FIRST TRACE GAS MEASUREMENTS USING FOURIER TRANSFORM INFRARED SOLAR ABSORPTION SPECTROSCOPY AT THE UNIVERSITY OF TORONTO ATMOSPHERIC OBSERVATORY

by

Aldona Wiacek

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Department of Physics
University of Toronto

To my Parents,

who sacrificed all things familiar

in order to give me the opportunity

to explore the unknown.

Abstract

First Trace Gas Measurements Using Fourier Transform Infrared Solar Absorption Spectroscopy at the University of Toronto Atmospheric Observatory

Aldona Wiacek
Doctor of Philosophy
Department of Physics
University of Toronto
2006

A new high-resolution (0.004 cm^{-1}) research-grade Fourier transform spectrometer was installed at the University of Toronto Atmospheric Observatory (TAO), and commissioned in October 2001. TAO fills a coverage gap in North America in the Network for the Detection of Atmospheric Composition Change (NDACC), and was accepted as a Complementary Station in March 2004 as a direct result of work presented in this thesis.

Trace gas concentrations are derived using an existing Optimal Estimation Method (OEM) retrieval algorithm implemented at TAO as part of this work. A set of prior constraints necessary for retrievals of vertical profiles of trace gases using the OEM approach was optimized, and includes *a priori* volume mixing ratio profiles suited to TAO's location, and the prior and measurement noise covariances. A three-year time series of O_3, NO, NO_2, HCl, HF, N_2O and CH_4 (exhibiting the known seasonal cycles) was submitted to the NDACC database for long-term monitoring, trend detection, and satellite validation purposes. A quantitative retrieval characterization was performed whereby the averaging kernels, weighting functions, measurement, smoothing and temperature error covariances of the retrievals of these seven gases were calculated. The degrees of freedom for signal ranged from 1.10 for NO_2 to 4.12 for N_2O.

A systematic study showed that the effect of several different layering schemes (implicit retrieval constraints) on the retrieved vertical profile is less than 5% for N_2O and

HF, and less than 2% for NO, NO_2, CH_4 and HCl; however, retrieved profiles of H_2O showed differences up to 13% when the layering scheme was varied. A maximum tropospheric layer thickness of 0.8 km is inferred for accurate H_2O retrievals.

Finally, the usefulness of Fourier Transform InfraRed (FTIR) observations for studying different regions of the atmosphere was demonstrated by 1) documenting a limited sensitivity of ground-based measurements to mesospheric-lower thermospheric concentrations of NO, 2) correlating anomalously high stratospheric HCl, HF and O_3 concentrations derived at TAO with the passage of a polar air filament above Toronto, and 3) showing very good agreement between monthly mean abundances of tropospheric CO between January 2002 – September 2003 and spatially coincident MOPITT satellite measurements.

Acknowledgements

Swept up in a very irrational impulse, I will acknowledge the solid support I have received throughout this Ph.D. project from the graduate generational timeship that is the MP605 couch. It has always been there for me, although on some days it was invaded by one of my officemates, whom I will also thank for putting up with me.

Getting down to serious business, I will acknowledge the support of the funding agencies, foundations, university and industry, which have contributed to my support, and the establishment of the Toronto Atmospheric Observatory: NSERC, ABB Bomem Inc., PREA, CFCAS, CFI, ORDCF, CRESTech, OGS, CSA, The Zonta Foundation, The Joseph Bazylewicz Fund, and the University of Toronto.

Next, I will thank all technicians, summer students and post-doctoral fellows who contributed to measurements, lab support and data analysis at TAO since 2000: C. Avis, P. Chen, B. Chugh, D. Edwards, N. Faridi, G. Hassanpour, C. Heald, M. Jensen, T. Kerzenmacher, K. MacQuarrie, O. Mikhailov, R. Saari, P. Rourke, R. Sullivan, D. Yashcov, and Y. Zhao.

I certainly will not forget the good FTIR (or not FTIR) people who patiently answered my emails about auxiliary programs to run SFIT-2, or provided me with data, or served as my sounding board: Arndt Meier, Nicholas Jones, Frank Hase, James Hannigan, Stephen Wood, Manu Mahieu, Richard Mittermeier, Hans Fast, Chris McLinden, Keith MacQuarrie, Jane Liu, and Holger Bremer.

I would also like to thank the Strong Group, my UofT FTIR Support Group, consisting of Jeff Taylor and Debra Wunch, and, of course my supervisor, Prof. Kimberly Strong, who has given me a tremendous amount of guidance and support, in her usual patient and kind manner.

Finally, I thank my Family, and my Husband Jake Hanley, without whom exploring the unknown doesn't mean much.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

The most general goal in the field of Atmospheric Science is to improve our understanding of, and hence our predictive capabilities regarding, the dynamical and chemical state of the atmosphere. To achieve this goal, atmospheric scientists strive to make computer models of the atmosphere agree well with increasingly precise and accurate observations by including more sophisticated representations of the underlying chemical, microphysical and dynamical processes. Only mature, i.e. validated, models of climate and atmospheric composition can be used with confidence to inform public policy regarding the control and mitigation of human impacts on the environment.

While the general large-scale features of dynamics and chemistry in the atmosphere – and many small-scale microphysical features – are captured by current models, outstanding questions remain. Some are discussed further in Section 1.2. In order to identify missing or poorly represented processes in atmospheric models, there exists a global Earth observation effort using satellite-, aircraft-, balloon- and ground-based platforms. The role of ground-based observatories is to provide long-term high-quality correlative measurements of trace gases, aerosols, and dynamical variables, which can be used for unique

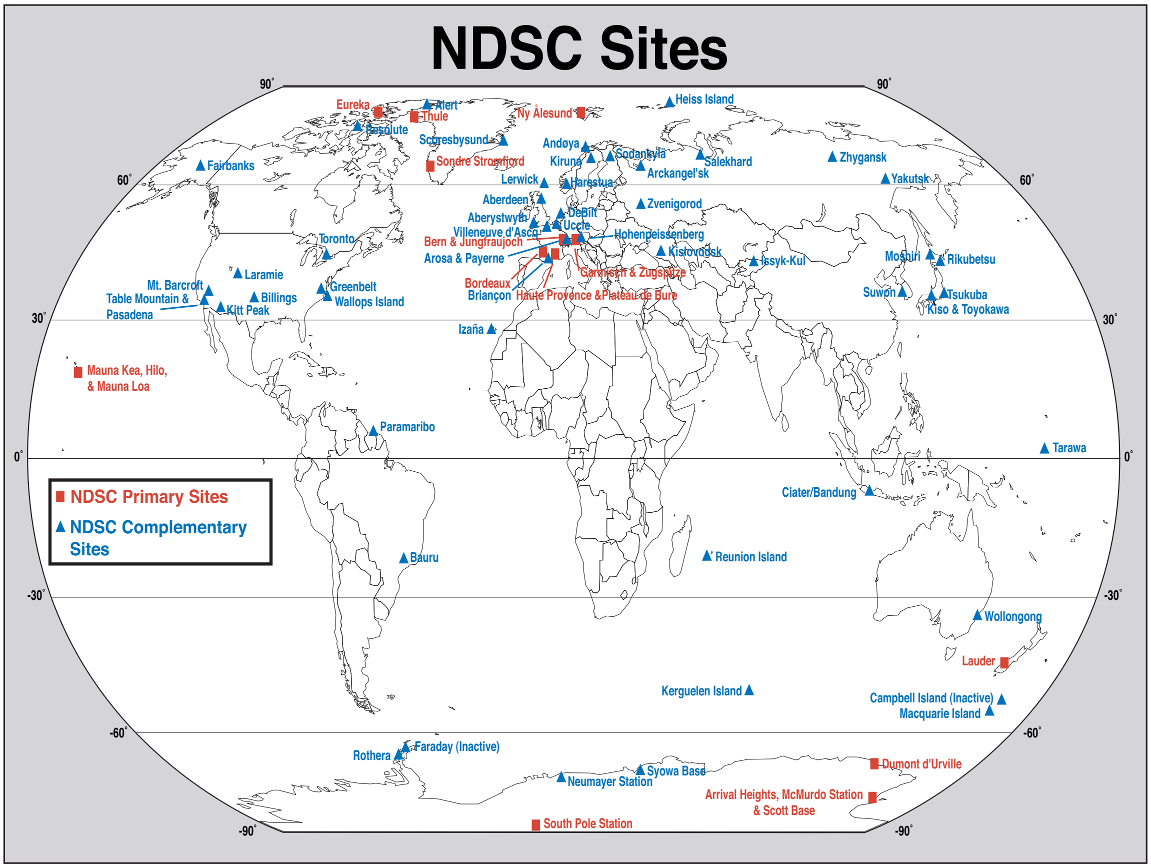


Figure 1.1 Global distribution of NDSC (now NDACC) Stations showing the newly established Toronto Atmospheric Observatory, which fills a coverage gap in mid-latitude North America (http://www.ndsc.ncep.noaa.gov/instr_sites/NDSCmap_20051031_US.pdf).

scientific studies (both long-term and short-term), as well as for satellite and model validation. In this context, the University of Toronto Atmospheric Observatory (TAO) was established in 2001, and has since become a Complementary Station in the global Network for the Detection of Atmospheric Composition and Change (NDAAC), formerly known as the Network for the Detection of Stratospheric Change (NDSC) described by *Kurylo and Zander* [2000] and shown in Figure 1.1. TAO fills a North American mid-latitude coverage gap within NDACC, which has provided a focal point for the high-resolution ground-based Fourier Transform InfraRed (FTIR) community, leading to continuous improvements in the measurement technique (e.g. instrument performance characterisation [*Hase et al.*, 1999; *Bernardo and Griffith*, 2005], new retrieval microwindows [e.g *Meier et al.*, 2004b], and an improved characterization of the interfering solar spectrum [*Hase et al., in prep*]), as well as to an increased return from the measurements in the form of retrievals of vertical profile information and a full characterization of retrieval errors [e.g. *Pougatchev et al.*, 1995; *Barret et al.*, 2002; *Schneider et al.*, 2005a]. Most importantly, ground-based FTIR spectra recorded at NDACC stations have been used to study chemical and dynamical processes in the polar [e.g. *Blumenstock et al.*, 2006; *Wood et al.*, 2004], mid-latitude [e.g. *Velazco et al.*, 2005], and tropical [*Schneider et al.*, 2005a, e.g.] atmosphere in the troposphere, stratosphere and mesosphere [*Wiacek et al.*, 2006a; *Kasai et al.*, 2005, *Velazco et al., submitted to GRL*]. Furthermore, numerous papers have been written on, e.g., trace gas trends [e.g. *Zander et al.*, 1994; *Rinsland et al.*, 2002a,b, 2003], biomass burning emissions [e.g. *Yurganov et al.*, 2005; *Paton-Walsh et al.*, 2004; *Zhao et al.*, 2002; *Rinsland et al.*, 2001], volcanic activity [e.g. *Notholt et al.*, 2003], and satellite validation [e.g. *Dils et al.*, 2006; *Sussmann and Buchwitz*, 2005; *Mahieu et al.*, 2005; *Barret et al.*, 2003].

The goals of this Ph.D. project were the commissioning of the primary instrument at TAO (a Fourier transform infrared spectrometer), its acceptance into NDACC, the implementation of retrieval algorithms, and the derivation of total and partial columns of a suite of trace gases from the infrared solar absorption spectra recorded by the spectrometer, after optimizing the retrieval procedure for TAO's location. In addition, a part of this dataset was compared to other observations, as appropriate or available, and several case studies of particular events were undertaken.

1.2 The Composition and Structure of the Atmosphere

The abundance of a gas in the Earth's atmosphere can be measured in the fractional unit of *Volume Mixing Ratio* (VMR), which refers to the fractional volume of a given gas per unit volume of air, e.g. the VMR of molecular Nitrogen (N_2) and Oxygen (O_2) in the dry atmosphere is 0.78 and 0.21, respectively, meaning that together, these two gases make up 0.99, or 99% of air. Ironically, N_2 and O_2 only absorb infrared radiation under special circumstances and are not easy to measure by the FTIR technique described in this work. Furthermore, it is the *trace* gases, which make up the remaining 1% of the atmosphere that are responsible for such high-profile environmental problems as the thinning of the life-sustaining ozone (O_3) layer, global warming, and urban smog. For example, the peak VMR of O_3 above Toronto (at an altitude of $\sim$35 km) is on average only 7x10^{-6}, or 7 *parts per million by volume* (ppmv), and the average VMR of water vapour (H_2O) near the surface in Toronto is only $\sim$0.01.

Throughout this work we will refer to various regions of the atmosphere, which are illustrated in Figure 1.2. The *troposphere* is the relatively well-mixed region of decreasing temperature between approximately 0-15 km, which contains all of the atmospheric features referred to as "weather." It is separated by the *tropopause* from the region of increasing temperature between approximately 15-50 km known as the *stratosphere*, on account of its being stably stratified. The stratosphere is separated from the *mesosphere* ($\approx$ 50-90 km) by the *stratopause*, and the mesosphere is separated from the *thermosphere* ($\approx$ 90-130 km) by the *mesopause*. The decreasing temperature of the troposphere with height is caused by the radiative cooling of a semi-opaque (*grey*) atmosphere heated from below, just as the cooling of the mesosphere occurs via the radiation emitted by molecules to space. The heating of the stratosphere is caused by the absorbtion of ultraviolet (UV) light by the O_3 layer centered at an altitude of approximately 22 km, while the heating

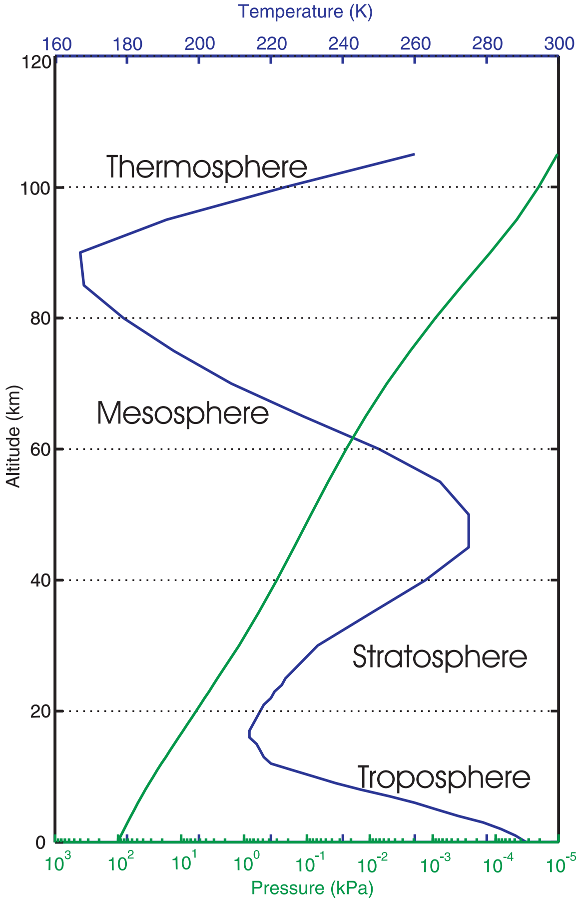


Figure 1.2 The 1976 US Standard Atmosphere (U.S. Government Printing Office, Washington, D.C., 1976.)

of the thermosphere is caused by the photodissociation and ionization of the ubiquitous molecular oxygen and nitrogen by even shorter wavelengths of UV light.

The troposphere is where greenhouse gases such as water (H_2O), carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), sulfur hexafluoride (SF_6), and chlorofluorocarbons (CFCs) enter the atmosphere and heat the Earth by trapping additional IR radiation that would otherwise escape to space. Molecules that have at least one H atom are removed by reaction with the hydroxyl radical (OH) in the troposphere, while those such as N_2O, SF_6 and CFCs are only photodissociated by UV light once they reach the stratosphere. CO_2, on the other hand, is stable and is only destroyed by hard UV ionization in the upper atmosphere, while tropospheric O_3, which is the next most important greenhouse gas after H_2O, CO_2 and CH_4, is not emitted at all, but rather it is produced entirely via photochemistry. Finally, carbon monoxide (CO) is sometimes termed an "indirect greenhouse gas" because of its influence on the OH radical, which in turn controls the concentration of other greenhouse gases, such as CH_4. While all of the above have some natural sources, the anthropogenic contribution to their budgets is large and the atmospheric abundances of these gases are continuing to increase, with the exception of CFCs controlled by the Montreal Protocol on Ozone Depleting Substances (1987) and its subsequent amendments [*IPCC*, 2001]. TAO has been established in part to provide long-term monitoring and trend-detection of tropospheric greenhouse gas abundances, which are key parameters in future climate scenarios predicted by 3-D General Circulation Models (GCMs) with coupled chemistry and dynamics (also known as Chemistry-Climate Models, or CCMs).

The chemistry of greenhouse gases is coupled to the chemistry of stratospheric O_3 loss via CFCs, N_2O, CH_4 and H_2O, which supply chlorine, nitrogen and hydrogen molecules to participate in the catalytic destruction of O_3 that accounts for 75% of the observed stratospheric O_3 loss [*Jacob*, 1999]. (It is important to distinguish between tropospheric O_3, which is a health hazard to animal and plant life, and stratospheric O_3, which protects this same life by shielding it from UV radiation.) The basic gas phase catalytic cycle is the same regardless of what radical is involved (e.g. ClO, NO, HO), hence in Reactions 1.1 and 1.2 it is symbolized by 'X'. Reaction 1.3 is the net result of the catalytic cycle. Since the initial radical is preserved, it is free to initiate another catalytic cycle, destroying two $O_x = O + O_3$ molecules each time.

$$X + O_3 \rightarrow XO + O_2 \tag{1.1}$$

$$XO + O \rightarrow X + O_2 \tag{1.2}$$

$$Net : O + O_3 \rightarrow O_2 + O_2. \tag{1.3}$$

Postulated first in 1930, the Chapman mechanism accounts for the remaining 25% of observed ozone loss [*Jacob*, 1999]. Reaction 1.4 is relatively slow and it initiates the cycle by producing 2O_x atoms. Reactions 1.5 and 1.6 are fast reactions which keep cycling O_x between O and O_3. Reaction 1.7 is another relatively slow reaction which consumes 2 O_x molecules. Thus the Chapman mechanism consists of 2 slow reactions which produce and destroy O_x and 2 fast reactions which cycle O_x.

$$O_2 + h\nu \rightarrow O + O \tag{1.4}$$

$$O + O_2 + M \rightarrow O_3 + M \tag{1.5}$$

$$O_3 + h\nu \rightarrow O + O_2 \tag{1.6}$$

$$O + O_3 \rightarrow 2O_2 \tag{1.7}$$

Chronologically, the Chapman mechanism was elucidated first, followed by the catalytic cycle involving HO_x=HO+HO_2 in the 1950's, and the catalytic cycle involving NO_x=NO+NO_2 in the 1960's, when supersonic flight was being considered for the stratosphere [*Jacob*, 1999]. (For a an extensive review of the concepts and history of stratospheric ozone depletion see *Solomon* [1999] and references to key historical papers included therein.) The significance of the ClO_x=ClO+ClO_2 catalytic cycle was recognized in the 1970's, eventually leading to a ban on CFC production initiated by the Montreal Protocol in 1987. The final piece of the puzzle regarding stratospheric O_3 loss came in the 1980's, when a new catalytic cycle involving ClO (with itself) was postulated to explain near total O_3 loss observed in the lower stratosphere over the Antarctic. It was discovered that the high ClO concentrations over the Antarctic polar region were caused by the near-titrating reaction involving the short-term and long-term chlorine reservoirs chlorine nitrate ($ClONO_2$) and hydrogen chloride (HCl), respectively:

$$ClONO_2 + HCl \quad \overrightarrow{PSC} \quad Cl_2 + HNO_3. \tag{1.8}$$

Cl_2 is photolyzed into 2 Cl atoms immediately upon formation (after the return of sunlight over the polar regions) and reacts with O_3 to form ClO and initiate the catalytic destruction cycle. An essential feature of Reaction 1.8 is that it proceeds on *Polar Stratospheric*

Cloud (PSC) surfaces, which are thought to be primarily solid nitric acid trihydrates (NAT; $HNO_3 \cdot 3H_2O$), followed by supercooled ternary solutions containing both nitric and sulfuric acids and water (STS; $H_2SO4/HNO_3/H_2O$). Referred to as Type Ia and Type Ib PSCs, respectively, these are found in cold regions of the polar stratosphere at temperatures $\lesssim$196 K. When temperatures fall below ~189 K, the formation of Type II PSCs, composed of water ice crystals, is also possible [*Peter*, 1997]. Recent research also indicates that Reaction 1.8 can proceed on aqueous sulfuric acid aerosols, which are ubiquitous in the whole stratosphere, not just in the polar regions (luckily this H_2SO_4 aerosol O_3 loss mechanism requires temperatures below 200 K to proceed [*Jacob*, 1999]). The NO_x and ClO_x families are coupled via the following reactions

$$HNO_3 + h\nu \quad \rightarrow \quad NO_2 + OH, \tag{1.9}$$

$$ClO + NO_2 + M \quad \rightarrow \quad ClONO_2 + M, \tag{1.10}$$

where M is a third gas that conserves momentum in the collision of the other two gases, making the reaction much more likely. This reaction chain, which shuts down polar O_3 loss by scavenging ClO, is itself shut down by the sedimentation of NAT PSCs.

At the present time, a topic of vigorous research is the natural variability of the dynamical factors that control the confinement and temperature of the Arctic stratosphere, and hence the extent of ozone depletion in this region [*IPCC*, 2001; *WMO*, 2003]. Increased confinement and lower Arctic temperatures also reduce the abundance of O_3 at northern mid-latitudes (after the breakup of the Arctic vortex), which is otherwise controlled by transport from production regions near the equator and destruction by the Chapman and catalytic mechanisms described above. For this reason, it is important to monitor the abundances of key species in stratospheric chemical and dynamical processes above Toronto: O_3, NO, NO_2, HF, HCl, $ClONO_2$, HNO_3, N_2O, CH_4 and CO. (While H_2O is important to the HO_x catalytic cycle, the FTIR measurement technique is not sensitive to its stratospheric column.)

The abundances of trace gases are also controlled by dynamical processes, which are different in the well-mixed troposphere and the statically stable stratosphere [*Shepherd*, 2003]. The ability to detect dynamical influences and decouple them from chemical influences using TAO FTIR data is limited without the aid of modeling tools in the interpretation of trace gas measurements. Some dynamical influences can be inferred from measurements of hydrogen fluoride (HF), which is extremely stable against chemical loss, or from measurements of CH_4 and N_2O, which are also relatively long-lived.

Finally, although FTIR measurements are not sensitive to tropospheric NO_x and O_3, which are key players in the formation of urban air pollution, TAO measurements have a role to play in the detection of urban and free tropospheric CO, and other trace gases which are involved in tropospheric chemistry, e.g. C_2H_6, C_2H_2, and HCN. The recently installed ground-level O_3 monitor and Brewer spectrophotometer (here used to measure total O_3 and aerosol optical depth) further enrich the TAO data set for studies of urban pollution and pollution transport. With the use of appropriate modeling tools, such as a regional Chemistry-Transport Model (CTM), TAO data can be used in integrated studies of urban pollution and tropospheric pollution transport, involving global satellite measurements as appropriate.

1.3 Thesis Outline and Candidate's Contribution to Collaborative Work

The remainder of this thesis is structured as follows. In Chapter 2, the measurement technique of FTIR solar absorption spectroscopy will be introduced. The Candidate played a key role in establishing TAO during her M.Sc. project. During her Ph.D. project, she worked closely with Bomem engineers, post-doctoral fellows (Dr. Yongjing Zhao and Dr. Dmitri Yashcov), another Ph.D. student (Jeffrey R. Taylor), and five summer students to characterize the instrument and to maintain its optimal alignment. Furthermore, between September 2001 and December 2005 the candidate recorded solar absorption spectra on 83 sunny days as part of a rotating schedule, representing approximately one quarter of all measurement days at TAO during this period.

Chapter 3 will show how total and partial column information is derived from the measured spectra using an Optimal Estimation Method (OEM). The candidate implemented a previously existing OEM code (SFIT-2) but constructed the measurement and *a priori* constraints and the *a priori* profiles herself (Chapter 4). She closely supervised the work of two summer students to write Matlab codes to streamline the conversion of spectra to the SFIT-2 format and also the creation of model atmospheres (density-weighted pressure and temperature, airmass, and VMR profiles) necessary in the inversion process. She herself wrote Matlab code to characterize individual retrievals in terms of their averaging kernels, weighting functions, and degrees of freedom for signal, following the work of others. She also calculated measurement, smoothing and temperature errors in the retrieved total columns and profiles. Her work has since been expanded on by Ph.D. student Jeffrey R. Taylor, and their two codes have been cross-checked for consistency of major outputs. All results in Chapters 3 and 4 were prepared by the Candidate.

Chapter 5 documents retrievals performed as part of the *Retrieval Algorithm Intercomparison Exercise* prescribed by the NDACC community in order for TAO to be granted the status of a Complementary observation station (Section 5.1). Toronto (43.66°N, 79.40°W), Wollongong (34.45°S, 150.88°E) and Lauder (45.04°S, 169.68°E) spectra were exchanged between the Candidate and collaborators from the University of Wollongong. Total column amounts were retrieved from all spectra by both teams and the similarity of the results was refereed by a third party, also from the University of Wollongong. All Toronto retrievals and analysis of discrepancies was performed by the Candidate. Section 5.2 shows the results and characterization of trace gas retrievals spanning the period between 2001-2004, submitted to the NDACC Database in June, 2005, and performed by the Candidate.

Chapter 6 presents the application of the candidate's research to the detection of meso-thermospheric NO (Section 6.1) and intrusions of polar air above Toronto (Section 6.2). In the latter section, TAO partial columns of HCl are also compared to Atmospheric Chemistry Experiment - Fourier Transform Spectrometer (ACE-FTS) satellite measurements. Finally, total columns of CO are compared to MOPITT (Measurements Of Pollution In The Troposphere) satellite measurements in Section 6.3. All results in this section are derived from work carried out independently by the Candidate, with the exception of the recording of Eureka spectra analyzed in Section 6.1 and the direct HCl comparisons with ACE-FTS measurements in Figure 6.6.

Finally, Chapter 7 summarizes the main conclusions of this thesis and provides recommendations for future work.

Chapter 2

Fourier Transform Infrared Spectroscopy

In solar absorption FTIR spectroscopy, the source of the radiation measured by the ground-based detection system is the Sun. The Sun's radiation spectrum is well-known from theoretical blackbody considerations and from independent spacecraft measurements. As this radiation passes through the atmosphere, it is absorbed at well-defined frequencies by all of the molecules present in the atmosphere. The total column amount of an absorbing gas is proportional to the depth of its unique spectral absorption features, while the vertical distribution of the gas can be inferred from the absorption features' shape.

In practice, both total column amounts and vertical profile distributions are derived in two distinct steps. First, an interferogram is recorded and transformed to give the underlying infrared spectrum. Second, a retrieval is performed on the spectrum whereby the *a priori* vertical distribution of the trace gas of interest is adjusted until the modeled and measured spectra agree to within noise levels. This chapter will describe how infrared spectra arise and how they are recorded at TAO. Chapter 3 will describe the details of the retrieval process alluded to in the second step.

2.1 Infrared Atmospheric Spectra

2.1.1 Energy transitions

The mid-infrared absorption spectra of molecules in the atmosphere are a result of simultaneous rotational and vibrational energy level transitions of these same molecules. Such spectra present themselves as bands of absorption with fine structure, where the central wavenumber of the band is determined by the type of vibrational transition taking place (fundamental, first overtone, etc.) while the band structure itself is determined by the allowed rotational transitions for that particular vibrational transition; the energy required to effect a rotational transition is much smaller than that required to effect a vibrational transition. We will now briefly consider rotational, vibrational and rotational-vibrational energy transitions in turn, following the treatment of *Banwell and McCash* [1994] and *Remedios* [1990].

2.1.1.1 Rotational transitions

Pure rotational transitions are observed for molecules with permanent electric dipole moments (e.g. HCl, CO, OCS) that can interact with the incident electromagnetic radiation. This excludes homonuclear diatomic molecules (e.g. O_2 and N_2, except where a collision-induced dipole moment is present), linear molecules with a center of symmetry (e.g. CO_2, C_2H_2) and spherical top molecules (e.g. CH_4). The selection rules for rotational transitions of *linear* molecules are $\Delta J = \pm 1$ (ΔJ is the change in the rotational quantum number) and typical energy separations between rotational levels are 10 cm^{-1} or less. This energy spacing is reduced by deviations of a molecule from the rigid rotor approximation (at high J values) and by heavy isotope substitution. Finally, the intensity of an observed transition is proportional to the transition probability, the population and degeneracy of the initial energy state and the path length of the sample.

2.1.1.2 Vibrational transitions

A molecule with N atoms has 3N degrees of freedom for its motion. Since two sets of three degrees of freedom are required to describe the translational and rotational motion of the center of mass of the molecule, this leaves 3N-6 degrees of freedom for vibration in the molecule (3N-5 for linear molecules since only two degrees of freedom are necessary to describe the rotational motion of the center of mass). Thus H_2O has three normal modes of vibration (ν_1, ν_2, ν_3) while CO_2 (as a linear molecule) has four (but two of its bending modes are degenerate), and, furthermore, each normal mode is allowed its own set of overtone bands. Like rotational transitions, vibrational transitions are also only observed if the vibration causes a change in the dipole moment, e.g. the symmetric stretch of CO_2 (O=C=O) is infrared-inactive, and homonuclear molecules such as O_2 and N_2 still generally remain infrared-inactive.

To a first approximation, a vibrating diatomic molecule can be described as a Simple Harmonic Oscillator (SHO) with the vibrational selection rule of $\Delta\nu = \pm 1$ ($\Delta\nu$ is the change in the vibrational quantum number) and a typical spacing between vibrational levels of 1000 cm^{-1}. Since the bonds of real molecules do not obey Hooke's law (due to, e.g. nonlinear dipole moment variation with atomic displacement) the SHO transforms into an Anharmonic Oscillator with different vibrational selection rules ($\Delta\nu = \pm 1, \pm 2, \pm 3, \ldots$). Transitions from energy levels other than the ground state are known as hot bands because they only become significant at high temperatures when the populations of higher energy levels increase; therefore, only the $\Delta\nu = \pm 1, \pm 2, \pm 3$ transitions (from the ground state) are of practical importance. As a further consequence of anharmonicity, the spacing between each vibrational energy level decreases with increasing vibrational energy, e.g. the HCl fundamental (ν_1) is found at 2886 cm^{-1}, the first overtone ($2\nu_1$) at 5668 cm^{-1} and the second overtone ($3\nu_1$) at only 8347 cm^{-1}.

Finally, in addition to the fundamental modes and their weaker overtones, combination and difference bands are also permitted, e.g. $\nu_1 + \nu_2$ or $2\nu_1 - \nu_2$, however, their intensities are generally very weak. The weak combination/difference or overtone bands can sometimes be greatly enhanced at the expense of a nearby (*accidentally degenerate*) fundamental band in a process referred to as *Fermi resonance*. In addition to changes in intensity, the frequency of the higher (lower) band can also be further raised (lowered).

2.1.1.3 Rotational-vibrational transitions

Under the Born-Oppenheimer approximation, a *diatomic* molecule undergoes vibrational and rotational transitions independently of one another, since the energy scales are separated by two to three orders of magnitude. (We neglect electronic transitions in this discussion, whose energy scales are another two to three orders of magnitude higher than vibrational energy scales.) The selection rules for simultaneous but independent rotational and vibrational transitions can be shown to be a combination of the selection rules for each transition

$$\Delta\nu = \pm1, \pm2, \pm3, \ldots \qquad \Delta J = \pm1. \tag{2.1}$$

Strictly speaking, $\Delta\nu = 0$ is allowed and corresponds to a purely rotational transition, however $\Delta J = 0$ is, in general, not allowed for a *diatomic* molecule, i.e. a vibrational transition must, in general, be accompanied by a rotational transition.

The two distinct branches of the resultant vibrational-rotational band shown in Figure 2.1 are referred to as the P and R branches, from left to right. They correspond to $\Delta J = -1$ and $\Delta J = +1$, respectively. By convention, they are labelled as P1, P2, P3, ... and R0, R1, R2, ... away from the band center, with the numbers signifying the rotational quantum number in the lower vibrational state (J''). P0 does not appear because it would mean transitioning to a negative rotational quantum number in the higher vibrational state ($J' = -1$). Upon closer inspection, it is evident that for high J'' in the P-branch the rotational lines become more sparsely spaced while for high J'' in the R-branch the rotational lines crowd closer together. This is a result of the failure of the Born-Oppenheimer approximation rather than deviations from the rigid rotor

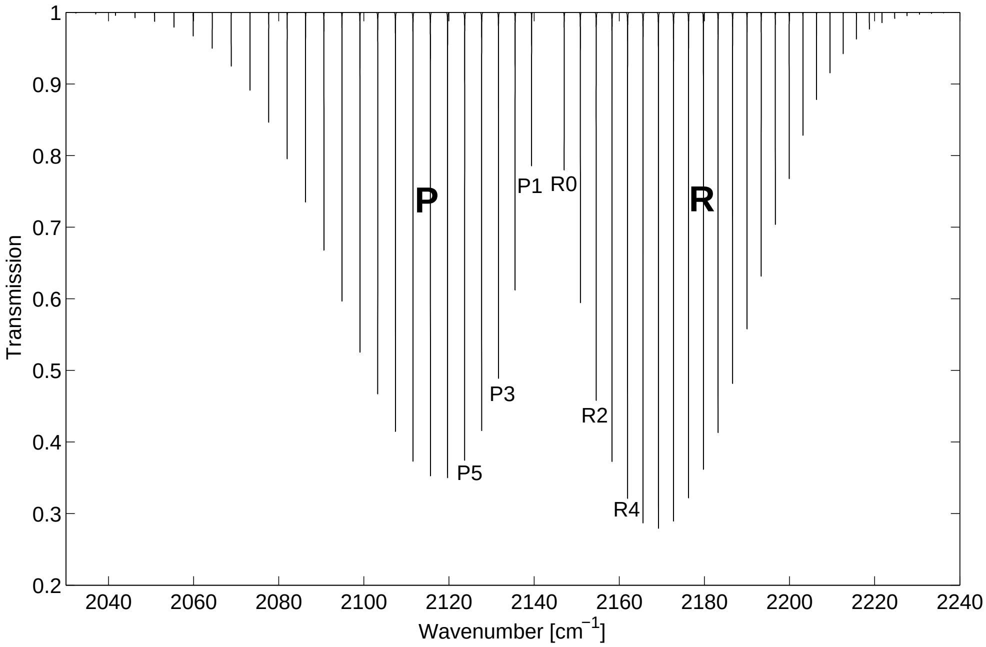


Figure 2.1 Simulated rotational-vibrational transmission spectrum of CO showing its PR band structure.

approximation.

The discussion of rotational-vibrational spectra up to this point applies only to *diatomic* molecules, which are by definition linear. It turns out that the rotational transitions of complex molecules also depend on the vibrational transition that the molecule is undergoing. For a *linear* polyatomic molecule (e.g. CO_2) if the vibration is causing a dipole moment change that is parallel to the principal axis of rotational symmetry then the selection rules remain the same (although the spectra may still become unresolved for heavy molecules where the rotational line spacing becomes very small). On the other hand, if the dipole moment change caused by the vibration has a component that is perpendicular to the principal axis of rotational symmetry then $\Delta J = 0$ ceases to be forbidden and a vibrational transition becomes possible without a rotational transition. Since such vibrational transitions can and do occur at all rotational levels, an intense Q-branch appears in the spectrum at the band center, between the P and R branches

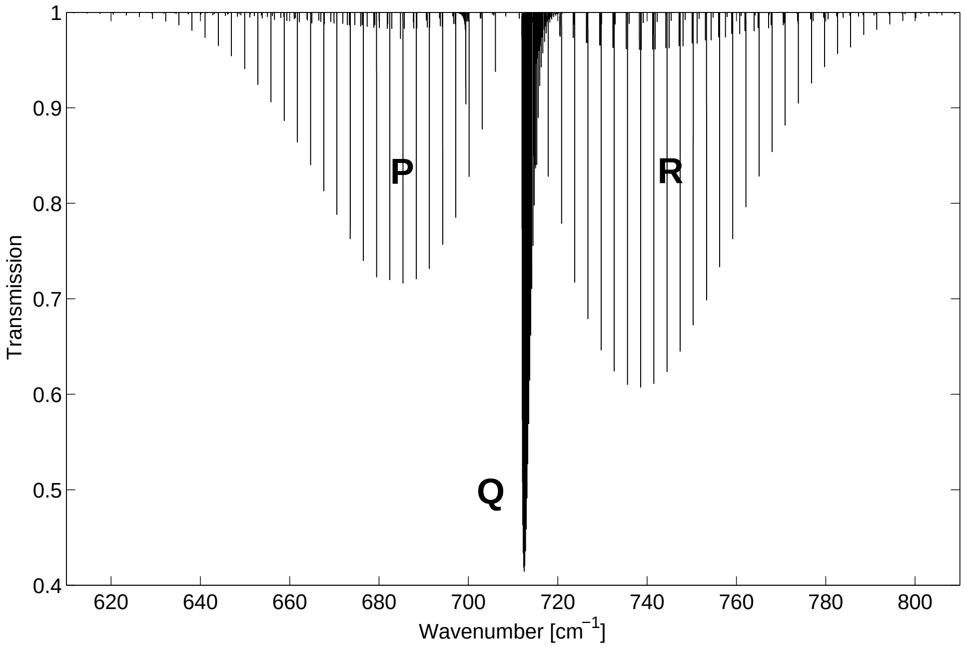


Figure 2.2 Simulated rotational-vibrational transition spectrum of HCN showing its PQR band structure.

(Figure 2.2). Both Figures 2.1 and 2.2 were simulated using the line-by-line GENSPECT code described by *Quine and Drummond* [2002].

Finally, the spectra of molecules with more complicated geometries (symmetric tops (e.g. CH_3Cl), asymmetric tops (e.g. O_3, H_2O), spherical tops (e.g. CH_4) are increasingly complex, and require the introduction of a supplementary rotational quantum number K, which is beyond the scope of this introduction.

2.1.2 Transition line strength and shape

This section follows the discussion of *Remedios* [1990], *Barret* [2003], *Notholt et al.* [2006] and *Rothman et al.* [1998]. The strength $S(\overline{\nu}_0, T)$ [cm^{-1}/(molec $\cdot$ cm^{-2}) = cm $\cdot$ molec^{-1}] of an absorption line with a central frequency $\overline{\nu}_0$ [cm^{-1}] (distinguished from the vibrational quantum number by the bar) in a rotational-vibrational band and at temperature

T [K] is given by

$$S(\overline{\nu}_0, T) = S(\overline{\nu}_0, T_0)\frac{Q(T_0)}{Q(T)}\frac{exp(-E''/(kT))}{exp(-E''/(kT_0))}\frac{(1 - exp(-hc\overline{\nu}_0/(kT)))}{(1 - exp(-hc\overline{\nu}_0/(kT_0)))}, \tag{2.2}$$

where T_0 is a reference temperature of 296 K, Q is the partition function, E'' is the lower state energy of the transition, h is the Planck constant, c is the speed of light, and k is the Boltzmann constant. The sensitivity of Equation 2.2 to temperature is strongly influenced by E''. More details on the calculation of vibrational and rotational partition functions can be found in *Fischer et al.* [2003], *Rothman et al.* [1998, App. A], and *Notholt et al.* [2006].

Theoretical rotational-vibrational transition lines are broadened by three underlying mechanical processes: 1) Lorentz (alternatively *pressure* or *collisional*) broadening (either due to air or self), 2) Doppler broadening and 3) natural broadening. Natural broadening refers to the broadening due to the Heisenberg uncertainty in the energy levels of the gas molecules; it is negligible in the infrared when compared to Lorentz or Doppler broadening in the atmosphere.

The absorption coefficient $k(\overline{\nu})$ $[1/(\text{molec} \cdot \text{cm}^{-2})]$ that describes a broadened transition line is modeled as the product of the line strength with the broadening function $f(\overline{\nu} - \overline{\nu}_0)$ $[1/\text{cm}^{-1}]$, normalized as per Equation 2.4

$$k(\overline{\nu}) = S(\overline{\nu}_0, T)\; f(\overline{\nu} - \overline{\nu}_0) \tag{2.3}$$

$$\int_0^\infty f(\overline{\nu} - \overline{\nu}_0) d\overline{\nu} = 1. \tag{2.4}$$

We will now discuss the first two broadening mechanisms in turn.

2.1.2.1 Lorentz broadening

Lorentz broadening occurs when gas molecules collide *strongly* with one another (formally known as the *binary impact* approximation, where the collision duration is assumed negligible compared to the time between collisions), which randomizes the time-domain phase of the SHO radiation emitted by the molecules. Since the Fourier transform of a truncated SHO emission is a sinc function, a single truncated SHO emission will appear as a sinc function in the spectral domain. In the limit of many simultaneous truncated emissions and many molecules, the resultant superposition of sinc functions can be shown to reduce to the Lorentzian functional form

$$f(\overline{\nu} - \overline{\nu}_0) = \frac{1}{\pi} \frac{\gamma_L}{(\overline{\nu} - \overline{\nu}_0 - \delta_L)^2 + \gamma_L^2}, \tag{2.5}$$

where

$$\gamma_L(T, p) = \left(\frac{T_0}{T}\right)^{n_{air}} \left[\gamma_{air}(T_0, p_0)(p - p_s) + \gamma_{self}(T_0, p_0)p_s\right]. \tag{2.6}$$

The pressure-dependent shift of the line center frequency, $\delta_L = \delta_{air}p$, is introduced in Equation 2.5 to account for the failure of the *strong* collision approximation at high pressures, where p is the pressure and δ_{air} is the coefficient of the air-broadened pressure shift [$\text{cm}^{-1} \cdot \text{atm}^{-1}$] measured at a reference temperature and pressure ($T_0 = 296$ K,$p_0 = 1$ atm). δ_L has a temperature dependence of the same form as γ_L, and n_{air} is used in the absence of other information. The Half-Width at Half-Maximum (HWHM) of the Lorentzian, γ_L [cm^{-1}], is calculated according to Equation 2.6 given the different self- and air-broadening coefficients γ_{air} and γ_{self} [$\text{cm}^{-1} \cdot \text{atm}^{-1}$], which are also measured at $T_0 = 296$ K and $p_0 = 1$ atm, the partial pressure of the gas, p_s, and the exponent of the temperature dependence of the broadening coefficients, n_{air}. Equation 2.6 assumes one temperature dependence for both the self- and air-broadening processes in the absence of other information (n_{air} takes on different values depending on the nature of the colliding partners and the particular transition).

The Lorentz broadening halfwidth dominates in the lowermost atmosphere but decreases rapidly with altitude due to its dependence on pressure, becoming comparable to Doppler broadening (see below) at ∼25 km, and negligible at ∼40 km. Depending on frequency, this broadening formalism breaks down many halfwidths away from the line center in the line wings, where the absorption lineshape is dominated by collisions of finite duration, among other effects.

2.1.2.2 Doppler broadening

Doppler broadening occurs when the mean free path of the gas molecules is large enough to allow their velocities to become high and Maxwell-Boltzmann-distributed. This leads to a smearing out of the absorption coefficient through the lineshape function which now has the Gaussian form of Equation 2.7, where M is the molecular mass and the HWHM is equal to $\sqrt{ln2}\gamma_D$ [cm^{-1}].

$$f(\overline{\nu}-\overline{\nu}_0)=\frac{1}{\sqrt{\pi}\gamma_D}\,exp\left(-\frac{(\overline{\nu}-\overline{\nu}_0)^2}{\gamma_D^2}\right) \tag{2.7}$$

$$\gamma_D(T)=\sqrt{\frac{2kT}{M}}\frac{\overline{\nu}_o}{c} \tag{2.8}$$

The Doppler broadening linewidth does not vary significantly throughout the atmosphere since it is not dependent on pressure. It becomes much bigger than the Lorentz broadening halfwidth in the upper stratosphere and above, where the mean free path is large. Finally, the Doppler broadening halfwidth is *reduced* as the mean free path decreases with increasing pressure in a process referred to as collisional or Dicke narrowing [*Dicke*, 1953]. Although this is a second-order effect compared to Lorentz broadening in the troposphere, it is of some consequence in the stratosphere, where such small changes in the Doppler halfwidth are not obscured by the rapidly decreasing Lorentz halfwidth [*Barret et al.*, 2005].

2.1.2.3 Voigt broadening

When both broadening processes are important, as is the case throughout the stratosphere, the Voigt line shape function is used:

$$f(x,y)=\frac{1}{\gamma_D}\frac{y}{\pi^{3/2}}\int_{-\infty}^{\infty}\frac{exp(-t^2)}{(x-t)^2+y^2}dt \tag{2.9}$$

where $x=(\overline{\nu}-\overline{\nu}_0)/\gamma_D$ and $y=\gamma_L/\gamma_D$. Equation 2.9 can be shown to be the convolution of the Lorentz and Doppler lineshape functions. Collisional (Dicke) narrowing and collisional line mixing are not usually accounted for in the Voigt formulation. Collisional line mixing refers to the case when neighbouring transitions cannot be considered independent of one another because the wings of the absorption features now overlap, as is the case under typical atmospheric pressures found at the surface. In collisional line mixing, the band strength is not altered, but rather it is redistributed as a function of frequency. The Voigt profile is used to represent absorption line shapes in the SFIT-2 retrieval algorithm described in detail in Chapter 3.

2.1.2.4 HITRAN spectral database

The HITRAN (HIgh Resolution TRANsmission) database is a research-standard database containing reference temperature and pressure line strengths ($S(\overline{\nu}_0)$) and pressure-broadening and shift parameters ($\gamma_{air}, \gamma_{self}, \delta_{air}$, along with their temperature dependence exponent n_{air}) for the vibrational-rotational transitions of the 39 most common atmospheric molecules. The database also contains lower state energy (E'') information necessary for the line strength calculation and several other quantum mechanical parameters.

HITRAN 1996 [*Rothman et al.*, 1998], HITRAN 2000 and 2001 [*Rothman et al.*, 2003], as well as HITRAN 2004 [*Rothman et al.*, 2005] exist concurrently, and new, more accurate measurements are continuously being compiled for future updates of the database. In addition to line-by-line parameters, cross-section data and aerosol indices of refraction are also included for some species in some pressure, temperature and wavenumber ranges.

2.1.3 Equation of radiative transfer

Having described the strength and shape of absorption features of the molecules of a gas at a specified temperature, pressure, and frequency, we will now describe a model for the transmission of solar radiation from the top of the atmosphere (TOA) to the surface. This section follows the discussion of *Kidder and VonderHaar* [1995] and *Stephens* [1994].

The simplifying assumption that scattering can be ignored at infrared frequencies makes our task easier and we can write

$$\frac{dI_{\overline{\nu}}(\theta,\phi)}{ds} = \sigma_a[B_{\overline{\nu}}(T) - I_{\overline{\nu}}(\theta,\phi)] \tag{2.10}$$

where $I_{\overline{\nu}}$ is the monochromatic radiance $[\mathrm{W} \cdot \mathrm{cm}^{-2}\mathrm{sr}^{-1}]$ as a function of solar zenith angle θ and azimuthal angle ϕ, s is the slant path distance [cm], σ_a is the volume absorption coefficient $[\mathrm{cm}^{-1}]$, and $B_{\overline{\nu}}(T)$ is the monochromatic blackbody emission at temperature T $[\mathrm{W} \cdot \mathrm{cm}^{-2}\mathrm{sr}^{-1}]$. Equation 2.10 is a first order linear ordinary differential equation known in the remote sounding community as Schwarzchild's Equation, which can be integrated to give

$$I_{\overline{\nu}}(0) = I_{\overline{\nu}}(TOA) \cdot exp\left(-\int_{TOA}^{0} \rho k_{\overline{\nu}} ds\right) + \int_{TOA}^{0} \rho k_{\overline{\nu}} B_{\overline{\nu}}(T) \cdot exp\left(-\int_{TOA}^{s} \rho k_{\overline{\nu}} ds''\right) ds \tag{2.11}$$

where ρ is the density of the absorbing gas $[\mathrm{molec} \cdot \mathrm{cm}^{-3}]$, and the product $\rho k_{\overline{\nu}}$ is equal to the volume absorption coefficient introduced previously. If we define $\mu = cos\theta$ and the vertical optical depth from height z_1 to z_2 as $\chi_{\overline{\nu}} = \int_{z_1}^{z_2} \rho k_{\overline{\nu}} dz$ then $d\chi_{\overline{\nu}} = \mu \rho k_{\overline{\nu}} ds$ and Equation 2.11 simplifies to

$$I_{\overline{\nu}}(0) = I_{\overline{\nu}}(TOA) \cdot exp\left(-\frac{\chi_{\overline{\nu}}^0}{\mu}\right) + \int_0^{\chi_{\overline{\nu}}^0} B_{\overline{\nu}}(T) \cdot exp\left(-\frac{\chi_{\overline{\nu}}^0 - \chi_{\overline{\nu}}}{\mu}\right) \frac{d\chi_{\overline{\nu}}}{\mu} \tag{2.12}$$

where $\chi_{\overline{\nu}}^0$ is the vertical *optical depth* from TOA to the surface. If we finally introduce the *transmission* between two vertical optical depths as $\tau_{\overline{\nu}}(\chi_1, \chi_2) = exp(-|\chi_2 - \chi_1|)$ then Equation 2.12 is further simplified to

$$I_{\overline{\nu}}(0) = I_{\overline{\nu}}(TOA) \cdot \tau_{\overline{\nu}}(0, \chi_{\overline{\nu}}^0)^{\frac{1}{\mu}} + \int_1^{\tau_{\overline{\nu}}(0,\chi_{\overline{\nu}}^0)} B_{\overline{\nu}}(T) \cdot \tau_{\overline{\nu}}(\chi_{\overline{\nu}}, \chi_{\overline{\nu}}^0)^{\frac{1}{\mu}-1} d\tau_{\overline{\nu}}. \tag{2.13}$$

When $\theta = 0$ then $\mu = 1$ and Equation 2.13 simply states that the radiance at the surface is equal to the radiance at TOA multiplied by the transmission from TOA to the surface, added to the radiance emitted by each layer multiplied by the transmission from that layer to the surface. In the SFIT-2 algorithm (described in detail in Chapter

3) the atmospheric emission term is ignored in the forward model on grounds that it is negligible compared to the large solar radiance [*Barret*, 2003]. A recent calculation by *Notholt et al.* [2006] has shown that the emission term is of order 0.5% of the solar radiance at the lowest wavenumbers accessible to a typical mid-infrared spectrometer ($\sim$800 cm^{-1}), and that it should be modeled at wavenumbers below $\sim$1500 cm^{-1} in order to separate it from the effects of detector nonlinearity correctly.

Finally, we note that the effects of atmospheric curvature and of refraction are not described by Equation 2.13. In practice these effects are taken into account by the auxiliary *fscatm* subroutine [*Gallery et al.*, 1983] of the SFIT-2 algorithm which performs detailed ray tracing calculations.

2.2 Fourier Transform Spectrometer

In the preceding section we described how infrared absorption spectra arise for a gas of a single molecule at a given temperature and pressure; furthermore, we outlined the equations of radiative transfer, which allow us to model the transmission of solar radiation from the top of the atmosphere to the observation point at the surface, taking into account the variable temperature, pressure, and absorber amount along the transmission path. The Fourier Transform Spectrometer (FTS), which has the ability to record solar absorption spectra with a very high resolution and over a broad wavenumber range (e.g. 0.004 cm^{-1} and 720-8500 cm^{-1} at TAO) is described next.

2.2.1 The ideal Fourier transform spectrometer

The FTS is based upon a Michelson interferometer, after its original designer [*Michelson*, 1891, 1892]. Following Figure 2.3, it is a device which uses a *beamsplitter* to divide incoming radiation into two beams; it then introduces a phase delay between the beams, and finally, recombines them before detection. The phase delay is introduced in one arm

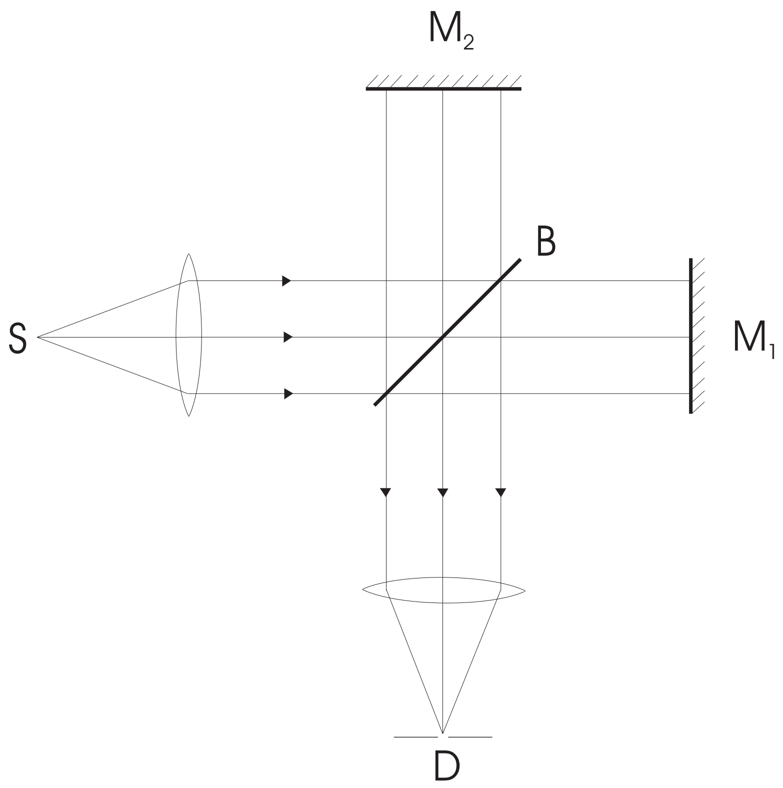


Figure 2.3 Michelson interferometer schematic after *Chamberlain* [1979], where S = point source, B = beamsplitter, M_1 = fixed mirror, M_2 = moving mirror, and D = detector.

of the interferometer by means of a moving mirror. If the source radiation is monochromatic (a delta-function in spectral space) and the moving mirror is scanned over a distance, the signal at the detector will vary sinusoidally as a function of mirror distance because the recombining beams experience successive cycles of constructive and destructive interference. Since the Fourier transform of a delta-function is a sine function, it is now clear that the FTS has performed such a transform on the incoming monochromatic radiation. In the case of polychromatic input radiation the interference pattern at the detector will appear complicated, but will in fact be a straightforward superposition of many sinusoidal interference patterns of varying frequency and amplitude, depending on the spectrum of the source. In Fourier transform spectroscopy, the interference signal recorded at the detector is called an *interferogram.* Its Fourier transform (carried out by an external computer) yields the spectrum of the source – in our case solar infrared radiation that has interacted with various trace gases present in the atmosphere. Finally, we note that a complementary beam is at all times reflected back to the source (its intensity is zero when the primary beam is at a maximum and vice versa), however, it is not often used in practice because of the difficulty of separating it from the source signal. When this beam is used, the technique is referred to (somewhat ambiguously) as *dual-beam interferometry.*

Fourier transform spectroscopy is a well known and widely applied technique for studying the spectroscopic properties of solid, liquid and gaseous matter. Other detection techniques can also be thought of as interferometric, with the prism spectrometer dividing the input beam into an infinite number of output beams, the grating spectrometer into N output beams (N is the number of grooves in the grating), and the Fabry-Perot into m beams (m is the *finesse* of the Fabry-Perot); however, the FTS divides the input beam into only two beams – the minimum required for interference. This arrangement gives the FTS the ultimate in throughput and efficiency, and consequently, according to *Bell* [1972], the concept of the two-beam interferometer is ideal and cannot be improved.

2.2.2 Mathematical basis

Following the intuitive treatment of *Griffiths and de Haseth* [1986], for a monochromatic source at wavenumber $\overline{\nu}$ (wavelength $\lambda = 1/\overline{\nu}$), we write the intensity recorded at the detector as a function of the optical path difference, $I'(\delta)$, as the sum of an AC component superimposed on a DC component equal to one half of the intensity as a function of wavenumber, $I(\overline{\nu})$:

$$I'(\delta) = \frac{1}{2}I(\overline{\nu})\ \{1 + cos2\pi\frac{\delta}{\lambda}\} = \frac{1}{2}I(\overline{\nu})\ \{1 + cos2\pi\overline{\nu}\delta\}. \tag{2.14}$$

In practice, one is primarily interested in the AC component (the DC component can be used to ascertain that the source strength is not changing during the measurement), and the interferogram, $I(\delta)$, is written as

$$I(\delta) = \frac{1}{2} I(\overline{\nu})\ H(\overline{\nu})\ cos 2\pi\overline{\nu}\delta. \tag{2.15}$$

The factor $H(\overline{\nu})$ is less than 1 and is included to account for the frequency-dependent beamsplitter, detector and amplifier losses, which are present in a non-ideal FTS, and which are constant from measurement to measurement. Finally, the function $B(\overline{\nu}) = I(\overline{\nu})\ H(\overline{\nu})$ is introduced to represent the intensity of the source at wavenumber $\overline{\nu}$ as modified by the non-ideal instrument characteristics (geometric non-idealities will be treated in the next section). Thus, in its simplest form, the interferogram of a monochromatic source at wavenumber $\overline{\nu}$ is written as

$$I(\delta) = \frac{B(\overline{\nu})}{2}\ cos\, 2\pi\overline{\nu}\delta. \tag{2.16}$$

If the source contains more than one frequency, the intensity at the detector can be expressed as an integral over all wavenumbers:

$$I(\delta) = \int_{-\infty}^{+\infty} \frac{B(\overline{\nu})}{2}\ cos\, 2\pi\overline{\nu}\delta\ d\overline{\nu}. \tag{2.17}$$

Equation 2.17 is one half of a cosine Fourier transform pair. The inverse transform is written as

$$B(\overline{\nu}) = 2 \int_{-\infty}^{+\infty} I(\delta)\ cos\, 2\pi\overline{\nu}\delta\ d\delta, \tag{2.18}$$

or equivalently (for I(δ) an even function) as

$$B(\overline{\nu}) = 4 \int_{0}^{\infty} I(\delta)\ cos\, 2\pi\overline{\nu}\delta\ d\delta. \tag{2.19}$$

Following the treatment of *Remedios* [1990], we can derive Equation 2.19 in a more rigorous manner from a consideration of the electric field vector of the input beam to the interferometer. The input beam intensity, I_o, is proportional to the time average ($\langle\ \ \rangle = \frac{1}{T}\int_{-\frac{T}{2}}^{+\frac{T}{2}} dt$) of the complex conjugate (*) product of the electric field component [*Chamberlain*, 1979, p. 75]:

$$I_o = \langle \vec{E}(\vec{r},t)^* \ \cdot\ \vec{E}(\vec{r},t) \rangle, \tag{2.20}$$

where $\vec{r}$ and t are position and time, respectively. We also recognize that the electric field vector can be expressed as

$$\vec{E}(\vec{r},t) = \int_{-\infty}^{+\infty} \sqrt{B(\nu)}\; e^{2\pi i(\nu t+\vec{r}\cdot\vec{k}(\nu))} d\nu, \tag{2.21}$$

where $\sqrt{B(\nu)}$ is the amplitude of frequency component ν (s^{-1}), and $k = \nu/c$ (m^{-1}) is the spatial frequency, taken in the direction of beam propagation. This formulation will become useful shortly.

If the beam in one arm of the interferometer is delayed by a time interval τ, then the electric field emerging from the spectrometer after the two beams are recombined is written as

$$\vec{E}_{out}(\vec{r},t) = \frac{\vec{E}(\vec{r},t)}{2} + \frac{\vec{E}(\vec{r},t+\tau)}{2} \tag{2.22}$$

and hence the intensity at the detector can be written as

$$I_{out} = \langle \left[\frac{\vec{E}(\vec{r},t)}{2} + \frac{\vec{E}(\vec{r},t+\tau)}{2}\right]^* \cdot \left[\frac{\vec{E}(\vec{r},t)}{2} + \frac{\vec{E}(\vec{r},t+\tau)}{2}\right]\rangle. \tag{2.23}$$

After expanding Equation 2.23, the intensity is written as

$$I_{out} = \frac{1}{T}\int_{-\frac{T}{2}}^{+\frac{T}{2}} [\frac{\vec{E}(\vec{r},t)^*}{2} \cdot \frac{\vec{E}(\vec{r},t)}{2} + \frac{\vec{E}(\vec{r},t+\tau)^*}{2} \cdot \frac{\vec{E}(\vec{r},t+\tau)}{2} + \\ \frac{\vec{E}(\vec{r},t)^*}{2} \cdot \frac{\vec{E}(\vec{r},t+\tau)}{2} + \frac{\vec{E}(\vec{r},t+\tau)^*}{2} \cdot \frac{\vec{E}(\vec{r},t)}{2}] \; dt, \tag{2.24}$$

which simplifies to

$$I_{out} = \frac{I_o}{2} + \frac{1}{T}\int_{-\frac{T}{2}}^{+\frac{T}{2}} \left[\frac{\vec{E}(\vec{r},t)^*}{2} \cdot \frac{\vec{E}(\vec{r},t+\tau)}{2} + \frac{\vec{E}(\vec{r},t+\tau)^*}{2} \cdot \frac{\vec{E}(\vec{r},t)}{2}\right] dt \tag{2.25}$$

when we recognize that the first term in Equation 2.24 is equal to the input beam intensity divided by 4, and that the second term is equal to the first when the averaging time T is sufficiently greater than the phase delay τ. Finally, upon substituting Equation 2.21 into Equation 2.25 we write

$$I_{out} = \frac{I_o}{2} + \frac{1}{T}\int_{-\frac{T}{2}}^{+\frac{T}{2}} \int_{-\infty}^{+\infty} \frac{B(\nu)}{4} [e^{-2\pi i(\nu t+\vec{r}\cdot\vec{k})+2\pi i(\nu(t+\tau)+\vec{r}\cdot\vec{k})} +$$

$$+e^{-2\pi i(\nu(t+\tau)+\vec{r}\cdot\vec{k})+2\pi i(\nu t+\vec{r}\cdot\vec{k})}]d\nu dt. \tag{2.26}$$

Expanding and simplifying the exponential terms reveals that the time variable vanishes (Equation 2.27). The exponential terms are further simplified into one cosine term, and the time average integral is simply equal to its argument when that argument has no time dependence (Equation 2.28). Equation 2.29 shows the final form of the expression for the intensity leaving the spectrometer after the interference of two beams separated by a time delay τ:

$$\begin{aligned} I_{out} &= \frac{I_o}{2} + \frac{1}{T}\int_{-\frac{T}{2}}^{+\frac{T}{2}}\int_{-\infty}^{+\infty}\frac{B(\nu)}{4}\left[e^{2\pi i\nu\tau}+e^{-2\pi i\nu\tau}\right]d\nu dt && (2.27)\\ &= \frac{I_o}{2} + \int_{-\infty}^{+\infty}\frac{B(\nu)}{4}\left[2cos2\pi\nu\tau\right]d\nu && (2.28)\\ &= \frac{I_o}{2} + \int_{-\infty}^{+\infty}\frac{B(\nu)}{2}\ \ cos2\pi\nu\tau\ \ d\nu. && (2.29)\end{aligned}$$

As before, we are only interested in the AC component of the intensity:

$$I(\tau) = \int_{-\infty}^{+\infty}\frac{B(\nu)}{2}\ \ cos2\pi\nu\tau\ \ d\nu. \tag{2.30}$$

Given $\tau = \delta/c$, $\nu = c\overline{\nu}$ and that the integral of intensity from $-\infty$ to $+\infty$ in frequency space is equal to the same integral in wavenumber space, we can now express Equation 2.30 in terms of δ and $\overline{\nu}$:

$$I(\delta) = \int_{-\infty}^{+\infty}\frac{B(\overline{\nu})}{2}\ \ cos2\pi\overline{\nu}\delta\ \ d\overline{\nu}, \tag{2.31}$$

which is identical to Equation 2.17 resulting from the intuitive derivation after *Griffiths and de Haseth* [1986]. The inverse cosine transform of Equation 2.31 is again identical with Equation 2.18. Since $I(\delta)$ is real, Equations 2.17 and 2.18 can also be written in terms of the full complex Fourier transform and the spectrum taken to be the real part of the transform:

$$I(\delta) = \int_{-\infty}^{+\infty} B(\overline{\nu})exp(i2\pi\overline{\nu}\delta)d\overline{\nu}, \tag{2.32}$$

$$B(\overline{\nu}) = \int_{-\infty}^{+\infty} I(\delta)exp(-i2\pi\overline{\nu}\delta)d\delta, \tag{2.33}$$

where the multiplicative constants have been omitted. Equation 2.33 cannot be implemented with a real spectrometer and data acquisition system because the former cannot scan to an infinite path difference and the latter cannot record the interferogram at infinitesimally small scan mirror displacements $d\delta$. As a consequence of these two limitations, there is a limit to the practical resolution of the spectrometer and to its spectral coverage, respectively. These and other limitations are now discussed in turn.

2.2.3 A real Fourier transform spectrometer

Two fundamental sources of non-ideality in an FTS are its finite path difference and its finite *Field-Of-View* (FOV). The former causes the *Instrumental Line Shape* (ILS) to be a *sinc* function ($sinc\ x = sin(x)/x$), while the latter broadens the ILS with a rectangular function and shifts it to lower wavenumbers. The ILS is, by definition, the instrument response to a delta-function radiation input, i.e. to purely monochromatic radiation. Another fundamental limitation, previously alluded to, arises due to the fact that the interferogram cannot be sampled and transformed continuously; this puts a finite limit on the maximum wavenumber accessible to the FTS.

2.2.3.1 Finite path difference

The effect of a finite maximum Optical Path Difference (OPD) achievable with the real spectrometer is modeled by multiplying the infinite interferogram by a boxcar function equal to 1 between -L and + L, and zero everywhere else (L = OPD$_{max}$). By the convolution theorem, this multiplication in interferogram space is equivalent to convolving the Fourier transform of the interferogram (the spectrum) with the Fourier transform of the boxcar function (the *sine cardinal* or *sinc* function). From the mathematical definition of the convolution operation, it follows that the delta-function spectrum of monochromatic radiation convolved with the sinc function introduced by the finite OPD will again yield a sinc function of the following form:

$$B(\overline{\nu}) = 2L\frac{sin(2\pi(\overline{\nu}-\overline{\nu}_0)L)}{2\pi(\overline{\nu}-\overline{\nu}_0)L} = 2L\ sinc_\pi(2(\overline{\nu}-\overline{\nu}_0)L). \tag{2.34}$$

$\mathrm{sinc}_\pi(x) = sin(\pi x)/(\pi x)$ is an alternative definition of the sinc function whose integral between $\pm\infty$ is 1. The first zeros of the sinc function (Figure 2.4) are separated by $1/L$ cm^{-1}, which introduces a practical limit on the resolution of the FTS. Moreover, for polychromatic input radiation, weak absorption features may become obscured by the pronounced negative sidelobes of the sinc function corresponding to a nearby strongly absorbing frequency. This problem is reduced through the technique of *apodization*. Instead of a boxcar, one can use, for example, a triangular apodization function which results in a sinc^2 Fourier transform. The sinc^2 function has greatly reduced positive sidelobes, however this is achieved at the cost of increasing the width of the central maximum: the distance between the first zeros increases from $1/L$ for the sinc function to $2/L$ for the sinc^2 function, thus reducing the overall resolution of the instrument. Many apodization functions exist, however none of them are capable of simultaneously reducing the side lobes and maintaining the line width.

Based on Figure 2.4 we can see that two monochromatic spectral absorption features of equal strength separated by $1/L$ [cm^{-1}] will be fully resolved, e.g. a 250-cm-OPD spectrometer will permit us to resolve features that are 0.004 cm^{-1} apart. However, the features will first *begin* to appear resolved at $0.66/L$, which is referred to as the *unapodized* resolution criterion, e.g. $0.66 \cdot 0.004$ cm^{-1} = 0.0026 cm^{-1}. This is in contrast to the Rayleigh resolution criterion, which was developed for the sinc^2 ILS of a grating spectrometer, and according to which two sinc^2-shaped absorption features are resolved when the peak of one is located at the first zero of the other. Using this criterion would not be correct for the case of the sinc-shaped ILS of the FTS as it would correspond to a separation of only $0.5/L$. Sometimes, the resolution criterion for an FTS is given as $0.73/L$, which results in a ~20% dip between the two absorption features, as is the case for the Rayleigh criterion applied to a grating spectrometer with the sinc^2 ILS.

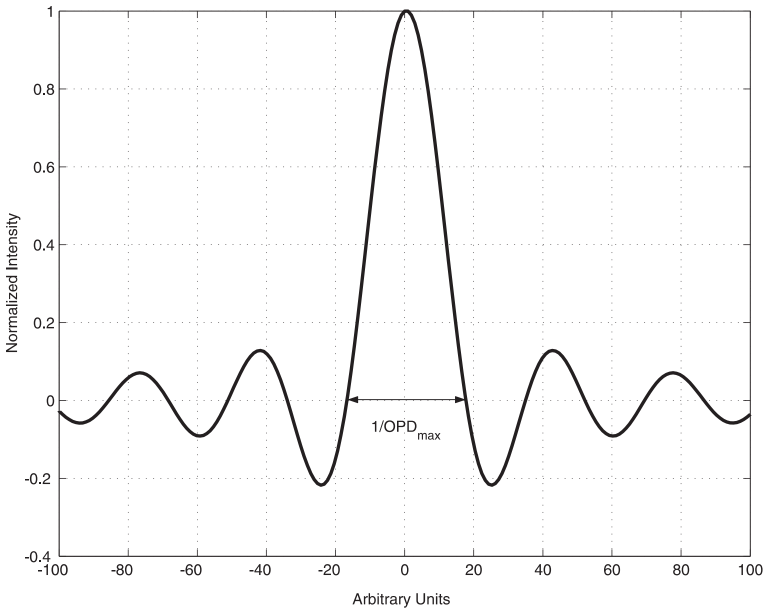


Figure 2.4 Instrumental lineshape of the sinc functional form introduced by a finite maximum optical path difference in the FTS.

2.2.3.2 Finite field-of-view

The FOV of an FTS is defined as the solid angle subtended by the entrance aperture at the collimating mirror (or equivalently at the collimating lens, as shown in Figure 2.3 for simplicity.) The solid angle subtended at the collimating lens by the point source of the ideal FTS is zero. While only a point source can produce a perfectly collimated beam and hence the ideal interference pattern, it reduces the energy throughput of the FTS to zero and makes the ideal FTS useless in practice. To further illustrate this, we now formally introduce the concept of throughput, following *Bell* [1972].

The radiant flux, $d\Phi$ [W], from a source of finite size ds [m^2] (inclined at an angle θ to the optical axis), striking a collimating optical element of size dA^* [m^2] (inclined at an angle ϕ to the optical axis), a distance r [m] away from this source is

$$d\Phi = \frac{B}{r^2}(cos\ \theta\ ds)(cos\ \phi\ dA^*), \tag{2.35}$$

where B is the proportionality constant known as the radiance or intensity [W m^{-2} sr^{-1}]. Reorganizing the terms, we recognize $cos\ \theta\ ds/r^2$ as dΩ, or the differential solid angle element subtended at the collimator by the source, and *cos* $\phi\ dA^*$ as the projected area of the collimator of size dA. We now write the radiant flux as

$$d\Phi = B\ dA\ d\Omega. \tag{2.36}$$

This general reasoning applies to any pair of lossless optical elements which serve as one another's source elements, depending on the direction of the beam; therefore, for an instrument comprising a series of such elements we can write the throughput, E [m^2 sr] (from the French word *étendue*) as

$$E \equiv (d\Phi/B)_{source} = dA\ d\Omega = dA'\ d\Omega' = dA''\ d\Omega'' = \cdots = (d\Phi/B)_{detector}, \tag{2.37}$$

where $'$ and $''$ denote successive optical element pairs. Defined in this way, the throughput of an FTS is identical at any point in a lossless instrument, from the source to the detector. It involves the product of the projected area of the beam and the solid angle subtended at this beam cross-section by the nearest optical element acting as the source. This result must also hold true from energy conservation arguments.

If, following *Chamberlain* [1979], we define the collimator aperture A_c, the solid angle subtended at the collimator by the source Ω_s, the source aperture A_s, the solid angle subtended at the source by the collimator Ω_c, and the focal length of the collimator f_c, then the relationship between these parameters is

$$f_c^2 = \frac{A_s}{\Omega_s} = \frac{A_c}{\Omega_c} \tag{2.38}$$

from the definition of the solid angle (assuming that the size of the collimator aperture, which coincides with the size of the projected collimator area, is not too large). The throughput can finally be written as

$$E = \Omega_c A_s = \Omega_s A_c, \tag{2.39}$$

which shows that a practical FTS must have a finite-sized entrance aperture for a non-zero throughput, even though this has the unwanted effect of causing a divergence in the beam and introducing further distortions in the ILS.

An extended source introduces further distortions in the ILS due to the fact that rays that are at an angle α to the principal optical axis experience a longer path difference ($\delta/cos\alpha$) for every mirror displacement δ. Therefore, at certain mirror displacements, off-axis rays will become out of phase with on-axis rays and the interference pattern will disappear at certain wavenumbers. Thus, the differential increment in the interference pattern power, $dS_{\overline{\nu}_0}$, produced by the off-axis rays of monochromatic radiation at $\overline{\nu}_0$ (and assuming that α is small such that $\cos\alpha \approx 1 - \alpha^2/2 = 1 - \Omega/2\pi$) is written as

$$dS_{\overline{\nu}_0}(\delta) = S_{\overline{\nu}_0}\left[1 + cos(2\pi\,\overline{\nu}_0\,\delta(1 - \frac{\Omega}{2\pi}))\right]\frac{d\Omega}{\Omega_s}, \tag{2.40}$$

where Ω is the solid angle of the cone of half-angle α, $d\Omega$ is the solid angle increment from α to $\alpha + d\alpha$, and Ω_s is the solid angle due to the maximum angle of divergence in the beam (dictated by the size of the source). The differential element is derived from a consideration of $S_{\overline{\nu}_0}$, i.e. the total power emitted by the source in a given spectral range [W], given by the product of the radiance [W m^{-2} sr^{-1}] and the throughput [m^2 sr], and also by making use of Equation 2.39:

$$dS_{\overline{\nu}_0} = I_{\overline{\nu}_0}\,dE = \frac{S_{\overline{\nu}_0}}{E}\,\Omega_c\,dA_s = \frac{S_{\overline{\nu}_0}}{\Omega_s\,A_c}\,d\Omega_s\,A_c = \frac{S_{\overline{\nu}_0}\,d\Omega_s}{\Omega_s}, \tag{2.41}$$

where $d\Omega_s$ corresponds to $d\Omega$ in Equation 2.40 (the subscript was retained for clarity), and dA_s is the annular differential area element between α and $\alpha + d\alpha$.

