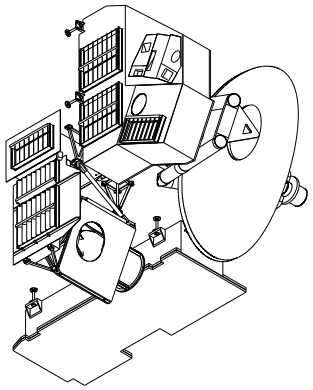
JPL D-16159 / CL #04-2043
EOS MLS DRL 601 (part 3)
ATBD-MLS-03

Earth Observing System (EOS)

Microwave Limb Sounder (MLS)

EOS MLS Retrieval Processes Algorithm Theoretical Basis



$$\mathbf{K}^{\mathrm{T}}\mathbf{S}_{\mathbf{y}}^{-1}\mathbf{K} = \begin{bmatrix} \times & \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & 0 & 0 & 0 \\ \times & \times & \times & \times & \times & 0 & 0 \\ \times & \times & \times & \times & \times & \times & 0 \\ \times & 0 & \times & \times & \times & \times & \times \\ \times & 0 & 0 & \times & \times & \times & \times \\ \times & 0 & 0 & 0 & \times & \times & \times \end{bmatrix}.$$

Nathaniel J. Livesey and W. Van Snyder

Version 2.0
July 19, 2004

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California, 91109-8099

Release Record

Version	Date	Comments
1.0	15 January 1999	Initial version.
1.1	15 October 1999	Released following formal review of Version 1.0 by NASA board reviewing the EOS CHEM Algorithm Theoretical Basis Documents. This document received top grade of 'A' from the review board. The board recommendations which apply to this document are italicized below, followed by the (non-italicized) responses of the MLS team. • *All aspects of the forward model should be examined...* As outlined in the overview document, a separate forward model ATBD is in preparation. • *A more complete and realistic evaluation of the retrieval algorithm CPU time requirements should be performed.* The concern during the review was the requirements for forward model CPU time, this will be discussed in the forward model ATBD. The CPU requirements for the inverse model are discussed in detail in this document. More detailed estimates will be given in later documents, based on early versions of the production code. No changes to the document were needed as a result of these recommendations. Changes from Version 1.0 are described below, reflect expected progress and correction of a few minor errors. • The 'secondary' geophysical products of Version 1.0 have now been defined as 'standard' data products. This resulted in updates to Table 1.1. • Also in Table 1.1 the product designation has been changed to use dashes rather than underscores, to reflect the implementation in software. • Added definition of g_0 in Section 3.4. • Modified Figure 3.6 to include effects of refraction. Caption modified accordingly. • Chapter on clouds largely rewritten to reflect new methodology. • Modified Section discussing data volumes as it is likely that the forward model radiance data will be produced and archived routinely. • Corrected Section C.2 to use $\log_2$ rather than ln. • Modified Table B.1 to use dashes rather than underscores, as in Table 1.1. • Minor grammatical corrections.

Version	Date	Comments
2.0	July 19, 2004	Updated to reflect the launch-ready state of the algorithms. The chapter on cloud handling was removed as that material is now covered in another document [Wu and Jiang, 2004]. (D. Wu removed as co-author as a result). Other significant changes: • Material on preconditioned conjugate gradient method removed as software now implements Cholesky decomposition. • Material on new method of processing chunks (looping over major frames) and Tikhonov smoothing added (Kronecker product material removed). • More details on optimization approach, expected χ^2 values, strategies for choosing Marquardt Levenberg parameters added. • More details of noisy products algorithm given. • Broad description of 'launch ready' configuration added to final chapter. Numerous other minor updates and changes made to reflect the algorithms as implemented.

Contents

Chapter 1
Introduction

1.1 The Microwave Limb Sounder experiment

EOS MLS is a successor to the MLS experiment [Waters et al., 1999; Barath et al., 1993] that formed part of the Upper Atmosphere Research Satellite (UARS), launched in September 1991 [Reber, 1993; Reber et al., 1993]. The instrument is designed to study aspects of the chemistry and dynamics of the atmosphere, from the upper troposphere to the mesopause. The microwave heterodyne technique is employed to observe thermal microwave emission from the Earth's limb in several spectral bands, designed to characterize emission from O_2 (used to obtain temperature and pressure information), O_3, H_2O, ClO, HCl, HNO_3, N_2O, CO, OH, SO_2, BrO, HOCl, HO_2, HCN and CH_3CN.

EOS MLS is one of four instruments on the EOS Aura spacecraft. The launch of the Aura platform is planned for July 2004. Aura will fly in a 98°inclined orbit, at a nominal height of 705 km, with a nominal orbital period of 98.9 minutes.

1.2 The aims of this document

This document describes the theoretical basis for the 'retrieval' algorithms to be used in the routine processing of data from the MLS instrument. The task of the retrieval algorithms is to convert calibrated MLS measurements of thermal limb emission into estimates of geophysical parameters such as temperature and composition.

The MLS calibrated radiance observations are known collectively as Level 1B data. Level 2 data typically describe vertical profiles of retrieved geophysical parameters at the nominal footprint of the MLS instrument. A description of the algorithms used to convert Level 1B to Level 2 data forms the bulk of this document. Level 3 data describe profiles of geophysical parameters on regular latitude/longitude grids. Most of the Level 3 products are generated by the application of gridding and/or averaging algorithms to the Level 2 data.

Some of the species MLS is designed to measure have very low mixing ratios and/or weak emission lines. This leads to a comparatively poor signal to noise ratio for the corresponding MLS radiance observations and resulting individual profiles of Level 2 data. The most scientifically useful observations of these 'noisy' products from MLS will be in the form of averages such as monthly mean maps or daily or monthly zonal means. These can be produced by appropriately averaging together individual Level 2 profiles. However, an alternative approach that amounts to directly retrieving these products from the Level 1B radiances can also be used. As this produces a Level 3 product it is strictly speaking part of the Level 3 processing, however, as it is based on retrieval calculations it is described in this document.

1.3 Related Algorithm Theoretical Basis Documents

An overview of the MLS instrument and data processing operations is given in Waters [2004]. The algorithms used in the Level 1 processing to calibrate the raw observations of microwave radiance made

Table 1.1: A list of the geophysical products produced by the algorithms described in the document. For a definition of the terms used see Waters [2004]. The products listed in bold type are the standard MLS data products, those in non-bold type are additional 'diagnostic' products produced by the algorithms. All the products, both standard and diagnostic, are produced using the same algorithms. The 'diagnostic' products are produced for special scientific and/or diagnostic purposes. Products marked † may be better processed using the 'noisy' products algorithm to produce Level 3 data for some or all of their vertical range, although Level 2 data for these will be produced routinely. Those denoted `[phase]` are named after their retrieval 'phase'. See section 6.4.2 for details. The SO_2 products, marked ‡, are not expected to be produced routinely by the software, unless significant volcanic activity has occurred and enhanced abundances are anticipated.

Geophysical products:

Temeprature	O3-640	**CO**
Temperature-[phase]	O3-2T5	CO-240
GPH	**HCl**	**HOCl**†
H2O	HCl-640	HOCl-640†
H2O-[phase]	**ClO**	**HCN**†
H2O-190	ClO-190	HCN-190†
RHI	ClO-640	**CH3CN**†
RHI-[phase]	**N2O**	CH3CN-190†
RHI-190	N2O-190	CH3CN-640†
HNO3	N2O-640	**SO2**‡
HNO3-190	**OH**†	SO2-190‡
HNO3-240	OH-2T5†	SO2-240‡
HNO3-640	**HO2**†	SO2-640‡
O3	HO2-640†	
O3-190	**BrO**†	
O3-240	BrO-640†	

by the instrument are described in Jarnot [2004]. An important component of the Level 2 processing algorithms is the *forward model*; the theoretical basis for this aspect of the data processing is given in Read et al. [2004], with additional information on special calculations required for the case of emission from mesospheric O_2 given in Schwartz [2004]. The impact of clouds on MLS observations and the use of MLS to measure cloud properties is described in Wu and Jiang [2004]. The gridding and averaging process used to convert most Level 2 data to Level 3 are given in Jiang [2004]. The precisions to be expected from the EOS MLS observations of geophysical parameters are given in Filipiak [2004].

1.4 EOS MLS data products for which this document applies

The algorithms described here are used in the production of all the EOS MLS Level 2 products except those related to cloud composition. For a complete list of the products see Table 1.1.

Chapter 2
Overview of EOS MLS Level 2 data processing

Most of this document describes the algorithms used in the production of the Level 2 data. This chapter gives a brief overview of the components of the algorithms and the software that implements them.

2.1 The aims of retrieval theory

The methods used to convert remote measurements of radiation emitted by the atmosphere into estimates of geophysical parameters are known as *retrieval algorithms*. Retrieval theory is a well-established field in atmospheric science and is covered in the standard literature [Rodgers, 1976, 1990, 2000]. A mathematical discussion of retrieval theory is given in Section 3.2. This section gives a brief outline of the fundamental principals involved.

In a retrieval algorithm a quantity known as the *state vector* is used to describe the current knowledge of the state of the atmosphere and relevant aspects of the instrumental calibration and state (known collectively as the measurement system). Typically the state vector is initialized with *a priori* information such as climatological datasets. Given the state vector, a *forward model* calculation can be applied to predict what radiances the instrument would observe, were the measurement system in the state described by the state vector. By comparing these predicted radiances with the radiances actually observed, and by making use of additional information provided by the forward model calculation (namely derivatives of radiance with respect to the state vector), the retrieval algorithm computes a better estimate of the state vector, i.e. one for which the predicted radiances will be closer to those observed. Often retrieval algorithms use an iterative approach to continuously refine the state vector until appropriate convergence has been achieved. For various reasons, including numerical stability, most retrieval algorithms include *virtual measurements*. These are typically *a priori* estimates of the contents of the state vector; usually, but not in all cases, the same estimates are used as initial values in the iterations.

2.2 Structure of the Level 2 data processing software

The main components of the Level 2 data processing software are shown in Figure 2.1. The principal components are the retrieval and forward model calculations described above. Before these are invoked, operational meteorological data are combined with climatological datasets and *a priori* knowledge of the state of the instrument, in order to construct an initial value for the state vector. In addition to the retrieval and forward model calculations, further processes produce other products (such as cloud parameters, tropopause pressure and stratospheric ozone column abundance). All these results are then output as Level 2 data along with appropriate diagnostics.

2.3 Output of Level 2 software

The most important output of the Level 2 software is the 'Level 2 Geophysical Product' (L2GP) files. These contain vertical profiles of retrieved atmospheric species (Temperature, ozone abundance etc.) on

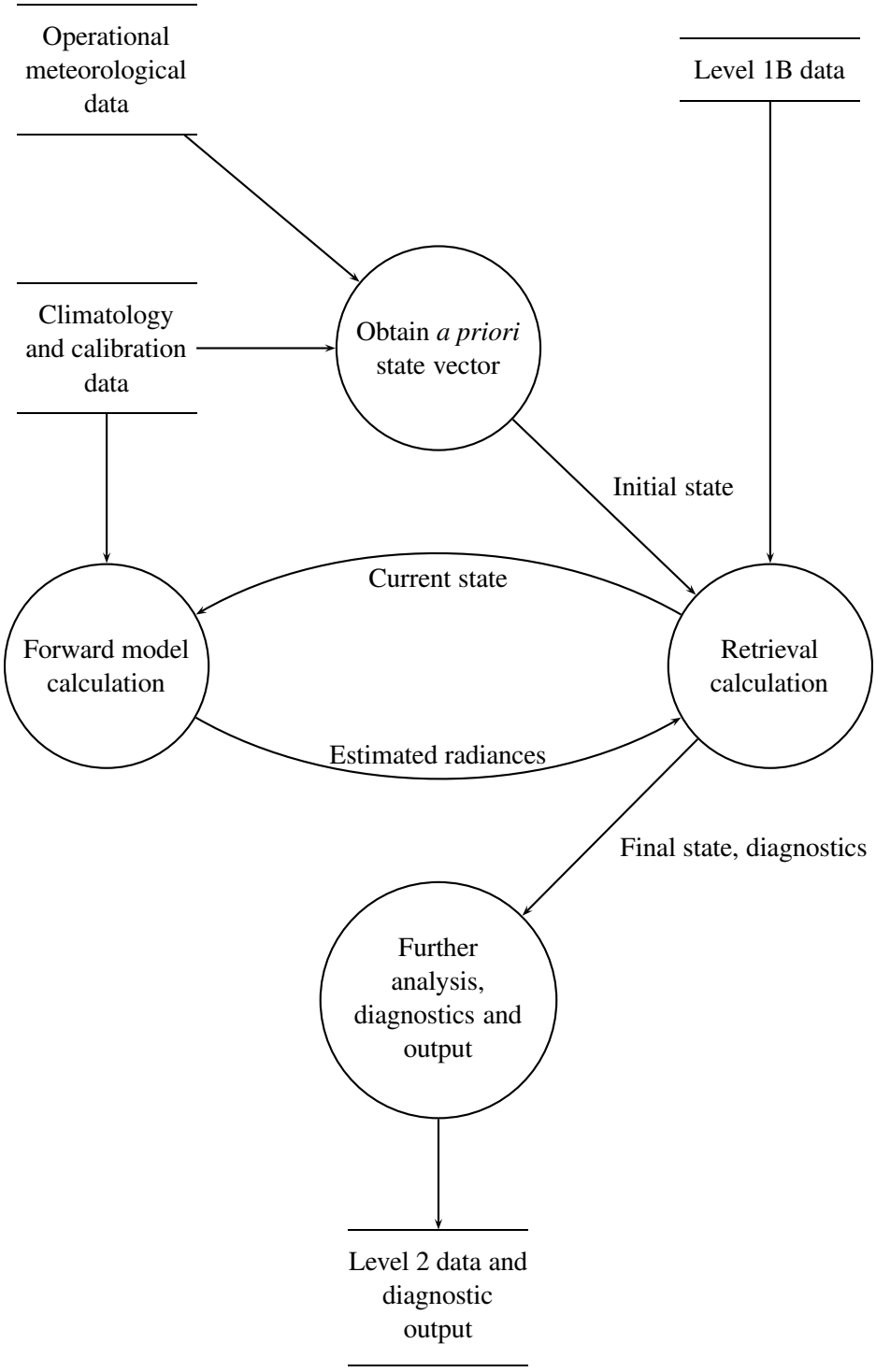


Figure 2.1: This figure shows the essential components and data flow for the EOS MLS Level 2 data processing software.

fixed pressure surfaces, spaced along the Aura orbit track, following the MLS measurement footprint. Each of the 'standard products' listed in bold in table 1.1 are output in a separate L2GP file. In addition, another L2GP file known as the 'DGG' (Diagnostic Products on a Geophysical Grid) file contains all the other 'diagnostic' products listed in table 1.1, along with other quantities that while not strictly describing geophysical parameters such as species abundances, are usefully stored on pressure surfaces like other L2GP quantities.

Another diagnostics file, the `L2AUX-DGM` file (Level 2 Auxiliary data on a Miscellaneous Grid), contains other diagnostic parameters, particularly those quantities that are stored on a 'minor frame' basis (see chapter 3 for more details of these). A similar file is the `L2AUX-CLD` file which contains some diagnostic information specific to MLS observations of cloud. The software also produces `L2FWM` files which contain radiance estimates from the forward model corresponding to the output data products. More details of the data files associated with each stage of the MLS data processing are given in Waters [2004].

2.4 Heritage of the MLS retrieval algorithms

The EOS MLS instrument is a successor to the MLS instrument that formed part of the Upper Atmosphere Research Satellite (UARS) mission, launched in September 1991. The data processing algorithms envisaged for EOS MLS have essentially the same theoretical basis as those implemented for UARS MLS (notably those used to produce the latest version, v5, of the UARS MLS dataset). For information on the earlier versions of the MLS retrieval algorithms see Fishbein et al. [1996]; Froidevaux et al. [1996] and Waters et al. [1996]. The version 5 data processing algorithms for UARS MLS are described in Livesey et al. [2003].

While the theoretical foundation of the EOS MLS retrieval algorithms is the same as that of the UARS algorithms, several aspects of the EOS MLS instrument design and intended scientific use differ from those of the UARS instrument, necessitating a somewhat different implementation of the algorithms.

The most significant of these differences is that the UARS MLS instrument observes limb emission in a direction perpendicular to the spacecraft flight direction, while the EOS MLS instrument observes emission from the region of the atmosphere directly ahead of the satellite. This geometry can, if properly exploited, yield significantly more information about the horizontal atmospheric variability along the measurement track. In order to capitalize on this information a very different approach to the retrieval process is required.

Another factor is that UARS MLS was designed to study processes mainly in the upper stratosphere. In recent years, scientific studies have become increasingly focused on the lower stratosphere and upper troposphere. The EOS MLS instrument targets this region specifically by the use of high bandwidth radiometers. Such observations, however, can be affected by the presence of clouds (though not to the same degree as observations made using infrared/visible techniques). In addition, in these regions, the retrieval problem is typically more non-linear than in the upper stratosphere, due to the increased optical depth of the atmosphere at the wavelengths under consideration. This increased non-linearity necessitates the use of more advanced forward model calculations than used for UARS MLS.

Chapter 3
The EOS MLS measurement system

This chapter outlines the physics behind the limb sounding technique and derives the algebraic expressions used in retrieval calculations. The aspects of the MLS instrument design that are relevant to the retrieval algorithms are summarized. The chapter then considers in detail how the state vector used in the MLS retrieval algorithms is constructed, and considers some implications of the MLS observation geometry.

3.1 The physics of limb sounding

Limb sounding of the atmosphere is a well established technique whereby the emission (or absorption in the case of occultation measurements) of electromagnetic radiation from the atmosphere at the limb of the earth is observed from a satellite platform. The radiance I observed by the instrument is given by the integral form of the radiative transfer equation

$$I = \int_t \int_\Omega \int_\nu A\left[\nu, \Omega(t)\right] \left\{ I_\infty(\nu, \Omega)\tau\left(\nu, \infty\right) + \int_{s=\infty}^{s=0} \tau\left(\nu, s\right) \frac{dB\left[\nu, T(s)\right]}{ds} ds \right\} d\nu \, d\Omega \, dt, \qquad (3.1)$$

where ν is frequency and Ω is solid angle, with $A\left(\nu, \Omega\right)$ describing the instrument's spectral and field of view response, and $\Omega(t)$ describing the movement of the MLS field of view as a function of time. $I_\infty(\nu, \Omega)$ represents the background emission, in this case the microwave background field. s is the distance along a given ray path, where the spacecraft is at $s = 0$; $T(s)$ is the atmospheric temperature along this ray path. $B(\nu, T)$ is the blackbody function, which describes the thermal emission of the atmosphere as a function of temperature and frequency (for this discussion local thermodynamic equilibrium has been assumed, and the effects of scattering have been neglected.) The quantity τ describes the transmission of the atmosphere from the point s to the spacecraft. This is defined by

$$\tau(\nu, s) = \exp\left[- \int_{s'=s}^{s'=0} k\left(\nu, T(s'), f(s')\right) \rho(s') \, ds' \right]. \qquad (3.2)$$

The quantity k is the absorption coefficient as a function of frequency, temperature and atmospheric composition described by the function $f(s')$. $\rho(s')$ is the atmospheric density.

3.2 Introduction to retrieval theory

Equation 3.1 gives an expression for the observed radiance as a function of the state of the atmosphere (i.e. its temperature and composition). The aim of the algorithms described in this document is to invert this calculation and obtain an estimate of the state of the atmosphere based on the observed radiances. Retrieval theory, the method by which these inverse calculations are constructed, has a great heritage in the remote sounding field (see Rodgers [1976, 1990, 2000]). The essential details of the subject, required for the MLS Level 2 algorithms, are given here.

3.2.1 The state vector, measurement vectors and χ^2

It is clearly impossible to invert Equation 3.1 to obtain the functional form of the atmospheric temperature and composition profiles, as this would involve obtaining infinite degrees of freedom from a finite series of measurements. The problem can only be solved if a *state vector* is used. The state vector $\mathbf{x}$ is an n element vector that describes aspects of the atmosphere and measurement system that affect the radiance measurements. Typically the state vector contains profiles of temperature and composition represented by a finite set of vertical levels. By describing the state using a vector with a finite number of elements, the retrieval task has been made tractable.

Measurements are also grouped into vectors, $\mathbf{y}_i$. Multiple vectors are used, as this explicitly indicates which sets of measurements are independent (i.e. have no covariance), and which are interdependent. For example, it is possible that radiometric noise may be correlated from channel to channel within an MLS spectral band, but not between radiometers. In such a case, one could choose to use separate vectors to represent the radiances from each band. Further reasons for this separation will become clearer when we introduce the concept of *virtual measurements* below. The covariance of the measurement vectors is represented by the matrix $\mathbf{S}_i$. In many cases, the measurement covariance matrices are purely diagonal. While it would be possible thus to split the corresponding measurement vectors up into separate one-element vectors, the grouping will typically be maintained for clarity. The computer programs that implement the retrieval algorithm ensure that unnecessary calculations are avoided in these cases, by considering only the diagonal elements of the matrix.

A key component of the retrieval algorithm is the *forward model*, a calculation that gives an estimate of the radiances that would be observed by the instrument, were the atmosphere in the state give by $\mathbf{x}$.

$$\hat{\mathbf{y}}_i = \mathbf{f}_i(\mathbf{x}), \tag{3.3}$$

where $\hat{\mathbf{y}}_i$ denotes an estimate of the vector $\mathbf{y}_i$. These functions are typically discrete forms of the radiative transfer integration in Equation 3.1. The aim of retrieval theory is to seek a value of the state vector that is a 'best estimate' of the true state of the atmosphere. The 'best estimate' is defined as that which most appropriately fits the observed measurements, by giving the minimum value of the quantity χ^2, defined by

$$\chi^2 = \sum_i \left[\mathbf{y}_i - \mathbf{f}_i(\mathbf{x})\right]^{\mathrm{T}} \mathbf{S}_i^{-1} \left[\mathbf{y}_i - \mathbf{f}_i(\mathbf{x})\right]. \tag{3.4}$$

This expression is simply the vector form of the more familiar definition of χ^2 in the scalar case as $\sum_i ([o_i - p_i]/s_i)^2$, where o is the observed measurement, p is the prediction, and s is the standard deviation of the uncertainty in the observed data. In this case, the differences between the measured and fitted observations are weighted according to their covariances, and summed.

3.2.2 Gauss-Newton iteration

Many different techniques exist for finding the minima of quantities such as χ^2. The most commonly used method in retrieval algorithms is the Gauss-Newton minimization, which gives an iterative expression for the best estimate of $\mathbf{x}$ according to

$$\mathbf{x}^{(r+1)} = \mathbf{x}^{(r)} - \left[\nabla^2\chi^2\right]^{-1} \nabla\chi^2, \tag{3.5}$$

where ∇ is the vector derivative operator

$$[\nabla]_i = \frac{\partial}{\partial x_i}, \tag{3.6}$$

and ∇^2 is the matrix second derivative operator

$$\left[\nabla^2\right]_{ij} = \frac{\partial^2}{\partial x_i \partial x_j}. \tag{3.7}$$

Note that much of the retrieval theory literature refers to the retrieved estimate of the state vector as $\hat{\mathbf{x}}$, with $\mathbf{x}$ describing the unknown true state of the atmosphere. However, in this document, due to the large number of subscripts, superscripts and diacritics that $\mathbf{x}$ will soon gain, the 'hat' has been omitted for clarity. The exception is in Section 3.2.6 where the distinction between $\mathbf{x}$ and $\hat{\mathbf{x}}$ is important.

The parenthetical superscripts indicate values from different iterations. The vector calculus identities shown in Appendix C.1, when applied to Equation 3.4 give

$$\nabla\chi^2 = -2\sum_i \mathbf{K}_i^{\mathrm{T}}\mathbf{S}_i^{-1}\left[\mathbf{y}_i - \mathbf{f}_i(\mathbf{x})\right] \tag{3.8}$$

$$\nabla^2\chi^2 = 2\sum_i \mathbf{K}_i^{\mathrm{T}}\mathbf{S}_i^{-1}\mathbf{K}_i\,, \tag{3.9}$$

where

$$\mathbf{K}_i = \frac{\partial \mathbf{f}_i(\mathbf{x})}{\partial \mathbf{x}} \tag{3.10}$$

are known as the matrices of *weighting functions* (in some literature referred to as the *Jacobians*) for the measurement vectors.

The Gauss-Newton iteration can thus be expressed as

$$\mathbf{x}^{(r+1)} = \mathbf{x}^{(r)} + \left[\sum_i \mathbf{K}_i^{\mathrm{T}}\mathbf{S}_i^{-1}\mathbf{K}_i\right]^{-1} \sum \mathbf{K}_i^{\mathrm{T}}\mathbf{S}_i^{-1}\left[\mathbf{y}_i - \mathbf{f}_i(\mathbf{x}^{(r)})\right]. \tag{3.11}$$

The covariance matrix of the solution can be shown to be given by

$$\mathbf{S}_{\mathbf{x}} = \left[\sum_i \mathbf{K}_i^{\mathrm{T}}\mathbf{S}_i^{-1}\mathbf{K}_i\right]^{-1}. \tag{3.12}$$

Note that higher order terms (i.e. those involving $\partial\mathbf{K}_i/\partial\mathbf{x}$) are neglected here. This approximation is generally acceptable except in situations known as 'large residual' problems, [Rodgers, 2000, for example] where the full *Newtonian* iteration method is required.

3.2.3 The need for virtual measurements

The matrix to be inverted in equation 3.11 is often known as the matrix of *normal equations* – a term arising from the 'least squares' family of problems. In many cases this inversion is impossible as the normal equation matrix is singular. This indicates that the 'direct' measurements (radiances etc.) have provided insufficient information to completely determine the state vector; there are some components (or, more correctly, eigenvectors) of the system about which no information has been obtained.

One solution to this problem is to introduce *virtual measurements*. These are additional measurement vectors included in the retrieval calculation in order to assure successful matrix inversion, and to ensure reasonable values for comparatively poorly-measured aspects of the system. In the MLS case, as is typical, these virtual measurements take the form of *a priori* estimates of the state vector or individual components of the state vector, constructed from datasets such as climatologies. The covariance of the *a priori* information is often chosen so as to limit the amount of bias it introduces in the resulting state vector.

Introducing *a priori* information in this manner can sometimes lead to incorrect interpretation of retrieved results. The covariance of the retrieved state vector should always be compared with the *a priori* covariance; if only a small amount of error reduction has been achieved, this indicates that the direct measurements (i.e. radiances etc.) have failed to contribute significant information to the knowledge of the state vector.

