SPECTROSCOPIC MEASUREMENTS OF TROPOSPHERIC HCN AND C_2H_6 OVER ADDIS ABABA

By

AMBACHEW ABEJE ALEMU

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN PHYSICS
AT
ADDIS ABABA UNIVERSITY
ADDIS ABABA, ETHIOPIA
JUNE 2011

© Copyright by AMBACHEW ABEJE ALEMU, 2011

ADDIS ABABA UNIVERSITY

DEPARTMENT OF

PHYSICS

Supervisor: _____

DR. GIZAW MENGISTU

Examiners: _____

PROF. A.V.GHOLAP

_____

DR. ELIAS LEWI

ADDIS ABABA UNIVERSITY

Date: **JUNE 2011**

Author: **AMBACHEW ABEJE ALEMU**

Title: **Spectroscopic Measurements of Tropospheric HCN and C_2H_6 over Addis Ababa**

Department: **Physics**

Degree: **M.Sc.** Convocation: **JUNE** Year: **2011**

Permission is herewith granted to Addis Ababa University to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions.

Signature of Author

THE AUTHOR RESERVES OTHER PUBLICATION RIGHTS, AND NEITHER THE THESIS NOR EXTENSIVE EXTRACTS FROM IT MAY BE PRINTED OR OTHERWISE REPRODUCED WITHOUT THE AUTHOR'S WRITTEN PERMISSION.

THE AUTHOR ATTESTS THAT PERMISSION HAS BEEN OBTAINED FOR THE USE OF ANY COPYRIGHTED MATERIAL APPEARING IN THIS THESIS (OTHER THAN BRIEF EXCERPTS REQUIRING ONLY PROPER ACKNOWLEDGEMENT IN SCHOLARLY WRITING) AND THAT ALL SUCH USE IS CLEARLY ACKNOWLEDGED.

For My Brother Melese Abeje and My Father.

Table of Contents

List of Figures

List of Tables

Abstract

The atmospheric concentrations of HCN and C_2H_6 are important indicators of tropospheric pollution and transport because they are primarily emitted from biomass burning and anthropogenic sources. Tropospheric column amounts and mixing ratios of C_2H_6 and HCN were retrieved from ground-based infrared solar spectra using a vertical profile retrieval algorithm (PROFFIT). The spectra were recorded with high spectral resolution Fourier transform infrared (FTIR) spectrometer at Addis Ababa (9.01^0 N, 38.76^0 E and 2.45 km asl), Ethiopia from May 2009 to February 2011. The selected microwindows are found to be good with a maximum residual error of less than 0.1% for HCN and less than 3.5% for C_2H_6 retrievals. We found that the degrees of freedom for HCN is 2 and 1.5 for C_2H_6 with a total statistical error 38.78% and 17.74% and a total systematic error 37.33% and 12.04% respectively. HCN is more sensitive in the upper atmosphere while C_2H_6 is sensitive in the lower atmosphere. The total column amount of HCN ranges from 9.83×10^{18} to 6.39×10^{19} while of C_2H_6 from 3.60×10^{19} to 1.70×10^{20} molecules/m^2. The maximum volume mixing ratio for both gases was found in March.

Acknowledgements

Above all, I would like to thank the almighty God for letting me accomplish this stage.
I am deeply indebted to express my sincere thanks to my instructor as well as advisor of this thesis Dr.Gizaw Mengistu for his guidance, assistance, supervision and contribution of valuable suggestions. His scientific excitement, integral view on research and overly enthusiasm has made a deep impression on me.
My infinite thanks goes to Samuael Takele, Endale Gemechu and Friewelega Geleta for their encouragement and tirless guidance in showing how to use all the softwares that were important for this research.
I also thank Mr. Eyale Bayable for his comment and encouragement from the start.
All atmospheric physics and space physics postgraduate students: Alemayehu Girma, Abebe Kebede, Temesgen Gebremariam, Jemal Seid, Tamene Mekonen, Gebreab Kidanu and Ephrem Tesfaye and there are other friends whose name is not mentioned here are very acknowledged for their support in idea, and computational and technical helps during this study.