The total power in the interference pattern is obtained by integrating Equation 2.40 over the solid angle, and consists of a constant and a variable term :

$$\begin{aligned} S_{\overline{\nu}_0}(\delta) &= \int_0^{\Omega_s} S_{\overline{\nu}_0}\left[1 + cos(2\pi\,\overline{\nu}_0\,\delta(1 - \frac{\Omega}{2\pi}))\right]\frac{d\Omega}{\Omega_s} && (2.42)\\ &= S_{\overline{\nu}_0} + S_{\overline{\nu}_0}\,sinc_\pi\left[\frac{\overline{\nu}_0\delta\Omega_s}{2\pi}\right]\,cos\left[2\pi\overline{\nu}_0\delta(1 - \frac{\Omega_s}{4\pi})\right]. && (2.43)\end{aligned}$$

The variable term is altered in two ways in comparison to Equation 2.14. The first effect of a finite FOV is to shift the spectral feature at wavenumber $\overline{\nu}_0$ to the lower wavenumber $\overline{\nu}_0(1 - \Omega_s/4\pi)$. The magnitude of this shift is of practical consequence in

FTIR spectroscopy and is accounted for after interferogram transformation. The second effect is more subtle: by the convolution theorem, multiplying the interferogram by a sinc function results in convolving the spectrum with a rectangular function of width $\overline{\nu}_0 \Omega_s / 2\pi$, which reduces the resolution of the instrument. This effect of the finite FOV is referred to as *self-apodization*. In practice, the width of the imposed sinc function (up to its first zero) is often broader than the maximum optical path difference achievable with the spectrometer; however, when this is not the case, e.g. at very high wavenumbers or very low throughputs that require a large entrance aperture, care must be taken not to record the interferogram past the point where the sinc envelope first drops to zero, as this reduces the interferogram signal while the noise increases [*Davis et al.*, 2001]. A common method used in practice is to record the interferogram so as to achieve the maximum fringe amplitude for the desired path difference (resolution) and maximum wavenumber of interest. This is achieved by setting the argument of the sine function in the sinc envelope to $\pi/2$, which yields the following expression for choosing the diameter of the optimum aperture, d:

$$d = 2f_c \sqrt{\frac{2}{\overline{\nu}_0 \ L}} = f_c \sqrt{\frac{8}{R}} \tag{2.44}$$

where R is the (wavenumber-dependent) resolving power of the FTS, defined as $R = \overline{\nu}_0 / \Delta\overline{\nu} = \overline{\nu}_0 \ L$ [*Davis et al.*, 2001].

2.2.3.3 Interferogram sampling and the discrete Fourier transform

The Discrete Fourier Transform (DFT) is invoked when the interferogram is sampled at N points separated by distance $\Delta\delta$ such that $N\Delta\delta = L$ (the maximum OPD). It then follows that $(N\Delta\delta)^{-1} = \Delta\overline{\nu}$, or alternatively that $\Delta\delta \ \Delta\overline{\nu} = 1/N$, and we can write the DFT pair as

$$F(k\Delta\overline{\nu}) = \Delta\delta \sum_{j=0}^{N-1} f(j\Delta\delta) e^{-i2\pi kj/N} \tag{2.45}$$

$$f(j\Delta\delta) = \Delta\overline{\nu} \sum_{k=0}^{N-1} F(k\Delta\overline{\nu}) e^{i2\pi kj/N}. \tag{2.46}$$

The properties of the discrete Fourier transform are equivalent to those of the continuous Fourier transform and this subject is treated extensively in many texts, e.g. *Bracewell* [1999]. In practice, the interferogram is padded with extra zeros until it is of length 2^m (a process known as *zero-filling*) so that it can be transformed using the Fast Fourier Transform (FFT) algorithm of *Cooley and Tukey* [1965]. Adding trailing zeros to the interferogram does not introduce more information into the spectrum because the additional points in the resulting spectrum are only interpolations between previously existing points. Nonetheless, it is important to be aware of the true instrument resolution dictated by $1/L$ and an apparently higher resolution introduced by the zero-filling process. In another post-processing step, the size of operational interferograms is reduced in practice by digital filtering techniques which re-sample the recorded interferogram at every nth point, while still respecting the Nyquist limit for the maximum frequency in the spectral band.

Until recently, all modern FTSs used the signal from an auxiliary interferogram produced by a co-located HeNe laser beam (15,800 cm^{-1} or 633 nm) to sample the source interferogram at regular intervals $\Delta\delta$. If the source interferogram is sampled once at each maximum of the HeNe laser interferogram, i.e. at 15,800 samples per cm, then the maximum detectable frequency is 7900 cm^{-1}, by the Nyquist Sampling Theorem. This is more than sufficient for mid-IR spectroscopy, where spectra are typically recorded in $<$ 1000 cm^{-1} bands between 650 - 4300 cm^{-1}. (To record spectra above 7900 cm^{-1}, the source interferogram has to be sampled more often (as is possible for the TAO FTS), e.g. twice per laser wavelength leads to a maximum detectable frequency of 15 800 cm^{-1}.) However, in a method developed by *Brault* [1996] known as the *sigma-delta* approach, the interferogram and the laser is oversampled in the time domain (at regular intervals of $\sim$ 50 kHz) and is later interpolated to regular intervals of OPD. The intervals can be made smaller than the laser fringe wavelength (known as *fringe subdivision*), thus increasing the free spectral range of the FTS. Additionally, the technique gives information about instantaneous velocity variations in the drive mechanism, which allows one to correct for any frequency dependent phase and amplitude response functions in the digital filter and signal electronics.

2.2.4 General error considerations

Systematic errors (and random noise) in the interferogram can be grouped into three broad categories: additive errors (ϵ_1), intensity errors (ϵ_2) and phase errors (ϵ_3). Following *Remedios* [1990], in the presence of all three error types, we write the interferogram as

$$\begin{aligned} I(\delta) &= [1+\epsilon_2(\delta)] \int_{-\infty}^{+\infty} B(\overline{\nu})\ \ exp[i2\pi\overline{\nu}(\delta+\epsilon_3(\overline{\nu}))]\ \ d\overline{\nu}+\epsilon_1 && (2.47)\\ &= [1+\epsilon_2(\delta)] \int_{-\infty}^{+\infty} B(\overline{\nu})\ \ exp[i2\pi\overline{\nu}\delta]\ exp[i\phi(\overline{\nu})]\ d\overline{\nu}+\epsilon_1, && (2.48)\end{aligned}$$

where $\phi(\overline{\nu}) = 2\pi\overline{\nu}\epsilon_3(\overline{\nu})$.

Constant additive errors in the interferogram manifest themselves as signal at zero wavenumber in the spectrum, whereas periodic additive errors manifest themselves as signal at the corresponding wavenumber. Bursts of signal in the interferogram due to multiple reflections within optical elements (known as *channeling*) result in spurious periodic signals in the spectrum. Channeling is eliminated with the use of wedged window and filter elements.

Intensity errors are real multiplicative factors which are symmetric about the Zero Path Difference (ZPD) location and which give rise to symmetric lineshape distortions in the spectrum, e.g. the fundamental limitations of finite mirror travel and finite source extent (self-apodization) lead to symmetric lineshape distortions. We will discuss additive and intensity (also referred to as *multiplicative*) errors again in the next section.

Phase errors multiply the true (purely real) spectrum by the complex factor $exp(i\phi(\overline{\nu}))$ in Equation 2.48, which implies an asymmetry (or *oddness*) has been introduced into the true and purely symmetric (or *even*) interferogram. This can be understood in terms of the mathematical property which states that any real function can be written as the sum of an even and an odd component. We also employ the fact that the complex Fourier transform of an even function is purely real and that of an odd function is purely imaginary. It then follows that the Fourier transform of an even function with a small odd component, i.e. an interferogram with a small degree of asymmetry, will have both a real and a small imaginary component, i.e. the true spectrum multiplied by $exp(i\phi(\overline{\nu}))$. The basic choice of recording a single-sided rather than a double-sided interferogram fundamentally destroys the interferogram's symmetry, thus introducing a phase error. Furthermore, asymmetries can be introduced by 1) digital sampling errors, i.e. usually there is no sampling location coincident with the ZPD location, 2) instrument optics, i.e. due to a mismatch in the beamsplitter compensator or dispersion in any of the filters or windows, and 3) instrument electronics, i.e. due to dispersion in electronic amplifier and filter circuits. The dispersive effects introduce asymmetries in the interferogram by causing frequency-dependent delays in the beam [*Davis et al.*, 2001].

2.2.4.1 Phase correction

Fortunately, phase errors do not degrade the information content of the interferogram, and they can be corrected for in a relatively straightforward manner. The complex spectrum of Equation 2.48 can be written as

$$B'(\overline{\nu}) = B(\overline{\nu})\ exp(i\phi(\overline{\nu})) = B(\overline{\nu})[cos(\phi(\overline{\nu})) + isin(\phi(\overline{\nu}))] = B'_R(\overline{\nu}) + iB'_I(\overline{\nu}). \qquad (2.49)$$

One could recover a real spectrum of the same magnitude as $B(\overline{\nu})$ by simply taking the modulus of the complex spectrum, $B'(\overline{\nu})$; however, this operation reduces the signal-to-noise ratio of the real spectrum because both the real and the imaginary components of $B'(\overline{\nu})$ contain an equivalent amount of noise, which is carried into the real spectrum by the modulus operation [*Davis et al.*, 2001]. Therefore, to recover the true spectrum, we find the phase angle $\phi(\overline{\nu})$, given by

$$\phi(\overline{\nu}) = arctan\left[\frac{B'_I(\overline{\nu})}{B'_R(\overline{\nu})}\right], \tag{2.50}$$

and multiply the complex spectrum by $exp(-i\phi(\overline{\nu}))$ (here implicitly assuming that intensity errors, $\epsilon_2(\overline{\nu})$), are negligible). In practice, this is complicated by the fact that we need a full-resolution double-sided interferogram to do this, and we are limited to recording single-sided interferograms in order to maximize the resolution for a given scanning range of the FTS. If we assume that the phase is stationary and a smooth function of wavenumber, then we can calculate its spectrum from a low-resolution double-sided interferogram and interpolate it to the full resolution of the FTS. Finally, we multiply the complex amplitude spectrum derived from a high-resolution single-sided interferogram by the negative of the interpolated phase spectrum in order to recover the true (purely real) amplitude spectrum at high resolution. Indeed, if the calculated phase spectrum was correct then the imaginary components of the true amplitude spectrum will be vanishingly small, providing a check of the method. This is the essence of the Mertz method of phase correction, whereby an asymmetric interferogram (high-resolution with a short component on the other side of ZPD) is recorded and the phase spectrum is derived from the symmetric low-resolution component of this uneven high-resolution interferogram [*Mertz*, 1965, 1967]. Instead of using only the single-sided portion of the interferogram in the complex Fourier transform, the Mertz method uses the entire asymmetric interferogram. The double-sided low-resolution region is multiplied by a symmetric weighting function in order to prevent it being counted twice in the transformation process.

In the mathematically equivalent Forman method [*Forman et al.*, 1966], the phase correction is performed in the interferogram domain. Namely, the interferogram of the phase spectrum is found by taking the inverse Fourier transform of $exp(i\phi(\overline{\nu}))$. Next, it is convolved with the recorded signal to give a phase-corrected interferogram (since the convolution operation can be performed on two series of unequal length, no explicit interpolation is necessary). Finally, the short low-resolution component is discarded, and the newly symmetrized one-sided interferogram is transformed using a cosine Fourier transformation to yield the real spectrum $B(\overline{\nu})$. The Forman method of phase correction is more accurate than the Mertz method because it does not use a symmetric weighting function; however, the Mertz method is much more computationally efficient because it does not use the costly convolution operation. An excellent description of the Mertz and Forman phase correction methods, including “a pictorial essay”, is found in *Griffiths and de Haseth* [1986].

2.2.5 Signal-to-noise ratio considerations

In what follows, we describe the two classic advantages of an FTS that give it a Signal-to-Noise Ratio (SNR) that is superior to, e.g. a grating spectrometer. Furthermore, we elaborate on the types of noise present in both the interferogram and spectral domains, and on the measurement parameter trade-offs used in FTIR spectroscopy in order to maximize SNR levels. Further details of the Fellgett and Jacquinot advantages are found in *Bell* [1972, Ch. 2], while a detailed discussion of noise is found in *Griffiths and de Haseth* [1986, Ch. 7], *Chamberlain* [1979, Ch. 8, 9] and *Davis et al.* [2001, Ch. 8].

2.2.5.1 Fellgett (multiplex) advantage

Historically, the Fellgett advantage [*Fellgett*, 1958] of a Fourier transform spectrometer pertained to the fact that, unlike a prism spectrometer or an older grating spectrometer, the Michelson interferometer receives information from its entire spectral range during *each* time element of a scan. Thus, if random noise independent of the signal level is assumed then the signal-to-noise ratio at the detector is proportional to the square root of the observation time, T:

$$SNR_{FTS} \propto T^{1/2}. \tag{2.51}$$

The situation was different for the older (scanning) grating spectrometer, which received information only on the narrow width of the spectrum that lay on the exit slit of the instrument at that particular time element of the scan. If there were M narrow spectral elements then the observation time for each one of them was T/M and the signal-to-noise ratio at the detector for a grating spectrometer was:

$$SNR_G \propto (T/M)^{1/2}. \tag{2.52}$$

The signal-to-noise-ratio of the FTS compared to the older grating spectrometer was thus proportional to $M^{1/2}$:

$$\frac{SNR_{FTS}}{SNR_G} \propto M^{1/2}. \tag{2.53}$$

Since modern grating spectrometers have multichannel detectors (linear diode arrays or charge-coupled devices), this advantage is now lost. The Fellgett advantage is also lost for an FTS in the visible spectral region where photon noise (see Section 2.2.5.3) dominates over detector noise, and hence is no longer independent of the signal level. When the noise is proportional to the square root of the signal intensity, the Fellgett advantage is precisely balanced by the increased signal noise (still taking the form of photon noise), and the signal-to-noise ratio for both an FTS and a grating spectrometer is

$$SNR \propto \sqrt{(T/M)\ I(\overline{\nu})}. \tag{2.54}$$

The above discussion is true strictly in the absence of multiplicative errors introduced by an unstable source. When source noise dominates (see Section 2.2.5.3), observing each spectral element simultaneously gives the FTS a multiplex *disadvantage* over a grating spectrometer proportional to $M^{-1/2}$.

2.2.5.2 Jacquinot (throughput or étendue) advantage

The Jacquinot advantage derives from the fact that in an FTS it is possible to have large throughput values (Equation 2.39) without a loss of resolution. This is achieved by increasing the area and the focal length of the collimating optics. *Bell* [1972] considers the throughput of a grating spectrometer, E_G, and that of a Michelson interferometer, E_{FTS}, and shows that the ratio of the two for the same measurement resolution is approximately equal to

$$\frac{E_{FTS}}{E_G} \simeq 2\pi \frac{f}{l}, \tag{2.55}$$

where f is the FTS collimating mirror focal length and l is the grating spectrometer slit length. (The slit length of a grating spectrometer determines the FOV while the narrower slit width determines the resolution.) At the time of Bell's book, this ratio was ∼200 for the best grating spectrometers, and it is still high today, e.g.∼97 for a 21-mm-slit spectrometer available commercially and a 324-mm-focal-length Bomem FTS, or ∼125 for a 418-mm-focal-length Bruker FTS. This result is consistent with the fact that, owing to its circular symmetry, the Michelson interferometer can collect larger amounts of energy than a grating spectrometer, for a given resolution. Unlike the Fellgett advantage, the Jacquinot advantage is never lost. Furthermore, the Jacquinot advantage is so great that in some applications it can outweigh the Fellgett disadvantage due to an unstable source.

2.2.5.3 Specific types of noise

Random and systematic noise enter the spectrum through disturbances in the signal source, the FTS scanning and alignment mechanism, the detector and its preamplifier, the Analog-to-Digital Converter (ADC), and finally, the Fourier transform and digital filtering operations. Historically, FTSs were slowly scanned (with a stepper motor, for example) and the noise at each scan location was suppressed with electronic filters having a time constant $\tau = RC$. Additionally, the signal being averaged was either amplitude or phase modulated, because of the better performance exhibited by AC amplifiers. In this configuration, the SNR in the interferogram depended on the number of coadded scans, but also on τ. Modern FTSs are all *rapid-scan* instruments that do not require amplitude or phase signal modulation, which results in higher signal throughput. The modulation is now performed by the scanning speed itself, e.g. energy at 1000 cm^{-1} results in 1000 signal fringes per centimeter of optical path difference and a 1 kHz AC signal when the scan speed is 1.0 cm/s. In a modern rapid-scanning FTS, the SNR is only determined by the number of scans coadded according to the simple theorem that for N independent measurements the SNR improvement is $\sqrt{N}$ over any single measurement.

Random noise that possesses a frequency-independent power spectrum is referred to as *white noise*, while random noise that has some frequency dependence is often termed *1/f noise*, although in practice f is often raised to a power slightly higher or lower than 1. As already alluded to in Equation 2.47, noise can be *additive* (independent of signal levels and simply added to the signal as is the case for detector noise), or *multiplicative* (having a dependence on the signal level). *Johnson noise*, due to the thermal noise in the preamplifier feedback resistor is an example of additive detector noise. The fundamental and unavoidable *photon noise*, which scales as the square root of the signal level, is a type of multiplicative noise, as is the very problematic *source noise* (e.g. due to clouds drifting in and out of the FTS FOV), which scales as the first power of the signal level and is to be avoided at all costs.

Photon noise arises from the random arrival of photons at the detector. (The term *shot noise* describes the similarly random nature of the subsequent emission of electrons by the detector element.) *Chamberlain* [1979] makes a distinction between two kinds of photon noise: *signal* photon noise and *background* photon noise. In solar absorption spectroscopy, background photons are thermal photons emitted by the FTS, and are therefore independent of the signal level. Historically, FTS noise was limited by the detectors, and accordingly, the Fellgett advantage held (provided that the source intensity was stable). Modern infrared detectors are much more sensitive, and without limiting the FTS FOV and employing band pass filters and exit apertures, the FTS noise would be photon-limited, and hence the Fellgett advantage would be lost. The term Background Limited Infrared Photoconductor (BLIP) is commonly used to describe such sensitive infrared detectors (BLIP can also refer to Background Limited In Performance). Even with the noise-limiting modifications in place, at high enough photon energies (visible and UV) photon noise increases to the point where it becomes the dominant noise source and the Fellgett advantage is lost. (At higher wavenumbers, where the energy per photon is higher, the noise due to the arrival or non-arrival of a photon increases.)

The theoretical discussion of FTS resolution implies that a longer path difference is more advantageous. However, when considering the practical signal – with its inherent noise – one finds that there exists an optimal path difference beyond which the signal does not increase but the noise grows. To see how this optimum path difference arises we must consider monochromatic and broadband interferograms in turn (in the additive noise regime). A monochromatic interferogram is sinusoidal for all path differences, and hence its signal power increases in proportion to the observation time, T, however, its noise power increases only as $\sqrt{T}$. On the other hand, a broadband interferogram decays quickly within an exponential envelope, and hence its signal power reaches an asymptotic limit, however, its noise power continues to increase as $\sqrt{T}$. Therefore, the SNR of a broadband interferogram is proportional to $1/\sqrt{T}$ for large path differences, and scanning beyond the optimum path difference reduces the overall interferogram SNR. Thus, beyond a certain path difference, one can only improve the SNR by increasing the observation time through coadding several interferograms. Another practical reason not to scan beyond a certain path difference is to avoid regions where the broadband interferogram oscillations drop below the lowest ADC bit. These digitization errors are reduced by employing narrow band filters, which reduce the interferogram centerburst at ZPD.

In summary, signal is proportional to source intensity, bandwidth and observation time, while noise is proportional to the inverse square root of bandwidth (more signal photons means less noise) and the square root of observation time. How noise relates to source intensity depends on the dominant source of noise in the system. A useful formulation adopted by *Davis et al.* [2001] is to write the noise as being proportional to I^k, where I is the source intensity, k=0 for additive noise, k=0.5 for multiplicative photon noise, and k=1.0 for multiplicative source noise.

Finally, the Root-Mean-Square (RMS) interferogram noise, ϵ_δ, is independent of the path difference in the spectrometer and the interferogram SNR can be estimated from

$$SNR_\delta = \frac{I(0)}{\epsilon_\delta}, \tag{2.56}$$

where $I(0)$ is the signal intensity at the interferogram centerburst. By the Fourier transform equal area theorem, $I(0) = \Delta\overline{\nu}\ N\ \overline{(B)}$, where N now refers to the number of points in the interferogram and $\overline{B}$ is the mean signal strength in the spectral band. Using this relation, *Davis et al.* [2001] show that the interferogram SNR is related to the spectrum SNR by

$$SNR_{\overline{\nu}} = \frac{B(\overline{\nu})}{\epsilon_{\overline{\nu}}} = \sqrt{\frac{2}{N}}\frac{B(\overline{\nu})}{\overline{B}}SNR_\delta, \tag{2.57}$$

where $\epsilon_{\overline{\nu}}$ is the uniformly distributed spectrum noise. Equation 2.57 is consistent with our previous discussion of an optimal resolution for a given signal bandwidth and SNR. Furthermore, it illustrates that the noise in the spectrum is proportional to the mean signal strength in the spectral band being observed, which places greater demands on interferogram quality in solar absorption and broadband spectra as opposed to emission and narrowband spectra.

2.2.5.4 SNR trade-off rules in FTIR spectroscopy

The above discussion already illustrates some aspects of how the SNR of an FTS depends on several parameters that must be traded off against one another. For example, we must balance the efficiency of recording broadband spectra that contain the spectral features of many gases against the dynamic range of our ADC and also against the optimal resolution dictated by the presence of additive noise. According to *Griffiths and de Haseth* [1986], for a given SNR, the observation time of a broadband source must be quadrupled if the maximum path difference is doubled.

In principle, the SNR of a rapid-scanning FTS operating at a constant velocity will increase as $\sqrt{T}$ or as $\sqrt{N}$, where N is the number of coadded scans. However, in practice the observation time must be limited to ~20 minutes or ~4 scans in order to avoid the serious consequences of a changing source intensity and a smearing of the atmospheric path due to a changing SZA. Another way to increase N is to increase the mirror scan speed (and hence modulation frequency), however, this option is of limited use above scan speeds of a few cm/s because it leads to scanning mirror instabilities that cause sampling and alignment errors. Another reason to increase the scan speed is that the specific detectivity (D^*) of InSb and MCT detectors (described in more detail below) increases with increasing modulation frequency. This occurs because the modulated signal frequencies are moved farther away from the source and detector noise frequencies, which exhibit $1/f$ characteristics, and because D^* is proportional to detector area (A), noise bandwith (Δf), and Noise Equivalent Power (NEP) as follows: $D^* = \sqrt{A\,\Delta f}/NEP$.

2.3 The TAO Facility and Bomem DA8 FTIR Spectrometer

TAO was established to acquire high-quality long-term measurements of trace gases for the purpose of understanding chemical and dynamical processes in the atmosphere and to validate models and satellite measurements of atmospheric constituents. The geographic position of TAO (43.66°N, 79.40°W) makes it well suited for measurements of mid-latitude stratospheric ozone, related species and greenhouse gases, while its urban setting enables measurements of tropospheric pollutants. The primary instrument at TAO is a high-resolution DA8 Research Grade Fourier Transform Spectrometer (FTS) manufactured by ABB Bomem Inc. (Québec, Canada). The FTS is complemented by a commercially available weather station (Vantage Pro Plus manufactured by Davis Instruments Corp.) that records local meteorological variables, UV radiation and solar irradiance, and (on a campaign basis) by a Differential Optical Absorption Spectrometer (DOAS) [*Bassford et al.*, 2001, 2005], and a SunPhotoSpectrometer (SPS) [*McElroy*, 1995]. A Meteorological Service of Canada (MSC) Brewer spectrometer was installed in March 2005, while a ground-level ozone monitor was installed in September 2005.

The FTS and its dedicated suntracker are located 174 m above sea level in downtown Toronto. The altitude-azimuth suntracker (manufactured by AIM Controls Inc., California, USA) is housed in an electronically controlled weatherproof enclosure and provides continuous active solar tracking throughout the day (passive tracking is also available). Figure 2.5 shows a schematic of the FTS and suntracker installation. To characterize the FTS ILS a dedicated blackbody (IR-563/301 manufactured by Infrared Systems Development Corp.) has also been installed at TAO and is coupled to the FTS through the same optics as the suntracker plus one additional removable 45° mirror (Figure 2.5).

FTIR solar absorption spectra are recorded under clear sky conditions, allowing for approximately 80 observation days per year in the first three years of operation. The measurements are semi-automated, involving an operator at start up and shut down to engage the suntracker, cool the detectors with liquid nitrogen, and initiate an automatic measurement sequence. The temporal coverage of the dataset is more sparse in the winter months due to decreased daylight hours and increased cloud cover.

2.3.1 FTS design, configuration and measurement parameters

The DA8 FTS is a modified Michelson interferometer (Figure 2.6) with a maximum optical path difference of 250 cm, providing a maximum apodized resolution of 0.004 cm^{-1}, equivalent to an unapodized resolution of 0.0026 cm^{-1}. Until recently, a dynamic alignment approach was used, whereby the position of the fixed mirror of the interferometer was adjusted to compensate for the deviations from alignment of the moving mirror; thus, the fixed mirror was not truly fixed. In the new design employed in the TAO DA8, the fixed mirror is permanently mounted and an adjustable flat mirror that folds the beam is added to the moving mirror arm of the interferometer (Figure 2.6). It is this

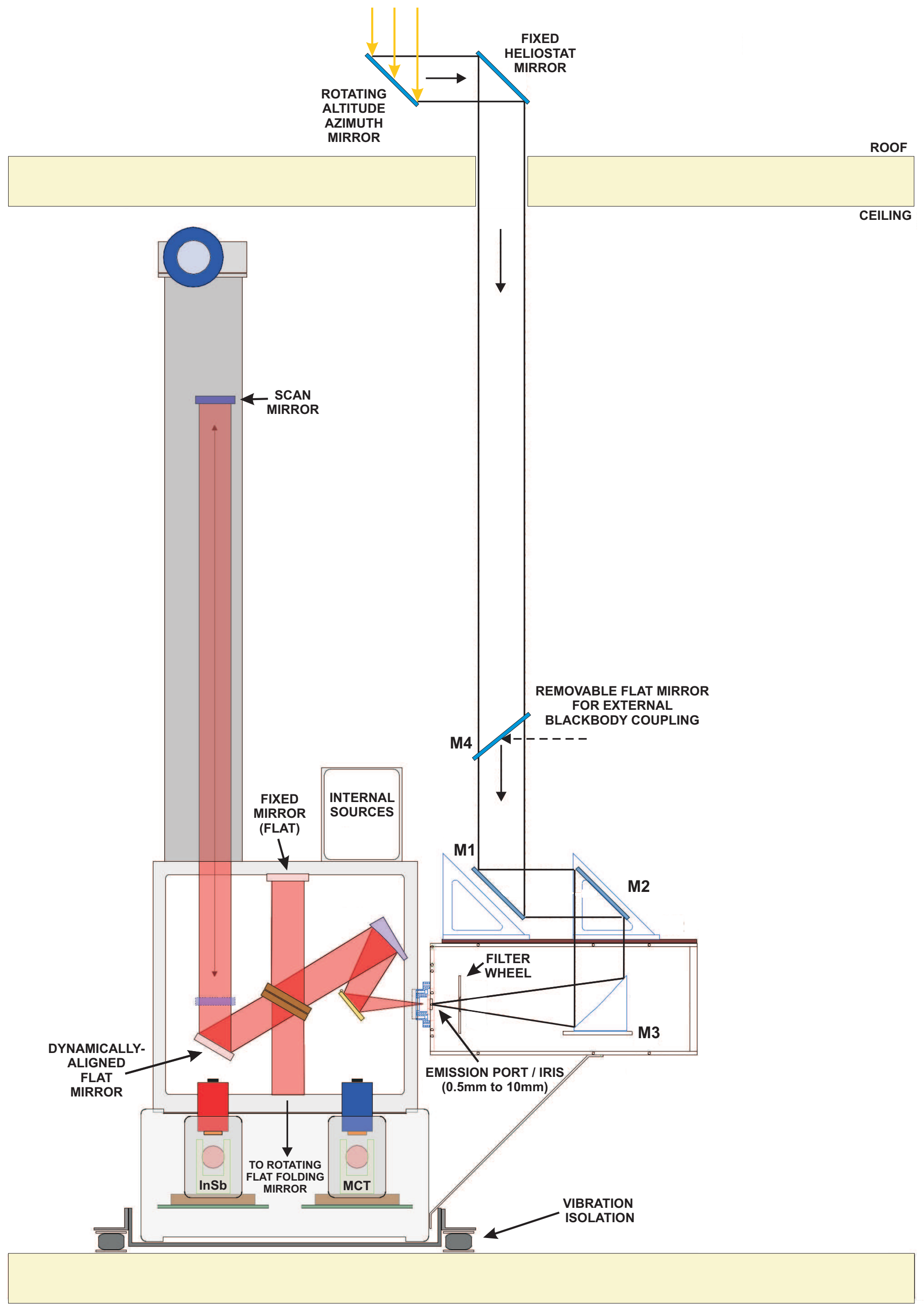


Figure 2.5 FTS and suntracker installation at TAO (adapted from ABB Bomem Inc. facility schematics). M1, M2 and M3 serve to couple the heliostat optics to the DA8 FTS. A removable 45° mirror (M4) couples the FTS to a blackbody for instrumental line shape testing.

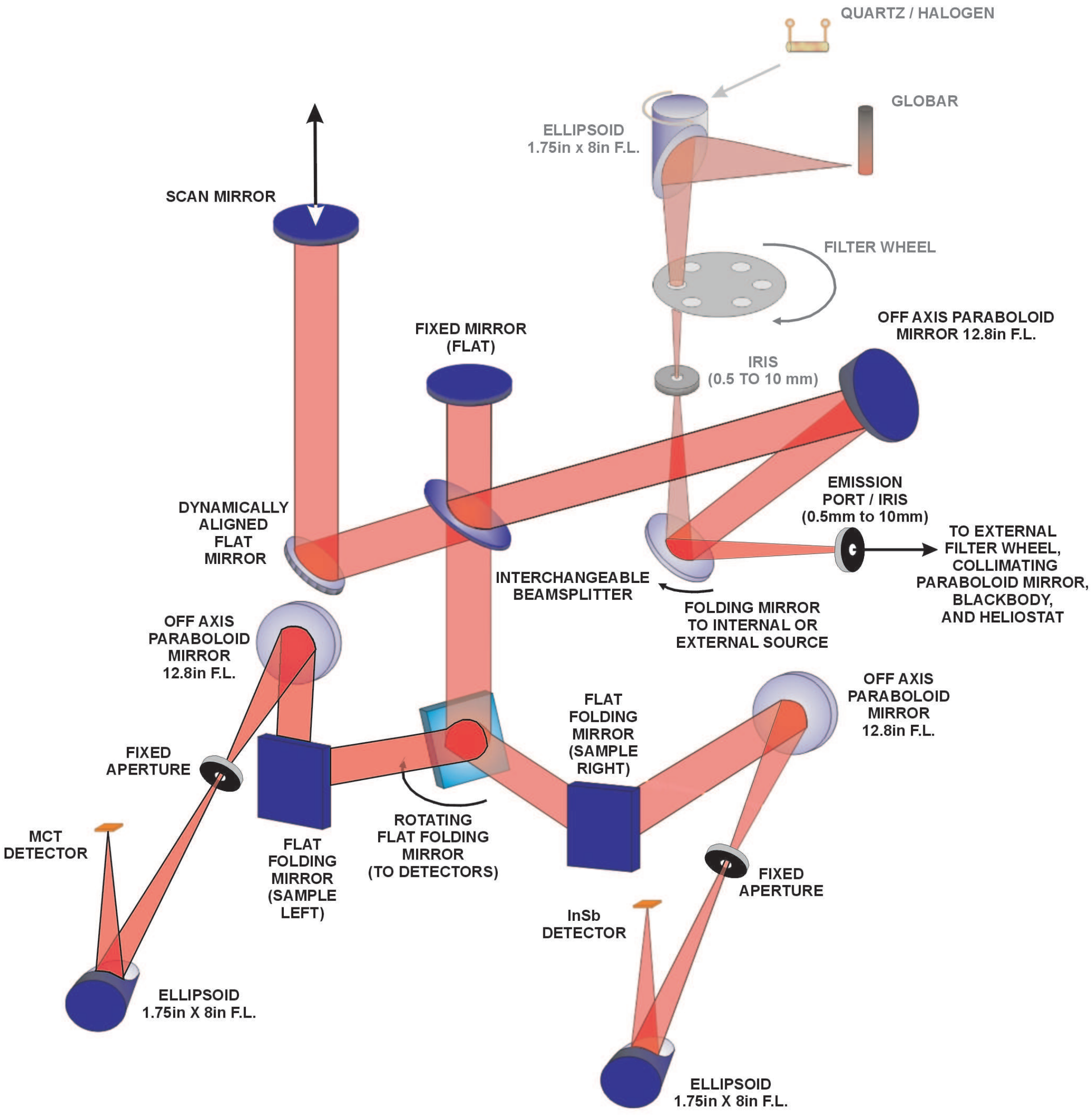


Figure 2.6 The DA8 FTS is a modified Michelson design (see text for details). A moving folding mirror selects between the emission port used for solar absorption measurements and blackbody line shape testing, and internal sources (shown in grey) used for some instrument alignment. (Adapted from ABB Bomem Inc. instrument schematics.)

NDACC Filter	Wavenumber Range (cm^{-1})	Wavelength Range (μm)	Trace Gases With Absorption Features In This Region
1	4000-4300	2.3-2.5	**HF**, CO ...
2	2900-3500	2.6-3.3	$\mathbf{H_2O}$, C_2H_2, HCN ...
3	2400-3100	3.3-4.1	$\mathbf{O_3}$, **HCl**, $\mathbf{N_2O}$, $\mathbf{CH_4}$, $\mathbf{NO_2}$, C_2H_6 H_2O, CO_2 ...
4	2000-2700	3.9-5.0	CO, OCS, N_2 ...
5	1500-2200	4.7-6.3	CO, **NO**, COF_2 ...
6	750-1350	7.4-14	O_3, $\mathbf{ClONO_2}$, $\mathbf{HNO_3}$, N_2O, CH_4, C_2H_2 C_2H_4, CFCs ...

Table 2.1 Spectral coverage of the NDACC filter set (NDACC targets shown in bold).

folding mirror that now provides dynamic alignment and compensates for any wobble of the moving mirror. At the same time, the dynamic alignment circuitry and actuation are kept stationary at the folding mirror. This arrangement results in a fixed optical axis through the beamsplitter (and a fixed focal point on the detector) as well as more stable modulation efficiency, which leads to an improved instrumental line shape and stability.

The FTS is currently equipped with Potassium Bromide (KBr) and Calcium Fluoride (CaF_2) beamsplitters, and Indium Antimonide (InSb) and Mercury Cadmium Telluride (HgCdTe or MCT) detectors, for nominal coverage of the spectral range from 750 to 8500 cm^{-1}. The system is also equipped with six narrow-band optical interference filters that are widely used in the NDACC InfraRed Working Group (IRWG). Table 2.1 summarizes trace gases commonly measured with this filter set, both detectors and the KBr beamsplitter. This FTS configuration (operating under vacuum) is used during regular solar absorption measurements. All spectra are recorded with boxcar apodization, nominally using the settings summarized in Table 2.2. The built-in globar and quartz halogen sources (Figure 2.6) as well as an external collimated blackbody are used during instrument characterization (described in the next section).

	Resolution (cm^{-1})	Scans Co-added[a]	Full FOV (mrad)[b]	Detector
Filters 1-5	0.004	4	1.54	InSb
Filter 6	0.005	5	4.63	MCT

Table 2.2 Measurement parameters used with the TAO FTS.

---

[a] Co-adding time is approximately 20 minutes in both cases to limit SZA changes.

[b] Defined as input aperture diameter divided by focal length of input optics (324 mm).

The InSb and MCT detectors are both 1.0 mm^2 photoconductive semiconductors that exhibit a half-maximum response between 1900-4500 cm^{-1} and 720-2200 cm^{-1}, respectively. Here the term *photoconductive* is used in the general sense, i.e. describing a material that releases charge carriers when photons are incident on it. The MCT detector is a bulk photoconductor, which releases electrons from the valence shell to the conduction band in response to incident IR radiation, while the InSb detector is a p-n junction, in which electron-hole charge pairs form in response to the IR radiation. Because of these different charge-carrier-forming mechanisms, the MCT detector is operated under a reverse bias in the preamplifying circuit and thus it creates a current, whereas the InSb detector is operated without a bias since it creates a voltage. This is why the former is referred to as *photoconductive* and the latter as *photovoltaic*. (Note the specific meaning of the term *photoconductive* in this case.) Prior to Nov. 2002, the FTS was operated with a 0.0625 mm^2 (0.25 mm x 0.25 mm) MCT detector element, which was suboptimal for observations of $\mathrm{ClONO_2}$ at 780 cm^{-1} with its higher 50% cut-on frequency of 790 cm^{-1}. (It was chosen because a detector's noise power is proportional to its area and bandwidth, thus giving the smaller detector element an SNR advantage.) Finally, MCT detectors experience increasing $1/f$ noise below the corner frequency (i.e. half power, or -3dB point) of (typically) 1kHz. Therefore, increasing the scan speed currently used for Filter 6 measurements at TAO (0.5 cm/s) may lead to SNR improvements in this spectral region.

2.3.2 FTS characterization

The alignment and performance of an FTS is characterized by its ILS, e.g. *Park* [1982] or *Goorvitch* [2000]. Accurate knowledge of the ILS is necessary in order to retrieve information on the vertical distribution of trace gases from FTIR spectra, which is contained in the spectroscopic absorption line shapes, e.g. *Schneider et al.* [2005a]. *Bernardo and Griffith* [2005] showed that accurate knowledge of the ILS can reduce errors in the retrieved absorber amount by as much as 12%. For these reasons the ILS of the TAO FTS has been monitored on a regular basis by means of calibrated low-pressure gas cell measurements. The candidate designed and built a gas filling station and performed qualitative ILS measurements with N_2O and HBr cells during her M.Sc. studies and the first year of her Ph.D. studies; however, quantitative results presented in this section are courtesy of Ph.D. candidate Jeffrey R. Taylor.

The use of N_2O and HBr cell measurements to determine the ILS of an FTS with a newly developed algorithm is described by *Hase et al.* [1999]. We apply the latest version of that retrieval software (LINEFIT v.9) to characterize the phase error and modulation efficiency of our instrument as a function of OPD from high SNR transmission spectra of N_2O and HBr absorption features. (HBr cell measurements can additionally be used to determine the ILS of an FTS directly while solar absorption spectra are being recorded since the atmosphere has no appreciable quantities of HBr [*Coffey et al.*, 1998].) For an ideally aligned FTS with perfect optics and an infinitesimal FOV the phase error is zero and the modulation efficiency is one at all path differences in the interferometer.

At TAO, low-pressure gas cell measurements can be made using a calibrated, 1050°C blackbody radiation source optically coupled to the emission port of the spectrometer, the internal globar source, or the sun (Figure 2.6). The gas cell is placed in the focused beam of radiation in the evacuated sample compartment (below the beamsplitter and immediately before the InSb detector) in order to limit cell temperature variations. Typically a spectrum consisting of 50 co-added scans is recorded at maximum resolution (0.004

cm^{-1}) with all other measurement settings matching those used for TAO atmospheric observations. Transmission spectra are generated by taking the ratio of a full cell spectrum to an empty beam spectrum for the HBr cell and an empty cell spectrum for the N_2O cell, thus eliminating systematic features associated with water vapour absorption and detector baseline nonlinearity. Selected microwindows in the transmission spectra, each containing a single HBr or N_2O absorption feature, are simultaneously fitted with the LINEFIT algorithm; the modulation efficiency and phase error as a function of OPD are adjusted until the spectral fit residual is minimized.

Low-pressure gas cell experiments are performed every three to four months at TAO and recent results (June 2005, courtesy of Ph.D. candidate Jeffrey R. Taylor), including spectral fits, retrieved ILS, modulation efficiency, and phase error are shown in Figure 2.7. Panels (c) and (d) show results obtained with six different HBr cells distributed among the NDACC community and brought to Toronto for the June2005 IRWG Meeting. The 2-cm-long sealed glass cells with 2.5-cm-diameter fused sapphire windows were created from the same gas distribution manifold in 2001 so as to contain ∼2 mbar of HBr [*Coffey et al.*, 2001]. The initial and final ambient temperature of the cells was monitored with a thermocouple inside the sample compartment, and the mean of the two values was used in the LINEFIT retrievals, with the largest observed difference for any given spectrum being 0.9 °C. The HBr spectra were recorded at 0.004 cm^{-1} resolution using the internal globar source and co-adding 50 scans over a period of 4.5 hours per spectrum. Background spectra consisted of 100 co-added scans. The RMS noise in the spectral fits is well below 1% and comparable to values reported by *Hase et al.* [1999]. The phase error is well below 0.1 radians for all optical path differences in the FTS and results in the small asymmetry visible in the ILS. The modulation efficiency drops to approximately 30% at high optical path differences in the scan and implies a reduction in the theoretical resolution (0.004 cm^{-1}) and a possible misalignment of the instrument.

Efforts to improve the alignment of the TAO FTS are ongoing, and the group is in-

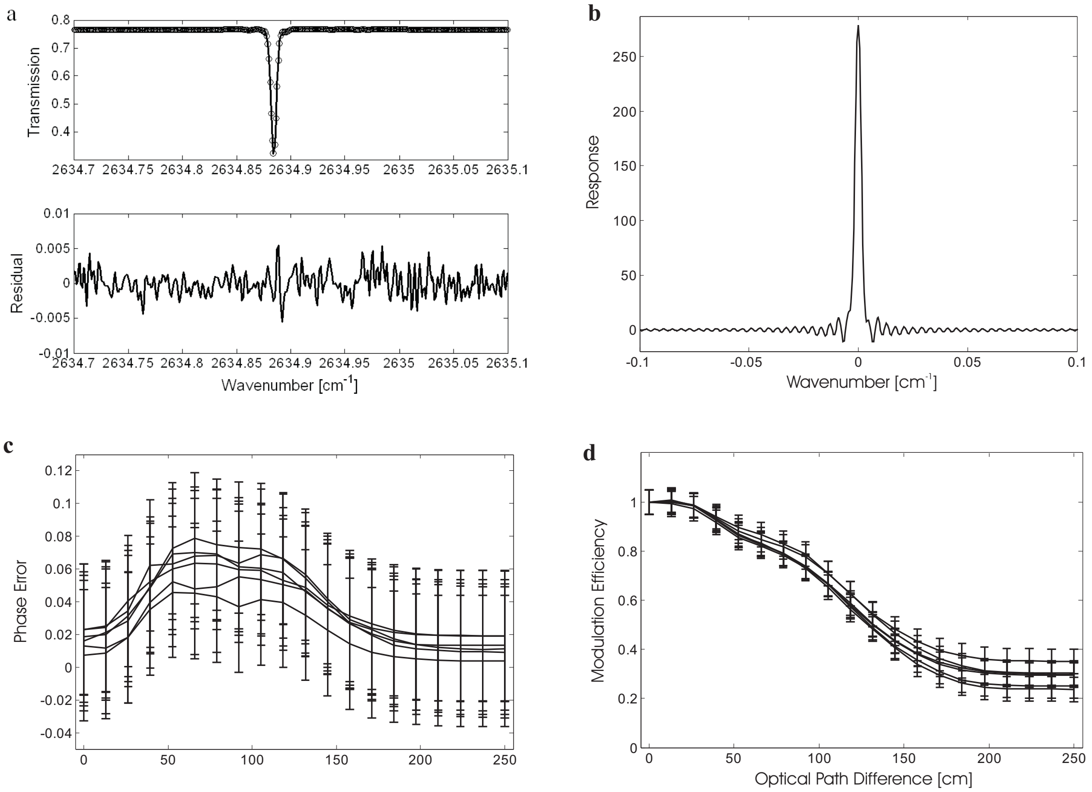


Figure 2.7 a. Observed (solid) and fitted (circles) HBr line (R4), with residual shown in black (the full fit microwindow is shown). b. Sample ILS retrieved by LINEFIT. c. Retrieved phase error from six different HBr cells (see text for details). d. Retrieved modulation efficiency from six different HBr cells.

vestigating the full effect of using the LINEFIT ILS parameters on retrieved trace gas profiles and the information content of the measurements. With a reduced modulation efficiency, a given atmospheric absorption feature will appear broader than it truly is, and the retrieved profile for the corresponding gas will shift to lower altitudes, where the Lorentz-broadening of spectral lines is greater. *Schneider et al.* [2005b] describe a calculated ascent of ∼115 m for a 2% increase in the modulation efficiency at maximum OPD. If we simply scale these numbers, the 70% reduction in Figure 2.7 (d) implies a ∼4 km descent in profiles retrieved at TAO without consideration of the ILS. This qualitative argument is illustrated in Figure 2.8, which shows O_3 profile retrievals performed without

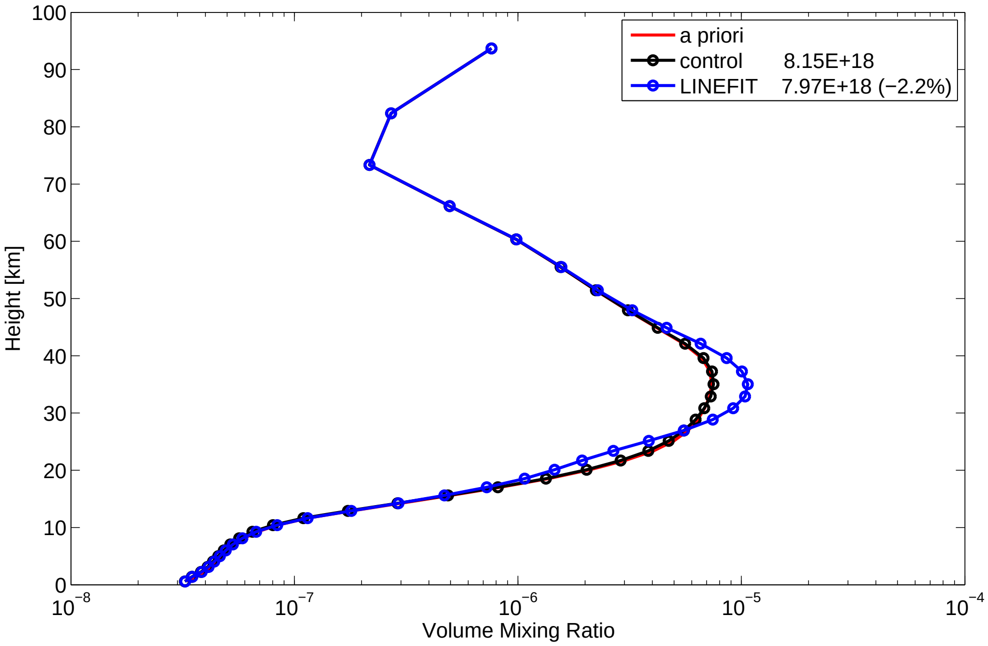


Figure 2.8 O_3 profile retrieved from spectrum 050902F3.S02 without and with the use of the LINEFIT ILS coefficients from Figure 2.7. The column amount [molec/cm^2] is reduced by 2.2% in the latter case.

and with the LINEFIT phase error and modulation efficiency coefficients for a spectrum recorded relatively close in time (Sept. 2005) to the ILS characterization measurements. When the LINEFIT coefficients are used in this retrieval, the total column amount is reduced by 2.2%, and the information content is reduced from 2.1 to 2.0 degrees of freedom for signal between 0-100 km. The maximum profile ascent is near the concentration peak of O_3 at ∼22 km, however, the maximum change in the retrieved VMR (∼50% increase, consistent with the calculations of *Schneider et al.* [2005b]) is near the VMR peak at ∼35 km. In general, the effects of an imperfect ILS become smaller for partial and total column retrievals, which primarily depend on the correct fitting of the area under the absorption curve, rather than fitting the details of the absorption line shape. Chapter 3 will now describe the details of the total column and vertical profile retrieval process used in this work, which does not apply the LINEFIT ILS coefficients from this point on.

Chapter 3

Trace Gas Retrievals

In the previous chapter we described how solar absorption spectra arise and how they are recorded with a Fourier transform infrared spectrometer. Furthermore, we described how the shape of individual spectral absorption features depends on pressure, which can be used to infer information about the vertical distribution of a trace gas. In this chapter we describe the subsequent analysis step known as the *retrieval*, wherein the most likely vertical distribution of a trace gas is found, subject to some *a priori* constraints. In the retrieval approach employed in the analysis of TAO spectra, the Optimal Estimation Method (OEM) formulation of *Rodgers* [2000] is used, whereby prior knowledge about the vertical distribution of a trace gas is combined with new information provided by the measured spectra in order to derive the new best estimate of the vertical distribution. The *a priori* and measurement noise covariance matrices provide the weighting assigned to each source of information; the former is estimated from knowledge of the *true* natural variability of a trace gas, while the latter is known from the measurement Signal-to-Noise Ratio (SNR). A *forward model* is used to connect the vertical distributions of trace gases, temperature and pressure with the observed spectra; it contains all of our knowledge of the measurement physics, i.e. of radiative transfer and FTS characteristics. Finally, the solution must be found iteratively due to the nonlinear nature of the forward model.

3.1 Inverse Theory

Inverse theory refers to the mathematical framework used in the inversion of complicated functions. In remote sounding of the atmosphere, the indirect quantities measured (e.g. radiances) must be inverted in order to derive the quantities that are desired (e.g. vertical trace gas or temperature or pressure distributions). Another example of a geophysical *inverse problem* is found in deriving profiles of the Earth's density from surface measurements of seismic wave arrival times. Whether the measurement is remote or not, in an inverse problem, we search for the best representation of the quantity of interest given some indirect measurements and any available prior information. We then characterize the information content and resolution of the retrieval and calculate how errors in our measurements and prior information propagate to the final solution, given the particular physics of the measurement.

The Optimal Estimation Method (OEM) as formulated by *Rodgers* [1976, 1990, 2000] is particularly well-suited to the under-determined problem of deriving a vertical profile of an atmospheric trace gas from a ground-based solar absorption spectrum. The problem is under-determined (*ill-posed*) in the trivial sense that we are seeking a continuous function from a finite set of measurements, however it remains under-determined even after we discretize the desired vertical profile. This occurs because the spectral absorption features contain only enough information to distinguish a few independent vertical layers, however, the atmosphere must be forward modeled on ∼40 layers in order to avoid gross errors in the simulated spectra due to the misrepresentation of trace gas volume mixing ratio, temperature, and pressure. The vertical resolution of ground-based spectra is fundamentally limited by observation geometry, maximum achievable spectral resolution and measurement SNR.

In the OEM formulation of the ground-based remote sounding problem, *a priori* knowledge, $\mathbf{x}_a$, of the n-dimensional *state vector*, $\mathbf{x}$ (vertical profile of the trace gas of interest) is combined with the m-dimensional *measurement vector*, $\mathbf{y}$ (high-resolution measurements of a spectral absorption feature) as a weighted mean. The (nonlinear) forward model $\mathbf{F}(\mathbf{x}, \mathbf{b})$ relates the state vector to the measurement vector, given some *model parameters*, $\mathbf{b}$

$$\mathbf{y} = \mathbf{F}(\mathbf{x}, \mathbf{b}) + \boldsymbol{\epsilon} = \mathbf{F}(\mathbf{x}_a, \mathbf{b}) + \mathbf{K}(\mathbf{x} - \mathbf{x}_a) + \boldsymbol{\epsilon} \tag{3.1}$$

where $\boldsymbol{\epsilon}$ is the random spectral measurement noise and $\mathbf{K} = \partial\mathbf{F}(\mathbf{x}, \mathbf{b})/\partial\mathbf{x}$, i.e. the linearization of $\mathbf{F}(\mathbf{x}, \mathbf{b})$ about some reference state, here taken to be $\mathbf{x}_a$. $K_{ij} = \partial F_i(\mathbf{x}, \mathbf{b})/\partial x_j$ is an element of an $m \cdot n$ matrix of derivatives of a forward model element with respect to a state vector element. $\mathbf{K}$ does not need to be square, and indeed, in the ground-based problem m is always greater than n, making the problem *mixed-determined*, i.e. both over-determined ($m > n$) and under-determined ($p < n$) at the same time. p is the *rank* of $\mathbf{K}$ and can be as large as n, however, as already alluded to, the information contained in ground-based spectra allows us to independently distinguish only a few (indeed only p) vertical layers, while the atmosphere must be represented by $n \approx 40$ state vector elements.

In algebraic terms, the rank of a matrix refers to the number of its linearly independent rows or columns, which form a *basis* for the *row space* or the *column space* of $\mathbf{K}$, respectively. When $p < n$, the row space of $\mathbf{K}$ does not span all of *state space*, and the column space of $\mathbf{K}$ (also of rank p) is also only a subspace of *measurement space.* The undetermined part of state space is called the *null space* of $\mathbf{K}$, as it contributes nothing to a measurement. According to *Rodgers* [2000, p. 17] *"If a retrieved state has components which lie in the null space, their values cannot have been obtained from the measurements,"* and therefore must have been obtained from the *a priori.*

We can determine the row and null space of $\mathbf{K}$ by performing a Singular Value Decomposition (SVD) of $\mathbf{K}$

$$\mathbf{K} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T \tag{3.2}$$

where the columns of $\mathbf{U}$ ($m \cdot p$) and $\mathbf{V}$ ($n \cdot p$) contain the left and right singular vectors of $\mathbf{K}$, respectively, and $\mathbf{\Lambda}(p \cdot p)$ is a diagonal matrix of the non-zero singular values of $\mathbf{K}$. The p columns of $\mathbf{V}$ form an orthogonal basis for a subspace of state space while the p columns of $\mathbf{U}$ form an orthogonal basis for the subspace of measurement space.

Having described some general aspects of inverse theory, and having posed the inverse problem encountered in ground-based remote sounding, we now turn to finding a specific solution for the state vector $\mathbf{x}$ from a weighted sum of prior knowledge about the state and new information introduced by the measurement, and using the error statistics of both sources of information as the weighting. We will consider linear problems first, followed by nonlinear problems in Section 3.1.2.

3.1.1 Linear problems

We define a *linear* problem as one that has a linear forward model. The present discussion also applies to *nearly linear* problems, i.e. problems in which the forward model can be linearized, and in which the linearization is then sufficient to find a solution without resorting to iterative techniques. Expanding Equation 3.1 to include a linearization about a set of model parameters $\mathbf{b}_a$ we write

$$\mathbf{y} - \mathbf{F}(\mathbf{x}_a, \mathbf{b}_a) = \frac{\partial \mathbf{F}}{\partial \mathbf{x}}(\mathbf{x} - \mathbf{x}_a) + \frac{\partial \mathbf{F}}{\partial \mathbf{b}}(\mathbf{b} - \mathbf{b}_a) + \boldsymbol{\epsilon} = \mathbf{K}(\mathbf{x} - \mathbf{x}_a) + \mathbf{K}_b(\mathbf{b} - \mathbf{b}_a) + \boldsymbol{\epsilon} \quad (3.3)$$

where $\mathbf{K}$ represents the sensitivity of the forward model to the true state and $\mathbf{K}_b$ represents the sensitivity of the forward model to its parameters. $\mathbf{K}$ is known by many names, including Jacobian, kernel, sensitivity kernel, tangent linear model, and *weighting function matrix*, which is a historical term that we will use.

Following *Rodgers* [1990], we also define an *inverse model*, $\mathbf{I}$, that operates on the measurements and makes use of the forward model parameters in order to produce the best estimate of the true state, $\hat{\mathbf{x}}$

$$\hat{\mathbf{x}} = \mathbf{I}(\mathbf{y}, \mathbf{b}). \quad (3.4)$$

We linearize this inverse model about the measurement expected given $\mathbf{x}_a$ – what Rodgers refers to as an *a priori* measurement, $\mathbf{y}_a$ – and the same set of model parameters

$$\hat{\mathbf{x}} - \mathbf{I}(\mathbf{y}_a, \mathbf{b}_a) = \frac{\partial \mathbf{I}}{\partial \mathbf{y}}(\mathbf{y} - \mathbf{y}_a) + \frac{\partial \mathbf{I}}{\partial \mathbf{b}}(\mathbf{b} - \mathbf{b}_a) \equiv \mathbf{G}(\mathbf{y} - \mathbf{y}_a) + \mathbf{G}_b(\mathbf{b} - \mathbf{b}_a). \tag{3.5}$$

The rows of the $n \cdot m$ *gain matrix*, $\mathbf{G}$, are commonly referred to as the *contribution functions* and represent the sensitivity of the retrieved state to the measurement. If we neglect the noise in the measurements and errors in the forward model parameters (to be discussed again in Section 3.3) we can write an expression for the retrieval *transfer function*, $\mathbf{T}$

$$\hat{\mathbf{x}} = \mathbf{I}(\mathbf{F}(\mathbf{x}, \mathbf{b}), \mathbf{b}) = \mathbf{T}(\mathbf{x}, \mathbf{b}), \tag{3.6}$$

which must also be linearized about the *a priori* state

$$\hat{\mathbf{x}} - \mathbf{T}(\mathbf{x}_a, \mathbf{b}) = \frac{\partial \mathbf{T}}{\partial \mathbf{x}}(\mathbf{x} - \mathbf{x}_a) = \frac{\partial \mathbf{I}}{\partial \mathbf{F}} \frac{\partial \mathbf{F}}{\partial \mathbf{x}}(\mathbf{x} - \mathbf{x}_a) = \mathbf{GK}(\mathbf{x} - \mathbf{x}_a). \tag{3.7}$$

If properly designed, a retrieval method will be unbiased and return the *a priori* when the input measurements are simulated from the *a priori* ($\mathbf{T}(\mathbf{x}_a, \mathbf{b}) = \mathbf{x}_a$), allowing us to finally write

$$\hat{\mathbf{x}} - \mathbf{x}_a = \mathbf{GK}(\mathbf{x} - \mathbf{x}_a) \equiv \mathbf{A}(\mathbf{x} - \mathbf{x}_a). \tag{3.8}$$

The rows of the $n \cdot n$ *averaging kernel matrix*, $\mathbf{A}$, give the sensitivity of the retrieved state to the true state, by definition,

$$\mathbf{A} = \mathbf{GK} = \frac{\partial \mathbf{I}}{\partial \mathbf{y}} \frac{\partial \mathbf{F}}{\partial \mathbf{x}} = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{x}}, \tag{3.9}$$

and are also used to calculate the sensitivity of the retrieval to the *a priori*, which follows from Equation 3.8:

$$\frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{x}_a} = \mathbf{I}_n - \mathbf{A}. \tag{3.10}$$

In order to find the form of $\mathbf{G}$ and the best estimate of the state, $\hat{\mathbf{x}}$, we must now consider the details of the error statistics of the *a priori* and the measurement, since these errors will weight the relative contribution of prior and measurement information to the solution.

The average and variance of a scalar, y, are defined as

$$\overline{y} = \int yP(y)dy, \qquad \sigma^2 = \int (y-\overline{y})^2 P(y)dy, \tag{3.11}$$

where $P(y)$ is the *Probability Density Function* (PDF), which we take to be Gaussian

$$P(y) = \frac{1}{(2\pi)^{1/2}\sigma} exp\left[-\frac{(y-\overline{y})^2}{2\sigma^2}\right]. \tag{3.12}$$

When $\mathbf{y}$ is a vector quantity we write the vector PDF, $P(\mathbf{y})$, as

$$P(\mathbf{y}) = \frac{1}{(2\pi)^{m/2}|\mathbf{S}_\epsilon|^{1/2}} exp\left[-\frac{1}{2}(\mathbf{y}-\overline{\mathbf{y}})^T \mathbf{S}_\epsilon^{-1}(\mathbf{y}-\overline{\mathbf{y}})\right], \tag{3.13}$$

where $\mathbf{S}_\epsilon$ is the matrix covariance of an ensemble of measurements ($\mathbf{y}_k$) about the average measurement ($\overline{\mathbf{y}}$), or the *measurement noise covariance matrix*, whose elements are defined as

$$\mathbf{S}_\epsilon(i,j) = \varepsilon[(y_i - \overline{y}_i)(y_j - \overline{y}_j)] \neq 0. \tag{3.14}$$

In Equation 3.14, ε represents the expectation value operator ($\varepsilon[q] = \int qP(q)dq$), $\overline{y}_i$ the i^{th} measured value averaged over the ensemble k (equal to the i^{th} simulated measurement in the absence of systematic errors), and $y_i - \overline{y}_i$ the i^{th} random measurement error ϵ_i. Switching to vector notation, we also define the matrix covariance of an ensemble of true states, $\mathbf{x}_i$, about the *a priori*, i.e. the *a priori covariance matrix*, $\mathbf{S}_a$, as

$$\mathbf{S}_a = \varepsilon[(\mathbf{x}-\mathbf{x}_a)(\mathbf{x}-\mathbf{x}_a)^T]. \tag{3.15}$$

According to *Rodgers* [2000, p. 25], *"[it is] less realistic, but convenient, to describe prior knowledge of* $\mathbf{x}$ *by a Gaussian PDF"*

$$P(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{S}_a|^{1/2}} exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{x}_a)^T \mathbf{S}_a^{-1}(\mathbf{x}-\mathbf{x}_a)\right]. \tag{3.16}$$

Although we do not possess knowledge of the true state, starting from Equation 3.13 we can write the conditional probability of the measurement given a certain true state as

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^{m/2}|\mathbf{S}_\epsilon|^{1/2}} exp\left[-\frac{1}{2}(\mathbf{y}-\mathbf{K}\mathbf{x})^T \mathbf{S}_\epsilon^{-1}(\mathbf{y}-\mathbf{K}\mathbf{x})\right]. \tag{3.17}$$

Finally, we must introduce Bayes' Theorem regarding joint and conditional probability of two variables, which states that

$$P(\mathbf{y}|\mathbf{x}) = \frac{P(\mathbf{x},\mathbf{y})}{P(\mathbf{x})}, \qquad P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{x},\mathbf{y})}{P(\mathbf{y})}, \tag{3.18}$$

and which allows us to write the quantity that we are interested in, that is the conditional probability of $\mathbf{x}$ given $\mathbf{y}$, i.e. the posterior PDF $P(\mathbf{x}|\mathbf{y})$, as

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y})}. \tag{3.19}$$

If we substitute Equations 3.13, 3.16, and 3.17 into Equation 3.19 and take the logarithm of the resulting expression we can finally write

$$-2lnP(\mathbf{x}|\mathbf{y}) = (\mathbf{y} - \mathbf{K}\mathbf{x})^T\mathbf{S}_\epsilon^{-1}(\mathbf{y} - \mathbf{K}\mathbf{x}) + (\mathbf{x} - \mathbf{x}_a)^T\mathbf{S}_a^{-1}(\mathbf{x} - \mathbf{x}_a) + c_1, \tag{3.20}$$

where $P(\mathbf{y})$, which does not depend on $\mathbf{x}$, has been incorporated into the constant c_1, along with the normalizing factors of $P(\mathbf{y}|\mathbf{x})$ and $P(\mathbf{x})$. Equation 3.20 is quadratic in $\mathbf{x}$, and hence it is possible to write it as

$$-2lnP(\mathbf{x}|\mathbf{y}) = (\mathbf{x} - \hat{\mathbf{x}})^T\hat{\mathbf{S}}^{-1}(\mathbf{x} - \hat{\mathbf{x}}) + c_2. \tag{3.21}$$

By comparing the terms quadratic in $\mathbf{x}$ between Equations 3.20 and 3.21, *Rodgers* [2000, p. 25] shows that the covariance, $\hat{\mathbf{S}}$, of the expected value, $\hat{\mathbf{x}}$, is given by

$$\hat{\mathbf{S}} = (\mathbf{K}^T\mathbf{S}_\epsilon^{-1}\mathbf{K} + \mathbf{S}_a^{-1})^{-1}, \tag{3.22}$$

while the expected value itself is given by equating terms that are linear in $\mathbf{x}^T$

$$\begin{aligned}
\hat{\mathbf{x}} &= (\mathbf{S}_a^{-1} + \mathbf{K}^T\mathbf{S}_\epsilon^{-1}\mathbf{K})^{-1}(\mathbf{S}_a^{-1}\mathbf{x}_a + \mathbf{K}^T\mathbf{S}_\epsilon^{-1}\mathbf{y}) && (3.23)\\
&= \mathbf{x}_a + (\mathbf{S}_a^{-1} + \mathbf{K}^T\mathbf{S}_\epsilon^{-1}\mathbf{K})^{-1}\mathbf{K}^T\mathbf{S}_\epsilon^{-1}(\mathbf{y} - \mathbf{K}\mathbf{x}_a) && (3.24)\\
&= \mathbf{x}_a + \mathbf{S}_a\mathbf{K}^T(\mathbf{K}\mathbf{S}_a\mathbf{K}^T + \mathbf{S}_\epsilon)^{-1}(\mathbf{y} - \mathbf{K}\mathbf{x}_a). && (3.25)
\end{aligned}$$

Equation 3.23 shows that the expected value of the posterior PDF is indeed the weighted sum of measurements and prior information. Equations 3.24 and 3.25 are alternate forms known as the n-form and m-form, according to the size of the matrix to be inverted in the process of finding $\hat{\mathbf{x}}$. Furthermore, starting from Equation 3.24 it is trivial to show that $\mathbf{G}$ must be equal to

$$\mathbf{G} = (\mathbf{S}_a^{-1} + \mathbf{K}^T\mathbf{S}_\epsilon^{-1}\mathbf{K})^{-1}\mathbf{K}^T\mathbf{S}_\epsilon^{-1} = \mathbf{S}_a\mathbf{K}^T(\mathbf{K}\mathbf{S}_a\mathbf{K}^T + \mathbf{S}_\epsilon)^{-1}, \tag{3.26}$$

thus allowing us to calculate $\mathbf{A}$ from Equation 3.9. Finally, we note that the solution to our inverse problem is the posterior distribution, $P(\mathbf{x}|\mathbf{y})$, of which $\hat{\mathbf{x}}$ is the expected value and $\hat{\mathbf{S}}$ is the covariance.

3.1.1.1 Optimal vs. suboptimal approaches

While the Bayesian approach provides us with a complete posterior PDF of the state vector, $P(\mathbf{x}|\mathbf{y})$, we need to select only one solution – logically, it should be one that optimizes something. Equations 3.23-3.25 describe the maximum *a posteriori* (MAP) solution, which is also equal to the expected value solution in the case of Gaussian (or symmetric) error statistics. However, we could have chosen a solution that is optimal in a different sense, e.g. one that minimizes the retrieval error variance. In the case of uncorrelated Gaussian measurement and *a priori* error statistics, the expected value of the state and its covariance are identical in the MAP and minimum variance (MV) solutions.

While we have not yet discussed individual sources of error (see Section 3.3), we can say at this point that it is possible to minimize a weighted sum of measurement and smoothing errors, where the weighting constant becomes a trade-off parameter between noise and resolution errors. While this approach is useful, it departs from optimality as defined for the MAP and MV solutions, which both optimize the total retrieval error covariance, $\hat{\mathbf{S}}$, in the case of uncorrelated Gaussian error statistics.

Another suboptimal approach (in the above sense) is to seek a solution with maximum vertical resolution. *Backus and Gilbert* [1970] minimized the *spread* of the averaging kernels subject to the constraint that their areas be equal to 1, initially without regard for noise or the departure of the retrieved profile from the true profile. *Conrath* [1972] developed a simultaneous minimization of noise and spread, and additionally examined constraints on the *resolving length*, i.e. the distance between the nominal and true height of an averaging kernel.

3.1.2 Nonlinear problems

In deriving the optimal solution and its error covariance (Equations 3.22 and 3.23-3.25) we have assumed Gaussian error statistics and a *linear* or *nearly linear* forward model, allowing us to find the optimal solution in one step. The inverse problem of deriving vertical profiles from infrared spectra is in fact *moderately nonlinear*, whereby the forward model linearization is sufficient to carry out an error analysis and retrieval characterization in the vicinity of the optimal solution, however it is insufficient to find the optimal solution in one step in the first place. (*Grossly nonlinear* problems are nonlinear even within the solution error bars, complicating matters even further.) In the moderately nonlinear inverse problem, the weighting function matrix **K** depends on the state, and hence, a solution must be found iteratively by evaluating **K** at each step of the iteration. We begin by taking the gradient (∇) of Equation 3.20 with respect to $\mathbf{x}$ and equating it to zero in order to find $\hat{\mathbf{x}}$:

$$\nabla_{\mathbf{x}}[-2lnP(\mathbf{x}|\mathbf{y})] = -[\nabla_{\mathbf{x}}\mathbf{F}(\mathbf{x})]^T\mathbf{S}_\epsilon^{-1}[\mathbf{y}-\mathbf{F}(\mathbf{x})] + \mathbf{S}_a^{-1}[\mathbf{x}-\mathbf{x}_a] = 0. \tag{3.27}$$

Setting $\mathbf{K}(\mathbf{x}) = \nabla_{\mathbf{x}}\mathbf{F}(\mathbf{x})$ and for $\mathbf{x} = \hat{\mathbf{x}}$ at the minimum we can write

$$-[\mathbf{K}(\hat{\mathbf{x}})]^T\mathbf{S}_\epsilon^{-1}[\mathbf{y}-\mathbf{F}(\hat{\mathbf{x}})] + \mathbf{S}_a^{-1}[\hat{\mathbf{x}}-\mathbf{x}_a] = 0, \tag{3.28}$$

which is an implicit equation for $\hat{\mathbf{x}}$ that must be solved numerically. The right hand side of Equation 3.27 is the gradient of the cost function, $\mathbf{g}(\mathbf{x})$, which is again differentiated in the vector-equivalent of *Newton's* method

$$\mathbf{x}_{i+1} = \mathbf{x}_i - [\nabla_{\mathbf{x}}\mathbf{g}(\mathbf{x}_i)]^{-1}\mathbf{g}(\mathbf{x}_i), \tag{3.29}$$

where the second derivative of the cost function, $\nabla_{\mathbf{x}}\mathbf{g}$, follows from Equation 3.27:

$$\nabla_{\mathbf{x}}\mathbf{g} = \mathbf{S}_a^{-1} + \mathbf{K}^T\mathbf{S}_\epsilon^{-1}\mathbf{K} - [\nabla_{\mathbf{x}}\mathbf{K}^T]\mathbf{S}_\epsilon^{-1}[\mathbf{y} - \mathbf{F}(\mathbf{x})]. \tag{3.30}$$

This method is also known as the *Inverse Hessian* method because the second derivative of a function is in some cases referred to as the *Hessian.* When the third term of Equation 3.30 is small – as is the case in what are known as *small residual* moderately nonlinear problems – the Newton method becomes the *Gauss-Newton* method and yields the following iterative solutions for $\mathbf{x}_{i+1}$, in complete analogy with Equations 3.23-3.25

$$\begin{aligned}
\mathbf{x}_{i+1} &= \mathbf{x}_i + (\mathbf{S}_a^{-1} + \mathbf{K}_i^T\mathbf{S}_\epsilon^{-1}\mathbf{K}_i)^{-1}\{\mathbf{K}_i^T\mathbf{S}_\epsilon^{-1}(\mathbf{y} - \mathbf{y}_i)) - \mathbf{S}_a^{-1}(\mathbf{x}_i - \mathbf{x}_a)\} && (3.31)\\
&= \mathbf{x}_a + (\mathbf{S}_a^{-1} + \mathbf{K}_i^T\mathbf{S}_\epsilon^{-1}\mathbf{K}_i)^{-1}\mathbf{K}_i^T\mathbf{S}_\epsilon^{-1}\{(\mathbf{y} - \mathbf{y}_i) + \mathbf{K}_i(\mathbf{x}_i - \mathbf{x}_a)\} && (3.32)\\
&= \mathbf{x}_a + \mathbf{S}_a\mathbf{K}_i^T(\mathbf{K}_i\mathbf{S}_a\mathbf{K}_i^T + \mathbf{S}_\epsilon)^{-1}\{(\mathbf{y} - \mathbf{y}_i) + \mathbf{K}_i(\mathbf{x}_i - \mathbf{x}_a)\}, && (3.33)
\end{aligned}$$

where $\mathbf{K}_i$ is taken to mean "$\mathbf{K}$ evaluated at $\mathbf{x}_i$" and $\mathbf{y}_i = \mathbf{F}(\mathbf{x}_i)$. Neglecting the $\nabla_{\mathbf{x}}\mathbf{K}^T$ term in Equation 3.30 is equivalent to assuming a Gaussian posterior PDF or linearity within the error bars of the iterated solution, i.e. that the problem is not grossly nonlinear. Finally, the covariance of the solution is evaluated as before, once $\hat{\mathbf{x}}$ is found

$$\hat{\mathbf{S}} = (\hat{\mathbf{K}}^T\mathbf{S}_\epsilon^{-1}\hat{\mathbf{K}} + \mathbf{S}_a^{-1})^{-1}, \tag{3.34}$$

and is the inverse of the Hessian (Equation 3.30) when we neglect the $\nabla_{\mathbf{x}}\mathbf{K}^T$ term ($\hat{\mathbf{K}}$ means "$\mathbf{K}$ evaluated at $\hat{\mathbf{x}}$"). The averaging kernel matrix is also obtained in an analogous manner as $\hat{\mathbf{A}} = \hat{\mathbf{G}}\hat{\mathbf{K}}$.