In order to simplify most later expressions in this document, the *a priori* information is explicitly separated from the other measurement vectors. The *a priori* state vector is denoted by $\mathbf{a}$, with covariance given by the matrix $\mathbf{S_a}$. The forward model for this quantity is simply $\mathbf{f}_a(\mathbf{x}) = \mathbf{x}$ giving a corresponding $\mathbf{K}$ matrix equal to the identity. This gives a modified form of Equation 3.11 as

$$\mathbf{x}^{(r+1)} = \mathbf{x}^{(r)} + \left[\mathbf{S_a}^{-1} + \sum_i \mathbf{K}_i^{\mathrm{T}}\mathbf{S}_i^{-1}\mathbf{K}_i\right]^{-1} \left\{\mathbf{S_a}^{-1}\left[\mathbf{a} - \mathbf{x}^{(r)}\right] + \sum_i \mathbf{K}_i^{\mathrm{T}}\mathbf{S}_i^{-1}\left[\mathbf{y}_i - \mathbf{f}_i(\mathbf{x}^{(r)})\right]\right\}. \quad (3.13)$$

The solution covariance in this case is given by

$$\mathbf{S_x} = \left[\mathbf{S_a}^{-1} + \sum_i \mathbf{K}_i^{\mathrm{T}}\mathbf{S}_i^{-1}\mathbf{K}_i\right]^{-1}. \quad (3.14)$$

In some retrieval situations there are elements of $\mathbf{x}$ for which the use of an *a priori* as a virtual measurement is inappropriate; for these elements, the corresponding rows and columns of $\mathbf{S_a}^{-1}$ are set to zero[1]

3.2.4 Use of virtual measurements to 'smooth' retrieved profiles

Covariance matrix approaches

In addition to the diagonal elements of the $\mathbf{S_a}$ matrix which specify the *a priori* variance of each element of the state vector, it is possible to use off-diagonal terms to describe correlations to be expected between adjacent elements (e.g., vertical levels). Terms such as these can be used to favor solutions where retrieved atmospheric profiles are smoother. This can be desirable when the vertical levels in the state vector are finer than the effective vertical resolution of the measurement system. In such cases the observed radiances may be equally well fitted with either smooth or oscillatory profiles. Smoothing terms such as these can be used to favor profiles that are arguably a more realistic reflection of the true atmospheric state.

Often off diagonal terms are of the form

$$\left[\mathbf{S_y}\right]_{ij} = \sqrt{\left[\mathbf{S_y}\right]_{ii}\left[\mathbf{S_y}\right]_{jj}} \exp\left(-\frac{\left|z_i - z_j\right|}{h}\right), \quad (3.15)$$

where i and j are indices of related elements of the state vector (for example two points in the temperature profile). z_i and z_j are their heights in some appropriate unit, and h is some characteristic length scale in that unit. Such terms arise from considering the profiles as Markov chains and amount to a constraint on the first derivative of the profile with respect to z. Such terms were used in the version 5 algorithms for the UARS MLS data processing.

[1] Note that strictly speaking, this makes $\mathbf{S_a}^{-1}$ singular, but it is never directly needed in any of these algorithms.

The Tikhonov approach

For EOS MLS a different smoothing constraint is typically imposed. The Tikhonov method [Tikhonov, 1963] can be used to impose constraints on higher order derivatives of profiles. Let us consider a simple case where the n element state vector $\mathbf{x}$ describes a single vertical profile of one species (e.g. Temperature) on evenly spaced levels. The 'second order' Tikhonov method constructs an $(n-2) \times n$ matrix $\mathbf{P}$ as

$$\mathbf{P} = \begin{bmatrix} -\frac{1}{4} & \frac{1}{2} & -\frac{1}{4} & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{4} & \frac{1}{2} & -\frac{1}{4} & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -\frac{1}{4} & \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & -\frac{1}{4} & \frac{1}{2} & -\frac{1}{4} \end{bmatrix} \tag{3.16}$$

The $n-2$ element vector $\mathbf{d} = \mathbf{Px}$ is then a measure of the second derivative of $\mathbf{x}$ with respect to level (i.e., approximate height, though it is affected by the vertical spacing). We can use this to construct another virtual measurement that 'the second derivative of $\mathbf{x}$ with respect to height should be small'. This adds terms such as $\mathbf{x}^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}\mathbf{Px}$ to the expression for χ^2. Typically some weighting of these terms is required, perhaps one that varies with height (i.e., level). This can be obtained by premultiplying $\mathbf{P}$ by a diagonal $(n-2) \times (n-2)$ matrix $\mathbf{W}$ to get a scaled 'regularization' matrix $\mathbf{R}$ (an alternative viewpoint is that the matrix $\mathbf{W}^{\mathrm{T}}\mathbf{W}$ is the error covariance for the virtual measurement that $\mathbf{d} \simeq 0$). In addition, it is often preferable to impose the smoothness constraint on the deviations of the state vector from the *a priori* vector, rather than on the state vector itself.

The Gauss-Newton iteration then becomes

$$\mathbf{x}^{(r+1)} = \mathbf{x}^{(r)} + \left[\mathbf{S}_{\mathbf{a}}^{-1} + \mathbf{R}^{\mathrm{T}}\mathbf{R} + \sum_i \mathbf{K}_i^{\mathrm{T}}\mathbf{S}_i^{-1}\mathbf{K}_i \right]^{-1} \left\{ \mathbf{S}_{\mathbf{a}}^{-1}\left[\mathbf{a} - \mathbf{x}^{(r)}\right] + \mathbf{R}^{\mathrm{T}}\mathbf{R}\left[\mathbf{a} - \mathbf{x}^{(r)}\right] + \sum_i \mathbf{K}_i^{\mathrm{T}}\mathbf{S}_i^{-1}\left[\mathbf{y}_i - \mathbf{f}_i(\mathbf{x}^{(r)})\right] \right\}, \tag{3.17}$$

with the solution covariance now given by

$$\mathbf{S}_{\mathbf{x}} = \left[\mathbf{S}_{\mathbf{a}}^{-1} + \mathbf{R}^{\mathrm{T}}\mathbf{R} + \sum_i \mathbf{K}_i^{\mathrm{T}}\mathbf{S}_i^{-1}\mathbf{K}_i \right]^{-1} \tag{3.18}$$

The Tikhonov method can easily be extended to higher order derivatives, and to state vectors containing heterogeneous data (e.g., temperature and composition), where constraints are imposed on individual species, but not between species. The $\mathbf{P}$ matrix for the qth derivative is $(n-q) \times n$, and is formed from the normalized order q binomial coefficients.

Section 4.2.6 describes the application of Tikhonov constraints to the MLS retrieval system, in particular its implementation in both the horizontal and vertical directions.

3.2.5 Retrieval phases and errors on constrained quantities

In many cases, the retrieval calculations are performed in separate phases, with the results from one phase being used in the forward model calculations for later phases. For example, in the UARS MLS data processing algorithms, a retrieval of temperature and tangent point pressure was obtained from the 63 GHz O_2 radiances. These results were then used in retrievals of the constituent information from the other spectral bands.

However, when performing a retrieval calculation in separate phases, the measurement covariance matrices $\mathbf{S}_i$ should be modified to allow for the fact that there are uncertainties in the knowledge of the previously-retrieved quantities. In the case of the example above, the error estimates for the radiances in, say, the ozone band should be inflated to account for our uncertainty in the previously retrieved temperature and tangent pressure. The modification should be made according to

$$\mathbf{S}_i \rightarrow \mathbf{S}_i + \mathbf{K}_{i[\mathbf{c}]}\mathbf{S}_\mathbf{c}\mathbf{K}_{i[\mathbf{c}]}^{\mathrm{T}}, \tag{3.19}$$

$\mathbf{S}_\mathbf{c}$ describes the covariance of the quantities $\mathbf{c}$ that were previously retrieved (i.e. the covariances obtained from Equation 3.18,) and the matrices $\mathbf{K}_{i[\mathbf{c}]}$ are the weighting functions for these quantities, such that

$$\left[\mathbf{K}_{i[\mathbf{c}]}\right]_{\alpha j} = \frac{\partial\left(\mathbf{y}_i\right)_\alpha}{\partial \mathbf{c}_j}, \tag{3.20}$$

where α and j are matrix and vector indices. This calculation is sometimes referred to as *constrained quantity error propagation*.

As described in Section 3.2.1, in many retrieval problems the measurement covariance matrices are diagonal (or can be assumed to be diagonal to a reasonable level of approximation.) This is the case for most of the MLS spectral bands (see the discussion in Sections 3.5.3 and 3.5.4). If a diagonal covariance matrix can be assumed, then computation time can be saved by optimizing the algorithm to take advantage of this fact. However, a constrained quantity error propagation calculation will typically produce a new set of measurement covariance matrices that are not diagonal. As these measurement covariance matrices have to be inverted as part of Equations 3.17 and 3.18, this can represent a significant amount of computational effort. In particular, in the case where the number of measurements greatly exceeds the size of the state vector (as in the MLS case), it is generally preferable to avoid constrained quantity error propagation. Instead, the most efficient approach is to retrieve the state of the entire system simultaneously, using all available measurements. In this manner, the measurement covariance matrices $\mathbf{S}_i$ remain diagonal. This does not preclude the use of phasing, however. Phases can be implemented in a different manner, such that more quantities are added to the state vector in each phase, rather than considering a completely different set of quantities each phase. Sections 4.4.2 and 6.4.2 deal with these issues in more detail.

3.2.6 Diagnosing retrieval performance

Averaging Kernels

When examining the results of a retrieval calculation it is important to check the retrieved error estimates and compare them with any *a priori* information as outlined in Section 3.2.3. One way to do this comparison is to look at the *averaging kernel* matrix. This is defined as

$$\mathbf{A} = \frac{\partial\hat{\mathbf{x}}}{\partial\mathbf{x}} = \left[\mathbf{S}_\mathbf{a}^{-1} + \mathbf{R}^{\mathrm{T}}\mathbf{R} + \sum_i \mathbf{K}_i^{\mathrm{T}}\mathbf{S}_i^{-1}\mathbf{K}_i\right]^{-1} \sum_i \mathbf{K}_i^{\mathrm{T}}\mathbf{S}_i^{-1}\mathbf{K}_i, \tag{3.21}$$

where for the sake of this discussion, $\mathbf{x}$ is the true state vector, and $\hat{\mathbf{x}}$ is the retrieved state from Equation 3.17. $\mathbf{A}$ describes the sensitivity of the retrieval to the true atmospheric state, as opposed to its sensitivity to the virtual measurements. One could consider it as the 'ratio' between the information contributed by the direct measurements alone and the total contributed by both the direct and virtual measurements.

Columns of the $\mathbf{A}$ matrix represent the response of the retrieval system to a 'delta function' disturbance in the atmosphere (i.e. a change in a single element of $\mathbf{x}$.) Rows of the matrix indicate the amount

each element of the true state vector has contributed to the retrieved estimate. The form of the averaging kernel rows provides a very useful description of the measurement system, and quantities such as the full width at half maximum of these rows are often invoked as metrics of resolution.

The use of χ^2 as a diagnostic

In addition to examining these quantities, the χ^2 value for a retrieval (or for each measurement vector independently) should also be examined. Ideally, the value of χ^2 should be comparable to the total number of real measurements. It is often divided by this number to yield a number that should be around one.

A significantly large value of χ^2 indicates that the radiance measurements have not been fitted to a sufficient level of accuracy. This can be due to errors in the forward model, or poor convergence in the retrieval algorithm. A significantly small χ^2 on the other hand usually indicates that the measurement precisions used are too pessimistic.

3.2.7 Other minimization techniques

In many cases, (e.g. for some of the MLS observations) the retrieval calculation is sufficiently linear that a small number of Gauss-Newton iterations can yield the correct result. However, in some cases, such as those where the system is moderately non-linear, and the initial value of the state vector is comparatively far from the solution, other techniques may be more appropriate.

The steepest descent approach is a simple algorithm that makes small steps each iteration in the direction of the steepest descent of the cost function (χ^2 in this case.)

$$\mathbf{x}^{(r+1)} = \mathbf{x}^{(r)} - \lambda^{-1} \nabla \chi^2, \tag{3.22}$$

where λ is a scalar value describing the size of step to be taken. This is typically a slow algorithm, as it takes no advantage of possible linearity in the system.

The Marquardt-Levenberg approach is faster, as it is a combination of the 'cautious' steepest descent method and the 'aggressive' Gauss-Newton method. As the iterations proceed, and the solution is approached, the steps taken each iteration typically become smaller, making linearity an increasingly better assumption, and allowing the minimization to become more aggressive. The iterations proceed according to

$$\mathbf{x}^{(r+1)} = \mathbf{x}^{(r)} - \left[\lambda \mathbf{I} + \nabla^2 \chi^2\right]^{-1} \nabla \chi^2. \tag{3.23}$$

For small values of λ this is equivalent to the Gauss-Newton iteration in Equation 3.5, while for large values of λ this is equivalent to steepest descent with a small step size.

Before a retrieval step is taken, it is possible to predict the value χ^2 would have at the new location, were the system truly linear (see Appendix C.3). A comparison of this value with that obtained once the step is taken yields a good measure of the non-linearity of the system. This information can be taken into account when choosing a value of λ for the next iteration. More details of this are given in Appendix F.

3.3 MLS Radiance measurements

The MLS instrument makes observations of microwave radiation in many different regions of the spectrum, in the frequency range from 118 GHz to 2.5 THz. The instrument consists of seven microwave radiometers, covering five different spectral regions. The signals from the radiometers are passed on to various spectrometers. The spectral coverage of the instrument is shown in Figure 3.1. In the MLS data

processing, radiance is measured in Kelvins and considered to be a *brightness temperature*. This is a quantity proportional to the observed radiance, which – in the long wavelength limit – is equal to the temperature a blackbody filling the MLS field of view would need to have to produce the same radiance.

In the standard operational mode, the instrument makes one complete vertical scan of the GHz and THz antennae over tangent heights between the surface and 95 km in ∼20 s (the GHz and THz scan rates differ somewhat to allow the THz antenna to spend more time observing the upper stratosphere than the GHz). Approximately 4.7 s are spent in calibration and antennae retrace activities, giving a repeat period of about 24.7 s. Each scan/calibrate/retrace activity is called a *major frame*. The 20 s limb scan is a continuous movement, as opposed to the 'stop and stare' scan that was used in UARS MLS. During the scan, 120 radiance integrations are performed, each of length ∼1/6 s. These integration periods are known as *minor frames*. The durations of the major and minor frames vary slightly to ensure that the latitudinal distribution of the scans is the same from orbit to orbit.

3.3.1 Behavior of the radiances

Figure 3.2 shows a set of calculated radiances from two of the EOS MLS spectral bands. The form of the radiance curves shown in the left plot is typical of the observations from limb sounding instruments. At high tangent point altitudes (the altitude of the point along the ray path closest to the Earth's surface), the atmospheric density along the limb path is very low, so little emission is observed. As the ray path descends through the atmosphere, emission becomes stronger as the atmosphere becomes thicker. Eventually, the atmosphere becomes sufficiently opaque that emission from lower regions in the atmosphere is absorbed by air at higher altitudes, and is thus never observed by the instrument. In these circumstances the radiances are said to be *saturated* or *blacked out*; this is the cause of the knee in the radiance curves. The saturated radiances are a measure of the temperature of the region of the atmosphere where the saturation takes place. Sometimes, the radiances continue to increase or decrease slightly as the tangent ray path is scanned further down. This is a geometrical effect. As the ray descends, the path length to a given height decreases; thus the saturation occurs at a different height in the atmosphere, leading to slightly different radiances, depending on the form of the temperature profile. The channels closer to the line centers saturate at higher altitudes than those on the wings of the lines, as the absorption at the frequencies closer to the line centers is stronger.

The curves shown in the right plot in Figure 3.2, corresponding to the 183 GHz H_2O observations, show a slightly different behavior. In this case two separate saturation process occur. This is simply explained by the fact that these observations are made by a 'double sideband' radiometer. The MLS radiometers work on the microwave heterodyne technique; they output an *intermediate frequency* (IF) signal, corresponding to the observed signal with the *local oscillator* (LO) frequency subtracted. The negative frequency components are folded over into positive IF space with a 180° phase shift. This is described in Figure 3.3. The intermediate frequency signal is given by

$$\langle \textit{Intermediate frequency signal} \rangle = \alpha \, \langle \textit{Upper sideband signal} \rangle + \beta \, \langle \textit{Lower sideband signal} \rangle \, . \quad (3.24)$$

The *sideband fractions*, α and β are measured during prelaunch calibration.

3.3.2 The importance of tangent pressure

As shown in Figure 3.2, the radiances observed by MLS depend strongly on the atmospheric pressure at the ray tangent point. This is due both to the large increases in atmospheric density with decreasing altitude, and to the fact that the spectral lines being measured are, at most altitudes, pressure broadened (an effect caused by molecules colliding with other molecules while emitting). At higher altitudes (in

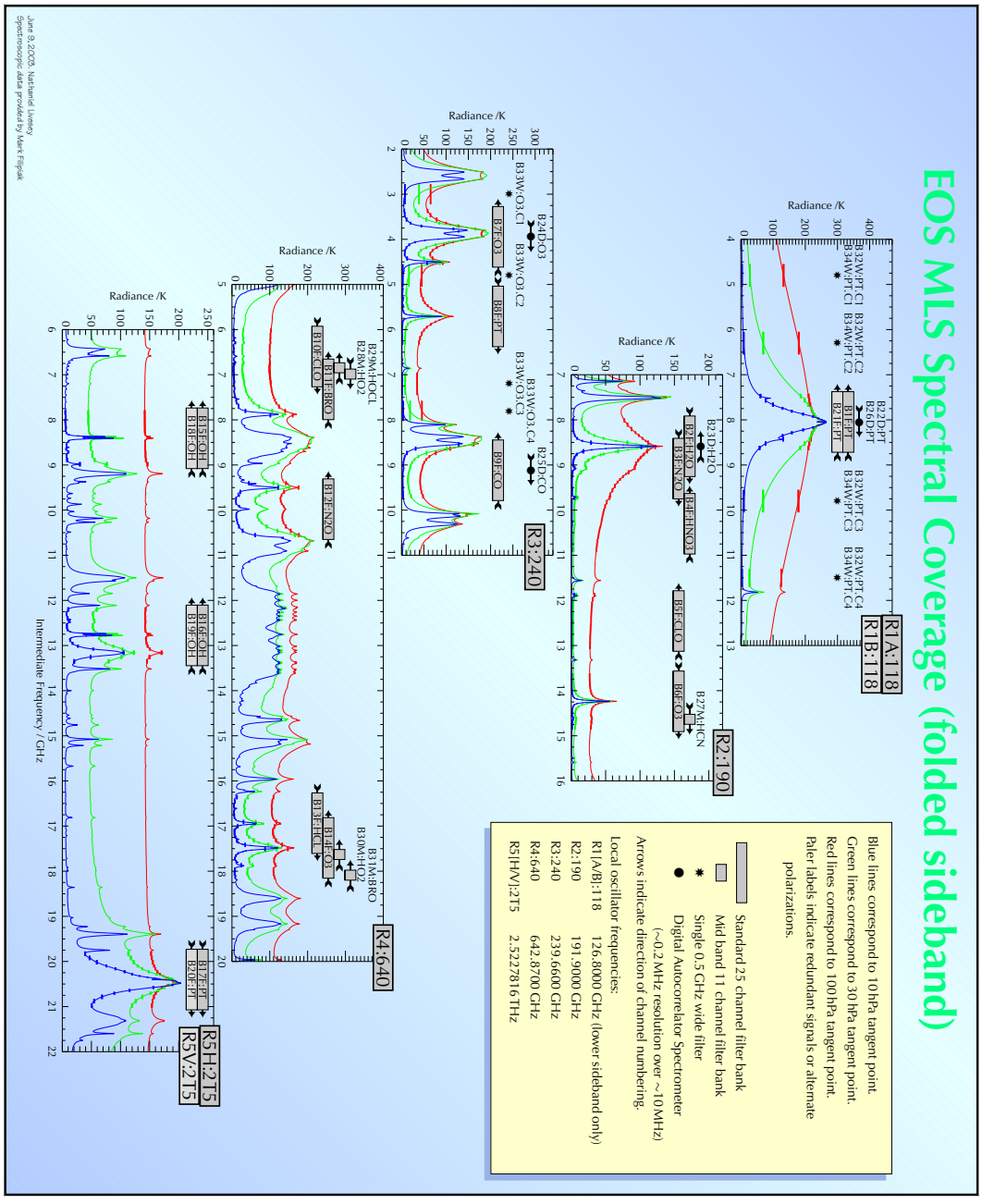
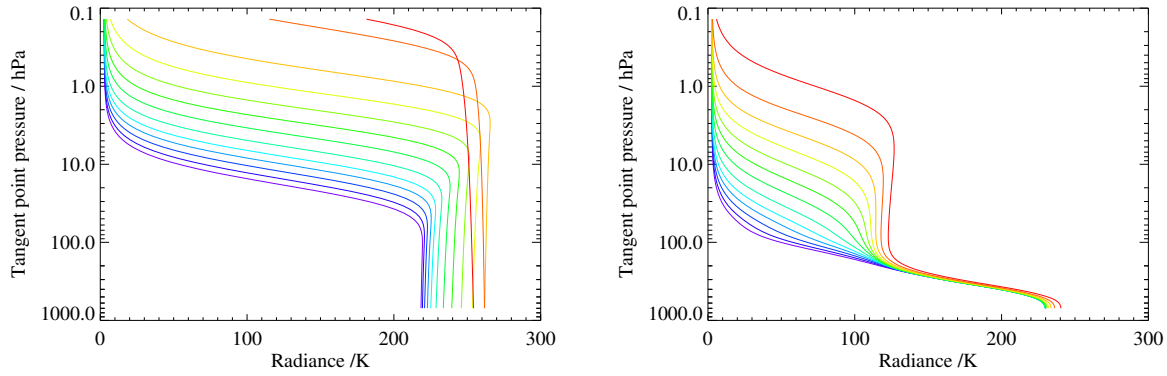


Figure 3.1: (color) EOS MLS Spectral coverage. See Appendix E for a description of the EOS MLS signal designation nomenclature.



(a) `R1A:118.B1F:PT`.

(b) `R2:190.B2F:H2O`.

Figure 3.2: (color) These plots show example radiance profiles for two different spectral bands in EOS MLS. The `R1A:118.B1F:PT` band is targeted at an O_2 line, and used to measure temperature and tangent pressure. The `R2:190.B2F:H2O` band is used to measure H_2O. Only the radiances for the first 13 channels in each band have been shown. The red lines correspond to the channels closest to the centers of the spectral lines; the purple lines correspond to those on the line wings. See Appendix E for a description of the EOS MLS signal designation nomenclature.

the mid to upper mesosphere), pressure broadening becomes less significant, and Doppler broadening due to the distribution of molecular velocities takes over as the dominant effect.

The use of tangent pressure is slightly complicated by the fact that the MLS instrument performs a continuous scan. The quantity stored in the state vector represents the tangent point pressure at the center time of the integration period. Currently, the forward model does not account for the motion of the tangent point during the scan, as the impact on the radiances is generally insignificant.

Quantities such as radiance and tangent pressure are somewhat distinct from explicitly geophysical quantities, such as temperature and composition, in that they are to a greater or lesser extent dependent on the state of the instrument (e.g. the pointing of the antenna.) These are often referred to as *minor frame quantities*, in that they vary from one MLS minor frame to the next. Where it is useful to draw distinctions between such quantities and strictly geophysical parameters, an arrow will be drawn over the relevant symbol, thus the tangent point pressure for minor frame i would be represented by the symbol $\vec{p}_i$.

3.4 Geometric measurements

In addition to microwave radiances, the MLS instrument's knowledge of the altitude of the limb tangent point $\vec{h}_i$, can be considered as a measurement. Given knowledge of the tangent point pressures $\vec{\mathbf{p}}$, the atmospheric temperature profile as a function of pressure $T(p)$, and the altitude h_0 of a fixed pressure surface p_0, a hydrostatic integration will serve as a forward model for such measurements.

$$\vec{h}_i = h_0 - \int_{p_0}^{\vec{p}_i} \frac{R(p)T(p)}{g(p)p} dp, \tag{3.25}$$

where g is the acceleration due to gravity and R is the atmospheric gas 'constant'.

The use of the additional information provided by these measurements is essential to supplying tangent pressure information in regions where the radiances are largely independent of pressure such as

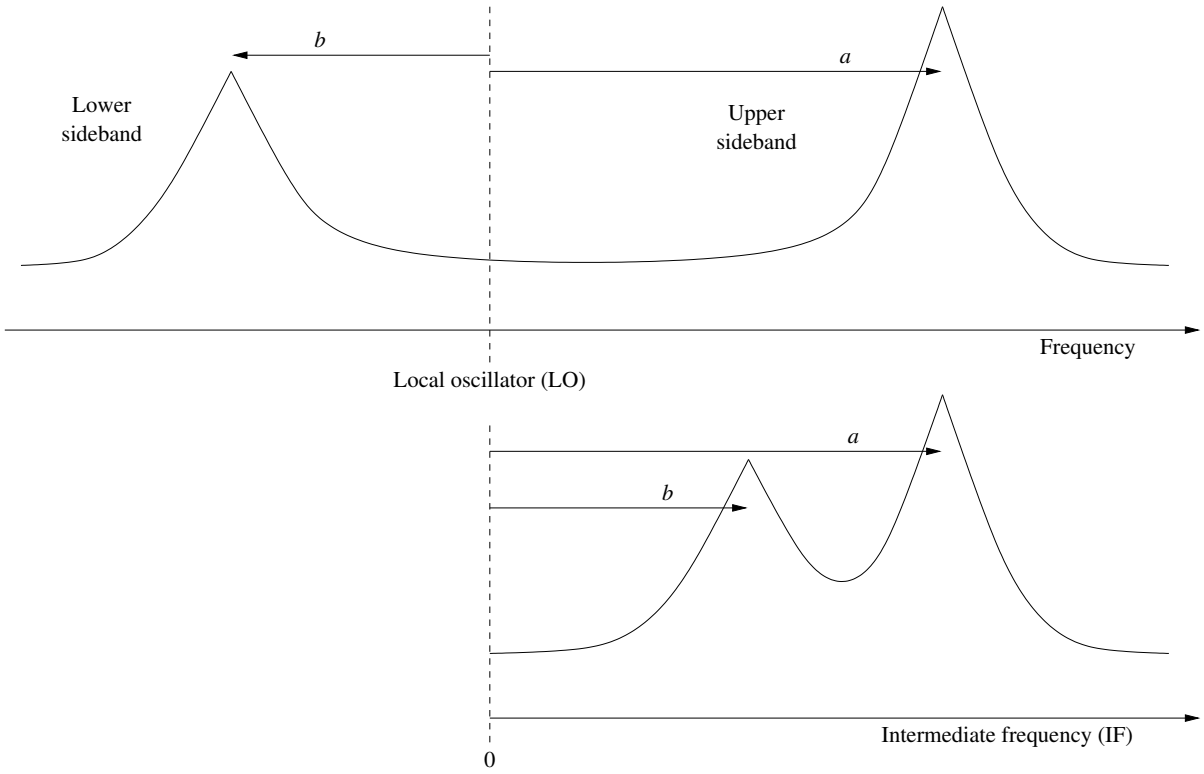


Figure 3.3: This figure shows the action of the microwave heterodyne technique. The observed radiances are combined in a non-linear element such as a diode, with a *Local Oscillator* (LO) signal. The resulting *Intermediate Frequency* (IF) signal is a combination of the signals in the upper and lower sidebands.

the lower altitude regions where the radiances are saturated, or the upper altitude regions where spectral lines are Doppler broadened rather than pressure broadened.