Introduction

The atmospheric concentrations of Hydrogen Cyanide (HCN) and Ethane (C_2H_6) are important indicators of tropospheric pollution and transport because they are primarily emitted from biomass burning and anthropogenic sources. C_2H_6 is precursor of tropospheric ozone (O_3) such that the budget of O_3 could be affected by changes in its tropospheric concentrations. Since HCN is a relatively inactive species, it is assumed to be a potential tracer of biomass burning emission. Thus, the issue of the resulting environmental pollution due to increases in the atmospheric concentrations of these trace gases and others with anthropogenic origin had been of great concern. Hence, we are interested in their distributions because continuous increase in their concentrations may adversely alter air quality, affect composition and chemistry of the atmosphere which can lead to health problems and long-term climate change. It has been observed that reactions with hydroxyl radicals (OH) are primary sinks of C_2H_6, while ocean uptake and reactions with $O(^1D)$ are part of the suggested sinks of HCN [1, 2]. Air pollution originates mainly from the burning of fuels such as petroleum products, ignite and coal, which are collectively classified as modern or commercial energy where as those from the burning of fuel wood, paddy husk, and charcoal are called traditional energy [3]. Human activities, like agricultural, industrial, transport and urbanisation activities, but also biomass burning (deforestation), lead directly to changes in the biogeochemical cycles, thereby changing the composition of the atmosphere (e.g. greenhouse gases, stratospheric ozone depleting gases, and toxic pollutants), and the chemistry of aquatic systems and soils (SO_4, NO_3, PO_4 heavy metals, pesticides and organic micro-pollutants). Evidently all these changes

differ with respect to scales in time (seasonal, annual, decadal and centurial) and space (local, regional, continental and global) [4].

Ethiopia is one of the East African countries. Its population is more than 70 millions. Majority of the people (>85%) live in rural area, where modern energy sources have not reached yet. So, this section of the population produces energy from biomass burning which results in introduction of trace gases and particulate matters like soot and ash. Ethiopia has been affected by droughts in different years which is caused by climate change. Accurate quantification of the amounts of trace gases and particulate matter emitted from vegetation fires and other sources of biomass burning (agricultural waste and biofuels) on a regional and global basis is required by a number of users, including scientists studying a wide range of atmospheric processes, national governments who are required to report greenhouse gas emissions, and those interested in quantifying the sources of air pollution that affect human health at regional scales [5].

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 scientific studies (both long-term and short-term), as well as for satellite and model validation. Most importantly, ground-based FTIR spectra recorded at International Network for the Detection of Atmospheric Composition Change (INDACC) stations have been used to study chemical and dynamical processes in the polar, mid-latitude and tropical atmosphere in the troposphere, stratosphere and mesosphere [6].

The main objective of this thesis is to determine the partial and total column amounts as well as mixing ratios (VMR) of HCN and C_2H_6 from solar spectra recorded over Ethiopia at Addis Ababa using high resolution FTIR spectrometer. Moreover, analysis of the derived HCN and C_2H_6 VMRs time series, their chemical and dynamical sources and sinks are investigated. In addition, the accuracy of derived HCN and C_2H_6 is characterized by the vertical resolution and sensitivity of the retrieval.

The thesis is divided into five chapters: In Chapter 1, we present definitions of atmospheric

tracers. We discuss about air pollution and biomass burning and its situation over Equatorial Africa generally and in Ethiopia particularly. Tropospheric biomass burning tracers such as HCN and C_2H_6 are also presented. Chapter 2 gives the measurement technique of FTIR solar absorption spectroscopy. This chapter covers infrared atmospheric spectra, Fourier transform spectrometer and the set up of FTIR spectrometer in Addis Ababa. Chapter 3 is about retrieval of trace gases which includes inverse theory, retrieval characterization, retrieval error analysis and retrieval algorithms. Results and discussion is presented in Chapter 4. And the last chapter, Chapter 5 describes the conclusion of this study.

