV is represented by a spread in the altitude domain and is assumed to be constant as a function of elevation scan angle. In the ORM versions up to ABC_1.2.1 the FOV was also assumed to be band-independent, however this approximation has been removed since ORM version ABC_1.2.2 [RD7]. The shape of the FOV is represented by a piecewise linear altitude distribution tabulated in the processor input files. The tabulated shape of the FOV used by the processor is an instrument parameter and is not subject to tuning, in fact, even if the adopted FOV shape has strong impact on the residuals at low altitudes (below 20 km), several concurring effects contribute to the residuals at low altitudes and it is impossible to discriminate the impact of a wrong FOV assumption among the other effects. For this reason the analysis of the residuals at low altitudes is seen only as a “verification” procedure and not as an alternative or complementary method to optimize the FOV-related processing setup parameters.
Therefore, the used input FOV shape will be the one that best fits MIPAS FOV measured experimentally either before (see e.g. RD2) or after launch. In fact, during in-flight operations the FOV response will be measured, however the measurement procedure (details are still to be defined) is expected to provide only rough estimates of the FOV pattern for the different spectral bands.
Another aspect linked to the FOV convolution operated in the ORM is the tuning of two user-defined parameters which establish the maximum separation (in tangent altitude) between the spectra simulated for the FOV convolution. These two parameters refer to the maximum separation of the simulated spectra below and above the tropopause altitude respectively. The finer is the tangent altitude grid of the simulated spectra, the more accurate is the FOV convolution and the longer is the required computing time. The two mentioned parameters have already been tuned using spectra simulated in worst-case atmospheric gradient conditions. However if the atmosphere actually observed by MIPAS will show vertical gradients larger than those used in the tests made before launch, a re-tuning will be required (vertical gradients are checked in Sect. 4.2.7).
If a re-tuning is necessary, the procedure consists in refining the grid of the simulated spectra until a further refinement does not change (within NESR/10, TBC) the radiances simulated in correspondence of the largest (T and H_2O) gradients.

Required Tools:

- OFM

Tool requirements arising from this procedure:

- None.

4.1.10 Measurement altitude range

The newly defined standard measurement scenario (6, 9, …, 42, 47, 52, 60, 68 km) extends to very low altitudes, in fact, considering the FOV width (~ 6 km for band "A" according to the measurements reported in RD2) air masses will be sampled down to 3 km. This implies that we will probably have to cope with effects that become important at low altitudes, namely:

1. possible lack of information below a certain altitude (e.g.: due to a low concentration of a target species or to very opaque atmosphere),
2. large horizontal gradients (e.g. temperature and water vapor)
3. the presence of extra absorption and / or scattering due to clouds that are not modeled in the ORM.

These problems will be quantified by first performing a retrieval that uses only measurements with tangent altitudes at 12 km and above. Then gradually measurements with lower tangent altitudes will be included in the analysis and the consistency of the retrieved profiles with the ones retrieved using only observations above 12 km will be checked. In case no significant upward error propagation is identified in presence of the above mentioned effects, the operational Level 2 processor could still process all the spectra of the standard scan, even if in this case the accuracy at low altitudes would be very poor. If upward error propagation in the retrieval is found to be significant, a recommendation will be issued suggesting not to process in real-time the measurements with tangent altitude below a given threshold.
The above procedure may be skipped if the pre-selection of the measurement altitude range operated by the MW/OM generation tools is found to be satisfactory.

Required Tools:

- ORM_SDC

Tool requirements arising from this procedure:

- The ORM_SDC shall be capable of ignoring observations (recorded in the input files) with tangent altitude below a user-defined threshold.

4.1.11 Line Of Sight (LOS) Variance Covariance Matrix

The engineering LOS data are updated at each scan and therefore constitute an independent source of information which can be routinely used in p, T retrievals. In hydrostatic equilibrium atmosphere it is possible to derive from p, T retrieved 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 tangent altitudes can be obtained. The differences between tangent altitudes obtained from p, T retrieval and the corresponding engineering estimates constitute the 'tangent heights corrections' vector. This is the correction to be applied to the assumed value of the tangent altitudes in order to obtain their correct value. The estimation of the tangent altitudes consists in weighting the retrieved tangent altitudes, with their VCM, with the engineering values, which are characterized by an a-priori VCM. The VCM associated to the engineering tangent altitudes is obtained from a simple algorithm [RD6] that simulates MIPAS pointing performance specifications. The compliance of this algorithm with the actual MIPAS pointing performance must be assessed by characterizing the differences between the engineering estimate of tangent altitudes and the tangent altitudes retrieved by the ORM (without making use of engineering LOS data). These differences will be characterized using the following approach. For the *k-th* tangent altitude we assume:

$$\left(z_{ret}(k) - z_{eng}(k)\right)^2 = \sigma^2_{ret}(k) + \alpha_k \cdot \sigma^2_{eng}(k) \qquad (2)$$

where *z* are the tangent altitudes, σ^2 are the variances and the subscripts *'ret'* and *'eng'* refer to the retrieved and the engineering quantities respectively. α_k is a coefficient that depends on the considered tangent altitude *k* and whose expectation value is 1 if the systematic errors affecting the retrieved tangent altitudes are negligible and the errors on the engineering pointings $\sigma^2_{eng}(k)$ have been correctly estimated.
For each tangent altitude *k*, the coefficient α_k will be calculated (from Eq. 2) for all the scans of the selected orbit and its average value $\overline{\alpha}_k$ will be calculated.
For each tangent altitude *k* the quantity $\sigma^2_{eng}(k) \cdot \overline{\alpha}_k$ will provide a new estimate for the engineering error at the *k-th* tangent altitude. If $\overline{\alpha}_k$ is significantly different from 1 the plausibility of the new error estimate must be verified with the engineers experts in MIPAS pointing system. If the verification fails, the understanding of the inconsistency between retrieved and engineering tangent altitudes will be attributed either to the presence of systematic errors neglected in Eq. (2) or to the baseline tested in Sect. 4.2.3.

Required Tools:
- Statistical tool

Tool requirements arising from this procedure:
- The statistical tool shall be capable of plotting the distribution of α_k for all the tangent altitudes *k* of the (standard) scan and to display the values of $\overline{\alpha}_k$.

4.2 Tests for critical level 2 baseline verification

In this section we report the description of the individual tests that will be done during the commissioning phase aiming at the verification of critical baselines, choices and approximations implemented in the Level 2 processor. Tests which were assigned "C" priority in RD1 are not reported here as they will not be done.
Several tests mentioned in this section involve the visual inspection of the residuals in the microwindows used for the retrieval. However in case a problem shows up in the analysis of a particular residual, a re-iteration of the analysis using an extended microwindow might be recommended. For this reason, as explained in Sect. 8, a flexible strategy must be adopted for data exchange between ESA and the ESL team. Furthermore the ESL team requires the test procedures identified in the present document to be carried-out using cross-section LUTs and irregular grids (IGs) optimized for accuracy of both "kept" and "skipped" spectral points in the selected MWs. This is because during the testing stage very important insights usually arise from the inspection of the whole residuals in the MWs, and therefore also the accuracy of "skipped" points is important at this stage.
A particular test named "Residuals and Error Correlation (REC) analysis" will be used to test several baselines. This analysis is described in [RD3] and will be used to test all those approximations which are expected to have significant impact on the retrieval accuracy (all items

for which error spectra have been calculated will be checked in this analysis). The analysis requires the error spectra as auxiliary input data.