There is, however, a complication, due to the refraction of the tangent ray by the atmosphere, which leads to a difference between the true ray tangent point altitude and the 'unrefracted' tangent point altitude as determined by geometry and reported by the instrument. The magnitude of this refraction is a function of atmospheric density (and thus temperature and pressure), and humidity. It is possible to 'invert' the effects of this refraction and convert our hydrostatic estimate of the tangent point altitude to the same 'unrefracted' view held by the instrument scan encoder. However, this inversion itself involves a complex series of iterations. A simpler approach is to consider the effect of refraction on the direct measurements that are the MLS tangent height observations.

This effect leads to a conceptual problem with the measurement system, in that the measurements depend on the value of the state vector, which is not strictly speaking valid in retrieval theory (only the predicted measurements obtained from the forward model should depend on the value of the state vector.) This is really only a matter of semantics; the situation can easily be remedied by defining the measurement as a *scan residual*, the difference between the hydrostatic and geometric / refraction calculations of tangent point altitude, defining the values of this measurement to be zero. A forward model is used to compute this difference. The covariance matrix for this measurement is constructed as a diagonal matrix, with the diagonal values being based on the precision expected for the Level 1B tangent point altitude data, and no off diagonal correlations. Full details of the hydrostatic calculations are given in Read et al. [2004]. In the software implementing these algorithms it is the geopotential rather than the geometric height of the reference pressure surface that is included in the state vector, as

this is the more useful retrieved quantity.

3.5 Construction of the MLS 'state vector'

Section 3.2 gave an introduction to the concept of the 'state vector', the vector $\mathbf{x}$ that describes all the aspects of the atmosphere and measurement system that can affect the direct measurements. The construction of the state vector is an essential part of the design of any retrieval system; its contents need to be chosen carefully to provide a complete set of independent parameters that describe the whole system. This section discusses the construction of the state vector for the MLS retrieval system.

Full details of the state vector used in the 'launch ready' version of the software that implements these algorithms are given in Appendix B.

3.5.1 State vector selection methodology and implementation

The construction of the MLS state vector is performed by studying the measurement system with reference to a set of simple criteria.

Firstly, the experimental objective is placed in the state vector. That is, profiles of atmospheric temperature $\mathbf{T}$, geopotential height $\mathbf{Z}$ and composition (concentration of targeted gases) $\mathbf{f}$ on fixed pressure surfaces.

At this point, any superfluous information in the state vector is removed. In this case the temperature information is redundant with most of the geopotential height profile as, given profiles of temperature and one geopotential height, the entire $\mathbf{Z}$ profile could be computed using hydrostatic balance. For this reason, the vector is reduced so as to contain the geopotential height of only one pressure surface.

Next, the primary source for direct information about the state vector is identified. In the MLS case, this direct information is the radiance observations $\vec{\mathbf{I}}$ and the scan residual measurements $\vec{\mathbf{r}}$.

The next set of quantities to be placed in the state vector represents any additional information needed in order to characterize the direct measurements. In this case, it is clear that tangent pressure information for each radiometer is essential if forward model estimates are required for both radiances and scan residual measurements. In order that the calculations may be more linear, this quantity is represented by the vector $\vec{\zeta} = -\log_{10}\left[\text{Tangent pressure / hPa}\right]$.

This whole process is somewhat iterative, in that these new quantities may themselves be dependent on further additional information. For example, the estimates of tangent pressure for each radiometer can be dramatically improved given knowledge of the angular offsets between the various radiometers and/or modules in the case of the GHz/THz module offsets. Thus these offsets are included in the state vector. These offsets are essentially constant, although in the case of the offset between the GHz and THz modules, there may be a slight orbital dependence. Additional redundancy may develop as more quantities are introduced into the state vector. It is important to recognize the source of this redundancy and attempt to eliminate it. A pictorial representation of this whole process, as applied to the EOS MLS measurement system is given in Figure 3.4.

As the state vector is constructed, attention needs to be paid to the possible need for *a priori* values for the elements added. In the case of the $\vec{\zeta}$ components, any *a priori* values would need to be based on the observed Level 1B tangent point height information. Clearly, as these heights are being used (in the form of geopotential heights) as direct measurements as described in Section 3.4, the use of an *a priori* virtual measurement for $\vec{\zeta}$ would result in the use of the same information twice. In fact, an *a priori* estimate for $\vec{\zeta}$ is unnecessary, as the measurement system already has enough information to completely describe $\vec{\zeta}$. The Level 1B tangent point height and the temperature elements of the state vector provide enough information to describe values of $\vec{\zeta}$ even in cases where all the radiances for a given minor frame

are missing. However, it is still desirable and correct to use this same height information in forming an initial guess for $\vec{\zeta}$.

3.5.2 Representation within the state vector

As described in Section 3.2.1, the state vector is designed to represent the functional form of the atmospheric temperature and composition. In the MLS case, this functional form is constructed using a set of *basis functions*. For example, in the case of the atmospheric temperature profile, the state vector is defined according to

$$T(z) = \sum_i T_i \eta_i(z), \tag{3.26}$$

where z is the log pressure vertical coordinate[2] and $T(z)$ is the functional form of temperature. T_i are the components of the state vector describing the temperature, and $\eta_i(z)$ are the basis functions.

For most of the MLS species, the n basis functions $\eta_0 \ldots \eta_{n-1}$ are defined as

$$\eta_i(z) = \begin{cases} 1 & \text{if } (i = 0 \text{ and } z < z_0) \text{ or } (i = n-1 \text{ and } z > z_{n-1}), \\ \frac{z_{i+1}-z}{z_{i+1}-z_i} & \text{if } z_i \le z < z_{i+1}, \\ \frac{z-z_{i-1}}{z_i-z_{i-1}} & \text{if } z_{i-1} \le z < z_i, \\ 0 & \text{otherwise} \end{cases} \tag{3.27}$$

A sample set of such functions is shown in Figure 3.5. These functions give a profile whose functional form is equivalent to linear interpolation in z between the values T_i on surfaces z_i, with no extrapolation beyond z_0 or z_{n-1}.

For some species, this linear interpolation form may not describe the true atmospheric profile with sufficient accuracy. An example of this is the case of tropospheric H_2O concentration. The vertical profile of H_2O in the upper troposphere shows rapidly decreasing abundance with increasing altitude. Such variations are better described by a basis that uses linear interpolation in log mixing ratio.

3.5.3 Continuum emission and 'baseline'

Most of the MLS observations rely on measurements of *spectral contrast*, that is, the retrieval algorithms determine the atmospheric composition by effectively comparing the radiances near the center of a spectral line with those in the wings. The absolute value of the radiances observed is not typically as important a factor in the MLS measurement system. There are many factors that determine such spectrally flat contributions to the MLS radiance observations. The term 'flat' in this context implies signals that may vary between the spectral bands, but will vary only slowly within the bands. Unexplained spectrally flat signals can arise in the measurement system from many different sources; these broadly divide into instrumental and forward model factors.

Instrumental contributions include unmodeled blackbody emission from the MLS antenna, and uncertainties in the knowledge of the instrumental field of view. In the case of the field of view, an important factor is the uncertainty in the signal received through sidelobes of the antenna pattern.

One mechanism whereby spectrally flat errors can be introduced into the forward model calculation is continuum emission, absorption or scattering by species whose abundance is not known to a high enough accuracy, and/or whose emission is not well known at the frequencies under consideration (for example water vapor in the lower stratosphere / upper troposphere.) While the water vapor profile from

[2]Do not confuse z with $\vec{\zeta}$; z is a fixed vertical coordinate, $\vec{\zeta}$ is a set of state vector elements describing the pressure at the limb path tangent points.

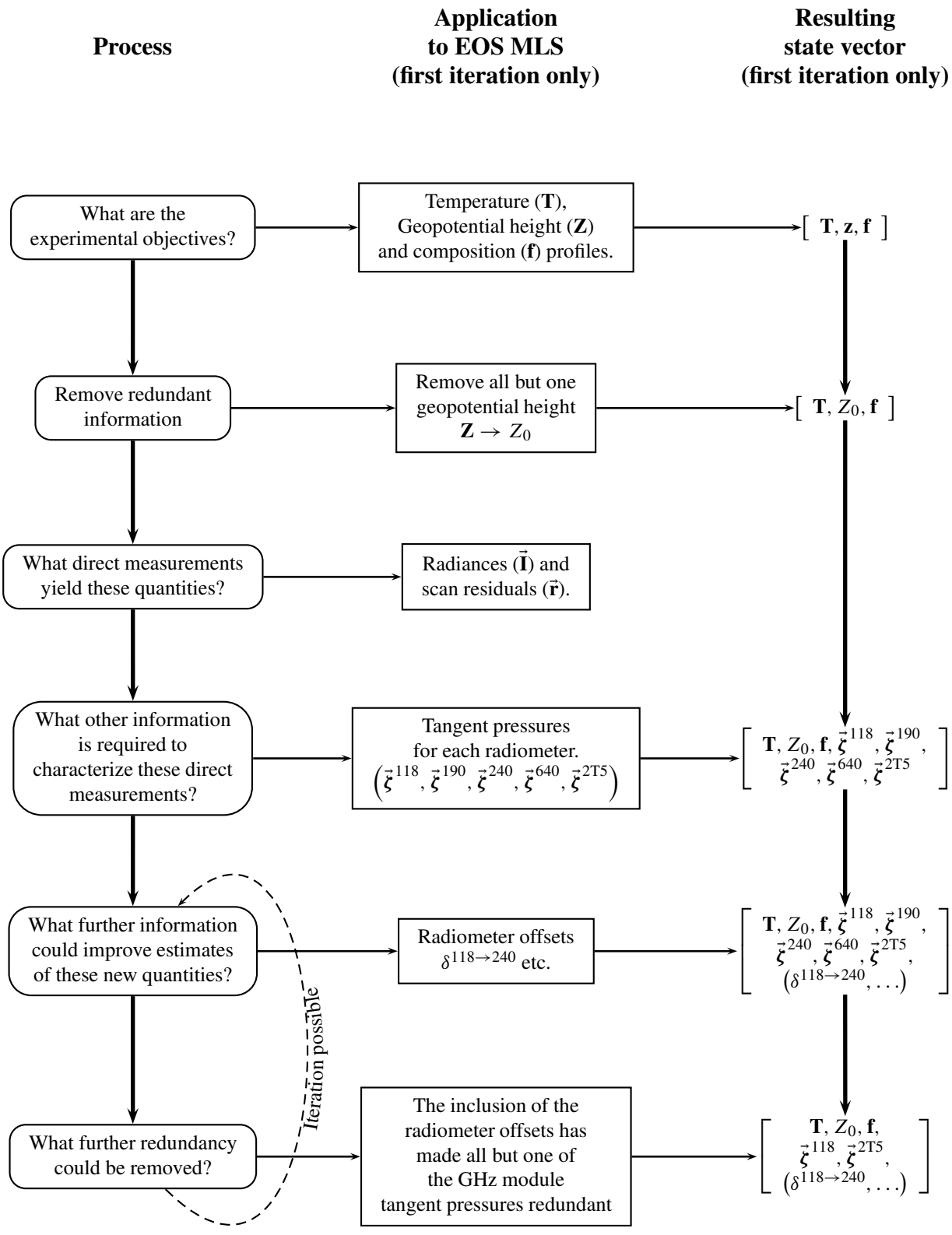


Figure 3.4: A pictorial representation of the EOS MLS state vector selection process

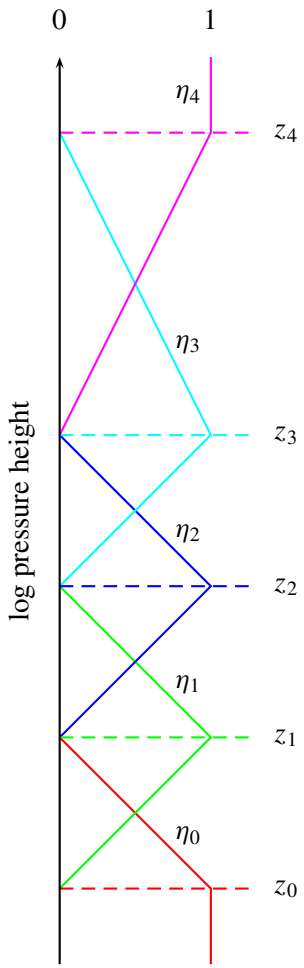


Figure 3.5: (color) This figure shows an example of the representation bases used in the MLS state vector. The functions η_i are shown by the colored lines, corresponding to log pressure heights z_i. The vertical resolution need not be constant; in this case, the resolution halves for the layer $z_3 \rightarrow z_4$.

about 500 hPa upward can be measured by MLS, it may not be known sufficiently well to account precisely for the spectrally 'flat' radiance at the lower altitudes with an accuracy that is consistent with the MLS measurement accuracy.

It is possible to model these errors through the use of off-diagonal terms in the MLS radiance measurement covariance matrices. However, as explained in Section 3.2.5, the computational effort required when dealing with such covariance matrices is vast. A more practical approach is to include elements in the state vector that model such effects, and retrieve them. This is the solution adopted in the MLS processing.

There are several possible approaches to modeling such spectrally flat contributions. For UARS MLS the state vector included a 'baseline' term for each band, which was a spectrally flat radiance added to all channels in a band, with independent values for each minor frame. For EOS MLS this would be somewhat less desirable, as the large number of minor frames would make this one of the larger components of the state vector. Instead, the current approach has a vertical baseline profile as a function of pressure for each radiometer. This is then interpolated to the minor frame tangent pressures in the state vector to determine the baseline to add.

While such a term can be useful for capturing instrumental artifacts, it is not so adept at dealing with atmospheric and spectroscopic problems. For example, if a thin cloud is present in the upper troposphere its impact on radiance observations with tangent pressures in the upper troposphere will depend on the optical depth of the channels in question. Only those channels optically thin enough to see down into the upper troposphere will be impacted by the cloud; those channels that saturate in the stratosphere or above will not be effected. However, a spectrally flat 'baseline' term would add uniformly to all the radiances regardless of optical depth.

An alternative approach that accounts for such effects is to include in the state vector a vertical profile of the mixing ratio of a fictitious species that has an extinction coefficient of unity at all frequencies. Optically thick radiances will, as with any atmospheric species, not be affected by changes in this parameter below the region where radiance saturation occurs. The drawback of this representation is that the impact of this parameter is very non-linear, especially when a negative abundance is required to best match the signals (as the contributions of this parameter to the radiances are of the form $\exp[-x]$.) Experience from testing of the EOS MLS algorithms has shown that this property makes the use of this parameter undesirable.

Other families of state vector elements are under consideration for later versions of the MLS data processing that may retain the property of accounting well for optical depth, while having a more linear impact on the MLS radiances.

3.5.4 Other sources of correlated radiance error.

In addition to the 'baseline' terms discussed above, there are additional mechanisms that can produce correlations in the errors on MLS radiance measurements. One such is the so-called $1/f$ noise described in Jarnot [2004]. In most cases, this effect is well modeled by the retrieval of the baseline terms described above. One possible exception may be cases where the spectral signal being sought is very weak. These are the measurements that will be used to derive the 'noisy' products. For a full description of the issues involved with these measurements, see Chapter 5. Another source of correlated errors in radiances may be the use of an interpolated value of the space radiance described in Jarnot [2004] to calibrate individual MLS radiance observations. Clearly, uncertainty in the knowledge of the true space radiance will yield correlated uncertainties in the observations of limb radiances.

Several of the spectral bands in the MLS instrument overlap, as shown in Figure 3.1, for example `R2:190.B2F:H2O` and `R2:190.B3F:N2O`. The noise on the MLS radiances is dominated by noise from the radiometers, rather than noise from the individual spectrometers. This means that there will be

a strong correlation in the errors on individual radiances from these two bands. If this is not modeled correctly, the retrieval algorithm will draw false inferences, believing it has two independent measures of the atmosphere, when in fact their noises are identical.

A full, non-diagonal, treatment of such cases is impractical, as described above. However, there are two possible alternative solutions. One could approximate the covariance matrix by increasing all the diagonal elements by a factor of $\sqrt{2}$. This prevents the retrieval algorithm from reading too much into the measurement system, but is an approximation, as it does not indicate the coupled nature of the problem. The prefered approach is simply to use data only from one of the two channels in regions of overlap. Typically one would choose to keep the channels with higher frequency resolution. There would probably remain either small regions of overlap, or small gaps, due to the lack of complete alignment in the channels for the two bands, but these are of little consequence.

The spectral response of the MLS signal channels is not completely rectangular as it would be in the ideal case. This leads to a very slight overlapping of spectral response between adjacent channels. While the detailed spectral response of the channels is considered fully in the forward model calculations, the small level of correlated uncertainty the overlap produces can safely be neglected in the retrieval error budget, as the overlap is very small.

3.5.5 A note on 'state vectors' and 'retrieval vectors'

So far the discussion in this section has assumed that all of the elements of the state vector will be retrieved by the MLS data processing algorithms. In reality, as described in Section 3.2.5, the retrieval algorithm will proceed in separate phases. For example a retrieval of temperature and pressure may be followed by a retrieval of tropospheric humidity and ozone abundance, followed in turn by a retrieval of stratospheric composition. Thus, as the retrieval progresses, some elements of the state vector may be constrained to values, either taken from a prior phase, or from some *a priori* information, as described in Section 3.2.5.

In the context of the standard literature (e.g., Rodgers [1976, 1990]), the 'retrieved' and 'constrained' quantities are distinguished by defining the forward model as

$$\hat{\mathbf{y}} = \mathbf{f}(\mathbf{x}, \mathbf{b}) , \tag{3.28}$$

where the vector $\mathbf{b}$ describes the quantities that are constrained. In this notation, the 'state vector' described in this section would be the combined vector $[\mathbf{x}, \mathbf{b}]$ (some times called the 'forward model' vector), while the 'retrieval vector' would be the vector $\mathbf{x}$ alone.

In most of this document the distinction is unnecessary, so the term 'state vector' will be used to loosely refer to either the 'retrieval vector' or the 'forward model' vector.

3.6 Observation geometry

The EOS MLS instrument observes radiances emitted from the limb of the earth in the plane described by the spacecraft orbit. This contrasts with the UARS MLS instrument, which observed limb emission perpendicular to the flight direction. This observation geometry has far reaching implications for improving the retrieval algorithms, best explained with reference to Figure 3.6. This figure makes it clear that the radiances observed in an individual limb scan are a function of the temperature and composition of the atmosphere over a range corresponding to several adjacent Level 2 profiles. In the UARS MLS case, while the horizontal scales involved were similar, spherical symmetry was assumed, as there was no direct way to obtain information about the horizontal variability of the atmosphere along the line of sight. Some instruments such as the *Improved Stratospheric And Mesospheric Sounder* instrument on

UARS [Taylor et al., 1993], which had the same measurement geometry as UARS MLS, adopted a two pass approach, whereby the profiles obtained from a first pass retrieval were mapped onto a grid. The horizontal gradients of this gridded field were then used as *a priori* information in a second pass of the retrieval process [Dudhia and Livesey, 1996].

In the EOS MLS case however, with the instrument looking along the flight direction, much more information on the horizontal variability of the atmosphere along the line of sight can be obtained. Spherical symmetry need no longer be assumed, and the retrieval task can be structured to account for, and provide information on, horizontal variability along the line of sight. The issues raised by this are covered in detail in the next chapter.

3.6.1 'Fundamental' coordinates

In constructing the state vector, it is very important to recognize which sets of quantities are independent, and which are related. For example, many of the aspects of the MLS measurement system are functions of earth radius and the acceleration due to gravity g for a given altitude. However, both of these quantities are themselves functions of latitude (and to a lesser extent, longitude.) It is important to acknowledge this interdependence; otherwise, the retrieval process can exhibit unphysical behavior due to the double bookkeeping taking place. The way to get around this problem is to construct a set of fundamental coordinates for the system. It is these coordinates that are then held constant when computing partial derivatives in the forward model.

Clearly, the development so far indicates that pressure is the most appropriate fundamental coordinate in the vertical direction. This does introduce additional complexity in the forward model, as changes in temperature as described on a pressure grid will, through hydrostatic balance, lead to a change in the altitude of that grid, and thus, changes in parameters such as g.

The appropriate horizontal coordinate should account for the inclination of the Aura orbit and the oblateness of the Earth. It is desirable to use a coordinate system within which the MLS scans are relatively evenly spaced, as many of the quantities used in the forward model will be precomputed on regular horizontal and vertical grids. Clearly latitude is not an appropriate candidate, as the scan spacing is not constant in latitude. In fact, at high latitudes there is a degeneracy in the latitude coordinate, due to the inclination of the orbit. For EOS MLS, the coordinate ϕ (the geodetic great circle angle along the orbit track), described in Figure 3.7, will be used as a fundamental coordinate.

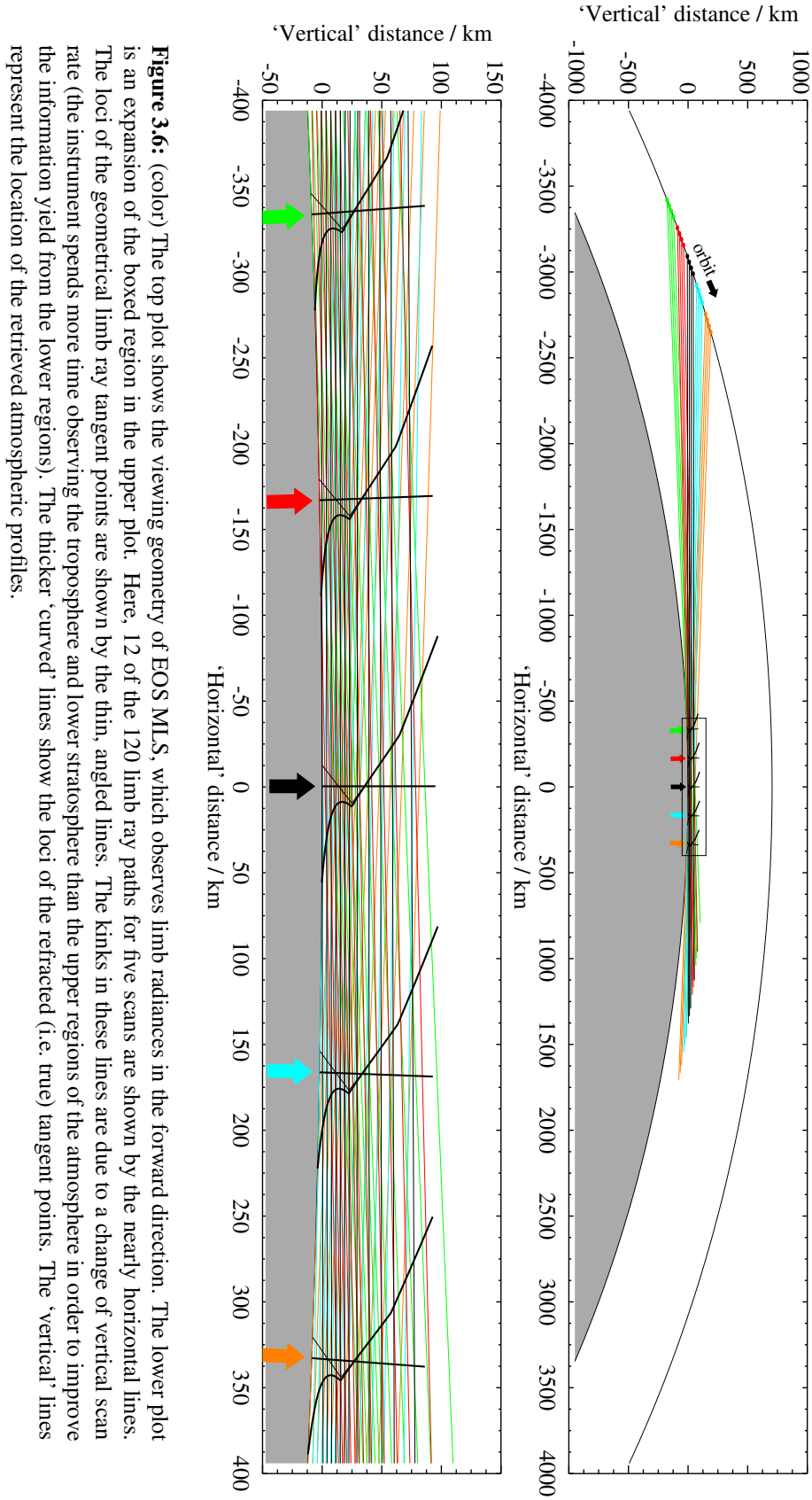


Figure 3.6: (color) The top plot shows the viewing geometry of EOS MLS, which observes limb radiances in the forward direction. The lower plot is an expansion of the boxed region in the upper plot. Here, 12 of the 120 limb ray paths for five scans are shown by the nearly horizontal lines. The loci of the geometrical limb ray tangent points are shown by the thin, angled lines. The kinks in these lines are due to a change of vertical scan rate (the instrument spends more time observing the troposphere and lower stratosphere than the upper regions of the atmosphere in order to improve the information yield from the lower regions). The thicker 'curved' lines show the loci of the refracted (i.e. true) tangent points. The 'vertical' lines represent the location of the retrieved atmospheric profiles.