3.2 Retrieval Characterization

We now turn to the characterization of the retrieval introduced in the previous section. This is an important step in any retrieval method, but it is *critical* in the highly underdetermined problem of retrieving vertical profiles of trace gases from ground-based solar absorption spectra. In this section we will outline concepts and methods used to determine the information content and vertical resolution of the measurement and retrieval. In Section 3.3 we will analyse how different sources of error propagate to the retrieved solution. Both sections follow the general approach of *Rodgers* [2000, Ch. 2 and Ch. 3], which can be applied to inverse methods of various designs.

3.2.1 Information content

It is quite conceivable that in the presence of random measurement noise, the number of independent pieces of information which can be extracted from the measurements can be reduced to less than p. This occurs when elements of the state that correspond to very small singular values of $\mathbf{K}$ become overwhelmed by random measurement noise, and are no longer measurable. When this is the case, we speak of the *effective rank* of the problem and its *effective row space*. To identify the effective row space we must compare the measurement error covariance with the covariance of the natural variability of the measurement vector. If a measurement's natural variability is smaller than the error in that measurement, then it is not detectable in practice and not part of the effective row space. In order to carry out this comparison, we must perform two transformations that eliminate the off-diagonal covariance elements which complicate our task. If we define

$$\tilde{\mathbf{x}} = \mathbf{S}_a^{-\frac{1}{2}}\mathbf{x} \qquad \tilde{\mathbf{y}} = \mathbf{S}_\epsilon^{-\frac{1}{2}}\mathbf{y} \qquad \tilde{\boldsymbol{\epsilon}} = \mathbf{S}_\epsilon^{-\frac{1}{2}}\boldsymbol{\epsilon}, \tag{3.35}$$

then

$$\tilde{\mathbf{y}} = \mathbf{S}_\epsilon^{-\frac{1}{2}}\mathbf{K}\mathbf{S}_a^{\frac{1}{2}}\tilde{\mathbf{x}} + \tilde{\boldsymbol{\epsilon}} \equiv \tilde{\mathbf{K}}\tilde{\mathbf{x}} + \tilde{\boldsymbol{\epsilon}}. \tag{3.36}$$

The covariances of $\tilde{\mathbf{x}}$ and $\tilde{\boldsymbol{\epsilon}}$ are now equal to unit matrices because, for example,

$$\mathbf{S}_{\tilde{\epsilon}} = \varepsilon[\tilde{\boldsymbol{\epsilon}}\tilde{\boldsymbol{\epsilon}}^T] = \mathbf{S}_\epsilon^{-\frac{1}{2}}\varepsilon[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T]\mathbf{S}_\epsilon^{-\frac{1}{2}} = \mathbf{S}_\epsilon^{-\frac{1}{2}}\mathbf{S}_\epsilon\mathbf{S}_\epsilon^{-\frac{1}{2}} = \mathbf{I}_m. \tag{3.37}$$

For this reason the above transformation of $\mathbf{y}$ is referred to as *pre-whitening* – because it turns noise into white noise. From Equation 3.36, we have that the prior covariance of the pre-whitened measurement vector is given by

$$\mathbf{S}_{\tilde{y}} = \varepsilon[\tilde{\mathbf{y}}\tilde{\mathbf{y}}^T] = \varepsilon[(\tilde{\mathbf{K}}\tilde{\mathbf{x}} + \tilde{\boldsymbol{\epsilon}})(\tilde{\mathbf{K}}\tilde{\mathbf{x}} + \tilde{\boldsymbol{\epsilon}})^T] = \tilde{\mathbf{K}}\tilde{\mathbf{K}}^T + \mathbf{I}_m. \tag{3.38}$$

According to Equation 3.37, $\mathbf{I}_m$ is the component of the covariance due to measurement noise, and hence $\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T$ must be the component of the covariance due to the variability of the state. However, $\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T$ is not normally diagonal, and one final transformation is necessary before we can compare it to $\mathbf{I}_m$. Using the SVD of $\tilde{\mathbf{K}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T$, we define

$$\mathbf{y}' = \mathbf{U}^T\tilde{\mathbf{y}}, \qquad \mathbf{x}' = \mathbf{V}^T\tilde{\mathbf{x}}, \qquad \boldsymbol{\epsilon}' = \mathbf{U}^T\tilde{\boldsymbol{\epsilon}}, \tag{3.39}$$

to give a fully diagonal forward model equation

$$\mathbf{y}' = \boldsymbol{\Lambda}\mathbf{x}' + \boldsymbol{\epsilon}'. \tag{3.40}$$

The covariances of $\boldsymbol{\epsilon}'$ and $\mathbf{x}'$ remain equal to unity (now of dimension $\mathbf{I}_p$), because both $\mathbf{U}^T\mathbf{U}$ and $\mathbf{V}^T\mathbf{V}$ are equal to $\mathbf{I}_p$, and finally, we can see that the covariance of $\mathbf{y}'$, $\mathbf{S}_{\mathbf{y}'}$, is equal to the diagonal matrix $\boldsymbol{\Lambda}^2 + \mathbf{I}_p$. The diagonal elements of $\boldsymbol{\Lambda}^2$ can now be compared to the diagonal elements of the transformed measurement noise covariance ($\mathbf{I}_p$). Elements of $\mathbf{y}'$ which vary more than the noise correspond to $\lambda_i^2 \gtrsim 1$, where λ_i represent the non-zero singular values of $\tilde{\mathbf{K}}$. Finally, we have that *"the number of independent measurements made to better than measurement error, the effective rank of the problem, is the number of singular values of* $\mathbf{S}_\epsilon^{-\frac{1}{2}}\mathbf{K}\mathbf{S}_a^{\frac{1}{2}}$ *which are greater than about unity,"* [*Rodgers*, 2000, p. 29].

In order to make the concept of information content more precise, *Rodgers* [2000] defines the *degrees of freedom for signal*, d_s, and the *degrees of freedom for noise*, d_n as follows:

$$d_s = \varepsilon[(\hat{\mathbf{x}} - \mathbf{x}_a)^T\mathbf{S}_a^{-1}(\hat{\mathbf{x}} - \mathbf{x}_a)], \tag{3.41}$$

$$d_n = \varepsilon[(\mathbf{y} - \mathbf{K}\hat{\mathbf{x}})^T\mathbf{S}_\epsilon^{-1}(\mathbf{y} - \mathbf{K}\hat{\mathbf{x}})]. \tag{3.42}$$

The meaning of these quantities becomes clear if we again consider the linear inverse problem characterized by Gaussian measurement and *a priori* statistics, wherein the most probable state minimizes

$$\chi^2 = (\mathbf{y} - \mathbf{Kx})^T \mathbf{S}_\epsilon^{-1} (\mathbf{y} - \mathbf{Kx}) + (\mathbf{x} - \mathbf{x}_a)^T \mathbf{S}_a^{-1} (\mathbf{x} - \mathbf{x}_a), \qquad (3.43)$$

with $\mathbf{y} - \mathbf{Kx} = \boldsymbol{\epsilon}$. We know that for the most probable state, the expected value of χ^2 is equal to the total number of degrees of freedom, which is equal to the number of measurements $[(m+n) - n = m]$. However, we have also seen that some measurements contribute information while others contribute only noise. The non-integer quantity d_s measures the extent to which the retrieved state has changed from the *a priori*, and therefore, the amount of *signal* provided by the measurement process. The non-integer quantity d_n measures how much the modeled signal differs from the measured signal, and therefore, the kind of random *noise* that has not been accounted for by the retrieval. In this way, d_s quantifies what portion of χ^2 is attributed to signal in the state vector, while d_n quantifies what portion of χ^2 is attributed to noise in the measurement vector. Finally, the two quantities can be written in two different forms, depending on whether we consider the measurement and state space described by $\mathbf{x}$ and $\mathbf{y}$, or by the transformed variables $\mathbf{x}'$ and $\mathbf{y}'$:

$$d_s = tr[(\mathbf{S_a^{-1}} + \mathbf{K^T S_\epsilon^{-1} K})^{-1} \mathbf{K^T S_\epsilon^{-1} K}] = tr[\mathbf{\Lambda}^2 (\mathbf{\Lambda}^2 + \mathbf{I}_m)^{-1}] = \sum_{i=1}^{m} \frac{\lambda_i^2}{1 + \lambda_i^2} \qquad (3.44)$$

$$d_n = tr[(\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1} \mathbf{S}_a^{-1}] + m - n = tr[(\mathbf{\Lambda}^2 + \mathbf{I}_m)^{-1}] = \sum_{i=1}^{m} \frac{1}{1 + \lambda_i^2}, \qquad (3.45)$$

where $tr(\mathbf{A})$ represents the mathematical trace operator. Note that $\mathbf{\Lambda}$ is now an $m \cdot m$ matrix (because it now also includes the $m - p$ singular values of $\tilde{\mathbf{K}}$ that are equal to zero), such that $d_s + d_n = m$ in both formulations. Furthermore, we note from Equation 3.44 that $d_s = tr(\mathbf{A})$, which hints at a relationship between the eigenvalues and eigenvectors of $\mathbf{A}$ and the singular values and singular vectors of $\tilde{\mathbf{K}}$. If we consider the eigenvalue decomposition of $\mathbf{A} = \mathbf{R \Delta L}^T$, where $\mathbf{R}$ and $\mathbf{L}$ contain the right and left eigenvectors, respectively, and $\mathbf{\Delta}$ is the diagonal matrix of the eigenvalues of $\mathbf{A}$, then it can be shown that $\mathbf{R} = \mathbf{S}_a^{\frac{1}{2}} \mathbf{V}$, $\mathbf{L} = \mathbf{S}_a^{-\frac{1}{2}} \mathbf{V}$, and $\mathbf{\Delta} = \mathbf{\Lambda}^2 (\mathbf{\Lambda}^2 + \mathbf{I}_p)^{-1}$. Since the eigenvalues of $\mathbf{A}$, δ_i, are equal to $\lambda_i^2/(1+\lambda_i^2)$, we can see that the number of independent measurements made to better than measurement error must also be equal to the number of eigenvalues of $\mathbf{A}$ that are of order 1.

In characterizing the information content of TAO retrievals, we take the trace of $\mathbf{A}$ over a certain vertical range of the atmosphere in order to determine the number of independently resolved pieces of information present in the partial column derived from this region. Typically, for a total column measurement (0-100 km) d_s ranges from ~1 for a weakly absorbing gas such as NO_2 to ~4 for a strongly absorbing gas such as N_2O. Finally, we note that the area of each averaging kernel (the sum of its elements) is ~1 in regions where the retrieval has good sensitivity to the true state. Otherwise, retrieved information comes from the *a priori*. The characterization of vertical information contained in high-resolution FTIR spectra is of ongoing interest to the NDACC; the more sophisticated vertical profile and partial column data products clearly have more uses in basic research and satellite validation applications than the more limited vertical columns. Results of TAO retrieval characterization for a suite of trace gases are discussed in Sections 4.3 and 5.2.

3.2.1.1 Other definitions of information content

In the linear Gaussian case, the inverse covariance matrix $\hat{\mathbf{S}}^{-1}$ is sometimes referred to as the *Fisher* information matrix, from the body of work on maximum likelihood estimation. This posterior information matrix is in turn the sum of the information matrices of the prior and of the measurement ($\mathbf{S}_a^{-1} + \mathbf{K}^T\mathbf{S}_\epsilon^{-1}\mathbf{K}$).

The *Shannon* information content of a measurement is defined as the logarithm of the factor by which the measurement improves the knowledge of a quantity over its known natural variability. For example, if a certain sea surface temperature that is known to have a natural variability of 4 K is measured with an accuracy of 1 K, then the measurement has improved the knowledge of this temperature by a factor of 4, thus providing 2 bits of information. In our multivariate case, the hyper-volume of the posterior PDF is reduced over the volume of the prior PDF by the act of measurement, and the Shannon information content is equal to the logarithm of the volume reduction factor. According to *Rodgers* [2000], in the Shannon information content definition, a Gaussian PDF has an *entropy* that is equal to the logarithm of the volume inside a surface of constant probability (apart from a constant which depends on the surface chosen), and therefore, the information content of a measurement can be expressed as

$$H = \frac{1}{2} ln|\hat{\mathbf{S}}^{-1}\mathbf{S}_a| = -\frac{1}{2} ln|\mathbf{I}_n - \mathbf{A}| = \frac{1}{2}\sum_{i=1}^{m} ln(1+\lambda_i^2). \tag{3.46}$$

This formulation has a direct relationship to the thermodynamic definition of entropy as the logarithm of the number of distinct internal states of a thermodynamic system consistent with a measured macro-state (e.g. pressure). Furthermore, *Rodgers* [2000] shows that given the mean and variance of a random variable, a Gaussian PDF for the distribution of this variable has maximum entropy. This is consistent with our desire to constrain the prior PDF of a variable as little as possible, and is the primary reason for defaulting to a Gaussian PDF (besides the convenience of dealing with a quadratic cost function as a result).

3.2.2 Vertical Resolution

In correspondence to several definitions of information, there exist several definitions of vertical resolution, apart from the vertical grid spacing on which the state vector is defined. As already alluded to in Section 3.2.1, in ground-based solar absorption measurements, grid spacing cannot be used to define resolution because the rank of $\mathbf{K}$ is much less than n, implying a high degree of correlation between retrieved state vector elements. Outside of the realm of ground-based FTIR measurements, various approaches have been taken to defining resolution, each of these considering measurement noise and/or total retrieval error in some fashion. For example, *Conrath* [1972] produced trade-off plots of retrieval error as a function of averaging kernel spread in order to choose the optimal resolution for his particular observing system. Another approach, discussed by *Rodgers* [2000], is to perform retrievals on simulated measurements wherein sinusoidal perturbations of a given amplitude but of varying frequency are superimposed on the *a priori* profile used in the retrieval; the amplitude of the retrieval response is compared to the total retrieval error in order to determine the degree to which the observing system is sensitive to the perturbation. In yet another definition, the reciprocal of d_s, i.e. *levels per degree of freedom*, can be used as a measure of resolution, since the d_s corresponding to each level is equal to the nominal peak of the averaging kernel at that level. If the averaging kernel is predominantly positive and has a unit area, then we can see how the inverse of its peak provides a measure of its resolution. Finally, the averaging kernel width itself can also be used as a measure of resolution – provided that the kernels do not have very significant positive or negative sidelobes – and this will be the approach used in the characterization of TAO retrievals.

As we have shown previously, the best estimate of the retrieved state is given by

$$\hat{\mathbf{x}} = \mathbf{x}_a + \mathbf{A}(\mathbf{x} - \mathbf{x}_a). \tag{3.47}$$

If a measurement system resolves each element of the true state vector with perfect sensitivity, $\mathbf{A}$ is equal to the identity matrix, the best estimate of the state does not depend in any way on the *a priori*, and the resolution of the observing system is equal to the vertical grid spacing. In ground-based FTIR measurements the diagonal elements of $\mathbf{A}$ are not unity, representing imperfect sensitivity to the true state at all heights, and off diagonal elements are present in all rows of $\mathbf{A}$, representing the inability to *independently* resolve all elements of the true state vector.

Since $A_{ij} = \partial\hat{x}_i/\partial x_j$, we can see that the i^{th} row of $\mathbf{A}$ characterizes the response of the i^{th} element of the retrieved state vector to perturbations in all elements of the true state vector. The rows of $\mathbf{A}$ are referred to as *smoothing functions*, or *averaging kernels*, or *resolution functions*. On the other hand, the j^{th} column of $\mathbf{A}$ characterizes the response of the retrieved state vector to an impulse applied at the j^{th} element of the true state vector. The columns of **A** are alternately referred to as *impulse response functions*, or *δ-function responses*, or *point spread functions.* A single δ-function response is calculated in a straightforward manner as the change in each element of a retrieval due to a linear perturbation in a single state vector element x_i. If x_1 to x_n are perturbed in turn, the n resultant δ-function responses make up the n columns of **A**, and the rows of this matrix then correspond to the averaging kernels. This is known as the perturbation method of calculating averaging kernels and is sometimes convenient when the weighting functions of the problem are not readily available to the data user.

As stated previously, the profiles of trace gases derived from ground-based FTIR spectra have a limited vertical resolution, and indeed, we only expect ∼1-4 independently resolved layers, depending on the trace gas in question. This corresponds to a vertical resolution of at best ∼10 km between the surface and ∼40 km, where pressure broadening is well resolved. As such, it is clear that ground-based FTIR retrievals cannot resolve any fine structure present in the profile. However, a more subtle point made by *Rodgers* [2000, p. 56], is that a low-resolution retrieval is also not sensitive to gradient discontinuities in the profile, and that these are contributed entirely by the *a priori*. Characterization of TAO retrieval resolution is presented for a suite of trace gases in Sections 4.3 and 5.2.

3.3 Retrieval Error Analysis

We resume the error analysis postponed in Section 3.1.1 and rewrite Equation 3.6 as

$$\hat{\mathbf{x}} = \mathbf{I}(\mathbf{F}(\mathbf{x},\mathbf{b}) + \boldsymbol{\Delta}\mathbf{f}(\mathbf{x},\mathbf{b},\mathbf{b}') + \boldsymbol{\epsilon},\mathbf{b}) = \mathbf{T}(\mathbf{x},\mathbf{b}), \tag{3.48}$$

where we have now distinguished between the forward *function* $\mathbf{f}(\mathbf{x},\mathbf{b},\mathbf{b}')$ and the forward *model* $\mathbf{F}(\mathbf{x},\mathbf{b})$. The former describes the complete physics of the problem, while the latter is either a convenient approximation to the forward function, or its best-known empirical representation in cases where the complete physics of the problem is not fully understood. $\mathbf{b}'$ represents parameters of the forward function that have been ignored in constructing the approximate forward model (where $\mathbf{F}(\mathbf{x},\mathbf{b})$ is the best-known empirical approximation, $\mathbf{b}'$ are unknown). It is assumed that $\mathbf{F}(\mathbf{x},\mathbf{b}) \simeq \mathbf{f}(\mathbf{x},\mathbf{b},\mathbf{b}')$ and that $\mathbf{\Delta f} = \mathbf{f}(\mathbf{x},\mathbf{b},\mathbf{b}') - \mathbf{F}(\mathbf{x},\mathbf{b})$. $\boldsymbol{\epsilon}$ is again the random measurement noise. Replacing the forward model in Equation 3.48 by a linearization about $\mathbf{x}_a$ and $\mathbf{b}_a$ (Equation 3.3), we have

$$\hat{\mathbf{x}} = \mathbf{I}[\mathbf{F}(\mathbf{x}_a,\mathbf{b}_a) + \mathbf{K}(\mathbf{x}-\mathbf{x}_a) + \mathbf{K}_b(\mathbf{b}-\mathbf{b}_a) + \Delta\mathbf{f}(\mathbf{x},\mathbf{b},\mathbf{b}') + \boldsymbol{\epsilon}, \mathbf{b}_a], \tag{3.49}$$

which upon linearization about $\mathbf{y}_a$ finally yields

$$\begin{aligned}
\hat{\mathbf{x}} &= \mathbf{I}[\mathbf{F}(\mathbf{x}_a,\mathbf{b}_a),\mathbf{b}_a] + \frac{\partial \mathbf{I}}{\partial \mathbf{F}}[\mathbf{K}(\mathbf{x}-\mathbf{x}_a) + \mathbf{K}_b(\mathbf{b}-\mathbf{b}_a) + \Delta\mathbf{f}(\mathbf{x},\mathbf{b},\mathbf{b}') + \boldsymbol{\epsilon}] && (3.50)\\
&= \mathbf{x}_a + \mathbf{GK}(\mathbf{x}-\mathbf{x}_a) + \mathbf{GK}_b(\mathbf{b}-\mathbf{b}_a) + \mathbf{G\Delta f}(\mathbf{x},\mathbf{b},\mathbf{b}') + \mathbf{G}\boldsymbol{\epsilon} && (3.51)\\
&= \mathbf{x}_a + \mathbf{A}(\mathbf{x}-\mathbf{x}_a) + \mathbf{G}\boldsymbol{\epsilon}_y, && (3.52)
\end{aligned}$$

where we assume that the retrieval method does not introduce a bias ($\mathbf{I}[\mathbf{F}(\mathbf{x}_a,\mathbf{b}_a),\mathbf{b}_a] = \mathbf{x}_a$) and that $\boldsymbol{\epsilon}_y$ now represents the total error due to forward model parameter errors, forward model error and random measurement noise ($\boldsymbol{\epsilon}_y = \mathbf{K}_b(\mathbf{b}-\mathbf{b}_a) + \mathbf{\Delta f}(\mathbf{x},\mathbf{b},\mathbf{b}') + \boldsymbol{\epsilon}$). Equation 3.51 can be rearranged as

$$\begin{aligned}
\hat{\mathbf{x}} - \mathbf{x} &= (\mathbf{A} - \mathbf{I}_n)(\mathbf{x}-\mathbf{x}_a) && (\textit{smoothing error})\\
&+ \mathbf{GK}_b(\mathbf{b}-\mathbf{b}_a) && (\textit{model parameter error})\\
&+ \mathbf{G\Delta f}(\mathbf{x},\mathbf{b},\mathbf{b}') && (\textit{forward model error}) \qquad (3.53)\\
&+ \mathbf{G}\boldsymbol{\epsilon} && (\textit{retrieval noise})
\end{aligned}$$

where $\hat{\mathbf{x}} - \mathbf{x}$ gives the error in $\hat{\mathbf{x}}$, and can be grouped into the *smoothing error* and the *retrieval error* ($\mathbf{G}\boldsymbol{\epsilon}_y$), in turn comprised of the *model parameter error*, *forward model error*, and *retrieval noise*. We note that *Rodgers* [2000] reserves the term *noise* for random measurement noise and its propagation to the retrieval, while he uses the term *error* more loosely to signify both systematic and total errors, the latter also including random noise. By taking the expectation value of Equation 3.53, we obtain the error covariance matrices

$$\begin{aligned} \mathbf{S}_s &= (\mathbf{A}-\mathbf{I})\mathbf{S}_a(\mathbf{A}-\mathbf{I})^T && (3.54) \\ \mathbf{S}_f &= \mathbf{G}\mathbf{K}_b\mathbf{S}_b\mathbf{K}_b^T\mathbf{G}^T && (3.55) \\ \mathbf{S}_m &= \mathbf{G}\mathbf{S}_\epsilon\mathbf{G}^T, && (3.56) \end{aligned}$$

where $\mathbf{S}_b = \varepsilon[(\mathbf{b}-\mathbf{b}_a)(\mathbf{b}-\mathbf{b}_a)^T]$, and where we have neglected the forward model error due to the complications in evaluating it without knowledge of $\mathbf{x}$ and $\mathbf{b}$. $\mathbf{S}_s$ is the covariance of the error caused by the smoothing of the true state by the averaging kernel, $\mathbf{S}_f$ is the covariance of the error due to inaccuracies in forward model parameters, and $\mathbf{S}_m$ is the covariance of the error due to random measurement noise propagating to the retrieval. Starting from the definition of $\mathbf{S}_s$, $\mathbf{S}_m$, and $\mathbf{G}$, it is straightforward to show that $\mathbf{S}_s + \mathbf{S}_m = \hat{\mathbf{S}}$, i.e. the total error (apart from model parameter error) is identical to the covariance of $P(\mathbf{x}|\mathbf{y})$, as shown in Equation 3.22.

While $\mathbf{S}_m$ is purely random and $\mathbf{S}_f$ can contain both random (e.g. temperature, solar zenith angle) and systematic (ILS, spectroscopic parameters) components, $\mathbf{S}_s$ is a good example of an error that is systematic on short time scales, but random on long time scales. *Rodgers* [2000, p. 49] makes a critical point about $\mathbf{S}_s$, i.e. that its estimate is only correct insofar as $\mathbf{S}_a$ represents the covariance of a real ensemble of atmospheric states, and $\mathbf{x}_a$ represents the true mean atmospheric state. Furthermore, he points out that *"to estimate it correctly, the actual statistics of the fine structure must be known. [...] If the real covariance is not available, it may be better to abandon the estimation of the smoothing error, and consider the retrieval as an estimate of a smoothed version of the state, rather than an estimate of the complete state [with a smoothing error]."*

3.3.1 Interpreting error covariances: Error patterns

Covariances are essentially multi-dimensional and correlated error bars wherein diagonal elements represent the simple variances of each of the elements of the retrieved state vector. It is common practice to display the variances or standard deviations of $\mathbf{S}_a$, $\mathbf{S}_s$, and $\mathbf{S}_m$, and $\hat{\mathbf{S}} = \mathbf{S}_s + \mathbf{S}_m$ on the same plot (Section 4.3 and 5.2). Retrieval noise variance decreases outside the region where weighting functions are large because this is where measurements contribute little to the retrieval and information is obtained from the *a priori*. Likewise, outside the weighting function region the variance of the smoothing error (and the total retrieval error) increases and approaches that of the *a priori*.

Although off-diagonal elements are hard to visualize, it is important to recognize that they provide additional information about the measurement. One can further visualize a covariance matrix, $\mathbf{S}$, by solving its eigenvalue equation ($\mathbf{S}\mathbf{l}_i = \lambda_i \mathbf{l}_\mathbf{i}$) for its eigenvalues ($\lambda_i$) and eigenvectors ($\mathbf{l}_i$). Because a covariance matrix is symmetric, $\mathbf{S}$ can be decomposed into

$$\mathbf{S} = \sum_i \lambda_i \mathbf{l}_i \mathbf{l}_i^T = \sum_i \mathbf{e}_i \mathbf{e}_i^T, \tag{3.57}$$

where the orthogonal vectors $\mathbf{e}_i = \lambda_i^{\frac{1}{2}} \mathbf{l}_i$ are what *Rodgers* [2000] refers to as *error patterns*. This is useful because the error in the state vector, $\boldsymbol{\epsilon}_x$, can be written as the sum of the error patterns multiplied by a random factor a_i of unit variance

$$\boldsymbol{\epsilon}_x = \sum_{i=1}^{n} a_i \mathbf{e}_i, \tag{3.58}$$

and because the retrieval contains only the *a priori* plus these error pattern shapes [*Rodgers*, 2000, p. 59].

Referring to Figure 3.1, long vertical wavelength error patterns of $\mathbf{S}_m$ have small amplitudes since the spectra provide good information about low vertical resolution variations of the atmosphere (Pattern 5); short wavelength error patterns of $\mathbf{S}_m$ can also have small amplitudes since high vertical resolution variations of the atmosphere are not captured by the spectral lineshapes, and the information at these wavelengths comes from

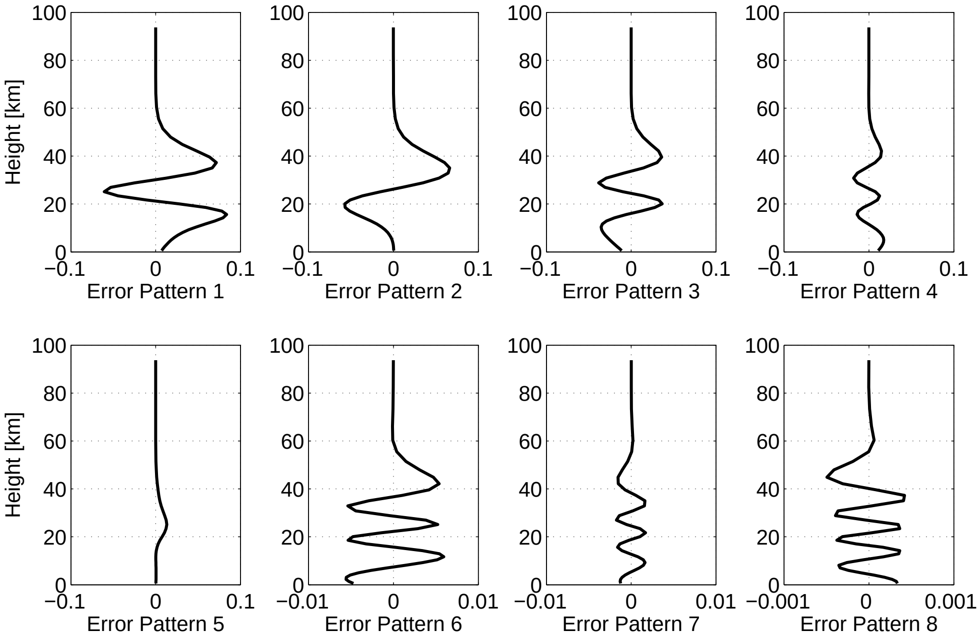


Figure 3.1 First eight error patterns of the retrieval noise covariance ($\mathbf{S}_m$) of a typical O_3 retrieval described in Chapter 5 (note the changing horizontal scale).

the *a priori* (Patterns 4, and 6-8); conversely, intermediate wavelength error patterns of $\mathbf{S}_m$ – on the scale of those found in **K** and **A** – have large amplitudes (Patterns 1-3). A similar discussion applies to the error patterns of $\mathbf{S}_s$, although it is more difficult to discern in Figure 3.2: very short wavelength patterns will generally have small amplitudes since $\mathbf{S}_a$ does not contain these very small scale structures; $\mathbf{S}_s$ error patterns with wavelengths that are somewhat shorter than those contained in **K** and **A** will contribute with significant amplitudes; and finally, long wavelength $\mathbf{S}_s$ error patterns will have small amplitudes where **K** and **A** show good sensitivity, but the amplitudes will grow outside of this region. Finally, $\mathbf{S}_s$ may be of full rank (thus yielding up to n error patterns), unlike $\mathbf{S}_m$.

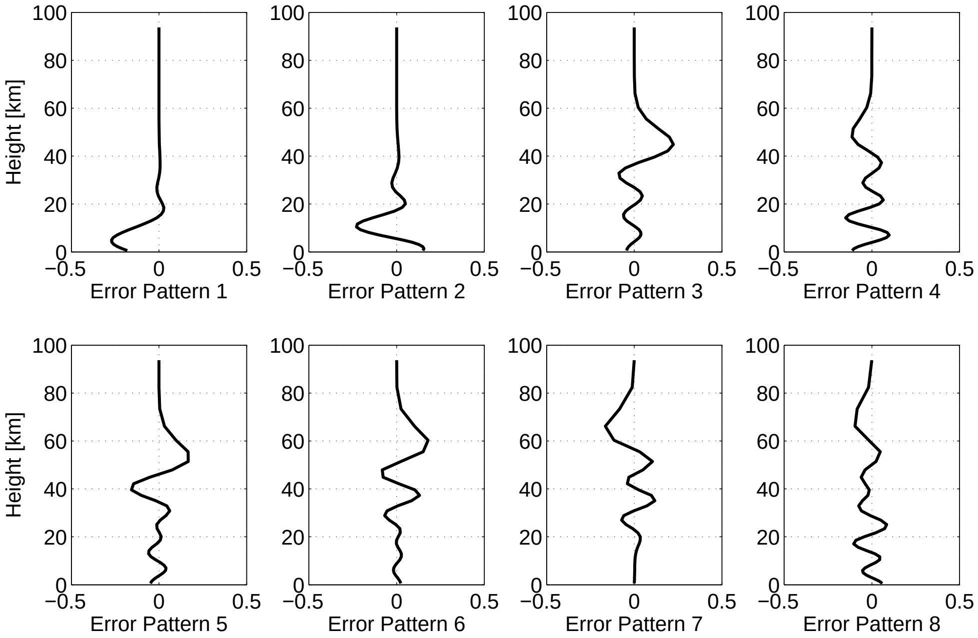


Figure 3.2 First eight error patterns of the smoothing error covariance ($\mathbf{S}_s$) of a typical O_3 retrieval described in Chapter 5.

3.3.2 Model parameters as state vector elements

When the forward model shows high sensitivity to a model parameter that varies considerably we have two options for dealing with this problem, which are equivalent in the linear case. First, the model parameter in question can be placed in the state vector and retrieved together with the vertical profile using the noise covariance $\mathbf{S}_\epsilon$. After the retrieval is performed, the contribution of the parameter in question to the retrieval error covariance is calculated as $\mathbf{G}\mathbf{K}_b\mathbf{S}_b\mathbf{K}_b^T\mathbf{G}^T$. In the second option, the model parameter is not retrieved, but the retrieval is performed with an effective error covariance $\mathbf{K}_b\mathbf{S}_b\mathbf{K}_b^T + \mathbf{S}_\epsilon$. Since the computationally expensive Jacobian calculation has already been performed, the model parameter can be retrieved afterwards with little additional cost. In the nonlinear case, however, it may be necessary to retrieve $\mathbf{x}$ and $\mathbf{b}$ together in order to obtain the correct weighting functions [*Rodgers*, 2000, p. 71].

In the analysis of TAO spectra, the first approach is taken, since the inverse problem is moderately nonlinear. An example of a retrieved forward model parameter is the background aerosol transmission slope. The 38 temperature parameters (necessary for the 38 retrieval layers at TAO) are neither retrieved nor accounted for by an effective error covariance during a retrieval (e.g. $\mathbf{K}_b\mathbf{S}_T\mathbf{K}_b^T + \mathbf{S}_\epsilon$, where $\mathbf{S}_T$ is a 38·38 temperature error covariance). However, their error contribution is estimated (post-retrieval) after calculating the m·38 temperature weighting function matrix by perturbation methods. The error analysis of TAO retrievals is presented for a suite of trace gases in Sections 4.3 and 5.2.

3.4 SFIT-2 Retrieval Algorithm

The above OEM formulation of *Rodgers* [1976, 1990, 2000] is implemented semi-empirically in the SFIT-2 algorithm developed at NASA Langley and the New Zealand Institute for Water and Atmosphere (NIWA) [*Pougatchev et al.*, 1995; *Connor et al.*, 1996; *Rinsland et al.*, 1998]. The term *semi-empirical* is used because the *a priori* covariance is based on prior measurements of the state or *ad hoc* smoothing constraints, and not on its theoretical behaviour [*Rodgers*, 2000, p. 99]. SFIT-2 has been successfully compared to another OEM retrieval algorithm (PROFFIT-9) developed by *Hase et al.* [2004].

Prior to performing the retrieval, we calculate a model atmosphere using the ray tracing program *fscatm*, documented extensively by *Gallery et al.* [1983]. Recent improvements to the code, including an electronic supplement, are described by *Meier et al.* [2004a]. *fscatm* converts pressure, temperature and VMR profiles specified at 63 layer boundaries to density-weighted effective pressure, temperature and VMR profiles within the 38 layers that form the vertical retrieval grid used at TAO (see Section 4.2 for grid details). The program also calculates airmass factors for each layer. Pressure and temperature profiles are obtained from the National Centre for Environmental Prediction (NCEP) reanalyses available through the NASA Goddard Space Flight Center automailer (hyperion.gsfc.nasa.gov/Data_services/automailer/index.html), and are supplemented with the 1976 U.S. Standard Atmosphere above ∼45 km. We examined the effect of using *in situ* temperatures and pressures from the weather station installed at TAO, instead of the lowest NCEP profile values, on tropospheric (CO) and stratospheric (HCl) retrievals; retrieved total column differences were negligible for both species, while retrieved VMR profile differences were well below 0.5% and 0.01%, respectively, at all heights, which is in turn well below the error estimates shown in Section 4.3. Finally, VMR profiles used as inputs to *fscatm* were constructed from a variety of measurements and model outputs (see Section 4.1 and Figure 4.1 for *a priori* details). A single *a priori* profile was used for each gas for all seasons in order to avoid discontinuities in the retrieved profiles and column amounts, however, this approach can attenuate the strength of any seasonal cycles for that gas, see, e.g. *Schneider et al.* [2005a].

During the iterative retrieval, the model atmosphere is used in each iteration to calculate the infrared absorption spectrum of the target gas (plus any significantly absorbing interfering species) in one or several small bandpasses or *microwindows* (∼1 cm^{-1} wide). Pressure- and temperature- dependent absorption line parameters are obtained from the HITRAN 2004 spectral database [*Rothman et al.*, 2005]. The absorption cross-sections are calculated assuming a Voigt line shape on a grid of 0.0005 cm^{-1}.

The retrieval forward model includes a wavenumber scale multiplier, resolution, field-of-view, any apodization imposed digitally during the recording of the interferogram (none at TAO), number of microwindows, a pressure-induced line shift (taken from linelist by default), as well as four parameters to do with empirical instrument apodization and phase error specifications. The empirical parameters can be none (ideal ILS apart from OPD and FOV limitations), or a set of values as a function of OPD derived from LINEFIT, or a polynomial or Fourier series of a user-specified order. Model parameters that can be retrieved include a wavenumber shift (either individual or common in each bandpass), differential wavenumber shift for each retrieved molecule, a constant instrument phase error, or finally, polynomial coefficients for instrument phase error and apodization.

The simulated spectrum (forward model) can also include absorption features originating in the atmosphere of the Sun, whose line position shift can be retrieved. SFIT-2 uses the solar absorption spectrum of Frank Hase [*Meier et al.*, 2004b], which includes both theoretical line calculations and observed line information. In addition to solar features, continuous broadband extinction features due to aerosols in the Earth's atmosphere are also modeled by background slope and curvature parameters, which can both be retrieved. Finally, the spectrum can also exhibit channeling or a saturated line offset (both due to instrument imperfections); these characteristics are also included in the forward model and can both be retrieved.

Commonly, in a given microwindow we retrieve only the scaling factors applied to the a priori profiles of major interfering species, while full vertical profiles are only retrieved for the target gas. Thus, the retrieved state vector $\mathbf{x}$ consists of 38 target gas profile elements, usually a few scaling factors for profiles of major interfering gases, and a few retrieved forward model parameters such as wavenumber shift, background slope, solar line shifts, and phase error.

Finally, the user must construct the measurement noise and *a priori* covariance matrices ($\mathbf{S}_\epsilon$, $\mathbf{S}_a$) which are used in the iterative retrieval scheme (Equation 3.31-3.33). Together with the weighting function matrix, $\mathbf{K}$, $\mathbf{S}_\epsilon$ and $\mathbf{S}_a$ determine the quantity and vertical distribution of the independent pieces of information in the retrieval (Equation 3.44). Furthermore, the covariance matrices determine the error budget via the gain matrix (Equation 3.26). As such, the importance of their realistic estimation in retrievals and in retrieval characterization and error analysis is widely recognized, and they are discussed in more detail in Section 4.1.

3.4.1 State and measurement vector implementation

SFIT-2 has been designed to operate on a transmission spectrum scaled by the average radiance in a microwindow. The state vector used in the retrieval is also scaled by the *a priori* and only the relative scaling factor is retrieved, $\mathbf{x}^{relative} = \mathbf{x}/\mathbf{x}^{apriori}$, where the indicated division is an element-wise operation [*Wood*, 2004]. Furthermore, both $\mathbf{S}_\epsilon$ and $\mathbf{S}_a$ are specified as relative quantities, as a result of which $\mathbf{K}$, $\mathbf{G}$ and $\mathbf{A}$ are also relative. The relative state vector averaging kernel matrix, A^r_{ij}, is related to the VMR state vector averaging kernel matrix, A_{ij}, by

$$A^r_{ij} = \frac{\partial \hat{x^r_i}}{\partial x^r_j} = \frac{\partial(\hat{x}_i/x^a_i)}{\partial(x_j/x^a_j)} = \frac{x^a_j}{x^a_i}\frac{\partial \hat{x}_i}{\partial x_j} = \frac{x^a_j}{x^a_i}A_{ij}. \tag{3.59}$$

This implementation has an implication for the calculation of the total column averaging kernel vector, $\mathbf{a}_c$. We define the total column, c, as $\sum_i x_i N_i$, where N_i is the partial column of air [molec/cm^2] calculated by *fscatm* for each layer. The averaging kernel for the total column is defined as

$$(a_c)_j = \frac{\partial \hat{c}}{\partial(x_j N_j)} = \frac{\partial}{\partial x_j}\frac{1}{N_j}\sum_i \hat{x}_i N_i = \frac{1}{N_j}\sum_i N_i \frac{\partial \hat{x}_i}{\partial x_j}, \tag{3.60}$$

and by making use of Equation 3.59 we can equivalently write

$$(a_c)_j = \frac{1}{N_j x^a_j}\sum_i N_i x^a_i A^r_{ij}. \tag{3.61}$$

To calculate the averaging kernel vector of a *partial* column, we multiply the *a priori* state vector (element-wise) with an operator containing ones in the partial column region of interest and zeros everywhere else. In analogy to a total column, we can also define a height-averaged volume mixing ratio

$$v = \frac{\sum_i x_i \Delta z_i}{\sum_i \Delta z_i} \tag{3.62}$$

and its corresponding averaging kernel vector, $\mathbf{a}_v$, as

$$(a_v)_j = \frac{\partial \hat{v}}{\partial x_j} = \frac{\partial}{\partial x_j}\frac{\sum_i \hat{x}_i \Delta z_i}{\sum_i \Delta z_i} = \frac{1}{\sum_i \Delta z_i}\sum_i \frac{\partial \hat{x}_i}{\partial x_j}\Delta z_i = \frac{1}{x^a_j \sum_i \Delta z_i}\sum_i x^a_i \Delta z_i A^r_{ij}, \tag{3.63}$$

where we have again made use of Equation 3.59. If the full averaging kernel matrix is ideal ($\mathbf{A} = \mathbf{I}$), then

$$(a_v)_j = A^r_{ij} \frac{\Delta z_i}{\sum_i \Delta z_i}, \tag{3.64}$$

which will only equal unity if the volume mixing ratio is averaged over one layer ($(a_v)_j = 0$ in all other layers). This is in contrast to the total or partial column averaging kernel vector, which will contain ones in the region of interest and zeros everywhere else in the case that $\mathbf{A} = \mathbf{I}$. Examples of both partial column and height-averaged VMR averaging kernels are shown in Section 4.3 and 5.2.

3.4.2 Auxiliary and retrieval software versions

The above description applies to *fscatm* v.2.05 and SFIT-2 v.3.91, although TAO was originally furnished with custom releases of *fscatm* v.2.03 and SFIT-2 v.3.81 by Dr. Arndt Meier. The version traceability and release documentation of both codes has greatly improved in recent years due to the efforts of the NDACC IRWG SFIT-2 development team. Nonetheless, we briefly recount the chronology of major code revisions for reference. First, the major *fscatm* code improvements introduced in v.2.03 are described in detail by *Meier et al.* [2004a]. The only functional improvement since this release has been the implementation of a flexible model atmosphere lid instead of the hardwired value of 100 km (the new lid value is written on the first line of the *zpt* file output by *fscatm*). This improvement was accommodated in SFIT-2 v.3.81, which also introduced two other major functional improvements. First, the use of *pseudolines* was implemented for 11 heavy molecules whose spectroscopic data are not entirely satisfactory at this point (N_2O_5, $ClONO_2$, CF_4, F12, F11, CCl_4, F22, $COCl_2$, SF_6, F142b, F113). Since pseudolines are an empirical spectroscopic solution (beyond the scope of this discussion), they render a characterization of profile retrievals meaningless, unless averaging kernels are calculated manually by perturbation methods. Second, v.3.81 included for the first time a calculation of self-broadening for all molecules, with implications for the forward modeling of water vapour, especially at low-altitude observation stations such as TAO. Next, SFIT-2 v.3.82 included many organizational and syntactic improvements, as well as a transition of the code from FORTRAN77 to FORTRAN90. Major functional improvements include the modeling and retrieval of channeling parameters, and the ability to incorporate LINEFIT ILS parameters in the forward model. Next, SFIT-2 v.3.90 introduced the new solar spectrum model of Frank Hase (see above) and changes to the *a priori* and measurement noise covariance matrices. First, the solar model used up to this point employed five retrievable parameters and is described in detail by *Rinsland et al.* [1998]. Second, the option to use exponential off-diagonal elements of a user-specified correlation length was added to the existing option of Gaussian off-diagonal elements in the *a priori* covariance matrix. Furthermore, the option to use a full matrix (or its inverse) was also added. Third, in addition to a static measurement noise covariance, the option to use a single scaling factor was added, which multiplies the noise derived from the residual between the first iteration ($\mathbf{y}_a$) and the measurement ($\mathbf{y}$), and thus allows the noise to be variable over an ensemble of retrievals. Finally, SFIT-2 v.3.91 (currently used at TAO) continues to improve on the solar model of Frank Hase (present linelist: 041123_S.TMP) and the input/output and fitting functionality of previous versions. The program's text control file (*binput*) is now sufficiently different to warrant a redesign of the auxiliary programs *prepare* and *specasc*, which has been performed by Dr. Arndt Meier and tested by the candidate.

3.4.3 Retrieval software outstanding issues

The SFIT-2 retrieval algorithm is a research tool whose functionality is continuously being expanded by the SFIT-2 development team within the IRWG of NDACC. The following sections describe new software and implementation issues that have been identified during the course of this thesis that have some impact on the characterization and retrieval of vertical profiles from ground-based FTIR spectra.

	[0-100 km]		[0-15 km]		[15-50 km]	
	$\mathbf{K}_a$	$\mathbf{K}_n$	$\mathbf{K}_a$	$\mathbf{K}_n$	$\mathbf{K}_a$	$\mathbf{K}_n$
Mean d_s	2.010	2.020	0.098	0.098	1.910	1.920
StDev	0.201	0.207	0.028	0.029	0.171	0.176
Mean S_s	2.580	2.580	18.30	18.40	1.280	1.290
StDev	0.154	0.158	0.450	0.463	0.182	0.184
Mean S_m	1.830	1.840	4.210	4.230	1.780	1.790
StDev	0.254	0.255	0.442	0.448	0.326	0.328
Mean %a	13.10	13.10	83.50	83.40	0.266	0.303
StDev	0.668	0.691	3.060	3.200	0.553	0.543
Mean S_T	2.320	2.310	0.789	0.767	2.710	2.680
StDev	0.149	0.144	0.123	0.097	0.149	0.151

Table 3.1 Effect of using $\mathbf{K}_a$ vs. $\mathbf{K}_n$ in offline vs. online characterization and error analysis on 2004 O_3 retrieval diagnostics (i.e. scalars in the case of total and partial columns) of 63 spectra. %a represents "% *a priori* contribution", and S_T is the temperature error.

3.4.3.1 Online vs. offline characterization and error analysis

The Matlab and IDL characterization and error analysis codes distributed in the NDACC community perform an off-line error analysis. That is, the codes use a pre-existing SFIT-2 forward model of the atmosphere (density-weighted pressure, temperature and VMR) to simulate a spectrum (using additional forward model parameters specified in the *binput* control file) and produce weighting functions for the zeroth retrieval iteration ($\mathbf{x}_a$, $\mathbf{y}_a$ and $\mathbf{K}_a$). The assumption implicit in this approach is that the problem is nearly linear, such that $\mathbf{K}_a \simeq \mathbf{K}_n$, where $\mathbf{K}_n$ represents $\mathbf{K}$ evaluated at the retrieved solution $\mathbf{x}_n$, as described in Equation 3.31. However, this assumption contradicts the use of an iterative retrieval in the first place. While the TAO characterization and error analysis codes have been implemented following the NDACC approach of offline error analysis, there is no functional limitation on performing the error analysis online, i.e. it is possible to instruct SFIT-2 to output weighting functions after the last iteration. Table 3.1 shows differences in total and partial column retrieval diagnostics produced with $\mathbf{K}_a$ and $\mathbf{K}_n$ for an ensemble of 63 O_3 retrievals for 2004. It is clear that the O_3 retrieval is closer to nearly linear rather than to moderately nonlinear, since the mean values are very similar for $\mathbf{K}_a$ and $\mathbf{K}_n$; as expected, there is marginally more scatter in mean values produced with $\mathbf{K}_n$. The two approaches to error analysis will produce diverging results as nonlinearity in the forward model grows (e.g. due to saturated absorption features) and also as the number of iterations in the retrieval increases (e.g. due to increased measurement noise).

3.4.3.2 Scaling of the state vector

In order to calculate $\mathbf{G}$ and $\mathbf{A}$ (Equations 3.26 and 3.9), the matrix $\mathbf{X} = \mathbf{S}_a^{-1} + \mathbf{K}^T\mathbf{S}_\epsilon^{-1}\mathbf{K}$ must be inverted, given SFIT-2 outputs of $\mathbf{S}_a^{-1}$, $\mathbf{S}_\epsilon^{-1}$, and $\mathbf{K}$. This inversion often triggers a Matlab warning about the conditioning of $\mathbf{X}$, i.e. that the ratio of its lowest to highest eigenvalue ($c = \lambda_{low}/\lambda_{hi}$) is of the order of machine precision ($\simeq 2.2^{-16}$), and that consequently $\mathbf{X}^{-1}$ may be inaccurate. Technically, if $c \sim 10^{-k}$, then the last k digits of of X_{ij} may be inaccurate. Since c ranges from 5.3^{-14} (for NO) to 7.7^{-17} (for HF), this appears to imply severe numerical computational errors for the 15-digit double-precision numbers used in the characterization and error analysis of TAO retrievals. Figure 3.3 shows the eigenvalue spectrum of the best- and worst-conditioned $\mathbf{X}$ encountered in all 2004 HF retrieval diagnostics calculations, which contains smoothly decreasing eigenvalues instead of abrupt truncations that could signal numerical problems. This is a hint that the Matlab condition number, c, is a pessimistic indicator of inversion accuracy.

Fortuitously, HF retrievals of non-TAO spectra performed with PROFFIT-9 were available for comparison, and the condition number of $\mathbf{X}$ was found to be nearly seven orders of magnitude higher at 2.2^{-9}. Since $\mathbf{S}_a^{-1}$ and $\mathbf{S}_\epsilon^{-1}$ were both specified diagonal (making them trivial to invert), the form of $\mathbf{K}^T\mathbf{K}$ was investigated further in both SFIT-2 and PROFFIT-9. It was revealed that the dynamic range of the diagonal elements of $\mathbf{K}^T\mathbf{K}$ (Figure 3.4) is also seven orders of magnitude less in PROFFIT-9 than in SFIT-2. In particular, the SFIT-2 wavenumber shift forward model parameter is responsible for

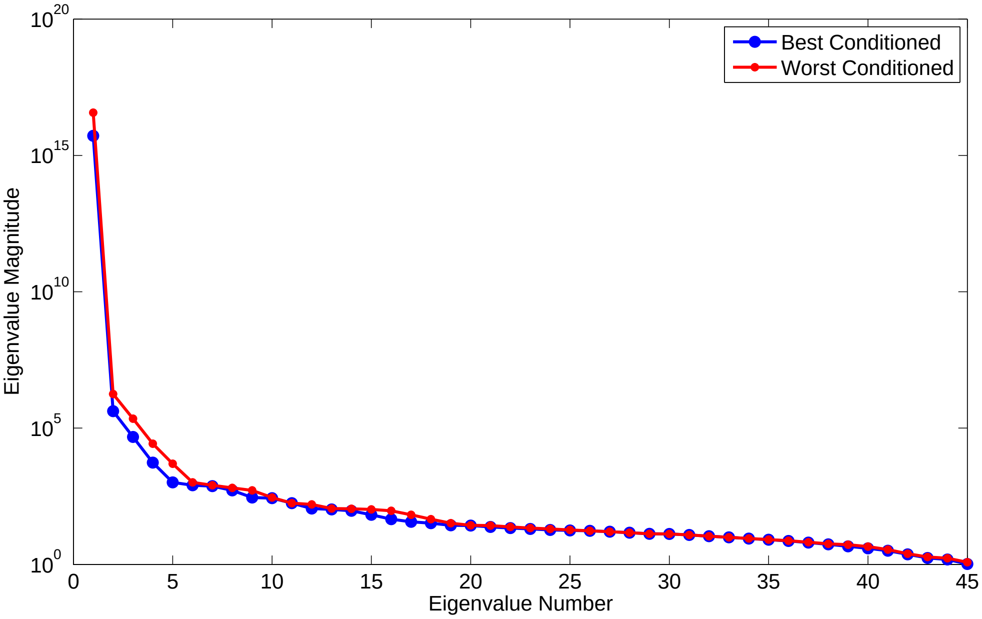


Figure 3.3 Eigenvalue Spectrum of $\mathbf{X}$ (see text for details) for the best- and worst-conditioned inversions in 2004 TAO HF retrieval diagnostics calculations.

the majority of this difference. While a high dynamic range of the diagonal elements of $\mathbf{K}^T\mathbf{K}$ is not sufficient to produce a poor condition number, in the presence of off-diagonal elements this will be the case. Figure 3.5 shows the large off-diagonal elements of $\mathbf{K}^T\mathbf{K}$ (re-scaled by taking the logarithm of the absolute value for increased contrast), which project the wavenumber shift parameter ($\mathrm{K}^T\mathrm{K}_{3,3}$) on other elements of the state vector, most notably on the vertical profile of HF and on the vertical profile scaling factor for the column fit of H_2O. The implication of this preliminary analysis is that only certain off-diagonal elements of $\mathbf{X}^{-1}$ are inaccurate given a poor condition number of $\mathbf{X}$. From Figure 3.5 we can see that the block matrix pertaining to the vertical profile elements used in the characterization and error analysis of TAO retrievals is likely well-conditioned, thus producing accurate retrieval diagnostics. However, if one wished to calculate the *interference error* between the vertical profile and other state vector elements (described by *Rodgers and Connor* [2003]), this error would likely not be accurate given the current scaling of the complete state vector. Fortunately, this problem can be overcome to a

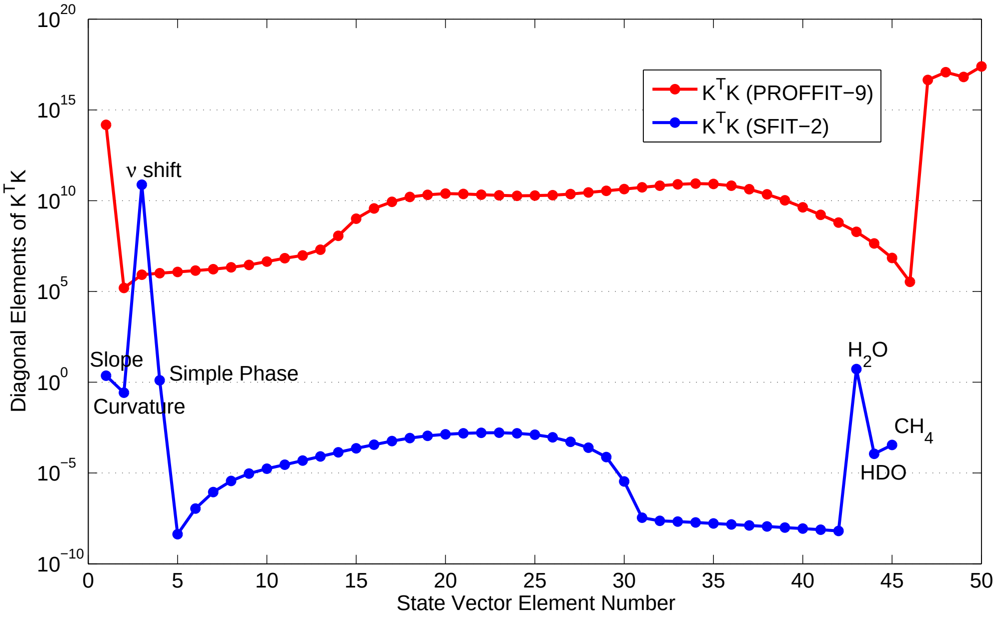


Figure 3.4 Diagonal elements of $\mathbf{K}^T\mathbf{K}$ for a PROFFIT-9 and SFIT-2 retrieval of HF. PROFFIT-9 has 1 "lead" and 4 "tail" state vector elements related to interfering species and other forward model parameters, while SFIT-2 has 4 "lead" and 3 "tail" elements.

large extent by re-scaling only the wavenumber shift parameter in SFIT-2. Since $\mathbf{X}$ is also inverted at each iteration in the retrieval process (Equation 3.32), the effect of such a change on the complete retrieved state vector of profile elements and forward model parameters ($\mathbf{x}$ and $\mathbf{b}$) should be examined.

3.4.3.3 Degrees of freedom for signal and *T15ASCII* file resolution

In the course of this work it was discovered that calculated d_s values depend on the resolution parameter found in the *T15ASCII* file that contains the SFIT-2 input spectrum, rather than on the resolution derived from the OPD parameter in the SFIT-2 *binput* control file (1/OPD). Since the *T15ASCII* file contains a spectrum that is zero-filled, its apparent resolution will always be higher than that derived from the *binput* file, resulting in higher d_s values. If the apparent resolution is described by 1/(X*OPD), then Figure 3.6 shows the effect of increasing X on d_s for a sample NO_2 retrieval. The increase of

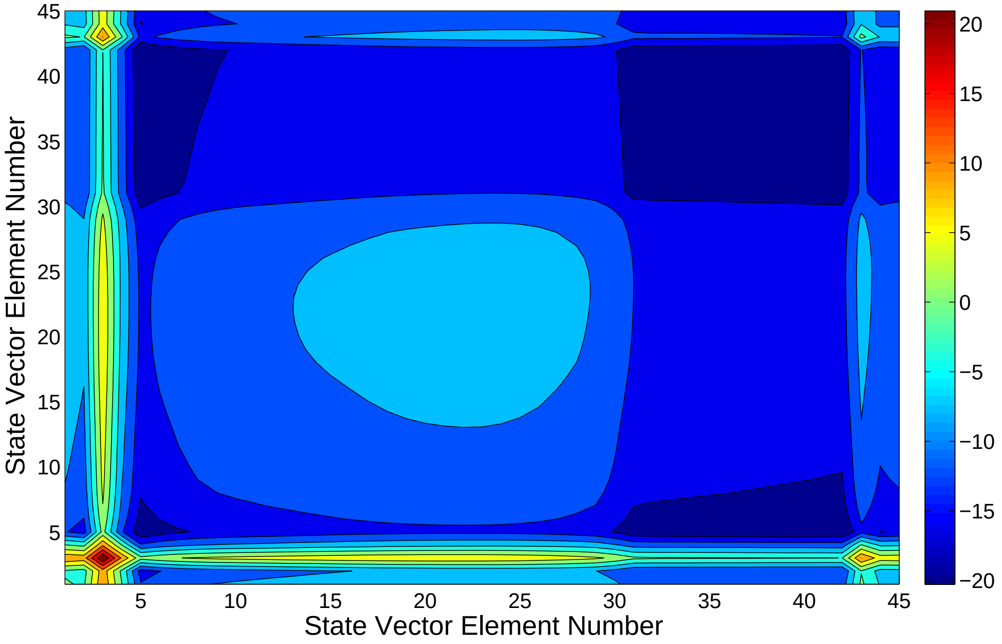


Figure 3.5 The natural logarithm of the absolute value of $\mathbf{K}^T\mathbf{K}$ for the best-conditioned TAO retrieval diagnostics calculation for HF in 2004, showing large off-diagonal amplitudes (see text for details). Plot origin is at (1,1).

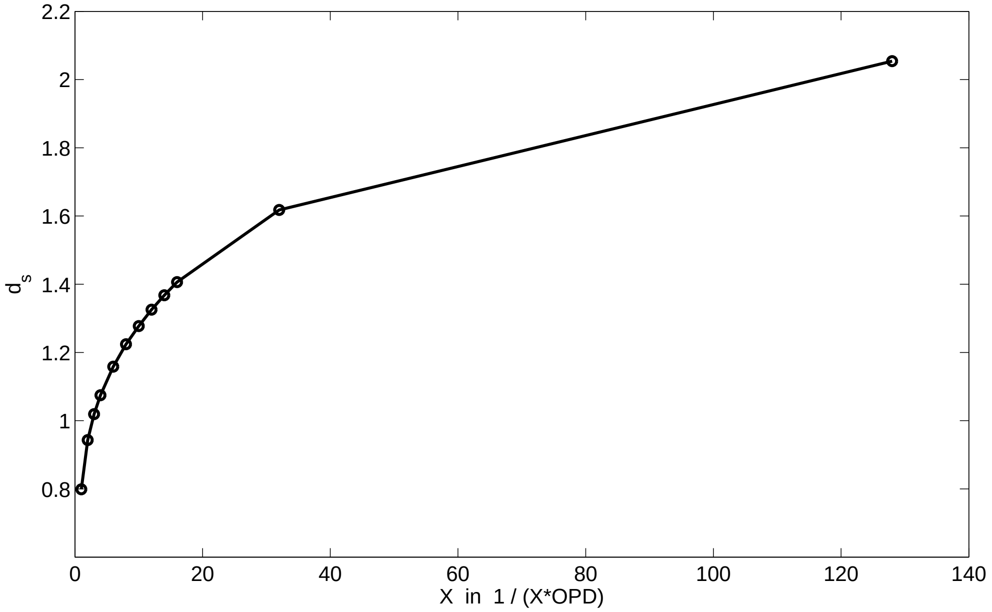


Figure 3.6 Increasing d_s for an increasing *T15ASCII* file resolution [1/(X*OPD)] used in the characterization of an NO_2 retrieval.

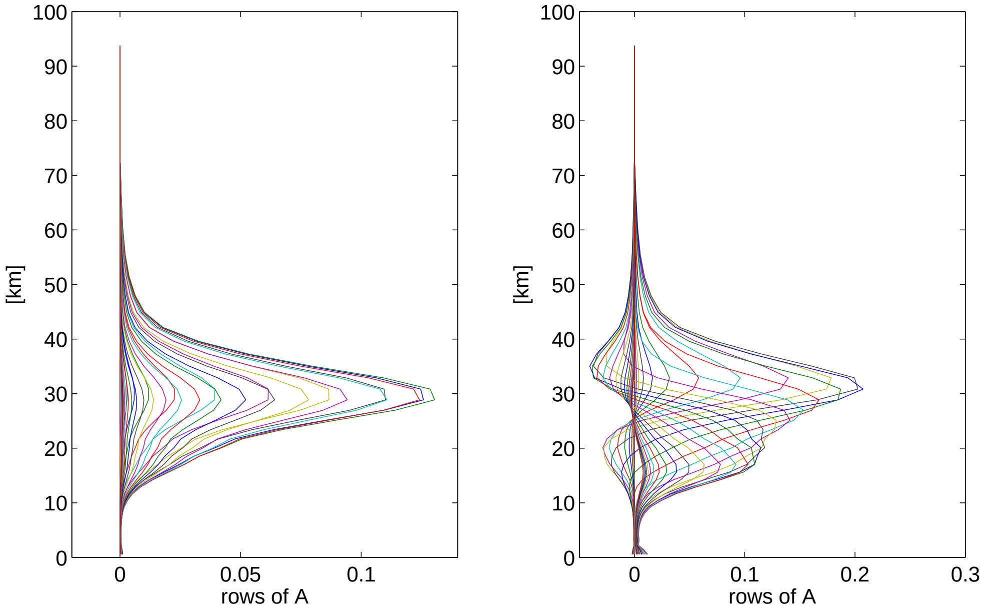


Figure 3.7 Increasing vertical resolution corresponding to X=1 (left) and X=128 (right) in Figure 3.6.

d_s appears unbounded and the averaging kernels corresponding to X=128 (Figure 3.7) show the development of a consistent second peak. X ranges from 2-4 for TAO filters 1-6, since each filter requires a different amount of zero-filling. The characterization of 2004 retrievals of seven gases submitted to the NDACC database in 2005 was carried out for both X=1 and X=[2-4], as appropriate for the filter used to record the corresponding spectrum. It was found that when calculated with X=1, d_s decreased as follows: N_2O 15%, CH_4 15%, O_3 22%, NO 22%, NO_2 24%, HF 13%, and HCl 18%. Additionally, the total column smoothing and measurement errors increased by up to 79% and 63%, respectively. This Fortran source code issue has been brought to the attention of the SFIT-2 development team within the NDACC, wherein the common approach is to perform all characterizations on simulated spectra and using X=3. The d_s values shown in this work are derived from the characterization of real spectra, where X=[2-4].

3.4.3.4 RMS of fit residual

The final issue is simply a note of caution for future reference. The previously-described *T15ASCII* file contains the absolute intensity values of the recorded spectrum in a given microwindow, which SFIT-2 then scales by the *average* intensity in this microwindow before performing a retrieval. After performing the retrieval, the RMS of the final iteration is written to the *detail* file, which records all information necessary to reconstruct a given retrieval, while the final simulated and measured spectra are written to a graphic *pbpfile*, after being scaled again, this time by the *highest* value in the microwindow. When the average value in a microwindow differs from the highest value (as is the case for broad absorption features), the RMS derived from each file will also differ. Independent of this difference, the RMS derived from the *detail* file spans all microwindows used in a retrieval, whereas the RMS derived from the *pbpfile* is usually different for each microwindow. This may cause problems if a retrieval is performed by combining two microwindows sufficiently far apart in the spectrum that their SNR values are significantly different from the average. The higher SNR microwindow will be under-fitted, while the lower SNR microwindow will be over-fitted, likely leading to spurious profile oscillations in the one retrieved solution. The 2001-2004 data submitted to the NDACC database were retrieved using individual microwindows, except for CH_4, where three separate microwindows were used; however, the SNR in these three separate microwindows was found to be comparable.

Chapter 4

Retrieval Constraints, Characterization, and Error Analysis

We now describe the construction of and some effects of the retrieval constraints that are central to our inverse problem: $\mathbf{x}_a$, $\mathbf{S}_a$, and $\mathbf{S}_\epsilon$. The most important function of $\mathbf{x}_a$ is to reasonably constrain the null space and the *near* null space of the solution. If the choice of *a priori* in this region of state space is incorrect, so will be the retrieved solution and its error estimates. This is further illustrated with retrievals of NO and NO_2 (Section 4.3.2 and 4.3.3). While we can determine the sensitivity of a retrieval to the *a priori* (Equation 3.10), the effects of an incorrect $\mathbf{S}_a$ and $\mathbf{S}_\epsilon$ on a retrieval must be assessed by numerical simulations. This is briefly illustrated using retrievals of O_3 from simulated spectra (Sections 4.1.2 and 4.1.3). In addition to these *explicit* retrieval constraints, in Section 4.2 we show some effects of the constraint *implicit* in the chosen state vector representation, i.e. the vertical layering scheme of the retrieval. Finally, in Section 4.3 we present the characterization and error analysis of the retrievals of eight major trace gases at TAO.

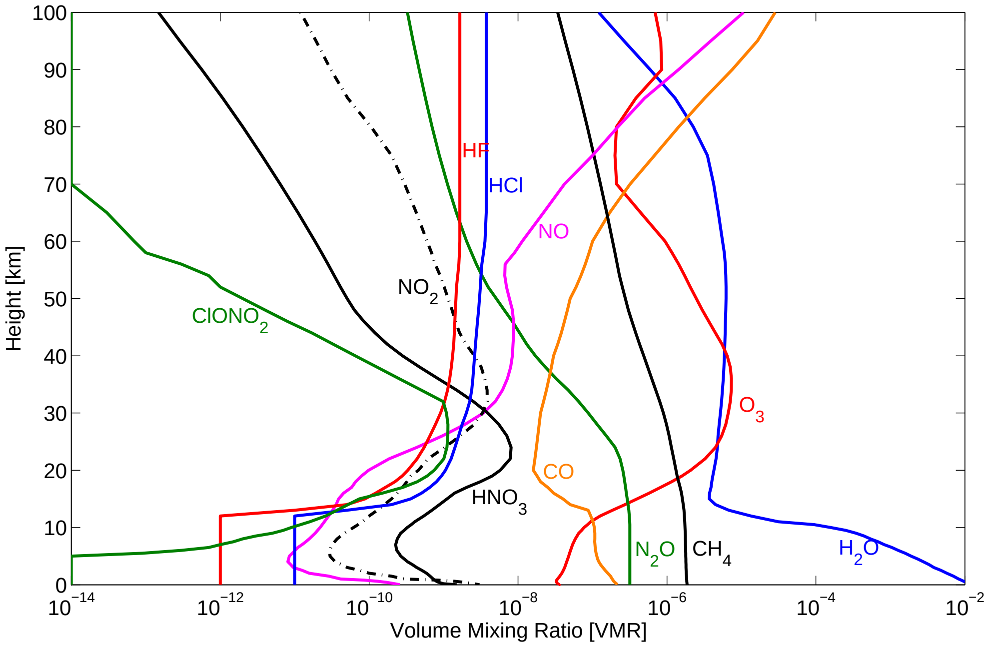


Figure 4.1 A priori VMR profiles used in TAO retrievals for the 10 NDACC target gases and CO, based on a combination of sources discussed in Section 4.1.

4.1 Explicit Constraints

4.1.1 Optimization of *a priori* profiles ($\mathbf{x}_a$)

TAO *a priori* profiles (Figure 4.1) were constructed from validated Upper Atmosphere Research Satellite (UARS) HALogen Occultation Experiment (HALOE) measurements [*Russell et al.*, 1993], where available, and supplemented with:

- v.3.0 mid-latitude daytime MIPAS [*Carli et al.*, 2004] reference climatology profiles constructed in 2001 and available from www.atm.ox.ac.uk/group/mipas/species,
- Jet Propulsion Laboratory MkIV FTIR balloon flight profiles made available at mark4sun.jpl.nasa.gov/balloon.html by Dr. Geoffrey Toon,
- *a priori* profiles constructed by Dr. Arndt Meier in 1998 for Complementary NDACC station Kiruna and provided for TAO in 2002.

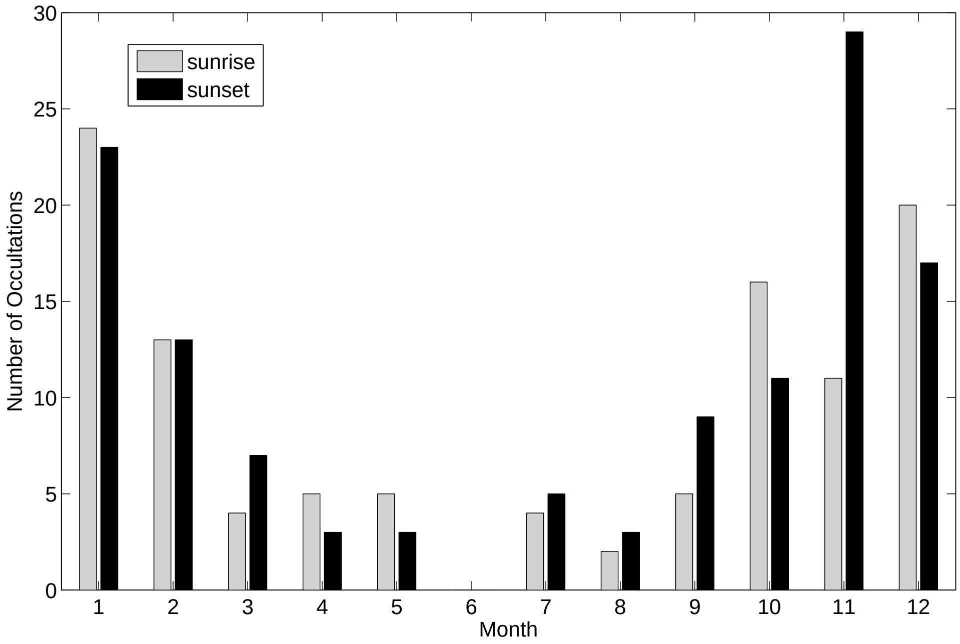


Figure 4.2 HALOE monthly coverage within a 500-km radius of Toronto (1991-2004).

HALOE data (v.19) was downloaded from haloedata.larc.nasa.gov/overpass/index.php and Matlab scripts were written to manipulate it, as well as the MIPAS, MkIV and Kiruna data sets. Beside VMR profiles, retrieved tropopause height was also extracted from the HALOE data, along with potential temperature, aerosol composition (% H_2SO_4), and aerosol concentration (1/cm^3). The Matlab scripts can be modified easily to extract aerosol density (g/cm^3), mean radius, distribution width, surface area, volume (μm^3/cm^3), effective radius, and finally, extinction broken down by the 5.26 μm, 3.46 μm, 3.40 μm, and 2.45 μm channels. The temporal coverage of the HALOE data set is heavily skewed towards the winter (Figure 4.2), therefore, seasonal means were produced first and then averaged to give an annual mean (sunset profiles were used due to their higher SNR). The annual mean tropopause height retrieved above Toronto was constructed in the same manner and was found to be 11.7 km for sunrise and 11.9 km for sunset data (Figure 4.3).

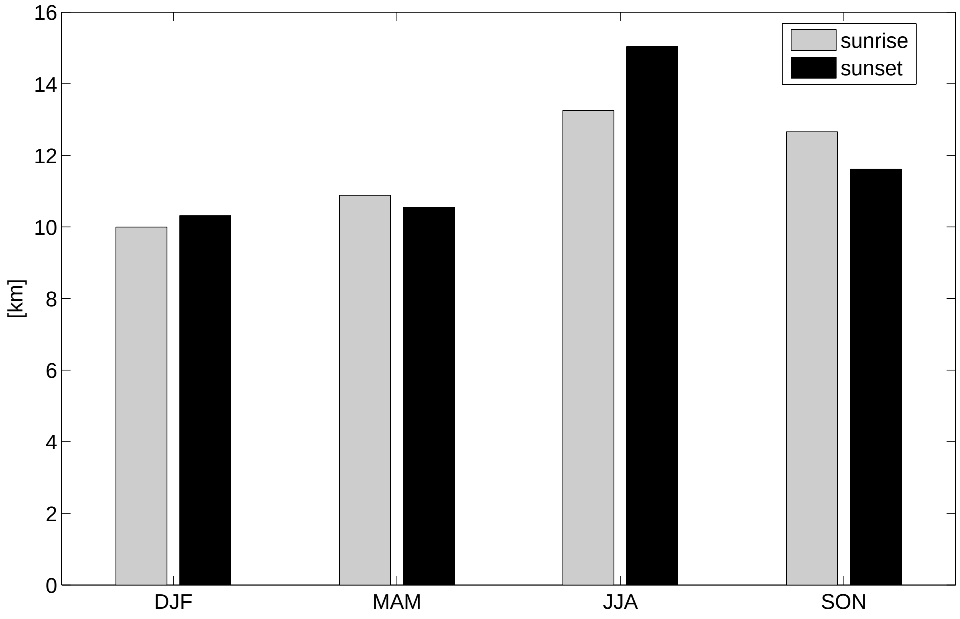


Figure 4.3 Seasonal means of tropopause heights retrieved from 1991-2004 HALOE data within a 500-km radius of Toronto.