4.2.1 Local thermodynamic equilibrium (LTE)

ORM assumes the atmosphere in local thermodynamic equilibrium (LTE). This means that the temperature of the Boltzmann distribution is equal to the kinetic temperature and the source function is the Planck function at the local kinetic temperature. This LTE model is usually valid at low altitudes where kinetic collisions are frequent.
Non-LTE effects cause a radiance higher or lower than that modeled in LTE. Non-LTE effects can sometimes be discriminated by the fact that they tend to decrease during the night.

Tests

The following different procedures have been identified for testing this baseline:

1. statistics of day-night variability of residuals for identifying possible Non-LTE features,
2. analysis of the non-standard measurement scenario named: “Upper Atmosphere” scenario, in which the scans are extended to high altitudes,
3. REC analysis of the residuals

In test 1. the residuals corresponding to a selected retrieval from ‘day’ observations will be compared with the residuals of a selected retrieval from ‘night’ measurements. The residuals will be compared in correspondence of the microwindows and altitudes which have the highest NLTE error quantifiers (reported in the MW databases). This test is considered successful if no unmodeled or badly-modeled features emerge from the analysis of the residuals. Test 1 and test 3. are of “A” priority and test 2. is of “B” priority, therefore test 2. will be done only if test 1. and / or test 3. is / are not successful. Test 2. consists in analyzing the residuals of day-time retrievals in correspondence of the microwindows which caused test 1. failure. In particular these, residuals will be analyzed for very high altitudes (Upper Atmosphere scenario) where NLTE is more likely to occur (vibrational temperature much greater than kinetic temperature). The output of tests 1. and 2. consists in a list of microwindows that are found to be affected by NLTE more significantly than expected on the basis of the NLTE quantifiers. Of course this list is “empty” if test 1. is successful. This list of spectral regions will be taken into account by the Oxford team while updating MW and OM data.

Required Tools

- Statistical tool
- Tool for REC analysis of the residuals

Required Auxiliary Data

No auxiliary data required.

Required Measurement Scenario

“Upper Atmosphere” special measurement scenario, for 2.

Tool requirements arising from this procedure:

- The statistical tool shall be capable of plotting the residuals for the individual MWs / altitudes and calculating their statistical moments (TBD which ones). Visualization of the partial chi-squares and statistical moments for “all” and only “kept” points, relating to the individual MWs / altitudes. Capability of working on residuals averaged over a user-defined set of limb-scanning sequences.

4.2.2 Horizontal homogeneity

Limb sounding attains good sensitivity due to the long path length of the observation, but this necessarily implies measurements of the average atmosphere over long horizontal distances. The horizontal length scale, for a typical limb sounding experiment, is of the order of several hundreds of kilometers and the assumption that the atmosphere is horizontally homogeneous over this distance may fail in some cases. The retrieval accuracy is particularly sensitive to horizontal temperature gradients (see [RD5]).
The problem causes large chi-squares and systematic errors in correspondence of large latitudinal gradients (poles and equator). The amplitude of the horizontal gradients can be calculated from the difference between profiles retrieved from subsequent scans.

Tests
The following procedures have been identified to verify this assumption:
1. Assessment of the correlations between chi-square of the fit at a given tangent altitude and latitudinal variation of temperature, H_2O and O_3 at the same altitude, in a time sequence of limb measurements,
2. Comparison of the profiles for different scans with ‘external’ information (correlative measurements, chemical models, etc.) in areas where large gradients occur,
3. REC analysis of the residuals (includes only the effect of T gradients)

The output of procedure 1. is a set of scatter plots correlating horizontal gradient and total chi-square at a given altitude. Each plot will contain as many points as many are the limb-scanning sequences of the considered orbit. The correlation between the considered quantities will be quantified by the linear correlation coefficient. The correlation will be considered significant if the linear correlation coefficient is different from zero consistently with its 1-σ error. The most important effects of the tested assumption are expected to arise from H_2O, O_3 and T gradients at low altitudes, therefore plots will be built correlating the horizontal gradient of these quantities at a selected (low) altitude with the average chi-square at the same altitude. In case the mentioned plots do not highlight a correlation between chi-square and gradient, it means that most likely the horizontal homogeneity assumption is compensated by a systematic deviation of the retrieved VMR from its true value. In this case test 2. will provide useful insights regarding the correlation between horizontal gradients and systematic VMR error. The output of tests 1. and 2. is an assessment of the impact of horizontal homogeneity assumption on MIPAS retrievals, no re-iteration of these tests are foreseen. Test 2. can be done only if external VMR (H_2O, O_3) and T information is available.
The confidence level of the above tests could be increased by extending the analysis to a set of several orbits, however this operation is seen as a consolidation process to which applies the comment reported in Sect. 4.1 (regarding computing workload during the commissioning phase).

Required Tools
- Statistical tool
- Tool for REC analysis of the residuals

Required Auxiliary Data
ECMWF (T, H_2O and O_3) ‘external’ profiles, for test 2. (TBC).

Required Measurement Scenario
Standard measurement scenario.

Tool requirements arising from this procedure:
- The statistical tool shall be capable of building scatter plots that, at a user-defined altitude, correlate the average chi-square at the selected altitude and the horizontal gradients of T, ozone and water vapor. Both the linear correlation coefficient and its statistical error shall be displayed in these plots.

- The statistical tool shall also be capable of plotting initial guess, retrieved and external profiles (TBC) for the various scans along the orbit.

4.2.3 Hydrostatic equilibrium

Hydrostatic equilibrium provides a relationship between temperature, pressure and geometrical altitude and is generally fulfilled in normal atmospheric conditions (especially in the stratosphere). It has to be noted that, with limb scanning, the profile of acquired tangent points is a slant profile. This is due both to the variation of the tangent point position with the elevation angle and because of the satellite motion (most important factor).
The assumption of hydrostatic equilibrium fails in the case of a vertical profile in presence of strong turbulence and in the case of slant profiles through a non-homogeneous atmosphere.
The problem can generate occasionally large chi-square values in p,T retrievals due to the fact that p and T variables are over-constrained.