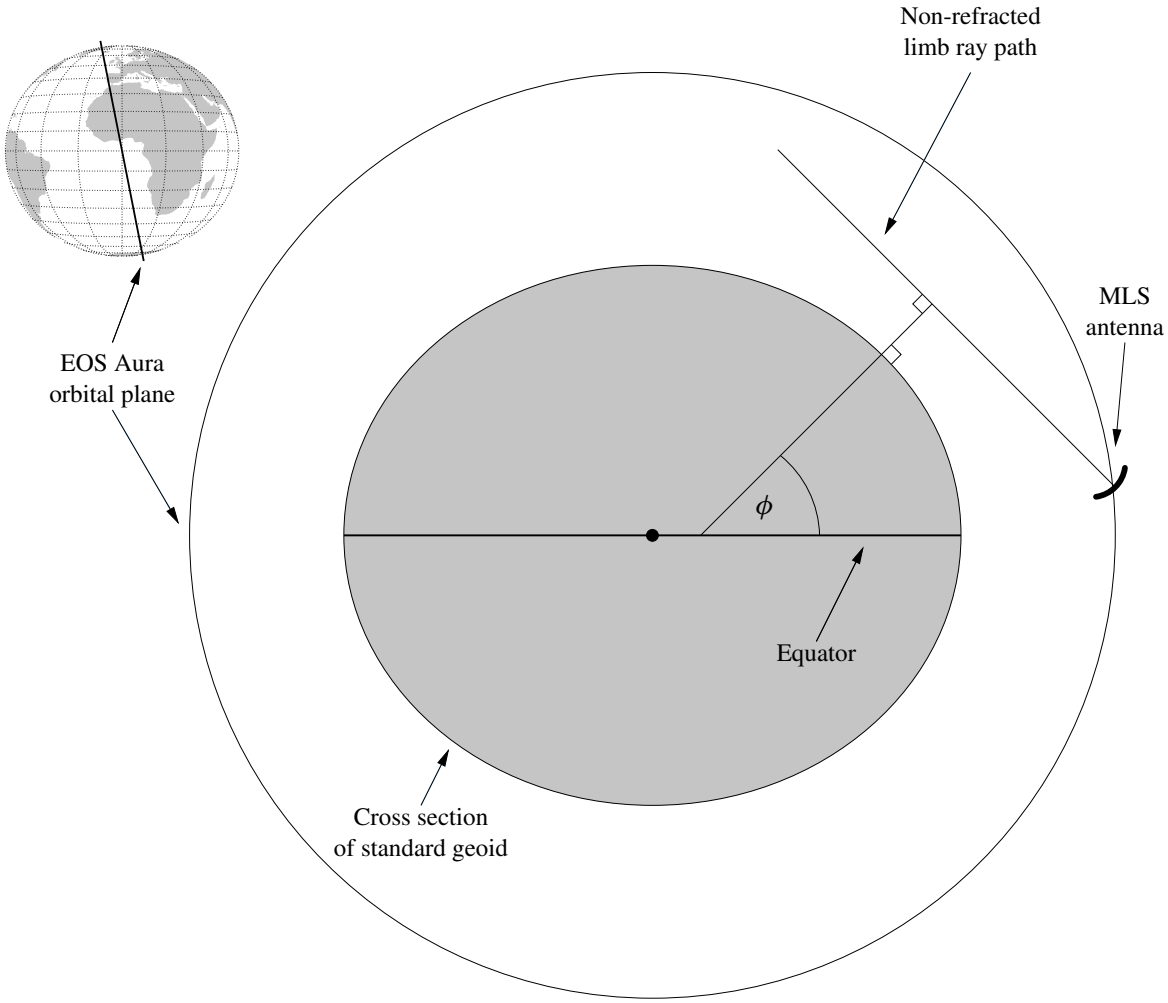


Figure 3.7: The observation geometry of MLS is affected by the oblateness of the earth, and the inclination of the orbit. The main part of this figure shows a cross section in the 98° inclined orbital plane. The master horizontal coordinate, ϕ, is defined as the angle between the normal to the geoid, normal to the limb ray path, and the equator. The radius of the orbit and the oblateness of the Earth have been exaggerated for clarity.

Chapter 4
The EOS MLS Level 2 data processing algorithms

Section 3.6 discussed the details of the EOS MLS observation geometry. A key point is that radiance observations from one scan are dependent on the state of the atmosphere over a horizontal range of order of a few profile spacings. This fact provides a useful tool for characterizing the details of atmospheric variability along the spacecraft flight direction, if it is explicitly included in the calculations.

In order to take advantage of this, horizontal homogeneity cannot be assumed in either the forward model or retrieval processes. The approach taken is to retrieve the data in blocks of contiguous profiles known as 'chunks'. Such a calculation is not prohibitive, due to the comparative sparsity of the matrices involved. This Chapter considers in detail the implementation of the retrieval calculation in this manner.

Other aspects of the implementation of the retrieval calculation are also discussed, including improving the efficiency of the calculation, and its numerical stability.

4.1 A simple one dimensional approach

Before discussing the full two-dimensional MLS retrieval system described above, it should be pointed out that simpler one dimensional approaches are possible. Equation 3.17, repeated here, gives an iterative expression for the retrieval operation

$$\mathbf{x}^{(r+1)} = \mathbf{x}^{(r)} + \left[\mathbf{S}_\mathbf{a}^{-1} + \mathbf{R}^\mathrm{T}\mathbf{R} + \sum_i \mathbf{K}_i^\mathrm{T}\mathbf{S}_i^{-1}\mathbf{K}_i \right]^{-1} \left\{ \mathbf{S}_\mathbf{a}^{-1}\left[\mathbf{a} - \mathbf{x}^{(r)}\right] + \mathbf{R}^\mathrm{T}\mathbf{R}\left[\mathbf{a} - \mathbf{x}^{(r)}\right] + \sum_i \mathbf{K}_i^\mathrm{T}\mathbf{S}_i^{-1}\left[\mathbf{y}_i - \mathbf{f}_i(\mathbf{x}^{(r)})\right] \right\}. \quad (4.1)$$

It is possible to ignore all the two dimensional issues discussed in Section 3.6. Instead the assumption is made that one complete scan of radiance observations depends only on one atmospheric profile, and that horizontal homogeneity can be assumed. This turns the retrieval calculation into N separate individual profile retrievals, where N is the number of profiles under consideration. This approach corresponds exactly to the UARS MLS version 5 retrieval algorithm, where retrievals are made independently on each major frame's data in turn.

4.2 Structure and sparsity in the MLS retrieval system

The MLS retrieval system processes data in 'chunks', with each chunk containing data for a fraction of a single orbit. For each chunk, the system retrieves data from M major frames worth of radiances. The state vector for the retrieval consists of N profiles[1]. The retrieval algorithm is tasked with finding the optimum value for all the state vector profiles that simultaneously matches the M radiance observations.

[1] In this discussion the term 'profile' (singular) is taken to refer to a complete set of vertical profiles of temperature and composition for a *single* location on the globe.

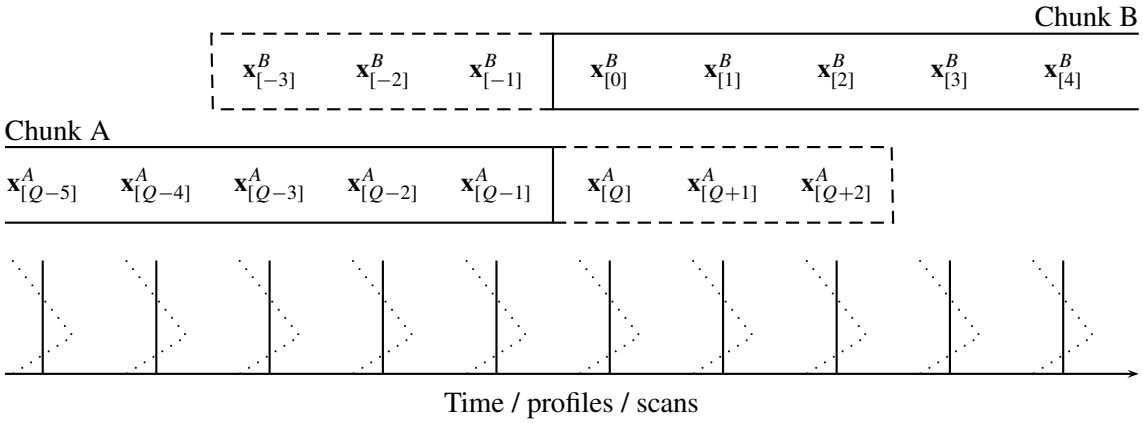


Figure 4.1: This figure shows an example of how the MLS profiles are retrieved in chunks. Each chunk consists of Q profiles, with an additional q profiles of overlap (3 in this case). The total number of profiles in a chunk N is thus $Q+2q$. Beyond the end of each chunk, horizontal homogeneity is assumed; the overlaps are included (where possible) to reduce the impact of this approximation on the retrieved data. The data in the overlapping regions could be compared for diagnostic purposes. For example, $\mathbf{x}^A_{[Q]}$ should be similar to $\mathbf{x}^B_{[0]}$.

In order to overcome edge effects, the chunks overlap slightly, with the retrievals in the overlap regions discarded.

Typically the horizontal spacing of the state vector profiles is chosen to match the horizontal spacing of the MLS scans, giving $N \simeq M$ (with differences between N and M mainly arising from slight variations in the scan spacing around the orbit). However, this is not a requirement; one could construct a state vector consisting of two profiles per scan, or one profile every two scans if necessary. In addition, the horizontal spacing of the state vector may vary from species to species within the vector. For simplicity, we will consider the $N = M$ arrangement shown in Figure 4.1

For the moment, we for clarity consider a system with only one measurement vector. The state vector consists of N profiles $\mathbf{x}_{[j]}$ each of length n. There are additional state vector elements $\mathbf{x}^\diamond$ that contain any 'constant' quantities such as spectroscopy and instrument calibration (e.g. sideband ratios), which may be sufficiently uncertain that it is appropriate to retrieve them.

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}^\diamond \\ \mathbf{x}_{[1]} \\ \mathbf{x}_{[2]} \\ \vdots \\ \mathbf{x}_{[N]} \end{bmatrix} \tag{4.2}$$

The goal of the retrieval algorithm is then to simultaneously obtain values for the entire vector $\mathbf{x}$. The main part of the retrieval calculation is the computation of the normal equation matrix $\left[\mathbf{S}_\mathbf{a}^{-1} + \mathbf{R}^\mathrm{T}\mathbf{R} + \mathbf{K}^\mathrm{T}\mathbf{S}_\mathbf{y}^{-1}\mathbf{K}\right]$, which is the topic of the rest of this section. The following section discusses the issues involved in inverting this matrix.

4.2.1 The weighting function matrices

In the same manner as we divided the state vector up in to sub-vectors, the measurement vector can be divided up into sub-vectors by major frame. Following from this, the weighting function matrix $\mathbf{K}$ can be divided into blocks, each corresponding to the derivative of one major frame's worth of radiance

measurements with respect to one state vector profile. Section 3.6 showed that in the $N = M$ case, the measurements from one scan are affected by ~ 5 adjacent profiles, 2 either side of the nominal position. This amounts to saying that the $\mathbf{K}$ matrix has a bandwidth $p = 2$, such that $\mathbf{K}_{[\alpha j]} = \mathbf{0}$ for all $|\alpha - j| > p$. As an illustration, consider a simpler case where $p = 1$, and $N = 6$; here the weighting function matrix has the block structure

$$\mathbf{K} = \begin{bmatrix} \times & \times & \times & 0 & 0 & 0 & 0 \\ \times & \times & \times & \times & 0 & 0 & 0 \\ \times & 0 & \times & \times & \times & 0 & 0 \\ \times & 0 & 0 & \times & \times & \times & 0 \\ \times & 0 & 0 & 0 & \times & \times & \times \\ \times & 0 & 0 & 0 & 0 & \times & \times \end{bmatrix}. \tag{4.3}$$

The $\times$ symbol indicates a non-zero block sub matrix, while 0 indicates a sub matrix that is identically zero. Each row of the matrix corresponds to a separate scan. The first column indicates the weighting functions for the $\mathbf{x}^{\diamond}$ information, while the remaining columns indicate profiles $1 \cdots N$. This type of matrix is known as a singly-bordered (i.e. one column fully non zero) block band diagonal matrix.

4.2.2 The $\mathbf{K}^{\mathrm{T}}\mathbf{S}_{\mathbf{y}}^{-1}\mathbf{K}$ matrix

As described in Sections 3.2.1 and 3.2.5 the measurement covariance matrix $\mathbf{S}_{\mathbf{y}}$ is diagonal. This gives a $\mathbf{K}^{\mathrm{T}}\mathbf{S}_{\mathbf{y}}^{-1}\mathbf{K}$ matrix that is doubly-bordered block band diagonal with a block bandwidth of $2p$. The example $\mathbf{K}$ matrix of Equation 4.3 gives

$$\mathbf{K}^{\mathrm{T}}\mathbf{S}_{\mathbf{y}}^{-1}\mathbf{K} = \begin{bmatrix} \times & \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & 0 & 0 & 0 \\ \times & \times & \times & \times & \times & 0 & 0 \\ \times & \times & \times & \times & \times & \times & 0 \\ \times & 0 & \times & \times & \times & \times & \times \\ \times & 0 & 0 & \times & \times & \times & \times \\ \times & 0 & 0 & 0 & \times & \times & \times \end{bmatrix}. \tag{4.4}$$

4.2.3 A sequential approach to the problem

The full $\mathbf{K}$ matrix in equation 4.3 represents a very large array of numbers, corresponding to memory usages of order tens of gigabytes. Clearly this too large to be easily stored in memory on any but the largest of computers. However, it is not necessary that the whole $\mathbf{K}$ matrix should be stored simultaneously. Let us briefly return to dividing the measurements up into separate vectors $\mathbf{y}_i$, and let the division be by major frame (i.e., scan). Our individual $\mathbf{K}_i$ matrices are then just single row blocks from the matrix in equation 4.3, and represent one Nth of the memory storage requirement.

$$\begin{array}{rcl} \mathbf{K}_1 & = & [\ \times\ \times\ \times\ 0\ 0\ 0\ 0\ ] \\ \mathbf{K}_2 & = & [\ \times\ \times\ \times\ \times\ 0\ 0\ 0\ ] \\ \mathbf{K}_3 & = & [\ \times\ 0\ \times\ \times\ \times\ 0\ 0\ ] \\ \mathbf{K}_4 & = & [\ \times\ 0\ 0\ \times\ \times\ \times\ 0\ ] \\ \mathbf{K}_5 & = & [\ \times\ 0\ 0\ 0\ \times\ \times\ \times\ ] \\ \mathbf{K}_6 & = & [\ \times\ 0\ 0\ 0\ 0\ \times\ \times\ ] \end{array} \tag{4.5}$$

Our sum over measurement vectors becomes a loop over MLS major frames. For each frame, the forward model radiances $\mathbf{f}_i(\mathbf{x})$ and weighting functions $\mathbf{K}_i$ are computed. Given this, the product $\mathbf{K}_i^{\mathrm{T}}\mathbf{S}_i^{-1}\mathbf{K}_i$ is formed and added to the normal equation matrix, along with the contributions of this major frame

to the $\sum_i \mathbf{K}_i^{\mathrm{T}} \mathbf{S}_i^{-1} [\mathbf{y}_i - \mathbf{f}_i(\mathbf{x})]$ term. The normal equation matrix itself is much smaller, typically on the order of a few hundred megabytes for 15° chunks, so storage of this is not a problem for a computer of a typical caliber. This recasting of the problem does not result in any decrease in the number of computations required, merely in the amount of memory used.

Note that this loop over scans occurs within the iteration loop. That is to say each iteration involves a loop over all the major frames in a chunk. In this way it can be regarded as a form of 'transpose' of the one dimensional approach discussed in Section 4.1 where the iterations were carried out on individual major frames in turn.

For the remainder of this chapter we will return to considering only one measurement vector for simplicity, while recognizing that the algorithms are implemented in software in the manner described above.

4.2.4 Sparsity in the individual block sub matrices

As well as being sparse in the block sense, many of the individual submatrices involved in the retrieval calculation are themselves sparse. However, only when the sparsity in a block submatrix is significant can appreciable savings be made. For example, while the ∂[Radiance]/∂[Composition] submatrices will typically be about 50% zero, the possible savings to be made in not multiplying by zero in this case would easily be outweighed by the burden of storing and perusing the matrix block in sparse form.

One set of block submatrices that will be highly sparse is those involving the tangent pressure $\vec{\boldsymbol{\zeta}}$ quantity. Each radiance is only dependent on the tangent pressure for its own MIF. Thus, for a 25 channel filter bank, with 120 MIFs, the full matrix size would be $(25 \times 120) \times 120 = 32\,320$, however, only $25 \times 120 = 3000$ (0.8%) of these values will be non zero. This is significant, as the $\vec{\boldsymbol{\zeta}}$ quantity, having 120 elements per scan, has far larger derivative matrices than, say, temperature where 40 – 50 elements per profile are typical.

4.2.5 The *a priori* covariance matrix

Section 3.2.3 detailed the use of the *a priori* to ensure the existence of a solution of the retrieval problem. Section 3.2.4 also detailed how the *a priori* covariance matrix can be used to apply some smoothing on the measurement system. The MLS retrieval algorithms do not use this approach, favoring the Tikhonov smoothing method as described in the next section. Omitting the smoothing terms from the *a priori* covariance matrix makes that matrix a purely diagonal one, with the diagonal elements representing the *a priori* variance of the state vector element in question. The computation of its inverse and its addition to the normal equations is therefore a trivial computational step.

4.2.6 The Tikhonov smoothing

Section 3.2.4 gave an introduction to the Tikhonov smoothing approach. In the MLS algorithms the smoothing is applied both in the vertical and horizontal domain. This leads to two distinct smoothing matrices $\mathbf{R}_v$ and $\mathbf{R}_h$. The $\mathbf{R}_v$ matrix describes relationships between data within individual vertical profiles and is therefore clearly block diagonal. The individual blocks of $\mathbf{R}_v$ have the same form as the $\mathbf{P}$ matrix of equation 3.16, scaled by a $\mathbf{W}_v$ describing the weight of the vertical smoothing terms in the retrieval.

The approach to choosing an appropriate value for the $\mathbf{W}_v$ is somewhat complex. As the characteristics of the MLS measurement system (precision and resolution) are a strong function of height, it is desirable to be able to adjust the smoothing weights for different height regions appropriately. However, the dimensions of $\mathbf{W}_v$ of $(n - q) \times (n - q)$ (where n is the number of levels in the profile and p is the

order of the Tikhonov smoothing) bear little relation to other vectors and matrices in the algorithms, and do not easily relate to a familiar coordinate system.

In the case of the MLS algorithms, we solve this by defining a n element vector $\mathbf{w}$, which is viewed as a measure of our *a priori* estimate of the standard deviation of the vertical qth derivative of the species. We then form the matrix $\mathbf{W}_v$ as a diagonal matrix with the vector obtained from the product abs $(\mathbf{P})\,\mathbf{w}$ in its diagonal elements (the operation abs $(\mathbf{P})$ in this case indicates the matrix formed by taking the absolute value of each element of $\mathbf{P}$). This is a simple way to arrive at $\mathbf{W}_v$ from a quantity more directly related to the state vector vertical grid, and can be viewed as a 'smoothing' of our weighting. The fact that $\mathbf{R}_v$ is block diagonal implies that $\mathbf{R}_v^{\mathrm{T}}\mathbf{R}_v$ is also block diagonal, making its computation and storage very efficient.

The matrix $\mathbf{R}_h$ is a little more complex, as it describes relationships between adjacent profiles in the state vector, and thus adjacent matrix blocks. For the second order case it can be expressed in the block form as the $(N-2)\times N$ matrix.

$$\mathbf{R}_h = \begin{bmatrix} -\frac{1}{4}\mathbf{W}_h & \frac{1}{2}\mathbf{W}_h & -\frac{1}{4}\mathbf{W}_h & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{4}\mathbf{W}_h & \frac{1}{2}\mathbf{W}_h & -\frac{1}{4}\mathbf{W}_h & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -\frac{1}{4}\mathbf{W}_h & \frac{1}{2}\mathbf{W}_h & -\frac{1}{4}\mathbf{W}_h & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & -\frac{1}{4}\mathbf{W}_h & \frac{1}{2}\mathbf{W}_h & -\frac{1}{4}\mathbf{W}_h \end{bmatrix} \tag{4.6}$$

where the $n \times n$ diagonal matrix $\mathbf{W}_h$ describes as a function of height the *a priori* standard deviation of our virtual measurement that the derivative is zero. The complicated approach used to construct $\mathbf{W}_v$ is not required here as there is no desire to change the horizontal smoothing weight in the horizontal direction, only in the vertical. Extensions to higher order derivatives result in large block bandwidths, as described in section 3.2.4. As with the $\mathbf{K}$ matrix, if $\mathbf{R}$ is block band diagonal with a bandwidth of q, the matrix $\mathbf{R}^{\mathrm{T}}\mathbf{R}$ will be block band diagonal with a $2q$ bandwidth.

4.3 Inverting the normal equation matrix – Cholesky decomposition

Once formed, the inverse of the normal equation matrix is required in equation 3.17. When solving matrix equations of the form $\mathbf{Ax} = \mathbf{b}$ for $\mathbf{x}$, where $\mathbf{x}$ and $\mathbf{b}$ are n element vectors, and $\mathbf{A}$ is an $n \times n$ matrix, one does not necessarily need to compute the matrix $\mathbf{A}^{-1}$. In the case of matrices that are symmetric and positive definite (i.e., has positive eigenvalues), such as the normal equation matrix, the *Cholesky decomposition* approach is appropriate. This expresses the matrix $\mathbf{A}$ in terms of the product of a lower triangular matrix and its transpose,

$$\mathbf{A} = \mathbf{L}\mathbf{L}^{\mathrm{T}}. \tag{4.7}$$

Computing the Cholesky factor matrix $\mathbf{L}$ is a somewhat complex calculation usually involving of order n^3 operations where n is the total size of the matrix [Golub and VanLoan, 1996, for example]. We will later see that in our case this reduces to Np^2n^3.

Given this decomposition, the solution to $\mathbf{Ax} = \mathbf{b}$ can be easily obtained when one recognizes that equations involving triangular matrices are easy to solve (involving order n^2 computations). We express our equation as $\mathbf{L}\mathbf{L}^{\mathrm{T}}\mathbf{x} = \mathbf{b}$, and define $\mathbf{y} = \mathbf{L}^{\mathrm{T}}\mathbf{x}$. We first solve the triangular system $\mathbf{Ly} = \mathbf{b}$ for $\mathbf{y}$, following which we solve $\mathbf{L}^{\mathrm{T}}\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$.

In our case, while the full inverse of $\mathbf{A}$ is not needed for individual iterations, it is needed at the end of the calculation, as it represents the covariance of the solution. It can be obtained by solving n $\mathbf{Ax} = \mathbf{b}$ problems with $\mathbf{b}$ being the individual columns of the $n \times n$ identity matrix. An alternative approach

Table 4.1: A summary of the operation counts required for each iteration of the retrieval method. The 'independent' method retrieves the N profiles independently, as described in Section 4.1. The 'full' and 'sparse' methods both retrieve the profiles in blocks of N, the sparse method taking advantage of the simplicity of the system. The $\mathbf{S_y}$ matrix is assumed diagonal; if this is not the case, the first operation increases by a factor of at least m (possibly Nm.) Less computationally intensive terms (e.g., the formation of $\mathbf{K}^T\mathbf{S}_\mathbf{y}^{-1}[\mathbf{y}-\mathbf{f}]$) have been omitted, as have those such as $\mathbf{S}_\mathbf{a}^{-1}$ which can be computed outside the iteration loop.

Operation	Independent	Full	Sparse
$\mathbf{K}^T\mathbf{S}_\mathbf{y}^{-1}\mathbf{K}$	Nn^2m	N^3n^2m	Np^2n^2m
Cholesky factor	Nn^3	N^3n^3	Np^2n^3

is to compute the inverse of the matrix $\mathbf{L}$ (a simple operation for triangular matrices) and then form $\mathbf{A}^{-1} = \left[\mathbf{L}^{-1}\right]^T\mathbf{L}^{-1}$. This has the advantage that if only the diagonal elements of $\mathbf{A}^{-1}$ are required (as is often the case), efficiency can be obtained by not computing the off diagonal elements of this product.

4.4 Increasing efficiency in the retrieval calculation

Even when taking maximum advantage of the sparsity of the matrices involved, the MLS retrieval algorithm can be a time-consuming calculation. The aim of this section is to consider various approaches to improving the speed of the calculation, and to outline the manner in which decisions would be arrived at regarding which compromises are appropriate to achieve a given saving.

4.4.1 Operation counts

The key to the issue of improving the speed of the retrieval calculations is the number of individual scalar operations required by the algorithm. However, the number of operations required by a calculation does not directly relate to the time it takes to perform the calculation on a computer. Issues such as cache filling, parallel processing etc. have a big impact on the efficiency of a calculation. The operation counts serve merely as a 'rule of thumb' measure of the size of the task involved.

Table 4.1 gives a rough summary the number of operations required for the various stages in the EOS MLS retrieval algorithm. The summary assumes no sparsity within the profile / scan blocks at either the submatrix level (e.g. no allowance for the fact that $\partial[\texttt{R1A:118.B1F:PT}]/\partial[\text{ClO}] = 0$) or within the submatrices (e.g. the issue with $\vec{\boldsymbol{\zeta}}$ discussed in Section 4.2.4.)

In the MLS case $m \gg n$ for all the measurement vectors. From this it is clear that the ability to assume that $\mathbf{S_y}$ is diagonal is crucial if the algorithm is to be efficient. This factor rules out the use of constrained quantity error propagation as explained in Section 3.2.5. It is clearly more efficient to attempt to retrieve the whole state vector from every measurement vector simultaneously, as opposed to proceeding in separate phases and propagating errors for the constrained quantities.

Given that the $\mathbf{S_y}$ matrix is diagonal, the most computationally intensive step in the retrieval process is the construction of the $\mathbf{K}^T\mathbf{S}\mathbf{y}^{-1}\mathbf{K}$ matrices, taking Np^2n^2m operations. There are several points of interest to note about this calculation. Firstly, it is linear in N, that is to say, (ignoring the overlaps) retrieving 40 profiles in one single chunk will involve the same computational effort as retrieving two 20 profile chunks. This means that the size of N is limited only by the memory capacity of the computer used, with the limitation being the size of the normal equation matrix and the elements of $\mathbf{K}$ corre-

sponding to one major frame. Secondly, while the operation is linear in m_i, it is quadratic in n and p. While p has a geometrical origin, one source of improved efficiency would be to curtail the bandwidth considered by the forward model and the retrieval. The length of the state vector, n can also be changed, being a factor of the vertical resolution and range of the state vector quantities. Being a quadratic term, a factor of two increase in retrieval performance can result from removing only 30% of the elements of $\mathbf{x}$. m_i can also be changed fairly easily by removing or combining radiances that contribute little information to the system.

4.4.2 Phasing revisited

Although the use of constrained quantity error propagation has been ruled out on the grounds of being too computationally intensive, this does not necessarily rule out the use of retrieval phasing. Phasing can still be a very useful tool in the algorithm. However, unlike in previous descriptions, the previously retrieved quantities are not constrained, rather new items are added to the pool of retrieved quantities.

In this manner, the more non-linear quantities (tangent pressure and temperature in particular), that will require more iterations for convergence can be retrieved alone in the earlier phases. The later phases include quantities such as stratospheric composition, which, if retrieved alone, would require a smaller number of iterations (1 or 2.) The non-linear quantities are still retrieved in these later phases. However, as good convergence for these was achieved in the earlier phases, only small revisions should be needed. By using this scheme, the number of iterations required for the full system should be minimized.

Taking this one stage further, it is also possible in the earlier phases of the retrieval process to constrain quantities without propagating errors for them, or assuming that the modified $\mathbf{S_y}$ matrix is diagonal. As these early stages are only used to find a suitable starting point for the final 'full' retrieval, there is no need for a comprehensive treatment of the error budget.

It is even possible to conceive of more complex systems whereby the temperature is retrieved at full resolution in an early phase, and the results output as the standard temperature product. In the later phases, where composition is the main goal, and temperature is less important (and has little impact on the radiances), the vertical resolution and/or range of the temperature profile could be decreased in the state vector, in order to reduce the computation effort.