It should be noted that prior to taking the seasonal averages all HALOE data analyzed in this work were filtered such that only retrieved quantities a factor of two bigger than their error estimates were considered, as per the v.2.0 HALOE data description document. This document also states that error estimates only consider noise error and error due to the aerosol correction, and become inaccurate below ∼20 km. Finally, the filtered data set was also interpolated onto the 0-100 km 63-layer grid used as input to the *fscatm* program, which in turn calculates density-weighted temperatures, pressures, volume mixing ratios, and airmass factors on the 0-100 km 38-layer grid used in SFIT-2 retrievals (Figure 4.4).

Volume mixing ratio profiles of O_3, NO, NO_2, HCl, and HF in the altitude region between approximately 15 and 65 km (up to 130 km for NO) were constructed from 123 HALOE solar sunset occultation profiles recorded between 1991 and 2004 and within a 500-km radius of TAO. Outside of this altitude region, the *a priori* profile of each species

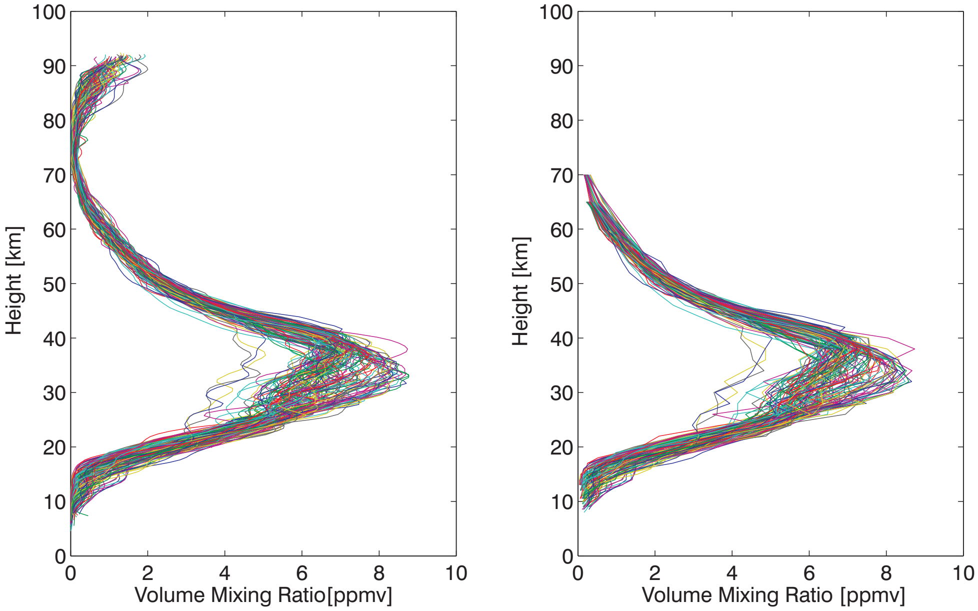


Figure 4.4 (left) All HALOE O_3 profiles within a 500-km radius of Toronto (1991-2004). (right) HALOE O_3 profiles filtered for random errors $< 50\%$ and interpolated onto a 63-layer grid.

was treated differently, as described below. The *a priori* profiles of the remaining major TAO target species are also discussed.

Ozone (O_3)

The final ozone profile adopted at TAO is shown in Figure 4.5. Outside of the altitude range covered by HALOE (15-70 km), the MIPAS profile was used, except below 1 km (Figure 4.6), where the MIPAS VMR was smoothly increased to the annual surface mean over Toronto (36 ppbv) obtained from the GEOS-CHEM v.5-04 model [*Bey et al.*, 2001], available at www.atmos.washington.edu/~jaegle/geoso3_start.html. As will be shown in Section 4.3, the retrieval is not sensitive to the true O_3 profile in the troposphere, therefore, this change only reduces the total column bias, given TAO's urban location. Finally, Figure 4.7 shows the range of O_3 profiles retrieved from all available MkIV balloon flights, together with their locations, for future reference.

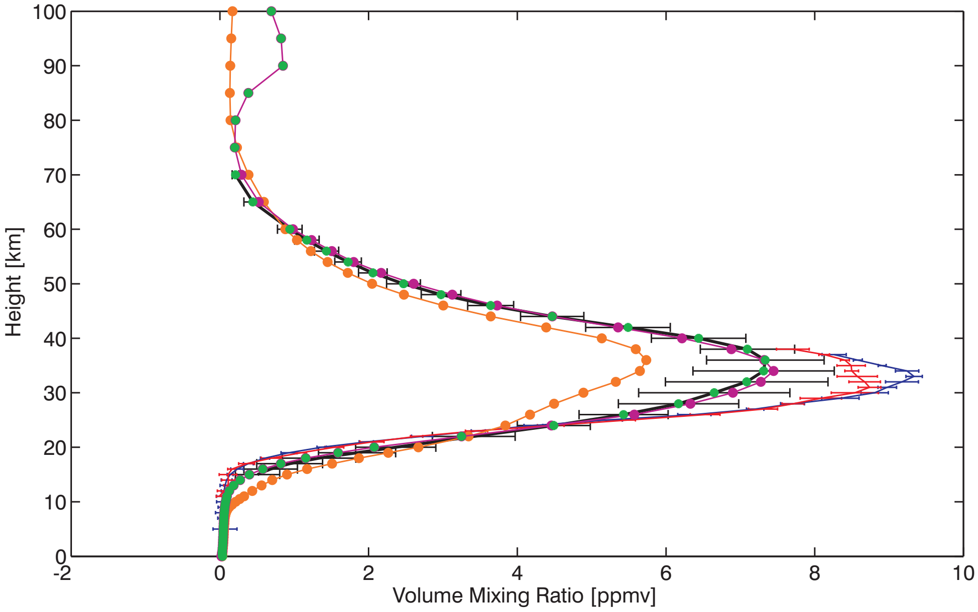


Figure 4.5 O_3 *a priori* profiles from Kiruna (orange), HALOE mean (black, with 1σ “error bars” showing variability), MIPAS (purple), 2003-2004 Ft. Sumner MkIV balloon flights (blue, red, with error bars showing calculated retrieval errors), and TAO (green).

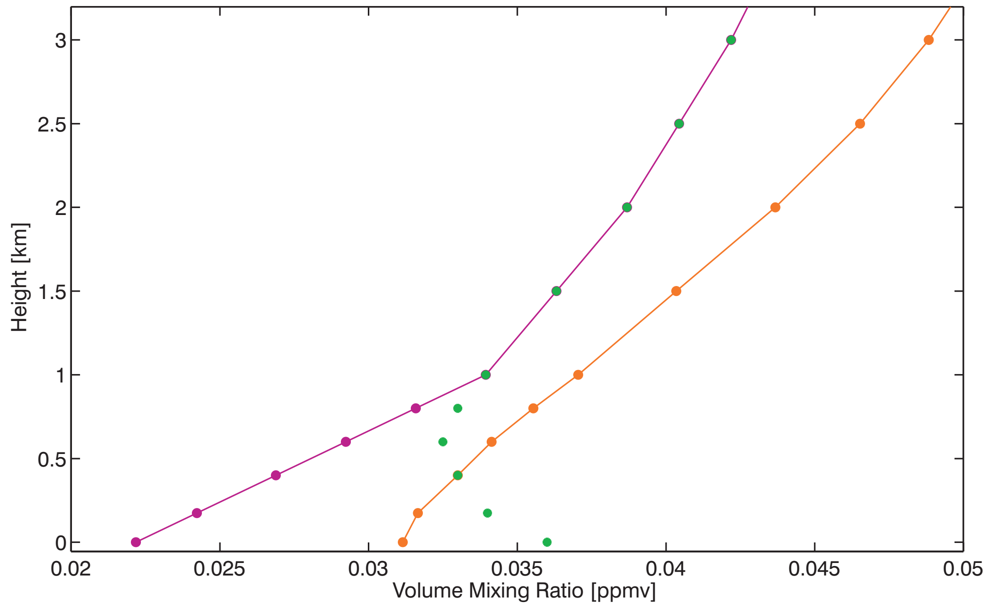


Figure 4.6 Surface values of O_3 *a priori* profiles from Kiruna (orange), MIPAS (purple), and TAO (green).

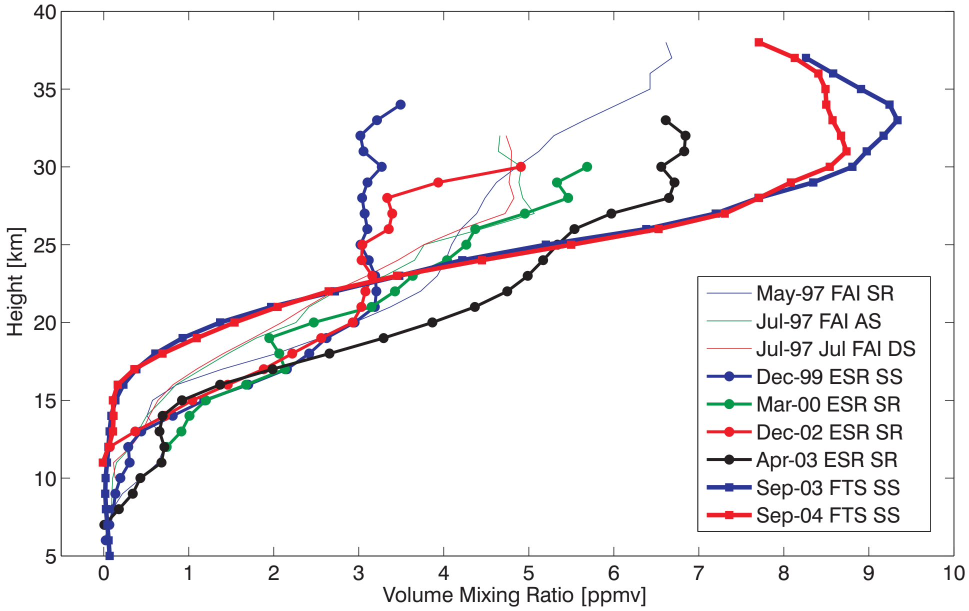


Figure 4.7 O_3 profiles retrieved from all available MkIV balloon flights. FAI = Fairbanks (AK), ESR = Esrange (Sweden), FTS = Ft. Sumner (NM), SR = sunrise, SS = sunset.

Nitric Oxide (NO)

The nitric oxide *a priori* profile was constructed following the same approach as for O_3. The final profile adopted at TAO (Figure 4.8) consists of the sunset HALOE profile (20-130 km) smoothly connected to the MIPAS profile (0-7 km). The retrieval grid was extended up to 130 km in the case of NO because a significant portion of the total column is found between 100-130 km (see Section 6.1 for details). Between 60-90 km the HALOE sensitivity to NO is low and none of the data met the 50% relative error criterion, hence the values were smoothly interpolated on a logarithmic scale. Finally, the MIPAS profile displays tropospheric NO mixing ratios that are more appropriate for TAO's urban location than those in the Kiruna profile. As is the case for O_3, the NO retrieval is not sensitive to the true profile in the troposphere, hence the *a priori* profile chosen in this region determines the total column bias.

Nitrogen Dioxide (NO_2)

The nitrogen dioxide *a priori* profile adopted at TAO is shown in Figure 4.9. It consists

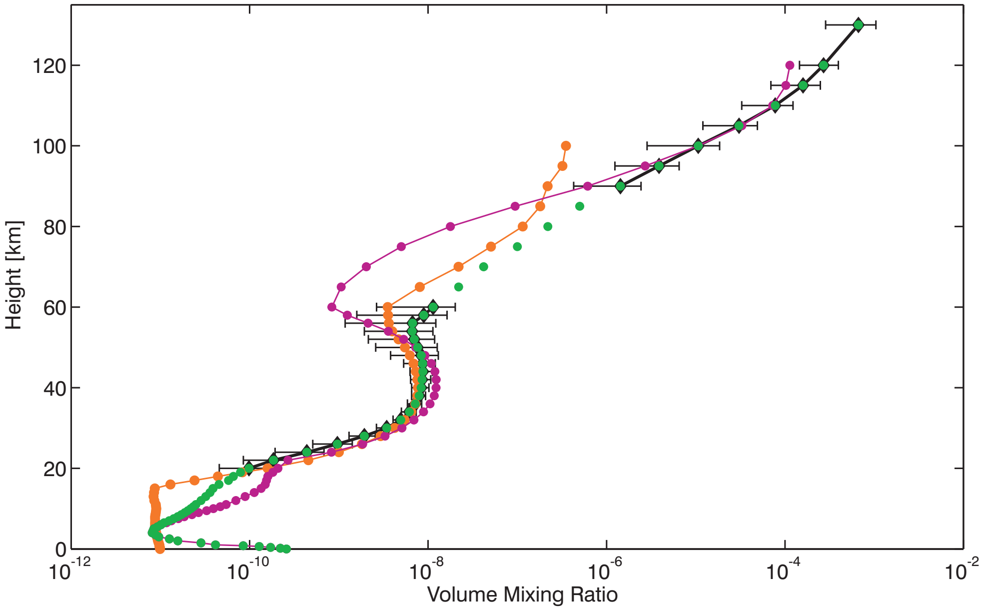


Figure 4.8 NO *a priori* profiles from Kiruna (orange), HALOE mean (black), MIPAS (purple), and TAO (green).

of the mean sunset HALOE profile (19-44 km) smoothly joined to the daytime MIPAS reference profile (0-15 km and 46-100 km). Since NO_2 has a very strong diurnal cycle involving conversion to and from NO, this *a priori* profile may need to be scaled to midday values using a photochemical box model in order to be truly optimal for TAO.

Hydrogen Chloride (HCl)

The hydrogen chloride *a priori* profile adopted at TAO is shown in Figure 4.10. It consists of the HALOE profile between 15-58 km, the 60 km HALOE profile value at all grid points above this altitude, and 10^{-11} in the troposphere (0-12 km). The tropospheric value is based on the extensive description of HCl retrievals presented in *Rinsland et al.* [2003]. (The MIPAS reference climatology does not include HCl.)

Hydrogen Fluoride (HF)

The hydrogen fluoride *a priori* profile adopted at TAO is shown in Figure 4.11. It consists of the HALOE profile between 15-54 km, the 60 km HALOE profile value at all grid points above this altitude, and 10^{-12} in the troposphere (0-12 km). The tropospheric value is

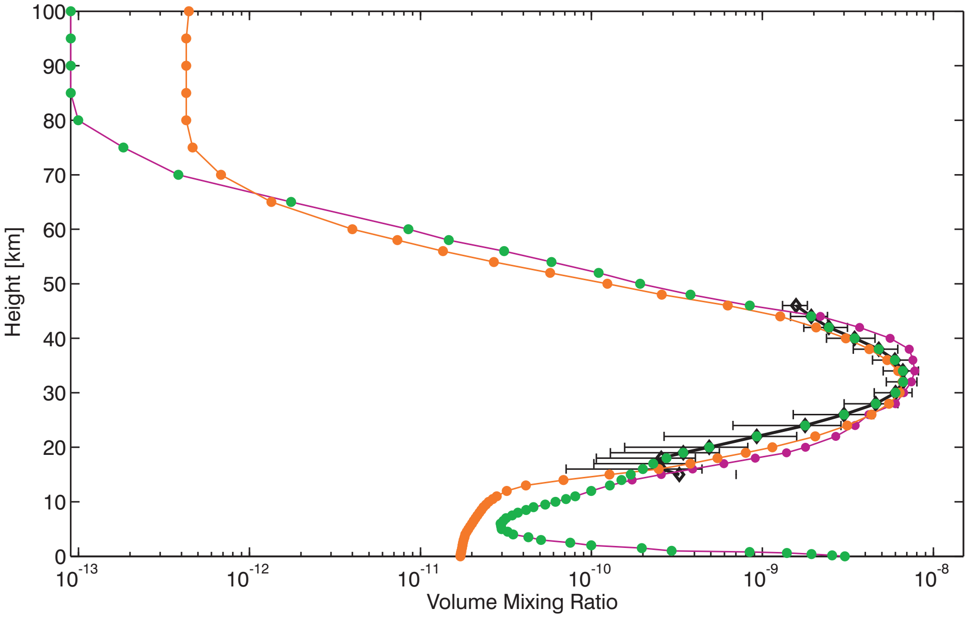


Figure 4.9 NO_2 *a priori* profiles from Kiruna (orange), HALOE mean (black), MIPAS (purple), and TAO (green).

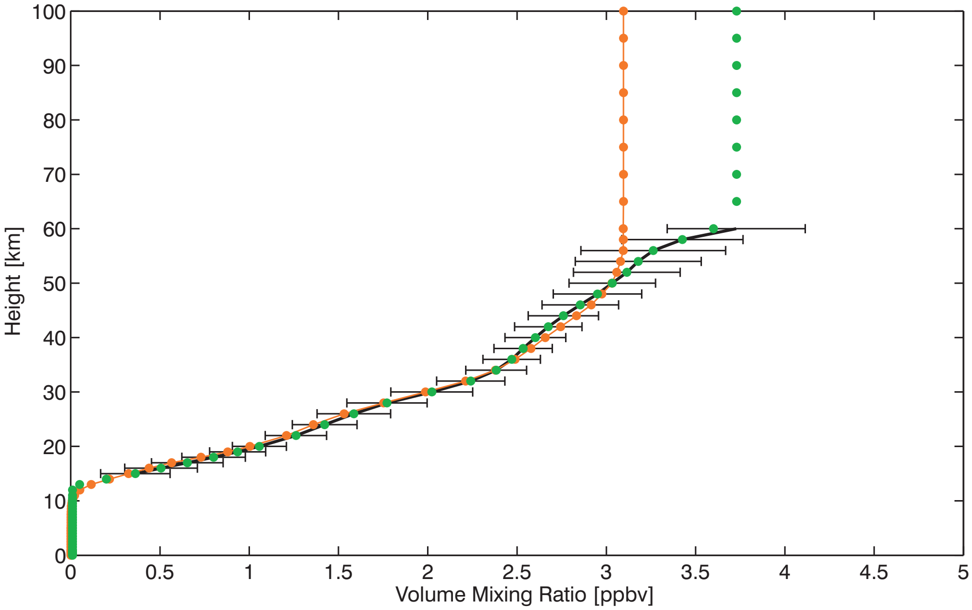


Figure 4.10 HCl *a priori* profiles from Kiruna (orange), HALOE mean (black), and TAO (green).

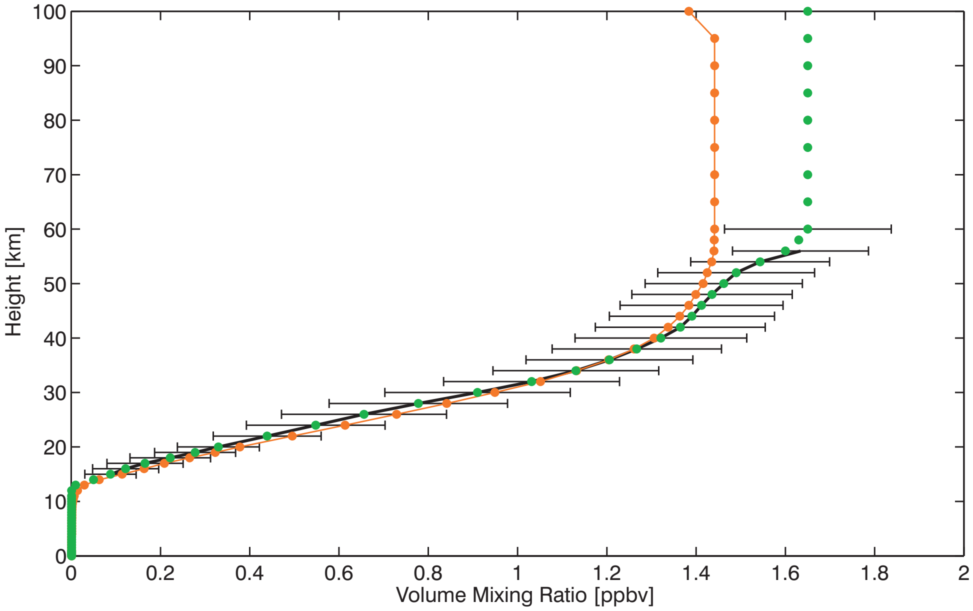


Figure 4.11 HF *a priori* profiles from Kiruna (orange), HALOE mean (black), and TAO (green).

based on HF retrievals described by *Rinsland et al.* [2002a]. (The MIPAS reference climatology does not include HF.)

Water vapour (H_2O)

As expected, the water vapour *a priori* profile constructed for Kiruna was too dry for Toronto in the troposphere, and its tropopause height was also too low. Lacking an ensemble of tropospheric water vapour soundings in the vicinity of downtown Toronto, the Kiruna profile was replaced with the MIPAS profile throughout the atmosphere (Figure 4.1). While the vast majority of the total column of water vapour is found in the troposphere, we note that the MIPAS profile was found to be nearly identical to the HALOE profile between 15-70 km – much better than within one standard deviation of the 2001-2004 HALOE mean. An alternative approach to retrievals of water vapour is to use daily *a priori* profiles from NCEP analyses as a pseudo-measurement above Toronto; this can be helpful when accurate water vapour results are required in order to improve the retrievals of other gases, however it was not used in this work.

Methane (CH_4)

Again, as expected, the methane *a priori* profile constructed for Kiruna was too clean in the troposphere, and its tropopause height was again too low. It too, was replaced with the MIPAS profile throughout the atmosphere, as it contains higher methane volume mixing ratios throughout the troposphere (∼1.8 ppmv), and further elevated values near the surface (∼1.9 ppmv), consistent with TAO's urban location (Figure 4.1). As was the case for water, the MIPAS and HALOE methane profiles were also found to be nearly identical between 15-70 km.

Nitrous Oxide (N_2O)

The Kiruna nitrous oxide *a priori* profile was replaced with a modified MIPAS profile throughout the atmosphere, which has a tropospheric N_2O concentration of 317 ppbv (Figure 4.1). The global mean surface N_2O VMR was reported as 314 ppbv 1998, with an annual increase of 0.8 pp [*IPCC*, 2001]. The Kiruna profile was very similar to the MIPAS profile, however, it lacked smoothness in some regions, and hence was not included. The tropopause height of the MIPAS profile was manually adjusted from ∼9 km to ∼12 km, which is more appropriate for Toronto.

Nitric Acid (HNO_3)

The Kiruna nitric acid *a priori* profile was replaced with the MIPAS reference profile between 1-100 km, as the stratospheric peak of HNO_3 over Kiruna was too high for Toronto. The GEOS-CHEM web tool was used to increase the MIPAS profile between 0-1 km from 0.18 ppbv to 1.5 ppbv – a value more consistent with TAO's urban location. The resultant profile adopted at TAO is shown in Figure 4.1

Chlorine Nitrate ($ClONO_2$)

The Kiruna chlorine nitrate profile was replaced with one constructed from an April 2003 MkIV balloon flight over Esrange, Sweden. The new TAO profile tapers to a tropospheric volume mixing ratio of 10^{-14} [*Rinsland et al.*, 2003], and is also consistent with the MIPAS

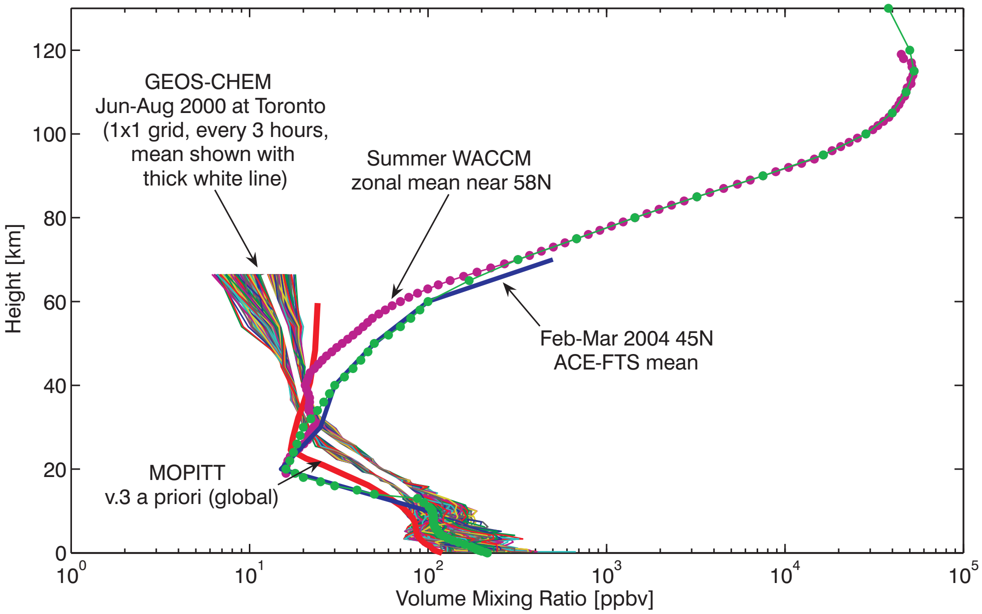


Figure 4.12 CO *a priori* profile adopted at TAO shown in green (see text for more details).

reference profile throughout the stratosphere (Figure 4.1).

Carbon Monoxide (CO)

Finally, the *a priori* profile of carbon monoxide shown in Figure 4.12 is a composite of several sources. Summer GEOS-CHEM model outputs over Toronto (every three hours) were obtained from a simulation performed by Dr. Qinbin Li on a 1x1° grid (personal communication, Dr. Colette Heald, 2004). Summer-time Whole Atmosphere Community Climate Model (WACCM) model zonal means near 58°N were obtained from *Forkman et al.* [2003, Figure 1], and winter ACE-FTS satellite measurement zonal means at 45°N were obtained from *Clerbaux et al.* [2005, Figure 3]. The MOPITT satellite *a priori* profile (applied globally) is shown for comparison. Finally, the CO *a priori* profile was also extended to 130 km at the time that the same modification was made to the NO *a priori* profile; however, unlike the case of NO, the fraction of the total column of CO found between 100-130 km is negligible.

TAO *a priori* profiles for trace gases not routinely retrieved (or minor interfering species) are presently identical with the Kiruna profiles. Since Kiruna is located above the Arctic Circle, the tropopause height and peak volume mixing ratios of such profiles likely need some further adjustments. Additional information is available in the MIPAS mid-latitude reference profile set for SO_2, OCS, SF_6, HCN, C_2H_6, C_2H_2, CCl_3F (CFC11), CCl_2F_2 (CFC12), CF_4 (CFC14), CHF_2Cl (CFC22), CCl_4, COF_2, ClO, HOCl, NH_3, H_2O_2, HO_2NO_2, and N_2O_5. Information on the same species (and a few others) is also available in the MkIV balloon flight data set, although the closest flight has been in Ft. Sumner (34.4°N, 104.2°W), New Mexico. Since new satellite measurements also exist for many of the above species, any modifications to their *a priori* profiles should be reconsidered in full, as needed.

4.1.2 Construction of *a priori* profile covariances ($\mathbf{S}_a$)

The construction of $\mathbf{S}_a$ is challenging in that knowledge of the *true* variability of the target trace gas from 0-100 km is required in order for the retrieval to be truly optimal. Since this information is commonly unknown, or only partially known, many *ad hoc* – and hence no longer optimal – approaches to constraining the retrieval have been investigated by different authors, e.g. *Eriksson* [2000], *Steck* [2002], *Hase et al.* [2004], and references therein. When a diagonal prior covariance is used in Equation 3.20, this minimizes the squared departure from the *a priori* and constitutes a *smoothing constraint* identical to that employed in Tikhonov regularization, e.g. *Eriksson* [2000]. A prior covariance with exponentially decreasing off-diagonal elements over a user-specified correlation length (only available in SFIT-2 v.3.91) is a special case of the first derivative operator squared – special in that it is now invertible [*Rodgers*, 2000, p. 218]. ($\mathbf{S}_a$ must be invertible when the retrieval algorithm implements Equation 3.31 or 3.32.) Such a prior covariance constrains the combination of the squared departure from the *a priori* and the squared first derivative of the retrieved profile. Finally, another common prior

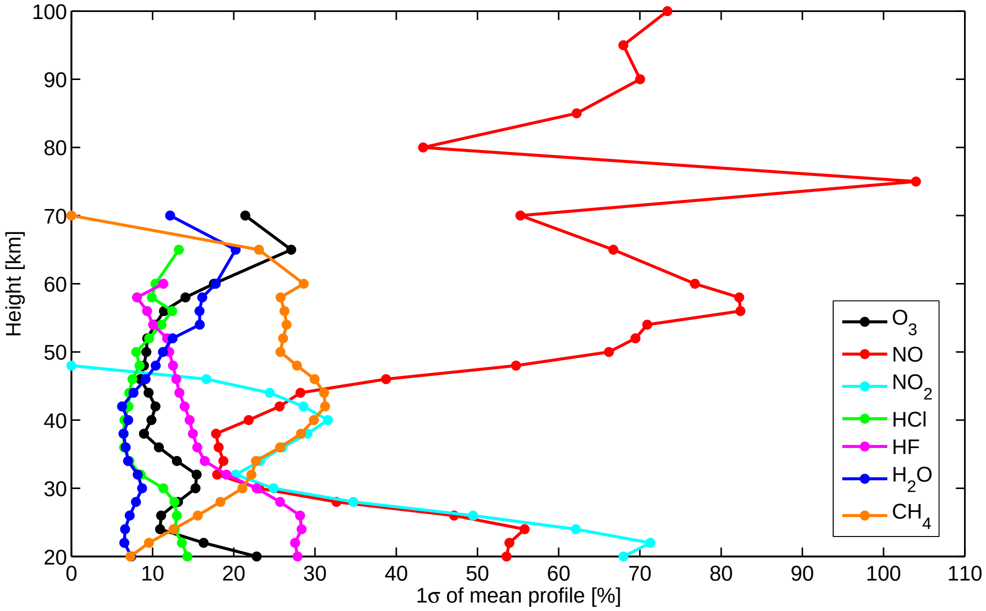


Figure 4.13 Standard deviations of mean HALOE profiles above Toronto.

covariance (available in all SFIT-2 versions) involves off-diagonal elements of Gaussian form, again employing a user-specified correlation length set to the standard deviation of the Gaussian function. This matrix is a special (invertible) case of the second derivative operator squared and places additional smoothing constraints on the profile curvature.

In order to estimate the magnitude of the diagonal elements of $\mathbf{S}_a$ used in operational TAO retrievals, we examined the standard deviation of the previously-described HALOE climatology as a function of height. The values shown in Figure 4.13 served as a rough guide in constraining the retrievals, and were typically chosen higher (see Section 5.2 for operational constraint details). Furthermore, we calculated correlation and covariance matrices for the ensembles of the seven species retrieved by HALOE. The correlation matrices again served as a very rough guide in choosing vertical correlation lengths for use in TAO retrievals. Figure 4.14 shows the correlation matrix for O_3 (between 20-65 km), which displays more complicated features than, e.g. exponentially decaying off-diagonal correlations. In particular, near the VMR peak at ∼35 km, the correlation length is

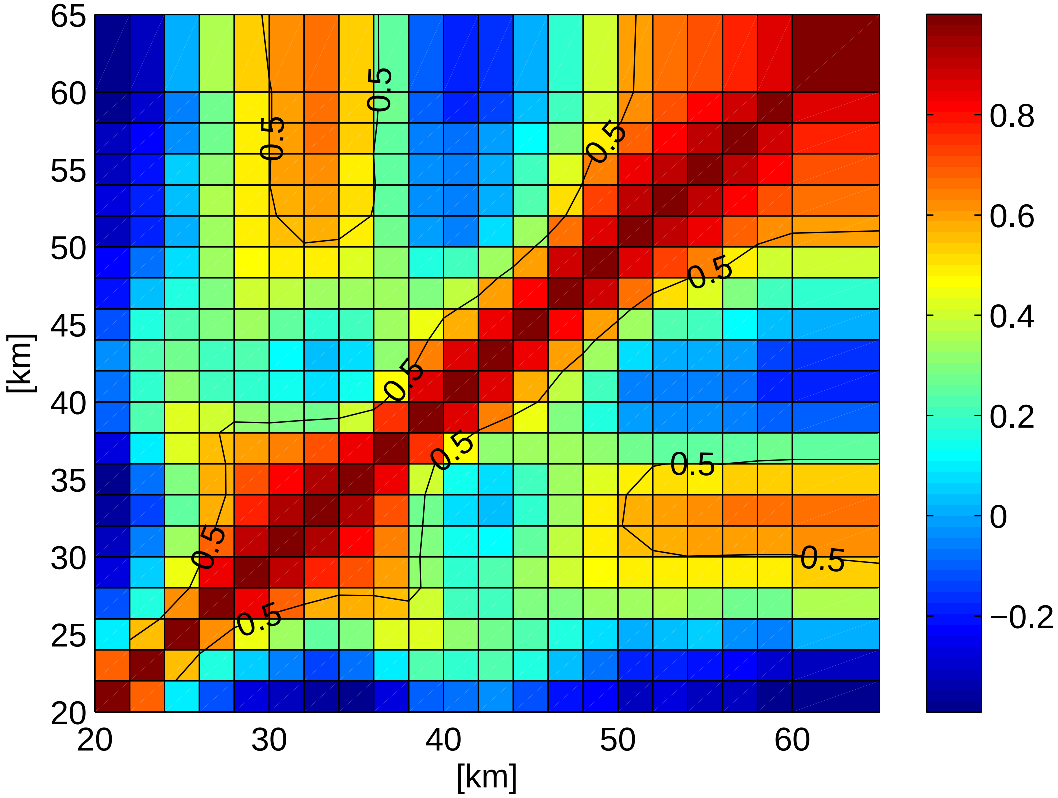


Figure 4.14 O_3 correlation matrix calculated from the HALOE climatology.

not symmetric, being 6 km above and 10 km below the peak. Furthermore, conspicuous "islands" of correlation exist between the VMR peak at ∼35 km and the VMR minimum near ∼70 km. Accordingly, we have carried out retrievals of O_3 from a simulated measurement, where the "true" profile is known, using three different $\mathbf{S}_a$ matrices: 1) diagonal with 100% standard deviation at all altitudes (*"Sa100"*), 2) 30% standard deviation for all diagonal elements and Gaussian off-diagonal elements with a vertical correlation length of 4 km between 0-100 km (*"SaNDSC"*), and 3) full covariance matrix constructed from the HALOE climatology covariance and supplementary information as described below (*"SaFull"*). The retrievals were performed on a simulated measurement with and without noise. The white noise added is representative of the average noise in TAO O_3 spectra near 3045 cm^{-1} (SNR=300), although we note that this value includes

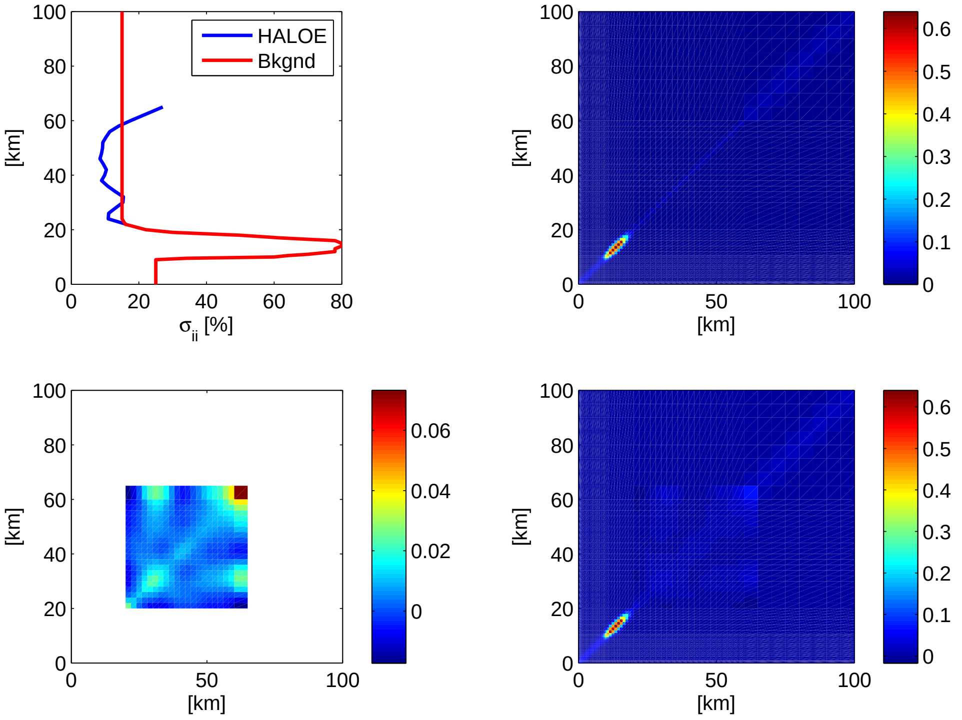


Figure 4.15 Top left: square roots of the diagonal elements of the HALOE covariance (bottom left) and the background covariance (top right). Bottom right: full covariance matrix constructed as described in the text.

systematic errors in practice.

Since the full covariance matrix based on the HALOE climatology only spans 20-65 km (Figure 4.15), it was supplemented with a background (*"Bkgnd"*) covariance spanning 0-100 km (Figure 4.15) constructed as follows. The diagonal elements of the background covariance (Figure 4.15) below 25 km were obtained from *Schneider et al.* [2005b, Figure 7], although we note that these values were in turn obtained from ozone sondes at Complementary NDACC station Izaña (28°N, 16°W). Above 25 km, the diagonal elements were set to 15%. Two different Gaussian correlation lengths were applied to the background covariance: 1.5 km (1.8 km HWHM) below 20 km [*Schneider et al.*, 2005b], and 5 km (6.0 km HWHM) above 65 km. In the full covariance matrix (Figure 4.15), elements of the background covariance for which HALOE covariances exist were replaced with the

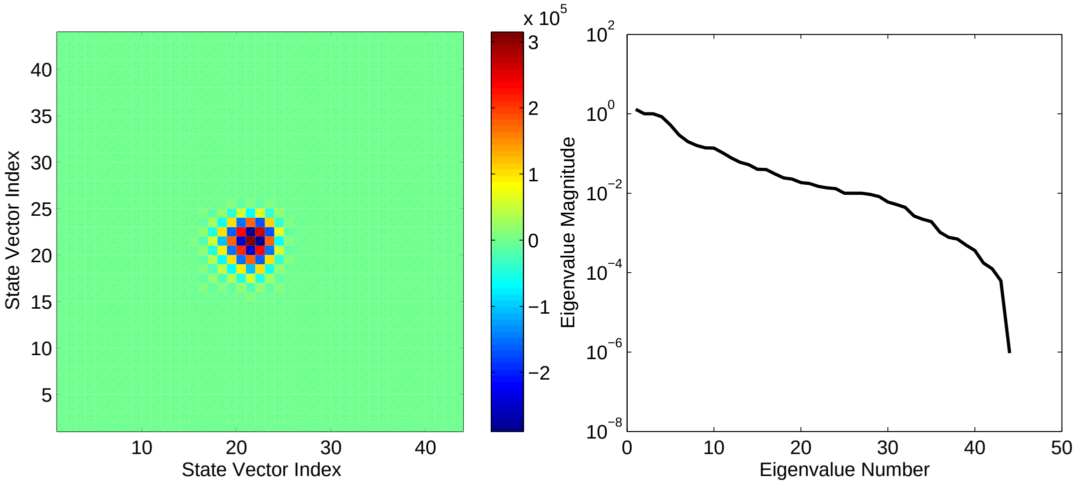


Figure 4.16 The inverse (left) and eigenvalue spectrum (right) of *SaFull*, showing its invertibility.

HALOE covariances. The full matrix was then interpolated onto the 38-layer retrieval grid used at TAO and further expanded to include the prior covariances of the non-profile state vector elements. These were the background slope and curvature in the spectrum, as well as the wavenumber shift and the covariances of the two interfering species whose profile scaling factors were being retrieved (CH_4 and H_2O). No correlation was assumed between these forward model parameters and the O_3 profile elements, as per how SFIT-2 treats the situation when a full covariance matrix is not used in the retrieval. As a final step, we ensured that *SaFull* was positive definite and hence invertible (Figure 4.16).

The *"true"* profile used to simulate the measured spectrum is shown in Figure 4.17, together with the *a priori* profile used in the retrievals and the retrieval results for the three different $\mathbf{S}_a$ matrices. The true profile is an extreme in the Toronto HALOE climatology (Jan. 1, 2002), used here to test results under challenging retrieval conditions. In the vicinity of the O_3 VMR peak, both *Sa100* and *SaNDSC* yield retrieval results that are closer to the true profile than when *SaFull* is used (Figure 4.18). Even though the diagonal elements of *SaNDSC* are much more constrained (30% vs 100%), it performs similarly to *Sa100*, presumably due to its 4 km vertical correlations, which introduce

	[0-100 km]			[0-15 km]			[15-50 km]		
	Sa100	SaNDSC	SaFull	Sa100	SaNDSC	SaFull	Sa100	SaNDSC	SaFull
d$_s$	2.60	2.09	1.37	0.14	0.12	0.16	2.46	1.97	1.21
δ	6.91	8.31	11.12	35.82	34.49	22.98	-3.54	-1.15	6.97
%a	11.90	12.70	15.20	82.50	81.80	75.50	-1.45	-0.43	3.67
S$_s$	3.00	2.50	1.16	25.60	18.00	8.53	2.68	1.16	1.55
S$_m$	1.85	1.57	1.24	6.10	4.44	5.34	1.79	1.49	1.01
S$_T$	1.95	2.15	2.15	0.67	0.76	1.93	2.27	2.56	2.33
S$_{tot}$	3.52	2.95	1.70	26.32	18.54	10.06	3.22	1.89	1.85
S$'_{tot}$	4.03	3.65	2.74	26.33	18.56	10.25	3.94	3.18	2.98
col	7.97	7.85	7.61	1.45	1.48	1.74	6.50	6.35	5.84

Table 4.1 Total and partial O$_3$ columns retrieved with each $\mathbf{S}_a$ (see text for details), and the corresponding information content (d_s, %a) and percent smoothing (S$_s$), measurement (S$_m$), and temperature (S$_T$) errors. %a = % contribution of *a priori*, S$_{tot}$ = $\sqrt{S_s^2 + S_m^2}$, S$'_{tot}$ = $\sqrt{S_s^2 + S_m^2 + S_T^2}$, col = total or partial column [x10^{18} molecules/cm^2], δ = % difference from true total or partial column: (true-col)/true*100.

more information into the prior covariance. *SaFull*, which has more realistic off-diagonal elements but more severely constrained diagonal elements (Figure 4.15), yields retrieval results that are much closer to the *a priori* at all heights near the VMR peak.

Table 4.1 shows the total and partial columns retrieved with each $\mathbf{S}_a$, as well as the corresponding d_s and errors. The 0-15 km partial column results are shown for completeness, since in Section 4.3 we will show that the weighting function tends to zero in this region and that the retrieval is in fact not sensitive to the tropospheric O$_3$ column. *Sa100* yields the highest total and stratospheric column d_s and the smallest % difference from the true total column, but also the highest overall smoothing and measurement errors. The converse is true of *SaFull*: it yields the lowest total and stratospheric column d_s and the largest % difference from the true total column, but the lowest overall smoothing error (except between 15-50 km, where *SaNDSC* performs better), and the lowest overall measurement error (except between 0-15 km, where *SaNDSC* again performs better).

Indeed, *SaFull* yields the lowest total error, both when S_T is included in the total (S'_{tot}) or not (S_{tot}). However, if we consider the region of greatest retrieval sensitivity as our region of interest (15-50 km), then *SaNDSC* yields total errors comparable to *SaFull*, especially S_{tot}, and furthermore, it reproduces the true stratospheric column most faithfully (-1.15% difference), with the lowest smoothing error (1.16%), the least % contribution from the *a priori* (-0.43%), and d_s that is a compromise between *Sa100* and *SaFull*.

While the impact of adding white noise does not appear very significant for the retrieved profiles (Figures 4.17 and 4.18), more insight can be gained from examining the quality of the spectral fits, which also gives more insight into the performance of each prior covariance in the retrieval. Figure 4.19 (fits to the noise-free simulated spectrum) shows clear differences between the retrieval performance of *Sa100* and *SaNDSC* on the one hand, and *SaFull* on the other. As expected, *Sa100* yields the lowest RMS residual (0.032%), closely followed by *SaNDSC* (0.046%), however, the RMS residual of *SaFull* is nearly an order of magnitude higher (0.202%). When white noise is introduced into the simulated spectrum, *Sa100* and *SaNDSC* lead to nearly identical RMS residuals (0.355% and 0.357%, respectively), now approximately an order of magnitude higher than the noise-free case. Furthermore, the gap between these two and *SaFull* (0.402%) becomes much smaller. The noise introduced into the simulated measurement is not large enough to mask some of the systematic features of the simulated spectrum corresponding to the profile retrieved using *SaFull*, unlike the case of *Sa100* and *SaNDSC* (Figure 4.20).

In summary, we have illustrated how in the case of O_3 retrievals, an $\mathbf{S}_a$ matrix with variances of equal magnitude on the diagonal and vertical correlations of Gaussian form (i.e. the *SaNDSC* case) leads to a good compromise between retrieval accuracy, precision, information content (d_s, %a), and the quality of spectral fits. The square roots of the variances were in this case approximately equal to twice the standard deviations derived from the HALOE climatology (∼30%), while the vertical correlation length was equal to an average value derived from the HALOE climatology between 20-65 km (∼4 km).

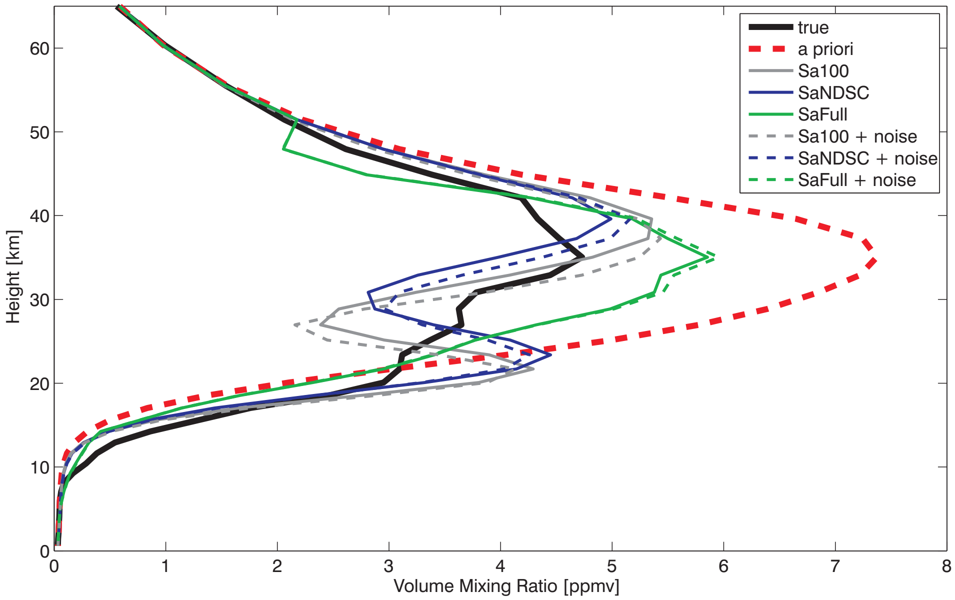


Figure 4.17 Retrievals of O_3 from a simulated spectrum (with and without noise) using differently constructed prior covariance matrices (see text for details).

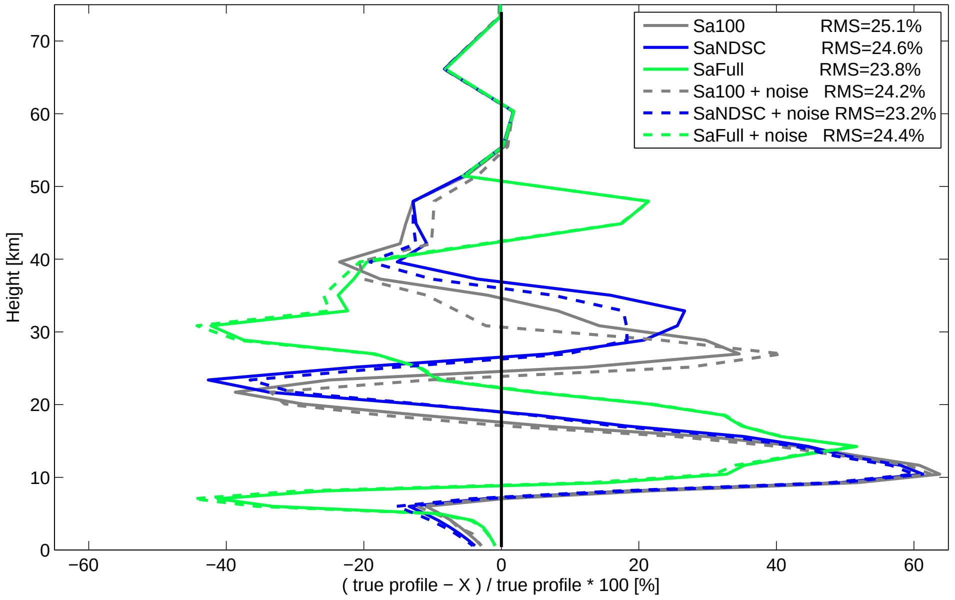


Figure 4.18 Percent difference between the retrievals in Figure 4.17 and the “true” profile.

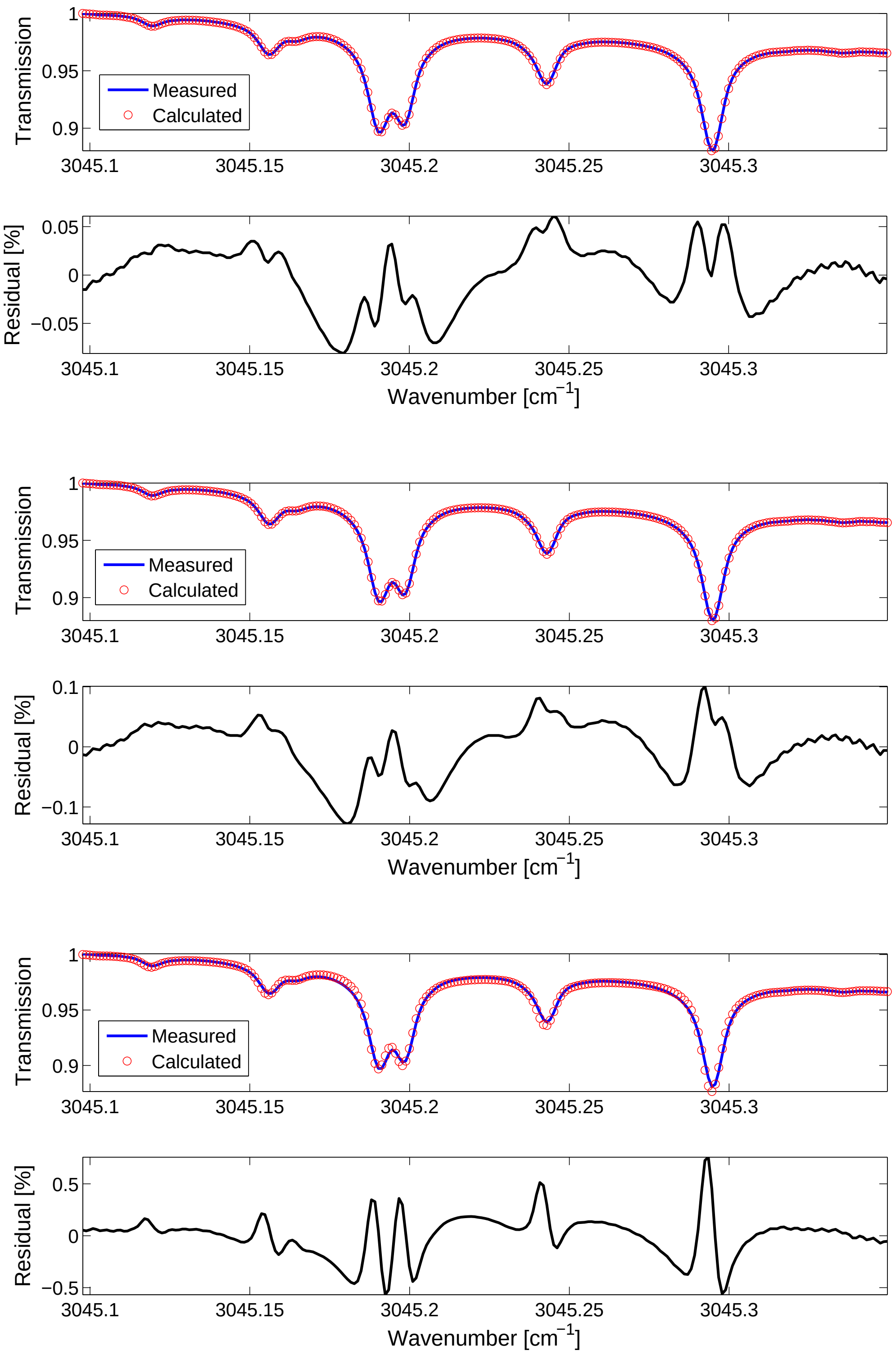


Figure 4.19 Spectral fits to simulated measurement without noise using *Sa100* (top), *SaNDSC* (middle), and *SaFull* (bottom). Fit RMS = 0.032%, 0.046%, and 0.202%, respectively. Note different scales for residual.

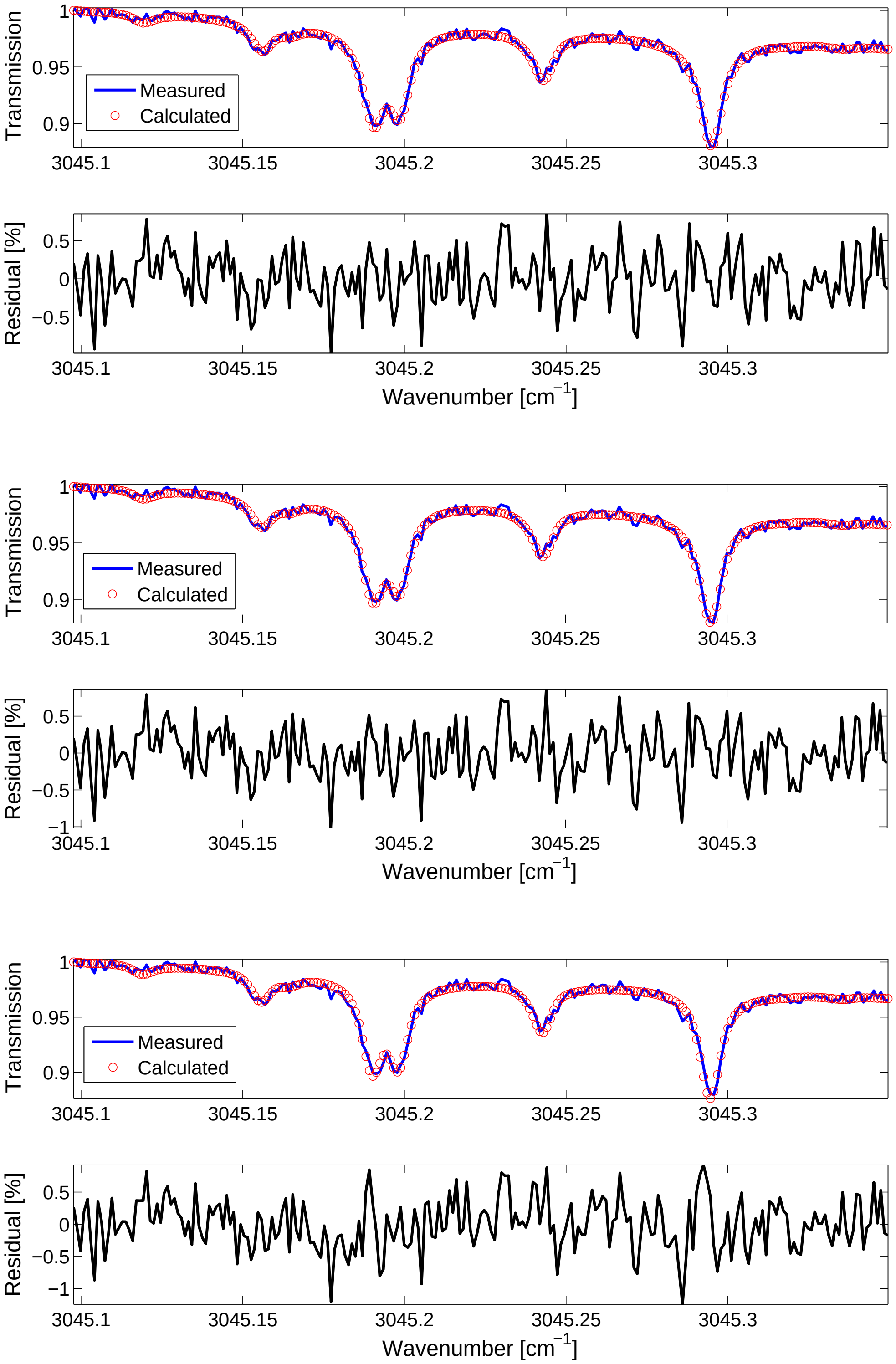


Figure 4.20 Spectral fits to simulated measurement with noise using *Sa100* (top), *SaNDSC* (middle), and *SaFull* (bottom). Fit RMS = 0.355%, 0.357%, and 0.402%, respectively.

The poorer retrieval performance of *SaFull* is in part due to the limited information available when constructing this prior covariance, and the need to make many *ad hoc* assumptions. While we have calculated full covariance and correlation matrices between 20-60 km for the other six species measured by HALOE (Figure 4.21), it is clear that their implementation into operational retrievals requires care and further research. At the present time, they appear more useful for estimating the vertical correlation lengths to be used in the retrievals. In particular, we note that the HALOE climatology is entirely inadequate for estimating the covariance matrices of tropospheric species (H_2O, CH_4, N_2O, CO, etc.) For this purpose we need access to a high-resolution CTM driven by real meteorological data.

Finally, lacking high-resolution CTM data for Toronto at the present time, we estimated the monthly mean VMR and standard deviation of the roof-top water vapour at TAO from one year (2003-2004) of continuous, densely sampled pressure, temperature and relative humidity measurements recorded by the weather station described in Chapter 2. (Water is an abundant interfering gas with ubiquitous spectral signatures.) The January and July mean VMR (0.00180 and 0.0175, respectively) brackets the surface value from the MIPAS reference profile (0.00973) adopted at TAO. The standard deviation of the weather station VMR in January and July is 54% and 21%, respectively, which puts a constraint on the magnitude of the profile scaling factor used when fitting water as an interfering species in the profile retrieval of another target gas.

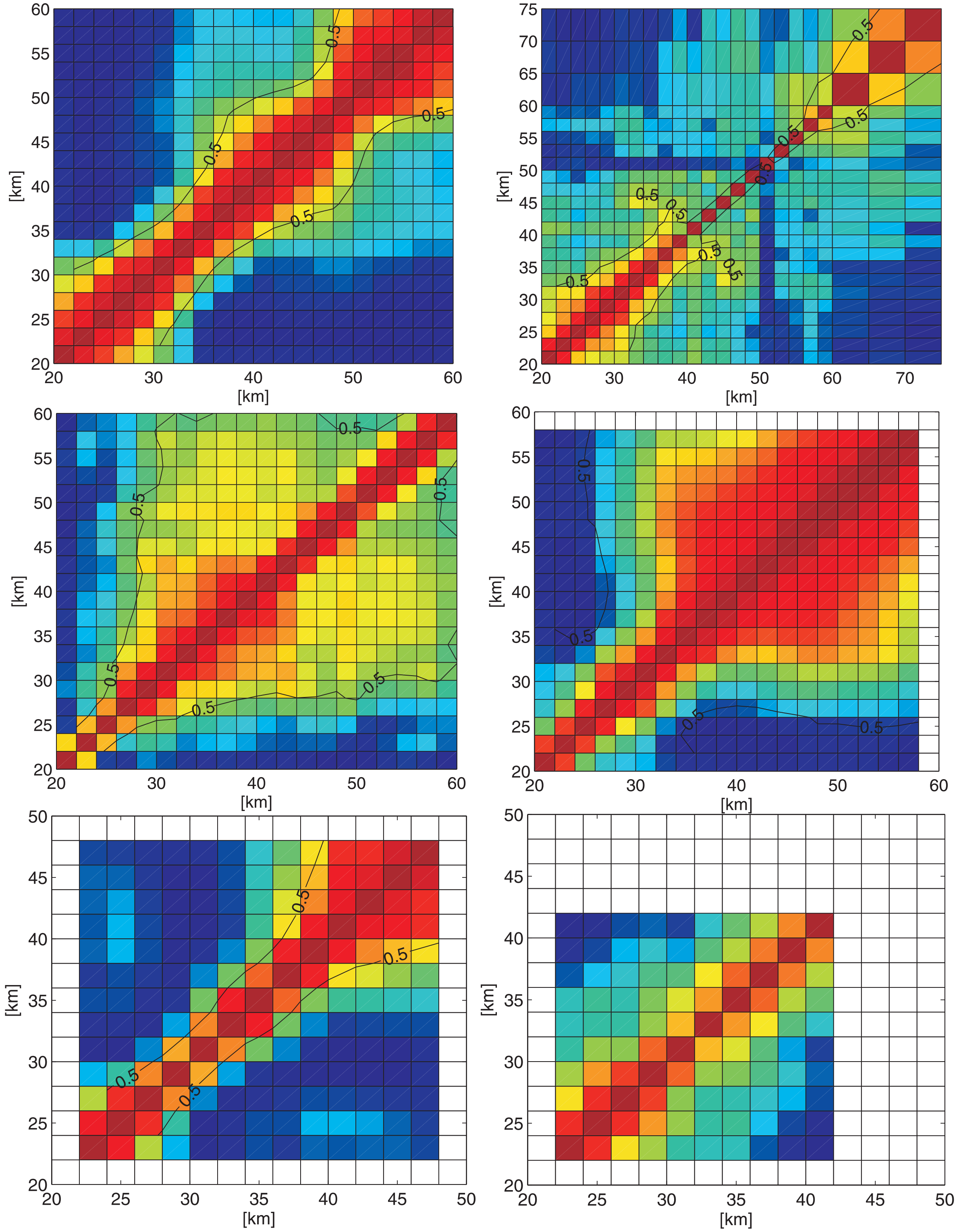


Figure 4.21 CH_4 (top left), H_2O (top right), HCl (middle left), HF (middle right), NO (bottom left), and NO_2 (bottom right) correlation matrices calculated from the HALOE climatology. Color scheme as for O_3 in Figure 4.14.

4.1.3 Construction of measurement noise covariances ($\mathbf{S}_\epsilon$)

The construction of $\mathbf{S}_\epsilon$ appears to be the least challenging of the explicit retrieval constraints, at least when we assume uncorrelated measurement errors in the spectral domain. This approximation breaks down when misalignments of the FTS cause the instrumental lineshape to deviate from that modeled for a given entrance aperture and maximum path difference. In this case, correlations are introduced between adjacent FTIR measurement channel (different wavenumber) errors, and a purely diagonal $\mathbf{S}_\epsilon$ discards some of the additional information available about the quality of the measurements. Table 4.2 briefly explores the effect of introducing off-diagonal $\mathbf{S}_\epsilon$ elements of Gaussian form with a spectral correlation length of 0.001 cm^{-1} (HWHM) – approximately equal to the halfwidth of the unapodized ILS – on the characterization and error analysis of the O_3 retrievals from Section 4.1.2. (It is not possible to explore the effects of a non-diagonal $\mathbf{S}_\epsilon$ on SFIT-2 partial and total column retrieval results because only a diagonal $\mathbf{S}_\epsilon$ is implemented in the code at the present time.) As compared to values in Table 4.1, the degrees of freedom for signal are always smaller and the % contribution from the *a priori* is always bigger. Furthermore, all errors are bigger when spectral correlations are introduced, with a few exceptions between 0-15 km, where the retrieval is not sensitive to the O_3 profile in the first place. This highlights the importance of correct FTS alignment or the full characterization of the non-ideal FTS ILS in order to characterize profile retrieval errors and information content accurately.

The second approximation concerning $\mathbf{S}_\epsilon$ involves not adding the term $\mathbf{K}_b\mathbf{S}_b\mathbf{K}_b^T$ to $\mathbf{S}_\epsilon$ to form an effective measurement covariance (Section 3.3.2), which then includes the errors of forward model parameters that are not being retrieved, e.g. the assumed temperature profile in the case of SFIT-2. While this approach is ubiquitous in the NDACC FTIR community that uses the SFIT-2 retrieval algorithm (and a commonly taken short cut according to *Rodgers* [2000, p. 111]), it is important not to lose sight of the fact that it is a deeply embedded approximation. Evaluating its effects on retrievals, character-

	[0-100 km]			[0-15 km]			[15-50 km]		
	Sa100	SaNDSC	SaFull	Sa100	SaNDSC	SaFull	Sa100	SaNDSC	SaFull
d_s	2.33	1.88	1.22	0.10	0.09	0.11	2.23	1.78	1.11
%a	13.00	13.50	16.90	86.20	85.10	79.60	-0.91	-0.10	4.97
S_s	3.20	2.67	1.36	26.20	18.50	8.23	2.83	1.35	1.65
S_m	2.12	1.84	1.36	5.15	3.83	4.48	2.21	1.87	1.35
S_T	1.97	2.13	1.99	0.54	0.67	1.24	2.36	2.52	2.24
S_{tot}	3.84	3.24	1.92	26.70	18.89	9.37	3.59	2.31	2.13
S'_{tot}	4.31	3.88	2.77	26.71	18.90	9.45	4.3	3.42	3.09

Table 4.2 Effect of introducing a spectral correlation length of 0.001 cm^{-1} (HWHM) into $\mathbf{S}_\epsilon$ on the information content and errors of simulated retrievals, as defined in Table 4.1.

ization, and error analysis is not possible at present since it requires incorporating the temperature-pressure-airmass forward model calculation program (*fscatm*) into the main forward model of SFIT-2, and thus constitutes a major software redesign.

In operational TAO retrievals, all diagonal elements of $\mathbf{S}_\epsilon$ are set to $1/(\text{SNR})^2$. (The SNR in a microwindow is calculated as the maximum signal divided by the RMS residual in the same microwindow.) A two-step retrieval approach was developed, whereby in the first step retrievals were performed on an ensemble of spectra using a single SNR value (known from experience, or obtained from the literature, or found by trial-and-error), and in the second step, the retrievals were repeated with a *"custom"* value of SNR for each spectrum. The custom SNR value was derived from the RMS residual of the fit in the first step. Implicit in this method is that the RMS residual of the first step fit be robust against changes in the ensemble SNR value. This is illustrated for the case of 2001-2004 HF retrievals performed using an ensemble SNR value of 150, 300, and 500. Figure 4.22 shows that while the ensemble SNR value varies by approximately a factor of 2 with respect to 300, the vast majority of the RMS fit residuals vary by less than 10%. This two-step approach was applied to all 2001-2004 retrievals submitted to the NDACC database, and the custom SNR values of the second step were used in all

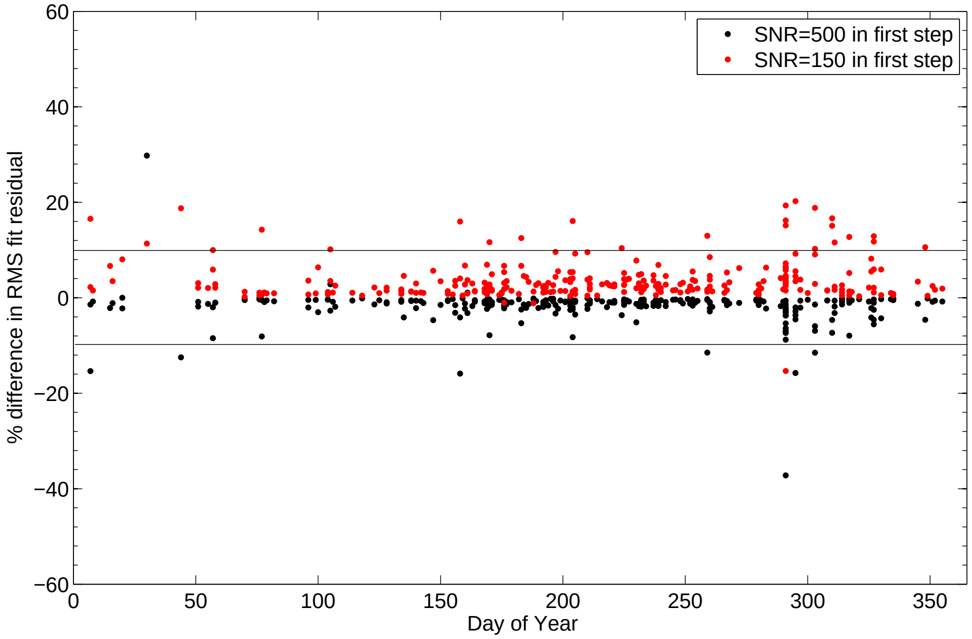


Figure 4.22 Percent difference in RMS fit residuals of the first retrieval step for the ensemble SNR set to 150 and 500 with respect to RMS values derived using an ensemble SNR of 300. Data set includes all 2001-2004 HF spectra (4038 cm^{-1}).

retrieval characterization and error analysis presented in Section 4.3 and 5.2.

Finally, the use of the residual of the first fit in order to determine the SNR for the second step retrieval also serves to reduce the SNR that may be observed in the raw spectra (peak signal in a microwindow divided by the noise in a saturated region of the spectrum) to an effective value that includes the systematic effects due to forward modeling errors, e.g. spectroscopy and lineshape. Such an approach de-weights the measurements and makes it less likely that the retrieved profiles will oscillate in order to compensate for systematic differences between the modeled and measured spectrum (although they can oscillate for other reasons).

4.2 Implicit Constraints

We now turn to the constraint implicit in the choice of the retrieval layering scheme. At TAO, two grid schemes must be distinguished. The first is a 63-layer scheme spanning 0-100 km and used to specify NCEP meteorological data and VMR profiles at its 63 layer boundaries. The ray tracing program *fscatm* converts these quantities into density-weighted values specified at 38 layer mid-points, again spanning 0-100 km, and forming the second layering scheme on which the retrieval is then performed. The location of the 38 layer mid-points was chosen following the practice of Complementary NDACC station Wollongong: the bottom-most mid-point was specified at 0.4 km above TAO's altitude (0.174 km above sea level) in order to yield an 0.8 km thick first layer; each successive mid-point was specified such that the layer thickness increased by 4%. This approach resulted in ∼2 km layer thicknesses in the lower stratosphere, ∼3 km in the upper stratosphere, and ∼10 km near the top of the retrieval grid at 100 km. Furthermore, this choice eliminated abrupt changes in layer thickness and resulted in a total number of layers that is computationally manageable. Retrieving up to the altitude of 100 km is a common – if arbitrary – practice in the NDACC, justified by the fact that all molecules of interest have negligible partial column contributions from above 100 km. NO is a notable exception to this rule (Section 6.1).

Historically, profile retrievals using FTIR spectra were first performed on a 29-layer vertical grid with a 1-km layer thickness in the troposphere, e.g. *Pougatchev et al.* [1995], and with the retrieval of only stratospheric species in mind. The choice of the number of layers used in a ground-based FTIR retrieval – and their optimal location – is discussed in some detail by *Meier et al.* [2004a]; however, that study focused primarily on retrievals of total column amounts using a nonlinear least squares fitting technique, and thus could not quantify grid spacing effects on retrievals of vertical profiles. *Hase et al.* [2004] reported the inadequacy of a 45-layer grid with a 1.2-km tropospheric layer thickness for optimal estimation retrievals of the vertical profiles of H_2O (they noted that the half-width of the

Total Layers	0-7 km	7-14 km	14-21 km	21-49 km	49-100 km
96	0.2	0.4	0.8	1.6	3.0
79	0.4	0.4	0.8	1.6	3.0
62	0.8	0.8	0.8	1.6	3.0
57	1.0	1.0	1.0	1.6	3.0
70	0.8	0.8	0.8	1.1	3.0
58	0.8	0.8	0.8	2.1	3.0
55	0.8	0.8	0.8	2.6	3.0

Table 4.3 Constant layer thicknesses [km] specified in 7-km regions in order to form seven different grid schemes (each table row) used in the forward modeling of transmission spectra.

features observed in the **A** matrix approached the grid spacing).

In addition to determining the minimum width of vertical structures accessible to the measurements, the choice of grid spacing affects profile retrievals via the density-weighted pressure and temperature profiles, which in turn determine the interpolation of spectroscopic parameters used to forward model the spectrum in the retrieval process. In the present study, seven retrieval grid schemes that employed various degrees of tropospheric and stratospheric sparseness were constructed (Table 4.3) in order to systematically determine the minimum grid spacing necessary for modeling the transmission spectra of H_2O (the trace gas exhibiting the strongest vertical profile gradients and the highest abundance) and other NDACC target gases, such that the transmission biases due to the grid spacing fall below typical noise levels in real spectra. We note that the purpose of this study was *not* to determine the minimum number of retrieval layers or their optimal spacing, which is a problem specific to each target gas.