Chapter 1

Atmospheric Tracers

Any detectable material that is conserved along trajectories may serve as a tracer. Tracer studies are extremely important in furthering our knowledge of atmospheric dispersion. These studies consist of release of a known quantity of a unique substance (the tracer), with measurements of that substance at one or more downwind sampling locations [7]. In this chapter we are going to deal with air pollution and biomass burning generally and over Equatorial Africa (EA) particularly. We also discuss biomass burning tracers (like tropospheric C_2H_6 and HCN) which are indicators of air pollution.

1.1 Air Pollution and Biomass Burning

Air pollution is due to both natural and human activities. Natural air pollution problems on the Earth are as old as the Earth itself. Volcanoes, fumaroles, natural fires, and desert dust have all contributed to natural air pollution. Anthropogenic air pollution problems have existed on urban scales for centuries and have resulted from burning of wood, vegetation, coal, oil, natural gas, waste, and chemicals. Humans first emitted air pollutants when they burned wood and cleared land (increasing windblown dust). Biomass burning (natural or anthropogenic) is a problem of long standing. Huge amounts of air pollution are produced worldwide by the annual burning of 3 billion metric tons of biomass such as wood, leaves, trees, grass and trash (Abelson). Biomass burning represents the

largest source of air pollution in many rural areas of the developed and developing world. Biomass burning is used to create heat, to clear forests, to dispose of leaves, crop stubble, trash and wood. Globally, biomass burning is estimated to produce 40% of the carbon dioxide, 32% of the carbon monoxide, 20% of the particulates, and 50% of the highly carcinogenic poly-aromatic hydrocarbons produced by all sources [8].

1.1.1 Air Pollution

Air pollution science has a shorter history as compared to, for instance, meteorological science, which is an old and established field. A condition of "air pollution"may be defined as a situation in which substances that result from natural and/or anthropogenic activities present at concentrations sufficiently high above their normal ambient levels to produce a measurable effect on humans, animals, vegetation, or materials. This definition could include any substance, whether noxious or benign; however, the implication is that the effects are undesirable. Natural air pollution has occurred on Earth since the planet's formation. Fires, volcanic eruptions, meteorite impacts, and high winds all cause natural air pollution. Anthropogenic air pollution problems have existed on urban scales for centuries and have resulted from burning of wood, vegetation, coal, oil, natural gas, waste, and chemicals. Traditionally, air pollution has been viewed as a phenomenon characteristic only of large urban centers and industrialized regions. It is now clear that dense urban centers are just "hot spots"in a continuum of trace species concentrations over the entire Earth. Both urban smogs and stratospheric ozone depletion by chlorofluorocarbons are manifestations of what might be termed in the broadest sense as air pollution [9].

1.1.2 Air Pollution over Equatorial Africa

The northern and southern countries in Africa receive little rainfall, below 1000 mm per year. The Equatorial countries receive over 1000 mm of rainfall on average. Higher speed winds are generally found in the southern and northern parts of the continent. Virtually

all countries in Africa receive over ten hours of sunshine a day, providing good potential for solar energy generation. The air in Africa, except in major industrial cities, is relatively clean.

Atmospheric resources offer a variety of opportunities for sustainable development. Air pollutants including soot and dust, greenhouse gases (GHG), chlorofluorocarbons, and heavy metals affect the quality of air and threaten the goods-and- services provided by the atmosphere. These pollutants contribute to climate change, climate variability, depletion of the ozone layer, and low air quality. Climate variability and climate change manifest themselves in global warming, and extreme weather events including floods, droughts, heat waves and typhoons. These phenomena affect human well-being through increasing the incidence of diseases, affect land and marine productive systems from which livelihoods are derived, and destroy infrastructure [10].