4.4.3 An 'Information' perspective on retrieval issues

The previous sections touched on the mechanisms whereby the retrieval algorithm can be made more efficient, namely, reducing the size of n or m. Clearly, some metric is needed to indicate which reduction schemes are preferable. One such metric is the *information content* of the retrieval system.

The information content of a system is a measure of the size (in fact 'inverse size') of the region in state space within which one is confident that the system is located. This region is described by the covariance matrix, which describes a hyperellipsoidal region in state space. The volume of this region is given by the product of the lengths of all the axes of the ellipsoid (give or take factors of π). These lengths are given by the square roots of the eigenvalues of the covariance matrix (recall that a covariance matrix is an inherently 'squared' quantity.) Thus the volume of state space is given by the square root of the product of the eigenvalues of the covariance matrix. Recall that the determinant of a matrix ($|\cdots|$) is equal to this product.

The information content (H) of a system is thus defined by the logarithm of the reciprocal of this volume, i.e.

$$H = \log_2 \left[\sqrt{|\mathbf{S}|}\right]^{-1} = -\frac{1}{2}\log_2 |\mathbf{S}| = \frac{1}{2}\log_2 \left|\mathbf{S}^{-1}\right|. \tag{4.8}$$

(Recall that $|\mathbf{M}|^{-1} = \left|\mathbf{M}^{-1}\right|$.) Base 2 is typically used for the logarithm, so that the information content can be described in terms of the number of 'bits' of information available.

In studies of retrieval systems it is useful to consider the information content added by the retrieval operation. Combining Equation 4.8 with Equation 3.18, describing the covariance of the retrieved product, gives the following expression for the information added by a retrieval calculation (recall that $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$):

$$\Delta H = [\text{Final information content}] - [\textit{a priori} \text{ information content}] \tag{4.9}$$

$$= \frac{1}{2}\log_2\left|\mathbf{S}_\mathbf{a}^{-1} + \mathbf{R}^\mathrm{T}\mathbf{R} + \sum_i \mathbf{K}_i^\mathrm{T}\mathbf{S}_i^{-1}\mathbf{K}_i\right| - \frac{1}{2}\log_2\left|\mathbf{S}_\mathbf{a}^{-1} + \mathbf{R}^\mathrm{T}\mathbf{R}\right| \tag{4.10}$$

$$\tag{4.11}$$

In practice, the determinant operation is highly numerically unstable, so it is best to apply a sequential approach to the calculation of the information content, adding measurements one at a time. The derivation of this calculation is somewhat complex; see Appendix C.2 for more details.

4.4.4 Vertical resolution

Given that the rate determining step in the retrieval calculation is order n^2m, cutting down the size of the state vector will give a significant increase in speed. The main way in which to decrease this size is to decrease the vertical resolution of the geophysical profiles in the state vector. However, too coarse a resolution will give a deterioration in information content. Figure 4.2 shows the trade-off between the vertical resolution and information gained for the ozone observations from band `R2:190.B6F:O3`.

The figure shows that little information is lost by going from twelve to six surfaces per decade for the O_3 profile. Given that this is a factor of two decrease in the state vector length, a quadrupling of inverse model speed could be seen if this were implemented. If similar savings can be made in all the other species, significant speed increases could be obtained. In addition, the vertical range of the profiles could be limited, or the resolution further degraded in certain regions of the profile (for example where the signal to noise ratio is poor.)

4.4.5 Radiance averaging / limiting

In addition to decreasing the resolution and / or range of the state vector components, the number of radiances used in the retrieval calculations can be reduced by either limiting the vertical range of the radiances used, or combining radiances from adjacent minor frames. Again, the information content of the retrieval system is the metric whereby decisions would be made regarding the most appropriate strategy. The form of the averaging or limiting would clearly vary from channel to channel, as the tangent point altitude range over which useful signals are obtained varies from channel to channel (see, for example, Figure 3.2.)

As a test case, a simple 'random walk' type algorithm has been implemented to determine the most advantageous scheme for a given reduction in the number of radiances used. Figure 4.3 shows the application of this algorithm to the `R2:190.B6F:O3` O_3 observations for the case where a 30% reduction in the effective number of radiances is required.

This example shows that rather than merging radiances from adjacent minor frames, the best way to reduce the number of radiances used in the retrieval calculation is to use the radiances at the full vertical resolution available, but over a limited vertical range, which varies from channel to channel.

There are other points of interest to note from this example. For example, while the channel closest to the line center would be thought to give the most information about the upper regions of the atmosphere, the calculation has chosen to concentrate on the information from the two pairs of channels

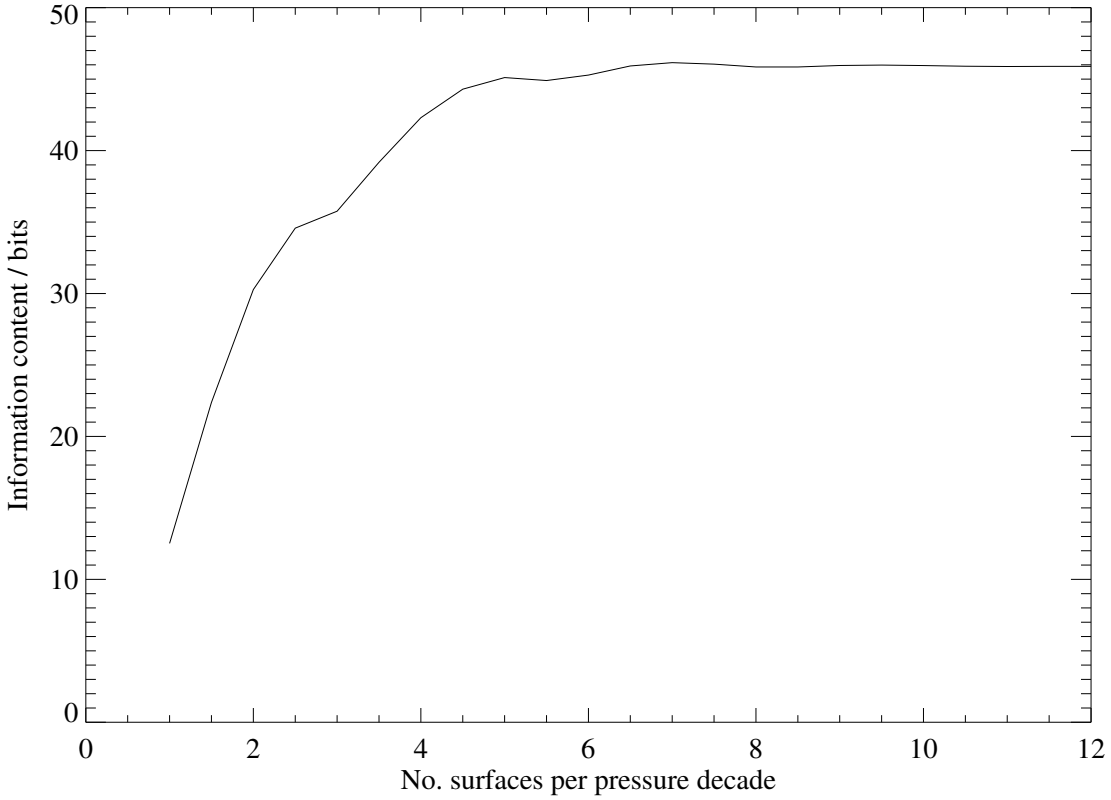
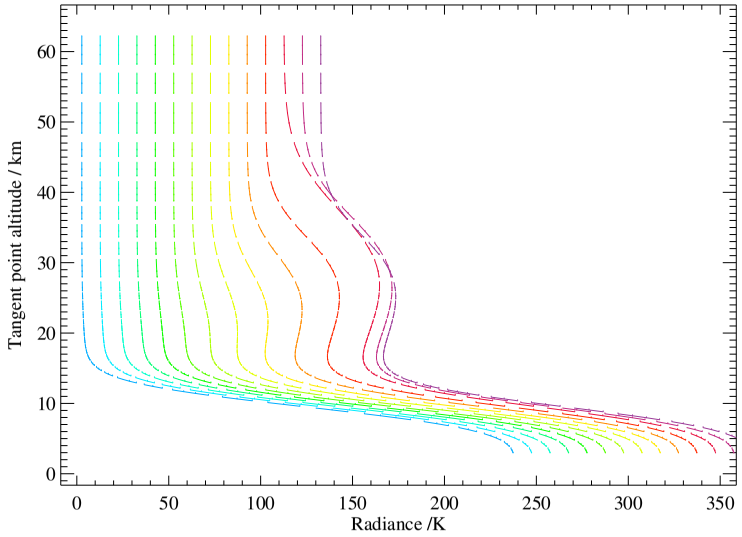


Figure 4.2: This plot shows the trade-off between resolution and information content for a system measuring O_3 from the `R2:190.B6F:O3` radiance observations (temperature and pressure are taken to be perfectly known.) The O_3 profile spans 1000 to 0.01 mb.

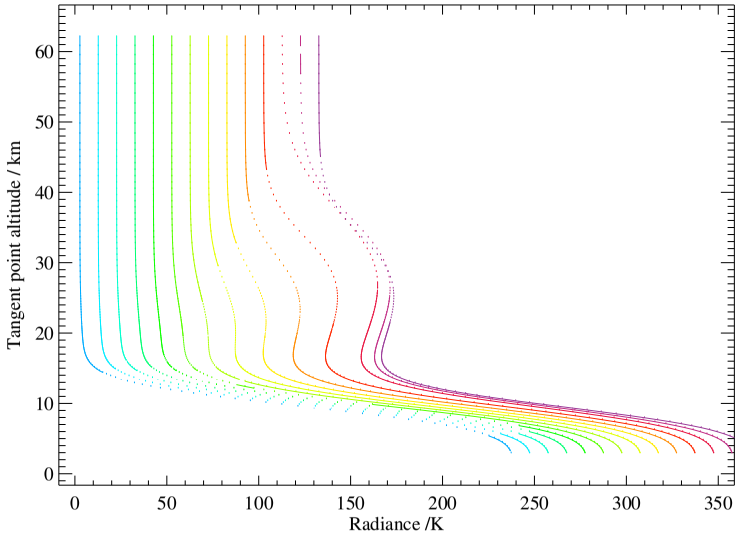
further away from the line center. This is due to the fact that the line center channel has a smaller band width and therefore a poorer signal to noise ratio than the channels further out. Figure 4.4 shows the optimum information content achieved by this search as a function of the fraction of the number of radiances used in the retrieval calculation.

In addition to averaging together radiances from multiple minor frames, it is also possible to average together radiances from different channels. While this is not discussed in detail here, a similar approach to the one outlined above can be used to arrive at optimum averaging schemes. This technique may be useful when using radiance observations from the digital autocorrelator spectrometers (DACs). These spectrometers provide $\sim$2 MHz resolution over $\sim$10 MHz, for each minor frame. The amount of information supplied by these measurements is very small compared to the data volume. Techniques such as these allow the autocorrelator data to be effectively used.

Note that reducing the size of the state vector can also increase the efficiency of the forward model calculation. However, reducing the number of radiances used does not necessarily speed up the forward model, as the field of view and frequency convolution calculations in the forward model require radiances at relatively high resolution over large vertical and frequency ranges, independently of how many radiances are actually required for the retrieval calculation.



(a) Initial down-sample scheme



(b) Final down-sample scheme

Figure 4.3: (color) This figure shows the outcome of a search for the 'optimum' radiance down-sampling scheme for a desired 30% reduction in the number of radiances used in a retrieval of ozone at six surfaces per decade from `R2:190.B6F:O3`. These plots show different schemes for averaging together radiance measurements in order to reduce the size of the measurement vector. Where radiances have been averaged together, they are connected with a line. Those radiances that have not been averaged are represented by a point.
The left hand figure shows the initial scheme used, with radiances being averaged together in groups of about three to achieve the 30% reduction in measurement vector size. The right hand figure shows the scheme chosen based on maximizing the information content of the resulting system, while keeping this 30% cut.
The maximum possible information content, were all the radiances used, is 45.29 bits, the information content for the initial system (left hand figure) was 34.59 bits. The final information content achieved for the optimal scheme was 38.22 bits.
Only the first 14 channels have been shown; the radiances curves have been successively offset by 10 K for clarity.

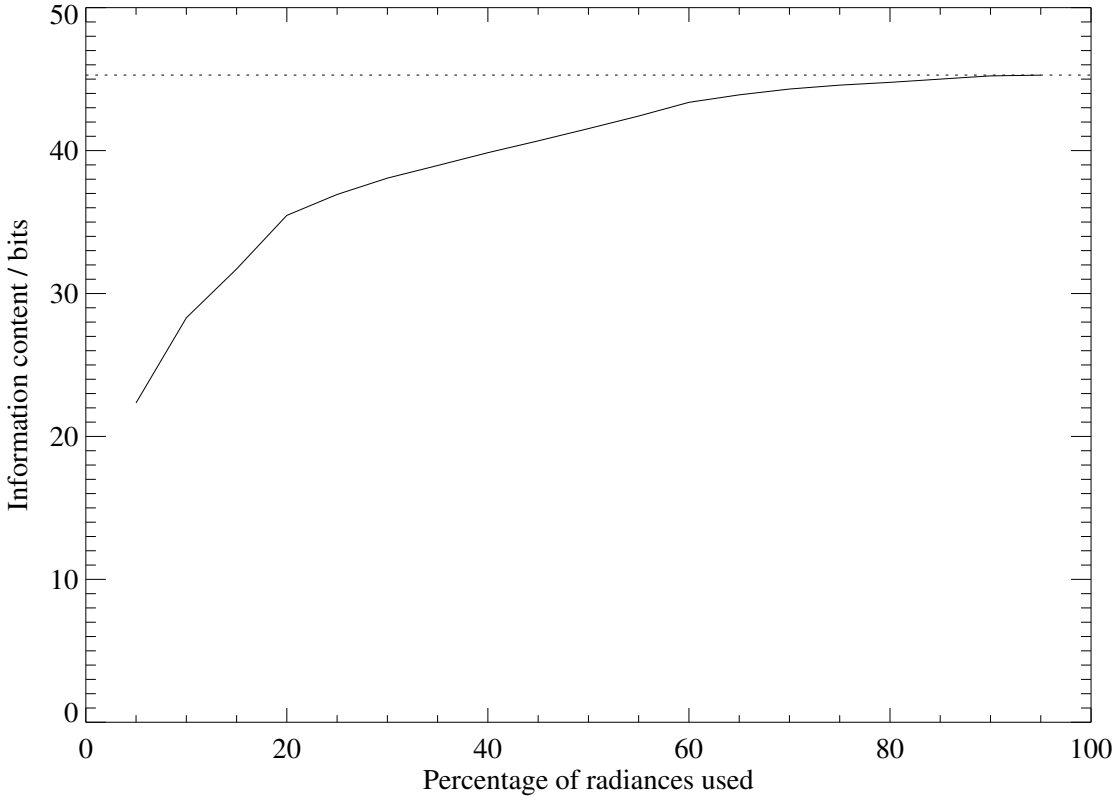


Figure 4.4: This figure shows the trade-off between the number of radiances used in the `R2:190.B6F:O3` ozone retrieval (at six surfaces per decade) and the optimum information content that can be obtained from the retrieval system using the given number of radiances.

4.4.6 Implementation of these schemes in the production processing

Section 6.4 details how the algorithms described in this document have been implemented in the 'launch ready' version of the MLS data processing software. In this implementation, the computational effort involved in the retrieval calculations has been found typically to account for less than 20% of the required effort, with the remaining time taken up almost entirely by the forward model. Accordingly, the issues discussed in this section have been given little emphasis in the algorithm development, as they would not lead to significant increases in overall performance. Instead, emphasis has been placed on improving efficiency in the forward model computations.

4.5 Numerical stability considerations

The issue of numerical stability is very important in these calculations. The contents of the state vector represents a huge dynamic range; consider the contrast between the dynamic ranges of temperature ($\sim$40 K) and ClO mixing ratio ($\sim 10^{-9}$). Care must be taken to avoid numerical round off errors when combining such quantities in sums.

The approach taken for the MLS retrieval algorithms is first to scale the measurement vectors by dividing each element by its corresponding error estimate, yielding a dimensionless value (the weighting function matrices are also scaled appropriately). This turns the $\mathbf{K}^{\mathrm{T}}\mathbf{S}_{\mathrm{y}}^{-1}\mathbf{K}$ calculation into $\widetilde{\mathbf{K}}^{\mathrm{T}}\widetilde{\mathbf{K}}$ where $\widetilde{\mathbf{K}}$

is the scaled weighting function matrix (as $\mathbf{S_y}$ is diagonal).

Once this scaling is performed, the normal equations can be safely formed. This is because the terms such as $\mathbf{S_a}^{-1}$ and $\mathbf{R}^T\mathbf{R}$ do not involve mixing different quantities within the state vector. Although the $\widetilde{\mathbf{K}}^T\widetilde{\mathbf{K}}$ terms do involve mixing of different quantities, this is in multiplication rather than addition. The resulting normal equation matrix does however still contain a large dynamic range (larger than the original state vector in fact). This needs to be addressed before Cholesky decomposition is attempted. The approach taken is to divide each row and column of the normal equation matrix by the square root of the corresponding diagonal element. This yields a scaled normal equation matrix with unity along the diagonal. Once the solution is obtained in this scaled space, the resulting state vector can be unscaled appropriately to return it to its appropriate physical meaning.

4.6 Testing for suitable convergence

In a Gauss-Newton iterative minimization, steps are taken on the assumption that the system is linear and that χ^2 therefore has a quadratic form. At each iteration it is possible to compute the χ^2 value that would be found at the minimum, were the problem truly linear (details of this calculation are given in Appendix C.3).

The convergence criterion used in the MLS retrieval algorithm compares the observed χ^2 to that predicted at the 'true' minimum each iteration (χ^2_{min}), and considers the retrieval converged when these agree to some appropriate degree. Currently a tolerance value of 2% is used (i.e., iterations cease when $\chi^2 < 1.02\chi^2_{\text{min}}$). In addition, a hard limit of a fixed number of iterations is also imposed, as there will be some cases where convergence is hard to achieve, even with many iterations.

Although the Marquardt-Levenberg approach does not rely on assumptions of linearity in choosing its step, it is still considering the same local conditions, and thus expects the same value of χ^2 to exist at the minimum, even though it may not choose to step to that location.

4.7 Summary of the retrieval algorithm

For clarity let us summarize all the steps involved in each iteration of the MLS retrieval algorithms. The algorithm is attempting to retrieve the state vector $\mathbf{x}$ given a set of measurement vectors $\mathbf{y}_i$ describing observations in different major frames, with associated diagonal covariance matrices $\mathbf{S}_i$. The *a priori* state is described by the vector $\mathbf{a}$ with it's associated covariance $\mathbf{S_a}$.

1. The first part of the algorithm is the computation of the forward model radiances and derivatives, the χ^2 statistic and the formation of the normal equation matrix, and the $\mathbf{S_a}^{-1}[\mathbf{a}-\mathbf{x}] + \mathbf{R}^T\mathbf{R}[\mathbf{a}-\mathbf{x}] + \mathbf{K}_i^T\mathbf{S}_i^{-1}[\mathbf{y}-\mathbf{f}]$ vector.

   The first terms considered in these are the contributions from the *a priori*. So, $\mathbf{S_a}^{-1}$ is added to the normal equation matrix, and $\mathbf{S_a}^{-1}[\mathbf{a}-\mathbf{x}]$ is added to the vector.

2. Following this, the contributions from the Tikhonov terms are included: $\mathbf{R}^T\mathbf{R}$ is added to the normal equation matrix, and $\mathbf{R}^T\mathbf{R}[\mathbf{a}-\mathbf{x}]$ is added to the vector.

3. Any required Marquardt-Levenberg term $\lambda\mathbf{I}$ is now added to the normal equations. The choice of value for λ is discussed in more detail in Appendix F.

4. The contributions from the radiances are then considered. As explained above, this takes the form of a loop over major frames. So the following stages are performed for each major frame in the chunk in turn.

(a) Firstly, the forward model is invoked to get an estimated set of radiances $\mathbf{f}_i$ corresponding to $\mathbf{y}_i$, and a corresponding set of weighting functions $\mathbf{K}_i$.

(b) The contributions of this major frame to the χ^2 statistic are computed.

(c) Following this $\mathbf{y}_i$, $\mathbf{f}_i$ and $\mathbf{K}_i$ are scaled by the measurement noise.

(d) The contributions of this major frame to the normal equation matrix and to the $\sum_i \mathbf{K}_i^{\mathrm{T}} \mathbf{S}_i^{-1} [\mathbf{y}_i - \mathbf{f}_i]$ vector is computed.

(e) The calculation proceeds to the next major frame.

5. Next we consider the χ^2 statistic arrived at above. For all but the first iteration, this is compared with that expected at the minimum (computed below); if suitable convergence is achieved, the iterations are halted and the final solution covariance (or its diagonal) is computed, along with any requested diagnostics such as averaging kernels.

6. For all but the first iteration, the χ^2 statistic is compared with that which was expected at this location (computed below). The level of agreement found is considered when choosing the Marquardt-Levenberg parameter for the next step.

7. The normal equation matrix is now scaled by its diagonal elements.

8. A Cholesky decomposition is performed on the normal equation matrix.

9. A change to $\mathbf{x}$ is computed (and appropriately unscaled), but not yet applied.

10. An estimate is made of the value of χ^2 at both the expected minimum and the current destination. This is described in more detail in Appendix C.3.

11. The step is taken and we return to the first stage.

There is a slight complication at step 6 where, if χ^2 is found to have increased as a result of the step taken, the retrieval may choose to retreat to the starting point and attempt a different step. This involves a slightly more complicated path through the steps above. The strategy for deciding when to do this is described in Appendix F.

Chapter 5
Related algorithms for EOS MLS 'noisy' products

5.1 Introduction

Some of the molecules EOS MLS is designed to observe have particularly small mixing ratios, and weak emission lines. The corresponding radiance observations thus have poor signal to noise ratios, leading to noisy retrievals. For these products, more useful results can be obtained by considering averaged products, such as daily zonal means, or monthly maps. There are several ways in which to compute such quantities.

5.2 Possible approaches

One approach is simply to retrieve the products in the same manner as all the others, and then use whatever averaging technique is appropriate afterwards. The disadvantage of this method is that, unless special care is taken, the *a priori* information can significantly bias the results, as it is included in each separate retrieval. This is the approach taken for the 'launch ready' version of the MLS data processing software (using appropriately large values for the *a priori* uncertainty for the species of interest).

A second approach is to average the radiances from the relevant bands in whatever manner is appropriate, and then to perform retrievals on the averaged radiances. This method has a profound problem however when the lines of interest are contaminated by strong emission from other, highly variable molecules. This is the case for example with the BrO observations in `R4:640.B31M:BrO`, which are close to a strong O_3 line.

5.3 The approach chosen

The best approach to this problem is to retrieve the averaged products as a separate task, after the main processing has occured. Rather than using averaged radiances as above, however, the full radiance data set for the relevant band is considered. Consider the iterative retrieval expression given in Equation 3.17

$$\mathbf{x}^{(r+1)} = \mathbf{x}^{(r)} + \left[\mathbf{S}_\mathbf{a}^{-1} + \mathbf{R}^\mathrm{T}\mathbf{R} + \sum_i \mathbf{K}_i^\mathrm{T}\mathbf{S}_i^{-1}\mathbf{K}_i\right]^{-1} \left\{\mathbf{S}_\mathbf{a}^{-1}\left[\mathbf{a} - \mathbf{x}^{(r)}\right] + \mathbf{R}^\mathrm{T}\mathbf{R}\left[\mathbf{a} - \mathbf{x}^{(r)}\right] + \sum_i \mathbf{K}_i^\mathrm{T}\mathbf{S}_i^{-1}\left[\mathbf{y}_i - \mathbf{f}_i(\mathbf{x}^{(r)})\right]\right\}. \quad (5.1)$$

In the linear (i.e. single iteration, with initial guess $\mathbf{x} = \mathbf{a}$) case, this reduces to

$$\mathbf{x} = \mathbf{a} + \left[\mathbf{S}_\mathbf{a}^{-1} + \mathbf{R}^\mathrm{T}\mathbf{R} + \sum_i \mathbf{K}_i^\mathrm{T}\mathbf{S}_i^{-1}\mathbf{K}_i\right]^{-1} \sum_i \mathbf{K}_i^\mathrm{T}\mathbf{S}_i^{-1}\left[\mathbf{y}_i - \mathbf{f}_i\left(\mathbf{a}\right)\right]. \quad (5.2)$$

Now for the case of the noisy products take $\mathbf{x}$ to be a specific component of an averaged dataset (e.g. a single profile corresponding to one latitude in a monthly zonal mean retrieval). Consider the measurement vectors $\mathbf{y}_i$ to represent each individual scan in the relevant spectral band that contributes to this component (e.g. all the scans in the latitude range under consideration that month.) The forward models for each scan use the previously retrieved values for the other molecules and parameters that affect the radiance measurements (O_3, temperature, tangent pressure etc.) as constrained quantities.

It is possible to take this method further by defining $\mathbf{x}_0$ as the value of the product retrieved by the standard Level 2 processing (we shall set aside the two dimensional issues for the moment and consider only a single profile retrieved from a single major frame). Let the vector $\mathbf{b}$ contain all the other aspects of the state retrieved by the Level 2 algorithms (ozone, temperature etc.).

It is clear therefore that

$$\mathbf{y}_i - \mathbf{f}_i\left(\mathbf{a}, \mathbf{b}\right) = \mathbf{y}_i - \left[\mathbf{f}_i\left(\mathbf{x}_0, \mathbf{b}\right) + \mathbf{K}\left(\mathbf{a} - \mathbf{x_0}\right)\right] \tag{5.3}$$

Equation 5.2 then reduces to:

$$\mathbf{x} = \mathbf{a} + \left[\mathbf{S}_\mathbf{a}^{-1} + \mathbf{R}^\mathrm{T}\mathbf{R} + \sum_i \mathbf{K}_i^\mathrm{T}\mathbf{S}_i^{-1}\mathbf{K}_i\right]^{-1}\left[\sum_i \mathbf{K}_i^\mathrm{T}\mathbf{S}_i^{-1}\left[\mathbf{y}_i - \mathbf{f}_i\left(\mathbf{x}_0, \mathbf{b}\right)\right] - \mathbf{K}_i^\mathrm{T}\mathbf{S}_i^{-1}\mathbf{K}_i\left[\mathbf{a} - \mathbf{x}_0\right]\right]. \tag{5.4}$$

Accordingly, if the Level 2 software can save the value of the $\mathbf{K}_i^\mathrm{T}\mathbf{S}_i^{-1}\mathbf{K}_i$ matrix and the $\mathbf{K}_i^\mathrm{T}\mathbf{S}_i^{-1}\left[\mathbf{y} - \mathbf{f}_i\left(\mathbf{x}_0\right)\right]$ vector for each retrieved profile, the 'noisy products' algorithm need not invoke any forward model calculations. All that is required is that values of the above matrix and vector are collated together appropriately, including the correction term $\mathbf{K}_i^\mathrm{T}\mathbf{S}_i^{-1}\mathbf{K}_i\left[\mathbf{a} - \mathbf{x}_0\right]$ where $\mathbf{x}_0$ is taken from the standard Level 2 product, and the final state computed as the result of equation 5.4.