The spectra were modeled in typical retrieval microwindows using the spectroscopic parameters of HITRAN 2004 and assuming a 250-cm OPD, 45° SZA, and a 1.54-mrad FOV, except for spectra in the F6 region, where the FOV was 4.63 mrad. In order to separate the effects of the grid spacing from those of the annual variation of the temperature and pressure profiles, 365 spectra were simulated using each grid scheme for the 365 NCEP pressure and temperature reanalyses for Toronto in 2004. It was confirmed that transmission variations due to the annual range of temperature and pressure parameters are much smaller than those due to the choice of grid spacing, and hence only annual averages of transmission are shown in Figures 4.23-4.30. What is being shown in these figures is the transmission difference (%) due to a given grid scheme with respect to the most dense grid scheme in the troposphere (left panel) and in the stratosphere (right panel). The transmission difference due to the most dense tropospheric or stratospheric grid scheme is also shown with respect to itself and is always zero in either figure panel. The legends in these figures are identical and refer to the layer thickness between 0-7 km in the left panel, and the layer thickness between 21-49 km in the right panel. All grids that are varied in the troposphere (stratosphere) have a 1.6 km (0.8 km) layer thickness in the stratosphere (troposphere), and form one row of Table 4.3.

Figure 4.23 shows that a tropospheric grid with a 1.0-km layer thickness between 0-7 km will lead to a systematic transmission difference of at most 0.4% with respect to a grid with a 0.2-km layer thickness between 0-7 km. (Implicit in such a comparison is the assumption that the most dense grid scheme is sufficient to capture all features of practical consequence in the true atmosphere.) In order to detect such a systematic feature with a relative SNR of at least 2, the spectrum would need to be recorded with an SNR of 500 (2*1/0.4%). Differences between the next least dense grid (0.8 km between 0-7 km) and the most dense grid would require an SNR of 1000 for detection, and finally, differences between the 0.4-km grid and the most dense grid would require an SNR of 5000 for detection. It appears that a tropospheric grid spacing of 0.8 km or thinner will lead to systematic spectral differences that are not distinguishable from the noise, although the situation will depend on the SNR typically found in a given microwindow at a given observation station (e.g. SNR is 200-1000 in TAO spectra). The right panel of Figure 4.23

shows that variations of the stratospheric grid spacing are of very little consequence for the H_2O absorption feature, as expected given the vertical distribution of H_2O, however, they affect the interfering CO_2 feature to a much greater extent. Nonetheless, an SNR of 2857 would be necessary in order to detect the largest of these systematic spectral differences with a relative SNR of 2, which is out of reach of a low-altitude observation station like TAO. Finally, this particular H_2O microwindow was chosen for this study because of its high lower state energy (E”=1962.5 cm^{-1}), which increases the sensitivity of the absorption feature to temperature and thus provides a conservative estimate of spectral variations with grid scheme.

Figure 4.24 proves that variations of the tropospheric grid spacing are of very little consequence for the absorption features of O_3, as expected. The stratospheric spacing of 2.6 km leads to biases with respect to the 1.1 km spacing that would require an SNR of 8000 for detection with a relative SNR of 2. Since this is beyond the state of the art in ground-based FTIR measurements at present, a layer spacing of the order of 2.6 km is appropriate for use in retrievals.

Figures 4.25 and 4.26 show that any of the tropospheric grid spacings are sufficient to represent the transmission spectra of CH_4 and N_2O faithfully given typical noise levels found in FTIR spectra. Somewhat surprisingly, the stratospheric grid spacing has a bigger effect on the absorption features of these gases, just like for the case of CO_2 in the H_2O microwindow. Nonetheless, an SNR of 4444 and 1429 would be required to detect systematic spectral differences between the most and least dense stratospheric grid spacing schemes for CH_4 and N_2O, respectively.

Figures 4.27 and 4.28 show that the stratospheric grid spacing of 2.6 km is sufficient to represent the transmission spectra of HCl and HF faithfully given the noise levels found in FTIR spectra. While the tropospheric grid spacing has a negligible impact on the transmission spectrum of HCl, it affects that of HF through the broadband H_2O absorption that forms the transmission background in this microwindow.

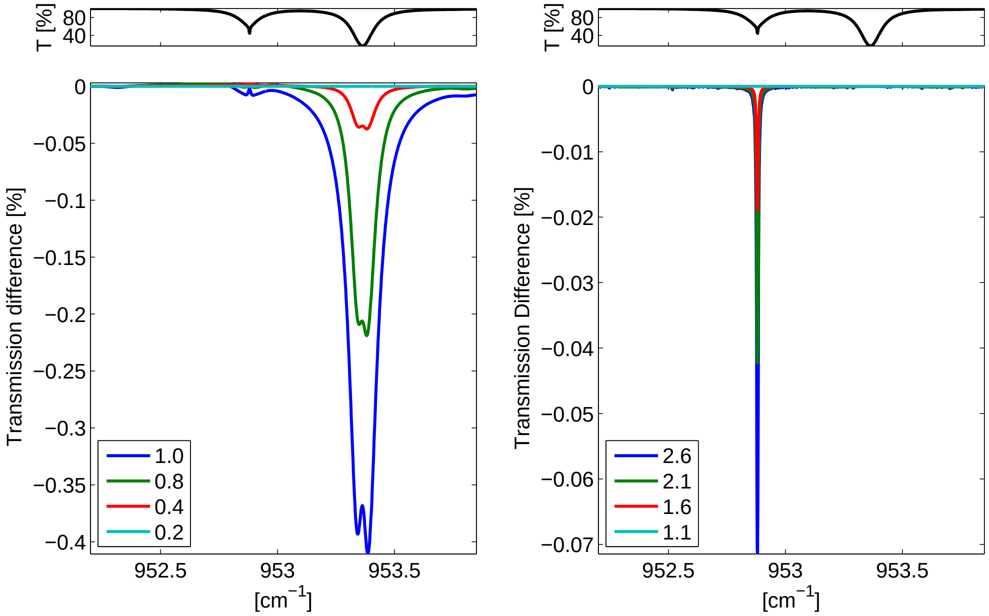


Figure 4.23 Transmission (top panels) and % differences for tropospheric (left) and stratospheric (right) grid variations in an H_2O microwindow (see text for legend details).

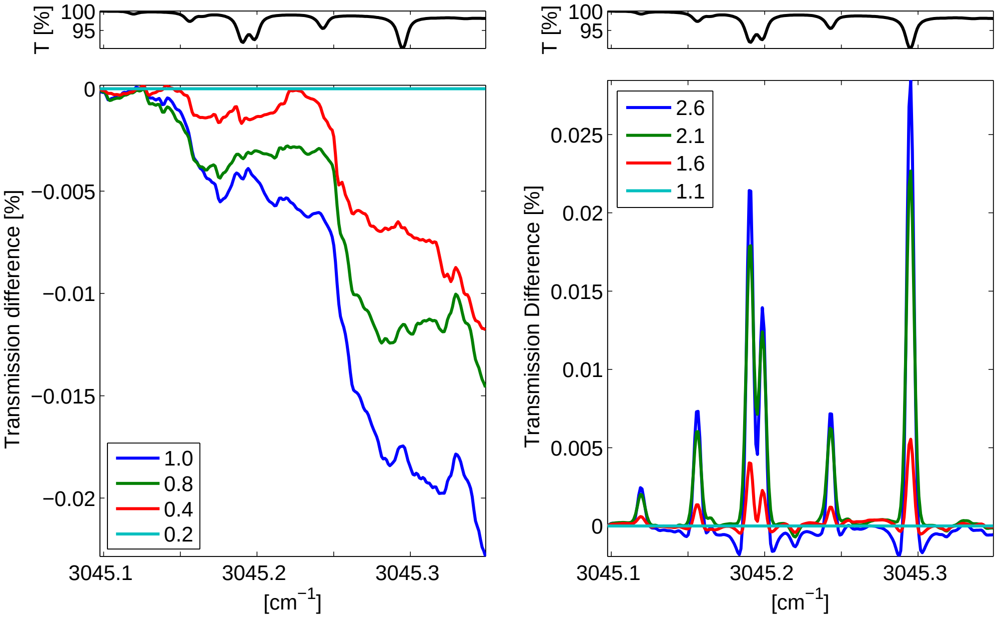


Figure 4.24 Transmission (top panels) and % differences for tropospheric (left) and stratospheric (right) grid variations in an O_3 microwindow (see text for legend details).

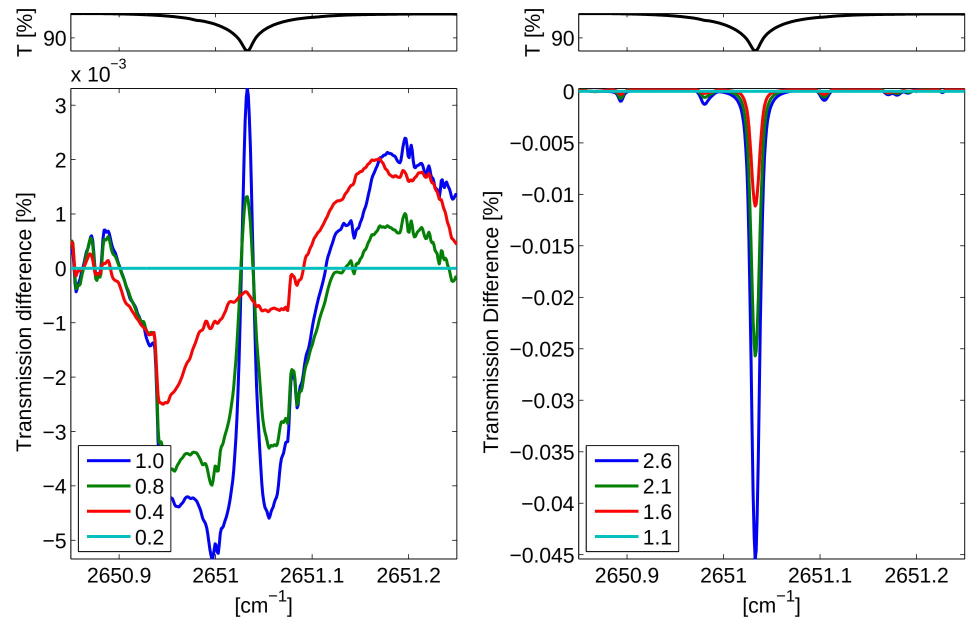


Figure 4.25 Transmission (top panels) and % differences for tropospheric (left) and stratospheric (right) grid variations in a CH_4 microwindow (see text for legend details).

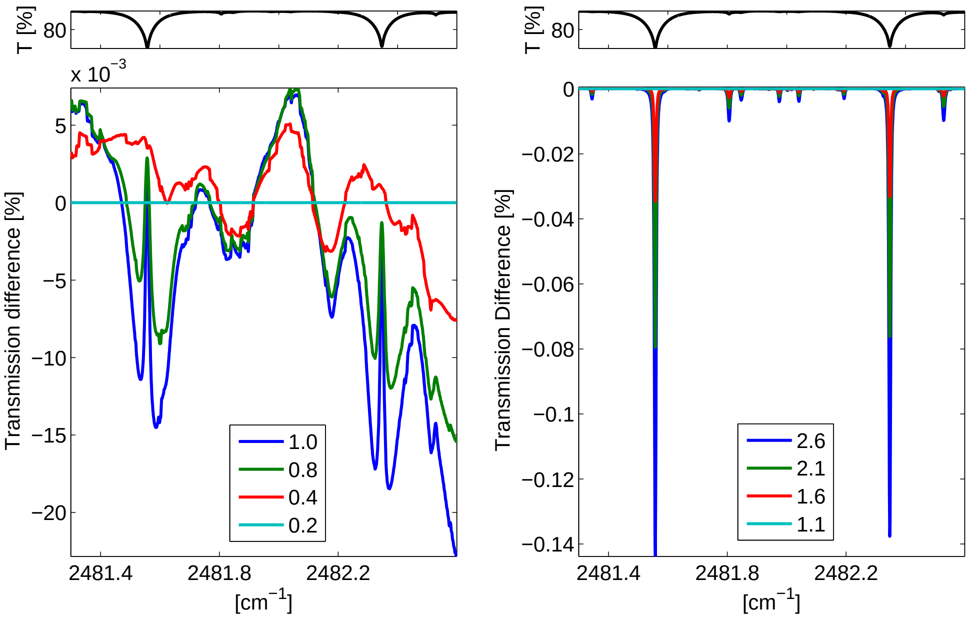


Figure 4.26 Transmission (top panels) and % differences for tropospheric (left) and stratospheric (right) grid variations in an N_2O microwindow (see text for legend details).

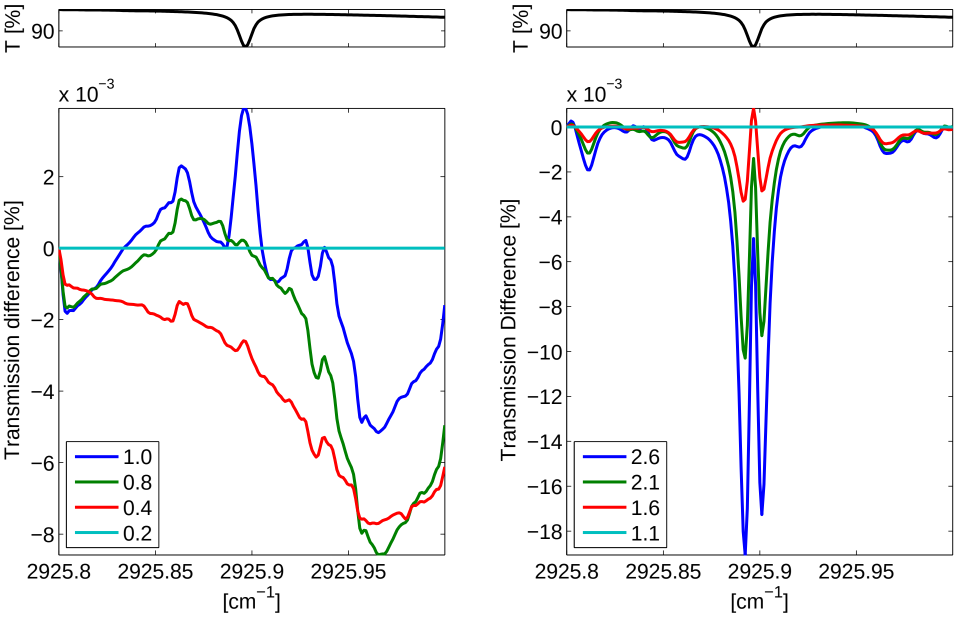


Figure 4.27 Transmission (top panels) and % differences for tropospheric (left) and stratospheric (right) grid variations in an HCl microwindow (see text for legend details).

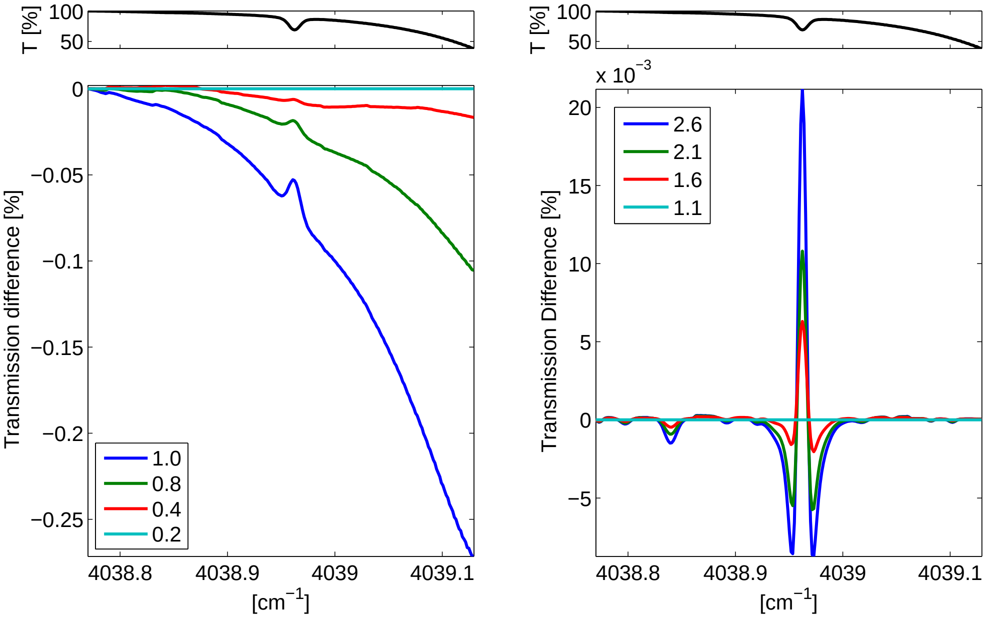


Figure 4.28 Transmission (top panels) and % differences for tropospheric (left) and stratospheric (right) grid variations in an HF microwindow (see text for legend details).

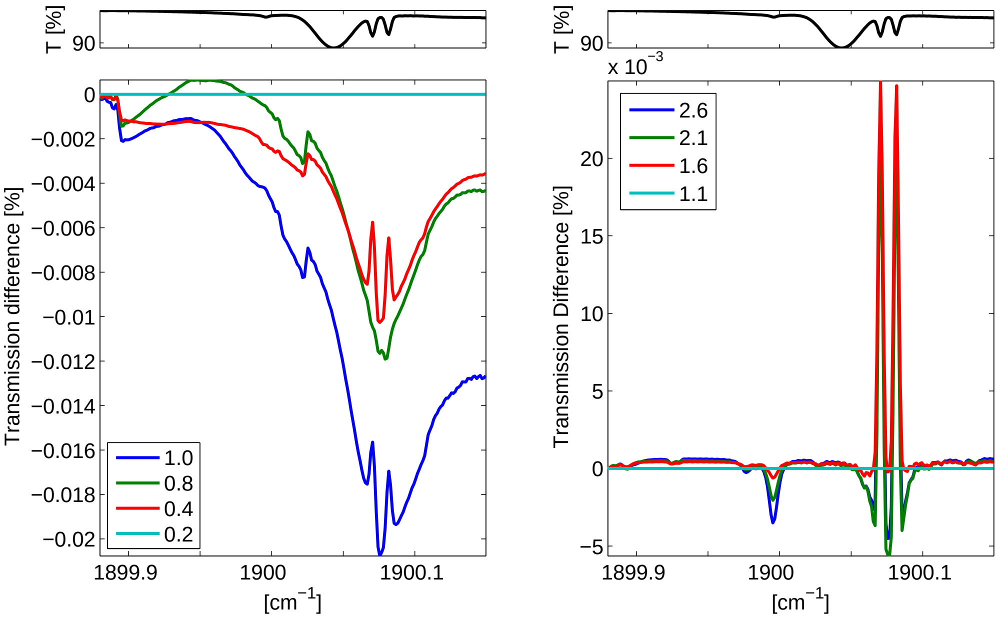


Figure 4.29 Transmission (top panels) and % differences for tropospheric (left) and stratospheric (right) grid variations in an NO microwindow (see text for legend details).

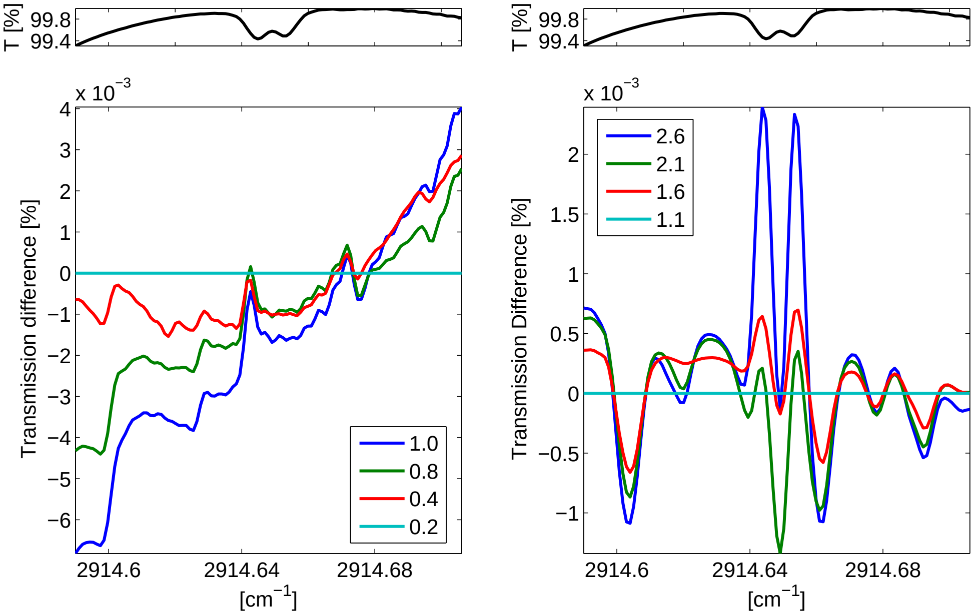


Figure 4.30 Transmission (top panels) and % differences for tropospheric (left) and stratospheric (right) grid variations in an NO_2 microwindow (see text for legend details).

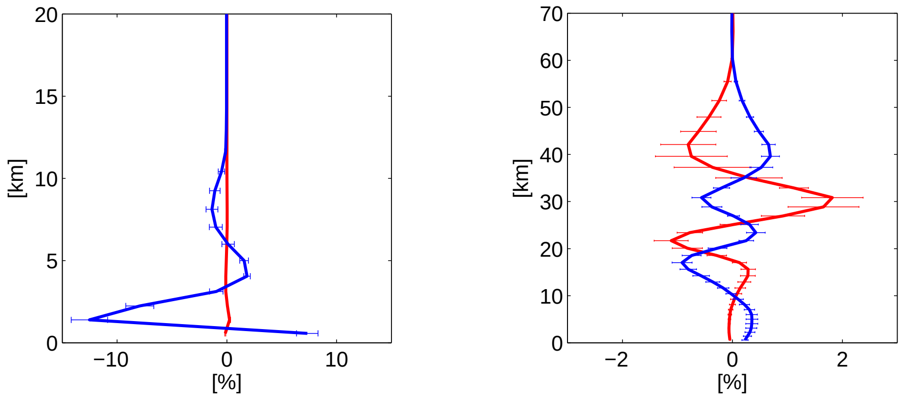


Figure 4.31 Mean difference in 365 H_2O (left) and O_3 (right) profiles retrieved on a 38 layer grid from spectra simulated on the most and on the least dense tropospheric grid (blue) and from spectra simulated on the most and on the least dense stratospheric grid (red). Error bars represent the standard deviation of the mean difference.

Finally, Figures 4.29 and 4.30 show that the stratospheric grid spacing of 2.6 km is sufficient to represent the transmission spectra of NO and NO_2 faithfully given the noise levels found in FTIR spectra. Although some systematic features are found in the difference spectra of NO for varying tropospheric grids, they are below FTIR detection levels. The difference spectra of NO_2 are even less affected by variations in the tropospheric grid spacing.

Although Figures 4.23-4.30 show that systematic differences in transmission spectra due to the choice of grid scheme are only significant in the case of H_2O (either as a target gas or as an interfering species), nonetheless, we studied the effect of all of the above spectral differences on retrievals of vertical profiles performed on the 38-layer retrieval grid used at TAO. For each of the eight target gases discussed above, 365 retrievals were performed on the 365 spectra simulated on the most and on the least dense stratospheric grid (*"stratospheric retrievals"*), and on the 365 spectra simulated on the most and the least dense tropospheric grid (*"tropospheric retrievals"*). The mean of the day-by-day differences between the two sets of tropospheric retrievals brackets the magnitude of the profile retrieval response to the highest level of systematic biases introduced into the spectra by the tropospheric grid spacing. This mean difference is shown in blue in Figure 4.31, and is ∼10% for H_20. The mean difference between the two sets of stratospheric retrievals is shown in red and is seen to be negligibly small for H_2O, in agreement with the size of the observed spectral biases due to differences in the stratospheric grid spacing.

The right panel of Figure 4.31 shows that both tropospheric and stratospheric grid differences propagate to O_3 retrievals, however, the effect is mostly below 1%, which is smaller than the 2-4% temperature error and the 7-9% retrieval noise error estimated for O_3 in Section 4.3 (Figure 4.34). Figure 4.32 shows the corresponding mean differences of retrieved profiles for the remaining gases discussed above. CH_4 shows mean profile differences below 2%, which are smaller than the ∼5% temperature and retrieval noise errors estimated in Figure 4.48. N_2O shows mean profile differences below 5%, which are comparable to the retrieval noise error and smaller than the ∼15% temperature error estimated in Figure 4.50. Both CH_4 and N_2O show a response to changes in the stratospheric grid (red curves) at ∼25 km, presumably due to the curious spectral biases introduced by the stratospheric grid spacing in Figures 4.25 and 4.26. As expected, HCl shows a smaller response to the tropospheric grid changes than to the stratospheric grid changes, and both responses' magnitudes are below the ∼3% temperature error and the ∼10% retrieval noise error in Figure 4.42. HF shows a very similar response to HCl to stratospheric grid variations, however, a much bigger response than HCl to tropospheric grid variations, presumably due to the considerable transmission bias introduced by the broad H_2O absorption feature. At a maximum of 4%, the HF response is bigger than the ∼3% temperature error but smaller than the ∼10% retrieval error shown in Figure 4.44. Finally, NO shows a bigger response than NO_2 to both tropospheric and stratospheric grid variations, however, the response magnitudes of both gases are much smaller than the ∼10% retrieval noise error estimate in Figures 4.36 and 4.40.

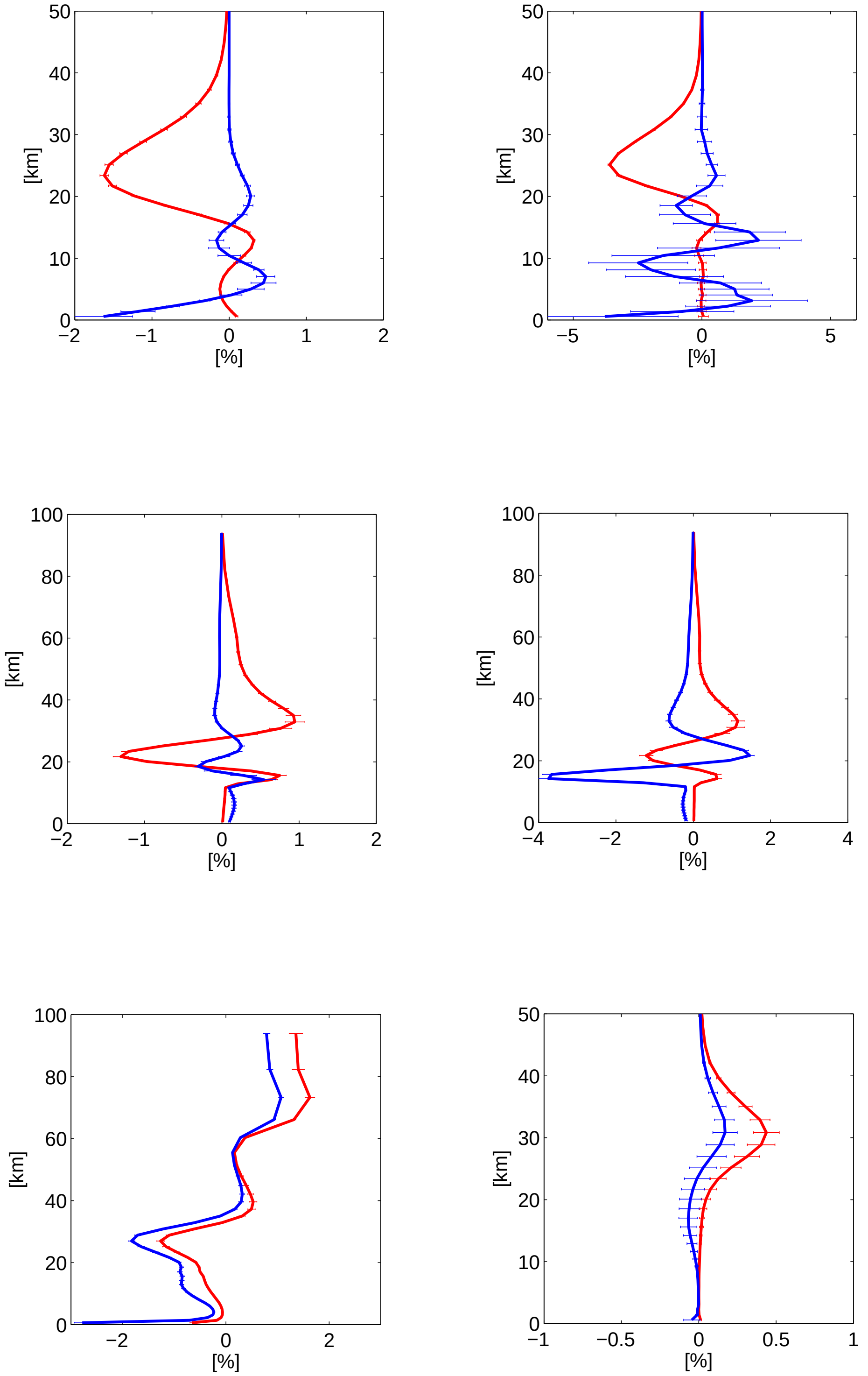


Figure 4.32 Same as for Figure 4.31, showing CH_4 (top left), N_2O (top right), HCl (middle left), HF (middle right), NO (bottom left) and NO_2 (bottom right).

In summary, the largest systematic bias is introduced into the transmission spectra of H_2O and HF due to the choice of the tropospheric grid spacing. A tropospheric layer thickness of 0.8 km should reduce this bias to levels below the SNR observed in the spectra. While all transmission biases – whether caused by tropospheric or stratospheric grid spacing – propagate to profile retrievals, the effect is below 5% for all gases except H_2O, where it is $\sim$10%.

4.3 Characterization and Error Analysis Results

In this section, we show a representative spectral fit and fit residual, as well as the corresponding weighting function matrix for a single retrieval of each of the eight trace gases routinely processed at TAO. We also show a six-panel figure with **1)** the averaging kernel matrix, **2)** the averaging kernels (the rows of $\mathbf{A}$) superimposed on the area of each kernel (divided by 10 for convenience and showing regions of good retrieval sensitivity when ∼0.1) and the % contribution of the *a priori* (also divided by 10 for convenience and showing regions of *poor* retrieval sensitivity when ∼0.1), **3)** the height-averaged volume mixing ratio kernels (Equation 3.63) for layers spanning 0-100 km, 0-15 km, and 15-50 km, unless indicated otherwise, **4)** the partial column kernels for the same altitude regions (Equation 3.61), **5)** the square roots of the diagonal elements of the $\mathbf{S}_a$, $\mathbf{S}_m$, $\mathbf{S}_T$, $\mathbf{S}_s$, and $\mathbf{S}'_{tot}$ error covariance matrices (Equations 3.54-3.56), showing the standard deviation of these errors at each profile altitude, and finally, **6)** the *a priori* and retrieved profiles for this particular target gas and analysis.

The indicated retrieval date and spectrum for each target gas was chosen from the 2004 ensemble of quality-controlled results based on having the best matrix conditioning, as described in Section 3.4.3.2. While the matrix conditioning varies from retrieval to retrieval, in practice, characterization and error analysis are only appreciably affected by $\mathbf{S}_\epsilon$, $\mathbf{S}_a$, the *a priori*, and the spectrum solar zenith angle; therefore, the results presented in this section are representative of the 2004 ensemble, and indeed, all TAO retrievals performed with similar constraints (especially a similar prior covariance). Section 5.2 will further characterize the range of retrieval sensitivity, information content and errors introduced by the custom SNR, a variable SZA, and also the annual range of pressure and temperature parameters. As per the discussion of Section 3.4.3.1, all retrievals were characterized with respect to the *a priori* profile, rather than the retrieved profile. Finally, all other major retrieval details are summarized in Section 5.2, for clarity.

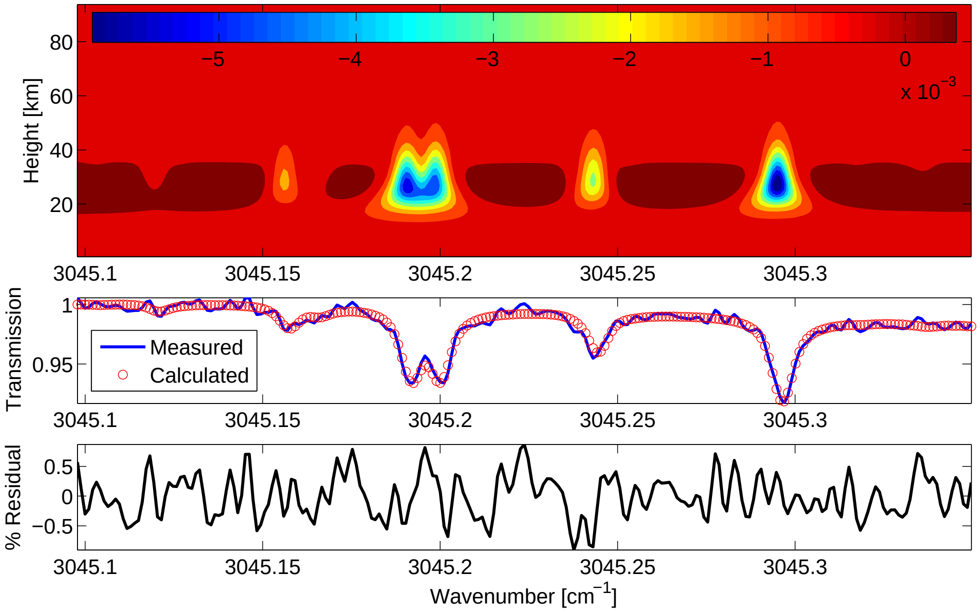


Figure 4.33 Sample fit (middle) and residual (bottom) for a transmission spectrum recorded on June 23, 2004. The weighting function matrix (top) shows the regions of vertical sensitivity due to each O_3 absorption feature in this microwindow.

4.3.1 O_3

O_3 retrievals at TAO are currently performed near 3045 cm^{-1} for lack of a long time series of high quality data in the Filter 6 region, which gives information on the commonly analyzed O_3 absorption feature near 1146 cm^{-1}. The RMS residual of the spectral fit shown in Figure 4.33 corresponds to the 2004 ensemble average SNR of ∼300, while the weighting function matrix clearly shows sensitivity to the vertical distribution of O_3 between 20-40 km. This is also expressed in the averaging kernel matrix shown in Figure 4.34. The height-averaged VMR kernels show the greatest sensitivity to the stratospheric region that matches the regions of sensitivity of $\mathbf{K}$ and $\mathbf{A}$, while the partial column kernels additionally show that the stratospheric kernel maintains sensitivity to the top of the atmosphere. In the ideal case, the partial column kernel for 15-50 km should tend to zero outside of this region; fortunately, the partial column of O_3 above 50

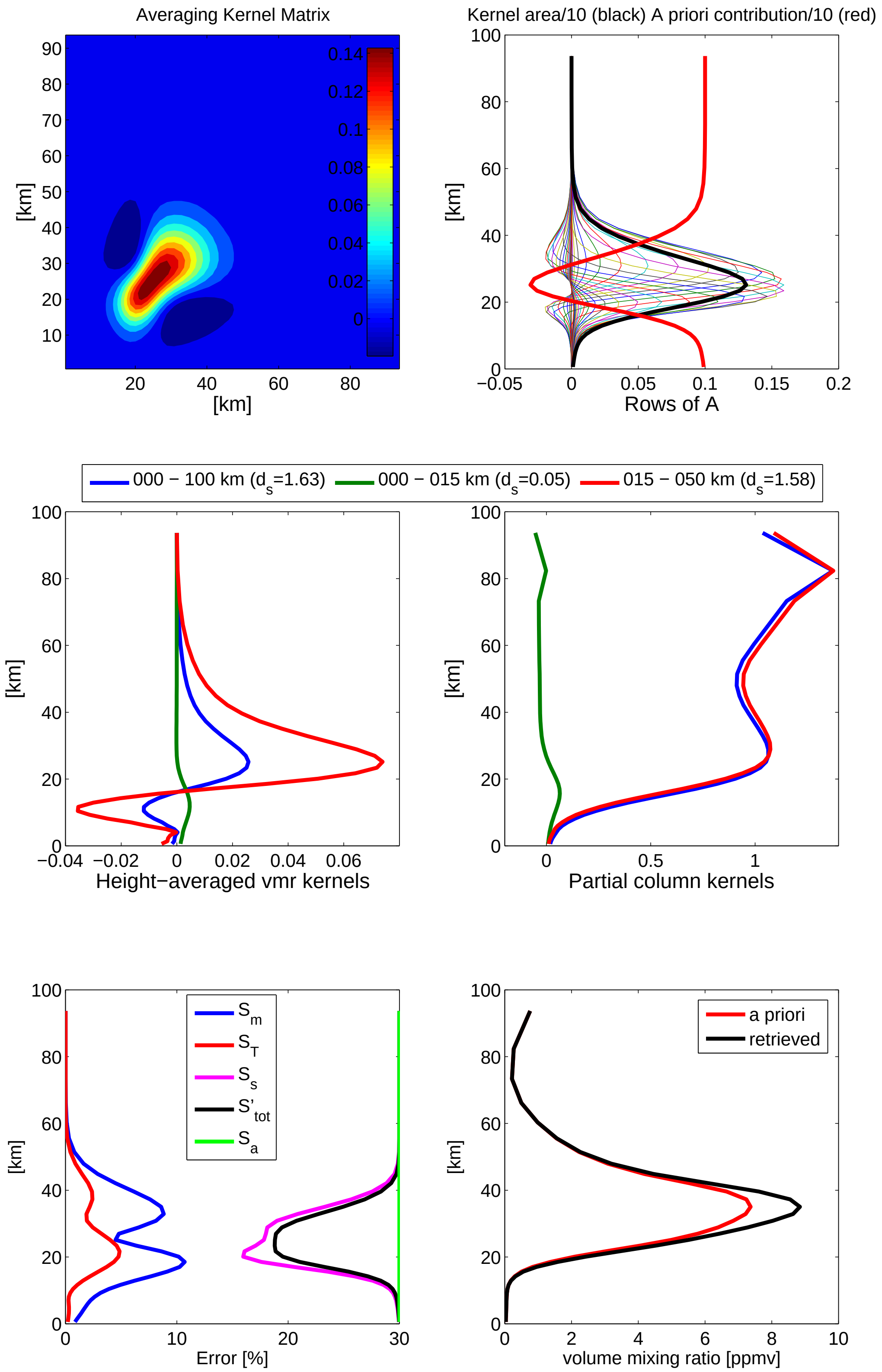


Figure 4.34 Characterization of the O_3 retrieval from Figure 4.33 (see text for details).

km is negligible. The total column degrees of freedom for signal ($d_s = 1.63$) are consistent with the shape and vertical distribution of the averaging kernels, which show more than one but less than two clearly resolved peaks.

The standard deviation of the retrieval noise error (S_m) increases in the region of sensitivity (15-50 km), while the standard deviation of the smoothing error (S_s) decreases, as per the discussion of Section 3.3.1. The smoothing error, and also the total error (S'_{tot}), shown in Figure 4.34 are conservative estimates, since the *ad hoc* prior covariance matrix used in the retrieval was also used in the characterization and error analysis calculations (i.e. HALOE-derived diagonal elements of $\mathbf{S}_a$ were less than 30%, as shown in Figure 4.15). The magnitude of the diagonal elements of S_a is shown in Figure 4.34, while its off-diagonal elements were of Gaussian form with a 4-km vertical correlation length. The standard deviations of the temperature error (S_T) are also shown and included in S'_{tot}, although we note again that the temperature profile is not a retrieved forward model parameter, nor is it a part of the effective measurement error covariance (Section 3.3.2).

4.3.2 NO

TAO retrievals of NO are currently performed near 1900 cm^{-1}, close to a prominent solar CO interference feature. The RMS residual of the spectral fit shown in Figure 4.35 corresponds to the 2004 ensemble average SNR of ∼200, although there is significant scatter in this value because of the highly variable concentration of H_2O, which controls the broad transmission level in this region of the spectrum. The weighting function matrix shows a sensitivity to the vertical distribution of NO between 25-50 km and again above 80 km, due to the exponentially increasing NO volume mixing ratio. (The sensitivity above 80 km was greatly attenuated when the Kiruna *a priori* profile of Figure 4.8 was used in the retrievals.) However, it is apparent from the averaging kernel matrix shown in Figure 4.36 that enhancements in the true profile above 80 km are in fact incorrectly attributed to the stratosphere. This phenomenon and the sensitivity of the ground-based

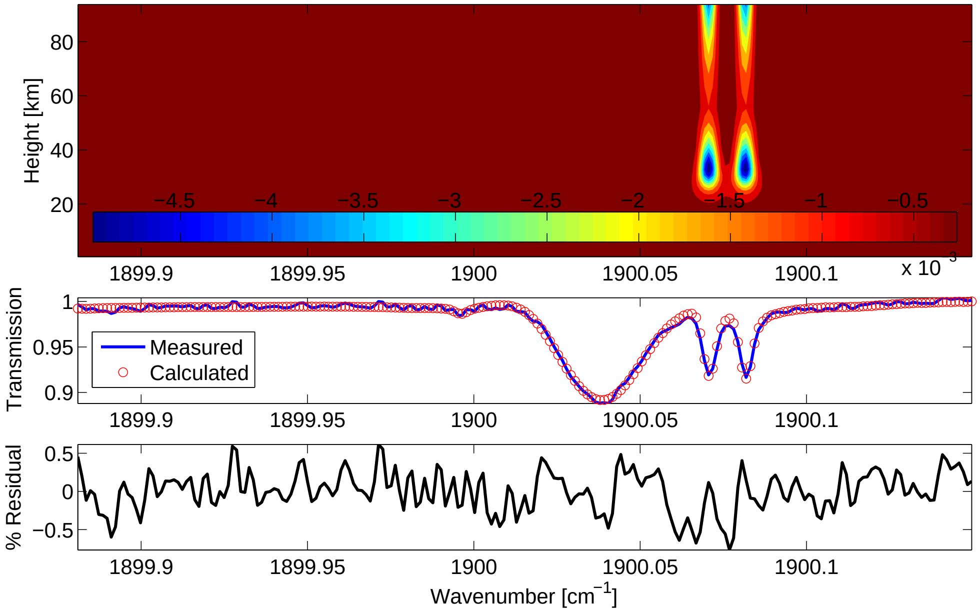


Figure 4.35 Sample fit (middle) and residual (bottom) for a transmission spectrum recorded on May 19, 2004. The weighting function matrix (top) shows the regions of vertical sensitivity due to each NO absorption feature in this microwindow.

FTIR retrieval to meso-thermospheric NO is discussed in detail in Section 6.1.

The height-averaged VMR kernel shown in Figure 4.36 is now less intuitive since the VMR profile of NO spans more than 5 orders of magnitude between 0-100 km. (The PROFFIT-9 OEM implementation of *Hase et al.* [2004] performs VMR profile retrievals on a logarithmic scale, however, this capability is not incorporated into SFIT-2 at present.) However, the stratospheric column kernel shows good sensitivity between 30-50 km and an increasing sensitivity to the top of the atmosphere, where the partial column of NO is not negligible as in the case of O_3. The mesospheric partial column kernel is shown on account of the sensitivity displayed in **K** and **A**, and it implies a very limited 50-100 km partial column sensitivity (d_s=0.10), spread over 30-100 km.

Finally, the errors are again conservatively estimated with an *ad hoc* retrieval prior covariance with 40% on the diagonal and a vertical correlation length of 4 km. While the temperature errors appear small, the retrieval noise error increases again above 60

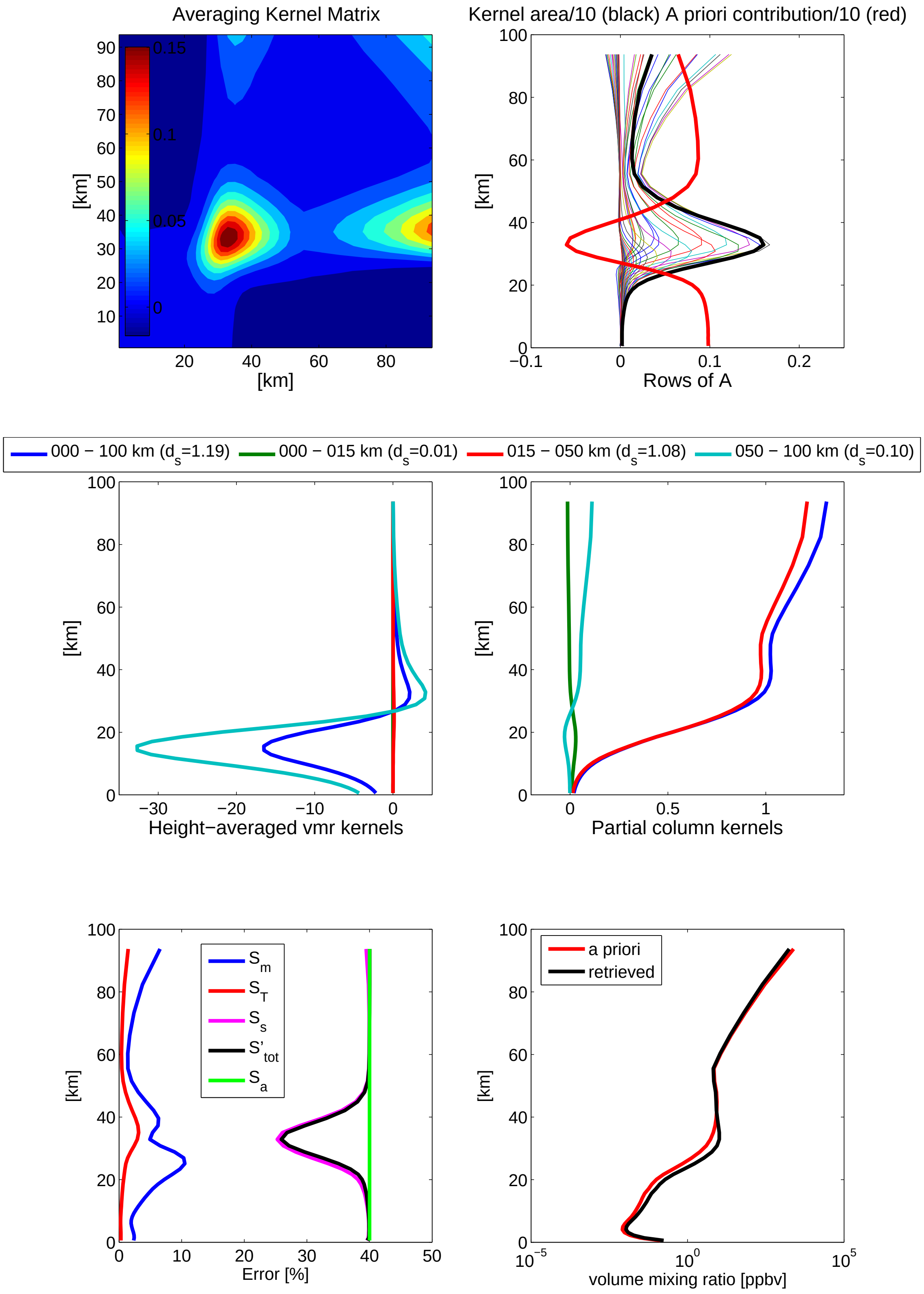


Figure 4.36 Characterization of the NO retrieval from Figure 4.35 (see text for details).

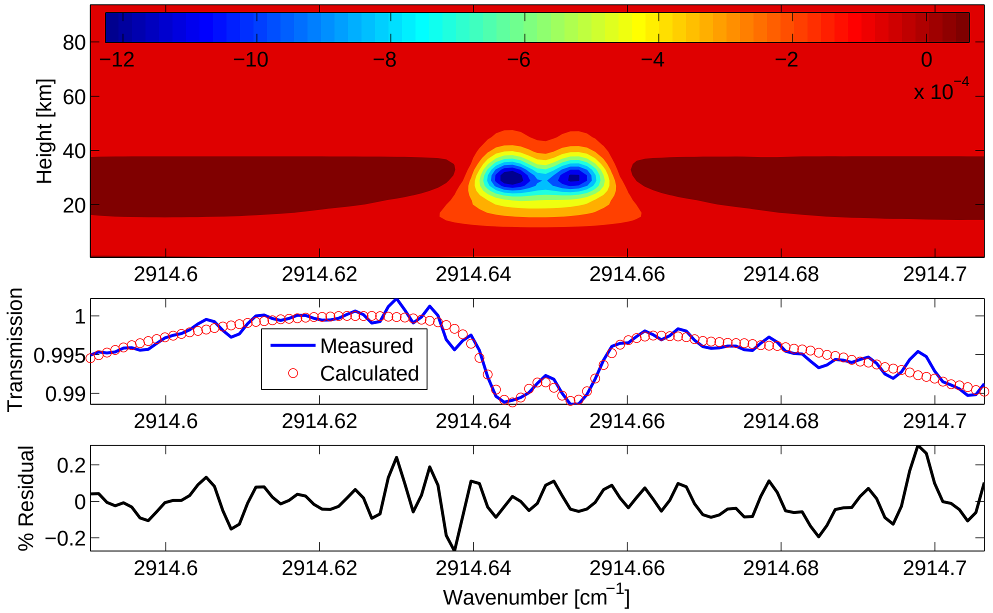


Figure 4.37 Sample fit (middle) and residual (bottom) for a transmission spectrum recorded on November 13, 2004. The weighting function matrix (top) shows the regions of vertical sensitivity due to each NO_2 absorption feature in this microwindow.

km, which is consistent with the sensitivity shown in **K**. The smoothing error remains nearly unchanged above 60 km, likely due to the low values of $\mathbf{S}_a$ used in the retrievals (40%) as compared to the HALOE values from Figure 4.13 (∼70%).

4.3.3 NO_2

NO_2 is currently retrieved from the very weak absorption feature near 2915 cm^{-1}, although the RMS residual of the spectral fits corresponds to an ensemble average SNR of ∼1000. Figure 4.37 shows the weak sensitivity of this retrieval to the vertical distribution of NO_2 between 20-40 km, which is again reflected in the averaging kernel matrix shown in Figure 4.40. The height-averaged VMR kernel for 15-50 km and the stratospheric column kernel show good sensitivity to this region. The total column degrees of freedom for signal (d_s=1.46) are again consistent with the shape of the averaging kernels, which show just slightly more than one clearly resolved peak.

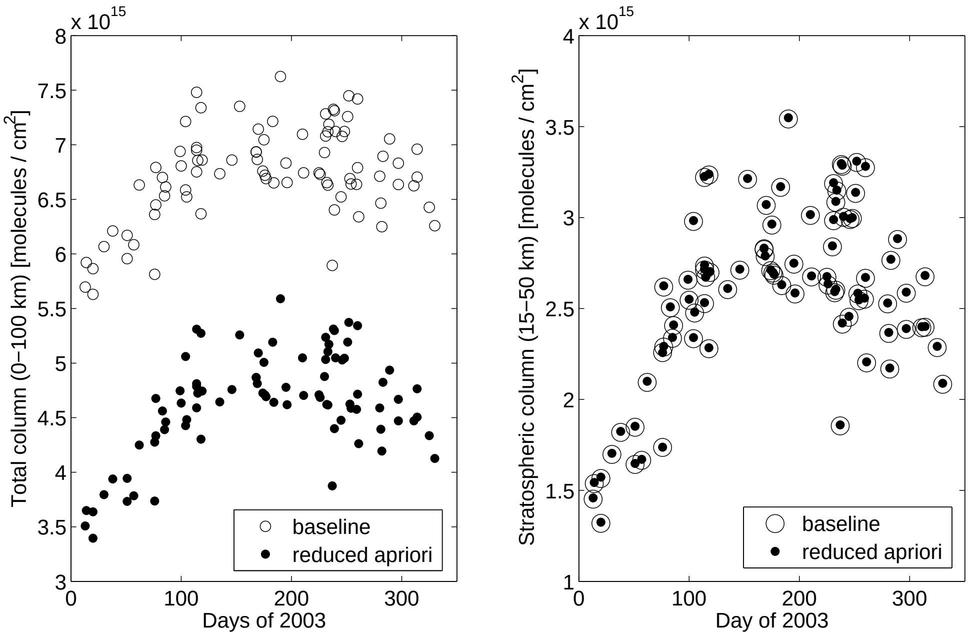


Figure 4.38 Response of total (left) and stratospheric (right) column retrievals of NO_2 to different tropospheric *a priori* profiles (see text for details).

The diagonal elements of the prior covariance used in this retrieval were set to 40% between 10-55 km and were tapered to 10% outside of this region; this is one approach to retrievals suggested in the NDACC IRWG community whereby the retrieval is constrained in the null space. This choice of prior covariance is seen to affect the partial column kernel characterization near 100 km, and the smoothing error near 10 km. Nonetheless, the biasing effect of this $\mathbf{S}_a$ is small because the retrieval does not have sensitivity to the troposphere in the first place, despite the high NO_2 VMR near the surface. To illustrate the latter point, retrievals were performed on an ensemble of 2003 TAO spectra with two *a priori* profiles: one as shown in Figure 4.9 (the "TAO" profile referred to as *"baseline"* here) and the other with tropospheric VMR values reduced to 0.5 ppbv (*"reduced a priori"*). Figure 4.38 shows that while the total column was greatly affected by this *a priori* change, the stratospheric column remained virtually identical. This is consistent with the fact that the tropospheric region is a part of the null space of the state space and hence the retrieved solution in this region is derived entirely from the *a priori*.

Conversely, the stratospheric column is well-determined by the measurements and shows the same seasonal cycle regardless of the tropospheric *a priori* values. Nonetheless, a change in the stratospheric portion of the *a priori* is expected to have a small biasing effect on the stratospheric column due to the non-ideal sensitivity of **A**.

Figure 4.39 illustrates why the tropospheric region falls in the null space, this time with a simulated spectrum of NO, to which the same reasoning applies. We show the difference in transmission between an NO spectrum simulated using a VMR of 10^{-11} in the troposphere versus the much higher *a priori* values from the MIPAS reference profile (Figure 4.8). The two NO peaks visible in Figure 4.35 are blended together due to pressure broadening near the surface, however, the maximum transmission difference is only 0.065 %, which would require an SNR of ∼3000 in our spectra in order to be detected with a relative SNR of ∼2. Since the TAO SNR in this region is on average 200, the tropospheric region of state space remains inaccessible to the measurements and falls in the null space. Consequently, in an urban setting, where concentrations of NO_x are high – and highly variable – it is misleading to quote the total column of either NO or NO_2, as only stratospheric values can be reported with confidence (although even further complications arise in the case of NO, as described in Section 6.1).

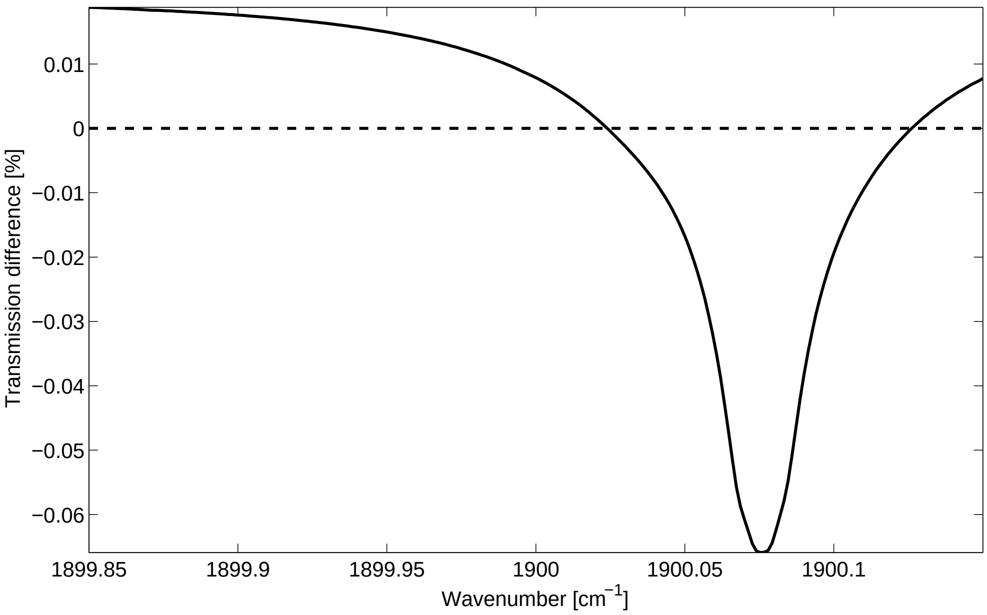


Figure 4.39 Simulated difference in the transmission spectrum of NO for two different tropospheric *a priori* values (see text for details).

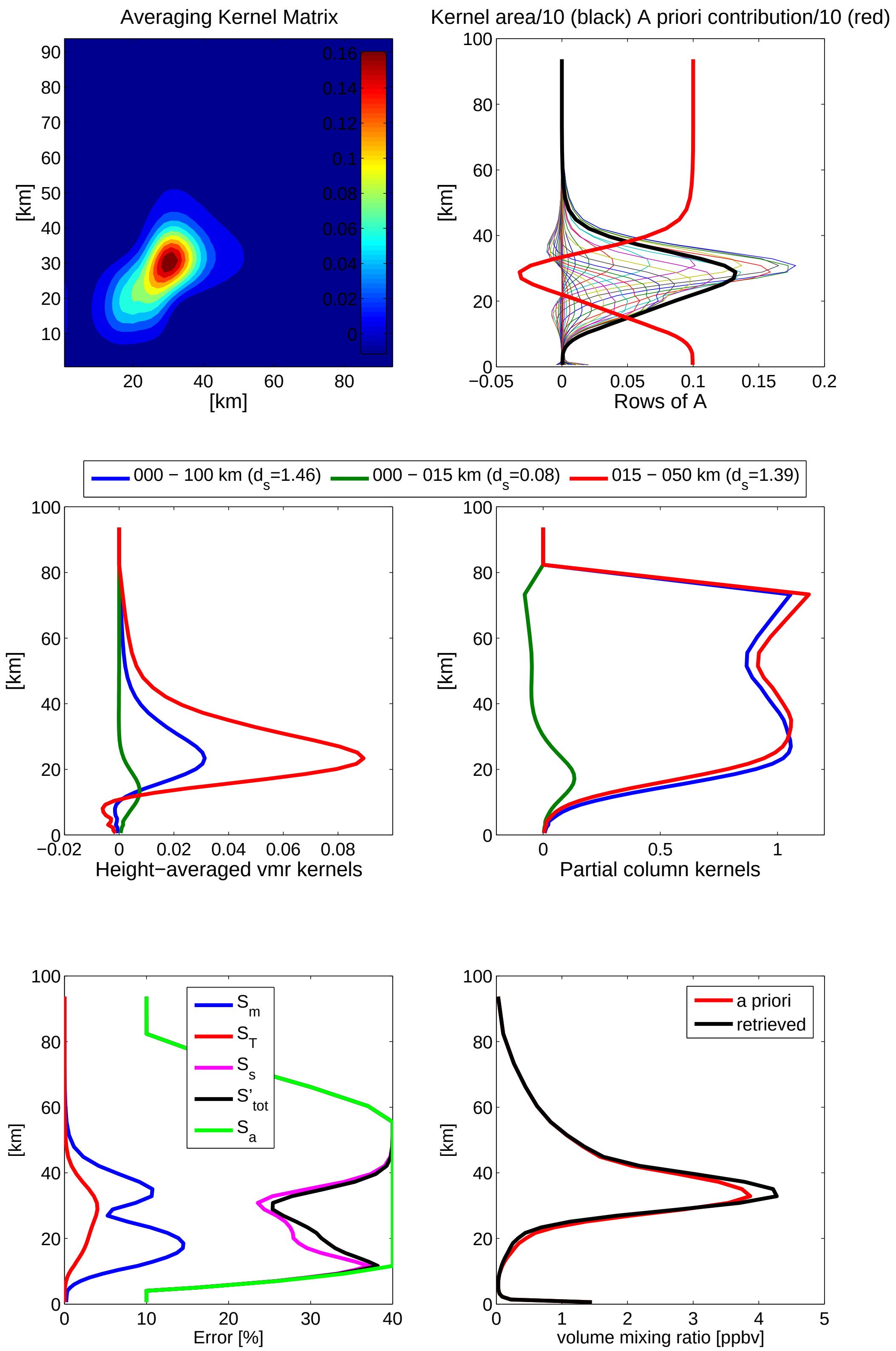


Figure 4.40 Characterization of the NO_2 retrieval from Figure 4.37 (see text for details).

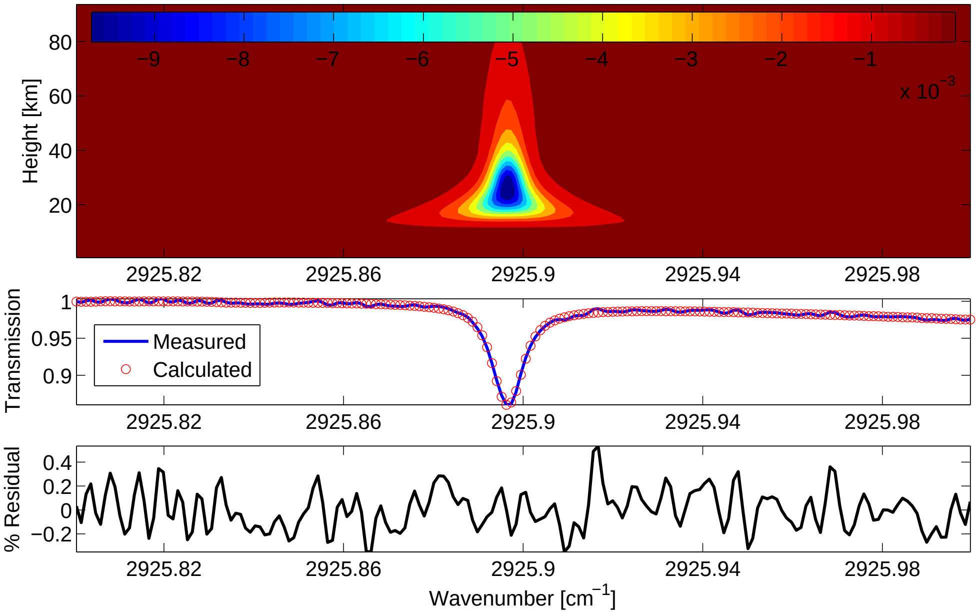


Figure 4.41 Sample fit (middle) and residual (bottom) for a transmission spectrum recorded on July 6, 2004. The weighting function matrix (top) shows the regions of vertical sensitivity due to the HCl absorption feature in this microwindow.

4.3.4 HCl

HCl retrievals at TAO are currently performed on the spectral absorption feature near 2925 cm^{-1} (superimposed on the wing of a CH_4 feature) recorded with an ensemble average SNR of ∼700. The weighting function matrix (Figure 4.41) shows good sensitivity to HCl throughout the stratosphere (15-50 km), which is also reflected in the averaging kernel matrix (Figure 4.42) and the stratospheric column kernel (d_s=2.41). While the tropospheric portion of the HCl VMR profile also falls in the retrieval null space, tropospheric HCl concentrations are not as variable or as large as those of NO_x, and hence quoting the total column of HCl is much more accurate.

The diagonal elements of $\mathbf{S}_a$ (with vertical correlations of 4 km) were also tapered to 10% outside of 10-55 km, with a small biasing effect on the partial column kernels and the smoothing error estimate. The temperature error estimate approaches approximately

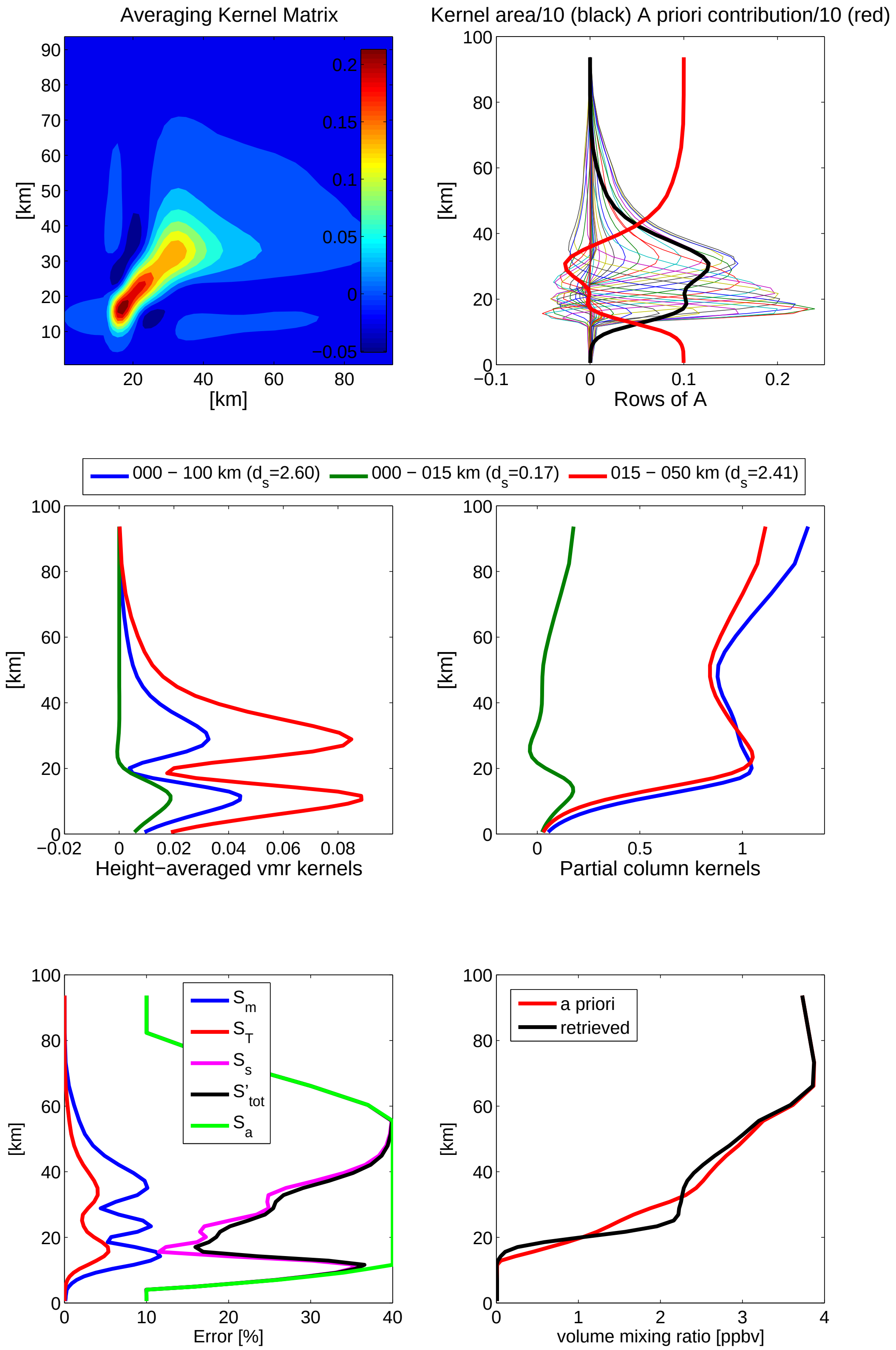


Figure 4.42 Characterization of the HCl retrieval from Figure 4.41 (see text for details).

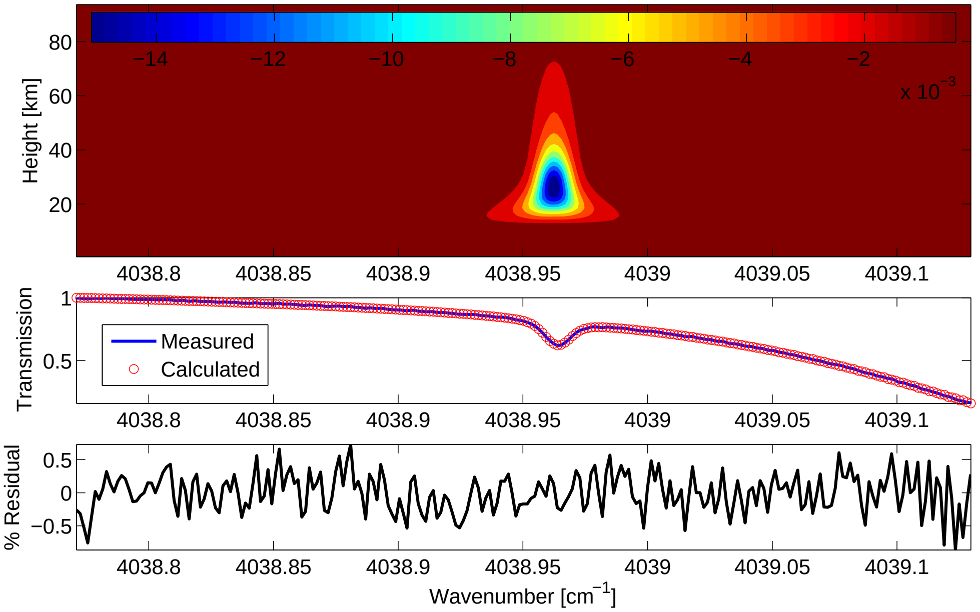


Figure 4.43 Sample fit (middle) and residual (bottom) for a transmission spectrum recorded on June 12, 2004. The weighting function matrix (top) shows the regions of vertical sensitivity due to the HF absorption feature in this microwindow.

half that of the retrieval noise error estimate, as in the case of O_3 retrievals (Figure 4.34).

4.3.5 HF

HF retrievals at TAO are currently performed on the spectral absorption feature near 4038 cm^{-1} (superimposed on the wing of a H_2O feature) recorded with an ensemble average SNR of ~300. The weighting function matrix (Figure 4.43) shows good sensitivity to HF throughout the stratosphere (15-50 km), which is also reflected in the averaging kernel matrix (Figure 4.44) and the stratospheric column kernel (d_s=2.17). Again, while the tropospheric portion of the HF VMR profile also falls in the retrieval null space, tropospheric HF concentrations are not as variable as those of NO_x, and hence quoting the total column of HF is much more accurate than in the case of NO_2 or NO.

The diagonal elements of $\mathbf{S}_a$ (with vertical correlations of 4 km) were also tapered to 10% outside of 10-55 km, with a small biasing effect on the partial column kernels

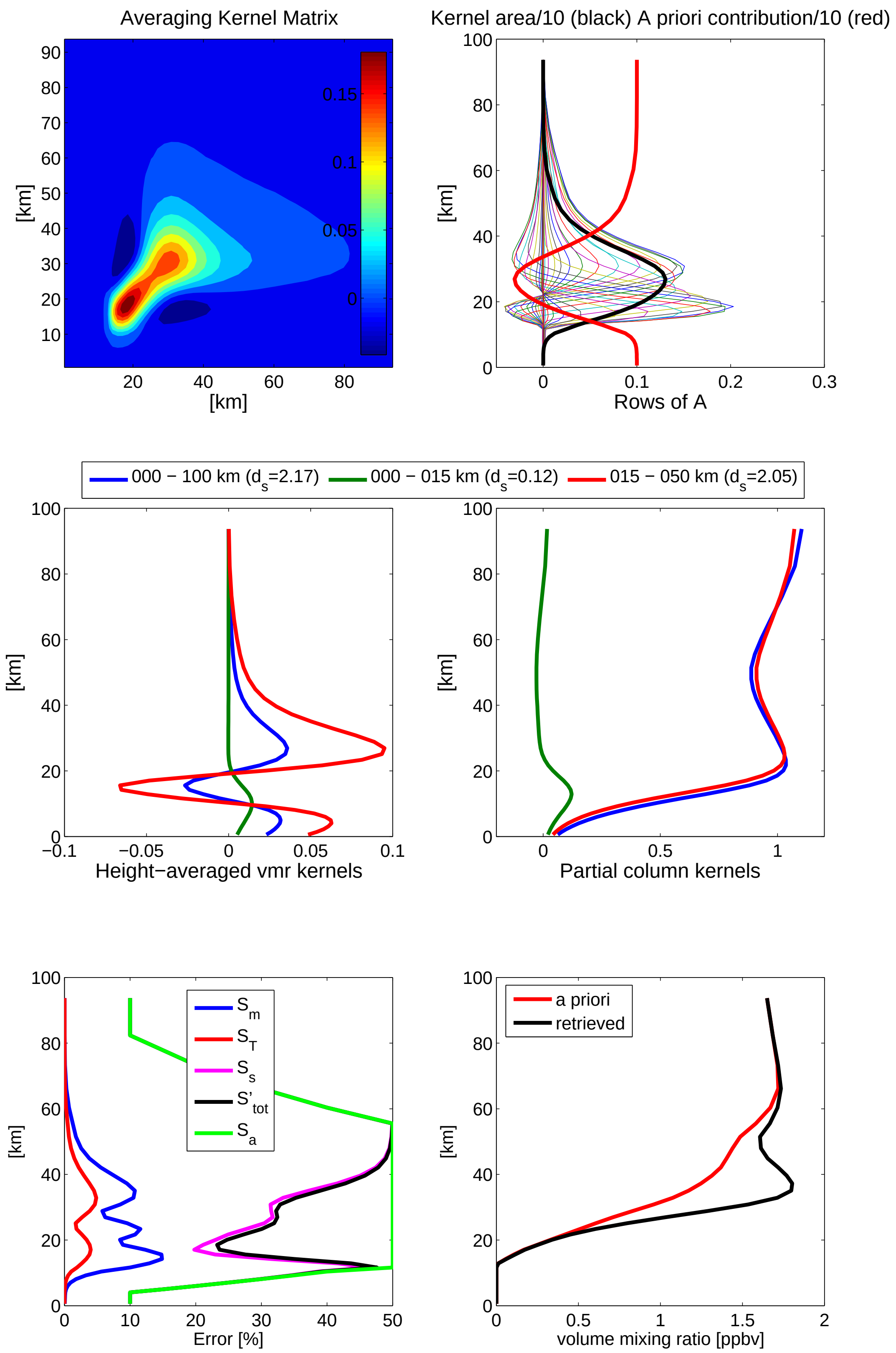


Figure 4.44 Characterization of the HF retrieval from Figure 4.43 (see text for details).