There are three broad sources of air pollution from human activities; stationary or point, mobile, and indoor air pollution. Industries, power plants, etc are the causes of stationary air pollution, where as indoor air pollution refers to pollutions from open fires for cooking and heating. It is mostly a problem in developing countries, and its effect become intense in rural areas. Mobile or Vehicular air pollution is particularly a serious problem in urban areas. Ethiopia is one of the countries of Equatorial Africa. Since most economic activities or transactions are concentrated in urban Ethiopia, the air pollution that resulted from these three sources is a serious problem. However, it seems that, in Ethiopia, compared to other developed countries, there is no serious problem related to air pollution. But this is only because of the fact that no body gives a serious attention to the effects of air pollution. Moreover, there is no study or reliable data regarding the effects of air pollution on the health condition of the people living in the country [11].

1.1.3 Biomass Burning

Biomass burning (BB) is the burning of evergreen forests, deciduous forests, woodlands, grassland, and agricultural land, either to clear land for other use, to stimulate grass growth, for forest management, or as a ritual. Forest fires are predominantly natural fires set by lightning but also unintentional or arson fires set by humans. About 90% of all fires worldwide are biomass burning fires; the rest are forest fires. Biomass burning produces gases, such as carbon dioxide, carbon monoxide, methane, nitrous oxide, oxides of nitrogen, and reactive organic gases. It also produces particles, such as ash, plant fibers, soil dust, inorganic compounds, organic matter, and soot. Organic matter (OM) consists of carbon and hydrogen based compounds and often contains oxygen (O), nitrogen (N), etc., as well [12].

1.1.4 Biomass Burning over Equatorial Africa

Biomass burning has long been recognized as a significant source of reactive species such as CO, hydrocarbons and NO_x and aerosols, which play an important role in the chemistry and radiative budget of the troposphere. In Africa, human activities are the primary cause of biomass burning, including forest clearance, savanna burning restoration and land management [13].

For long-lived trace gases such as CO, pyrogenic convection into the free troposphere results in polluted air-masses traveling long distances where signatures of such events have previously been observed in measurements made around Africa . Another important trace species linked to BB is tropospheric ozone (O_3), which is formed via the photochemical oxidation of NO_2 released during burning events. For instance, it has previously been shown that enhanced concentrations of tropospheric (O_3) in the lower atmosphere over Equatorial Africa during July can be attributed directly to emissions from fires near the equator, Ethiopia is within this region [14].

Agricultural and agro-industrial residues constitute 15% of the total energy consumed in Ethiopia [15]. Residues are mostly used in the domestic sector for cooking and baking, using very low efficiency devices. Residue supply is seasonal and residue use as fuel is also seasonal. In different parts of the country, various types of crops are cultivated and, as a result, a considerable volume of crop residues is also produced. Generally, for use as fuel, crops with a higher residue-to-seed ratio provide the largest volume of potential biomass. However, it is often not desirable, socially and environmentally acceptable or, indeed, economically viable to divert all types of biomass residue for fuel [15].

The intentional burning of land results in a major source of combustion products to the atmosphere. Most of this burning occurs in the tropics. Emissions from burning vegetation are typical of those from any uncontrolled combustion process and include CO_2, CO, NO_x, CH_4 and non-methane hydrocarbons, and elemental and organic particulate matter [9]. For this study we have selected four interrelated tropospheric tracers of air pollution discussed in the following subsections. The retrieved results and discussion of the first two (HCN and C_2H_6) measured by high resolution Fourier transform infrared spectrometer (FTIRS) at Collage of Natural Science, Arat Killo Campus, of Addis Ababa University (located around 9.01^0 N, 38.76^0 E and 2.45 km asl), Ethiopia are presented in chapter four of this thesis.

1.1.5 Tropospheric Hydrogen Cyanide, Ethane, Carbon Monoxide and Nitric Oxide

Biomass burning is a significant source of tropospheric CO, non-methane hydrocarbons (NMHCs), and HCN. Carbon monoxide (CO), ethane (C_2H_6), and acetylene (C_2H_2) are important tropospheric ozone (O_3) precursors. Changes in the tropospheric concentration of these molecules could affect the regional and global tropospheric O_3 budget [1]. So, let us see how O_3 is produced from HCN, C_2H_6, CO and NO and its effect in the troposphere.