Extending this to allow it to follow from a full two dimensional Level 2 calculation is achieved by defining $\mathbf{x}$ to be the mean of all or several profiles in the chunk and collapsing together the appropriate block columns of the $\mathbf{K}_i$ matrices. The issue of errors on constrained quantities has not been considered for this problem. It is possible that these should be considered, and a non-diagonal form for $\mathbf{S}_i$ used. This would make it harder for the 'noisy products' algorithm to avoid invoking forward model calculations. The issue will be investigated when these algorithms are developed.

Chapter 6
Implementation of the MLS Level 2 algorithms

This chapter describes the implementation and configuration of the MLS Level 2 algorithms as part of the MLS data processing software. It broadly divides into two parts. The first describes relevant aspects of the implementation of the algorithms given in chapters 3 and 4 in software, including issues such as data products, quality control and performance optimization. The second part describes in more detail the particular configuration of that software chosen for the 'launch ready' version of the MLS data processing system.

6.1 Products from the MLS Level 2 software

As described in chapter 3 the EOS MLS instrument observes the atmosphere in five different regions of the microwave spectrum (using seven radiometers). Many atmospheric species measured by MLS (e.g., ozone, nitric acid etc.) have spectral signatures in more than one of the radiometers. In the absence of significant systematic errors, the optimum retrieved observations of these species would be obtained by including in the retrieval calculation all the relevant radiance measurements. Such a data product would capitalize on the strengths of all the individual radiance observations, and represent the best observation of that species possible from the MLS data.

However, if systematic biases exist between the observations made by the different radiometers, these will hinder the ability of the retrieval algorithms to fit the different radiances concurrently. Such biases may arise due to uncertainties in the relative alignment of the radiometers, or perhaps due to uncertainties in the spectroscopy of molecules in the various spectral regions. For this reason it is desirable to retrieve estimates of the species separately from each radiometer, at least until any systematic effects are sufficiently understood and characterized.

Another reason why such separate products may be useful is that there may be periods where one of the MLS radiometers is switched off (though there is expected to be enough power available to run the usual complement continuously). The characteristics of a 'combined' product would change significantly as a result of such events. Scientists wishing to do detailed studies of trends in atmospheric species abundances would be better advised to consider the products retrieved from individual radiometer observations as being less sensitive to such changes in instrument configuration.

To avoid confusion for typical users of MLS data, the software will always output a set of 'standard' products. In the longer term, it is planned that these would be the results of a simultaneous retrieval from all the radiometers as described above. However, such a retrieval is considered too ambitious for the 'launch ready' setup. Instead the 'standard products' are chosen to be simply a copy of one of the products retrieved from an individual radiometer. For example, the 'standard ozone' product is a copy of the 'ozone retrieved from the 240 GHz radiometer' product.

6.2 Quality control, exception handling and related issues

6.2.1 Quality of retrieved data

In addition to retrieving an optimum state vector, the MLS data processing algorithms compute an estimated precision for each element of the state vector. This uncertainty is taken from the diagonal elements of the solution covariance matrix in equation 3.18.

As described in Section 3.2.3, the uncertainty on the retrieved result should always be compared with the uncertainty given on the *a priori* information. This comparison, along with the uncertainty information itself form a major part of quality control. As a simple flag to data users, the uncertainty is set negative if it is greater than half of the *a priori* uncertainty. In the longer term, there may be additional factors chosen to trigger the software to flag data with negative precision estimates, though none are currently implemented.

Another source of quality control information is the χ^2 information for the radiance measurements. Cases where the retrieval has failed to converge, or has converged on an inappropriate solution, or where the radiances are poor for some reason, may be indicated by a high value of χ^2. A complete set of χ^2 statistics are produced by the data processing algorithms, giving the values of χ^2 for each major frames worth of radiance observations for each band for each phase.

However, these χ^2 parameters are difficult for the typical user of MLS data to interpret. As an aid to the user, the χ^2 statistics for the radiances thought to be most relevant to the quality of an individual product are summarized in a parameter called `quality`, which has one value for each profile for that product. A value of `quality` of one would indicate excellent data (i.e., a low χ^2 for the relevant radiances); values of zero would indicate unacceptable data. The intermediate value of `quality` that should be considered as a 'cutoff' will be indicated to the users in the data quality document produced for each version of the MLS software.

In addition to the `quality` field, a parameter known as `status` is supplied with every profile of each species. This is a bit field with each bit indicating a factor users should take into account when using this profile. Meanings of bits will include 'do not use this profile' or 'this profile may have been affected by thick clouds'. As with `quality`, the detailed use of this information will be described in a data quality document for each MLS data version.

6.2.2 Bad or missing radiances

If a radiance observation is missing or marked bad for whatever reason, it is simply not included in the retrieval calculation. If several consecutive radiances are missing, such as a whole major frame's radiances for a band, retrieval is still possible, as the *a priori* information, along with information from the adjacent scans, will influence the retrieval for the corresponding profile. The retrieved uncertainties, however, reflect the comparative lack of information for the corresponding profile.

If several consecutive major frames' worth of radiances are missing, the retrieval algorithm takes this into account when dividing the dataset into chunks, using the boundaries of the region of missing data as the edges of the chunks.

6.2.3 Numerical exceptions

The retrieval calculations described here are sufficiently well posed and numerically stable (when scaled as described in Section 4.5) that occurrences such as division by zero, or requesting the square root of a negative number should never occur. For this reason, no special handling is needed for such events; any attempt to perform such a calculation will be indicative of a 'bug' in the program, and so should simply bring the processing to an immediate halt with an appropriate error message.

6.3 Suitability of the software to modern computer architectures

Chapter 4 described how the Level 2 software operates by dividing the input data into chunks of a fraction of an orbit. It was shown how (neglecting the effect of overlaps), the CPU time required to process N profiles in a chunk scaled linearly in N. The limiting factor on the size of the chunks therefore becomes the memory capacity of the computer concerned. It has also been shown that, by looping over the major frames within a chunk, the storage of the complete matrix of weighting functions is not required, only those relevant to individual major frames. The main storage limitations then become the storage of a fraction of the weighting functions, and of the normal equation matrix.

For the retrieval configuration chosen for the 'launch ready' software, this equates to a few hundred megabytes of memory, but generally less than one gigabyte. Additional memory ($\sim$200 Mb) is typically required for the forward model calculations and calibration/spectroscopy information. This amount of memory is a perfectly reasonable requirement for the typical contemporary computer and certainly not an amount usually associated with a 'supercomputer'.

The computation time is a more significant limitation. It is clearly essential to be able to process a complete day's worth of data in less than a day of real time in order to be able to keep up with the MLS observations. The simplest way to do this is to use a cluster of computers, with each computer in the cluster tasked with retrieving data from a single chunk. A 'master' task overseas the individual chunk tasks and arranges for the joining together of the results from each chunk into the final product.

6.4 The 'launch ready' configuration

As described above, the MLS retrieval algorithms proceed in a series of phases. The software that implements them is very flexible, and the choice of these phases can easily be changed. This section describes the configuration of these phases and the products they produce for the v1.4 'launch ready' version of the level 2 software.

6.4.1 Form of the output products

Section 4.4.4 detailed how choice of vertical resolution and range for the retrieved products can improve the efficiency of the retrieval calculation (forward model efficiency is also related to state vector size). In practice, scientific factors have been the main basis for choosing the state vector resolution. The grid chosen is related to the 'UARS' pressure levels, having six surfaces per decade change in pressure starting from 1000 hPa, corresponding to a spacing of about 2.5 km. Above 0.1 hPa, the resolution decreases to three surfaces per decade change in pressure.

In addition a 'high resolution' grid is used for some products, having twelve surfaces per decade from 1000 hPa to 22 hPa, reverting to the usual grid above this level. The products on this grid are produced from a special 'high resolution' phase of the retrieval designed to retrieve temperature and water vapor only. At the time of writing, it is not planned to include this phase in the operational retrieval software, though it may be run for periods of special interest.

All of the geophysical data products are produced on a horizontal grid having exactly 1.5 degree spacing in orbit geodetic angle (the 'fundamental' horizontal coordinate described in section 3.6.1), centered on the equator. This corresponds to a spacing along the track of around 160 km, and is a good match for the approximately 1.5 degree spacing of the MLS scans.

6.4.2 The Core, Core+Rn approach

The phasing for the flight-ready algorithms has been implemented in what is known as the 'Core, Core+Rn' approach. Several initial phases constitute 'Core', which together arrive at good estimates for temperature, geopotential height and tangent pressure for the GHz radiances using the 'R1A' (118 GHz) radiance observations. In addition, part of Core retrieves upper tropospheric humidity using some of the 190 GHz R2 radiances. The later 'Core+Rn' phases retrieve temperature, geopotential height and tangent pressure as in Core, along with species from the various radiometers. For example, 'Core+R2' retrieves temperature and tangent pressure, along with H_2O, N_2O, HNO_3, ClO, O_3, HCN and CH_3CN from the 'R2' (240 GHz) radiometer and the same 118 GHz radiances as used in Core.

Section 4.4.2 described how this approach to phasing – where the state and measurement vectors are augmented with additional products – is preferable to one where in the later retrievals the species retrieved earlier (e.g., temperature and tangent pressure) are constrained to the previously obtained values. In the latter method, either an extremely costly constrained quantity error calculation is required, or some approximation is needed, which can lead to an inaccurate error budget for the retrieved products.

It is certainly very beneficial to take the former approach when considering tangent pressure which, as has been explained before, is the main determinant of MLS radiance observations. Taking this approach also works well for temperature. However, upper tropospheric humidity is a very non-linear variable and difficult to retrieve under some (notably particularly wet) conditions. It was found in early testing that continuing to retrieve this in some of the later phases could lead to instability in the retrievals, which remains to be explored and understood.

Accordingly, our 'Core', 'Core+Rn' approach is not strictly in line with the description in Section 4.4.2, as retrievals of upper tropospheric humidity are omitted in Core+R3, Core+R4 and Core+R5, and this parameter is constrained to the results of Core+R2. The impact of neglecting error propagation for the constrained H_2O abundances has been shown to be insignificant in pre-launch testing. Figure 6.1 shows the phases implemented in the launch ready version of the software.

More details on the performance of the launch ready software can be found in the data quality document [Livesey, 2004]. More details of the implementation of the software are given in the functional requirements document [Livesey, 1999b] and the Level 2 users' guide [Livesey, 1999a] (currently under revision).

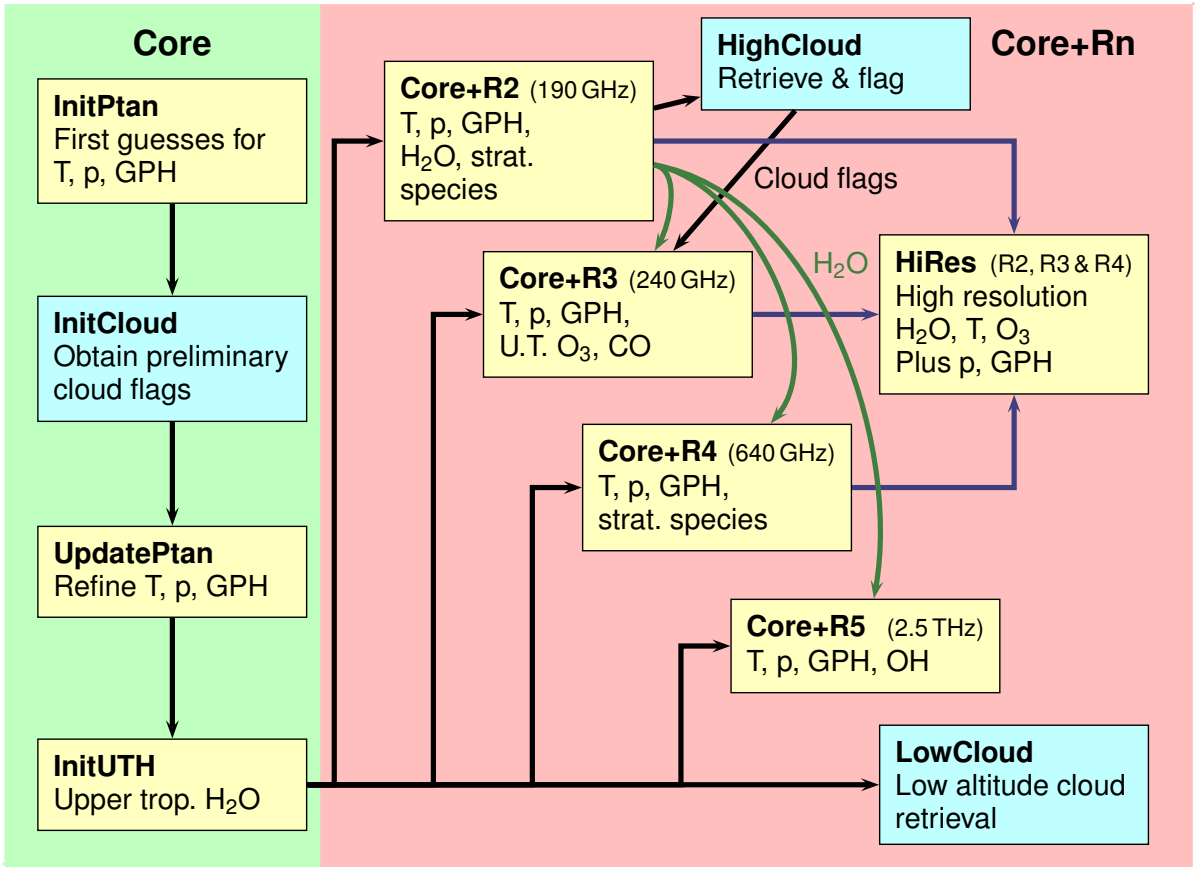


Figure 6.1: The phases implemented in the launch ready version of the MLS Level 2 data processing software. The high resolution phase is generally omitted for speed, but may be run for periods of special interest.

Appendix A
Algorithms for other MLS products

A.1 Tropopause pressure

One of the diagnostic quantities produced by the MLS Level 2 processing is the pressure at the tropopause. This quantity is very useful in many dynamical studies undertaken using MLS data, particularly those involving stratosphere/troposphere exchange. There are many different definitions of the tropopause. Currently the algorithms implement the World Meteorological Organization definition to obtain an estimate based on retrieved MLS temperature profiles. An alternative convention is the 'cold point' tropopause, which is simply the pressure of the coldest part of the temperature profile.

Other possible definitions that label a particular value of potential vorticity or potential temperature are often used. However, these definitions are not useful in the context of the MLS Level 2 software, which is unaware of these quantities. Later versions however may obtain these from meteorological datasets and produce a corresponding estimate of the tropopause pressure for each MLS retrieved profile.

A.2 Column products

In addition to providing profiles of atmospheric species, the Level 2 processing also outputs stratospheric column abundances of selected species (e.g. ozone.) These are obtained by integrating the abundance profiles from the tropopause (as obtained above) to the top of the retrieval range, and then converting the product into appropriate units (e.g. Dobson units).

The full details of the calculations can be found in the next section, which is a modified reproduction of an earlier document by W.G. Read and J.W. Waters.

A.3 Column abundances of MLS profiles.

This is to document expressions for column abundances in the vertical profiles retrieved from MLS.

Let $N(z_1, z_2)$ be the vertical column of molecules (per square meter) between heights z_1 and z_2. This is given by:

$$N(z_1, z_2) \quad = \quad \int_{z_1}^{z_2} f(z)\, n(z)\, \mathrm{d}z, \tag{A.1}$$

where $f(z)$ is the volume mixing ratio at height z of the species being considered, and $n(z)$ is the *total* ('air') number density at z.

We convert to pressure p, which is the vertical coordinate for MLS retrievals, by using hydrostatic equilibrium:

$$\mathrm{d}p(z) \quad = \quad -\rho(z)\, g(z)\, \mathrm{d}z, \tag{A.2}$$

where $\rho(z)$ is the mass density of air at z and $g(z)$ is the gravitational acceleration. Converting from mass density ρ to number density n, and neglecting the small variation[1] in g, (A.2) becomes

$$\mathrm{d}p(z) \;=\; -\frac{M\,g}{A}\,n(z)\;\mathrm{d}z, \tag{A.3}$$

where M is the 'effective' molecular weight of air and A = Avogadro's number (6.022×10^{23} molecules / mole). Using $n(z)\,\mathrm{d}z$ from (A.3) in (A.1) gives

$$N(z_1, z_2) = N(p_2, p_1) \;=\; \frac{A}{Mg}\int_{p_2}^{p_1} f(p)\;\mathrm{d}p, \tag{A.4}$$

where p_1 is the pressure at z_1 and p_2 is the pressure at z_2, and it is assumed that $p_1 > p_2$.

If the mixing ratio $f(p)$ is a constant between p_1 and p_2 then equation (A.4) becomes

$$N(p_2, p_1) \;=\; \frac{A}{Mg}\,f\,\Delta p \tag{A.5}$$

$$\;=\; \frac{(6.022\times10^{23}\ \text{molecules/mole})}{(28.97\ \text{g/mole})(9.71\ \text{ms}^{-2})}\,f\,\left(\Delta p \text{kg}\,\text{m}^{-1}\,\text{s}^{-2}\right) \tag{A.6}$$

$$\;=\; 2.14\times10^{24}\,f\,\Delta p\ \text{molecules}/m^2\ (\text{for } p \text{ in Pascals}) \tag{A.7}$$

$$\;=\; 2.14\times10^{26}\,f\,\Delta p\ \text{molecules}/m^2\ (\text{for } p \text{ in hPa}), \tag{A.8}$$

where

$$\Delta p \;\stackrel{\text{def}}{=}\; p_1 - p_2, \tag{A.9}$$

$g = 9.71\,\text{m/s}^2$ (representative of 45° latitude and 30 km altitude) and 28.97 g/mole for the molecular weight of air[2] have been used in (A.6). For $f = 1$ we obtain the vertical column of 'air' and (A.8) shows that an air layer of 'thickness' 1 hPa contains 2.14×10^{26} molecules per square meter.[3]

The mixing ratio profiles retrieved from MLS are given by

$$f(p) \;=\; \sum_j f_j\,\eta_j(p), \tag{A.10}$$

where $\eta_j(p)$ are the 'basis functions' used for representing the profile and f_j are retrieved coefficients. Putting (A.10) in (A.4) gives

$$N(p_1, p_2) \;=\; \frac{A}{Mg}\int_{p_2}^{p_1}\sum_j f_j\,\eta_j(p)\;\mathrm{d}p \tag{A.11}$$

$$\;=\; \frac{A}{Mg}\sum_j f_j\int_{p_2}^{p_1}\eta_j(p)\;\mathrm{d}p. \tag{A.12}$$

The basis functions currently used for MLS retrievals are triangular in log p and are sketched in figure A.1.

[1] The effective gravitational acceleration at Earth's surface varies 0.5% between $9.780\,\text{m/s}^2$ at the equator and $9.832\,\text{m/s}^2$ at the pole, and is $9.806\,\text{m/s}^2$ at 45° latitude. It also decreases 1% for each 32 km increase in height, for heights much less than Earth's radius. See, for example, Fleagle and Businger [1963] and Hess [1959].

[2] See, for example, Hess [1959].

[3] Note that this is independent of temperature, which determines the *height* thickness of the layer (through the gas law).

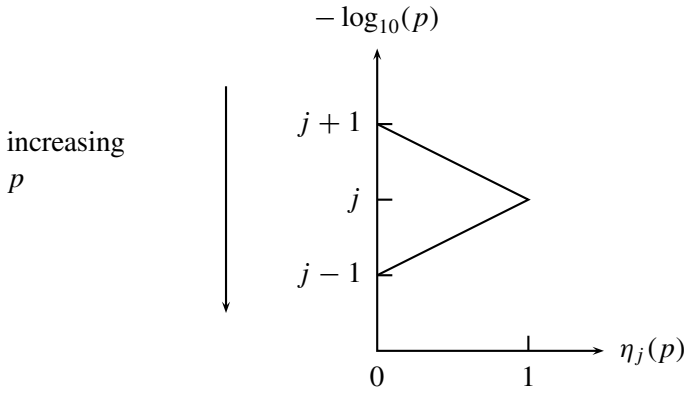


Figure A.1: The MLS basis functions.

The mathematical expressions for these basis functions are

$$\eta_j(p) = 0 \quad \text{for } p < p_{j+1} \tag{A.13}$$

$$\eta_j(p) = (\log_{10} p - \log_{10} p_{j+1})/\Delta_{j+}^{\log_{10} p} \quad \text{for } p_j \geq p \geq p_{j+1} \tag{A.14}$$

$$\eta_j(p) = (\log_{10} p_{j-1} - \log_{10} p)/\Delta_{j-}^{\log_{10} p} \quad \text{for } p_{j-1} \geq p \geq p_j \tag{A.15}$$

$$\eta_j(p) = 0 \quad \text{for } p > p_{j-1}, \tag{A.16}$$

where

$$\Delta_{j+}^{\log_{10} p} = \log_{10} p_j - \log_{10} p_{j+1} \tag{A.17}$$

$$\Delta_{j-}^{\log_{10} p} = \log_{10} p_{j-1} - \log_{10} p_j, \tag{A.18}$$

Using (A.13)–(A.18), the integral in (A.12) can be evaluated. It is broken up into that for the 'bottom' portion of the basis function ranging between $j-1$ and j (indicated by superscript '$-$'), and the 'top' portion ranging between j and $j+1$ (indicated by superscript '$+$'). We assume that the integration limits p_1, p_2 occur *outside* the region where $\eta_j(p)$ is non-zero.

First, for integration over the the 'bottom' portion of $\eta_j(p)$:

$$
\begin{aligned}
I_j^- &\stackrel{\text{def}}{=} \int_{p_j}^{p_{j-1}} \eta_j(p)\ \mathrm{d}p && \text{(A.19)}\\
&= \frac{1}{\Delta_{j-}^{\log_{10} p}} \int_{p_j}^{p_{j-1}} \left(\log_{10} p_{j-1} - \log_{10} p\right)\ \mathrm{d}p && \text{(A.20)}\\
&= \frac{1}{\Delta_{j-}^{\log_{10} p}\ \ln 10} \left\{ \ln p_{j-1} \int_{p_j}^{p_{j-1}} \mathrm{d}p - \int_{p_j}^{p_{j-1}} \ln p\ \mathrm{d}p \right\} && \text{(A.21)}\\
&= \frac{1}{\Delta_{j-}^{\log_{10} p}\ \ln 10} \left\{ \left(p_{j-1} - p_j\right) \ln p_{j-1} - \left(p \ln p - p\right) \Big|_{p_j}^{p_{j-1}} \right\} && \text{(A.22)}\\
&= \frac{1}{\Delta_{j-}^{\log_{10} p}\ \ln 10} \left\{ p_{j-1} \ln p_{j-1} - p_j \ln p_{j-1} - p_{j-1} \ln p_{j-1} + p_{j-1} + p_j \ln p_j - p_j \right\} && \text{(A.23)}\\
&= \frac{1}{\Delta_{j-}^{\log_{10} p}\ \ln 10} \left\{ -p_j \ln p_{j-1} + p_{j-1} + p_j \ln p_j - p_j \right\} && \text{(A.24)}\\
&= -p_j \left\{ \frac{\log_{10} p_{j-1} - \log_{10} p_j}{\Delta_{j-}^{\log_{10} p}} \right\} + \frac{p_{j-1} - p_j}{\Delta_{j-}^{\log_{10} p}\ \ln 10} && \text{(A.25)}\\
&= -p_j + \frac{1}{\ln 10}\ \frac{\Delta_{j-}^{p}}{\Delta_{j-}^{\log_{10} p}} && \text{(A.26)}\\
&= -p_j + \frac{\Delta_{j-}^{p}}{\Delta_{j-}^{\ln p}}, && \text{(A.27)}
\end{aligned}
$$

where

$$
\begin{aligned}
\Delta_{j-}^{p} &\stackrel{\text{def}}{=} p_{j-1} - p_j && \text{(A.28)}\\
\Delta_{j-}^{\ln p} &\stackrel{\text{def}}{=} \ln p_{j-1} - \ln p_j. && \text{(A.29)}
\end{aligned}
$$

(A.18) has been used in (A.26), and

$$
\int \ln p\ \mathrm{d}p = p \ln p - p \qquad \text{(A.30)}
$$

has been used in (A.22), and

$$
\log_{10} p = \frac{\ln p}{\ln 10} \qquad \text{(A.31)}
$$

has been used in (A.21,A.25,A.27).