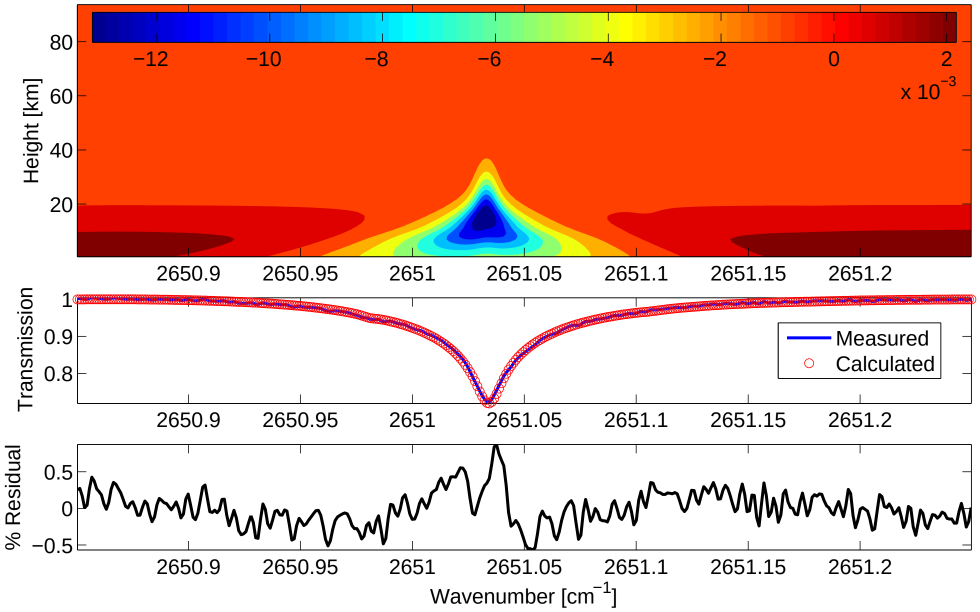


Figure 4.45 Sample fit (middle) and residual (bottom) for a transmission spectrum recorded on January 19, 2004. The weighting function matrix (top) shows the regions of vertical sensitivity due to the CH_4 absorption feature in this microwindow.

and the smoothing error estimate. While S_T is comparable to that of HCl, S_m is higher, consistent with the lower average SNR in HF spectra.

4.3.6 CH_4

The optimization of retrievals of CH_4 from TAO spectra is an active topic of research presently pursued by Ph.D. Candidate Jeffrey R. Taylor. Here we present the characterization of a CH_4 retrieval based on absorption features contained in three separate microwindows between 2650 and 2675 cm^{-1} (Figures 4.45-4.47). The average SNR across the three microwindows is ∼500 in 2004, as determined from the *detail* file, and ∼600 as determined from the *pbpfile* file (Section 3.4.3.4). The individual SNR derived from the *pbpfile* for each microwindow is ∼600, ∼720, and ∼550, therefore, the use of a single custom SNR derived from the *detail* file for all three microwindows is justified. (If this is not the case for a particular multi-microwindow retrieval, SFIT-2 allows the user to de-weight any region of the combined spectrum to the desired SNR level.)

The three weighting function matrices show good sensitivity to CH_4 up to ∼35 km, with the most strongly absorbing line showing the highest values of sensitivity near the line center. The averaging kernel matrix mirrors this sensitivity (Figure 4.48), and the height-averaged VMR kernel also shows the highest sensitivity to the troposphere (d_s=2.5) and a substantial sensitivity to the stratosphere (d_s=1.15). The partial column kernels additionally reveal that the retrieval overestimates stratospheric column enhancements but compensates for them in the troposphere, thus “conserving” the total column, which displays a nearly ideal kernel.

The % *a priori* contribution, the kernel area and also the smoothing error estimate all display a conspicuous kink at 15 km, which is entirely user-induced (there is also a discontinuity in the averaging kernel matrix near this altitude). In the course of investigating the extent of the prominent oscillations displayed by CH_4 retrievals (also noted by other NDACC groups retrieving CH_4 profiles), the prior covariance was changed from a diagonal matrix with 20% at all heights to one that has a 4 km vertical correlation length, but only between 0-15 km (shown here). This choice was made based on the assumption that the troposphere is well-mixed. Although the effect on retrieval characterization and error analysis is striking, we note that the oscillatory character of the average retrieved profile (based on 79 quality-controlled spectra analyzed in 2004) was not visibly affected by the use of this modified $\mathbf{S}_a$. Moreover, the oscillations in the average profile were not visibly affected by adding the use of custom SNR values in addition to the tropospheric $\mathbf{S}_a$ described above; this result is consistent with the fact that oscillations are a normal feature of ground-based FTIR retrievals, although spurious oscillations can occur when the retrieval SNR truly exceeds the real noise levels in the spectrum. However, one retrieval setting did have an impact on the average profile retrieved in 2004: the SFIT-2 switch to retrieve the simple phase of the FTS ILS. While the 2004 mean RMS spectral fit residual remained at 0.179% for the original retrievals, as well as those with tropospheric

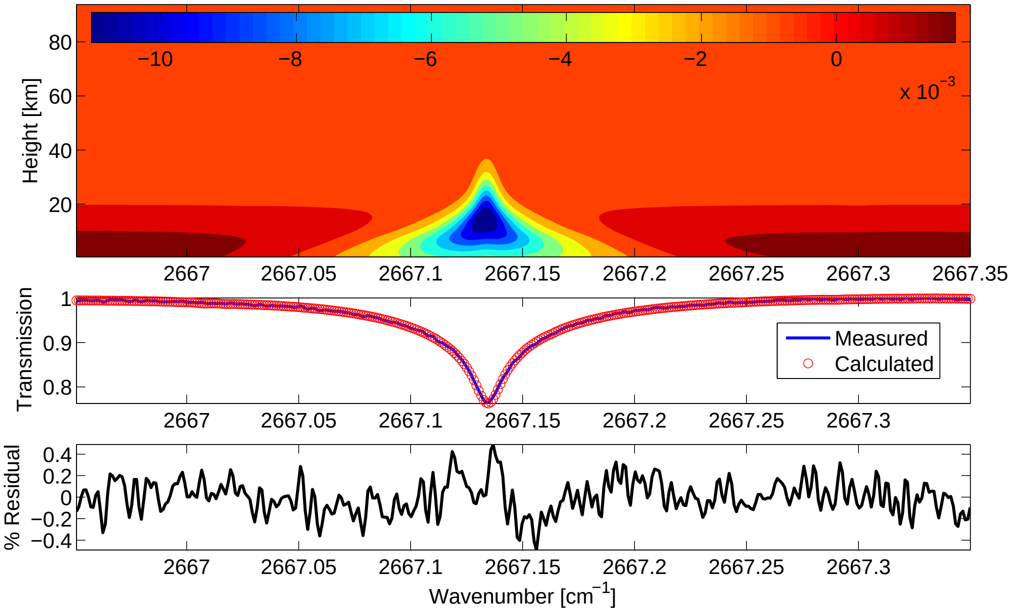


Figure 4.46 As in Figure 4.45, but for second microwindow used in CH_4 retrieval.

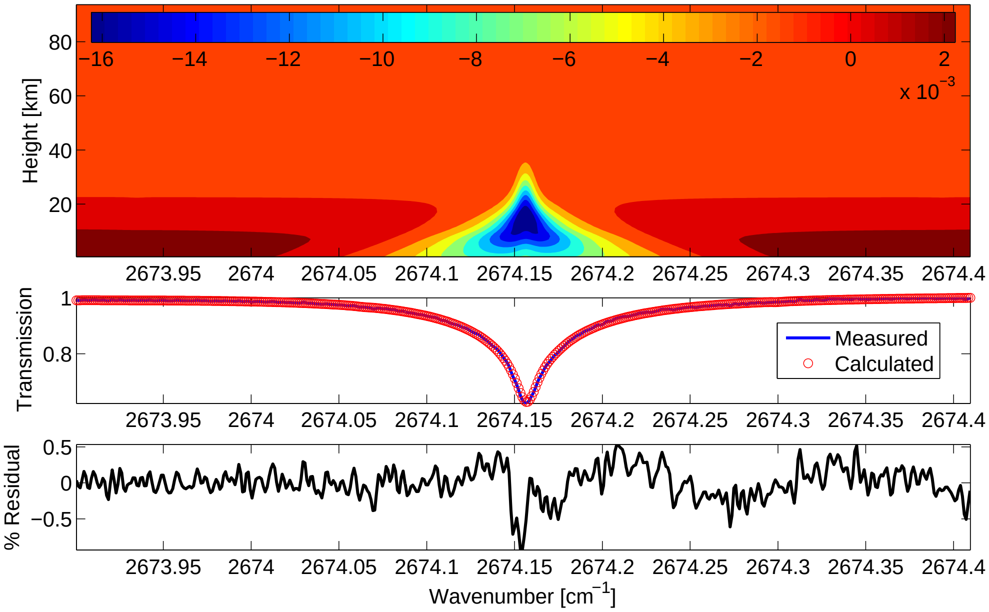


Figure 4.47 As in Figure 4.45, but for third microwindow used in CH_4 retrieval.

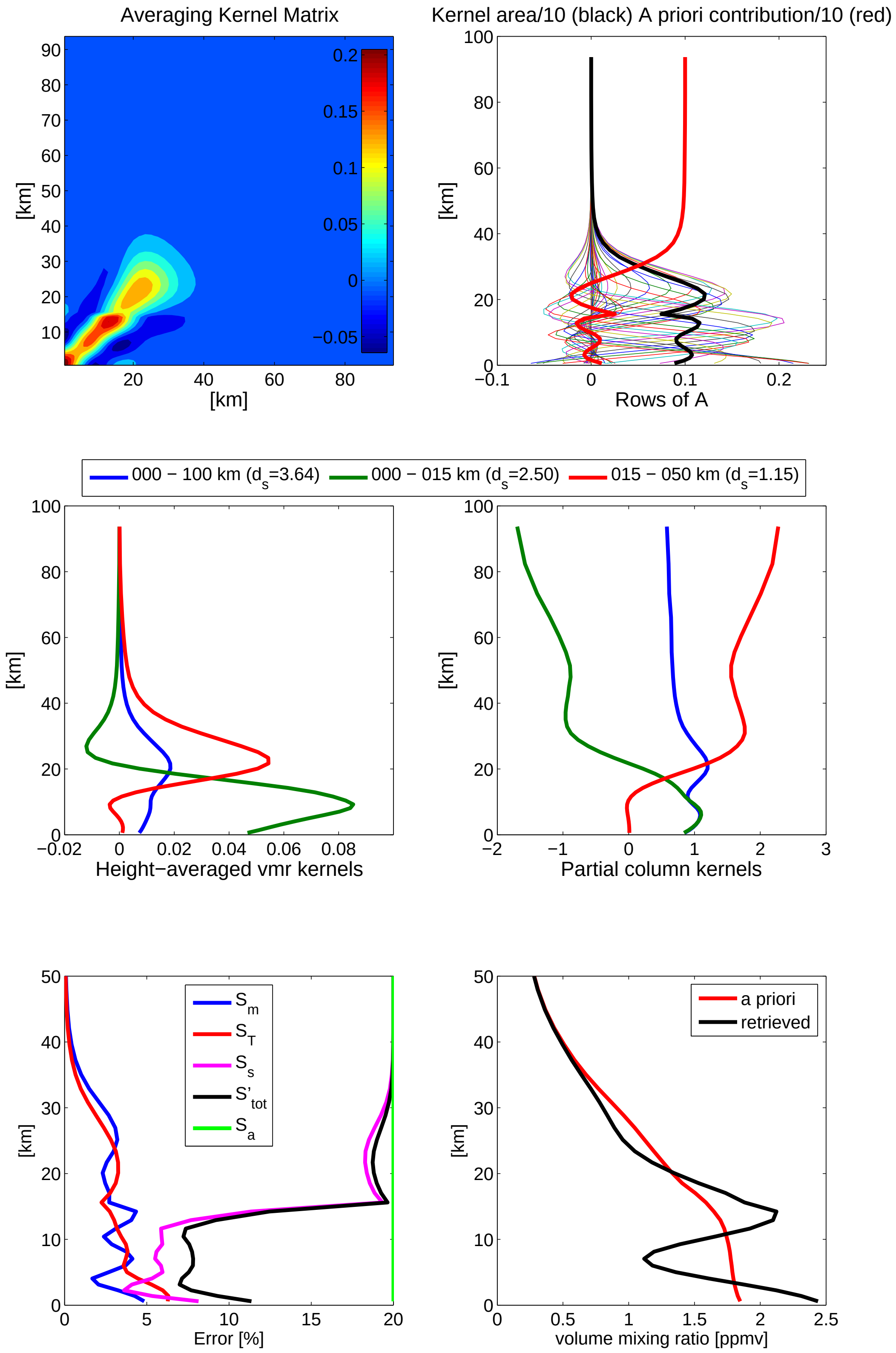


Figure 4.48 Characterization of CH_4 retrieval from Figures 4.45-4.47 (see text for details).

correlations and then custom SNR values (applied cumulatively), the further addition of the ILS parameter to the state vector reduced the mean RMS spectral fit residual to 0.168% and increased the mean retrieved total column by 1.8%. It is possible that the ILS was used as a compensating parameter for shortcomings in the spectral database, in particular the known failures of the Voigt model in the presence of line mixing. Clearly, the characterization of the FTS ILS has some bearing on retrievals of total column amounts, and a much greater impact on the retrievals of partial columns and vertical profiles, which rely heavily on modeling the ILS correctly.

Finally, the temperature error in this retrieval is comparable to the retrieval noise error at all heights, and becomes bigger near the surface. In Section 5.2.6, it is shown to be the dominant source of error in retrievals of total and tropospheric CH_4 columns.

4.3.7 N_2O

N_2O retrievals at TAO are currently performed near 2481 cm^{-1}. The RMS residual of the spectral fit shown in Figure 4.49 corresponds to the 2004 ensemble average SNR of $\sim$300. The weighting function matrix shows good sensitivity to the vertical distribution of N_2O between 0-30 km, with a small reduction near the very center of the absorption feature due to 70% saturation. The averaging kernel matrix expresses a similar sensitivity to N_2O (Figure 4.50), with the height-averaged VMR kernel also showing the highest sensitivity to the troposphere (d_s=2.77) and a substantial sensitivity to the stratosphere (d_s=1.19). Similar to CH_4, this retrieval overestimates stratospheric column contributions (stratospheric kernel > 1) but seems to compensate for them in the troposphere.

The retrieval characterization and error estimates suffer from a kink near 15 km for the same reasons as in the case of CH_4. A similar investigation of oscillations in retrieved N_2O profiles revealed that they remain unchanged by the introduction of vertical correlations between 0-15 km, and later by the use of custom SNR values in each retrieval in 2004. The addition of ILS fitting had a much smaller effect in the case of N_2O, possibly because

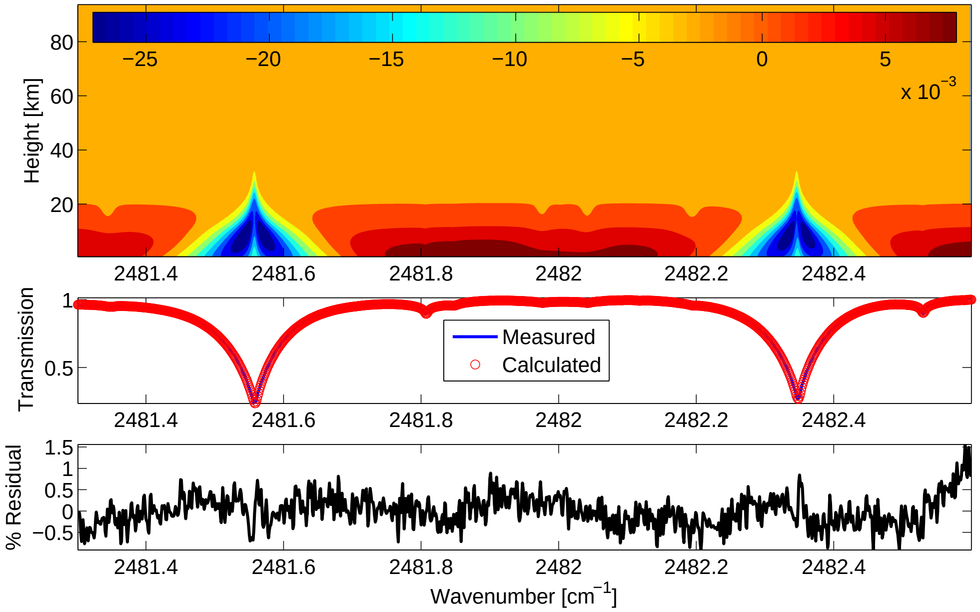


Figure 4.49 Sample fit (middle) and residual (bottom) for a transmission spectrum recorded on November 12, 2004. The weighting function matrix (top) shows the regions of vertical sensitivity due to each N_2O absorption feature in this microwindow.

line mixing is less of an issue for this less abundant trace gas. Finally, similarly to CH_4, the estimated temperature error is large near the surface, and again becomes the dominant source of error in the retrievals of total and partial N_2O columns (Section 5.2.7).

4.3.8 CO

TAO retrievals of CO are performed using the three microwindows shown in Figures 4.51-4.53. The RMS residual of the spectral fits corresponds to an ensemble average SNR of ∼250. Systematic features in the residuals are due to difficulties in modeling the ubiquitous interferences from solar CO lines, as reported by other investigators, e.g. *Zhao et al.* [2002]. This aspect of CO retrievals can now be further optimized with the use of SFIT-2 v.3.91 (recently implemented at TAO), which employs the improved solar line database and theoretical model of Dr. Frank Hase.

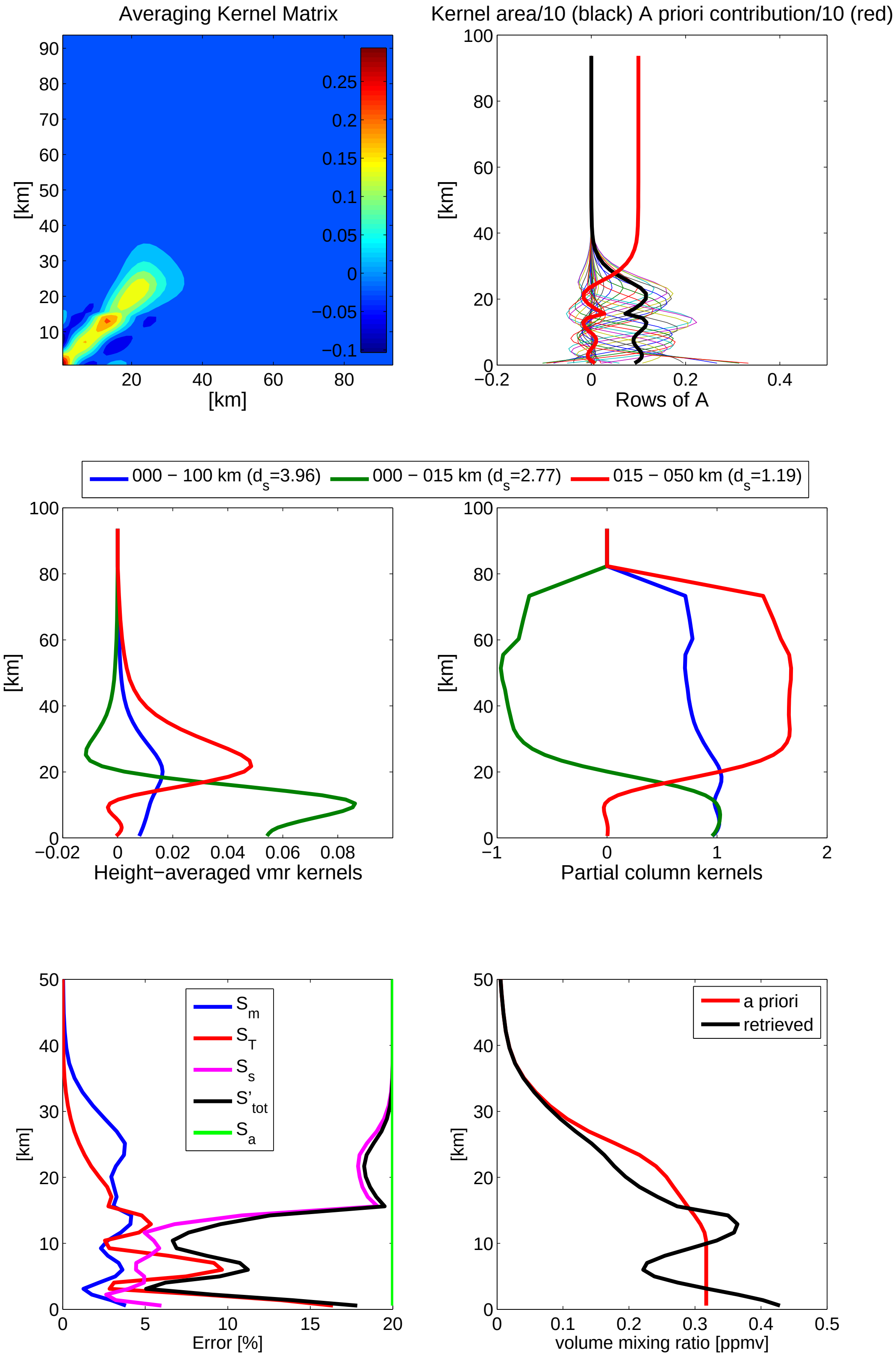


Figure 4.50 Characterization of the N_2O retrieval from Figure 4.49 (see text for details).

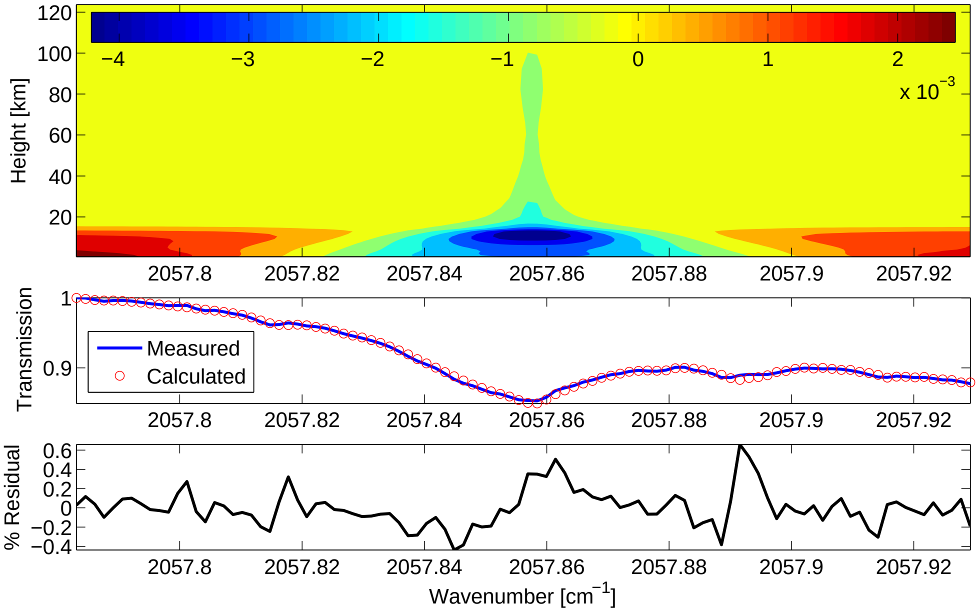


Figure 4.51 Sample fit (middle) and residual (bottom) for a transmission spectrum recorded on May 19, 2004. The weighting function matrix (top) shows the regions of vertical sensitivity due to the CO absorption feature in this microwindow.

The use of the saturated CO absorption feature near 2158 cm^{-1} greatly increases retrieval sensitivity to the lowermost troposphere, which may or may not be desired, depending on the purpose of the data. For example, when comparisons with MOPITT satellite measurements of CO are desired (Section 6.3), this microwindow can be excluded from the analysis of TAO spectra in order to reduce the sensitivity to the surface, and hence match MOPITT's vertical sensitivity more accurately. On the other hand, if comparisons with SCIAMACHY (Scanning Imaging Absorption Spectrometer for Atmospheric Chartography) satellite measurements are desired, then retaining this microwindow will match retrieval sensitivity to SCIAMACHY's greater ability to sense the surface [*Bovensmann et al.*, 1999].

The retrieval and characterization shown in Figure 4.54 was performed with SFIT-2 v.3.91 on a vertical grid from 0-130 km in order to ascertain any upper atmospheric sensitivity to CO, which also possesses an exponentially increasing VMR profile, like NO.

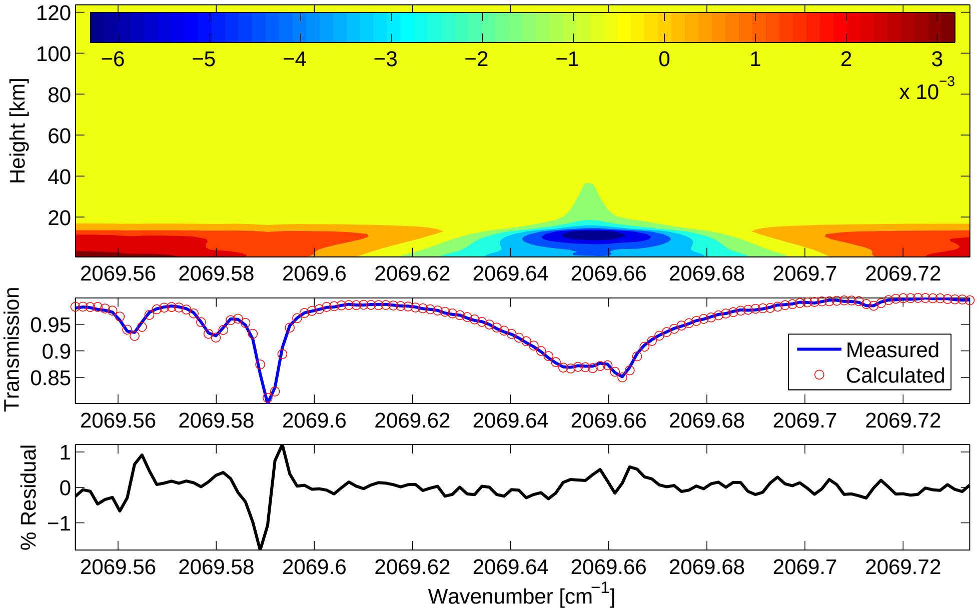


Figure 4.52 As in Figure 4.51, but for second microwindow used in retrieval.

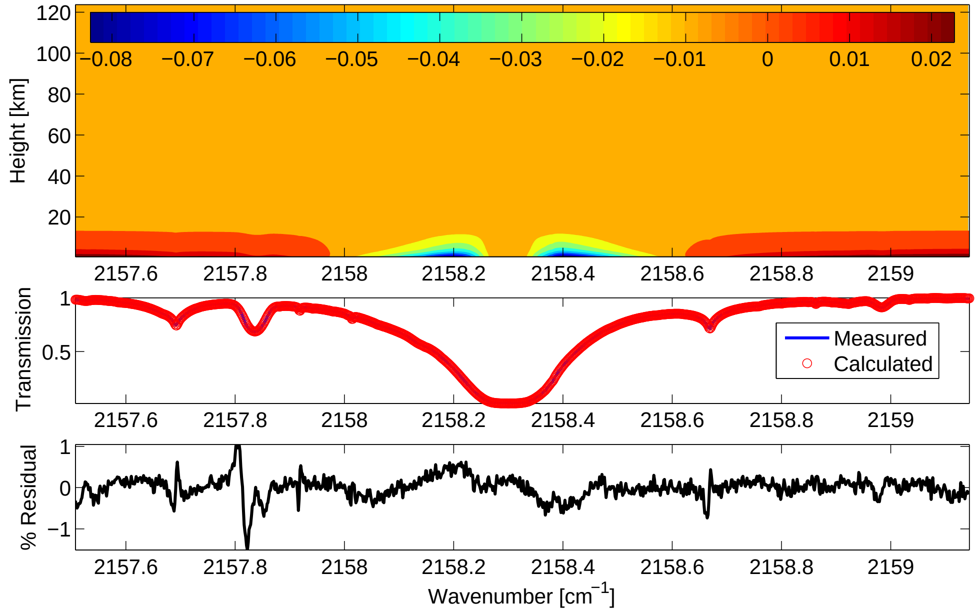


Figure 4.53 As in Figure 4.51, but for third microwindow used in retrieval.

	[0-130 km]		[0-15 km]		[15-50 km]	
d_s (σ)	2.61	(0.09)	2.43	(0.06)	0.17	(0.04)
%a (σ)	0.83	(0.29)	-0.94	(0.14)	65.13	(4.67)
S_s (σ)	0.09	(0.01)	0.21	(0.01)	7.10	(0.12)
S_m (σ)	0.28	(0.02)	0.29	(0.02)	1.86	(0.12)
S_T (σ)	2.59	(0.05)	2.66	(0.06)	0.66	(0.06)
S'_{tot} (σ)	2.61	(0.05)	2.69	(0.06)	7.38	(0.09)

Table 4.4 2004 average CO total and partial column retrieval information content and percent errors, as defined in Table 4.1 (σ = one standard deviation of the mean.)

Although measurement error increases slightly near 80 km, indicating some sensitivity to CO at this altitude, this sensitivity is not distinguishable in **K** and **A** against the tropospheric backdrop. Furthermore, it is not expressed in the height-averaged or partial column kernels, which show that virtually all of the information contained in the total column (d_s=2.52) originates in the troposphere (d_s=2.38) and stratosphere (d_s=0.13). However, we note one other ground-based FTIR study where, with some tuning, the investigators were able to detect the winter-time downwelling of CO into the stratosphere over Alaska [*Kasai et al.*, 2005].

Finally, the smoothing error estimate shown in Figure 4.54 is conservative on account of the retrieval and characterization being performed using a diagonal $\mathbf{S}_a$ with a 20% standard deviation at all heights and without vertical correlations. The retrieval noise error is surpassed by the temperature error near the surface, the latter again being the dominant source of error in retrievals of total and tropospheric columns. Table 4.4 summarizes the errors, degrees of freedom for signal, and % contribution from the *a priori* for an ensemble of CO retrievals performed on 2004 spectra (as compared to d_s for a single spectrum in Figure 4.54). Such a table of diagnostics is provided in Chapter 5 for each gas submitted to the NDACC database in 2005.

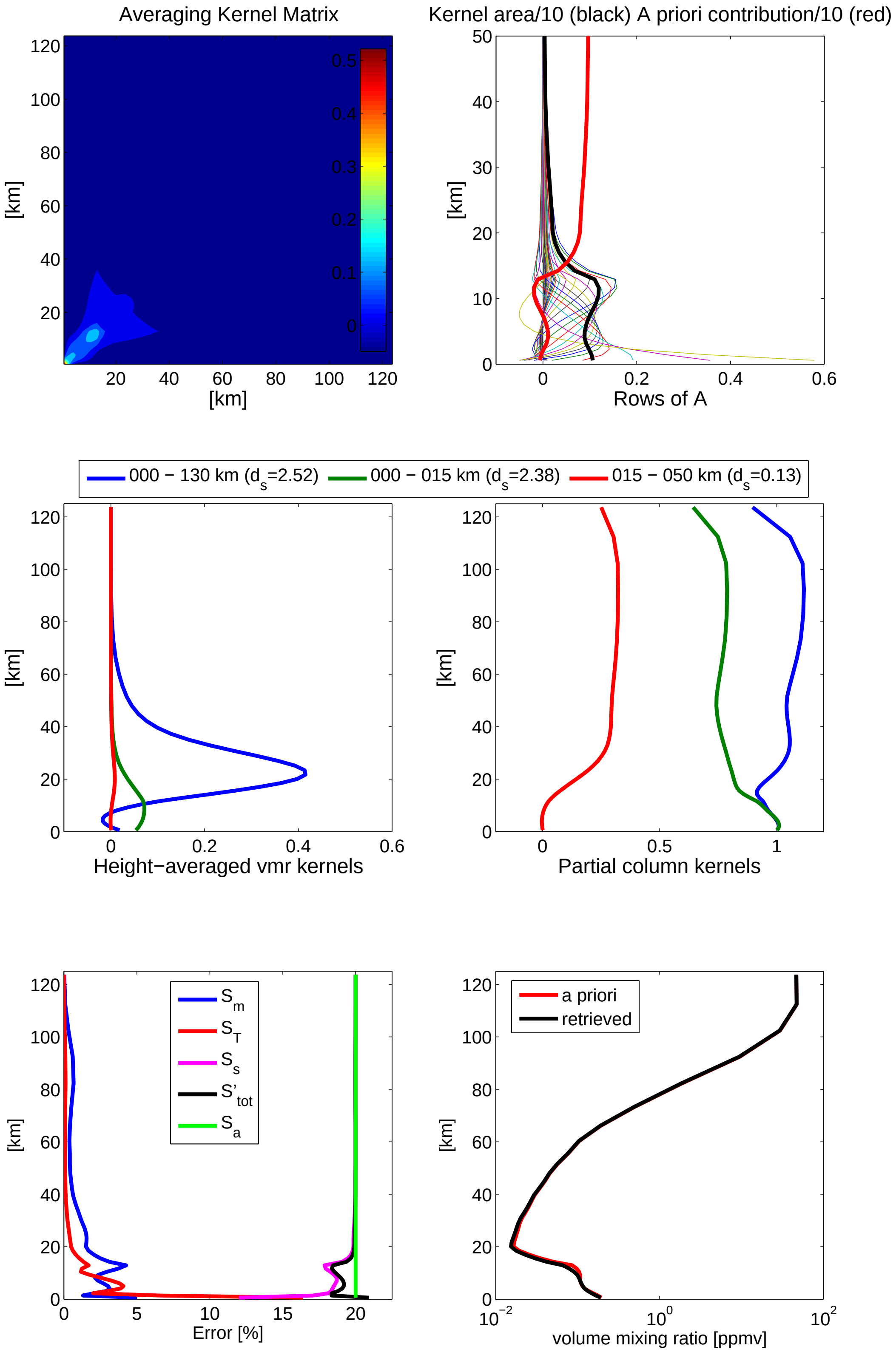


Figure 4.54 Characterization of the CO retrieval from Figure 4.51 (see text for details).

Chapter 5

Retrievals for NDACC Certification and Database Submission

In the first section of this chapter we present the results of a *Retrieval Algorithm User Intercomparison* exercise, on the basis of which TAO was granted the status of a Complementary observation station of NDACC (then, the NDSC) in March, 2004. (This section was published in the refereed literature [*Wiacek et al.*, 2006b] and is reproduced here with permission of the American Meteorological Society.) The purpose of this exercise was to ascertain the data analysis and retrieval characterization capabilities at TAO, in lieu of a side-by-side instrument intercomparison, which is technically a requirement for admission into NDACC. This requirement was waived as it was not possible to arrange for the travelling FTS used in previous NDACC intercomparisons to be brought to Toronto. However, the TAO FTS has undergone an extended side-by-side instrument intercomparison known as the mini-MANTRA campaign (May-Sep. 2005), wherein spectra were recorded at TAO with two other FTS instruments using a shared solar beam. The first FTS was a low-resolution (0.02 cm^{-1}) Bomem DA5 recently refurbished by Ph.D. Candidate Debra Wunch. The second FTS was an adaptation of the ACE-FTS satellite instrument (0.02 cm^{-1}) known as PARIS-IR, and also manufactured by ABB Bomem, Inc. Preliminary results of that study are encouraging and a publication on this work, led by another TAO Ph.D. Candidate Jeffrey R. Taylor, is currently in preparation.

In the second section of this chapter, we characterize TAO retrievals performed on spectra recorded between Oct. 2001 - Dec. 2004, which were submitted to the NDACC database in June, 2005. This section builds on the characterization of individual retrievals shown in Section 4.3. The noise and information content characteristics of TAO data are discussed further using a full year of observations (2004), and finally, time series of trace gases spanning 2001-2004 are shown.

5.1 Retrieval Algorithm User Intercomparison

Side-by-side instrument and algorithm intercomparisons comprise a part of the NDACC instrument and site validation protocol (http://www.ndsc.ncep.noaa.gov/protocols/appendix4.html). Instrument side-by-side intercomparisons of retrieval results have been described by *Paton-Walsh et al.* [1997], *Goldman et al.* [1999], *Griffith et al.* [2003] and *Meier et al.* [2005]. Since these studies included different retrieval algorithms (sometimes employing different retrieval techniques), they did not focus on details of differences due to different retrieval constraints. *Hase et al.* [2004] focus specifically on an intercomparison of two different software implementations of the OEM retrieval and they describe results for matched and unmatched constraints. Our study differs in that we describe how the results obtained by different users can differ when the same retrieval algorithm is applied to the same set of spectra. However, the root causes of any differences in retrieval results are, of course, the implicit and explicit constraints necessitated by the OEM approach.

In ideal circumstances, a side-by-side instrument intercomparison is highly desirable, provided that instrument alignment and retrieval constraint differences can be accounted for. However, since it was not possible to arrange for the travelling FTS that was used in past NDACC intercomparisons to be shipped to Toronto, a validation exercise was designed to demonstrate the retrieval capabilities of the Toronto group (and also the Bremen candidate NDACC Complementary observation station, which will not be described here) as follows. Investigators at the University of Toronto and Wollongong exchanged a small set of quality-controlled spectra recorded at Toronto (43.66°N, 79.40°W), Wollongong, Australia (34.45°S, 150.88°E) and Lauder, New Zealand (45.04°S, 169.68°E). The Wollongong spectra were also recorded using a 250-cm-OPD Bomem DA8 FTS previously described by *Rinsland et al.* [2001] and retrofitted with the design described in this work, while the Lauder spectra were recorded using a 257-cm-OPD Bruker 120M FTS previously described by *Rinsland et al.* [2002c]. Although the spectra were recorded by three different instruments, each team analyzed all spectra using the SFIT-2 algorithm, retrieving O_3, N_2O, and HCl from their own spectra and from the spectra of the other team. The analyses performed by Dr. Nicholas Jones at Wollongong became the controls, while Prof. David W. T. Griffith (also at Wollongong) refereed the exercise. Corresponding to spectra from each site, a set of *a priori* profiles of VMR, temperature, and pressure was agreed upon and used by both teams; the retrieval microwindows and major interfering species were also prescribed in the blind phase of the intercomparison, and the spectral database used was that documented in *Rothman et al.* [2003]. The *a priori* VMR profiles of the target gases (O_3, N_2O, HCl) were allowed to vary at all heights in a full profile retrieval, whereas the *a priori* VMR profiles of the major interfering species (e.g. CO_2, H_2O) were scaled by a single (retrieved) factor. $\mathbf{S}_\epsilon$ and $\mathbf{S}_a$ were left to the discretion of the investigators in the blind phase of the intercomparison, as were any other retrieved forward model parameters and minor interfering species. In the open phase of the intercomparison, all of the above were matched to the parameters used in the control analysis performed at Wollongong. For example, the Toronto spectra were analyzed using $\mathbf{S}_\epsilon$ of 440, 380, and 350 for O_3, HCl, and N_2O in the blind phase, and 200, 200 and 150, respectively, in the open phase. Similarly, in the blind phase, Toronto spectra were analyzed using diagonal $\mathbf{S}_a$ matrices with 20% standard deviations at all heights for all three target

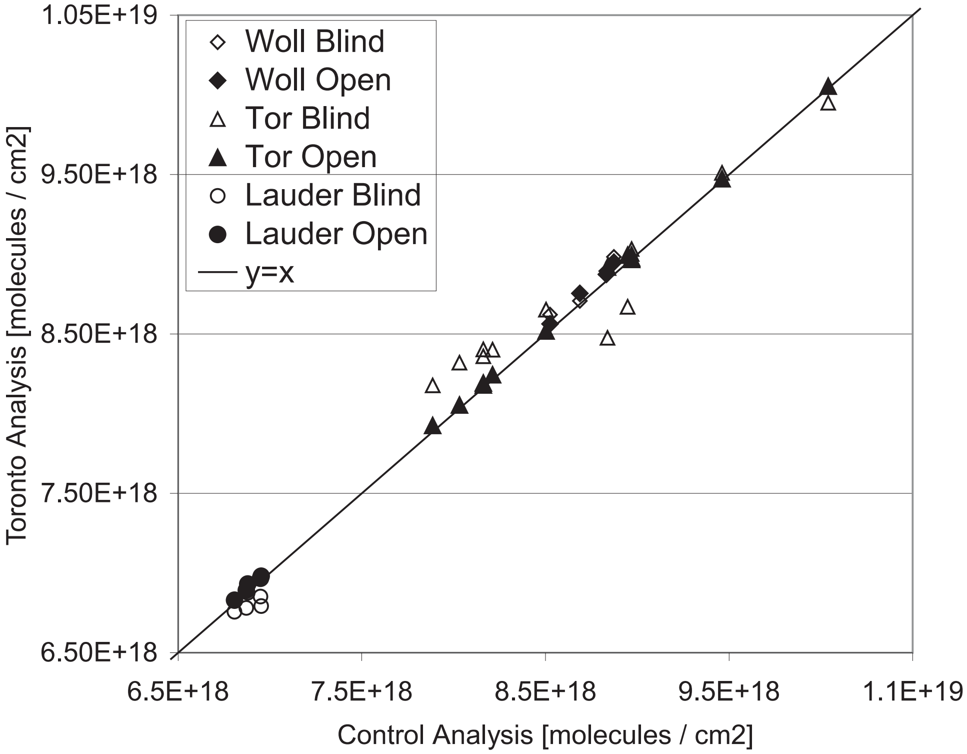


Figure 5.1 Total columns of O_3 retrieved during the NDACC validation exercise.

gases, while in the open phase a 4-km correlation length was applied for O_3 retrievals, none was used in HCl retrievals, and a 3-km correlation length was again used in N_2O retrievals. Furthermore, the diagonal elements of $\mathbf{S}_a$ were not constant as a function of height for all three target gases in the open phase, and the standard deviations ranged between 0-100%. Finally, very similar constraint differences were present in the blind and open phase retrievals of O_3, HCl, and N_2O from Wollongong and Lauder spectra.

The blind and open phase total column retrieval results for O_3, N_2O and HCl are shown in Figures 5.1-5.3, respectively, for spectra from all three locations that were analyzed. Our results show that two different users analyzing the same spectra with the same auxiliary *a priori* information (VMR, pressure and temperature profiles) but different *a priori* and measurement noise covariances, as described above, (some) different retrieved forward model parameters (e.g. transmission spectrum background slope and curvature fitting, wavenumber shift fitting), and (some) different minor interfering

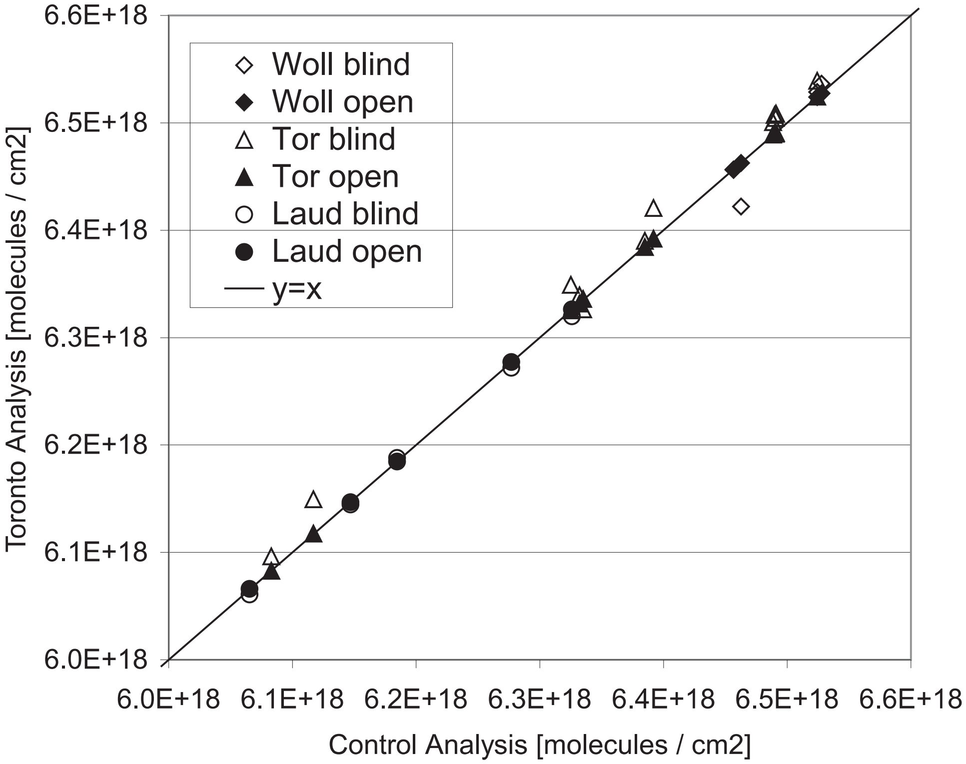


Figure 5.2 Total columns of N_2O retrieved during the NDACC validation exercise.

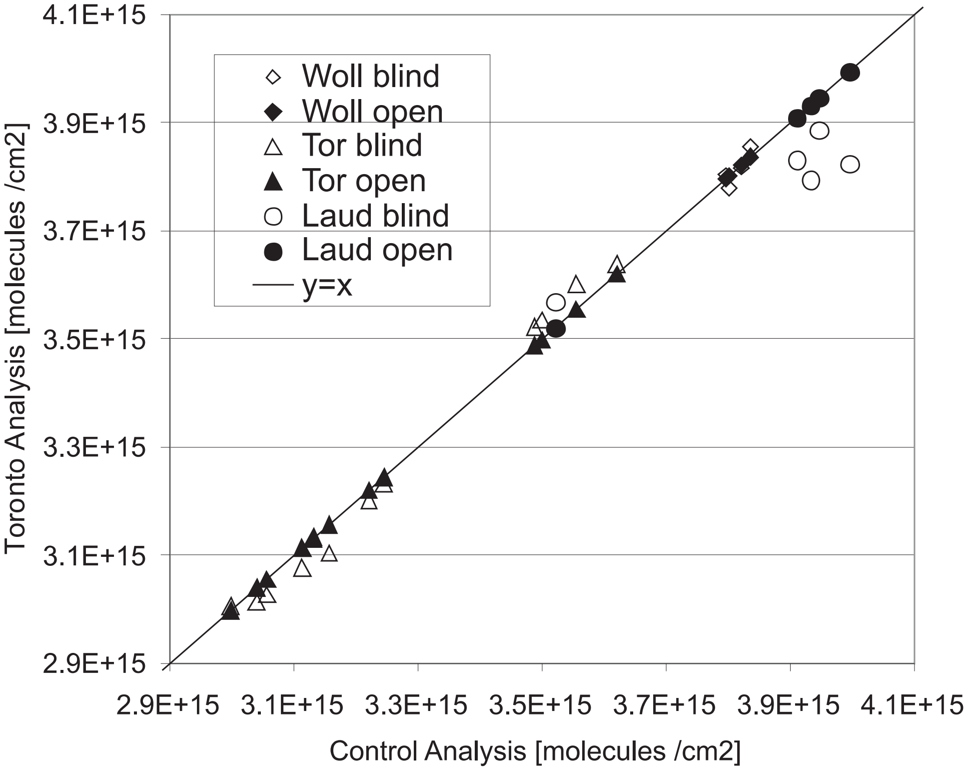


Figure 5.3 Total columns of HCl retrieved during the NDACC validation exercise.

species (e.g. OCS) can expect average differences of 2.1% for O_3, 1.6% for HCl, and 0.3% for N_2O. For comparison, *Schneider et al.* [2005a] summarize total column random errors due to SZA, temperature profile, measurement noise and smoothing as 2.6% for O_3, 2.2% for HCl, and 0.8% for N_2O. (A detailed discussion of error analysis in FTIR measurements is presented in *Schneider et al.* [2005b], and in Sections 4.3 and 5.2.) Only after controlling for all differences due to retrieval constraints, minor interfering species and retrieved forward model parameters do the average differences in the open phase of our intercomparison fall below 1% (0.83% for O_3, 0.92% for HCl, and 0.05% for N_2O). The remaining differences are ascribed to small differences in the retrieval grid (38 layers at Toronto vs. 36 layers at Wollongong), minor differences in the versions of the SFIT-2 software used, and the propagation of rounding errors to the retrieved solution. The effect of HCl retrieval parameters was studied in some detail and differences in retrieved total column amounts of up to 5% were observed over a range of retrieval parameter values such as different *a priori* covariances of profile scaling factors for interfering species (N_2O, CH_4, H_2O), the omission or inclusion of a minor interfering species (OCS), different values of the *a priori* wavenumber scale multiplier and slightly different microwindow boundaries. While no strict generalizations to other species are practical or appropriate because of the nonlinear nature of the retrieval, this number is representative of expected total column differences if the retrieval parameters that are used in the same software implementation of a retrieval algorithm are somewhat different.

Our total column differences below 1% are consistent with the findings of *Meier et al.* [2005], although our two studies are not precisely comparable since a non-linear least squares fitting approach was used by *Meier et al.* [2005] to derive the total column amounts, as opposed to the OEM approach used here to derive the full profile first and later integrate it using the airmass profile as a weighting. However, our results support the general conclusions of *Meier et al.* [2005], which are that instrumental differences (for well-aligned spectrometers) are often smaller than systematic differences due to differences in the details in the analysis approach, which are in turn smaller than day-to-day variability of the observed gases. More precisely, our study also points to the importance of the consistency in the analysis approach (including all elements of the retrieved state vector and all constraining information) for the detection of small trends in long-term data sets of variable quantities.

Finally, the two retrieval grids used in the Toronto and Control analyses contained a similar number of layers and were constructed in a similar fashion, i.e. with thin layers (∼1 km) in the troposphere and gradually thickening layers in the stratosphere and above. The results of a further investigation of six other commonly employed retrieval grids containing 29-45 layers suggest that total column differences remain smaller than 1% in retrievals performed on these different grids. This led to the more detailed study of this implicit retrieval constraint presented in Section 4.2.

5.2 NDACC Database Results (2001-2004)

We will now discuss the time series, information content, and errors of the seven gases submitted to the NDACC database in June, 2005. The ensemble of 2004 results is used to illustrate the variability of retrieval diagnostics over a representative range of SNR, SZA, and temperature and pressure profiles, caused by the weakly nonlinear nature of the forward model. Since calculations were performed using $\mathbf{K}_a$ rather than $\mathbf{K}_n$ (Section 3.4.3.1), the true variability of retrieval diagnostics is likely to be slightly larger.

5.2.1 O_3

The three-year time series of ozone total (0-100 km), stratospheric (15-50 km) and tropospheric (0-15 km) columns is shown in Figure 5.4. The error bars on this plot and other time series in Section 5.2 correspond to values submitted to the NDACC database, and represent the % partial column measurement and smoothing errors added in quadrature. These individual errors for a given partial column are in turn calculated by multiplying the respective full error covariance matrices by the corresponding density-weighted partial column operator. Temperature errors are not included in the time series plots because they were calculated in November, 2005; however, average values are given in tabular form, together with measurement and smoothing errors. O_3 total columns are highly correlated with stratospheric columns, which is consistent with this being the primary region of sensitivity of the retrieval (Figures 4.33 and 4.34). The tropospheric columns are included for reference and show small variations throughout the year. The primary signature in this figure is the mid-latitude seasonal cycle of O_3, which exhibits a spring-time maximum. Because the data has been quality controlled at this point (see below), the apparent outliers are likely to be signatures of dynamic variability above Toronto. In particular, the highest total column measurement of 2004 (March 22^{nd}) corresponds to an intrusion of polar air above Toronto, which is discussed in more detail in Section 6.2.

Figure 5.5 shows the RMS spectral residual of all O_3 retrievals performed between 2001 and 2004 using the two-step retrieval approach discussed in Section 4.1.3. All retrievals were first performed using an SNR of 300, and then repeated a second time using the custom SNR derived from the spectral residual of each fit in the first step. In the case of O_3, the average SNR of 300 corresponds to an average RMS spectral residual of 0.333%, which was multiplied by 1.2 in order to to form an objective – if somewhat arbitrary – rejection criterion of 0.4% for the entire dataset. One could in principle apply a more complicated statistical analysis, e.g. a trimmed mean and standard

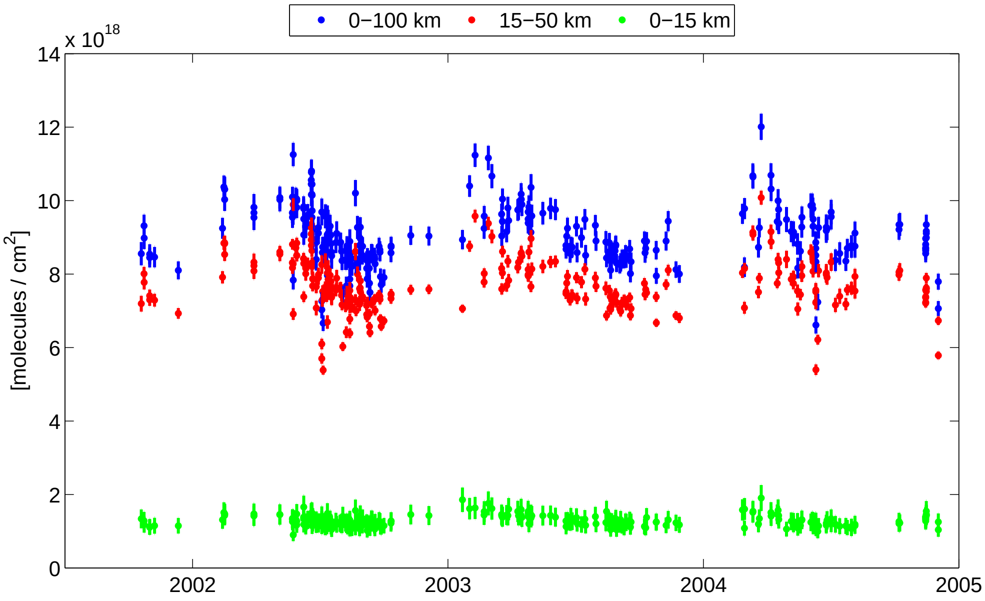


Figure 5.4 2001-2004 O_3 total and partial column time series recorded at TAO.

deviation, however, the simple procedure above yielded results that appear satisfactory for operational retrievals in terms of rejection rates and the general quality of retained spectral fits. In Figure 5.5 we can see that the RMS residual of spectra recorded in 2001 was highly variable, leading to a rejection rate of 84%. This is partly due to the suboptimal alignment of the FTS during its commissioning phase, and also due to the use of various "unorthodox" settings for the recording of the first TAO spectra. After a re-alignment and other FTS improvements, the rejection rate dropped to 11% in 2002 and 15% in 2003. The 2004 increase to 39% is likely attributable to the partial automation of the recording of TAO spectra in August, 2003, and hence reduced operator scrutiny of the acquisition system.

Figure 5.6 shows the general increase of d_s for O_3 as a function of increasing SNR and SZA, although the quantification of these relationships is confounded by a rather strong relationship between between SNR and SZA in TAO spectra. With some exceptions, the SNR generally decreases for increasing SZA due to the attenuation of background

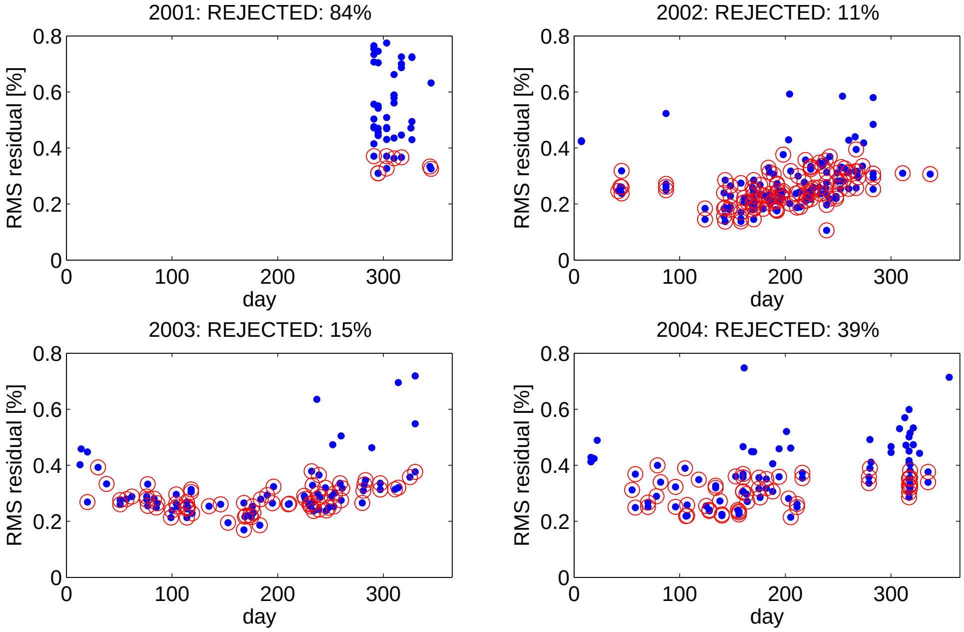


Figure 5.5 Spectral residuals of 2001-2004 O_3 fits at 3045 cm^{-1}. Values below the rejection criterion (0.4%) are circled in red.

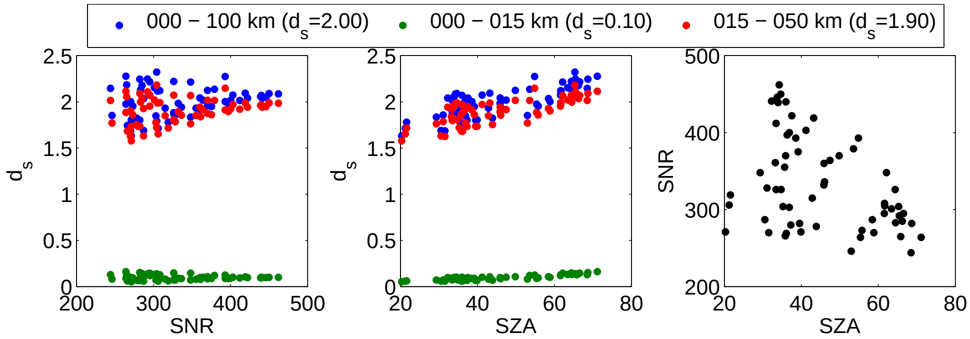


Figure 5.6 2004 O_3 total and partial column d_s as a function of SNR and SZA (legend shows mean values). Relationship between SNR and SZA at TAO is shown for reference.

	[0-100 km]		[0-15 km]		[15-50 km]	
d_s (σ)	2.00	(0.17)	0.10	(0.03)	1.90	(0.14)
%a (σ)	13.09	(0.64)	83.54	(3.00)	0.29	(0.51)
S_s (σ)	2.58	(0.14)	18.35	(0.45)	1.29	(0.18)
S_m (σ)	1.83	(0.25)	4.21	(0.44)	1.79	(0.32)
S_T (σ)	2.33	(0.09)	0.79	(0.12)	2.73	(0.09)
S'_{tot} (σ)	3.94	(0.22)	18.86	(0.34)	3.52	(0.22)

Table 5.1 2004 average O_3 total and partial column retrieval information content and percent errors, as defined in Table 4.1 (σ = one standard deviation of the mean.)

transmission levels having a larger impact than the increase in the depth of spectral absorption features caused by an increased path length at higher SZAs.

Finally, Figure 5.7 shows the 2004 average averaging kernel matrix and averaging kernels, which are very similar to the particular matrix and kernels singled out in Figure 4.34. Also shown are all retrieved profiles in 2004 as well as the average partial column kernels, and vertical profiles of average measurement, temperature and smoothing errors, and finally, the smoothing error (repeated for clarity) and the total error (the square roots of the diagonal elements of the error covariance matrices described in Section 3.3). The error bars represent the standard deviations of the annual averages. The kernels and errors are again qualitatively similar to those displayed for a single spectrum in Figure 4.34. Table 5.1 summarizes the average d_s, % contribution of the *a priori*, and errors for total, stratospheric and partial columns, with the standard deviations of the averages also shown in brackets. The smoothing error dominates for O_3, as is the case for other stratospheric gases. In general, the characterization and error analysis of the retrieval of O_3 appears to have a weak dependence on the state (relatively low variability of errors and kernels throughout the year), indicating that the inverse problem is only weakly nonlinear.

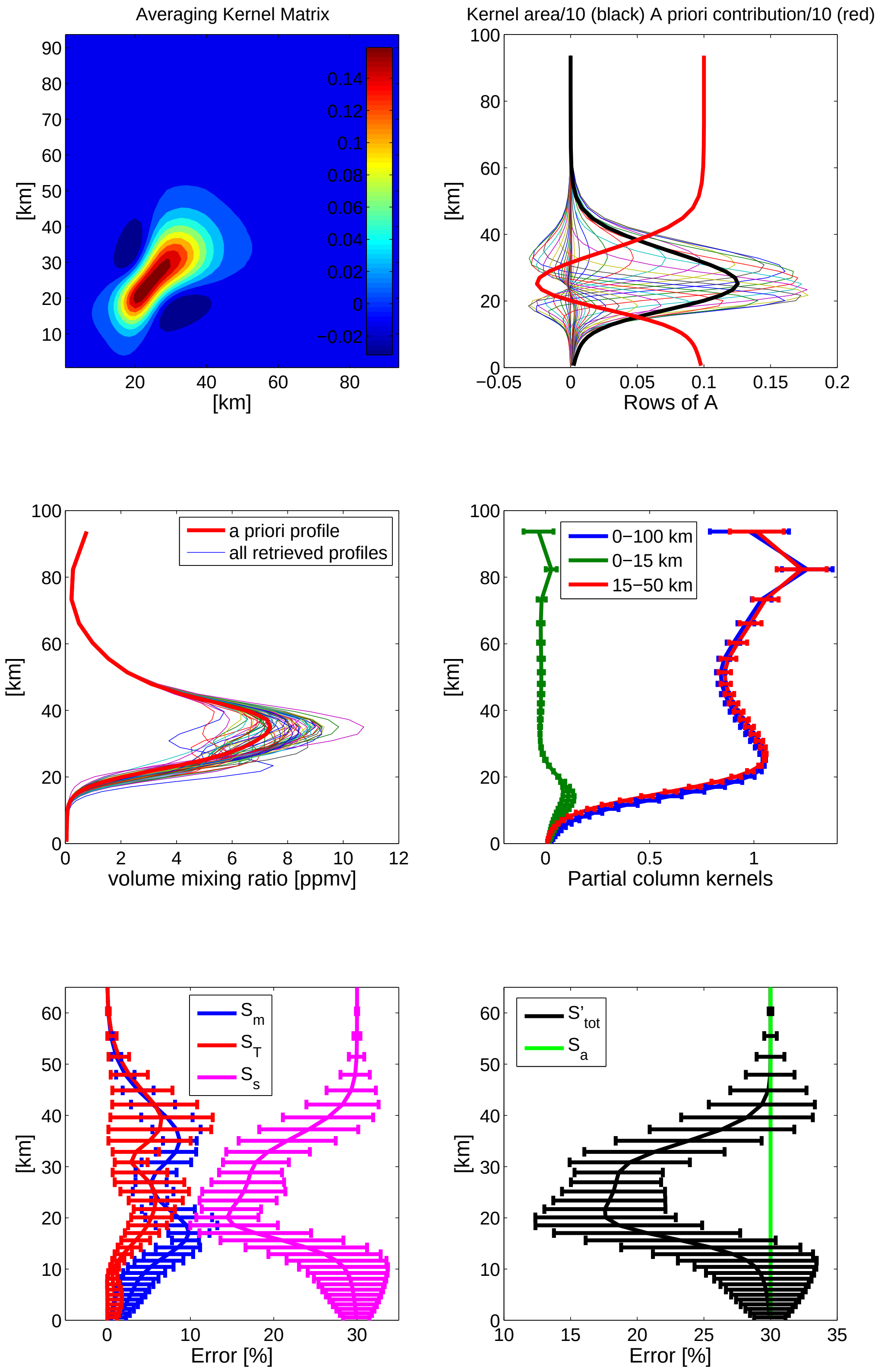


Figure 5.7 2004 average retrieval characterization for O_3 (error bars show 1σ).

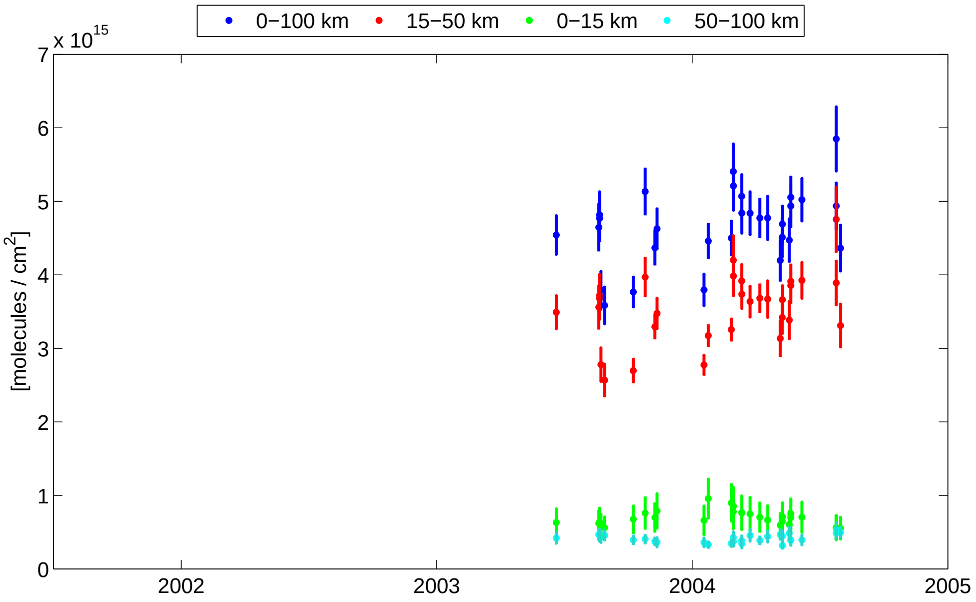


Figure 5.8 2001-2004 NO total and partial column time series recorded at TAO.

5.2.2 NO

The total, stratospheric and tropospheric columns of NO spanning approximately one year are shown in Figure 5.8, with error bars calculated the same as for O_3. This dataset has reduced temporal coverage because Filter 5 measurements were not made priori to the middle of 2003. The total columns are again highly correlated with stratospheric columns, and there is some response in the tropospheric columns, which is likely spurious given the shape of **A** and **K** in Figures 4.35 and 4.36. The mesospheric columns are shown for reference, since they are discussed in more detail in Section 6.1.

The d_s in Figure 5.9 shows a similar relationship to SNR and SZA as in the case of O_3, however, there is no longer a clear relationship between SNR and SZA. This is likely due to the fact that the SNR is most strongly affected by the highly variable concentrations of H_2O in this spectral region, also leading to higher spectrum rejection rates (Figure 5.10) than in the case of O_3. Rejection rates may be higher in the summer of 2003 because when automation was first implemented, NO measurements were primarily recorded near

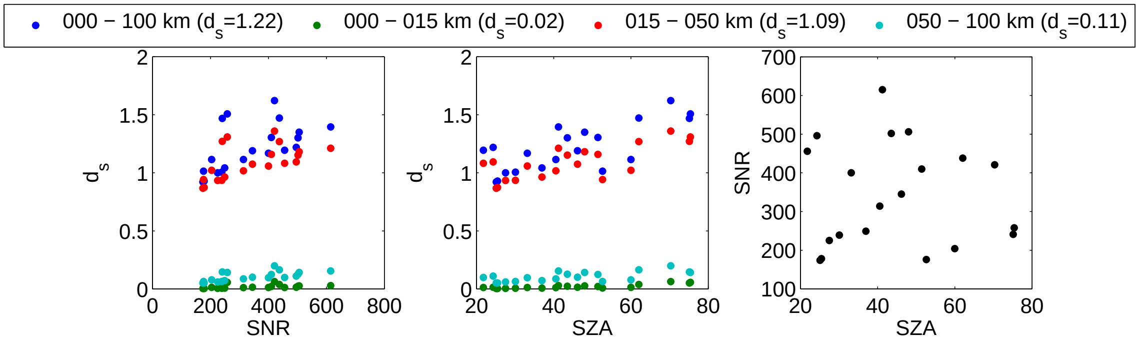


Figure 5.9 2004 NO total and partial column d_s as a function of SNR and SZA (legend shows mean values). Relationship between SNR and SZA at TAO is shown for reference.

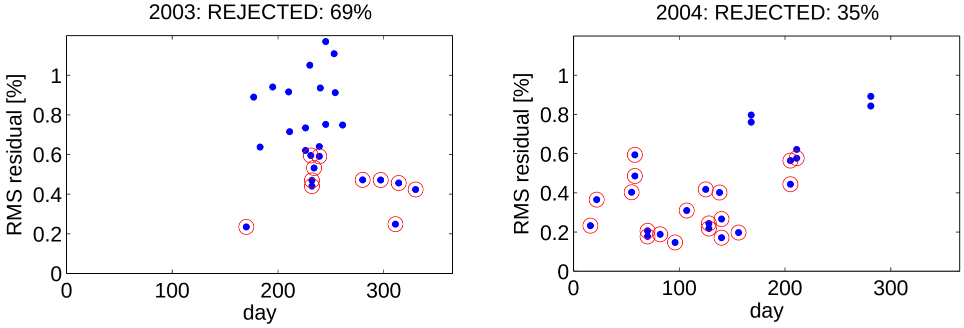


Figure 5.10 Spectral residuals of 2003-2004 NO fits at 1900 cm^{-1}. Values below the rejection criterion (0.6%) are circled in red.

noon, when the suntracker is known to be prone to drifting.

Finally, Figure 5.11 shows the 2004 average averaging kernel matrix, which is very similar to the particular matrix singled out in Figure 4.36. Also shown are all retrieved profiles in 2004 as well as the average partial column kernels, and vertical profiles of average measurement, temperature, smoothing and total error standard deviations. (As in the case of O_3, the error bars represent the standard deviations of the annual averages.) They are again qualitatively similar to the kernels and errors displayed for a single spectrum in Figure 4.36. Table 5.2 summarizes the average d_s, % contribution of the *a priori*, and errors for total, stratospheric and partial columns, with the standard deviations of the averages also shown in brackets.

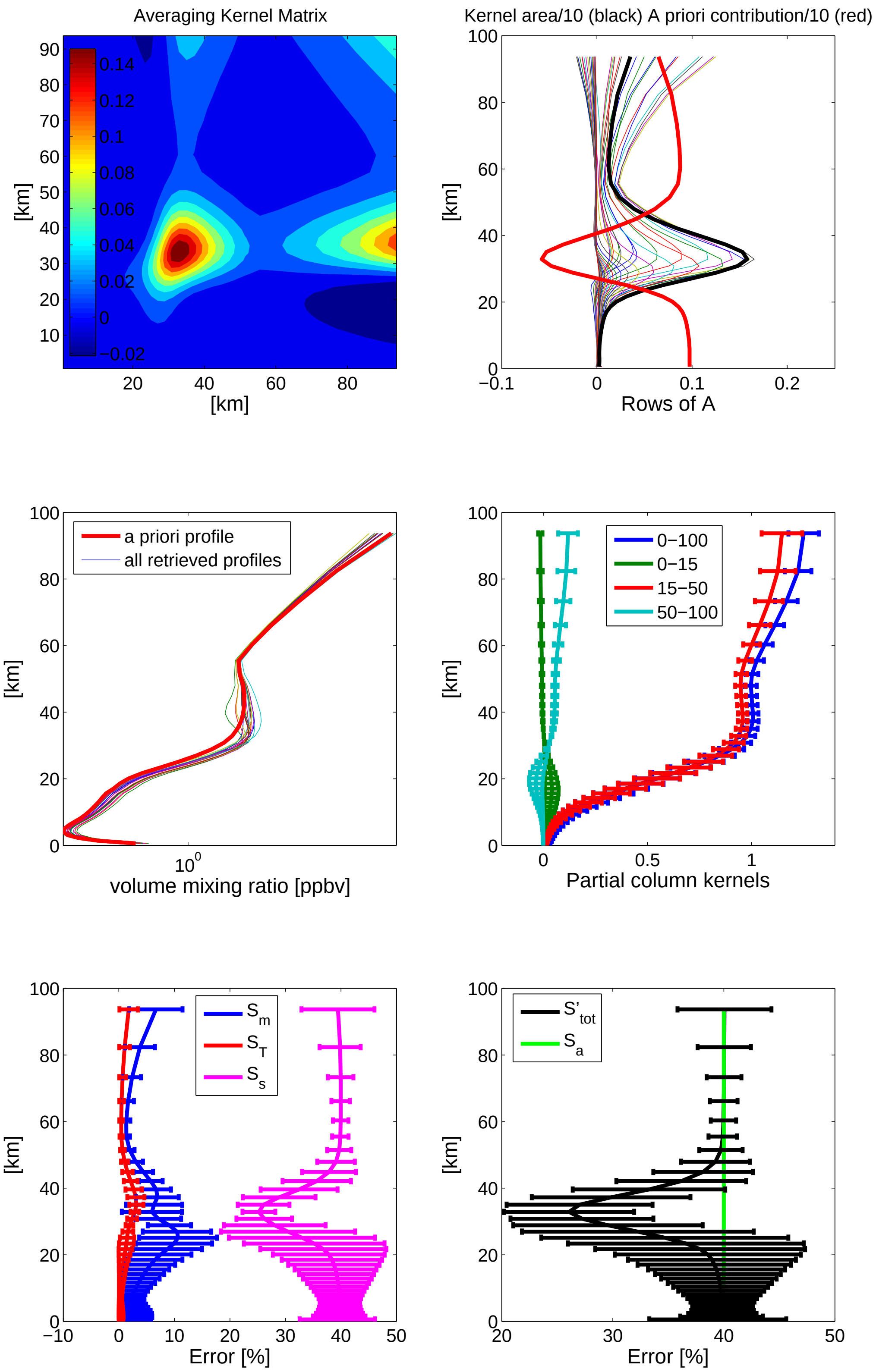


Figure 5.11 2004 average retrieval characterization for NO (error bars show 1σ).

	[0-100 km]		[0-15 km]		[15-50 km]		[50-100 km]	
d_s (σ)	1.22	(0.21)	0.02	(0.02)	1.09	(0.15)	0.11	(0.04)
%a (σ)	19.63	(2.68)	97.00	(1.64)	-12.47	(4.35)	76.95	(2.29)
S_s (σ)	5.36	(0.17)	29.44	(0.25)	4.95	(0.60)	16.68	(0.54)
S_m (σ)	2.87	(1.12)	2.59	(1.11)	4.22	(1.42)	3.50	(0.88)
S_T (σ)	1.45	(0.03)	0.32	(0.19)	1.96	(0.07)	0.94	(0.40)
S'_{tot} (σ)	6.32	(0.63)	29.57	(0.24)	6.84	(1.34)	17.10	(0.33)

Table 5.2 2004 average NO total and partial column retrieval information content and percent errors, as defined in Table 4.1 (σ = one standard deviation of the mean.)