Test
The test procedure consists in the assessment of the correlations between the horizontal temperature gradients and the tangent altitude corrections (i.e. the difference between engineering and retrieved estimates of the tangent altitude separation) obtained from p,T retrieval. For a TBD (user-defined, low) altitude a scatter plot will be built correlating the horizontal temperature gradient with the tangent altitude correction at the same altitude. A plot will contain as many points as many scans are measured along the considered orbit. The correlation between the considered quantities will be quantified by a linear correlation coefficient. The correlation will be considered significant if the linear correlation coefficient is different from zero consistently with its 1-σ error. In principle, as many plots as many are the sweeps in the standard scan can be built; in practice, only the plots corresponding to the lowest altitudes will provide significant information.
This test is successful if the tangent altitude corrections obtained in correspondence of the largest temperature gradients are consistent with the engineering pointing errors and no evident correlation between horizontal gradients and tangent altitude corrections arise from the analysis of the mentioned plots.
The aim of this test is to quantify systematic errors in pT retrievals due to assumption of hydrostatic equilibrium. The method will be based on comparisons of retrieved tangent height corrections with engineering pointing information. However no alternative methods will be developed to overcome hydrostatic equilibrium assumption. No re-iterations are foreseen for this test.

Required Tools
Statistical tool.

Required Auxiliary Data
No particular auxiliary data required.

Required Measurement Scenario
Standard measurement scenario.

Tool requirements arising from this procedure:

- The statistical tool shall be capable to build, for a user-defined altitude (sweep index) a scatter plot that correlates the horizontal temperature gradient (at the current altitude) with the tangent altitude correction. This plot shall contain as many points as many are the scans analyzed in the considered orbit. The statistical tool shall also be able to calculate both the linear correlation coefficient of the plotted data and its statistical error. It shall be possible to build such a plot for all the tangent altitudes of the standard scan.

4.2.4 Errors in the VMR profiles of interfering species

The ORM fits first pressure and temperature, then the VMRs of the target species in sequence. This means that in each retrieval the VMR of the interfering species, i.e. the species whose VMR is not fitted in the current retrieval, are assumed as known. Interference from lines of the non-target gases, as well as from lines of target species of previous retrievals, introduce a systematic error due to the imperfect knowledge of their VMR profiles.
The problem generates systematic errors in the simulations in correspondence of the spectral regions in which the contribution of the interfering species to the total emission is significant.

Tests
The following different procedures are planned to identify the problem:

1. The interference of water vapor lines in p,T microwindows will be analyzed in this test. After the completion of p,T retrieval and H_2O retrieval, p,T retrieval will be performed again with the new H_2O profile. This loop on p,T + H_2O retrievals will be stopped when the current retrieved profiles are consistent (within their error) with the profiles at the previous iteration. This procedure will be applied to the analysis of all the scans of the selected orbit. The output of this test is a recommendation on whether the tested iterative procedure should be introduced also in the operational processing or not.
2. REC analysis of the residuals.

Required Tools
- ORM_SDC
- Tool for REC analysis of the residuals

Required Auxiliary Data
The error spectra are required to perform test 2. (see beginning of Sect. 4.1).

Required Measurement Scenario
Standard measurement scenario.

Tool requirements arising from this procedure:
- The ORM_SDC shall be able to loop over p,T and H2O until the mentioned convergence criterion is met.

4.2.5 Errors in the spectroscopic parameters

As in the case of the VMR profiles of interfering species, the spectroscopic parameters are assumed to be known, therefore their errors cause systematic retrieval errors. The spectroscopic parameters associated with the most relevant spectroscopic parameters are the line strength, the line position and the pressure broadening coefficients. The line position error can also depend on pressure shift. Since pressure shift is only significant at very low altitudes that were not considered as the primary objective of MIPAS, pressure shift is presently neglected in the line-by-line calculations operated with the ORM. On the other hand, LUTs include pressure shift information only for lines where it is included in HITRAN96, i.e. only for some H_2O and CH_4 lines. Therefore this effect could show-up at very low altitudes even when using LUTs.
Errors in the spectroscopic parameters can show-up with characteristic shapes of residuals. For wrong line strength, the shape of residuals looks like a line shape, but this effect on the residual can be masked by compensations of the VMR. For wrong line position, the residuals have a first derivative shape and, for wrong line broadening the residuals have a second derivative shape.
Of course these are only the direct effects on individual residuals, in a retrieval these effects are masked by the concurring effects of other parameter errors.

Test

The following different procedures are planned:

1. statistical analysis of residuals must be performed to discover in the residuals possible patterns typical of spectroscopic errors. This type of analysis will be performed on residuals averaged over a TBD set of scans of the selected orbit. The patterns typical of spectroscopic errors will be identified using the following approach:
   - a microwindow will be selected at a given altitude
   - three scatter plots will be built correlating:
     a) residual with spectrum
     b) residual with first derivative of the spectrum with respect to frequency
     c) residual with second derivative of the spectrum with respect to frequency
   - Each plot will contain as many points as many are the sampled points of the selected microwindow. The correlation between the considered quantities will be quantified by a linear correlation coefficient. The correlation will be considered significant if the linear correlation coefficient is different from zero consistently with its 1-σ error. If a significant correlation exists, the considered microwindow must be excluded from list of the operational microwindows.
   - The above procedure will be repeated for all the used MWs.

   Note that, compared to the REC analysis of the residuals, this test permits to discriminate the various types of spectroscopic errors.
2. REC analysis of the residuals

In case of evidence of specific errors in the spectroscopic database, the spectroscopic data should be corrected if more accurate data are available (see WP 9421 of the CCN5).

Required Tools

- Statistical tool for the visualization of the residuals and cross-correlation plots for test 1.
- Tool for REC analysis of the residuals

Required Auxiliary Data

No particular auxiliary data required.

Required Measurement Scenario

Standard measurement scenario.

Tool requirements arising from this procedure:

- For a user-defined MW / altitude, the statistical tool shall be able to build the plots relating to the above test 1. and to calculate both the linear correlation coefficient of the plotted data and its statistical error.

4.2.6 Line-mixing

Line mixing corresponds to the deviation of measured line shape from the Voigt function. This effect occurs when collisions between radiating molecules and the broadening gas molecules of the same species cause the transfer of population between the rotational-vibrational states. This effect is neglected both in the spectroscopic model of the ORM and in the LUT calculation. Spectral regions affected by line mixing are avoided with an appropriate choice of microwindows.

The most evident effect of line mixing is a transfer of intensity from line wings to the line center.

Tests

The following tests are planned:

1. Critical evaluation of (averages of) residuals corresponding to MWs located in proximity of the CO_2 Q-branches, in which there is a reduction of the cross section. This test is considered successful if residuals corresponding to MWs located in the Q-branches are within the NESR. If this test is not successful, the output is a list of spectral regions in which the line-mixing effect has not been correctly modeled while calculating the error spectra used for MW selection.
2. REC analysis of the residuals

Required Tools
- Statistical tool
- Tool for REC analysis of the residuals

Required Auxiliary Data
No auxiliary data required.

Required Measurement Scenario
Standard measurement scenario.

Tool requirements arising from this procedure:
- The statistical tool shall be capable of visualizing the residuals for all microwindows and tangent altitudes. It shall be possible to analyze residuals averaged over a user-defined set of scans.

4.2.7 Field of View (FOV)

The effect of the input parameters that define the FOV shape and the vertical grid at which simulated spectra are calculated for the FOV convolution was already discussed in Sect. 4.1.9. A task deriving from those considerations is to check the vertical atmospheric gradients actually present in the real atmosphere measured by MIPAS. If these gradients are larger than those assumed for the worst-case simulations made before launch for tuning the FOV-related processing setup parameters, a re-tuning is needed.