Now integration over the the 'top' portion of $\eta_j(p)$ is performed in a similar fashion:

$$I_j^+ \stackrel{\text{def}}{=} \int_{p_{j+1}}^{p_j} \eta_j(p)\ \mathrm{d}p \tag{A.32}$$

$$= \frac{1}{\Delta_{j+}^{\log_{10} p}} \int_{p_{j+1}}^{p_j} \left(\log_{10} p - \log_{10} p_{j+1}\right)\ \mathrm{d}p \tag{A.33}$$

$$= \frac{1}{\Delta_{j+}^{\log_{10} p}\ \ln 10} \left\{ \int_{p_{j+1}}^{p_j} \ln p\ \mathrm{d}p - \ln p_{j+1} \int_{p_{j+1}}^{p_j} \mathrm{d}p \right\} \tag{A.34}$$

$$= \frac{1}{\Delta_{j+}^{\log_{10} p}\ \ln 10} \left\{ (p \ln p - p) \Big|_{p_{j+1}}^{p_j} - \left(p_j - p_{j+1}\right) \ln p_{j+1} \right\} \tag{A.35}$$

$$= \frac{1}{\Delta_{j+}^{\log_{10} p} \ln 10} \left\{ p_j \ln p_j - p_j - p_{j+1} \ln p_{j+1} + p_{j+1} - p_j \ln p_{j+1} + p_{j+1} \ln p_{j+1} \right\} \tag{A.36}$$

$$= \frac{1}{\Delta_{j+}^{\log_{10} p}\ \ln 10} \left\{ p_j \ln p_j\ -\ p_j + p_{j+1} - p_j \ln p_{j+1} \right\} \tag{A.37}$$

$$= p_j \left\{ \frac{\log_{10} p_j - \log_{10} p_{j+1}}{\Delta_{j+}^{\log_{10} p}} \right\} - \frac{p_j - p_{j+1}}{\Delta_{j+}^{\log_{10} p}\ \ln 10} \tag{A.38}$$

$$= p_j - \frac{1}{\ln 10}\ \frac{\Delta_{j+}^{p}}{\Delta_{j+}^{\log_{10} p}} \tag{A.39}$$

$$= p_j - \frac{\Delta_{j+}^{p}}{\Delta_{j+}^{\ln p}}, \tag{A.40}$$

where

$$\Delta_{j+}^{p} \stackrel{\text{def}}{=} p_{j+1} - p_j \tag{A.41}$$

$$\Delta_{j+}^{\ln p} \stackrel{\text{def}}{=} \ln p_j - \ln p_{j+1}. \tag{A.42}$$

The vertical column N_j represented by a *single* retrieval coefficient f_j is given by the j^{th} term in (A.12), with the integral evaluated between $j+1$ and $j-1$:

$$N_j = \frac{A}{Mg} f_j \int_{p_{j+1}}^{p_{j-1}} \eta_j(p)\ \mathrm{d}p \tag{A.43}$$

$$= \frac{A}{Mg} f_j \left[I_j^- + I_j^+ \right] \tag{A.44}$$

$$= \frac{A}{Mg} f_j \left[\frac{\Delta_{j-}^{p}}{\Delta_{j-}^{\ln p}} - \frac{\Delta_{j+}^{p}}{\Delta_{j+}^{\ln p}} \right]. \tag{A.45}$$

The EOS MLS retrievals use basis functions having

$$\Delta_{j+}^{\log_{10} p} = \Delta_{j-}^{\log_{10} p} = \frac{1}{12} \tag{A.46}$$

for all j. This gives

$$\Delta_{j+}^{\ln p} = \Delta_{j-}^{\ln p} = \frac{1}{12} \ln 10 = 0.1919 \tag{A.47}$$

and

$$\Delta^p_{j-} = \left(10^{1/12} - 1\right)\, p_j = 0.2115\, p_j \tag{A.48}$$

$$\Delta^p_{j+} = \left(1 - 10^{-1/12}\right)\, p_j = 0.1746\, p_j\,. \tag{A.49}$$

The column represented by the lower portion of the basis function is then

$$N^-_j = \frac{A}{Mg}\, f_j\, I^-_j \tag{A.50}$$

$$= \frac{A}{Mg}\left[\frac{12}{\ln 10}\left(10^{1/12} - 1\right) - 1\right] f_j\, p_j \tag{A.51}$$

$$= 0.102\, \frac{A}{Mg}\, f_j\, p_j \tag{A.52}$$

$$= 2.18\times 10^{25}\, f_j\, p_j \text{molecules/m}^2 \text{ (for } p_j \text{ in hPa)}, \tag{A.53}$$

and the column represented by the upper portion is

$$N^+_j = \frac{A}{Mg}\, f_j\, I^+_j \tag{A.54}$$

$$= \frac{A}{Mg}\left[1 - \frac{12}{\ln 10}\left(1 - 10^{-1/12}\right)\right] f_j\, p_j \tag{A.55}$$

$$= 0.090\, \frac{A}{Mg}\, f_j\, p_j \tag{A.56}$$

$$= 1.92\times 10^{25}\, f_j\, p_j \text{molecules/m}^2 \text{ (for } p_j \text{ in hPa)} \tag{A.57}$$

The column represented by the complete basis function, (A.45), becomes

$$N_j = \frac{A}{Mg}\, \frac{12}{\ln 10}\, \left(10^{1/12} + 10^{-1/12} - 2\right)\, f_j\, p_j \tag{A.58}$$

$$= 0.192\, \frac{A}{Mg}\, f_j\, p_j \tag{A.59}$$

$$= 4.12\times 10^{25}\, f_j\, p_j \text{molecules/m}^2 \text{ (for } p_j \text{ in hPa)}. \tag{A.60}$$

53% of the column represented by the basis function is in the lower portion and 47% is in the upper portion.

The total vertical column represented in the currently-retrieved MLS profiles is then

$$N = 4.12\times 10^{25} \sum_j f_j\, p_j \text{molecules/m}^2 \text{ (for } p_j \text{ in hPa)}. \tag{A.61}$$

For ozone it is convenient to express the column in Dobson Units (1 DU = 2.687×10^{20} molecules/m^2) and the mixing ratio in ppmv (1 ppmv = 1×10^{-6}). Equation (A.61) then becomes

$$\text{DU} = 0.153 \sum_j f_j\, p_j \text{(for } p_j \text{ in hPa and } f_j \text{ in ppmv)}. \tag{A.62}$$

The column *between* retrieval levels i and u is

$$N(p_i, p_u) = N^+_i + N^-_u + \sum_{j=i+1}^{j=u-1} N_j\,. \tag{A.63}$$

If f_j is a constant for all j, it can be shown that (A.63) reduces to (A.5).

The column between *adjacent* retrieval levels j and $j+1$ is

$$\begin{aligned} N(p_j, p_{j+1}) &= N_j^+ + N_{j+1}^- && \text{(A.64)} \\ &= \frac{A}{Mg}\left[1 - \frac{12}{\ln 10}\left(1 - 10^{-1/12}\right)\right] f_j\, p_j \\ &\quad + \frac{A}{Mg}\left[\frac{12}{\ln 10}\left(10^{1/12} - 1\right) - 1\right] f_{j+1}\, p_{j+1} && \text{(A.65)} \\ &= \frac{A}{Mg}\left[1 - \frac{12}{\ln 10}\left(1 - 10^{-1/12}\right)\right] f_j\, p_j \\ &\quad + \frac{A}{Mg}\left[\frac{12}{\ln 10}\left(1 - 10^{-1/12}\right) - 10^{-1/12}\right] f_{j+1}\, p_j && \text{(A.66)} \\ &= \frac{A}{Mg}\Big[0.090\, f_j + 0.084\, f_{j+1}\Big]\, p_j, && \text{(A.67)} \end{aligned}$$

where $p_{j+1} = 10^{-1/12}\, p_j$ has been used in (A.66).

It is interesting to compare (A.67) with the approximate expression that computes the column between two levels by assuming a constant mixing ratio equal to the average of retrieved values at j and $j+1$. This approximate expression is

$$\begin{aligned} N(p_j, p_{j+1}) &\approx \frac{A}{Mg}\Big[0.5\, f_j + 0.5\, f_{j+1}\Big]\Big[p_j - p_{j+1}\Big] && \text{(A.68)} \\ &\approx \frac{A}{Mg}\Big[0.5\, f_j + 0.5\, f_{j+1}\Big]\Big[1 - 10^{-1/12}\Big] p_j && \text{(A.69)} \\ &\approx \frac{A}{Mg}\Big[0.087\, f_j + 0.087\, f_{j+1}\Big] p_j. && \text{(A.70)} \end{aligned}$$

If the mixing ratio *is* constant between j and $j+1$, then (A.70) and (A.67) give the same answer — as they should. The error in (A.70) depends upon the the difference in f_j and f_{j+1}; worst-case error, when either f_j or f_{j+1} is zero, is approximately 10%.

Table A.1 summarizes expressions useful in calculating column abundances in the profiles retrieved from MLS.

Table A.1: Some useful expressions in calculating column abundances from retrieved MLS profiles.

Column abundance	Molecules/m^2 (note 1)	Dobson Units (note 2)
in lower portion of $\eta_j(p)$	$2.18\times10^{25}\, f_j\, p_j$	$0.071\, f_j\, p_j$
in upper portion of $\eta_j(p)$	$1.92\times10^{25}\, f_j\, p_j$	$0.080\, f_j\, p_j$
in all of $\eta_j(p)$	$4.12\times10^{25}\, f_j\, p_j$	$0.153\, f_j\, p_j$
between levels j and $j+1$	$\left(1.92 f_j + 1.80 f_{j+1}\right)\times10^{25} p_j$	$\left(0.071\, f_j + 0.067 f_{j+1}\right) p_j$

Note 1: f_j in vmr and p_j in hPa.
Note 2: f_j in ppmv and p_j in hPa.

Appendix B
Content of the EOS MLS state vector

Section 3.5 discussed the construction and contents of the EOS MLS state vector. This appendix gives full details of all the components of the state vector currently envisaged. These details can be found in Table B.1. Some terms used in the table are defined here.

Each row of the table describes a separate component of the state vector. The first column gives the name of the quantity (as used in the software). The meaning of the quantity is given in the second column. The third column describes the *type* of the quantity. The following types are used:

Minor frame quantities such as tangent pressures have one value or set of values for each EOS MLS minor frame.

Orbital quantities are typically represented as some function (e.g. an interpolation) of a coordinate such as ϕ, possibly with some additional long-term trend superimposed.

Constant quantities are literally constant in that a single value is used for all profiles / major frames.

Atmospheric profile quantities describe vertical profiles of geophysical parameters on fixed pressure surfaces.

Surface quantities are simply atmospheric quantities on one particular pressure surface.

The fourth column describes the units for the quantity, and the last column describes the expected usage of the quantity.

Retrieved quantities are routinely retrieved in the MLS data processing (though not necessarily in all phases.)

Derived quantities are derived from other (typically retrieved) quantities in the state vector. For example column abundances of ozone are derived from the retrieved ozone profile.

Constrained quantities are, as the name implies, constrained to an *a priori* value. However, in most cases errors are not propagated for these quantities.

Characterized quantities are constrained (as above) in routine processing. However, these quantities have sufficient impact on the direct measurements that their values could be retrieved in 'off-line' studies undertaken to characterize the instrument. These 'optimal' values can then be constrained in production processing.

Table B.1: Contents of EOS MLS state vector. See the text for an explanation of the terms and notes

Name	Meaning	Type	Units	Usage
Tangent point geometry				
pTanGHz	Nominal tangent pressure for the GHz radiance observations	Minor frame	$-\log_{10}$ [hPa]	Retrieved
pTanTHz	Nominal tangent pressure for the THz radiance observations	Minor frame	$-\log_{10}$ [hPa]	Retrieved
phiTanGHz	Nominal tangent geodetic angle for the GHz radiance observations	Minor frame	degrees	Constrained
phiTanTHz	Nominal tangent geodetic angle for the THz radiance observations	Minor frame	degrees	Constrained
Temperature				
Temperature	Atmospheric temperature	Atmospheric profile	K	Retrieved
Temperature-HR	Temperature on the 'high resolution' vertical grid	Atmospheric profile	K	Retrieved
Tropopause pressure				
tpPressure	Tropopause pressure	Surface	hPa	Derived
Geopotential height terms				
refGPH	Geoptential height of a 100 hPa reference surface	Surface	m	Retrieved
GPH	Geopotential height profile (computed from Temperature and refGPH)	Atmospheric profile	m	Derived
GPH-HR	Geopotential height profile on the 'high resolution' vertical grid	Atmospheric profile	m	Derived
Species abundances				
BrO	BrO mixing ratio	Atmospheric profile	vmr	Retrieved
CH3Cl	CH_3Cl mixing ratio	Atmospheric profile	vmr	Retrieved
And so on for CH3CN, CH3COCH3, CH3OH, CO, ClO, H2CO, H2O2, H2S, HCN, HCl, HNO3, HO2, HOCl, N2, N2O, NO, NO2, O2, O3, OCS, OH and SO2				
H2O-HR	H_2O mixing ratio on the 'high resolution' vertical grid	Atmospheric profile	vmr	Retrieved
Relative humidity				
RHI	Relative humidity with respect to ice	%	Derived	

continued on next page...

. . . continued from previous page

Name	Meaning	Type	Units	Usage
RHI-HR	Relative humidity with respect to ice on the 'high resolution' vertical grid	%	Derived	
Column abundances				
Column-BrO	Column abundance of BrO above the tropopause as defined by tpPressure	Surface	Dobson units	Derived
And so on for CH3Cl, CH3CN, CH3COCH3, CH3OH, CO, ClO, H2CO, H2O2, H2S, HCN, HCl, HNO3, HO2, HOCl, N2, N2O, NO, NO2, O2, O3, OCS, OH and SO2				
Cloud properties				
IWC	Ice water content	Atmospheric profile	$g\,m^{-3}$	Derived/retrieved
IWC-Fine	Ice water content on the 'fine' horizontal grid.	Atmospheric profile	$g\,m^{-3}$	Derived/retrieved
LWC	Liquid water content	Atmospheric profile	$g\,m^{-3}$	Constrained
For more information on these and other cloud properties see Wu and Jiang [2004].				
Ratios for isotopes and excited states				
isotopeRatioBrO-79-O	Isotope ratio for $Br^{79}O$	Constant	Dimensionless	Constrained
And so on for BR-79-O-V1, BR-81-O, BR-81-O-V1, CH3COCH3, CH3CN, CH3CL-35, CH3CL-37, CH3OH, CO, C-13-O, CO-17, CO-18, CL-35-O, CL-35-O-V1, CL-37-O, CL-37-O-V1, COF2, CL-35-OOCL-35, CL-35-OOCL-37, CL-35-ONO2, CL-37-ONO2, HBR-79, HBR-81, HCL-35, HCL-37, DCL-35, DCL-37, HCN, HCN-V2, HC-13-N, HCN-15, HCOOH, HF, DF, H2CO, H2C-13-O, H2CO-18, H2O, H2O-R1A, H2O-R1B, H2O-R2, H2O-R3, H2O-R4, H2O-R5H, H2O-R5V, H2O-V2, H2O-17, H2O-17-V2, H2O-18, H2O-18-V2, D2O, HDO, HDO-18, H2O2, H2S-32, H2S-32-O4, HNO3, HNO3-V5, HNO3-V6, HNO3-V7, HNO3-V8, HNO3-V9, HOBR-79, HOBR-81, HOCL-35, HOCL-37, HO2, HOONO2, NO, NO2, N2, N2O, N2O-V1, N2O-V2, N2O-2V2, N-15-NO, NN-15-O, N2O-18, O, OCO-18, OCL-35-O, OCL-37-O, OCS-32, OCS-34, OC-13-S-32, OH, O-18-H, OD, O2, O2-V1, SNGLT-DLTA-O2, O-17-O, O-18-O, O3, O3-V1-3, O3-V2, O3-2V2, O3-V1-3-V2, O3-ASYM-O-17, O3-SYM-O-17, O3-ASYM-O-18, O3-ASYM-O-18-V2, O3-SYM-O-18, O3-SYM-O-18-V2, S-32-O2, S-32-O2-V2, S-33-O2, S-34-O2				
Extinction terms				
ExtinctionR1A	Extinction in the 118 GHz spectral region (radiometer R1A)	Atmospheric profile	km^{-1}	Constrained
And so on for ExtinctionR1B, ExtinctionR2, ExtinctionR3, ExtinctionR4, ExtinctionR5H and ExtinctionR5V				
Baseline terms				
BaselineR1A	Additive radiance baseline for the 118 GHz spectral region (radiometer R1A)	Atmospheric profile	K	Retrieved

continued on next page. . .

. . . continued from previous page

Name	Meaning	Type	Units	Usage
And so on for `BaselineR1B`, `BaselineR2`, `BaselineR3`, `BaselineR4`, `BaselineR5H` and `BaselineR5V`				
`BaselineFineR1A`	Additive radiance baseline for the 118 GHz spectral region (radiometer R1A) on a fine horizontal grid, used in cloud retrievals.	Atmospheric profile	K	Retrieved
And so on for `BaselineFineR1B`, `BaselineFineR2`, `BaselineFineR3`, `BaselineFineR4`, `BaselineFineR5H` and `BaselineFineR5V`				
Sideband fractions				
`LimbSidebandFraction1L`	'Limb' sideband fraction for the lower sideband of band 1	Constant	dimensionless	Constrained
`LimbSidebandFraction1U`	'Limb' sideband fraction for the upper sideband of band 1	Constant	dimensionless	Constrained
And so on for bands 2–34.				
Elevation offset angles				
`Elev1L`	Elevation offset angle for the lower sideband of band 1	Constant	degrees	Constrained
`Elev1U`	Elevation offset angle for the upper sideband of band 1	Constant	degrees	Constrained
And so on for bands 2–34.				
Magnetic field terms				
`magneticField`	Earth's magnetic field	Atmospheric profile (xyz-vector)	Gauss	Derived/Constrained
`fieldAzimuth`	Azimuth angle for Earth's magnetic field	Atmospheric profile	degrees	Derived
`fieldElevation`	Elevation angle for Earth's magnetic field	Atmospheric profile	degrees	Derived
`fieldStrength`	Strength of Earth's magnetic field	Atmospheric profile	Gauss	Derived
Geometric / orbital terms				
`orbitInclination`	Inclination of spacecraft orbit	Constant	degrees	Constrained.
`scGeocAlt`	Spacecraft geocentric altitude	Minor frame	m	Constrained.
`tngtGeocAltGHz`	Tangent point geocentric altitude for GHz radiances	Minor frame	m	Constrained.
`tngtGeocAltTHz`	Tangent point geocentric altitude for THz radiances	Minor frame	m	Constrained.
`scECI`	Spacecraft location in Earth Centered Inertial (ECI) frame	Minor frame (xyz vector)	m	Constrained.
`tngtECIGHz`	Tangent point location for GHz radiances in ECI frame	Minor frame (xyz vector)	m	Constrained.

continued on next page. . .

. . . continued from previous page

Name	Meaning	Type	Units	Usage
tngtECITHz	Tangent point location for THz radiances in ECI frame	Minor frame (xyz vector)	m	Constrained.
scVelECR	Spacecraft velocity in Earth Centered Rotating (ECR) coordinates	Minor frame (xyz vector)	$m\,s^{-1}$	Constrained.
losVelGHz	Line of sight (LOS) velocty for GHz radiances	Minor frame	$m\,s^{-1}$	Constrained.
losVelTHz	Line of sight velocty for THz radiances	Minor frame	$m\,s^{-1}$	Constrained.
ECRtoFOV	Rotation matrix for converting ECR to field of view coordinate system	Minor frame (9 element matrix)	dimensionless	Constrained.
Other minor terms				
earthReflectivity	Earth surface albedo	Constant	Dimensionless	Constrained
spaceRadiance	Microwave background temperature	Constant	K	Constrained
.				

Appendix C
Details of formulae used in this document.

C.1 Calculus of vectors and matrices

Many of the manipulations in retrieval theory involve differentiating an expression with respect to a vector quantity. If the expression is a scalar, a vector of derivatives is obtained, each element representing the derivative of the scalar with respect to the corresponding element of the vector, as in

$$\left[\frac{\partial a}{\partial \mathbf{x}}\right]_i = \frac{\partial a}{\partial x_i}. \tag{C.1}$$

The derivative of a vector quantity with respect to another vector quantity can be represented by a matrix, according to

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = D_{ij} = \frac{\partial y_i}{\partial x_j}. \tag{C.2}$$

One can also construct vector equivalents for the various rules commonly associated with scalar calculus. In the following derivations consider vectors $\mathbf{x}$ and $\mathbf{y}$, with $\mathbf{y}$ depending on $\mathbf{x}$ according to $\mathbf{D} = \partial\mathbf{y}/\partial\mathbf{x}$. Many of the retrieval theory calculations involve evaluating expressions such as

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{A}\mathbf{y}, \tag{C.3}$$

where $\mathbf{A}$ is a constant matrix. In order to evaluate these, it is necessary to consider individual components of the result.

$$\left[\frac{\partial}{\partial \mathbf{x}}\mathbf{A}\mathbf{y}\right]_{ij} = \frac{\partial}{\partial x_j}\sum_k A_{ik}y_k = \sum_k A_{ik}\frac{\partial y_k}{\partial x_j} = \sum_k A_{ik}D_{kj} = [\mathbf{A}\mathbf{D}]_{ij}\ . \tag{C.4}$$

Another identity, involving the derivative of a scalar quantity with respect to a vector, is also common in retrieval problems. Again, the solution is easily found by considering components:

$$\left[\frac{\partial}{\partial \mathbf{x}}\mathbf{y}^{\mathrm{T}}\mathbf{A}\mathbf{y}\right]_i = \frac{\partial}{\partial x_i}\sum_{jk} y_j A_{jk} y_k = \sum_j \frac{\partial y_j}{\partial x_i}A_{jk}y_k + y_j A_{jk}\frac{\partial y_k}{\partial x_i}. \tag{C.5}$$

A little thought is required before this expression can be recast into a matrix; several possible expressions result, one of which is

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{y}^{\mathrm{T}}\mathbf{A}\mathbf{y} = \mathbf{D}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{y} + \mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{y}. \tag{C.6}$$

C.2 Details of the incremental information content calculation

Section 4.4.3 introduced the concept of 'information content' of a system, being related to the determinant of the covariance matrix. Computing the determinant of a matrix is an inherently unstable

calculation[1]; instead a method is derived here that computes the information content by considering the incremental improvements made by introducing individual measurements one by one.

The incremental information content calculation is described in Rodgers [1996]; the derivation is summarized here for clarity, and to set it in the context of the MLS case.

Equation 3.14 gave the covariance of the retrieved state vector as $\mathbf{S_x} = \left[\mathbf{S_a}^{-1} + \sum_i \mathbf{K}_i^{\mathrm{T}}\mathbf{S}_i^{-1}\mathbf{K}_i\right]^{-1}$ (we are neglecting the Tikhonov smoothing terms for simplicity here). In this case we consider systems with only one measurement vector; this gives the covariance as

$$\mathbf{S_x} = \left[\mathbf{S_a}^{-1} + \mathbf{K}^{\mathrm{T}}\mathbf{S_y}^{-1}\mathbf{K}\right]^{-1}. \tag{C.7}$$

It can be shown (through a somewhat complex series of manipulations, see Rodgers [2000]) that this is equivalent to

$$\mathbf{S_x} = \mathbf{S_a} - \mathbf{S_a}\mathbf{K}^{\mathrm{T}}\left[\mathbf{K}\mathbf{S_a}\mathbf{K}^{\mathrm{T}} + \mathbf{S_y}\right]^{-1}\mathbf{K}\mathbf{S_a}. \tag{C.8}$$

Now consider the case where the measurement covariance matrix $\mathbf{S_y}$ is diagonal, and the measurements are entered sequentially as scalar values with variances σ_j^2. In this case, Equation C.8 can become an iterative expression

$$\begin{aligned}\mathbf{S}_{\mathbf{x}}^{(j)} &= \mathbf{S}_{\mathbf{x}}^{(j-1)} - \frac{\mathbf{S}_{\mathbf{x}}^{(j-1)}\mathbf{k}_j^{\mathrm{T}}\mathbf{k}_j\mathbf{S}_{\mathbf{x}}^{(j-1)}}{\mathbf{k}_j\mathbf{S}_{\mathbf{x}}^{(j-1)}\mathbf{k}_j^{\mathrm{T}} + \sigma_j^2} \\ &= \mathbf{S}_{\mathbf{x}}^{(j-1)}\left[\mathbf{I}_n - \frac{\mathbf{k}_j^{\mathrm{T}}\mathbf{k}_j\mathbf{S}_{\mathbf{x}}^{(j-1)}}{\mathbf{k}_j\mathbf{S}_{\mathbf{x}}^{(j-1)}\mathbf{k}_j^{\mathrm{T}} + \sigma_j^2}\right],\end{aligned} \tag{C.9}$$

where $\mathbf{S}_{\mathbf{x}}^{j}$ is the covariance of the solution after introducing measurement j with $\mathbf{S}_{\mathbf{x}}^{(0)} = \mathbf{S_a}$, and $\mathbf{k}_j$ is the weighting function for the jth measurement, i.e. the jth row of $\mathbf{K}$.

The additional information contributed by measurement j is, from Equation 4.8, given by

$$\delta H_j = \frac{1}{2}\log_2\left|\mathbf{S}_{\mathbf{x}}^{(j-1)}\right| - \frac{1}{2}\log_2\left|\mathbf{S}_{\mathbf{x}}^{(j)}\right|, \tag{C.10}$$

which, when applying the identities $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$ and $\left|\mathbf{I} + \mathbf{a}\mathbf{b}^{\mathrm{T}}\right| = 1 + \mathbf{b}^{\mathrm{T}}\mathbf{a}$, where $\mathbf{a}$ and $\mathbf{b}$ are column vectors, gives

$$\delta H_j = -\frac{1}{2}\log_2\left[1 - \frac{\mathbf{k}_j^{\mathrm{T}}\mathbf{S}_{\mathbf{x}}^{(j-1)}\mathbf{k}_j}{\mathbf{k}_j^{\mathrm{T}}\mathbf{S}_{\mathbf{x}}^{(j-1)}\mathbf{k}_j + \sigma_j^2}\right] = \frac{1}{2}\log_2\left[1 + \frac{\mathbf{k}_j^{\mathrm{T}}\mathbf{S}_{\mathbf{x}}^{(j-1)}\mathbf{k}_j}{\sigma_j^2}\right]. \tag{C.11}$$

This calculation is significantly more stable and also more efficient than computing the full determinant of the solution covariance.

C.3 Computing 'expected' and 'minimum' χ^2 values.

Equation 3.4 defined χ^2 as

$$\chi^2 = \sum_i \left[\mathbf{y}_i - \mathbf{f}_i(\mathbf{x})\right]^{\mathrm{T}}\mathbf{S}_i^{-1}\left[\mathbf{y}_i - \mathbf{f}_i(\mathbf{x})\right].$$

[1]The determinant is the product of the eigenvalues of the matrix, consider the case of a 100×100 matrix whose eigen values are $\sim 10^{-6}$. This would have a determinant of 10^{-600} — too small to be represented on most computer architectures.

For clarity in this discussion, rather than breaking out the *a priori* and smoothing terms, or having a sum over a series of measurement vectors, we will collapse all our measurements, both real and virtual, together into one large measurement vector. This equates to stacking all the $\mathbf{K}_i$ matrices together into one large long thin matrix. Thus χ^2 is given by

$$\chi^2 = [\mathbf{y} - \mathbf{f}]^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} [\mathbf{y} - \mathbf{f}] . \tag{C.12}$$

Let us assume that a step $\mathbf{x} \rightarrow \mathbf{x} + \delta$ is being considered in the minimization. Let us also define $\Delta = (\mathbf{y} - \mathbf{f})$, giving $\chi^2 = \Delta^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \Delta$. The Gauss-Newtonian iteration with the Marquardt-Levenberg approach would define δ as

$$\delta = \left[\lambda \mathbf{I} + \mathbf{K}^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \mathbf{K}\right]^{-1} \mathbf{K}^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \Delta, \tag{C.13}$$

where λ is the Marquardt-Levenberg scalar. In the assumption that the system is linear, one would expect the vector $\mathbf{f}$ to change by $\mathbf{K}\delta$. This would modify χ^2 to

$$\hat{\chi}^2 = [\Delta - \mathbf{K}\delta]^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} [\Delta - \mathbf{K}\delta] \tag{C.14}$$

Expanding this out we obtain

$$\begin{aligned} \hat{\chi}^2 &= \Delta^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \Delta - \Delta^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \mathbf{K}\delta - \delta^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \Delta + \delta^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \mathbf{K}\delta \\ &= \chi^2 - 2\delta^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \Delta + \delta^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \mathbf{K}\delta. \end{aligned} \tag{C.15}$$

This can be computed efficiently by the software because the vector $\mathbf{K}^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \Delta$ and the matrix $\mathbf{K}^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \mathbf{K}$ has been computed as part of the formation of the normal equations.