5.2.3 NO_2

Figure 5.12 shows a clear seasonal cycle in the total and stratospheric column of NO_2, with a summer-time maximum. Tropospheric columns are high because of high *a priori* values but they remain virtually unchanged throughout the year, which is consistent with the shape of **K** and **A**. The RMS residuals shown in Figure 5.13 show much higher rejection rates in 2004, which may be due in part to less scrutinized automation, however, this may also be symptomatic of some drift in spectrometer alignment, which may be affecting this very weak absorption feature significantly.

Figure 5.14 shows d_s as a function of SNR and SZA, and reveals the general lack of a relationship between SNR and SZA. Figure 5.15 shows the 2004 average retrieval characterization and errors, as for the other gases, which is again similar to the particular matrix singled out in Figure 4.40. Table 5.3 summarizes the average d_s, % contribution of the *a priori*, and errors for total, stratospheric and partial columns, with the standard deviations of the averages again shown in brackets. The information content and % *a priori* contribution are consistent with the seasonal cycle response of the stratospheric and tropospheric columns shown in Figure 5.12.

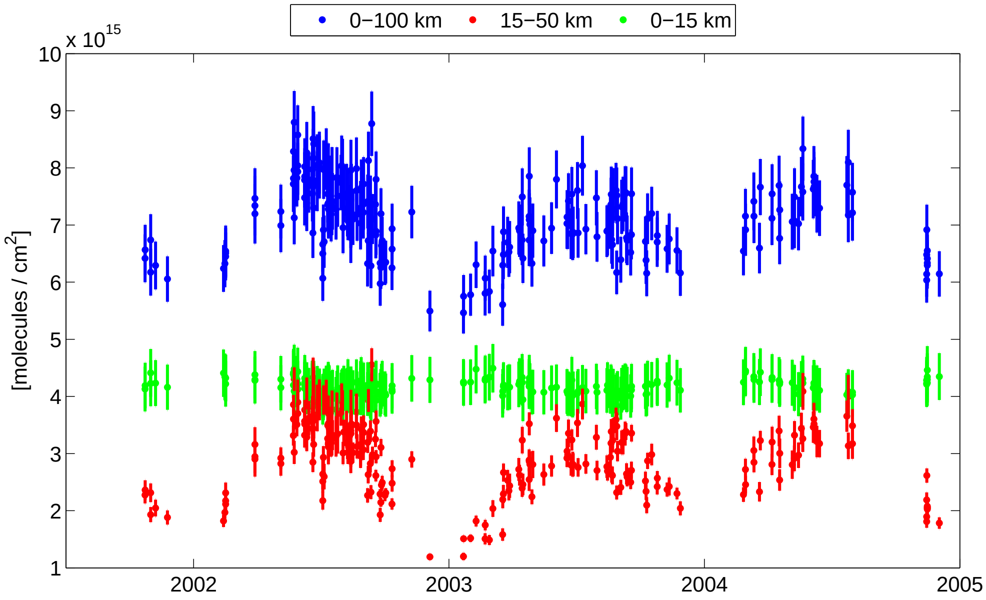


Figure 5.12 2001-2004 NO_2 total and partial column time series recorded at TAO.

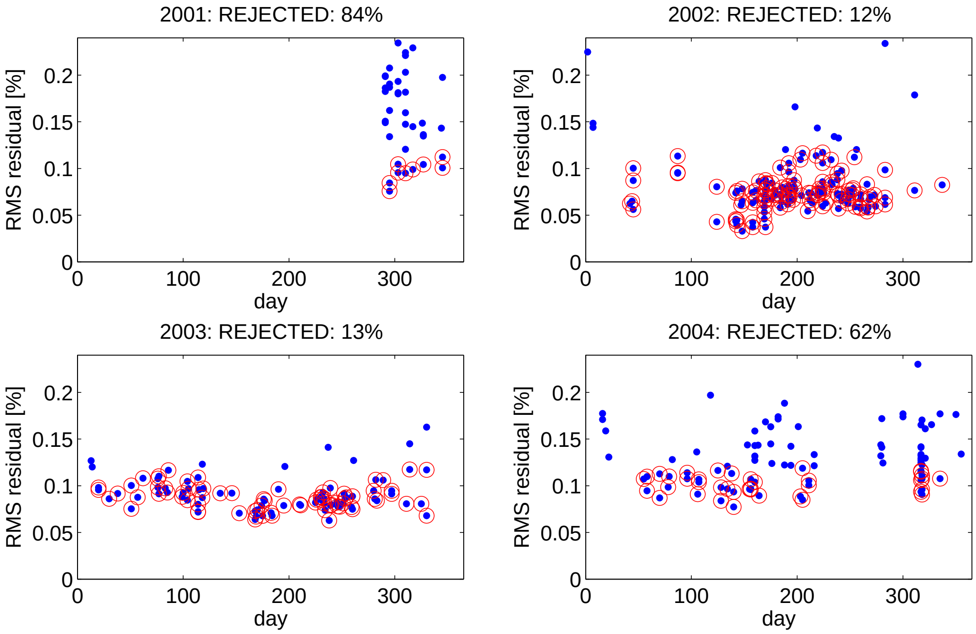


Figure 5.13 Spectral residuals of 2001-2004 NO_2 fits at 2915 cm^{-1}. Values below the rejection criterion (0.12%) are circled in red.

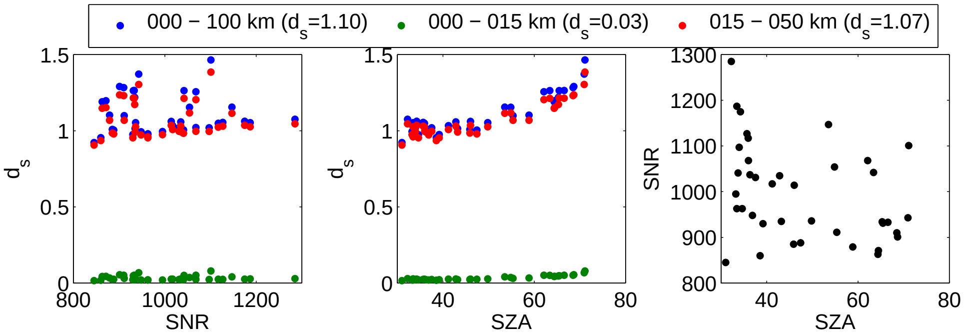


Figure 5.14 2004 NO_2 total and partial column d_s as a function of SNR and SZA (legend shows mean values). Relationship between SNR and SZA at TAO is shown for reference.

	[0-100 km]		[0-15 km]		[15-50 km]	
d_s (σ)	1.10	(0.13)	0.03	(0.01)	1.07	(0.11)
%a (σ)	66.28	(1.15)	97.61	(0.30)	8.26	(3.77)
S_s (σ)	6.24	(0.07)	9.46	(0.05)	4.00	(0.53)
S_m (σ)	2.28	(0.40)	0.73	(0.13)	5.88	(1.18)
S_T (σ)	0.98	(0.03)	0.07	(0.01)	2.68	(0.12)
S'_{tot} (σ)	6.73	(0.16)	9.49	(0.04)	7.62	(1.13)

Table 5.3 2004 average NO_2 total and partial column retrieval information content and percent errors, as defined in Table 4.1 (σ = one standard deviation of the mean.)

5.2.4 HCl

Figure 5.16 clearly shows a seasonal cycle in the total and stratospheric column of HCl, with a spring-time maximum. The highest total column recorded in 2004 corresponds to an intrusion of polar air above Toronto, which is discussed in more detail in Section 6.2. The remaining outliers are also likely indicators of dynamic variability over Toronto and should be investigated on a case by case basis. The tropospheric columns show some response throughout the year, which is possible given the shape of **K** and **A** for this retrieval and the limited information originating in this region (d_s=0.24). The RMS residuals shown in Figure 5.17 again show much higher rejection rates in 2004, which may be due in part to less scrutinized automation or some drift in spectrometer alignment.

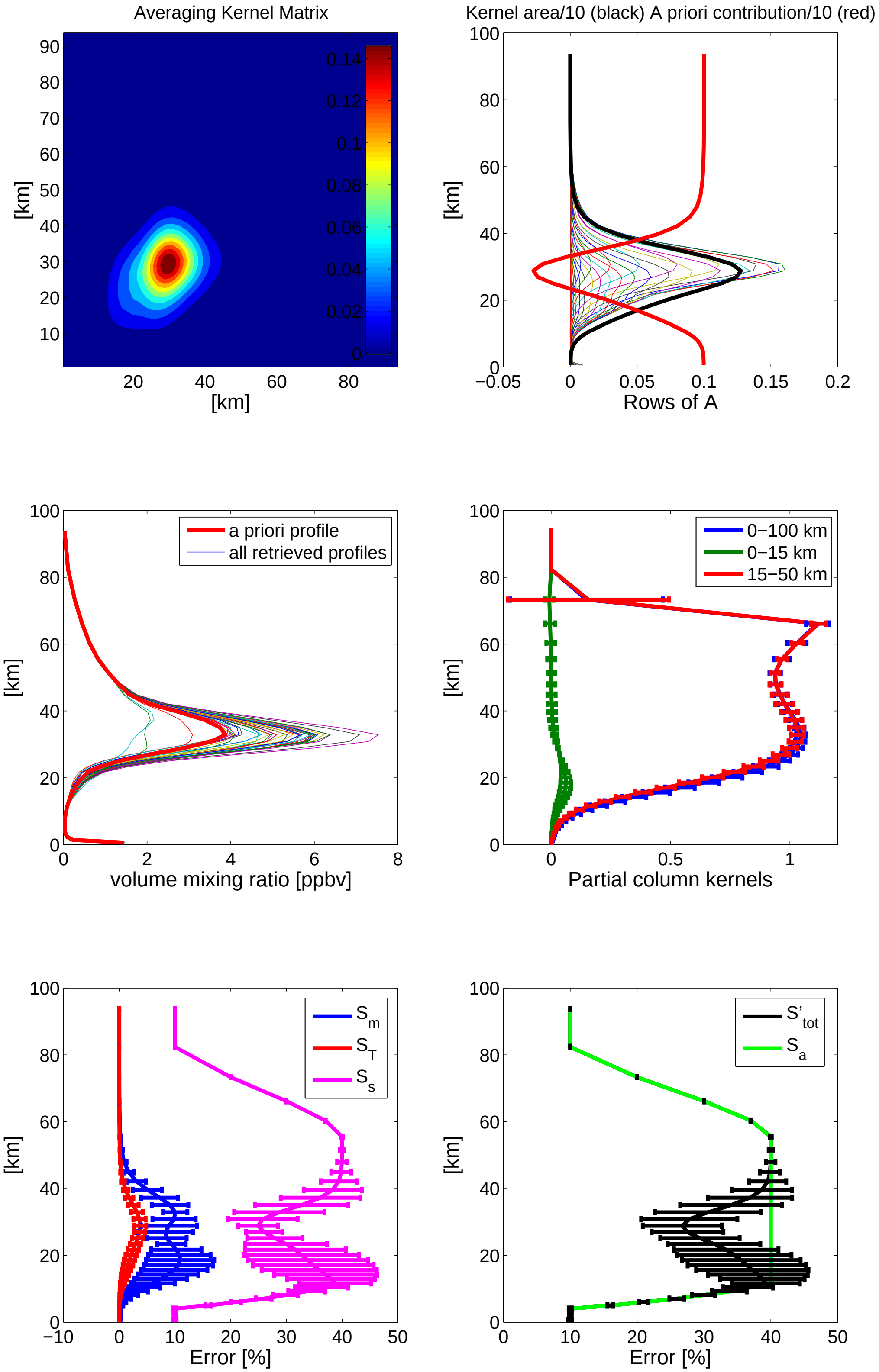


Figure 5.15 2004 average retrieval characterization for NO_2 (error bars show 1σ).

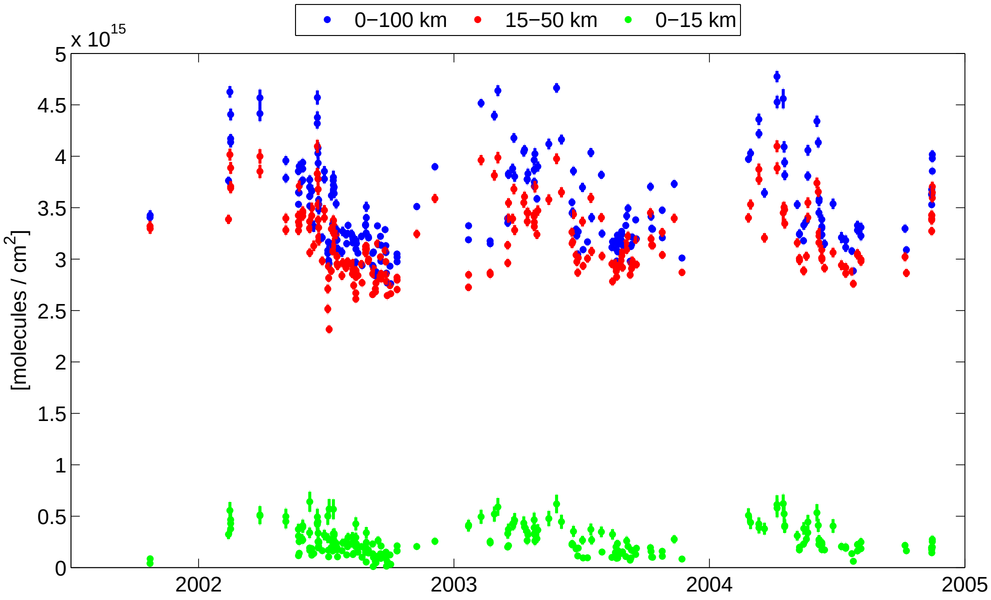


Figure 5.16 2001-2004 HCl total and partial column time series recorded at TAO.

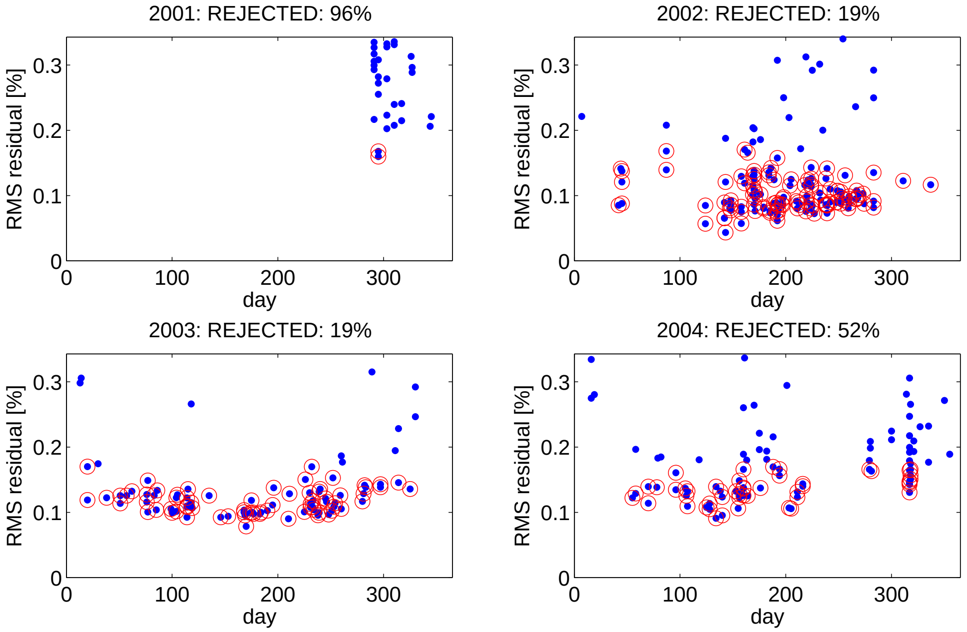


Figure 5.17 Spectral residuals of 2001-2004 HCl fits at 2925 cm^{-1}. Values below the rejection criterion (0.17%) are circled in red.

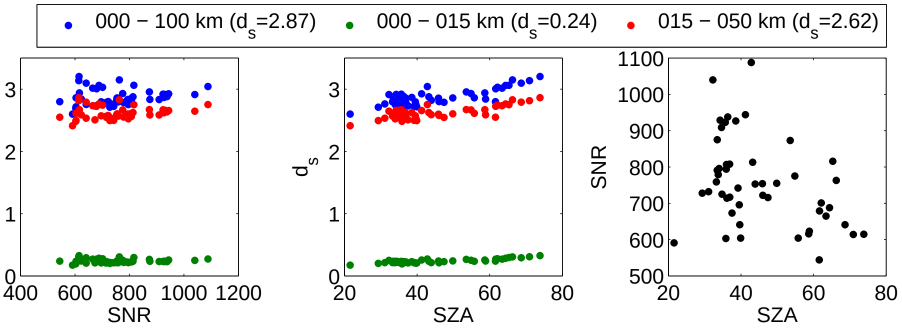


Figure 5.18 2004 HCl total and partial column d_s as a function of SNR and SZA (legend shows mean values). Relationship between SNR and SZA at TAO is shown for reference.

	[0-100 km]		[0-15 km]		[15-50 km]	
d_s (σ)	2.87	(0.13)	0.24	(0.03)	2.62	(0.10)
%a (σ)	5.09	(0.31)	52.64	(2.72)	-2.64	(0.07)
S_s (σ)	0.77	(0.09)	14.02	(0.91)	1.22	(0.05)
S_m (σ)	0.96	(0.16)	6.37	(0.24)	0.72	(0.10)
S_T (σ)	2.42	(0.04)	3.17	(0.23)	2.45	(0.04)
S'_{tot} (σ)	2.72	(0.06)	15.73	(0.86)	2.83	(0.04)

Table 5.4 2004 average HCl total and partial column retrieval information content and percent errors, as defined in Table 4.1 (σ = one standard deviation of the mean.)

Figure 5.18 shows d_s as a function of SNR and SZA, as well as the weak relationship between SNR and SZA. Figure 5.19 shows the 2004 average retrieval characterization, as for the other gases, which is again similar to the particular matrix singled out in Figure 4.42. The average errors shown in Figure 5.19 are also similar to those in Figure 4.42, and Table 5.4 again summarizes the average d_s, % contribution of the *a priori*, and errors for total, stratospheric and partial columns, with the standard deviations of the averages again shown in brackets.

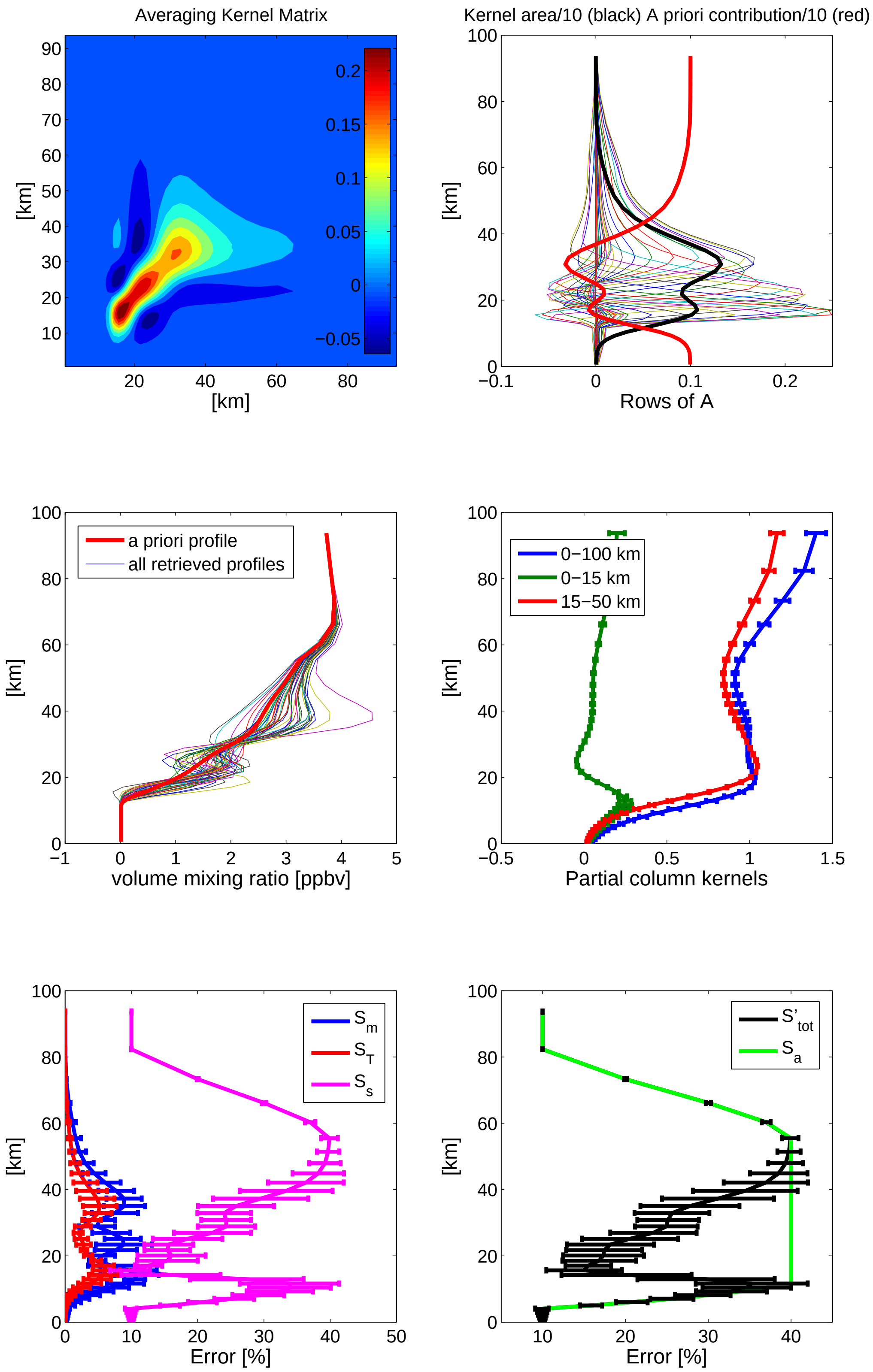


Figure 5.19 2004 average retrieval characterization for HCl (error bars show 1σ).

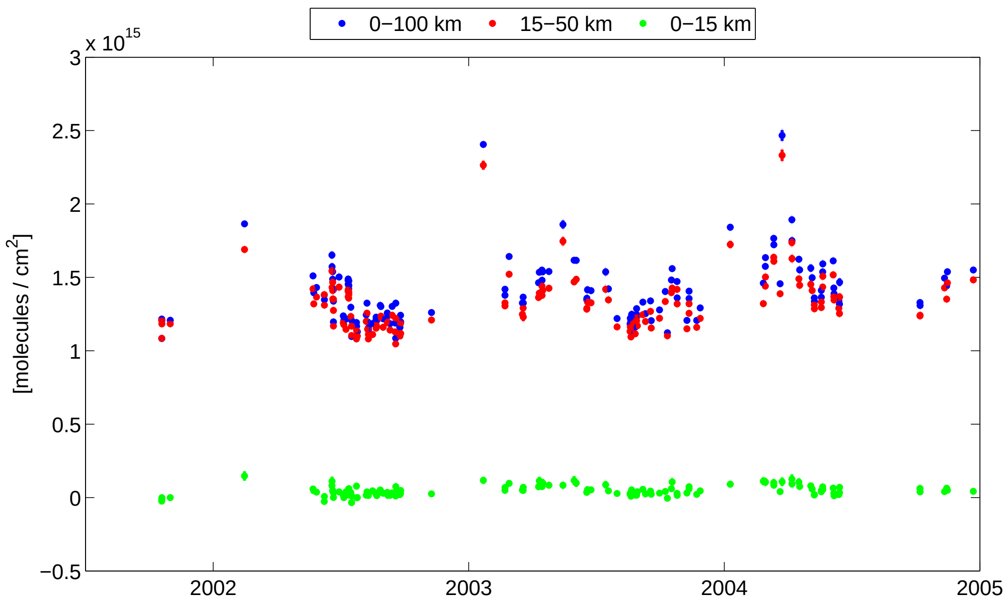


Figure 5.20 2001-2004 HF total and partial column time series recorded at TAO.

5.2.5 HF

Figure 5.20 gives some indication of the spring-time maximum seasonal cycle in the total and stratospheric columns of HF. The highest total column recorded in 2004 corresponds to an intrusion of polar air above Toronto, which is discussed in more detail in Section 6.2. The remaining outliers are also likely indicators of dynamic variability over Toronto and should be investigated on a case by case basis. Unlike HCl, the tropospheric columns of HF show virtually no response throughout the year, which is consistent with the amount of retrieval information originating in this region (d_s=0.03). Figure 5.21 shows overall higher rejection rates than, e.g. O_3, again with the 2004 values being the highest. The fact that the HF absorption feature lies on the wing of a strong H_2O feature is responsible for the variability in the SNR and the high rejection rates, as in the case of NO. Furthermore, since the retrieval of HF has been documented to suffer from systematic problems rooted in ILS distortions by *Griffith et al.* [2003], a detailed study of its behaviour and error budgets appears necessary, likely resulting in changes to the rejection criterion.

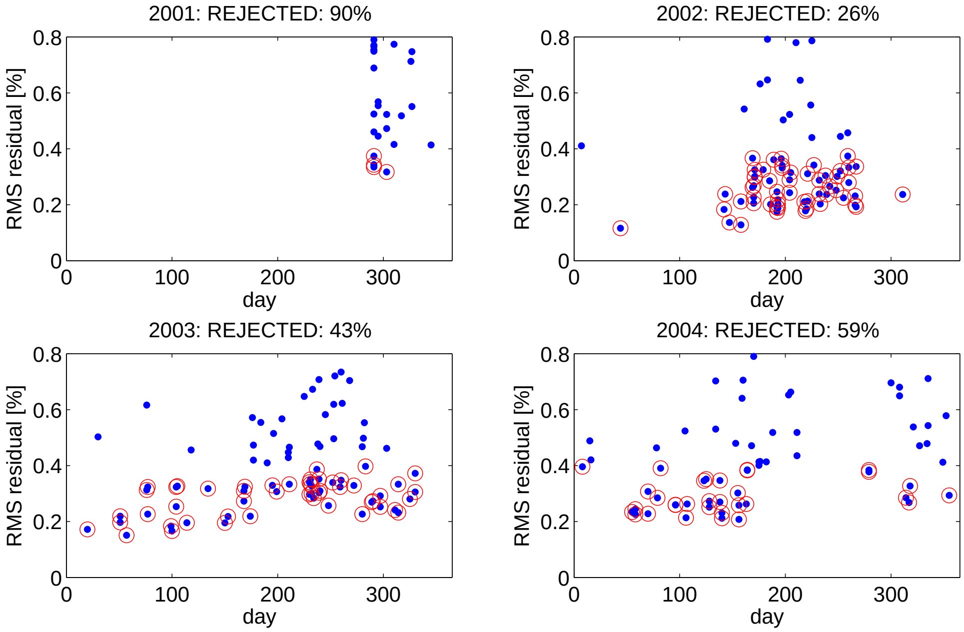


Figure 5.21 Spectral residuals of 2001-2004 HF fits at 4038 cm^{-1}. Values below the rejection criterion (0.4%) are circled in red.

Figure 5.22 shows d_s as a function of SNR and SZA, as well as the general lack of a relationship between SNR and SZA. The average errors shown in Figure 5.23 are also similar to those in Figure 4.44, and Table 5.5 again summarizes the average d_s, % contribution of the *a priori*, and errors for total, stratospheric and partial columns, with the standard deviations of the averages again shown in brackets.

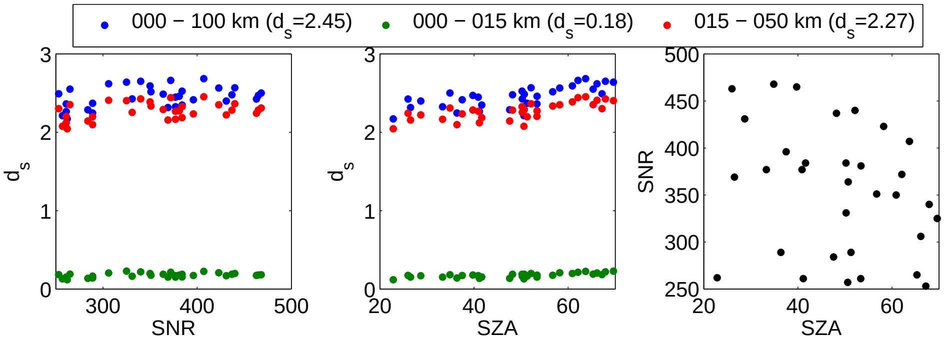
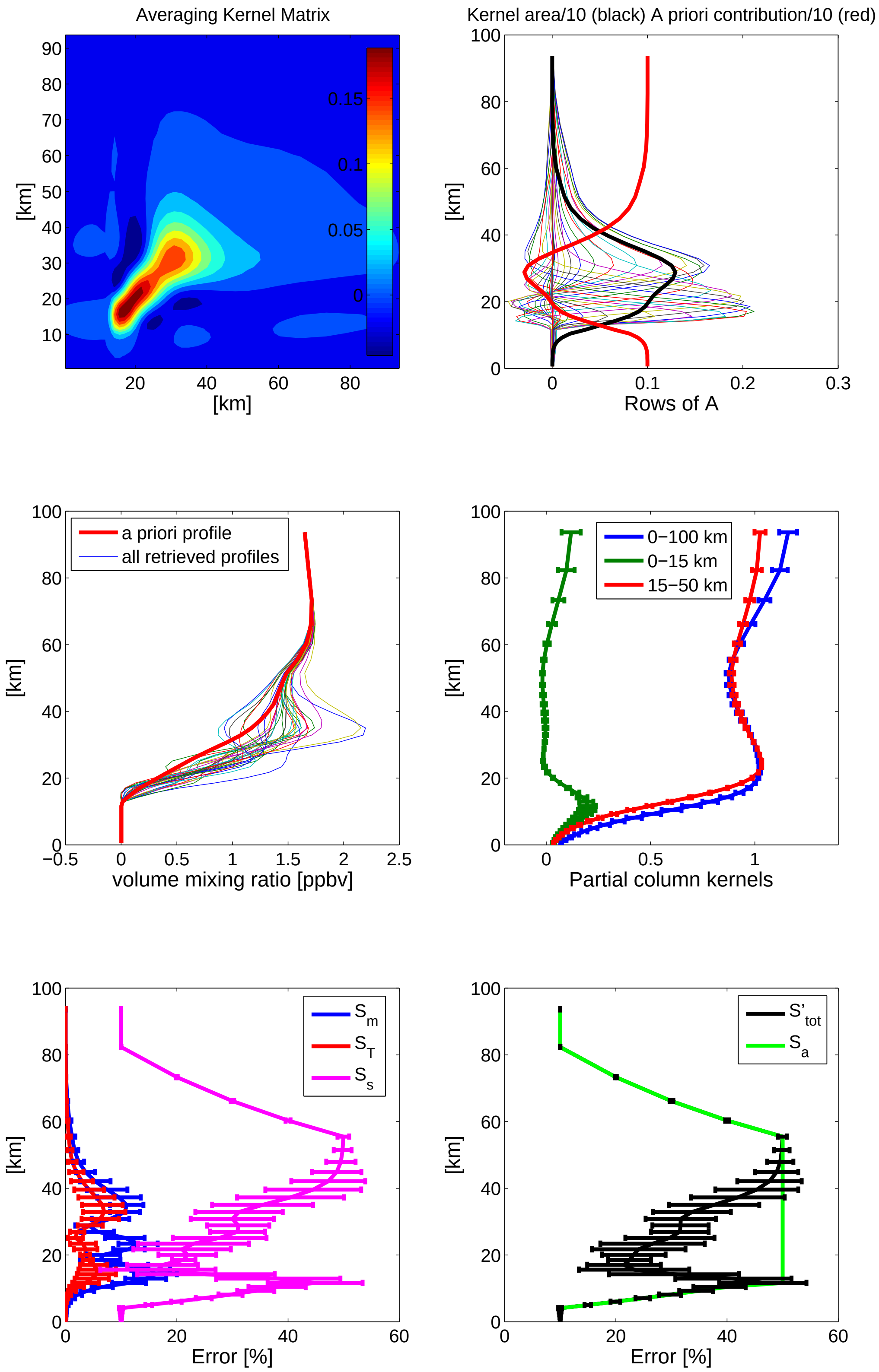


Figure 5.22 2004 HF total and partial column d_s as a function of SNR and SZA (legend shows mean values). Relationship between SNR and SZA at TAO is shown for reference.

	[0-100 km]		[0-15 km]		[15-50 km]	
d_s (σ)	2.45	(0.14)	0.18	(0.03)	2.27	(0.11)
%a (σ)	2.80	(0.27)	53.81	(5.26)	-2.73	(0.14)
S_s (σ)	0.75	(0.14)	23.22	(1.78)	1.30	(0.03)
S_m (σ)	1.00	(0.17)	11.72	(0.40)	0.77	(0.20)
S_T (σ)	2.22	(0.06)	3.88	(0.79)	2.19	(0.03)
S'_{tot} (σ)	2.56	(0.07)	26.33	(1.47)	2.67	(0.07)

Table 5.5 2004 average HF total and partial column retrieval information content and percent errors, as defined in Table 4.1 (σ = one standard deviation of the mean.)


Averaging Kernel Matrix

Kernel area/10 (black) A priori contribution/10 (red)

Figure 5.23 2004 average retrieval characterization for HF (error bars show 1σ).

5.2.6 CH_4

The three-year time series of CH_4 total and partial columns in shown in Figure 5.24, with the error bars smaller than the markers in most cases. A striking feature of this time series are the notable high outliers observed in the Summers of 2002, 2003 and 2004. The spectral fits of June 8^{th} and July 19^{th}, 2004, were re-examined and cannot be rejected based on the objective criterion of RMS residual. The retrieved profiles show large oscillations, however, this is – to an extent – a normal feature of ground-based FTIR retrievals. Moreover, the retrievals of N_2O from the same spectra (Filter 3) also show slightly elevated values (Figure 5.28), but ordinary spectral fits. It is not clear at this point whether these elevated values are the result of a subtle problem with the recorded spectra or a real (but transient) feature of the urban atmosphere on those days.

Figure 5.25 shows the RMS residuals of the spectral fits, with higher rejection rates in 2004, as for the other gases. Figure 5.26 shows the 2004 average retrieval characterization, as for the other gases, which is again similar to the particular matrix singled out in Figure 4.48. The average errors shown in Figure 5.26 are also similar to those in Figure 4.48, and Table 5.6 summarizes the average d_s, % contribution of the *a priori*, and errors for total, stratospheric and partial columns, with the standard deviations of the averages again shown in brackets. Finally, Figure 5.27 shows d_s as a function of SNR and SZA, as well as the relationship between SNR and SZA. A clear separation between d_s values is observed, with ∼1 piece of information from the stratosphere and ∼2 from the troposphere.

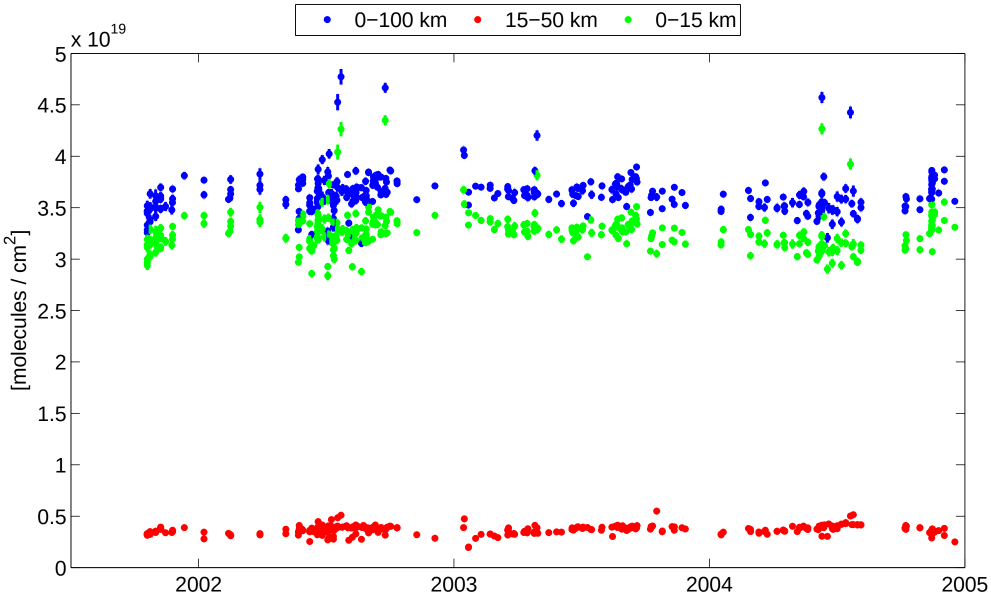


Figure 5.24 2001-2004 CH_4 total and partial column time series recorded at TAO.

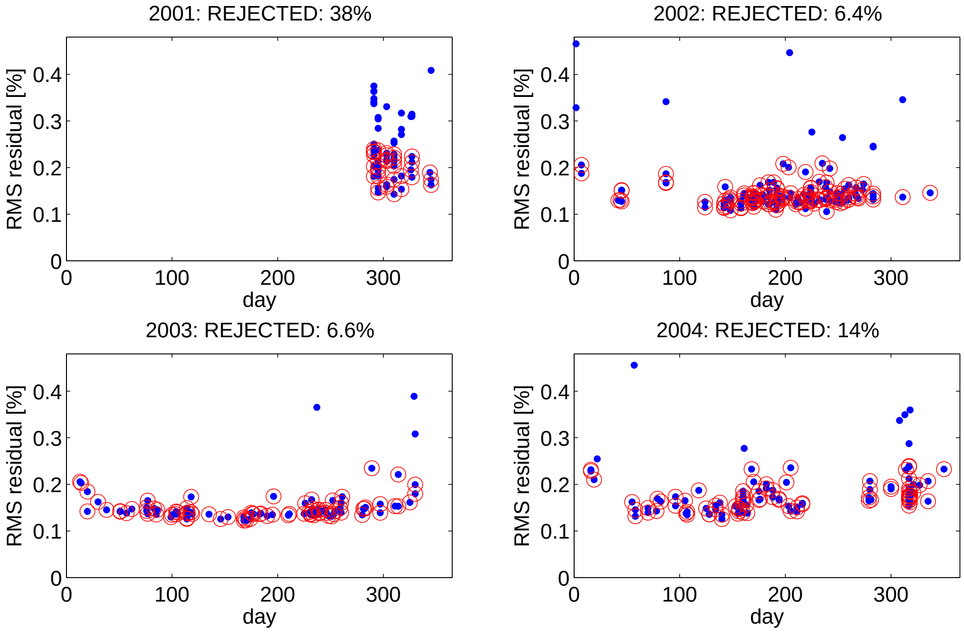


Figure 5.25 Spectral residuals of 2001-2004 CH_4 fits at 2651 cm^{-1}. Values below the rejection criterion (0.24%) are circled in red.

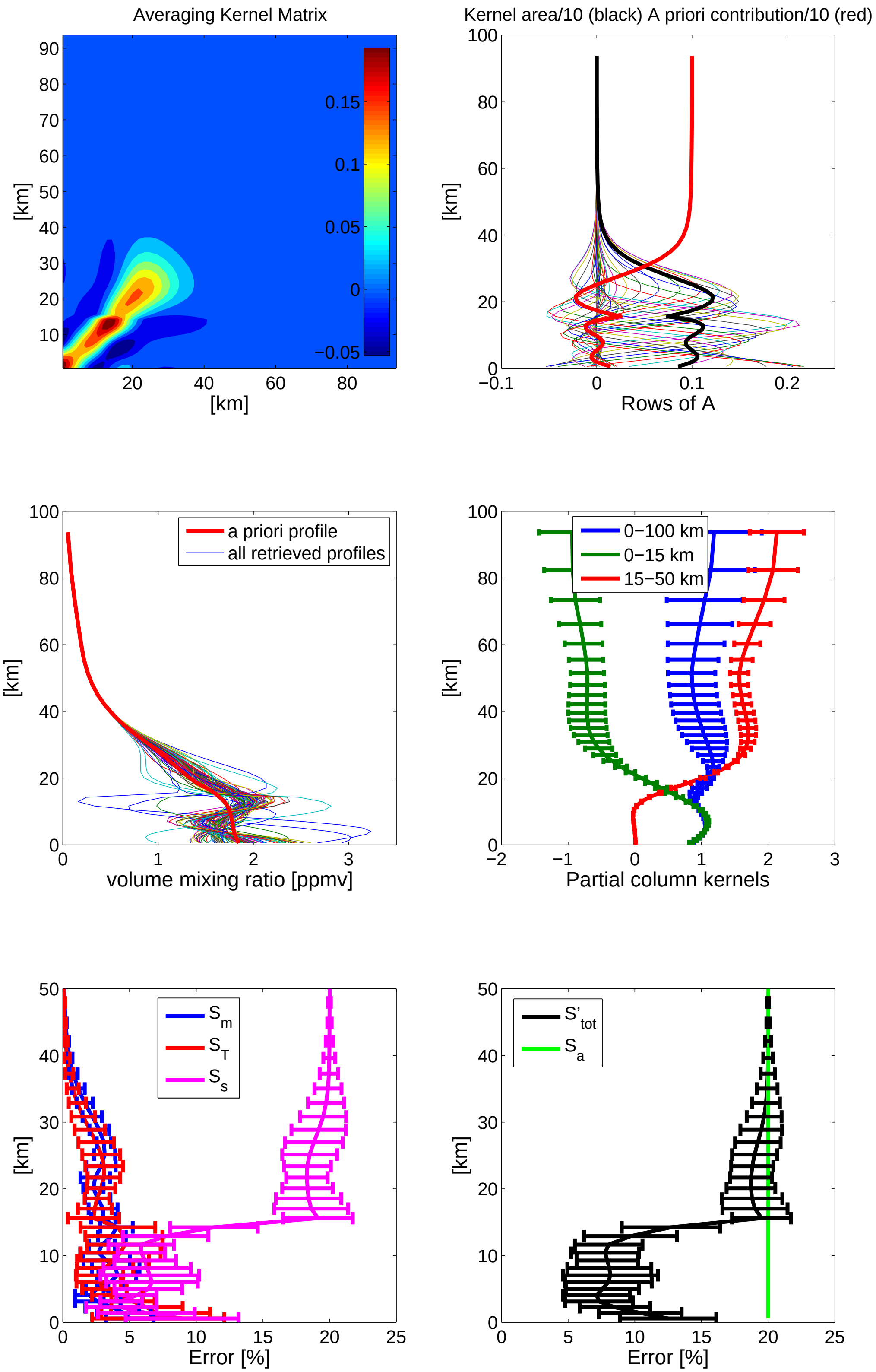


Figure 5.26 2004 average retrieval characterization for CH_4 (error bars show 1σ).

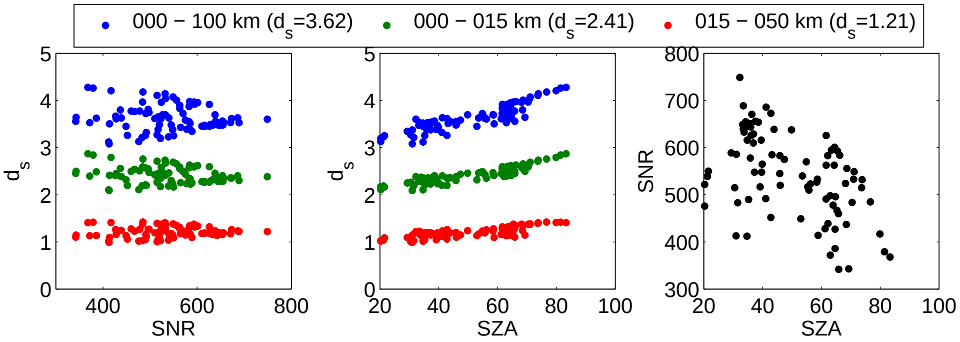


Figure 5.27 2004 CH_4 total and partial column d_s as a function of SNR and SZA (legend shows mean values). Relationship between SNR and SZA at TAO is shown for reference.

	[0-100 km]		[0-15 km]		[15-50 km]	
d_s (σ)	3.62	(0.27)	2.41	(0.17)	1.21	(0.11)
%a (σ)	1.03	(0.59)	0.65	(0.51)	4.60	(1.48)
S_s (σ)	0.52	(0.14)	0.80	(0.12)	4.20	(0.15)
S_m (σ)	0.80	(0.18)	0.84	(0.16)	1.62	(0.11)
S_T (σ)	3.09	(0.11)	3.42	(0.15)	2.18	(0.10)
S'_{tot} (σ)	3.24	(0.05)	3.62	(0.09)	5.00	(0.14)

Table 5.6 2004 average CH_4 total and partial column retrieval information content and percent errors, as defined in Table 4.1 (σ = one standard deviation of the mean.)

5.2.7 N_2O

The three-year time series of N_2O total and partial columns is shown in Figure 5.28, with the error bars smaller than the markers in most cases. The scatter in the total columns is too big to discern any seasonal pattern, however, there may be a weak maximum in stratospheric N_2O column amounts, which would have to be dynamic in origin, since N_2O destruction is expected to be higher in the summer-time stratosphere due to an increased solar insolation. A similar result has been reported for both N_2O and CH_4 by *Schneider et al.* [2005a], also using FTIR measurements, however, a seasonal cycle of CH_4 is not clear from Figure 5.24. A portion of the scatter of both N_2O and CH_4 total columns can be explained by variations in the tropopause height [*Washenfelder et al.*, 2003].

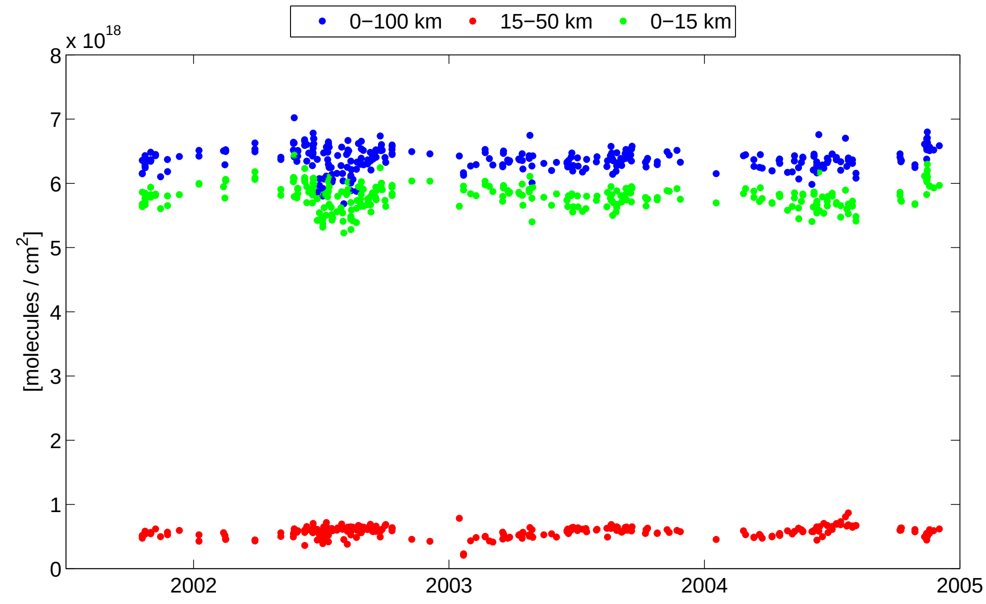


Figure 5.28 2001-2004 N_2O total and partial column time series recorded at TAO.

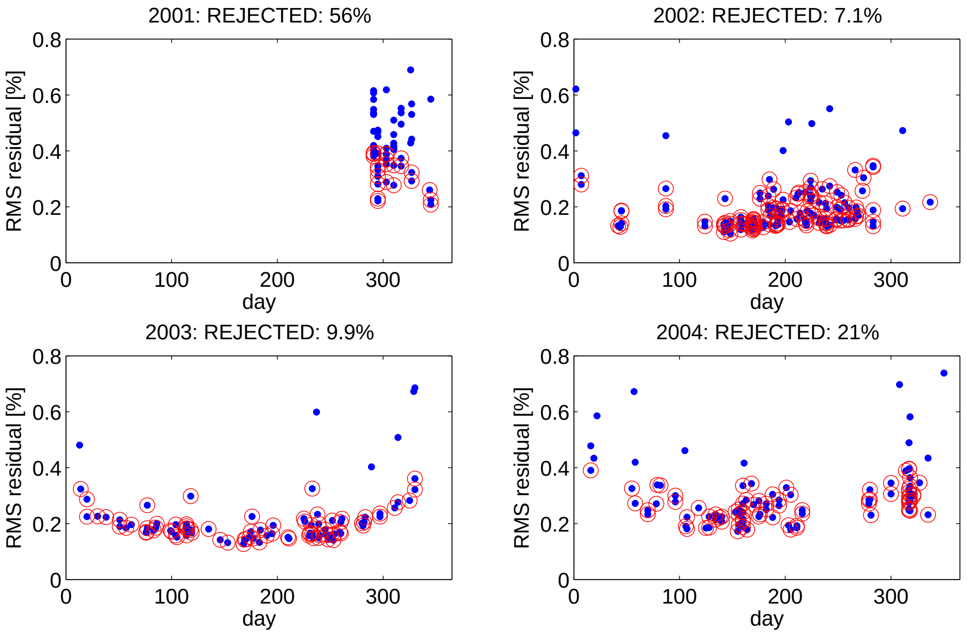


Figure 5.29 Spectral residuals of 2001-2004 N_2O fits at 2481 cm^{-1}. Values below the rejection criterion (0.4%) are circled in red.

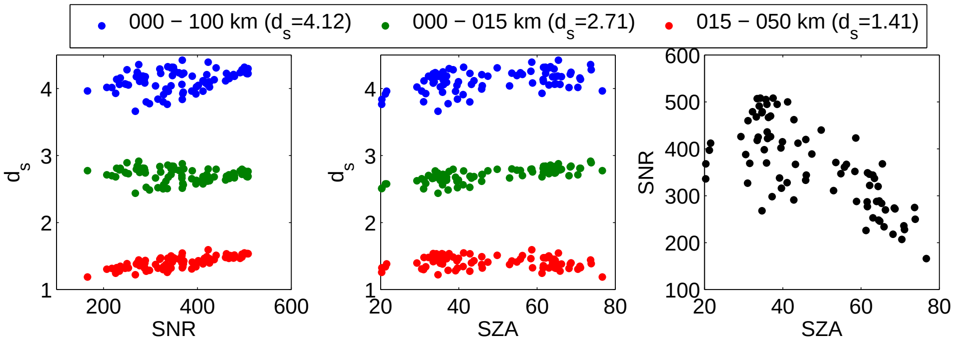


Figure 5.30 2004 N_2O total and partial column d$_s$ as a function of SNR and SZA (legend shows mean values). Relationship between SNR and SZA at TAO is shown for reference.

	[0-100 km]		[0-15 km]		[15-50 km]	
d_s (σ)	4.12	(0.16)	2.71	(0.10)	1.41	(0.09)
%a (σ)	0.14	(0.05)	-0.62	(0.05)	7.44	(0.44)
S_s (σ)	0.14	(0.03)	0.51	(0.05)	4.13	(0.15)
S_m (σ)	0.27	(0.05)	0.36	(0.03)	1.81	(0.04)
S_T (σ)	1.87	(0.12)	2.20	(0.18)	1.86	(0.42)
S'_{tot} (σ)	1.90	(0.12)	2.29	(0.17)	4.90	(0.13)

Table 5.7 2004 average N_2O total and partial column retrieval information content and percent errors, as defined in Table 4.1 (σ = one standard deviation of the mean.)

Figure 5.29 shows the RMS residuals of the spectral fits, with higher rejection rates in 2004, as for the other gases. Figure 5.30 shows d_s as a function of SNR and SZA, as well as the relationship between SNR and SZA. A clear separation between d_s values is observed, with ∼1 piece of information from the stratosphere and ∼2 from the troposphere. Figure 5.31 shows the 2004 average retrieval characterization, as for the other gases, which is again similar to the particular matrix singled out in Figure 4.50. The average errors shown in Figure 5.31 are also similar to those in Figure 4.50, and Table 5.7 summarizes the average d_s, % contribution of the *a priori*, and errors for total, stratospheric and partial columns, with the standard deviations of the averages again shown in brackets. The average stratospheric d_s in 2004 (1.41) is consistent with the ability to detect partial column trends in this region of the atmosphere, as is the case for CH_4 (d_s=1.21).

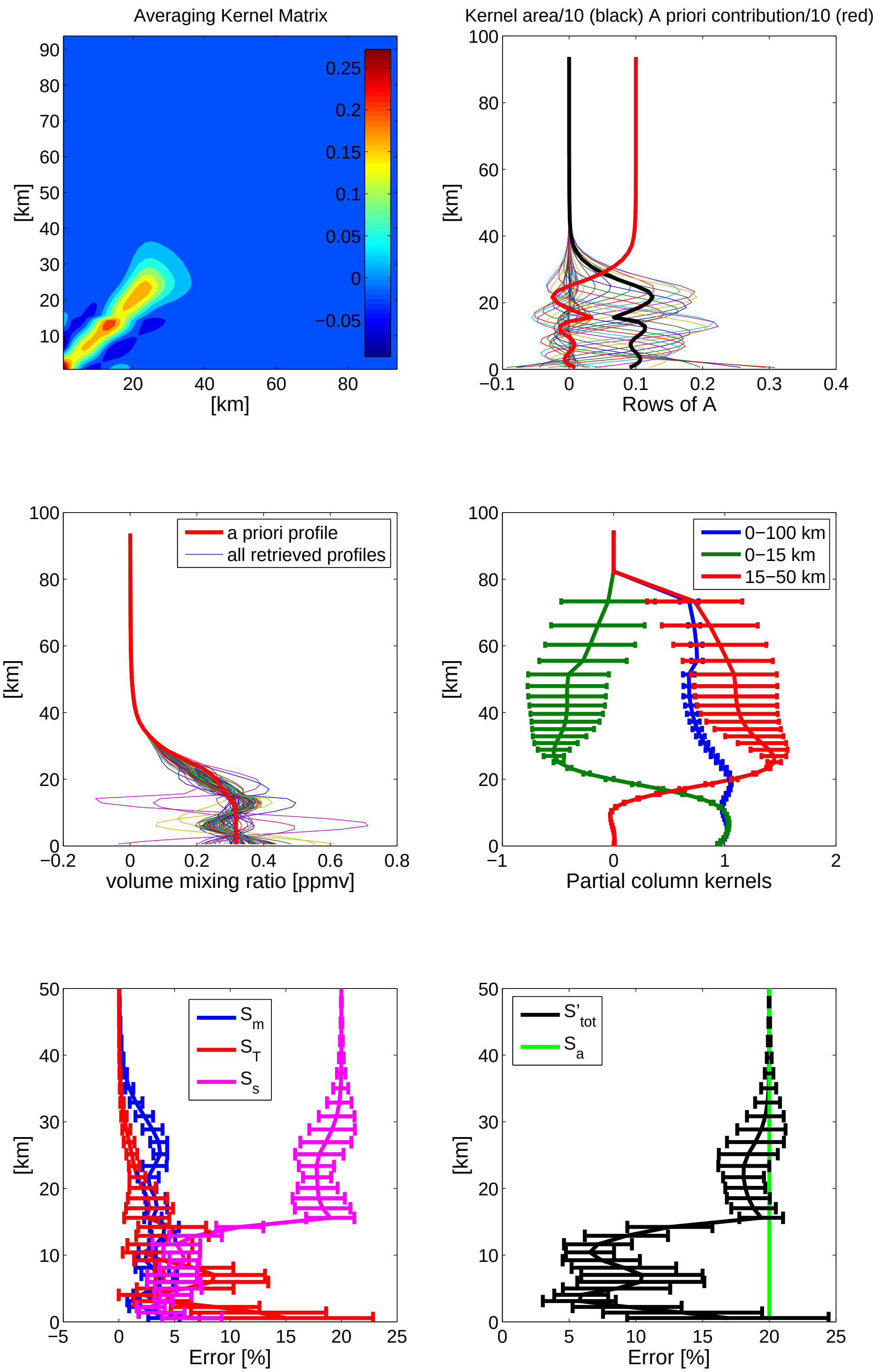


Figure 5.31 2004 average retrieval characterization for N_2O (error bars show 1σ).

Chapter 6

Applications and Comparisons with Other Data

In this final results chapter, we show the usefulness of TAO measurements for the detection of atmospheric constituents between 0-100 km. First, in Section 6.1 we outline retrieval method modifications necessary for the detection of Mesospheric-Lower Thermospheric (MLT) NO concentrations (50-130 km), and we apply the technique to spectra recorded at Toronto and Eureka during a period of increased solar activity and subsequent NO enhancements. This section was published in the refereed literature [*Wiacek et al.*, 2006a] and is reproduced here with permission of the American Geophysical Union. Second, we show comparisons of stratospheric column measurements of HCl at TAO with measurements performed by the ACE-FTS satellite instrument [*Mahieu et al.*, 2005], which show good qualitative agreement between the two observation platforms. A strong signature of the intrusion of polar air into mid-latitudes is identified in this data set and analyzed further in Section 6.2. Finally, in Section 6.3 we show good agreement between a ∼1.5 year time series of monthly mean total column measurements of CO at TAO and by the MOPITT satellite instrument. Combined, Sections 6.1-6.3 illustrate the usefulness of the growing TAO spectral record for the detection of trace gases in the troposphere and stratosphere, and in the case of NO, in the MLT region.

In the remainder of this chapter, we will make use of Ertel's *Potential Vorticity* (PV, $\mathrm{K} \cdot \mathrm{m}^2 \cdot \mathrm{kg}^{-1} \cdot \mathrm{s}^{-1}$), which is a quantity that is conserved when following a fluid's motion in the case of negligible friction (i.e. its material derivative, $D/Dt = \partial/\partial t + \overrightarrow{V} \cdot \overrightarrow{\nabla}$, is equal to zero, where $\overrightarrow{V} = u\overrightarrow{i_x} + v\overrightarrow{i_y} + w\overrightarrow{i_z}$ is the fluid's velocity field), and hence it is a useful tracer of the motion of air [*Houghton*, 2002]. Ertel's PV is defined on a potential temperature surface, θ (K), as

$$\frac{\zeta_\theta + f}{\rho} \frac{\partial \theta}{\partial z}, \tag{6.1}$$

where ζ_θ is the fluid's *relative* vorticity with respect to the Earth's surface (s^{-1}), here evaluated on a particular potential temperature surface, f is the *planetary* vorticity due to the Earth's rotation (s^{-1}), and ρ is the fluid density ($\mathrm{kg} \cdot \mathrm{m}^{-3}$). The potential temperature is defined as the temperature that an air parcel at a pressure p and at a temperature T would have if it were brought to the surface pressure, p_0, adiabatically:

$$\theta = T \left(\frac{p_0}{p} \right)^{\kappa}, \tag{6.2}$$

where $\kappa = (c_p - c_v)/c_p$, and where c_p and c_v are the specific heat of air at constant pressure and at constant volume, respectively. ζ_θ is the vertical component of the curl of the wind field ($\overrightarrow{\nabla} \times \overrightarrow{V}$) evaluated on a potential temperature surface

$$\zeta_\theta = \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right)_\theta, \tag{6.3}$$

where x and y denote the zonal (longitude) and meridional (latitude) directions in the local Cartesian co-ordinate system, and u and v denote the fluid velocity components in the same directions. Finally, the planetary vorticity is given by $f = 2\Omega sin\phi$, where Ω is the angular velocity of the Earth and ϕ is the latitude. Together, the sum of planetary and relative vorticity is referred to as the *absolute* vorticity.

Given temperature and wind fields, e.g. from NCEP or the European Centre for Medium-range Weather Forecasting (ECMWF), a scalar field of Ertel's PV can be calculated for a given potential temperature surface, which corresponds to a narrow range of heights in the atmosphere. High PV values are observed in polar stratospheric regions during the winter, after the formation of the stratospheric vortex. When high PV values are found at mid-latitudes in the winter or spring, this is a signature of the intrusion of polar air into these regions, as will be illustrated with PV maps shown in Section 6.2.

6.1 Retrievals of Meso-Thermospheric Nitric Oxide

We report the first detection of mesospheric-lower thermospheric (MLT, 50–130 km) NO from ground-based FTIR solar absorption spectra using Lorentz- and Doppler-broadened solar absorption lines in the stratosphere and in the MLT, respectively. We present the first characterization of vertical sensitivity in the FTIR NO retrieval and show that MLT NO partial columns can be retrieved with $\sim$1 independent piece of information using a climatological NO profile extending up to 130 km. The information content analysis also improves the characterization of stratospheric partial column retrievals and is relevant to NO results obtained at other NDACC FTIR sites. We apply our approach to spectra recorded at Complementary NDACC site Toronto (43.66°N, 79.40°W) during the solar storms of Oct–Nov 2003 and at Primary NDACC site Eureka (80.05°N, 86.42°W) during Feb–Mar 2004. MLT NO enhancements are found at Eureka, while possible enhancements at Toronto cannot be attributed to a particular altitude.

6.1.1 Introduction

NO is an important atmospheric trace constituent. In the troposphere it is a precursor of O_3 formation, whereas in the stratosphere it participates in its catalytic destruction. Increased MLT NO serves as a proxy for increased solar activity, with mesospheric NO produced by Solar Proton Events (SPE) and thermospheric NO produced by, e.g. energetic electrons or increased X-ray activity [e.g. *Semeniuk et al.*, 2005]; X-ray sources of NO need not be confined to the magnetic pole. NO produced by any of these mechanisms can be transported into the stratosphere during polar night by downwelling meridional circulation where it can later cause significant O_3 destruction [e.g. *Callis et al.*, 1996; *Jackman et al.*, 2005; *Rozanov et al.*, 2005]. *Rinsland et al.* [2005] recently showed the long-lasting effects on NO_x(=NO+NO_2) via the powerful Oct–Nov 2003 SPEs [*Woods et al.*, 2004] and subsequent NO_x production.

In this study we outline a retrieval approach used with the well-characterized optimal estimation-based retrieval algorithm SFIT-2 (v.3.91) [e.g. *Hase et al.*, 2004; *Rinsland et al.*, 1998], which enables the detection of MLT NO. The proper treatment of the MLT NO spectral signature in the inversion process also leads to an improved characterization of stratospheric partial column NO retrievals. We characterize the retrieval performed on a discrete 41-layer grid in terms of vertical resolution and information content from 0–130 km and, finally, apply our approach to spectra recorded at Toronto (54°N geomagnetic latitude) during the period of enhanced solar activity in Oct–Nov 2003 and at Eureka during the previously-described chemical enhancements of Feb-Mar 2004 [*Natarajan et al.*, 2004; *Orsolini et al.*, 2005; *Randall et al.*, 2005; *Rinsland et al.*, 2005; *Seppälä et al.*, 2004]. The Eureka measurements were performed during the 2004 ACE Arctic Validation campaign [*Kerzenmacher et al.*, 2005] at Environment Canada's Arctic Stratospheric Observatory (AStrO). The spectra were recorded using the Meteorological Service of Canada FTS housed at AStrO since 1993 [*Donovan et al.*, 1997], which is also a Bomem DA8 FTS.

6.1.2 Retrieval approach

The NO fundamental band absorption feature near 1900 cm^{-1} used in ground-based solar absorption FTIR spectroscopy is known for its relatively low and highly variable SNR, primarily due to broad absorption features by H_2O. *Notholt et al.* [1995] have reported ground-based vertical column densities of NO retrieved from this interval using a non-linear least squares spectral fitting algorithm. The approximate VMR *a priori*

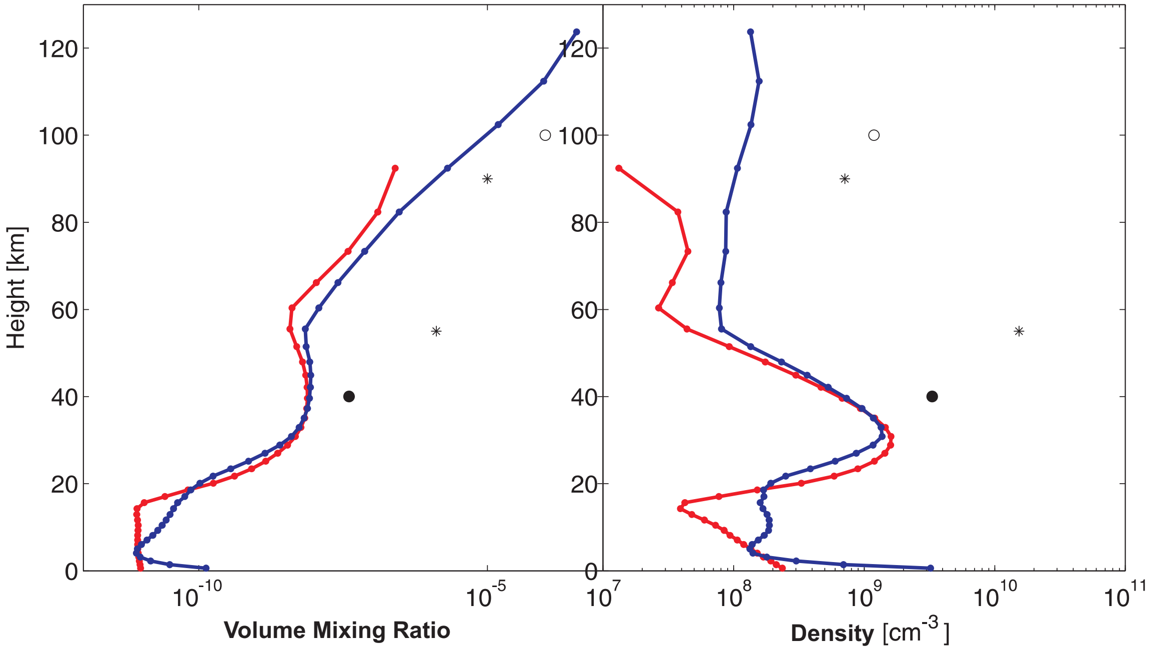


Figure 6.1 The NO *a priori* profile used in this study (blue; see text for details). The NO *a priori* profile after *Notholt et al.* [1995] (red) and satellite measurements of NO enhancements are shown for reference (○ = HALOE 75°S Nov 2003; ● = HALOE 71°N Apr 2004; ∗ = ACE-FTS 80°N Feb 2004).

profile used by *Notholt et al.* is shown in Figure 6.1, together with the revised and vertically extended NO VMR *a priori* profile used in our study, which is based on 123 HALOE sunset profiles between 1991 and 2004 within ±5° N-S and E-W of Toronto. Within each HALOE profile, only measurements with random errors less than 50% of the retrieved value were used to calculate seasonal means, which were then averaged to give an annual mean. At altitudes below 20 km, the MIPAS reference profile for NO was used (www.atm.ox.ac.uk/group/mipas/species/no.html); however, the retrieval is not sensitive to the high densities found near the surface since their spectral signatures are highly pressure broadened and well below the detection limits dictated by the SNR. Figure 6.1 also shows satellite measurements of NO_x enhancements highlighted by, e.g. *Semeniuk et al.* [2005], some nearly two orders of magnitude larger than normal.

We have extended the retrieval grid lid from 100 to 130 km since approximately

Table 6.1 Amount and percent (in brackets) of total NO column (x10^{14} molec/cm^2) found in each region of the atmosphere for the *a priori* VMR profile used in this study and (A) 0-100 km retrieval grid, (B) 0-130 km retrieval grid, (C) 10-fold VMR enhancement between 90-130 km, (D) 10-fold VMR enhancement between 50-90 km (both enhancements consistent with satellite measurements in Figure 6.1).

	Tropo 0–15	Strato 15–50	Meso 50–90	Thermo 90–130[a]	Total 0–130[a]
(A)	0.55 (17)	2.24 (69)	0.34 (10)	0.15 (05)	3.27 (100)
(B)	0.55 (15)	2.24 (61)	0.34 (09)	0.57 (15)	3.70 (100)
(C)	0.55 (06)	2.24 (26)	0.35 (04)	5.60 (64)	8.74 (100)
(D)	0.55 (09)	2.24 (35)	3.06 (47)	0.59 (09)	6.45 (100)

---

[a] except in (A) where only 90-100 km and 0-100 km

74% of the thermospheric column (90–130 km) and 11% of the total column (0–130 km) are found between 100–130 km during undisturbed conditions (Table 6.1). This upper limit captures the thermospheric peak of NO and NO variability [e.g. *Barth et al.*, 2003]. Assuming a 10-fold VMR enhancement between 90–130 km typical of enhanced solar activity or wintertime auroral activity, as much as 53% of the total column of NO can be found between 100–130 km (64% between 90–130 km), which will lead to errors in the retrieved vertical distribution of NO when the retrieval lid is only at 100 km. The effect of setting the NO VMR to 10^{-11} between 50-130 km (a depletion relative to the VMR profile in Figure 6.1) on the forward modelled spectrum is shown in Figure 6.2. Neglecting NO in the MLT in this manner results in ∼6% higher transmission levels for SZA of 85°. Conversely, the effect of a 10-fold enhancement in the MLT NO VMR profile results in ∼40% lower transmission levels (Figure 6.2). When transmissions are modeled at SZA=45°, the peak differences are still well within the detection limits of the ground-based FTIR technique at ∼2% and 17%, respectively. Finally, the impact of extending the retrieval grid above 130 km could become non-negligible during conditions of strong enhancement.

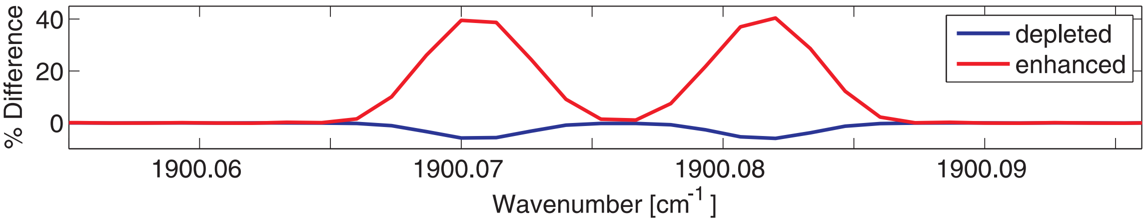


Figure 6.2 Modelled transmission differences (SZA=85°) for an *a priori* profile that is depleted or enhanced in NO between 50–130 km (see text for details).

Table 6.2 Retrieval intervals (cm^{-1}), interfering species and SNR for Toronto (T) and Eureka (E).

	(E)	SNR (E)	(T)	SNR (T)
1875.795–1875.830	NO	500	–	–
1900.055–1900.100	NO	500	NO	320
1900.500–1900.540	NO	500	NO	250
1903.050–1903.200	NO	500	NO, H_2O	150
1906.120–1906.175	NO	500	–	–
1975.283–1975.408	H_2O, O_3	500	–	–

Our retrieval approach consists of simultaneously fitting the high-resolution (0.004 cm^{-1}) solar absorption spectrum of NO in the microwindows listed in Table 6.2. NO absorption features were identified using the line-finding program of *Notholt et al.* [2006]. Using the spectroscopic parameters of HITRAN 2004 [*Rothman et al.*, 2005] and pressure and temperature profile information from daily sonde launches (available only for Eureka) and NCEP (obtained from the Goddard Automailer, science@hyperion.gsfc.nasa.gov), NO was fitted as a profile while the *a priori* profiles of H_2O and O_3 were scaled by a single factor in the retrieval. In each microwindow, we also retrieved a single wavenumber shift, a broad slope and a curvature parameter; interfering solar CO absorption features were also modelled and retrieved.

The measurement covariance matrix ($\mathbf{S}_\epsilon$) for Eureka retrievals was set to be diagonal and to correspond to a SNR of 500, which is on average appropriate for all microwindows and spectra. Toronto retrievals (only performed over microwindows with sufficiently high

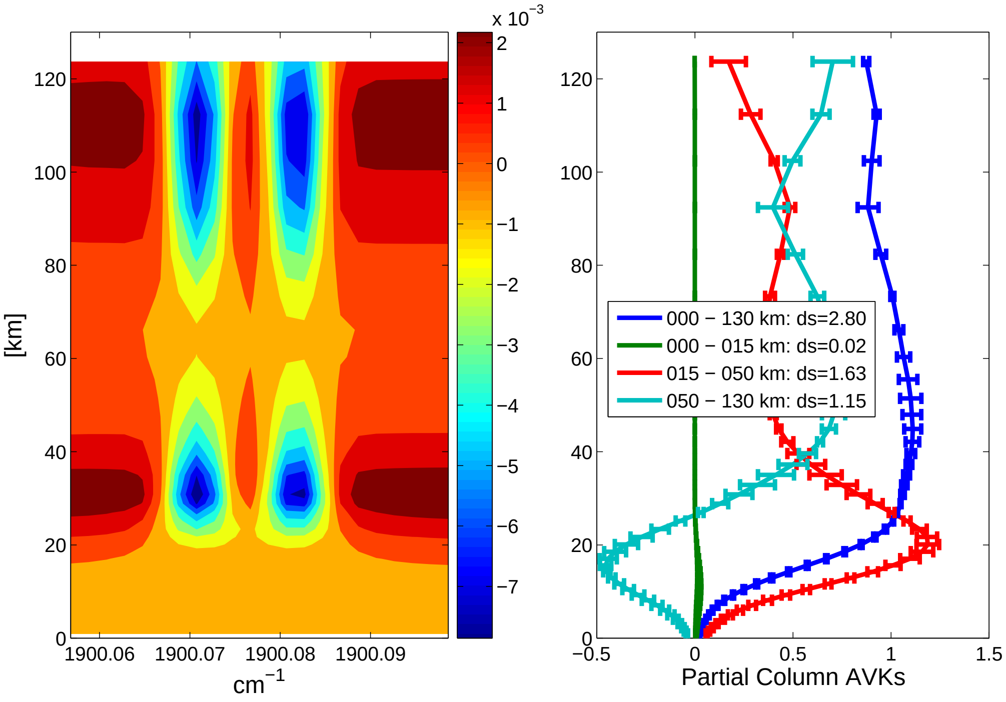


Figure 6.3 Left: average (for all Eureka data) NO weighting function matrix in one microwindow. Right: average Eureka partial column averaging kernels for NO (error bars show std. deviation).