Required Tools
- ORM_SDC

Required Auxiliary Data
Measured FOV patterns (for the different spectral bands) represented in the elevation scan angle domain.

Required Measurement Scenario
Standard measurement scenario.

Tool requirements arising from this procedure:
- The ORM_SDC shall be capable of handling a tangent-altitude- dependent FOV.

4.2.8 Apodized Instrument Line Shape (AILS)

Measured spectra are apodized with the Norton-Beer strong function before Level 2 processing in order to reduce the interference of far lines. The AILS is obtained by convoluting the measured ILS with the apodization function. The AILS, which is an input of the forward model, is assumed to be tangent altitude independent and does not take into account the instrument responsivity and phase error corrections as the retrieval is performed from calibrated and phase-error corrected spectra provided by Level 1b processing.
If the ILS assumed for the AILS calculation does not match the real ILS, the fit can reproduce the “area” of the measured features but not their shape. In particular, the problem will be evident in the

spectra relating to high tangent altitudes where other atmospheric "line shape" effects are less pronounced. Therefore, the presence of errors in the AILS modeling will be evident from the study of residuals at high altitudes. Also a significant bias in the altitude correction is expected to be caused by errors in the AILS (see results of tests with MIPAS-B2 data reported in [RD4]).

Tests
The following tests are planned:
1. REC analysis of the residuals. This analysis will eventually highlight a correlation between the residuals and the second derivative (wrt frequency) spectra.
2. fit of an additional parameter characterizing the instrument line shape

Test 2. will be done only if triggered by the results of test 1. Test 2. will quantify the error on the assumed ILS width and it can be considered successful if the retrieved value for the parameter characterizing the ILS is such that it leaves unchanged the ILS within the retrieval error associated to the parameter itself. If the retrieved parameter changes significantly the ILS shape, a recommendation will be issued to revise the procedure used in Level 1b to retrieve the ILS.

Required Tools
- ORM_SDC
- Tool for REC analysis of the residuals

Required Auxiliary Data
Test 1. requires the use of ILS error spectra

Required Measurement Scenario
Standard measurement scenario.

Tool requirements arising from this procedure:
- The ORM_SDC shall include the option of fitting a ILS-broadening parameter for test 2.

4.2.9 Frequency Calibration

Frequency calibration is performed in Level 1b in order to assign the correct frequency scale to the measured spectra. A possible imperfect frequency calibration in Level 1b introduces a systematic error in the observed spectra (i.e. all measured lines are shifted with respect to the simulated ones).

Tests
The following procedures are planned:
1. REC analysis of the residuals. This analysis will eventually highlight a correlation between the residuals and the first derivative (wrt frequency) spectra.
2. Fit of a frequency calibration parameter for each MIPAS spectral band, i.e. a factor scaling the frequency step.

Test 2. will be done only if test 1. highlights a significant correlation between the residuals and the first derivative spectra. Test 2. can be considered successful if the retrieved frequency scaling parameters (one for each spectral band) are equal to one within their retrieval error. If test 2. is not successful, a recommendation will be issued to revise the frequency calibration algorithm implemented in Level 1b.

Required Tools
- ORM_SDC
- Tool for REC analysis of the residuals

Required Auxiliary Data
Error spectra are required for test 1.

Required Measurement Scenario
Standard measurement scenario.

Tool requirements arising from this procedure:
- the ORM_SDC shall include an option for fitting a frequency scaling parameter different for each MIPAS spectral band, for test 2.

4.2.10 Intensity Calibration

The main effect of an error in the intensity calibration consists of a scaling factor applied to the spectrum. This scaling factor may depend on both the spectral band and on the direction of the interferometer sweep ("forward or "reverse"). This effect is visible through different residuals for different spectral bands and is more evident when saturated lines are considered. The impact of calibration error on the residuals is expected to be very similar to the effect of a temperature error.

Test
The following different procedures are considered for problem characterization:
1. REC analysis of the residuals. This analysis will highlight possible correlations between the residuals and the spectra themselves.
2. fit of an intensity scaling factor, one factor for each spectral band, different factors for "forward" and "reverse" sweeps.

Test 2. will be done only if test 1. highlights significant correlations between the residuals and the spectra. Test 2. can be considered successful if the value of the retrieved scaling factors are equal to one within the retrieval error associated to the scaling factors themselves. If test 2. is not successful, a recommendation will be issued to revise the intensity calibration algorithm implemented in Level 1b.
Note that an intensity calibration error can not be easily identified from test 2., because the intensity calibration is likely to be partially compensated by a change either in the fitted VMR or in temperature.

Required Tools
- ORM_SDC
- Tool for REC analysis of the residuals

Required Auxiliary Data
Error spectra are required for test 1.

Required Measurement Scenario
Standard measurement scenario.

Tool requirements arising from this procedure:
- the ORM_SDC shall include an option for fitting an intensity scaling parameter different for each MIPAS spectral band. Different scaling factors shall be used for "forward" and "reverse" sweeps.

4.2.11 Zero-level calibration

Causes of instrument zero level offset are internal emission of the instrument, scattering of light into the instrument or third order non-linearity of the detectors. All these causes of offset are corrected during the calibration step in Level 1b data processing.
In the ORM, a limb scanning angle independent offset is fitted for each microwindow in order to compensate for the residual uncorrected instrument offset. If the instrument has a limb angle dependent offset, the ORM corrects only partially for it.

An altitude dependent offset probably can not be seen in the residuals because cross talks are possible with intensity calibration errors and atmospheric continuum retrieval. The evidence is hidden in the inconsistency of the retrieved quantities.

Test
Fit of instrumental offset as a function of both tangent altitude and microwindow, for a subset of microwindows used in the retrieval (only for the MWs containing sufficient information to discriminate between offset and atmospheric continuum). This test is considered successful if the differences between the target profiles retrieved using a tangent altitude- independent and dependent offset are less than the retrieval errors of these quantities. If this test is not successful, a recommendation will be issued, suggesting to include in MIPAS Level 2 processor a functionality for fitting an instrumental offset both altitude- and MW- dependent (or a recommendation to improve Level 1b algorithm that corrects for the instrumental offset, in case not all the analyzed MWs contain information sufficient to allow the retrieval of a tangent height-dependent offset).

Required Tools
- ORM_SDC

Required Auxiliary Data
- None.

Required Measurement Scenario
Standard measurement scenario.

Tool requirements arising from this procedure:
- the ORM_SDC shall include the capability of fitting an instrumental offset dependent on both MW and tangent altitude (for a set of user-selected MWs).