1.1.5.1 Hydrogen Cyanide (HCN)

HCN (also called formonitrile) is a highly volatile, colorless, and extremely poisonous liquid and gas [16]. An HCN concentration of 300 mg/m^3 in air will kill a human within about 10 minutes. It is estimated that hydrogen cyanide at a concentration of 3500 ppmv (about 3200 mg/m^3) will kill a human in about 1 minute [16]. The toxicity is caused by the cyanide ion, which halts cellular respiration by inhibiting an enzyme in mitochondria called cytochrome c oxidase. There are many naturally occurring substances yielding cyanide in their seeds, such as the pit of the wild cherry. It usually occurs in combination with plant sugars. The tuberous edible plant of the spurge family called cassava (also known as manioc, mandioc, or yuca) were used by primitive peoples to produce HCN for poison darts and arrows. HCN is produced by other plants, bacteria and fungi [17]. HCN is a minor constituent of the atmosphere emitted primarily from biomass burning. In the troposphere, it is lost mainly by ocean uptake presumably through biological activity, with a small loss by reaction with OH. Having a relatively simple behaviour and a tropospheric lifetime of several months (2-4 months), HCN is considered to be a good tracer of biomass burning [18]. According to Lary [17] the products of the atmospheric oxidation of HCN are NO, CO and O_3. Reactions of CH_4 and CH_3OH with N are found to be important sources of HCN. Including the pseudohalogen chemistry in HCN budget gives a small increase in ozone and total reactive nitrogen (NO_y):

$$HCN + OH \longrightarrow H_2O + CN \tag{1.1.1}$$

$$CN + O_2 \longrightarrow NCO + O \tag{1.1.2}$$

$$NCO + h\nu \longrightarrow N + CO \tag{1.1.3}$$

$$N + O_2 \longrightarrow NO + O \tag{1.1.4}$$

$$2O + 2O_2 \longrightarrow 2O_3 \tag{1.1.5}$$

$$O_3 + h\nu \longrightarrow O(^1D) + O_2 \tag{1.1.6}$$

$$O(^1D) + H_2O \longrightarrow 2OH \tag{1.1.7}$$

$$Net : HCN + 3O_2 + 2h\nu \longrightarrow CO + O_3 + NO + OH \tag{1.1.8}$$

1.1.5.2 Ethane C_2H_6

Ethane is the second (i.e. next to methane) most abundant organic trace gas in the background troposphere. With mixing ratios ranging from a fraction of a part per billion to a few parts per billion it is roughly 3 orders of magnitude less abundant than methane. However, because of its higher reactivity and the possibility of forming more complex intermediate the oxidation of ethane is important to tropospheric chemistry. Thus the existence of ethane in the background troposphere opens up chemical pathways in atmospheric chemistry which would not be accessible from methane oxidation alone. Ethane can also serve to test predictions of atmospheric transport and chemistry models and to check the strength and geographic distribution of sources of atmospheric trace gases. The tropospheric lifetime of a few months is long enough to reduce the random variability in the mixing ratios in the background troposphere to an acceptable level but short enough to allow pronounced seasonal cycles and latitudinal gradients [19]. The oxidation of ethane produces ozone with other intermediate products such as PAN [20]:

$$C_2H_6 + 12O_2 + 6h\nu \rightarrow H_2 + 2H_2O + CO_2 + 6O_3 \tag{1.1.9}$$

The reaction of C_2H_6 which takes place in 24 steps, produces an intermediate product proxyacetylnitrate $CH_3C(O)OONO_2$ (PAN), which is a strong plant toxin and the most familiar eye-irritating component of smog.