The χ^2 expected at the minimum were the system linear is simply given by the above expression with $\lambda = 0$. Expanding δ and simplifying we find the pleasingly simple result:

$$\begin{aligned} \hat{\chi}^2 &= \chi^2 - 2\Delta^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \mathbf{K}^{\mathrm{T}} \left[\mathbf{K}^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \mathbf{K}\right]^{-1} \mathbf{K}^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \Delta + \\ &\qquad \Delta^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \mathbf{K}^{\mathrm{T}} \left[\mathbf{K}^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \mathbf{K}\right]^{-1} \mathbf{K}^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \mathbf{K} \left[\mathbf{K}^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \mathbf{K}\right]^{-1} \mathbf{K}^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \Delta \\ &= \chi^2 - \Delta^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \mathbf{K}^{\mathrm{T}} \left[\mathbf{K}^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \mathbf{K}\right]^{-1} \mathbf{K}^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \Delta \\ &= \chi^2 - \delta^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} \mathbf{S}_{\mathbf{y}}^{-1} \Delta \end{aligned} \tag{C.16}$$

Appendix D
Notation conventions.

While the use of a consistent notation convention is desirable, it should not be achieved at the expense of reduced clarity. It would be inappropriate, for example, to use any character other than g to describe the acceleration due to gravity. Hence, while there are exceptions to these conventions, in general the rules hold throughout the document.

Scalars: Scalars are represented by italic characters, e.g. α, i, n, M.

Vectors: Vectors are shown as bold lower case characters, such as $\mathbf{x}$. To describe individual elements of a vector, the corresponding italic character is subscripted, so $[\mathbf{x}]_i = x_i$. Where the bold character is subscripted with an italic index, this indicates a specific vector from a set of vectors. So $\mathbf{y}_i$ is the i'th measurement vector.

Matrices: Bold upper case characters indicate matrices (e.g. $\mathbf{A}$). Again, subscripts on corresponding italic characters indicate individual elements, so $[\mathbf{A}]_{ij} = A_{ij}$. Also, as before, where the bold character is subscripted, this indicates a particular matrix in a family of matrices (so $\mathbf{K}_i$ is the i'th weighting function matrix.) In the case of covariance matrices, bold subscripts are used to indicate the covariance of a particular vector, thus $\mathbf{S}_\mathbf{x}$ is the covariance of $\mathbf{x}$ ($\mathbf{S}_i$ is a shorthand for $\mathbf{S}_{\mathbf{y}_i}$.)

Subscripts: In order to improve clarity, latin characters are typically used to subscript quantities in state space, with Greek characters subscripting measurement space quantities, thus $K_{\alpha i} = \partial y_\alpha / \partial x_i$.

Iterative processes: In iterative process, the value of a quantity for a particular iteration is indicated by a parenthetical superscript, thus $\mathbf{x}^{(r+1)} = \mathbf{x}^{(r)} + \ldots$.

Minor frame quantities: It can often be useful to distinguish 'atmospheric' quantities from 'instrumental' ones. In most cases the atmospheric quantities are represented by profiles on fixed pressure surfaces. Instrumental quantities are typically dependent on the minor frame (i.e. radiance integration period). Where such distinction is useful, minor frame quantities are indicated by an arrow placed over the symbol. For example the tangent point pressure for minor frame i is indicated by $\vec{\zeta}_i$.

Subvectors and submatrices: Much of this work deals with subsections of matrices and vectors. These are indicated with bracketed subscripts. Examples of these are $\mathbf{x}_{[j]}$, the j'th subvector of $\mathbf{x}$, and $\mathbf{K}_{[\alpha i]}$, the matrix $\partial \mathbf{y}_{[\alpha]} / \partial \mathbf{x}_{[i]}$

Scaled quantities: The $\widetilde{\ }$ symbol is used to indicate quantities that have been scaled for numerical stability. Thus $\widetilde{\mathbf{x}}$ represents the scaled state vector.

Linerization points: Where a linearization point has been chosen for the system the $\star$ superscript is used to indicate the use of linearized values. So, for example $\mathbf{y} \simeq \mathbf{y}^\star + \mathbf{K}^\star [\mathbf{x} - \mathbf{x}^\star]$.

Appendix E
EOS MLS signal designation nomenclature

E.1 Motivation

The EOS MLS instrument contains seven radiometers observing five different spectral regions. The signals from these radiometers are subdivided into multiple bands, each of which is observed by a different spectrometer. The instrument contains a switch network that allows most spectral bands to be observed by one of two different spectrometers (or in many cases, both simultaneously). This switch network is present to provide both flexibility for power saving modes of instrument operation, and some redundancy.

The complexity of this system is such that a complete nomenclature scheme is essential for a clear understanding of the instrument. Such a system has been devised with the intention that it is used in all the aspects of the instrument and software, from hardware drawings to science data processing software.

E.2 The nomenclature scheme

The scheme consists of up to five fields, in the form

<Radiometer>.<Band>.<Switch>.<Spectrometer>.<Channel>

Such a specification has many useful properties; in particular, fields can be ignored if they are not relevant to the specification. For example, in the Level 2 software, the user could specify that ozone is to be retrieved from `R2:190.B6F:O3`, without needing to specify a switch and/or spectrometer, as either of the two alternatives will be appropriate.

Earlier fields can also be ignored. For example, the instrument command and data handling system will typically only considers the channel and spectrometer fields, as the switch, band and radiometer information are of little relevance to instrument data handling activities.

The following subsections explain each of the fields in the specification.

E.2.1 Radiometers

The instrument consists of seven radiometers, measuring five different spectral regions. These are called `R1A`, `R1B`, `R2`, `R3`, `R4`, `R5H` and `R5V`. `R1A` and `R1B` are redundant 118 GHz radiometers. `R5H` and `R5V` are alternate polarizations of the 2.5 THz signal. As a courtesy to the reader, the radiometer field can also contain an indication of the frequency, separated from the number by a colon. Thus the full names for the radiometers are: `R1A:118`, `R1B:118`, `R2:190`, `R3:240`, `R4:640`, `R5H:2T5`, and `R5V:2T5`. The frequency information can be omitted, but if present it shall be correct.

E.2.2 Bands

The MLS spectral bands are numbered sequentially throughout the instrument; the numbering does not restart at one for each new radiometer. The band specification begins with the letter `B`, followed by the

number of the band. Following that, there is an optional character indicating whether the upper (`U`) or lower (`L`) sideband of the radiometer is being considered. If this is omitted, a 'folded' measurement is assumed. Following this optional character is a compulsory one indicating the type of spectrometer used for this band. Thus, `F` indicates a standard 25 channel filter bank, `M` a 'mid-band' 11 channel filter bank, `D` a digital autocorrelator spectrometer, and `W` a set of four individual wide filters.

Following this (separated by a colon) there can be additional courtesy information to the user describing the primary target of the band (e.g. `O3` for ozone, `PT` for temperature/pressure.)

E.2.3 Switch

Nearly all of the MLS spectral bands can be routed to one or two different spectrometers through a switch network. Most bands have a direct route to a spectrometer through no switch; this is designated by `S0`. Alternatively, most bands can be routed to an alternative spectrometer through one of five switches; such cases are designated `S1`$\cdots$`S5`.

E.2.4 Spectrometer

Following the switch field, the spectrometer type and number is indicated. The instrument contains four types of spectrometers. There are nineteen 'standard' 25 channel filter banks, designated `FB25-1` $\cdots$ `FB25-19`; five 'mid-band' 11 channel filter banks (`MB11-1` $\cdots$ `MB11-5`); four digital autocorrelation spectrometers (`DACS-1` $\cdots$ `DACS-4`); and three sets of 4 individual wide filters (`WF4-1` $\cdots$ `WF4-3`.)

E.2.5 Channels

Channels are simply specified by a `C`, followed by a number. Channels are numbered from one, except in the DACS where the numbering starts from zero.

E.2.6 General comments

This nomenclature system deliberately contains much redundancy designed to improve clarity; however, it may be trimmed down if the user wishes.

- Radiometers may be specified without their frequency information.
- Bands may be specified without their intended target, or upper/lower sideband information.
- The switch and spectrometer information are redundant; thus one may be omitted if desired.
- If a band is specified the radiometer specification is redundant and may be omitted. This is discouraged, however, as clarity is lost.

E.3 The valid MLS signals

Tables E.1 and E.2 list all the various radiance signals that can be measured by EOS MLS.

Table E.1: The nominal MLS measurement set.

Radiometer	Band	Switch	Spectrometer	Channels
R1A:118	B1F:PT	S0	FB25-1	C1···C25
R2:190	B2F:H2O	S0	FB25-2	C1···C25
R2:190	B3F:N2O	S2	FB25-3	C1···C25
R2:190	B4F:HNO3	S0	FB25-4	C1···C25
R2:190	B5F:CLO	S0	FB25-5	C1···C25
R2:190	B6F:O3	S0	FB25-6	C1···C25
R3:240	B7F:O3	S0	FB25-7	C1···C25
R3:240	B8F:PT	S3	FB25-8	C1···C25
R3:240	B9F:CO	S0	FB25-9	C1···C25
R4:640	B10F:CLO	S0	FB25-10	C1···C25
R4:640	B11F:BRO	S0	FB25-11	C1···C25
R4:640	B12F:N2O	S4	FB25-12	C1···C25
R4:640	B13F:HCL	S0	FB25-13	C1···C25
R4:640	B14F:O3	S0	FB25-14	C1···C25
R5H:2T5	B15F:OH	S5	FB25-15	C1···C25
R5H:2T5	B16F:OH	S0	FB25-16	C1···C25
R5H:2T5	B17F:PT	S0	FB25-17	C1···C25
R5V:2T5	B18F:OH	S0	FB25-18	C1···C25
R5V:2T5	B19F:OH	S0	FB25-19	C1···C25
R1A:118	B22D:PT	S1	DACS-1	C0···C128
R2:190	B23D:H2O	S0	DACS-2	C0···C128
R3:240	B24D:O3	S0	DACS-3	C0···C128
R3:240	B25D:O3	S0	DACS-4	C0···C128
R2:190	B27M:HCN	S0	MB11-1	C1···C11
R4:640	B28M:HO2	S0	MB11-2	C1···C11
R4:640	B29M:HOCL	S0	MB11-3	C1···C11
R4:640	B30M:HO2	S0	MB11-4	C1···C11
R4:640	B31M:BRO	S0	MB11-5	C1···C11
R1A:118	B32W:PT	S0	WF4-1	C1···C4
R3:240	B33W:O3	S0	WF4-2	C1···C4

Table E.2: The alternate MLS measurement set.

Radiometer	Band	Switch	Spectrometer	Channels
R1A:118	B1F:PT	S3	FB25-8	C1···C25
R1B:118	B21F:PT	S4	FB25-12	C1···C25
R1B:118	B21F:PT	S3	FB25-8	C1···C25
R2:190	B2F:H2O	S2	FB25-3	C1···C25
R2:190	B4F:HNO3	S2	FB25-3	C1···C25
R2:190	B5F:CLO	S2	FB25-3	C1···C25
R3:240	B7F:O3	S3	FB25-8	C1···C25
R3:240	B8F:PT	S2	FB25-8	C1···C25
R3:240	B9F:CO	S3	FB25-8	C1···C25
R4:640	B13F:HCL	S4	FB25-12	C1···C25
R4:640	B14F:O3	S4	FB25-12	C1···C25
R5H:2T5	B16F:OH	S5	FB25-15	C1···C25
R5H:2T5	B17F:PT	S5	FB25-15	C1···C25
R5V:2T5	B18F:OH	S5	FB25-15	C1···C25
R5V:2T5	B19F:OH	S5	FB25-15	C1···C25
R5V:2T5	B20F:PT	S3	FB25-8	C1···C25
R5V:2T5	B20F:PT	S5	FB25-15	C1···C25
R1B:118	B26D:PT	S1	DACS-1	C0···C128
R1B:118	B34W:PT	S0	WF4-3	C1···C4

Appendix F
Some notes on the Gauss Newton minimizer

This appendix constitutes some notes on the implementation of the Gauss-Newton minimizer used in the Level 2 software, with a particular focus on strategies for choosing the Marquardt-Levenberg parameters. The notes were written by Fred Krogh, the author of the Gauss-Newton code that is invoked in the Level 2 software.

F.1 Introduction

We wish to minimize $\varphi = \frac{1}{2}||\mathbf{f}(\mathbf{x})||^2$. Some components of $\mathbf{f}$ are of the form $m(\mathbf{x}, \mathbf{p}) - b(\mathbf{p})$, where m is given function of the unknown vector $\mathbf{x}$, and we have observations $b(\mathbf{p})$ for given values of the known vector $\mathbf{p}$. There are similar *a priori* components that are not based on actual observed values, but rather on average values from past observations. Still other components come from Tikonov regularization. These components of $\mathbf{f}$ have the form $D\mathbf{x} = 0$, where D is a difference operator over a subset of the x_i. In this note we are neglecting the scaling that applies to these various parts.

The gradient of φ is given by $\mathbf{g} = J^T\mathbf{f}$ where $J = \partial\mathbf{f}/\partial\mathbf{x}$. A Newton method finds a local minimum of φ by iteratively solving the system $(\partial\mathbf{g}/\partial\mathbf{x})\,\boldsymbol{\delta}\mathbf{x} = -\mathbf{g}$, that if convergent will result in satisfying our goal of having $\mathbf{g} = \mathbf{0}$.

This Newton method requires forming $\partial\mathbf{g}/\partial x_i = J^T\partial\mathbf{f}/\partial x_i + (\partial J^T/\partial x_i)\mathbf{f}$, $i = 1, \ldots,$ n. The second term in this expression is significantly more work to obtain than the first, and when dropped in the computations, the method is referred to as a Gauss-Newton method. This method works very nearly as well as the full Newton method if φ is sufficiently small at the solution, or $\mathbf{f}$ has small second derivatives.

The Gauss-Newton method is thus given by the iteration $J_k^T J_k\boldsymbol{\delta}\mathbf{x}_k = -J_k^T\mathbf{f}_k$, $\mathbf{x}_{k+1} = \mathbf{x_k} + \boldsymbol{\delta}\mathbf{x}_k$. With either this or the full Newton method, the iterates are likely to diverge due to nonlinearities in $\mathbf{g}$. The Levenberg-Marquardt method reduces the size of $\boldsymbol{\delta}\mathbf{x}_k$ by minimizing on each step $\frac{1}{2}||\mathbf{f}_k + J_k\boldsymbol{\delta}\mathbf{x}_k||^2 + \frac{1}{2}\lambda||\boldsymbol{\delta}\mathbf{x}_k||^2$ for a choice of λ that gives a change small enough so that $\mathbf{f}$ is adequately approximated by a linear function.

Almost all books and papers that discuss this method recommend choosing λ large enough to get a decrease in $||\mathbf{f}||$. Thus if a given λ results in $||\mathbf{f}_{k+1}|| \geq ||\mathbf{f}_k||$, the solution is redone with a larger value for λ or something else is done to use a smaller $\boldsymbol{\delta}\mathbf{x}$. In Hanson and Krogh [1992] a number of principles are given for deciding how big a move is desirable when solving a nonlinear least squares problem. The most important principle is to allow for $||\mathbf{f}||$ to increase. Of course if this is done, some means must be available to keep things from getting out of hand.

The basic idea is to allow for arbitrarily large increases in $||\mathbf{f}||$ as long as there is a solution to the current linearization of the problem within a neighborhood of the current point that gives a better result than the best seen so far. The size of this neighborhood is then decreased as long as no improvement is observed. Thus ultimately either we get a new best result or the size of the neighborhood shrinks to the point where the linearized problem no longer gives an improvement over the best seen before. In practice when the latter case occurs, it tends to occur early. The method for selecting a value for λ to get a move of a given size is given in Moré and Sorensen [1983].

F.2 Determining When to Return to a Best x

Upon getting a new φ, and the value to be used for λ when the last φ was not the best seen to date we need to decide whether to give up on the current branch. Using variable names from the code DNWT,

FN	$\sqrt{2\varphi} = ||\mathbf{f}||$.
FNB	The smallest value for FN seen so far.
FNMIN	The residual after solving the linear problem, including the contribution of the Levenberg Marquardt stabilization to the residual, see ρ below.
FNXE	$\rho^2 - \lambda||\boldsymbol{\delta}\mathbf{x}||^2$, *i.e.* the value for $||\mathbf{f}||^2$ that would be expected on the next iteration if $\mathbf{f}$ were a linear function.

The variable FNXE is needed to decide if we should give up and is also used to decide if behavior is sufficiently linear to justify using a larger neighborhood. Consider the least squares problem

$$\begin{bmatrix} J \\ \sqrt{\lambda} I \end{bmatrix} \boldsymbol{\delta}\mathbf{x} \cong \begin{bmatrix} -\mathbf{f} \\ 0 \end{bmatrix} \quad \equiv \quad \begin{bmatrix} U \\ 0 \end{bmatrix} \boldsymbol{\delta}\mathbf{x} \cong \begin{bmatrix} -\widehat{\mathbf{f}} \\ \rho \end{bmatrix} \tag{F.1}$$

where an orthogonal matrix Q is used to triangularize the augmented matrix $\begin{bmatrix} J & -\mathbf{f} \\ \sqrt{\lambda} I & 0 \end{bmatrix}$.

The normal equations can be looked at similarly.

$$\left[J^T J + \lambda I\right] \boldsymbol{\delta}\mathbf{x} = -J^T \mathbf{f} \quad \equiv \quad U^T U \boldsymbol{\delta}\mathbf{x} = -J^T \mathbf{f} = -U^T \widehat{\mathbf{f}}, \tag{F.2}$$

where in this case U is obtained using a Cholesky factorization. If row signs are changed in Eq. (F.1) to make the diagonal entries of U positive the matrices are identical. Since the Cholesky factorization is both cheaper and significantly easier to code in this application it is the approach used.

Clearly, $\rho^2 = ||\mathbf{f}||^2 - ||\widehat{\mathbf{f}}||^2$, and $\widehat{\mathbf{f}}$ is given by $U^{-T}(-J^T\mathbf{f})$, a quantity that is needed when solving for $\boldsymbol{\delta}\mathbf{x}$, and ρ^2 is the value expected for $||\mathbf{f}||^2 + \lambda||\boldsymbol{\delta}\mathbf{x}||^2$, since $\boldsymbol{\delta}\mathbf{x}$ is set to make $\widehat{\mathbf{f}} = \mathbf{0}$. Thus we have the data necessary to compute FNXE = $\rho^2 - \lambda||\boldsymbol{\delta}\mathbf{x}||^2$. If FNXE > FNB^2 then the current branch is discarded and the code makes a gradient move from the point that generated the value of FNB.

There is another rule that can cause the code to give up as well. If we have gotten an improvement in FN on the last iteration, but not a new best, and we have used as small a λ as a the lower limit allowed, and the angle between the current and previous corrections is not too big, then it is assumed that current convergence is too slow and the code goes back to the previous best solution to continue the iterations.

F.3 Selecting λ

The code has so many decision points that it is very difficult to write down in any intelligible way how λ is selected. Generally speaking, it is made smaller when we get a new best, and bigger when things are not going well. The code gets more optimistic about taking big moves when the current correction to $\mathbf{x}$ is close being a multiple of the previous one, and more conservative when the current correction is in a very different direction.

The code makes use of a "relative" λ, that is one that is multiplied by a norm of J, as well as λ. It also maintains a minimal value for λ that is set to 0 on a new best, and is increased when things do not appear to be going well.

The FNXE above is only part of the story. This variable is also set to $\frac{1}{2}||\mathbf{f}||^2$. FNXE can also be set one other way after gradient moves.

The code assumes good linear behavior if $||\mathbf{f}||^2 < 1.125 * \text{FNXE}$, where in most cases this FNXE is $\frac{1}{2}||\mathbf{f}||^2$ from the previous evaluation.

Bibliography

F. T. Barath, M. C. Chavez, R. E. Cofield, D. A. Flower, M. A. Frerking, M. B. Gram, W. M. Harris, J. R. Holden, R. F. Jarnot, W. G. Kloezeman, G. J. Klose, G. K. Lau, M. S. Loo, B. J. Maddison, R. J. Mattauch, R. P. McKinney, G. E. Peckham, H. M. Pickett, G. Siebes, F. S. Soltis, R. A. Suttie, J. A. Tarsala, J. W. Waters, and W. J. Wilson. The Upper Atmosphere Research Satellite Microwave Limb Sounder Experiment. *Journal of Geophysical Research*, 98(D6):10,751–10,762, 1993.

A. Dudhia and N. J. Livesey. Validation of the Improved Stratospheric and Mesospheric Sounder temperature measurements. *Journal of Geophysical Research*, 101(D6):9795–9809, 1996.

M. Filipiak. Precision estimates for the geophysical parameters measured by EOS MLS. Technical report, University of Edinburgh, Department of Meteorology, 2004.

E. F. Fishbein, R. E. Cofield, L. Froidevaux, R. F. Jarnot, T. Lungu, W. G. Read, Z. Shippony, J. W. Waters, I. S. McDermid, T. J. McGee, U. Singh, M. Gross, A. Hauchecorne, P. Keckhut, M. E. Gelman, and R. M. Nagatani. Validation of UARS Microwave Limb Sounder temperature and pressure measurements. *Journal of Geophysical Research*, 101(D6):9983–10,016, 1996.

R. G. Fleagle and J. A. Businger. *An Introduction to Atmospheric Physics*. Academic Press, 1963.

L. Froidevaux, W. G. Read, T. A. Lungu, R. E. Cofield, E. F. Fishbein, D. A. Flower, R. F. Jarnot, B. P. Ridenoure, Z. Shippony, J. W. Waters, J. J. Margitan, I. S. McDermid, R. A. Stachnik, G. E. Peckham, G. Braathen, T. Deshler, J. Fishman D. J. Hoffmann, and S. J. Oltmans. Validation of UARS Microwave Limb Sounder ozone measurements. *Journal of Geophysical Research*, 101(D6): 10,017–10,060, 1996.

G. H. Golub and C. F. VanLoan. *Matrix Computations, Third edition*. Johns Hopkins University Press, 1996.

R. J. Hanson and F. T. Krogh. A quadratic-tensor model algorithm for nonlinear least-squares problems with linear constraints. *ACM Trans. Math. Sw.*, 18(2):115–133, 1992.

S. L. Hess, editor. *An Introduction to Theoretical Meteorology*. Krieger Publishing Company, 1959. (Holt and Rinehart and Winston).

R. F. Jarnot. EOS MLS level 1 data processing algorithm theoretical basis. Technical report, Jet Propulsion Laboratory, 2004. D-15210.

Y. Jiang. Level 3 mapping algorithm theoretical basis. Technical report, Jet Propulsion Laboratory, 2004. JPL D-18911.

N. J. Livesey. EOS MLS Level 2 Software Users' Guide. Technical report, Jet Propulsion Laboratory, 1999a. JPL D-18029.

N. J. Livesey. Functional Requirements for the EOS MLS Level 2 Data Processing Software. EOS MLS internal document, Jet Propulsion Laboratory, 1999b. D-18027.

N. J. Livesey. Data quality document for the EOS MLS version 1.5 level 2 dataset. Technical report, Jet Propulsion Laboratory, 2004.

N. J. Livesey, W. G. Read, L. Froidevaux, J.W. Waters, H.C. Pumphrey, D.L. Wu, M.L. Santee, Z. Shippony, and R.F. Jarnot. The UARS Microwave Limb Sounder version 5 dataset: Theory, characterization and validation. *Journal of Geophysical Research*, 108(D13):4378, 2003. doi:10.1029/2002JD002273.

J. J. Moré and D. C. Sorensen. Computing a trust region step. *SIAM Journal on Scientific Computing*, 4:553–572, 1983.

W. G. Read, Z. Shippony, and W. V. Snyder. Microwave Limb Sounder forward model algorithm theoretical basis document. Technical report, Jet Propulsion Laboratory, 2004. JPL D-18130.

C. A. Reber. The Upper Atmosphere Research Satellite (UARS). *Geophysical Research Letters*, 20 (12):1215–1218, 1993.

C. A. Reber, C. E. Trevathan, R. J. McNeal, and M. R. Luther. The Upper Atmosphere Research Satellite (UARS) mission. *Journal of Geophysical Research*, 98(D6):10,643–10,647, 1993.

C. D. Rodgers. Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. *Reviews of Geophysics and Space Physics*, 14(4):609–624, 1976.

C. D. Rodgers. Characterisation and error analysis of profiles retrieved from remote sounding measurements. *Journal of Geophysical Research*, 95(D5):5587–5595, 1990.

C. D. Rodgers. Information content and optimisation of high spectral resolution measurements. In P. B. Hays and J. Wang, editors, *Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research II*, pages 136–147. SPIE, 1996.

Clive D. Rodgers. *Inverse methods for atmospheric science, theory and practice*. World Scientific, 2000.

M. J. Schwartz. MLS mesosphere-specific forward model algorithm theoretical basis document. Technical report, Jet Propulsion Laboratory, 2004. JPL D-28534.

F. W. Taylor, C. Rodgers, J. Whitney, S. Werrett, J. Barnett, G. Peskett, P. Venters, J. Ballard, C. Palmer, R. Knight, P. Morris, and T. Nightingale. Remote sensing of atmospheric structure and composition by pressure modulation radiometry from space: The ISAMS experiment on UARS. *Journal of Geophysical Research*, 98(D6):10,799–10,814, 1993.

A. N. Tikhonov. On the solution of incorrectly stated problems and a method of regularization. *Dokl. Acad. Nauk SSSR*, 151:501, 1963.

J. W. Waters. An overview of the EOS MLS experiment. Technical report, Jet Propulsion Laboratory, 2004. D-15745.

J. W. Waters, W. G. Read, L. Froidevaux, R. F. Jarnot, R. E. Cofield, D. A. Flower, G. K. Lau, H. M. Pickett, M. L. Santee, D. L. Wu, M. A. Boyles, J. R. Burke, R. R. Lay, M. S. Loo, N. J. Livesey, T. A. Lungu, G. L. Manney, L. L. Nakamura, V. S. Perun, B. P. Ridenoure, Z. Shippony, P. H. Siegel, and R. P. Thurstans. The UARS and EOS Microwave Limb Sounder (MLS) experiments. *Journal of the Atmospheric Sciences*, 56:194–217, 1999.

J. W. Waters, W. G. Read, L. Froidevaux, T. A. Lungu, V. S. Perun, R. A. Stachnick, R. F. Jarnot, R. E. Cofield, E. F. Fishbein, D. A. Flower, J. R. Burke, J. C. Hardy, L. L. Nakamura, B. P. Ridenoure, Z. Shippony, and R. P. Thurstans. Validation of UARS Microwave Limb Sounder ClO measurements. *Journal of Geophysical Research*, 101(D6):10,091–10,127, 1996.

D. L. Wu and J. H. Jiang. EOS MLS algorithm theoretical basis for cloud measurements. Technical report, Jet Propulsion Laboratory, 2004. JPL D-19299.