4.2.12 Interpolation of the profiles

In the ORM the retrieved discrete values of the vertical profile are determined in correspondence of the so-called “retrieval grid”, equal to the grid of the measured tangent altitudes (or to a sub-set of them). Within these discrete points, an interpolated value of the profiles is computed in the forward model, whenever required. Since this retrieval grid is rather coarse ($\approx$3 km step), the choice of the most appropriate interpolation rule is critical.
The interpolation rules adopted in the ORM are the following ones: the independent variable is pressure, temperature and VMR profiles are linearly interpolated in log[p] (which roughly corresponds to linear interpolation in altitude). Temperature, pressure and altitude are constrained by the hydrostatic equilibrium.
Regarding continuum cross-section profiles, whenever an interpolation is required, a linear interpolation in pressure is used.
The retrieved profile above the highest tangent altitude is obtained scaling the initial-guess profile by the same quantity used for the highest fitted point; the same procedure is applied below the lowest fitted point. The interpolation rules have, in any case, some degree of arbitrariness and for this baseline we do not have a rigorous procedure that can be used to validate our choice. Therefore we only propose a test for the characterization of the impact of profiles extrapolation outside the measurement range (test 1.) and a verification procedure (test 2.) that is a minimum requirement for the retrieved profiles to allow intercomparisons with other measurements.

Tests
1. Errors due to extrapolation of the profiles in the altitude regions not explored by the scan:
   The newly defined standard measurement scenario includes also measurements with tangent altitudes as high as 68 km, therefore the extrapolation of the profiles above the highest tangent

altitude should have a very small effect. In any case, repeating the retrieval with initial guess profiles having different shapes above (below) the highest (lowest) tangent altitude of the scan will provide the final assessment of the error induced by profiles extrapolation. If available, the analysis of a measurement scenario extended to very high altitudes, such as the “Upper Atmosphere” scenario, would provide a reference for this test. This test is considered successful if the profiles retrieved using different profile shapes outside the scanned altitude range differ by a negligible amount compared to the profile total error. If this test is not successful the possibility of either further extending the standard MIPAS scan to high altitudes or to fit additional profile points above the highest tangent altitude of the scan must be considered.

2. “dynamics” special measurement scenario (when available): analysis of a scan that uses a elevation scanning step finer (e.g. 2 km) than that of the nominal scenario. Two independent retrievals will be carried-out using every other limb measurement and the resulting profiles compared.
   This test is successful if the profiles obtained from the two mentioned analyses (odd and even sweeps) are consistent (within the errors). This test is a minimum requirement for the retrieved profiles that will support intercomparisons with profiles derived from other instruments than MIPAS. If this test is not successful, the possibility of using a different interpolation scheme (e.g. exp(z) interpolation of VMR) in the operational Level 2 processor must be considered.

Remark: the “dynamics” special measurement scenario is not foreseen during the commissioning phase.

Required Tools
- ORM_SDC

Required Auxiliary Data
No auxiliary data required.

Required Measurement Scenario
“Upper Atmosphere” special measurement scenario, for test 1 (desirable, but not strictly required).
“Dynamic” special measurement scenario, for test 2.

Tool requirements arising from this procedure:
- None.

4.2.13 Continuum retrieval

Tests regarding the correctness of continuum retrieval have been cancelled. The test proposed in the draft version of the present document, with "empty" MWS specifically selected for continuum retrieval seemed an unpractical complication after discussion at PM #15 of the ORM study (see PO-MN-ESA-GS-1189). If the concern of the ORM team regarding the quality of continuum retrieval will persist, some tests will be done using "hand-made" MWs.

4.2.14 Initial Guess

For the analysis of a given scan, a first guess of the following atmospheric profiles must be supplied as input to the ORM: pressure, temperature, VMR profiles of target and interfering gases, continuum profiles for the microwindows used in both p,T and VMR retrievals.
These profiles are used in the different retrievals either as a first guess of the profiles that are going to be retrieved, or as assumed profiles of the atmospheric model (profiles of interfering species for all retrievals and p, T profiles in the case of VMR retrievals).
The Level 2 processor baseline is to use for T and VMR the ‘optimal estimation’ profile, obtained applying optimal estimation between ‘a-priori’ and ‘most recent measurement’. The initial guess

continuum profiles are always obtained only from the models, because continuum-related parameters are retrieved with very large errors.
The optimal estimation method consists in weighting the retrieved profile, with its VCM, with the 'a-priori' profile characterized by a "large" VCM. The optimal estimation of the profiles has to be determined not only at the beginning of each scan analysis, but also after each VMR retrieval, because the retrieved VMR profile is used as assumed profile in the subsequent retrievals. The errors to be associated with the VCM of the a-priori profiles must be determined on the basis of realistic estimates of the error affecting available a-priori profiles. These estimates are available from both ECMWF archives and from climatological studies (see e.g. WP5000 of CCN5 of the present study).
The correctness of this error estimate will be validated using the same approach adopted for the characterization of the errors on the engineering estimates of the tangent altitudes (see Sect. 4.1.11).
At a given tangent altitude *k* we assume:

$$\left(x_{ret}(k) - x_{ig}(k)\right)^2 = \sigma_{ret}^2(k) + \beta_k \cdot \sigma_{ig}^2(k) \qquad (3)$$

where x represents a profile, σ^2 are the variances and the subscripts '*ret*' and '*ig*' refer to the retrieved and the initial guess quantities respectively. β_k is a coefficient that depends on the considered tangent altitude *k* and whose expectation value is 1 if the (total) errors on both the initial guess and retrieved profiles have been correctly estimated.
For each tangent altitude *k*, the coefficient β_k will be calculated (from Eq. 3) for all the retrievals of the selected orbit and its average value $\overline{\beta}_k$ will be calculated. The quantity $\overline{\beta}_k$ will quantify the uncertainty related to the error associated with the a-priori estimates of the profiles.
Since only industry has the full visibility of the pre-processor function calculating the initial guess profiles, the validation procedure described above could be done directly by industry (TBC) with the support of the ESL team.

Required Tools:

- Statistical tool (TBC)

Tool requirements arising from this procedure:

TBC: The statistical tool shall be capable of plotting the distribution of β_k for all the tangent altitudes *k* of the (standard) scan and to display the values of $\overline{\beta}_k$.

5. Hierarchy of operations

The operations described in Sect.s 4.1 and 4.2 will be carried-out during the commissioning phase according to the sequence of operations proposed at the beginning of Sect. 4. In the present section we analyze the hierarchy of these operations and we associate to them a responsible person who will perform the test. Furthermore we estimate the time expected to be necessary for the individual operations.

5.1 Tuning of processing setup parameters: hierarchy

In principle, all processing setup parameters but the "measurement altitude range" and the "retrieval grid" can be tuned in parallel. The strategy is therefore first to tune these two parameters which have a significant cross-talk with the other parameters and secondly to optimize the other parameters. The convergence criteria will be optimized only at the final step, all the other processing parameters will be optimized using pre-defined conservative convergence criteria.
Table 1. shows, for each parameter to be tuned, the person who will perform the tuning operations and the number of working days expected to be necessary for the procedure.
The whole process of tuning processor setup parameters is expected to last 4 weeks.