SNRs) used a variable $\mathbf{S}_\epsilon$ to account for the varying SNR in the spectra due to strong absorption by H_2O. Finally, the NO *a priori* covariance matrix ($\mathbf{S}_a$) was set to be diagonal and 100% on all levels for both sites.

6.1.3 Retrieval characterization

When the NO *a priori* VMR profile is increased to agree with the HALOE climatology (which is much lower than the observed enhancements) and the retrieval grid is extended to 130 km then sensitivity to the MLT region increases in the NO retrieval. The weighting function matrix (Figure 6.3) shows two distinct regions of vertical sensitivity: in the stratosphere (mixed Lorentz and Doppler broadening) and in the MLT (Doppler broadening). While each partial column region is characterized by more than one independent piece of information (degrees of freedom for signal (d_s) > 1), the partial column averaging kernels are not ideally separated in altitude (Figure 6.3); therefore, the retrieved partial columns are not completely uncorrelated. This is expected since the spectral signatures of mixed Lorentz and Doppler broadening and pure Doppler broadening are of comparable width, which can be inferred from Figure 6.3.

It is instructive to contrast these findings with strato-mesospheric retrievals of CO reported by *Kasai et al.* [2005]. Since CO is mainly a tropospheric species, there is a sharp contrast between the broad tropospheric Lorentz lineshapes and the narrow Doppler lineshapes of higher altitudes. Accordingly, the averaging kernels for that retrieval are better separated in altitude. In the case of NO, whose peak density is normally found in the stratosphere, the Lorentz broadening is already narrow and comparable to Doppler broadening, leading to potential mis-assignment of altitude in the retrieval. If the *a priori* profile is unchanged from that used by *Notholt et al.* [1995] then the weighting function matrix will tend to zero in the MLT and any NO enhancements in this region will automatically be assigned to the stratosphere, regardless of the $\mathbf{S}_\epsilon$ and $\mathbf{S}_a$ employed in the retrievals.

The total column averaging kernel for Eureka retrievals (Figure 6.3) shows good sensitivity from 20–130 km, while for Toronto retrievals (not shown) it drops off above 100 km and reaches ~0.75 at 120 km. This is due to generally smaller SZAs and SNRs in Toronto spectra and is reflected in lower d_s values (0–130 km: 1.67; 0–15 km: 0.03; 15–50 km: 1.06; 50–130 km: 0.58). Furthermore, the average *a priori* contribution to the retrieved partial columns at Eureka was calculated as 2.3%, 94.1%, -2.7%, and 5.7% between 0–130 km, 0–15 km, 15–50 km, and 50–130 km, respectively. Finally, the average errors in the same (Eureka) partial columns due to errors in the temperature profile were calculated as 1.9%, 2.6%, 3.2%, and 6.7%, respectively, assuming a 100 K^2 (25 K^2) temperature variance above (below) 50 km and a 5 km correlation length [see, e.g. *Rodgers*, 2000]. Since larger temperature variations are possible in the MLT region, their effect on

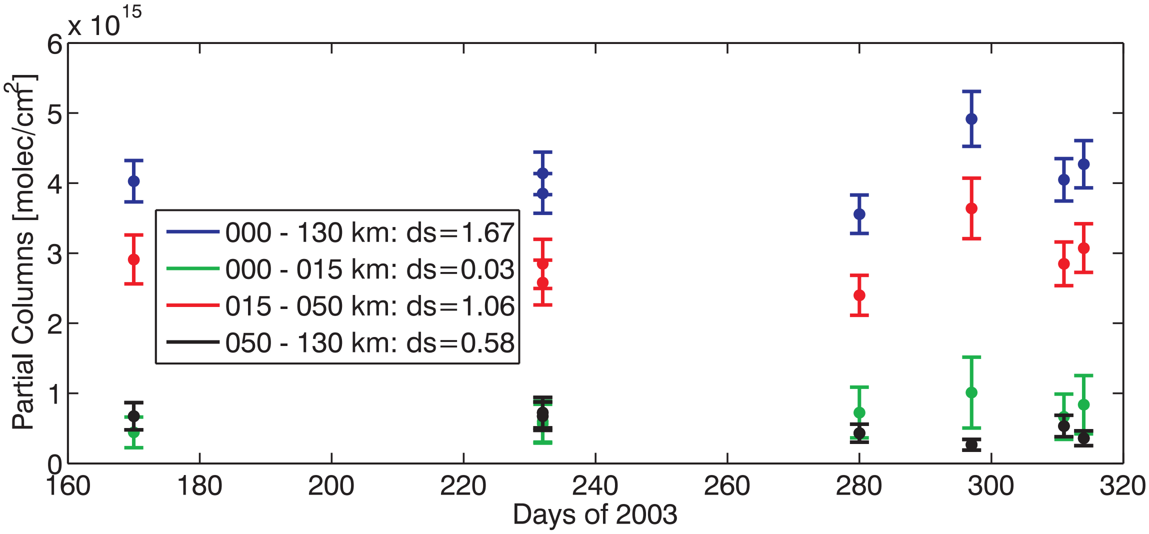


Figure 6.4 Partial columns of NO retrieved over Toronto (SPEs occurred between day 291-309).

retrievals should be investigated further.

6.1.4 Observations of NO enhancements

Temporal sampling of Toronto NO measurements was poor in late 2003 but spectra were recorded on Jun 19, Aug 20, Oct 7, Oct 24, Nov 07, and Nov 10 (Figure 6.4). The Oct 24 record was well into the period of enhancements but before the super flare of Oct 28; the Nov 7 and Nov 10 records follow the super flare of Nov 5 [*Woods et al.*, 2004]. These fall total columns are as high as those recorded in the summer, possibly indicating a detection of the known MLT NO enhancements; however, in Figure 6.4, any enhancements in Oct–Nov 2003 are attributed to the stratosphere (stratospheric partial columns are well-correlated with total columns). This is not consistent with the seasonal decrease of stratospheric NO towards the winter at mid-latitudes. In this context, the total column values recorded in Oct–Nov 2003 may indicate MLT enhancements.

Photochemical box model calculations provided by C. A. McLinden (personal communication) indicate that the stratospheric column of NO increases by ∼18% in July and

40% in January from 10AM to 4PM. This is not enough to explain the elevated total column values over Toronto in Oct–Nov 2003, especially given that all 2003 spectra were recorded between 10AM–2PM (2004 Eureka spectra were recorded near 9AM and noon).

Finally, the noontime photochemical lifetime of NO_x in the mesosphere and thermosphere at sunlit mid-latitudes is 1-2 days [*Brasseur and Solomon*, 1986; *Jackman et al.*, 2005]. In the lower mesosphere and upper stratosphere the lifetime of NO_x increases to weeks. Thus, given Toronto's geomagnetic latitude and the presence of powerful SPEs and secondary NO sources (X-rays and energetic electrons) that all persisted for about two weeks, it is reasonable to expect that NO would remain enhanced long enough to be measured at Toronto on Oct 24, Nov 07 and Nov 10.

For spectra recorded at Eureka, the total columns are elevated but not correlated with stratospheric columns as in Toronto data (Figure 6.5). High values retrieved in the MLT before day 70 also correspond to high PV values, albeit at lower altitudes (1 mb, from Goddard Automailer). After this date, the MLT partial columns decrease rapidly over ∼5 days, seemingly due to disturbed dynamic conditions. Photochemical destruction is also playing a role in the decreasing MLT concentrations (polar sunrise occurs on day 51); however, *Rinsland et al.* [2005] have reported an elevated-NO_x e-folding time of 15 days between day 46–83, thus we still expect to observe elevated MLT concentrations during our measurement interval. After day 70 we observe increasing stratospheric partial columns associated with the release of NO_x from HNO_3 and N_2O_5.

6.1.5 Summary and conclusions

We have demonstrated that the ground-based solar absorption FTIR retrieval has a sensitivity to MLT NO partial columns (d_s ∼1) provided that: 1) a reasonable VMR *a priori* profile is used, 2) the retrieval grid is extended to 130 km, and 3) the spectra are recorded at high SNRs and SZAs. A full averaging kernel and information content analysis reveals that when the above conditions are not met then enhancements in MLT partial columns

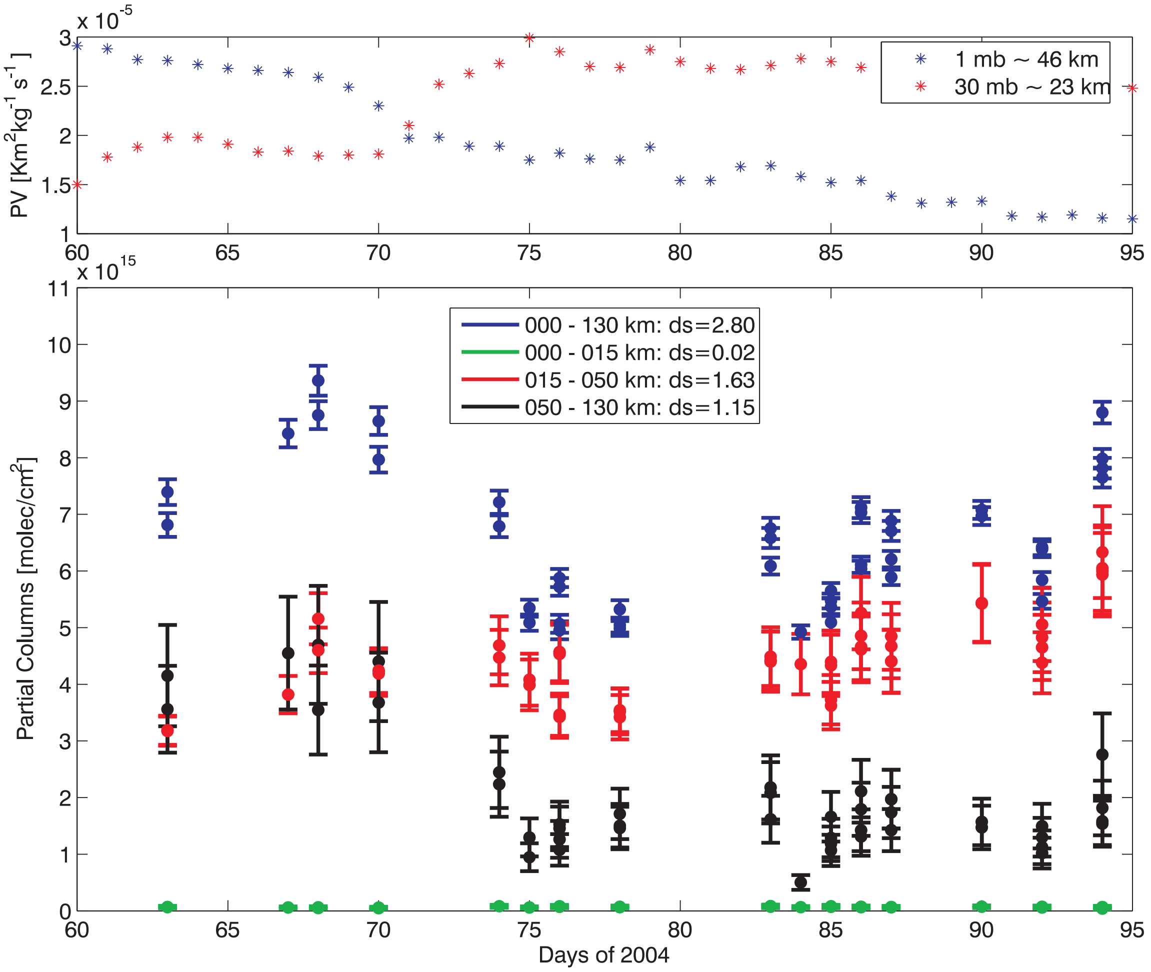


Figure 6.5 NO partial columns retrieved over Eureka. PV indicates disturbed dynamic conditions near day 70.

will be incorrectly assigned to the stratosphere. Further increases to the information content and vertical resolution of NO retrievals are possible given measurements with an increased spectral resolution and a higher signal to noise ratio, as well as through the use of a more detailed forward model of the NO absorption lineshape in the retrieval process.

The ground-based FTIR retrieval provides an additional means of constraining satellite measurements of baseline and SPE-enhanced MLT NO concentrations, especially at NDACC sites and during campaigns operating in the vicinity of the magnetic pole, where MLT NO production is high.

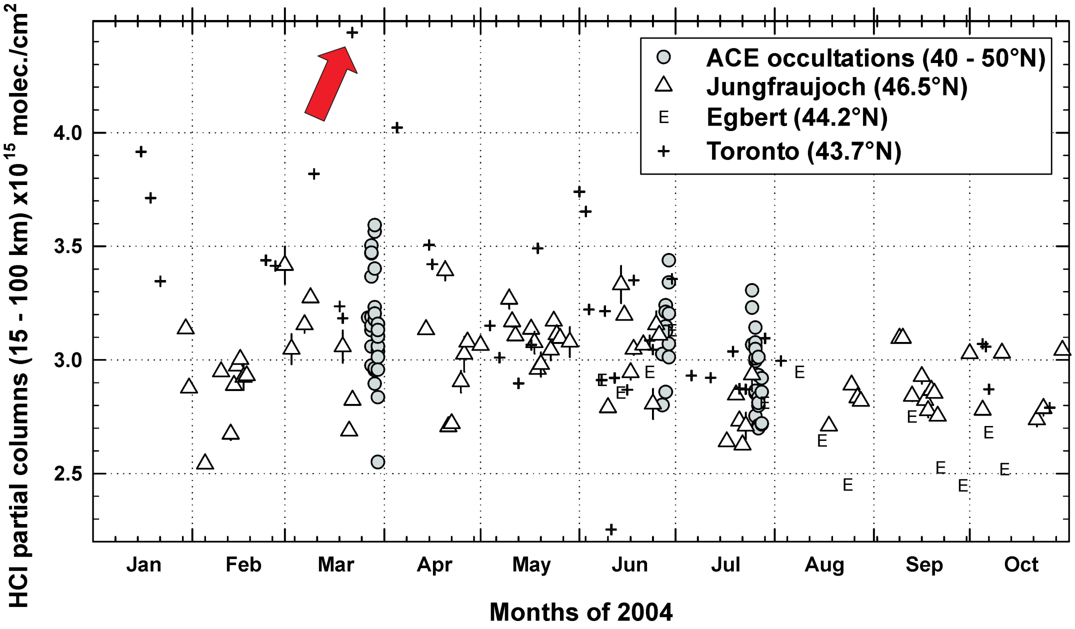


Figure 6.6 Comparison of stratospheric HCl columns recorded by ground-based stations and by the ACE-FTS satellite instrument. An exceptionally high HCl value recorded above Toronto is highlighted with an arrow (adapted from *Mahieu et al.* [2005])

6.2 Detection of Polar Intrusions Above Toronto

The 2004 time series of HCl stratospheric columns retrieved from spectra recorded at TAO was compared with ACE-FTS satellite measurements (launched in August 2003, and described by *Bernath et al.* [2005]) by *Mahieu et al.* [2005]. Figure 6.6 shows good qualitative agreement between the two observation platforms, as well as an exceptionally high value of HCl retrieved above Toronto on March 22^{nd}, 2004. The quality of this measurement was re-checked and found to be comparable with the rest of the dataset; furthermore, it was shown to correspond to a high stratospheric column of the long-lived trace gas HF (also retrieved at TAO on the same day), suggesting that both high values were caused by dynamic variability. Figure 6.7 shows a high level of correlation between same-day TAO measurements of HCl and HF in 2004, with the March 22^{nd} measurement highlighted for reference, which suggests that both species are primarily controlled by dynamics above Toronto.

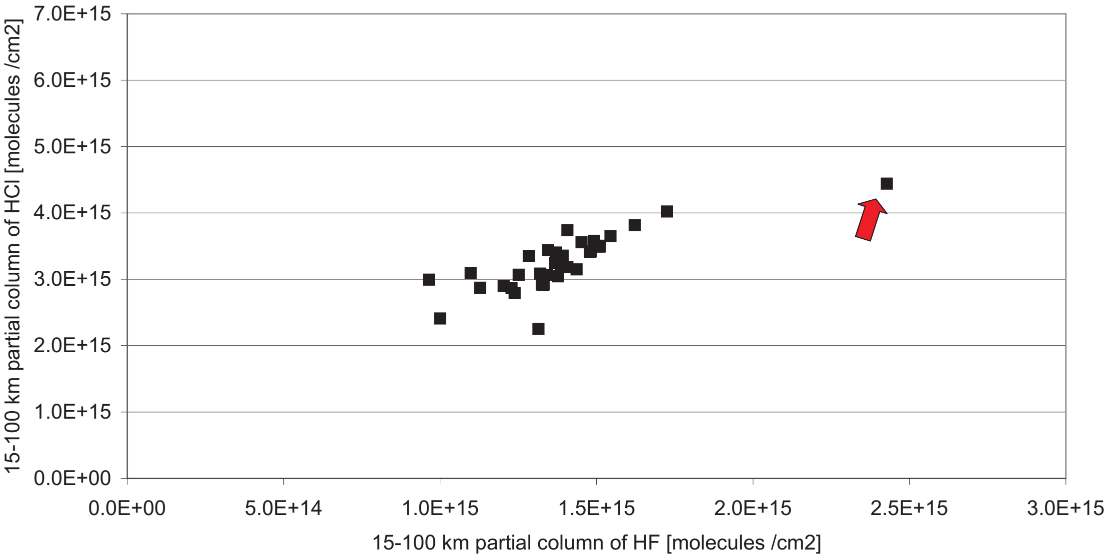


Figure 6.7 Same-day TAO observations of HCl and HF in 2004. The red arrow is highlighting the outlying point from Figure 6.6 observed on March 22^{nd}, 2004.

The extent of dynamical vs. chemical control of decreasing total ozone amounts at mid-latitudes is an unsolved scientific question [*WMO*, 2003], where proposed mechanisms include *in situ* chemical destruction, changes in the intensity of the Brewer-Dobson circulation, and the transport to mid-latitudes of polar air depleted in O_3 during wintertime stratospheric vortex disturbances or the final spring-time vortex breakup (see, e.g. *Marchand et al.* [2003], *Millard et al.* [2003], and references therein). Furthermore, the Northern Hemisphere (NH) total column trend observed in the winter/spring is twice as large as that observed in the summer/fall ($\sim$4% per decade vs. $\sim$2% per decade between 1997-2001 relative to pre-1980 values), while the Southern Hemisphere (SH) mid-latitudes display relatively little seasonal dependence in a trend of $\sim$6% per decade [*WMO*, 2003, Ch.4]. The attribution of these trends to dynamical vs. chemical causes originating in the polar regions is complicated by the fact that chemical O_3 depletion within the polar vortex is tightly coupled to vortex dynamics, with higher interannual variability observed in both the dynamics and chemistry of the NH vortex. Studies to quantify the transport of polar air to NH mid-latitudes are ongoing [e.g. *Hauchecorne* *et al.*, 2002; *Godin et al.*, 2002], and show that a cold and stable Arctic vortex leads to few vortex filament intrusions into mid-latitudes, however, that these intrusions can have a non-negligible local effect because of higher O_3 loss in a cold and stable vortex. Finally, a high-resolution three-dimensional chemistry-transport model is required to accurately predict disturbances in the polar vortex and the subsequent chemical processing, which can persist into the summer at mid-latitudes [*Durry and Hauchecorne*, 2005].

Since the 2003-2004 winter was warm [*Manney et al.*, 2005], we expect to see relatively little conversion of HCl to $ClONO_2$, and subsequent conversion of $ClONO_2$ to active chlorine via heterogeneous reactions on PSC surfaces. As a result, O_3 depletion during this winter is also expected to be relatively low. Additionally, the Arctic vortex was also highly disturbed in 2003-2004 [*Manney et al.*, 2005], and hence we expect an increased frequency of vortex (and vortex filament) excursions into mid-latitudes. Indeed, the ECMWF and MIMOSA 475 K PV maps shown in Figure 6.8 reveal a vortex filament with PV values greater than $30\text{x}10^{-6}\text{Km}^2\text{s}^{-1}\text{kg}^{-1}$ sweeping over Toronto on March 22^{nd}, 2004. MIMOSA stands for *Modélisation Isentrope du transport Méso-échelle de l'Ozone Stratosphérique par Advection* and is a high-resolution advection model of potential vorticity developed at the Service d'Aéronomie [*Hauchecorne et al.*, 2002]. MIMOSA uses six-hourly ECMWF analyses at 1.125° latitude and longitude resolution, and is able to predict the location of polar filaments with an accuracy better than 100 km. It has been used to forecast high-PV events at mid-latitudes and to send real-time alerts to a ground-based network of ozone lidars and ozonesonde stations, e.g. the Observatoire Haute-Provence (OHP), which is located at a latitude identical to TAO, in France. MIMOSA PV maps are available at www.aerov.jussieu.fr/∼fgoutail/MIMOSA/ARCHIVES/.

Finally, in the context of well-correlated and elevated HCl and HF values, a warm winter, a disturbed polar vortex, and PV maps showing a vortex filament over Toronto at the 475 K θ surface, we can conclude – and confirm – that an intrusion of polar air into mid-latitudes has occurred. The TAO measurements also show an elevated total

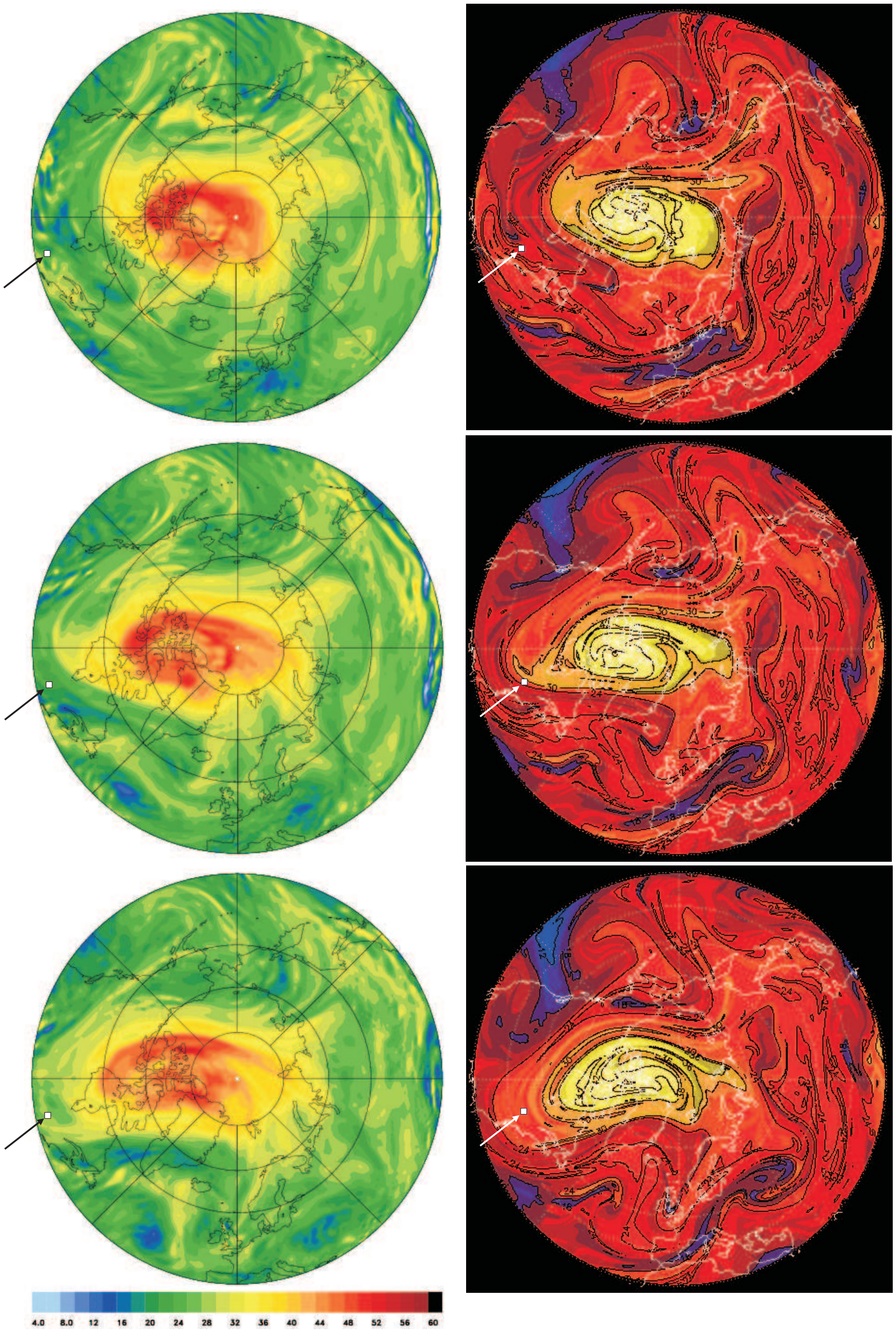


Figure 6.8 ECMWF (left, 40-90°N, 0Z) and MIMOSA (right, 30-90°N, 12Z) PV maps ($10^{-6}\mathrm{Km^2s^{-1}kg^{-1}}$) on the 475 K θ surface for March 21st (top), March 22nd (middle), and March 23rd (bottom). Toronto is marked with a white square.

column of O_3 on March 22^{nd}, consistent with this scenario (Figure 5.4). In summary, the usefulness of TAO data to help constrain dynamical vs. chemical ozone loss at mid-latitudes is clear. Given advanced warning of a polar air event, efforts to record both Filter 1 and Filter 3 spectra can be increased in order to expand the data set coinciding with these events. Filter 6 measurements will also be useful for the detection of $ClONO_2$ and HNO_3, as well as the strong O_3 absorption feature at 1146 cm^{-1}. Finally, FTIR measurements can be combined with ground-based DOAS and Brewer spectrophotometer measurements at TAO, FTIR measurements at the nearby Egbert Environment Canada site, complementary satellite measurements, as well as suitable modeling tools in order to study mid-latitude O_3 loss.

6.3 Comparisons with MOPITT CO Measurements

We now turn to comparisons of TAO total columns of Carbon Monoxide (CO) with total columns derived from the Measurements Of Pollution In The Troposphere (MOPITT) satellite instrument [*Drummond and Mand*, 1996; *Deeter et al.*, 2003]. CO is a poisonous tropospheric gas that is oxidized in the presence of NO to produce toxic ground-level O_3 – a key constituent of photochemical smog, which both causes and exacerbates respiratory difficulties in humans. CO and NO are produced during incomplete combustion in cars, energy production, and other industrial processes in urban centers. CO is additionally produced by natural processes such as savannah and forest wildfires, and by human activities such as the burning of wood fuel and prescribed agricultural biomass burning. The chemical processing of CO is a major sink for the OH radical, a key oxidant in the troposphere responsible for its self-cleansing capacity. A reduction of OH leads in turn to increased concentrations of greenhouse gases such as CH_4 that would otherwise have been chemically processed by OH. For this reason, CO is termed as an "indirect greenhouse gas" [*IPCC*, 2001] and is deemed responsible for significant radiative forcing

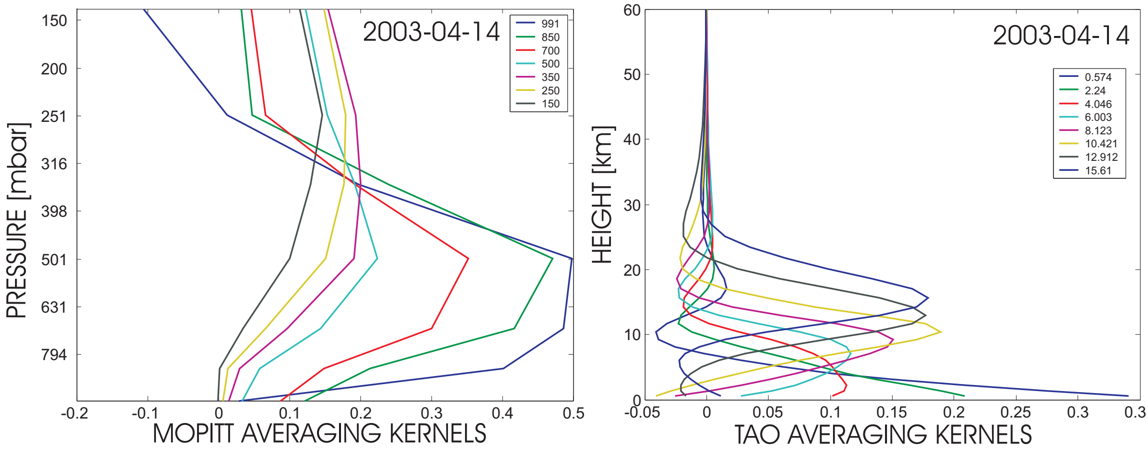


Figure 6.9 MOPITT and TAO averaging kernels for 2003-04-14 (nominal kernel heights in mbar and in km, respectively). Lowest MOPITT level at 991 mbar or $\simeq$150 m; highest MOPITT level at 150 mbar or $\simeq$13 km. (MOPITT kernels courtesy of Dr. Jane Liu.)

on the atmosphere. Because the lifetime of CO in the troposphere is 1-2 months (before it is further oxidized to CO_2) it is an excellent tracer of the tropospheric transport of pollutants, e.g. from Asia to North America [e.g. *Heald et al.*, 2003], or from North America to Europe [e.g. *Li et al.*, 2002].

Since March 2000, CO has been monitored with three-day global coverage from NASA's Terra satellite using the MOPITT instrument, which was developed by the University of Toronto and COM-DEV International Ltd. MOPITT is a gas correlation filter radiometer which derives total column and limited vertical information from the 4.7 μm (2100 cm^{-1}) absorption feature of CO. MOPITT's 22 km^2 footprint represents a high horizonal resolution that is suitable for the detection of a species with large horizontal gradients in the troposphere. Its sun-synchronous orbit means that it crosses a given latitude circle at the same time each day, in TAO's case at ~11:30 AM and again at ~3:30 AM. Although MOPITT retrievals are reported on seven pressure surfaces, the averaging kernels shown in Figure 6.9 reveal that it is primarily sensitive to the free troposphere and that it provides only ~1-2 independent pieces of vertical information [*Deeter et al.*, 2004a]. In comparison, Figure 6.9 shows that TAO retrievals can resolve

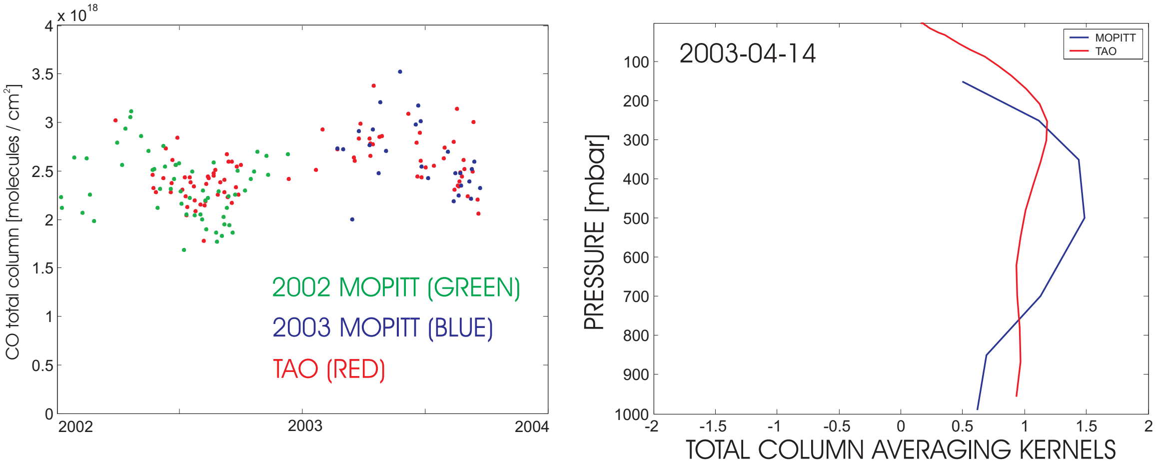


Figure 6.10 Daytime total columns measured by MOPITT and TAO in 2002-2003. A sample total column averaging kernel shows MOPITT's reduced sensitivity to the lowermost troposphere, in contrast to the sensitivity of TAO measurements.

up to two independent layers in the troposphere, and show considerable sensitivity to the lower stratosphere.

In the total column comparison below, we use daytime v.3 MOPITT column data (courtesy of Dr. Jane Liu and Dr. Holger Bremer) with a coincidence criterion of ± 0.5° in latitude and longitude. The TAO retrieval was simultaneously performed in the 2057 cm^{-1} and 2069 cm^{-1} microwindows (Figures 4.51 and 4.52), and another microwindow containing only a strong solar CO absorption feature at 2112.15 cm^{-1}. The last microwindow was used to better fit the solar CO absorption features which interfere in the first two microwindows. Finally, the saturated CO microwindow that displays the strongest sensitivity near the surface (Figure 4.53) was not used in this retrieval in order to better match the overall sensitivity to CO displayed by MOPITT. Figure 6.10 shows good qualitative agreement between all spatially coincident total columns measured by MOPITT and TAO from Jan. 2002 - Sep. 2003, and illustrates the reduced temporal coverage of both observation platforms in the winter due to cloud interference. Monthly means were calculated from both data sets and are shown in Figure 6.11. The two data sets agree within one standard deviation of the monthly mean for all months where this

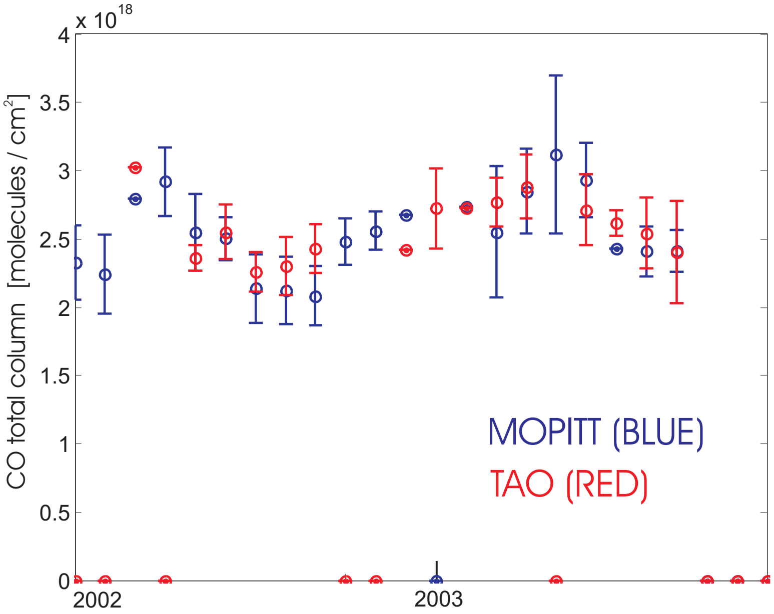


Figure 6.11 MOPITT and TAO monthly means for 2002-2003. Error bars indicate standard deviation of monthly mean. Zero standard deviations indicate not more than one observation, while zero monthly means indicate no observations for that month.

quantity could be calculated (see figure text), demonstrating the usefulness of TAO data for the investigation of tropospheric trace gases.

Although the agreement shown in Figure 6.11 is encouraging, the differences between MOPITT and TAO total column averaging kernels illustrate the need to employ the averaging kernel formalism of *Rodgers and Connor* [2003], especially in a rigorous comparison of measurements coincident in both space and time. In order to compare the retrieved profiles of MOPITT and TAO (and the total columns which are the integrals of the profiles), it is not correct to assume that one profile is so highly resolved that it can represent the “true” retrieved profile in Equation 3.47. Indeed, since MOPITT is a nadir sounder that “looks down” and since TAO is a ground-based instrument that “looks up”, their vertical resolution is comparable (except near the surface), and relatively poor in comparison to some other observation platforms. *Rodgers and Connor* [2003] outline an approach whereby one transforms a vertical profile retrieved at TAO such that it is directly comparable to a MOPITT vertical profile, while taking into consideration TAO averaging kernels, and the MOPITT *a priori* profile together with its prior covariance. This approach has been implemented by *Barret et al.* [2003] in a comparison of MOPITT CO columns with observations from the Alpine NDACC Station Jungfraujoch (46°N, 8°E, 3580 m a.s.l.). It remains to be implemented at TAO, although it is expected to play a larger role in the comparisons given TAO's low altitude. The complications due to the low-resolution averaging kernels of both ground-based FTIR observations and MOPITT are the reason why the global network of NDACC FTIR observation stations did not play a central role in MOPITT validation [*Deeter et al.*, 2004b; *Emmons et al.*, 2004], although some comparisons were carried out [e.g. *Barret et al.*, 2003; *Yurganov et al.*, 2005].

Chapter 7

Conclusion

7.1 Summary of Results

Fulfilling one of the goals of this thesis, a new high-resolution ($0.004\,\mathrm{cm}^{-1}$) research-grade Fourier transform spectrometer and its accompanying suntracker were installed at the Toronto Atmospheric Observatory, commissioned in October, 2001. The new facility fills a coverage gap in North America in the international Network for the Detection of Atmospheric Composition and Change. The FTS performance was characterized through blackbody and low-pressure HBr and N_2O gas cell measurements, and subsequent improvements were made to its alignment capabilities, the MCT detector, and system automation. The Candidate helped design operational and maintenance lab protocols, serviced the system, and recorded infrared spectra on 83 observation days (representing approximately one quarter of all measurements) between October 2001 – December 2005, as part of a rotating schedule. Finally, she worked closely with two summer students to streamline computer codes for the pre-processing of TAO spectra.

Also in fulfillment of the goals of this thesis, the Candidate implemented an existing Optimal Estimation Method retrieval algorithm (SFIT-2), examined its performance in depth and outlined avenues for future improvements, and finally, optimized a set of prior constraints necessary for the retrievals of vertical profiles and partial columns of trace gas abundances using the OEM approach. She created a set of *a priori* Volume Mixing Ratio (VMR) profiles, suited to TAO's location, for O_3, NO, NO_2, HCl, HF, N_2O, CH_4, H_2O, $ClONO_2$, HNO_3, and CO by making use of independent measurements (HALOE, MIPAS, MkIV, Kiruna, and other datasets.) Next, she performed retrievals on simulated spectra in order to determine optimal prior covariances for the retrievals of O_3, and determined that a diagonal prior covariance with a vertical correlation length between adjacent layers of 4 km leads to O_3 partial column retrievals with a good compromise between retrieval error, accuracy, and information content. This approach – using a 4 km vertical correlation length but different diagonal covariance elements – was also applied to the retrievals of NO, NO_2, CH_4, N_2O, HCl and HF (in the case of CH_4 and N_2O, the correlation was only applied between the altitudes of 0-15 km).

Next, the Candidate wrote computer code to perform a quantitative averaging kernel and error analysis of a large number of retrieval results, given weighting function outputs of SFIT-2. She used this capability in retrievals of HCl, O_3 and N_2O, which served as the basis for the NDACC Steering Committee's decision to grant TAO's application to become a Complementary NDACC observation station in March, 2004 [*Wiacek et al.*, 2006b]. The above retrievals were compared to a set of control analyses produced from the same spectra by scientists at the University of Wollongong, Australia, and the differences between the total column amounts retrieved by these different investigators spurred a more systematic investigation of the effect of the vertical layering scheme in retrievals. It was shown that the effect of several different layering schemes on the retrieved vertical profile is less than 5% for N_2O and HF, and less than 2% for NO, NO_2, CH_4, HCl and HF; however, retrieved profiles of H_2O showed differences of up to 13% when the layering scheme was varied.

Finally, the Candidate applied the retrieval and characterization codes to a three-year time series of O_3, NO, NO_2, HCl, HF, N_2O and CH_4, measured in five NDACC filters.

Retrieved tropospheric, stratospheric and total column concentrations were submitted to the NDACC database (www.ndsc.ncep.noaa.gov) for long-term monitoring and trend detection purposes, after designing and applying measures to ensure that only the highest quality retrievals were selected for archiving. Trace gas time series derived from TAO spectra by the Candidate exhibit the known seasonal cycles, providing a partial validation of the measurements. A quantitative retrieval characterization was performed whereby the averaging kernels, weighting functions, measurement, smoothing and temperature errors of the retrievals of these seven gases were examined in detail. In the course of this analysis, it was confirmed that TAO spectra are not sensitive to the tropospheric concentrations of stratospheric species (O_3, NO, NO_2, HF, and to a lesser extent, HCl); furthermore a relatively strong sensitivity to the stratospheric concentration of tropospheric species (CH_4 and N_2O) was also confirmed. The degrees of freedom for signal ranged from 1.10 for NO_2 (1.07 between 15-50 km) to 4.12 for N_2O (2.71 between 0-15 km). The error estimates are conservative, given that the *ad hoc* prior covariance was used in the calculations.

Finally, the Candidate demonstrated the capabilities of the FTIR observations for studying different regions of the atmosphere. She defined and documented a limited sensitivity of ground-based measurements to the Mesospheric-Lower Thermospheric concentrations of NO, performing a retrieval characterization on spectra recorded at TAO, and also on higher-SNR spectra from the Arctic NDACC station Eureka, where enhancements of NO were also expected [*Wiacek et al.*, 2006a]. Next, a case study of anomalously high stratospheric HCl, HF and O_3 concentrations was examined more closely and it was confirmed that these concentrations were rooted in the transient passage of a filament of polar air over Toronto. Lastly, abundances of the tropospheric trace gas CO were retrieved for the period of January 2002 – September 2003 and compared in a monthly mean sense to spatially coincident measurements by the MOPITT satellite instrument. Excellent agreement was found between the two observation platforms, again demonstrating the quality and usefulness of the growing TAO dataset.

7.2 Recommendations for Future Work

Instrument and Measurements

1. Maintaining the FTS in a well-aligned and well-characterized state by performing regular inspections of the optics and analyzing low-pressure gas cell spectra with both LINEFIT and SFIT-2.

2. Continuing to extend the spectral record and submitting further total, tropospheric and stratospheric column data to the NDACC database.

3. Recording (optimal) Filter 6 spectra in order to obtain more sensitive measurements of O_3, and the only feasible IR measurements of HNO_3, $ClONO_2$, SF_6, and CFCs.

4. Modifying the suntracker design to perform active solar tracking on a small quadrant sensor placed in the center of the converging solar beam (following the approach used at the University of Wollongong).

5. Replacing the near-obsolete PCDA acquisition software with an integrated system that allows for sophisticated measurement automation with built-in quality assurance steps, such as examining raw interferograms prior to co-adding.

Retrieval Algorithm

1. Expanding the range of species retrieved routinely at TAO (e.g. C_2H_6, C_2H_2, HCN, OCS, SO_2, NH_3, HCHO) and constructing reasonable prior constraints for these new retrievals from independent information.

2. Re-examining the rejection criteria used to balance automation vs. manual intervention in operational retrievals, e.g. assessing the quality of spectra more stringently prior to retrievals.

3. Completing the retrieval error budget, with a particular focus on the effect of the ILS on profile retrievals, but also calculating retrieval sensitivity to SZA and spectroscopic parameters.

4. Investigating the use of daily *a priori* profiles for water vapour from NCEP analyses in order to improve the retrievals of neighbouring gases; a related task is to investigate the use of *a priori* profiles and covariances defined on equivalent latitudes and potential temperature surfaces (for all gases, not just water vapour).

5. Further developing SFIT-2 in coordination with NDACC, in particular, investigating the root cause of the d_s discrepancy between measured and simulated spectra.

Interpretation of Retrievals

1. Completing the ongoing comparison of TAO data with forced, climatological, and data-assimilated runs of a GCM (CMAM) in order to study mid-latitude O_3 and greenhouse gas trends, as well as changing transport patterns, which are revealed in tracer-tracer correlations, such as CH_4-N_2O.

2. Continuing ongoing satellite validation activities. In particular, extending the 2.5-year time series of CO retrievals and performing a rigorous comparison to MOPITT retrievals, possibly incorporating additional satellite measurements, e.g. SCIAMACHY.

3. Examining the TAO record for further evidence of polar intrusions (e.g. stratospheric CH_4 and N_2O signatures) and obtaining alerts of future events using the MIMOSA system currently in place for Observatoire Haute Provence, if possible. Estimating the percent of polar intrusions detected with the relatively sparse temporal coverage of solar absorption measurements by comparing measurements and high-resolution model calculations of dynamical tracers.

4. Combining TAO data with a regional air quality model or a CTM (e.g. GEOS-CHEM) and back trajectory calculations to study urban air pollution at TAO's unique location within NDACC. Comparisons of CH_4 retrievals with the nearby Egbert FTS are ongoing, however, the present investigation is focused on differences introduced by the retrieval method and assumes constant atmospheric conditions at both sites; back trajectory calculations are necessary to confirm whether Egbert and Toronto indeed experience similar airmasses, and whether the influence of TAO's urban location can be detected.

5. Using TAO measurements in combination with Brewer data and our ground-level O_3 monitor to discern the tropospheric O_3 burden.

6. Exploring the feasibility and usefulness of stable C, N, O, and H isotope measurements.

Bibliography

Backus, G. E., and J. F. Gilbert, Uniqueness in the inversion of inaccurate gross earth data, *Phil. Trans. R. Soc. Lond.*, *A266*, 123–192, 1970.

Banwell, C. N., and E. M. McCash, *Fundamentals of Molecular Spectroscopy*, 4th ed., McGraw-Hill Book Company, Toronto, 1994.

Barret, B., Inversion et caractérisation de profils de constituants atmosphériques à partir de mesures FTIR sol, Ph.D. thesis, Université Libre de Bruxelles, 2003.

Barret, B., M. DeMazière, and P. Demoulin, Retrieval and characterization of ozone profiles from solar infrared spectra at the Jungfraujoch, *J. Geophys. Res.*, *107 (D24)*, 4788, doi:10.1029/2001JD001298, 2002.

Barret, B., M. DeMazière, and E. Mahieu, Ground-based FTIR measurements of CO from the Jungfraujoch: characterization and comparison with in situ surface and MOPITT data, *Atmos. Chem. Phys.*, *3*, 2217–2223, 2003.

Barret, B., D. Hurtmans, M. R. Carleer, M. DeMazière, E. Mahieu, and P.-F. Coheur, Line narrowing effect on the retrieval of HF and HCl vertical profiles from ground-based FTIR measurements, *J. Quant. Spectrosc.*, *95*, 499–519, 2005.

Barth, C. A., K. D. Mankoff, S. M. Bailey, and S. C. Solomon, Global observations of nitric oxide in the thermosphere, *J. Geophys. Res.*, *108 (A1)*, 1027, doi: 10.1029/2002JA009458, 2003.

Bassford, M. R., C. A. McLinden, and K. Strong, Zenith-Sky Observations of Stratospheric Gases: The Sensitivity of Air Mass Factors to Geophysical Parameters and the Influence of Tropospheric Clouds, *J. Quant. Spectrosc. Radiat. Transfer*, *68*, 657–677, 2001.

Bassford, M. R., K. Strong, C. A. McLinden, and C. T. McElroy, Ground-based measurements of ozone and NO_2 during MANTRA 1998 using a Zenith-sky spectrometer, *Atmosphere-Ocean*, *43 (4)*, 325–338, 2005.

Bell, R. J., *Introductory Fourier Transform Spectroscopy*, Academic Press, New York, 1972.

Bernardo, C., and D. W. T. Griffith, Fourier transform spectrometer instrument lineshape (ILS) retrieval by Fourier deconvolution, *J. Quant. Spectrosc. Radiat. Transfer*, *95*, 141–150, 2005.

Bernath, P. F., et al., Atmospheric Chemistry Experiment (ACE): Mission overview, *Geophys. Res. Lett.*, *32*, L15S01, doi:10.1029/2005GL022386, 2005.

Bey, I., D. J. Jacob, R. M. Yantosca, J. A. Logan, B. D. Field, A. M. Fiore, Q. Li, H. Y. Liu, L. J. Mickley, and M. G. Schultz, Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, *J. Geophys. Res.*, *106 (D19)*, 23,073–23,096, doi:10.1029/2001JD000807, 2001.

Blumenstock, T., G. Kopp, F. Hase, G. Hochschild, S. Mikuteit, U. Raffalski, and R. Ruhnke, Observation of unusual chlorine activation by ground-based infrared and microwave spectroscopy in the late Arctic winter 2000/01, *Atmos. Chem. Phys.*, *6*, 897–905, 2006.

Bovensmann, H., J. P. Burrows, M. Buchwitz, J. Frerick, S. Noël, V. V. Rozanov, K. V. Chance, and A. P. H. Goede, SCIAMACHY: Mission Objectives and Measurement Modes, *J. Atmos. Sci.*, *56 (2)*, 127–150, 1999.

Bracewell, R., *The Fourier Transform and Its Applications*, 3rd ed., McGraw-Hill, New York, 1999.

Brasseur, G., and S. Solomon, *Aeronomy of the Middle Atmosphere*, 2nd ed., D. Reidel Publishing Co., Boston, 1986.

Brault, J. W., New approach to high-precision fourier transform spectrometer design, *Appl. Opt.*, *35 (16)*, 2891–2896, 1996.

Callis, L. B., D. N. Baker, M. Natarajan, J. B. Blake, R. A. Mewaldt, R. S. Selesnick, and J. R. Cummings, A 2-D model simulation of downward transport of NO_y into the stratosphere: Effects on the 1994 austral spring O_3 and NO_y, *Geophys. Res. Lett.*, *23 (15)*, 1905–1908, 1996.

Carli, B., et al., First results of MIPAS/ENVISAT with operational Level 2 code., *Adv. Space Res.*, *33*, 1012–1019, doi:10.1016/S0273-1177(03)00584-2, 2004.

Chamberlain, J., *The Principles of Interferometric Spectroscopy*, John Wiley and Sons, Toronto, 1979.

Clerbaux, C., P.-F. Coheur, D. Hurtmans, B. Barret, M. Carleer, R. Colin, K. Semeniuk, J. C. McConnell, C. Boone, and P. Bernath, Carbon monoxide distribution from the ACE-FTS solar occultation measurements, *Geophys. Res. Lett.*, *32*, L16S01, doi: 10.1029/2005GL022394, 2005.

Coffey, M. T., A. Goldman, J. W. Hannigan, W. G. Mankin, W. G. Schoenfeld, C. P. Rinsland, C. Bernardo, and D. W. T. Griffith, Improved vibration-rotation (0-1) HBr line parameters for validating high resolution infrared atmospheric spectra measurements, *J. Quant. Spectrosc. Radiat. Transfer*, *60*, 863–867, 1998.

Coffey, M. T., J. W. Hannigan, W. G. Mankin, A. Goldman, and T. M. Stephen, Hbr reference cells for the infrared spectrometers of the Network for the Detection of Stratospheric Change, NDSC Science Symposium, Arcachon, France, 2001.

Connor, B. J., N. B. Jones, S. W. Wood, J. G. Keys, C. P. Rinsland, and F. J. Murcray, Retrieval of HCl and HNO_3 Profiles from Ground-based FTIR Data Using SFIT2, Paper presented at the XVIIIth Quadrennial Ozone Symposium, L'Aquila, Italy, Int. Ozone Commission, 485-488, 1996.

Conrath, B. J., Vertical resolution of temperature profiles obtained from remote radiation measurements, *J. Atmos. Sci.*, *29*, 1262–1271, 1972.

Cooley, J. W., and O. W. Tukey, An Algorithm for the Machine Calculation of Complex Fourier Series, *Math. Comput.*, *19*, 297–301, 1965.

Davis, S. P., M. C. Abrams, and J. W. Brault, *Fourier Transform Spectrometry*, Academic Press, New York, 2001.

Deeter, M. N., L. K. Emmons, and J. C. Gille, Vertical resolution and information content of CO profiles retrieved by MOPITT, *Geophys. Res. Lett.*, *31*, L15112, doi: 10.1029/2004GL020235, 2004a.

Deeter, M. N., et al., Operational carbon monoxide retrieval algorithm and selected results for the MOPITT instrument, *J. Geophys. Res.*, *108 (D14)*, 4399, doi: 10.1029/2002JD003186, 2003.

Deeter, M. N., et al., Evaluation of operational radiances for the Measurements of Pollution in the Troposphere (MOPITT) instrument CO thermal band channels, *J. Geophys. Res.*, *109*, D03308, doi:10.1029/2003JD003970, 2004b.

Dicke, R., The Effect of Collisions upon the Doppler Width of Spectral Lines, *Phys. Rev.*, *89*, 472–473, 1953.

Dils, B., et al., Comparisons between SCIAMACHY and ground-based FTIR data for total columns of CO, CH_4, CO_2 and n_2o, *Atmos. Chem. Phys.*, *6*, 1953–1976, 2006.

Donovan, D. P., et al., Ozone, Column ClO, and PSC measurements made at the NDSC Eureka observatory (80°N, 86°W) during the spring of 1997, *Geophys. Res. Lett.*, *24 (22)*, 27092712, 1997.

Drummond, J. R., and G. S. Mand, The measurements of pollution in the troposphere (MOPITT) instrument: Overall performance and calibration requirements, *J. Atmos. Oceanic Technol.*, *13*, 314–320, 1996.

Durry, G., and A. Hauchecorne, Evidence for long-lived polar vortex air in the mid-latitude summer stratosphere from in situ laser diode CH_4 and H_2O measurements, *Atmos. Chem. Phys.*, *5*, 14671472, 2005.

Emmons, L. K., et al., Validation of Measurements of Pollution in the Troposphere (MOPITT) CO retrievals with aircraft in situ profiles, *J. Geophys. Res.*, *109*, D03309, doi:10.1029/2003JD004101, 2004.

Eriksson, P., Analysis and comparison of two linear regularization methods for passive atmospheric observations, *J. Geophys. Res.*, *105 (D14)*, 18,157–18,167, 2000.

Fellgett, P., A propos de la théorie du spectromètre interférentiel multiplex, *J. Phys. Radium*, *19*, 187–191, 1958.

Fischer, J., R. R. Gamache, A. Goldman, L. S. Rothman, and A. Perrin, Total internal partition sums for molecular species in the 2000 edition of the HITRAN database, *J. Quant. Spectrosc. Radiat Transfer*, *82*, 401–412, 2003.

Forkman, P., P. Eriksson, A. Winnberg, R. R. Garcia, and D. Kinnison, Longest continuous ground-based measurements of mesospheric CO, *Geophys. Res. Lett.*, *30 (10)*, 1532, doi:10.1029/2003GL016931, 2003.

Forman, M. L., W. H. Steel, and G. A. Vanasse, Correction of Asymmetric Interferograms Obtained in Fourier Spectroscopy, *J. Opt. Soc. Am.*, *56*, 59–63, 1966.

Gallery, W. O., F. X. Kneizys, and S. A. Clough, Air Mass Computer Program for Atmospheric Transmittance/Radiance Calculation: FSCATM, *Environmental Research Papers No. 828 (AFGL-TR-83-0065)*, Air Force Geophysics Laboratory, 1983.

Godin, S., M. Marchand, A. Hauchecorne, and F. Lefèvre, Influence of Arctic polar ozone depletion on lower stratospheric ozone amounts at Haute-Provence Observatory (43.92°N, 5.71°E), *J. Geophys. Res.*, *107 (D20)*, 8272, doi:10.1029/2001JD000516, 2002.

Goldman, A., et al., Network for the Detection of Stratospheric Change Fourier transform infrared intercomparison at Table Mountain Facility, November 1996, *J. Geophys. Res.*, *104 (D23)*, 30,481–30,503, 1999.

Goorvitch, D., Calculation of instrument functions, *J. Quant. Spectrosc. Radiat. Transfer*, *67*, 253–257, 2000.

Griffith, D. W. T., N. B. Jones, B. McNamara, C. Paton-Walsh, W. Bell, and C. Bernardo, Intercomparison of NDSC Ground-Based Solar FTIR Measurements of Atmospheric Gases at Lauder, New Zealand, *J. Atmos. Oceanic Technol.*, *20*, 1138–1153, 2003.

Griffiths, P. R., and J. A. de Haseth, *Fourier Transform Infrared Spectrometry*, John Wiley and Sons, Toronto, 1986.

Hase, F., T. Blumenstock, and C. Paton-Walsh, Analysis of the instrumental line shape of high-resolution Fourier transform IR spectrometers with gas cell measurements and new retrieval software, *Appl. Opt.*, *38*, 3417–3422, 1999.

Hase, F., J. W. Hannigan, M. T. Coffey, A. Goldman, M. Höpfner, N. B. Jones, C. P. Rinsland, and S. W. Wood, Intercomparison of retrieval codes used for the analysis of high-resolution, ground-based FTIR measurements, *J. Quant. Spectrosc. Radiat. Transfer*, *87*, 25–52, 2004.

Hauchecorne, A., S. Godin, M. Marchand, B. Heese, and C. Souprayen, Quantification of the transport of chemical constituents from the polar vortex to midlatitudes in the lower stratosphere using the high-resolution advection model MIMOSA and effective diffusivity, *J. Geophys. Res.*, *107 (D20)*, 8289, doi:10.1029/2001JD000491, 2002.

Heald, C. L., et al., Asian outflow and trans-Pacific transport of carbon monoxide and ozone pollution: An integrated satellite, aircraft, and model perspective, *J. Geophys. Res.*, *108 (D24)*, 4804, doi:10.1029/2003JD003507, 2003.

Houghton, J., *The Physics of Atmospheres*, 3rd ed., Cambridge University Press, New York, 2002.

IPCC, *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, New York, 2001.

Jackman, C. H., M. T. DeLand, G. J. Labow, E. L. Fleming, D. K. Weisenstein, M. K. W. Ko, M. Sinnhuber, J. Anderson, and J. M. Russell, The influence of the several very large solar proton events in years 2000-2003 on the neutral middle atmosphere, *Adv. Space Res.*, *35*, 445–450, doi:10.1016/j.asr.2004.09.006, 2005.

Jacob, D. J., *Introduction to Atmospheric Chemistry*, Princeton University Press, Princeton, New Jersey, 1999.

Kasai, Y. J., T. Koshiro, M. Endo, N. B. Jones, and Y. Murayama, Ground-based measurement of strato-mesospheric CO by a FTIR spectrometer over Poker Flat, Alaska, *Adv. Space Res.*, *35*, 2024–2030, doi:10.1016/j.asr.2005.04.099, 2005.

Kerzenmacher, T. E., et al., Measurements of O_3, NO_2 and Temperature during the 2004 Canadian Arctic ACE Validation Campaign, *Geophys. Res. Lett.*, *32*, L16S07, doi:10.1029/2005GL023032, 2005.

Kidder, S. Q., and T. H. VonderHaar, *Satellite Meteorology*, Academic Press, Toronto, 1995.

Kurylo, M. J., and R. J. Zander, The NDSC – Its status after ten years of operation, Paper presented at the XIXth Quadrennial Ozone Symposium, Sapporo, Japan, Hokkaido Univ., 2000.

Li, Q., et al., Transatlantic transport of pollution and its effects on surface ozone in Europe and North America, *J. Geophys. Res.*, *107 (D13)*, 4166, doi: 10.1029/2001JD001422, 2002.

Mahieu, E., et al., Comparisons between ACE-FTS and ground-based measurements of stratospheric HCl and $ClONO_2$ loadings at northern latitudes, *Geophys. Res. Lett.*, *32*, L15S08, doi:10.1029/2005GL022396, 2005.

Manney, G. L., K. Krüger, J. L. Sabutis, S. A. Sena, and S. Pawson, The remarkable 2003-2004 winter and other recent warm winters in the Arctic stratosphere since the late 1990s, *J. Geophys. Res.*, *110*, D04107, doi:10.1029/2004JD005367, 2005.

Marchand, M., S. Godin, A. Hauchecorne, F. Lefèvre, and M. Chipperfield, Influence of polar ozone loss on northern midlatitude regions estimated by a high-resolution chemistry transport model during winter 1999/2000, *J. Geophys. Res.*, *108 (D5)*, 8326, doi:10.1029/2001JD000906, 2003.

McElroy, C. T., A spectroradiometer for the measurement of direct and scattered solar irradiance from on-board the NASA ER-2 high-altitude research aircraft, *Geophys. Res. Lett.*, *22 (11)*, 1361–1364, 1995.

Meier, A., A. Goldman, P. S. Manning, T. M. Stephen, C. P. Rinsland, N. B. Jones, and S. W. Wood, Improvements to air mass calculations for ground-based infrared measurements, *J. Quant. Spectrosc. Radiat. Transfer*, *83*, 109–113, 2004a.

Meier, A., G. C. Toon, C. P. Rinsland, A. Goldman, and F. Hase, Spectroscopic Atlas of Atmospheric Microwindows in the Middle Infra-Red, 2nd Edition, *IRF Technical Report 048 ISSN 0284-1738*, Swedish Institute of Space Physics, 2004b.

Meier, A., et al., Evidence of reduced measurement uncertainties from an FTIR instrument intercomparison at Kiruna, Sweden, *J. Quant. Spectrosc. Radiat. Transfer*, *96*, 75–84, 2005.

Mertz, L., *Transformations in Optics*, John Wiley and Sons, New York, 1965.

Mertz, L., Auxiliary Computation for Fourier Spectroscopy, *Infrared Phys.*, *7*, 17–23, 1967.

Michelson, A. A., Visibility of Interference-Fringes in the Focus of a Telescope, *Phil. Mag.*, *31*, 256–259, 1891.

Michelson, A. A., On the Application of Interference Methods to Spectroscopic Measurements, *Phil. Mag.*, *34*, 280, 1892.

Millard, G. A., A. M. Lee, and J. A. Pyle, A model study of the connection between polar and midlatitude ozone loss in the Northern Hemisphere lower stratosphere, *J. Geophys. Res.*, *108 (D5)*, 8323, doi:10.1029/2001JD000899, 2003.

Natarajan, M., E. E. Remsberg, L. E. Deaver, and J. M. R. III, Anomalously high levels of NO_x in the polar upper stratosphere during April 2004: Photochemical consistency of HALOE observations, *Geophys. Res. Lett.*, *31*, L15113, doi:10.1029/2004GL020566, 2004.

Notholt, J., A. Meier, and S. Peil, Total Column Densities of Tropospheric and Stratospheric Trace Gases in the Undisturbed Arctic Summer Atmosphere, *J. Atmos. Chem.*, *20 (3)*, 311–332, 1995.

Notholt, J., G. Toon, N. Jones, D. Griffith, and T. Warneke, Spectral line finding program for atmospheric remote sensing using full radiation transfer, *J. Quant. Spectrosc. Radiat. Transfer*, *97*, 112–125, 2006.

Notholt, J., et al., Enhanced upper tropical tropospheric COS: Impact on the stratospheric aerosol layer, *Science*, *300 (5617)*, 307–310, 2003.

Orsolini, Y. J., G. L. Manney, M. L. Santee, and C. E. Randall, An upper stratospheric layer of enhanced HNO_3 following exceptional solar storms, *Geophys. Res. Lett.*, *32*, L12S01, doi:10.1029/2004GL021588, 2005.

Park, J. H., Effect of interferogram smearing on atmospheric limb sounding by Fourier transform spectroscopy, *Appl. Opt.*, *21*, 1356–1366, 1982.

Paton-Walsh, C., N. B. Jones, S. Wilson, A. Meier, N. Deutscher, D. W. T. Griffith, R. Mitchell, and S. Campbell, Trace gas emissions from biomass burning inferred from aerosol optical depth, *Geophys. Res. Lett.*, *31*, L05116, doi:10.1029/2003GL018973, 2004.

Paton-Walsh, C., et al., An uncertainty budget for ground-based Fourier transform infrared column measurements of HCl, HF, N_2O, and HNO_3 deduced from results of side-by-side instrument intercomparisons, *J. Geophys. Res.*, *102 (D7)*, 8867–8873, 1997.

Peter, T., Microphysics and heterogeneous chemistry of polar stratospheric clouds, *Annu. Rev. Phys. Chem.*, *48*, 785–822, 1997.

Pougatchev, N. S., B. J. Connor, and C. P. Rinsland, Infrared measurements of the ozone vertical distribution above Kitt Peak, *J. Geophys. Res.*, *100 (D8)*, 16,689–16,697, 1995.

Quine, B. M., and J. R. Drummond, GENSPECT: a line-by-line code with selectable interpolation error tolerance, *J. Quant. Spectrosc. Radiat. Transfer*, *74*, 147–165, 2002.

Randall, C. E., et al., Stratospheric effects of energetic particle precipitation in 2003-2004, *Geophys. Res. Lett.*, *32*, L05802, doi:10.1029/2004GL022003, 2005.

Remedios, J. J., Spectroscopy for Remote Sounding of the Atmosphere, Ph.D. thesis, University of Oxford, 1990.

Rinsland, C. P., A. Meier, D. W. T. Griffith, and L. S. Chiou, Ground-based measurements of tropospheric CO, C_2H_6, and HCN from Australia at 34°S latitude during 1997-1998, *J. Geophys. Res.*, *106 (D18)*, 20,913–20,924, doi:10.1029/2000JD000318, 2001.

Rinsland, C. P., R. Zander, E. Mahieu, L. S. Chiou, A. Goldman, and N. B. Jones, Stratospheric HF column abundances above Kitt Peak (31.9°N latitude): trends from 1977 to 2001 and correlations with stratospheric HCl columns, *J. Quant. Spectrosc. Radiat. Transfer*, *74*, 205–216, 2002a.

Rinsland, C. P., C. Boone, R. Nassar, K. Walker, P. Bernath, J. C. McConnell, and L. Chiou, Atmospheric Chemistry Experiment (ACE) Arctic stratospheric measurements of NO_x during February and March 2004: Impact of intense solar flares, *Geophys. Res. Lett.*, *32*, L16S05, doi:10.1029/2005GL022425, 2005.

Rinsland, C. P., et al., Northern and southern hemisphere ground-based infrared spectroscopic measurements of tropospheric carbon monoxide and ethane, *J. Geophys. Res.*, *103 (D21)*, 28,197–28,217, 1998.

Rinsland, C. P., et al., Ground-based infrared spectroscopic measurements of carbonyl sulfide: Free troposphere trends from a 24-year time series of solar absorption measurements, *J. Geophys. Res.*, *107 (D22)*, 4657, doi:10.1029/2002JD002522, 2002b.

Rinsland, C. P., et al., Multiyear infrared solar spectroscopic measurements of HCN, CO, C_2H_6, and C_2H_2 tropospheric columns above Lauder, New Zealand (45°S Latitude), *J. Geophys. Res.*, *107 (D14)*, 4185, doi:10.1029/2001JD001150, 2002c.

Rinsland, C. P., et al., Long-term trends of inorganic chlorine from ground-based infrared solar spectra: Past increases and evidence for stabilization, *J. Geophys. Res.*, *108 (D8)*, 4252, doi:10.1029/2002JD003001, 2003.

Rodgers, C. D., Retrieval of Atmospheric Temperature and Composition From Remote Measurements of Thermal Radiation, *Rev. Geophys. Space Phys.*, *14*, 609–624, 1976.

Rodgers, C. D., Characterization and Error Analysis of Profiles Retrieved From Remote Sounding Measurements, *J. Geophys. Res.*, *95*, 5587–5595, 1990.

Rodgers, C. D., *Inverse Methods for Atmospheric Sounding: Theory and Practice*, World Scientific Publishing Co. Pte. Ltd., New Jersey, 2000.

Rodgers, C. D., and B. J. Connor, Intercomparison of remote sounding instruments, *J. Geophys. Res.*, *108 (D3)*, 4116, doi:10.1029/2002JD002299, 2003.

Rothman, L. S., et al., The HITRAN molecular spectroscopic database and HAWKS (HITRAN Atmospheric Workstation): 1996 edition, *J. Quant. Spectrosc. Radiat. Transfer*, *60*, 665–710, 1998.

Rothman, L. S., et al., The HITRAN molecular spectroscopic database: edition of 2000 including updates through 2001, *J. Quant. Spectrosc. Radiat. Transfer*, *82*, 5–44, 2003.

Rothman, L. S., et al., The HITRAN 2004 molecular spectroscopic database, *J. Quant. Spectrosc. Radiat. Transfer*, *96*, 139–204, 2005.

Rozanov, E., L. Callis, M. Schlesinger, F. Yang, N. Andronova, and V. Zubov, Atmospheric response to NO_y source due to energetic electron precipitation, *Geophys. Res. Lett.*, *32*, L14811, doi:10.1029/2005GL023041, 2005.

Russell, I. J. M., et al., The halogen occultation experiment, *J. Geophys. Res.*, *98 (D6)*, 10,777–10,797, 1993.

Schneider, M., T. Blumenstock, M. P. Chipperfield, F. Hase, W. Kouker, T. Reddmann, R. Ruhnke, E. Cuevas, and H. Fischer, Subtropical trace gas profiles determined by ground-based FTIR spectroscopy at Izaña (28°N, 16°W): Five-year record, error analysis, and comparison with 3-D CTMs, *Atmos. Chem. Phys.*, *5*, 153–167, 2005a.

Schneider, M., T. Blumenstock, F. Hase, M. Höpfner, E. Cuevas, A. Redondas, and J. M. Sancho, Ozone profiles and total column amounts derived at Izaña, Tenerife Island, from FTIR solar absorption spectra, and its validation by an intercomparison to ECC-sonde and Brewer spectrometer measurements, *J. Quant. Spectrosc. Radiat. Transfer*, *91*, 245–274, 2005b.

Semeniuk, K., J. C. McConnell, and C. H. Jackman, Simulation of the October-November 2003 solar proton events in the CMAM GCM: Comparison with observations, *Geophys. Res. Lett.*, *32*, L15S02, doi:10.1029/2005GL022392, 2005.

Seppälä, A., P. T. Verronen, E. Kyrölä, S. Hassinen, L. Backman, A. Hauchecorne, J. L. Bertaux, and D. Fussen, Solar proton events of October-November 2003: Ozone depletion in the Northern Hemisphere polar winter as seen by GOMOS/Envisat, *Geophys. Res. Lett.*, *31*, L19107, doi:10.1029/2004GL021042, 2004.

Shepherd, T. G., Large-scale atmospheric dynamics for atmospheric chemists, *Chem. Rev.*, *103*, 4509–4531, 2003.

Solomon, S., Stratospheric Ozone Depletion: A Review of Concepts and History, *Rev. Geophys.*, *37 (3)*, 275316, 1999.

Steck, T., Methods for determining regularization for atmospheric retrieval problems, *Appl. Opt.*, *41*, 1788–1797, 2002.

Stephens, G. L., *Remote Sensing of the Lower Atmosphere: An Introduction*, Oxford University Press, New York, 1994.

Sussmann, R., and M. Buchwitz, Initial validation of ENVISAT/SCIAMACHY columnar CO by FTIR profile retrievals at the Ground-Truthing Station Zugspitze, *Atmos. Chem. Phys.*, *5*, 14971503, 2005.

Velazco, V., J. Notholt, T. Warneke, M. Lawrence, H. Bremer, J. R. Drummond, A. Schulz, J. Krieg, and O. Schrems, Latitude and altitude variability of carbon monoxide in the Atlantic detected from ship-borne Fourier transform spectrometry, model, and satellite data, *J. Geophys. Res.*, *110*, D09306, doi:10.1029/2004JD005351, 2005.

Washenfelder, R. A., P. O. Wennberg, and G. C. Toon, Tropospheric methane retrieved from ground-based near-IR solar absorption spectra, *Geophys. Res. Lett.*, *30 (23)*, 2226, doi:10.1029/2003GL017969, 2003.

Wiacek, A., N. B. Jones, K. Strong, J. R. Taylor, R. L. Mittermeier, and H. Fast, First detection of meso-thermospheric Nitric Oxide (NO) by ground-based FTIR solar absorption spectroscopy, *Geophys. Res. Lett.*, *33*, L03811, doi:10.1029/2005GL024897, 2006a.

Wiacek, A., J. R. Taylor, K. Strong, R. Saari, T. E. Kerzenmacher, N. B. Jones, and D. W. T. Griffith, Ground-based solar absorption FTIR spectroscopy: a novel optical design instrument at a new NDSC Complementary Station, characterization of retrievals and first results, *J. Atmos. Oceanic Technol.*, *in press*, 2006b.

WMO, Scientific Assessment of Ozone Depletion: 2002, *Global Ozone Research and Monitoring Project – Report No. 47*, World Meteorological Organization, 2003.

Wood, S. W., Error Group Summary, NDACC InfraRed Working Group, Queenstown, New Zealand, 2004.

Wood, S. W., R. L. Batchelor, A. Goldman, C. P. Rinsland, B. J. Connor, F. J. Murcray, T. M. Stephen, and D. N. Heuff, Ground-based nitric acid measurements at Arrival Heights, Antarctica, using solar and lunar Fourier transform infrared observations, *J. Geophys. Res.*, *109 (D18)*, D18307, doi:10.1029/2004JD004665, 2004.

Woods, T. N., F. G. Eparvier, J. Fontenla, J. Harder, G. Kopp, W. E. McClintock, G. Rottman, B. Smiley, and M. Snow, Solar irradiance variability during the October 2003 solar storm period, *Geophys. Res. Lett.*, *31*, L10802, doi:10.1029/2004GL019571, 2004.

Yurganov, L. N., et al., Increased Northern Hemispheric carbon monoxide burden in the troposphere in 2002 and 2003 detected from the ground and from space, *Atmos. Chem. Phys.*, *5*, 563–573, 2005.

Zander, R., et al., Secular trend and seasonal variability of the column abundance of N_2O above the Jungfraujoch station determined from IR solar spectra, *J. Geophys. Res.*, *99 (D8)*, 16,745–16,756, 1994.

Zhao, Y., et al., Spectroscopic measurements of tropospheric CO, C_2H_6, C_2H_2, and HCN in northern Japan, *J. Geophys. Res.*, *107 (D18)*, 4343, doi:10.1029/2001JD000748, 2002.