Ref. Sect. in this doc.	Ref. to [RD8]	Tuned parameter	Who will do the tuning	Est. duration [work. days]
4.1.10	PS 2.5, 2.6	Measurement altitude range	P.Raspollini	5
4.1.6	PS 2.5, 2.6	Retrieval Grid	P.Raspollini	5
4.1.2	PS 2.5, 2.6	λ_{in}, λ^*, $\lambda/$, imxiterm	M.Ridolfi	5
4.1.3	PS 2.5, 2.6	Regularization parameters	B.Dinelli	5
4.1.4	PS 2.7	Eigenvalues thresholds	B.Dinelli	5
4.1.1	PS 2.5, 2.6 / AX 2.6	Continuum-related parameters	C.Piccolo	15
4.1.7	PS 2.5, 2.6	Linear-in-tau method	M.Carlotti	5
4.1.8	PS 2.5, 2.6	Tropopause altitude	P.Raspollini	5
4.1.9	PS 2.4	FOV-related parameters	S.Ceccherini	5
4.1.11	PS 2.5, 2.6 / AX 2.1	VCM LOS	M.Ridolfi	10
4.1.5	PS 2.5, 2.6	Convergence criteria	M.Ridolfi	5

Table 1: tuning of processing setup parameters

5.2 Tests for critical - Level 2 baseline verification: hierarchy

In principle, all the tests planned for critical – Level 2 baseline verification can be carried-out in parallel. Each test will provide information regarding the impact of a particular baseline on the accuracy of retrieval results, however at this stage no re-iteration of the tests is foreseen. Re-iterations can only take place at a higher level as explained at the beginning of Sect. 4.

Table 2. shows, for each test planned for critical-baseline verification, the type of inputs and outputs, the person who will perform the tuning operations and the number of working days expected to be necessary for the procedure.

In total, the operations required to carry-out the tests for critical – Level 2 baseline verification are expected to last 5 weeks.

Table 2 (1/3): Test of critical baselines, instrument effects

Tested baseline	Ref. to [RD8]	Test	Input[1]	Output	Duration [days]	Who
FOV	MV_2_11	Check of atmospheric gradients	I1	O1	5	S.Ceccherini
ILS	MV_2_12	Fit of a ILS parameter	I1 + I2	O2	5	P.Raspollini
		REC analysis	I1	O2	-	V.Jay
Freq. calib.	MV_2_15	Fit of a freq. Scaling parameter	I1 + I2	O3	5	V.Tenna
		REC analysis	I1	O3	-	V.Jay
Intensity calib.	MV_2_16	Fit of a intensity scaling parameter	I1 + I2	O4	5	M.Prosperi
		REC analysis	I1	O4	-	V.Jay
Instr. offset	MV_2_17	2D instr. Offset retrieval	I1	O5	5	M.Ridolfi

[1] See the list of various types of inputs and outputs reported after the table

Table 2 (2/3): Test of critical baselines, general systematic errors

Tested baseline	Ref. to [RD8]	Test	Input	Output	Duration [days]	Who
NLTE	MV_2_1	Day-night variability of residuals	I1	O6	5	V.Tenna
		Analysis of “upper atmosphere” scenario	I3	O6	10	P.Raspollini
		REC analysis	I1	O6	-	V.Jay
Refraction	MV_2_6	Recalculate engineering tang. alt. Using retrieved T	I1	O7	5	S.Ceccherini
Spectroscopic parameters	MV_2_8	Analysis of residuals	I1 + I2	O8	10	C.Piccolo
		REC analysis	I1	O8	-	V.Jay
Line mixing	MV_2_10	Analysis of the residuals in the Q-branches of CO_2, if included in the selected MWs	I1 + I2	O9	10	S.Ceccherini
		REC analysis of the residuals	I1 + I2	O9	-	V.Jay
Interpolation / extrapolation rules	?	Use IG profiles with different shapes	I1 or I4	O10	5	B.Dinelli
		Try scan with 2 km elevation steps	I4	O10	5	B.Dinelli

Table 2 (3/3): Test of critical baselines, localized effects

Tested baseline	Ref. to [RD8]	Test	Input	Output	Duration [days]	Who
Horizontal homogeneity	MV_2_2	Correlation between χ^2 and $\nabla T, \nabla H_2O, \nabla O_3$	I1	O12	5	M.Carlotti
		Comparison with external info	I1 + I5	O12	10	M.Carlotti
		REC analysis	I1	O12	-	V.Jay
Hydrostatic equilibrium	MV_2_3	Correlation between ΔZ_{RET} and ∇T	I1	O13	5	M.Ridolfi
Interfering species	MV_2_4	PT + H2O loop	I1	O14	10	B.Dinelli
		REC analysis	I1	O14	-	V.Jay
IG profiles	?	Characterization of used IG data	I1	O15	5	C.Piccolo

List of test inputs

I1 - MIPAS Level 1b spectra relating to a measured orbit
I2 - MIPAS Level 1b spectra obtained from averages of measurements (TBD which measurements must be averaged)
I3 - MIPAS Level 1b spectra relating to a measurement of an orbit of “upper atmosphere” scenario
I4 - MIPAS Level 1b spectra relating to a measurement of an orbit with fine (2km) elevation scanning steps (special measurement scenario “dynamics mode”)
I5 - ECMWF (T, H_2O and O_3) ‘external’ profiles

List of test outputs

O1 - If observed gradients are larger than those used for tuning FOV-related settings this test fails and a re-tuning of FOV-related parameters is needed
O2 - characterization of the error affecting the ILS as determined in Level 1b processor, recommendations to Level 1b expert teams
O3 - characterization of the frequency calibration error, possible recommendation to Level 1b expert teams
O4 - characterization of the intensity calibration error, possible recommendation to Level 1b expert teams
O5 - characterization of the Level 1b instrumental offset correction, possible recommendation to upgrade the Level 2 processor for fitting a both tangent altitude- and MW- dependent offset
O6 - characterization of LTE assumption, possible recommendation to discard (in the MW selection process) particular MWs affected by NLTE
O7 - characterization of the error affecting the retrieved tangent altitudes
O8 - characterization of spectroscopic error: list of spectral regions “potentially” affected by spectroscopic errors
O9 - characterization of the error introduced the Line Mixing model used for the generation of cross-section LUTs
O10 - characterization of the error due to interpolation / extrapolation of the atmospheric profiles, possible recommendation to change the interpolation scheme implemented in the ORM (Level 2 processor)
O11 - characterization of the quality of continuum retrieval, possible recommendations for retrieval algorithm improvements
O12 - characterization of the error due to horizontal homogeneity assumption in the retrieval, possible recommendation on how to overcome the horizontal homogeneity assumption in the Level 2 processor
O13 - characterization of the error introduced by hydrostatic equilibrium assumption on the retrieved profiles
O14 - characterization of the error due to contribution of interfering species in the selected MWs
O15 - characterization of the error affecting initial guess (and therefore also the a-priori) profiles.

4. Time chart

In the present section we report the overall time chart of the commissioning phase operations.

BAR CHART OF ORM VALIDATION ACTIVITIES

TASK	Man x days	TIME: weeks 1	2	3	4	5	6	7	8	9	10	11	12
DELIVERY OF LEVEL 1B DATA													
1. TUNING OF PROCESSING SETUP PARAMETERS													
1.1 Measurements altitude range	5												
1.2 Retrieval grid	5												
1.3 λ_{in}, λ^{*}, λ/, imxiterm	5												
1.4 Regularization parametrs	5												
1.5 Eigenvalues thresholds	5												
1.6 Continuum-related parameters	15												
1.7 Liner-in-tau method	5												
1.8 Tropopause altitude	5												
1.9 FOV-related parameters	5												
1.10 VCM LOS	10												
1.11 Convergence crtiteria	5												
OPTIMIZED SET OF SETUP PARAMETERS													
REFERENCE RETRIEVAL													
2. TESTS OF CRITICAL BASELINES													
2.1 REC analysis	10												
2.2. Test atmospheric gradients (FOV-related PS2)	5												
2.3 Fit of a ILS parameter	5												
2.4 Fit of a frequency scaling parameter	5												
2.5 Fit of a intensity scaling parameter	5												
2.6 2D instrument offset retieval	5												
2.7 NLTE day-night variability of residuals	5												
2.8 NLTE analysis of ‘upper atmosphere’ scenario	10												
2.9 Refraction	5												
2.10 Spectroscopic parameters: analysis of residuals	10												
2.11 Line mixing	5												
2.12 Try scan with 2 km elevation steps	5												

BAR CHART OF ORM VALIDATION ACTIVITIES

TASK	Man x days	TIME: weeks / days	1	2	3	4	5	6	7	8	9	10	11	12
2.13 Use IG profiles with different shapes	5													
2.14 Continuum: analysis of residuals	5													
2.15 Correlation between χ^2 and ∇T, ∇H_2O, ∇O_3	5													
2.16 Comparison with external info	10													
2.17 Hydrostatic equilibrium	5													
2.18 Interfering species: PT+H20 loop	10													
2.19 Characterization of used IG profiles	5													
3. CONTINGENCY FOR SPECIFIC ISSUES	40													
ORM VALIDATION AND CHARACTERIZATION														
4.PREPARATION OF ESA REPORT	20													
REPORT TO ESA AND MEETING														
5. CORRECTION OF ORM APPROXIMATIONS	40													
START OF POSSIBLE REITERATION														

8. Data exchange strategy

The tuning of processing setup parameters described in Sect. 4.1 and the tests for critical-baseline verification explained in Sect. 4.2 will be carried-out on the basis of the MIPAS measurements relating to one orbit. This orbit of data (Level 1b and Level 2 data) will be selected by the ESL team among a set of orbits supplied by ESA in Compact Disks (CDs). The Level 1b and Level 2 products will be read from the CDs and converted in the ORM format using alternatively:

1. An ESA-supplied tool to be installed on the SUN workstation available at IROE
2. An IROE-developed tool

Solution 1. is to be preferred even if its feasibility (computer requirements) must be assessed. Solution 2. May allow extra flexibility, but requires validation of the IROE tool against the ESA tool.

The tuning and test procedures identified in the present document will be carried-out by the ESL team using cross-section LUTs and irregular grids (IGs) optimized for accuracy of both "kept" and "skipped" spectral points in the selected MWs. This is because during the testing stage very important insights may arise from the inspection of the whole residuals in the MWs, and therefore also the accuracy of "skipped" points is important in this case. The ESL team also requires Level 2 data retrieved by the industrial prototype processor using the "conservative" set of LUTs and IGs mentioned above.

Appendix A: Definition of a quantifier for characterization of retrieval stability and convergence performance

In a stable retrieval, the chi-square decreases monotonically as a function of the iteration number an approaches asymptotically its minimum value. For practical purposes this behavior of the chi-square value can be assumed to be exponential:

$$\chi^2(i) = \left(\chi^2(0) - \chi^2(\infty)\right) \cdot \exp\left[-\frac{i}{\tau}\right] + \chi^2(\infty) \qquad \text{(A1)}$$

where:

i= iteration index,
$\chi^2(i)$= chi-square at iteration i
$\chi^2(0)$= initial chi-square
$\chi^2(\infty)$= asymptotic value of the chi-square
τ = chi-square life-time

If a retrieval is not "intrinsically" unstable, the chi-square reaches its minimum value after a finite number of iterations (not greater than 10 in the worst cases) therefore, both $\chi^2(\infty)$ and $\chi^2(0)$ are known. As long as $\chi^2(i) \neq \chi^2(\infty)$ (this condition imposes a threshold on the maximum value of i), at each iteration i we can calculate $\tau(i)$ from Eq. (A1) as:

$$\tau(i) = -i \cdot \left[\ln\left(\frac{\chi^2(i) - \chi^2(\infty)}{\chi^2(0) - \chi^2(\infty)}\right)\right]^{-1} \qquad \text{(A2)}$$

The behavior of $\chi^2(i)$ and of $\tau(i)$ are schematically represented in Fig. A1. The index i counts both Gauss and Marquardt iterations.

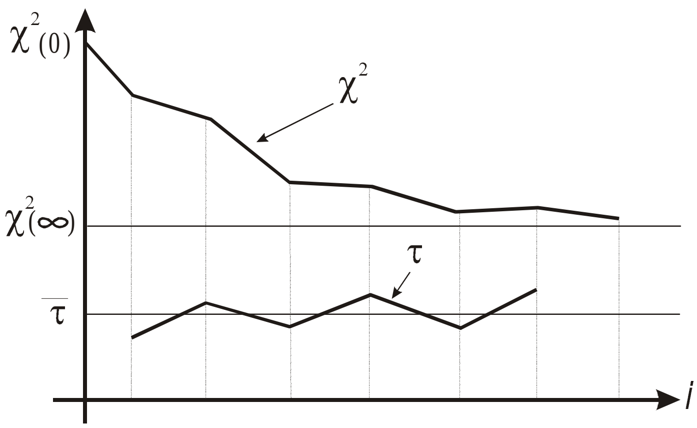

Fig. A1: schematic representation of the behavior of $\chi^2(i)$ and of $\tau(i)$ as a function of the iteration index i.

The "stability" of the retrieval, being connected with the monotonic behavior of the chi-square decay, is well quantified by the r.m.s. of $\tau(i)$ that will be indicated with σ_τ. The smaller is σ_τ the

more stable is the retrieval. The mean value of τ during the iterations (indicated with $\bar{\tau}$) quantifies the speed of the convergence. The smaller is $\bar{\tau}$ the faster is the convergence.
Therefore, a good merit indicator that quantifies at the same time both the speed and the stability of the retrieval is the quantity M defined as:

$$\mathrm{M} = \sigma_\tau \cdot \bar{\tau} \qquad \text{(A3)}$$

M can be defined as a function of the number of iterations. The number of iterations must be sufficiently small in order to avoid numerical problems in the calculation of $\tau(i)$ from Equ. (A2).
The smaller is M, the more fast and stable is the convergence.
Whenever a parameter controlling the strength of a constraint is to be tuned, the optimization process shall consider that the chi-square and the merit parameter M defined in Eq.(A3) behave as schematically shown in Fig. A2.

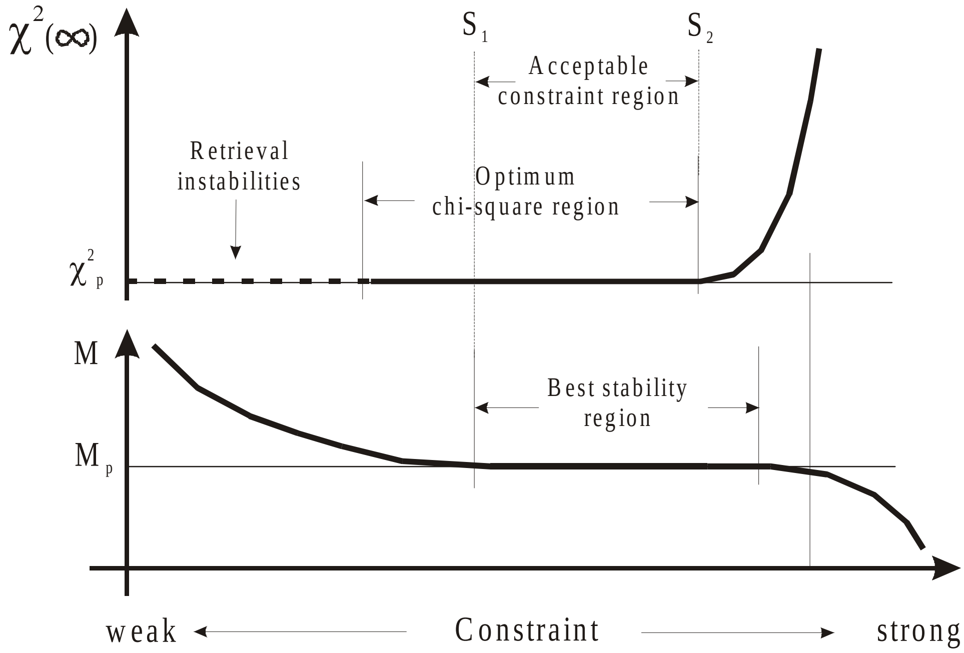

Fig. A2: Schematic representation of the behavior of $\chi^2(\infty)$ and M as a function of the strength of the constraint.

The optimum strength of the constraint must be chosen in a region in which the convergence is sufficiently stable and the final chi-square is the minimum attainable. By convention, we establish that this region is delimited by constraints of strengths S_1 and S_2 defined as follows:

- S_1 is such that a constraint with strength less than S_1 provides either $\chi^2(\infty) > 1.1 \cdot \chi_p^2$ or M > 1.1 M_p.
- S_2 is such that a constraint with strength greater than S_2 provides $\chi^2(\infty) > 1.1 \cdot \chi_p^2$.

Where χ_p^2 and M_p are defined in Fig. A2.
Whenever the strength of a constraint is to be optimized considering all the scans of a selected orbit, the plot of Fig. A2 must be constructed for each scan. S_1 and S_2 will fluctuate in the plots depending on the considered scan, however the regions spanned by S_1 and S_2 should always be separated by a region in which the strength lies of an optimum constraint suitable for all the scans of the considered orbit.

Appendix B: Summary table of the processing setup parameters to be tuned

Sect.	Par. name	Description	Units	Ref. val.	Min. val.	Max. val.
		Continuum constraints				
4.1.1	Rzc0	Altitude above which cont.=0	km	80	29	120
4.1.1	Rucl	Alt. above which cont. not fitted	km	36	26	68
4.1.1	Rconint	Umbrella radius (alt.,MW)	cm^{-1}	10	3	30
4.1.1	Rperc	Tight contiguity MWs	-	0.1	0.01	1
		Levenberg – Marquardt parameters				
4.1.2	Rlambdain	Initial Marquardt damping factor	-	0.01	0.001	1
4.1.2	Rlambdamult	Factor multiplying rlambda	-	10	1	100
4.1.2	Rlambdadiv	Factor dividing rlambda	-	10	1	100
		Regularization parameters				
4.1.3	Rl1	Global strength of regularization	-	5	0.1	100
4.1.3	Rd0_p	Diag. of P regul. Operator	-	2		
4.1.3	Rd0_t	Diag. of T regul. Operator	-	2		
4.1.3	Rd0_v	Diag. of VMR regul. Operator	-	2		
4.1.3	Rd0_c	Diag. of cont. regul. Operator	-	2		
4.1.3	Rd0_o	Diag. of offset regul. Operator	-	2		
4.1.3	Rd1_p	Off-diag. of P regul. Operator	-	-1		
4.1.3	Rd1_t	Off-diag. of T regul. Operator	-	-1		
4.1.3	Rd1_v	Off-diag. of VMR regul. Operator	-	-1		
4.1.3	Rd1_c	Off-diag. of cont. regul. Operator	-	-1		
4.1.3	Rd1_o	Off-diag. of offset regul. Operator	-	-1		
		Matrix inversion				
4.1.4	Dtineig	Threshold on eigenvalues	?	1.0d-20	1.0d0	1.0d-40
		Convergence criteria				
4.1.5	Rconvc(1)	Max diff between chi and chi_lin	-	1.05	1.001	1.5
4.1.5	Rconvc(2)	Max variation of T	K	0.1	0.01	2
4.1.5	Rconvc(3)	Max variation of P	-	0.005	0.001	0.1
4.1.5	Rconvc(4)	Max variation of VMR	-	0.005	0.001	0.1
		Retrieval grid				
4.1.6	Lfit	Retrieval grid	-	All “T”		
		Layering of atmosphere				
4.1.7	Rt1	Max T variation below rzt12	K	5	1	15
4.1.7	Rt2	Max T variation above rzt12	K	15	1	20
4.1.7	Rzt12	Alt. at which T thres. are exchanged	Km	56	0	120
4.1.7	Rhwvar	Max relative HW variation	-	1.2	1.001	2
		FOV convolution				
4.1.8	Rtropopause	Tropopause altitude	Km	14	5	50
4.1.9	Rint	Max t.a. dist between simul tropop				
4.1.9	Rintup	Max t.a. dist betw simul abv tropop				

….. continued

Sect.	Par. name	Description	Units	Ref. val.	Min. val.	Max. val.
Altitude range of the measurements used for the inversion						
4.1.10	?	Measurement altitude range	Tuned while building MW/OM data			
Algorithm used to build initial guess / assumed profiles (L2 pre-processor function)						
4.1.11	?	Error of a-priori profiles	%	100	10	1000
Tool for building the VCM of LOS engineering data						
4.1.12	Sig04s	Short term stability of pointing	km	0.115		
4.1.12	T04s	Time interval for short term stability	s	4.0		
4.1.12	Sig75s	Long term stability of pointing	km	0.33		
4.1.12	T75s	Time interval for long term stability	s	75.0		
4.1.12	Sig_tot	Max absolute pointing error	km	1.0		
4.1.12	Speed	Speed of the interferometer	cm/s	5.0		
4.1.12	T_ta	Turn-around time of the interferom.	s	0.5		