1.1.5.3 Carbon Monoxide (CO)

Carbon monoxide is a tasteless, colorless, and odorless gas that is toxic to humans and animals exposed to it for one hour at mixing ratios above about 700 ppmv. Exposure to 300 ppmv for one hour causes headaches. In urban regions away from freeways, CO

mixing ratios are typically 210 ppmv. On freeways and in traffic tunnels, mixing ratios can rise to more than 100 ppmv. In the free troposphere, CO mixing ratios vary from 50 to 150 ppbv. A major source of CO is incomplete combustion by automobiles, trucks, and airplanes. Wood burning and grass burning are also important sources. Natural sources of CO are plants and biological activity in the oceans [12]. According to Gezahegn [21] through rapid reaction with hydroxyl radicals in the troposphere, oxidation of non-methane hydrocarbons (NMHCs) contributes large quantities of CO to the troposphere (e.g the above reaction of HCN with OH radical).

Although CO is the most abundantly emitted pollutant gas in urban air, it does not play a major role in photochemical smog formation. Photochemical smog is characterized by the buildup of ozone and related products. CO does not produce much ozone in urban air. The major sink of CO is chemical conversion to carbon dioxide (CO_2) like [21]:

$$2CO + O_2 \rightarrow 2CO_2, \qquad CO + O_3 \rightarrow CO_2 + O_2 \tag{1.1.10}$$

1.1.5.4 Nitric Oxide (NO)

Nitric oxide is also a free radical. It is a colorless gas emitted from soils, plants, and combustion processes and produced by lightning and chemical reaction. Combustion sources include aircraft, automobiles, oil refineries, and biomass burning. The primary sink of NO is chemical reaction. A typical mixing ratio of NO in the background troposphere near sea level is 5 pptv. In the upper troposphere NO mixing ratios increase to 2060 pptv. In urban regions, NO mixing ratios reach 0.1 ppmv in the early morning but decrease significantly by midmorning [12]. According to Crutzen and Birks [20], when the concentration of NO in the troposphere exceeds 1/4000 that of ozone, the oxidation of CO increases the concentration of ozone:

$$CO + 2O_2 + h\nu \rightarrow CO_2 + O_3 \tag{1.1.11}$$

From the above tropospheric (C_2H_6, HCN, CO and NO) trace gases we can observe that the production of tropospheric O_3. While stratospheric ozone shields us from ultraviolet

radiation, in the troposphere this irritating, reactive molecule damages forests and crops; destroys nylon, rubber, and other materials; and injures or destroys living tissue. It is a particular threat to people who exercise outdoors or who already have respiratory problems. Ozone affects plants in several ways. High concentrations of ozone cause plants to close their stomata. These are the cells on the underside of the plant that allow carbon dioxide and water to diffuse into the plant tissue. This slows down photosynthesis and plant growth. Ozone may also enter the plants through the stomata and directly damage internal cells. Rubber, textile dyes, fibers, and certain paints may be weakened or damaged by exposure to ozone. Some elastic materials can become brittle and crack, while paints and fabric dyes may fade more quickly. When ozone pollution reaches high levels, pollution alerts are issued urging people with respiratory problems to take extra precautions or to remain indoors. Smog can damage respiratory tissues through inhalation. Ozone has been linked to tissue decay, the promotion of scar tissue formation, and cell damage by oxidation. It can impair an athlete's performance, create more frequent attacks for individuals with asthma, cause eye irritation, chest pain, coughing, nausea, headaches and chest congestion and discomfort. It can worsen heart disease, bronchitis, and emphysema [22].

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 where by 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 describes how infrared spectra arise and how they are recorded at Addis Ababa, Ethiopia.

2.1 Infrared Atmospheric Spectra

Infrared (IR) spectroscopy is measurement of the absorption of different IR frequencies by a sample positioned in the path of an IR beam. Certain frequencies (energies) of a radiation are absorbed when it passes through a sample, leading to vibrational and rotational motions in the molecules. The position of such absorption is measured in wave-number ν (cm^{-1}) (number of waves per unit length), or wavelength λ (m). The IR region of the electromagnetic spectrum is the region lying between the visible and microwave regions (Fig. 2.1), having wavenumbers from roughly 13000 to 10 cm^{-1} or wavelengths from 0.78 to 1000 μ m. The IR region is divided into three smaller areas, Near

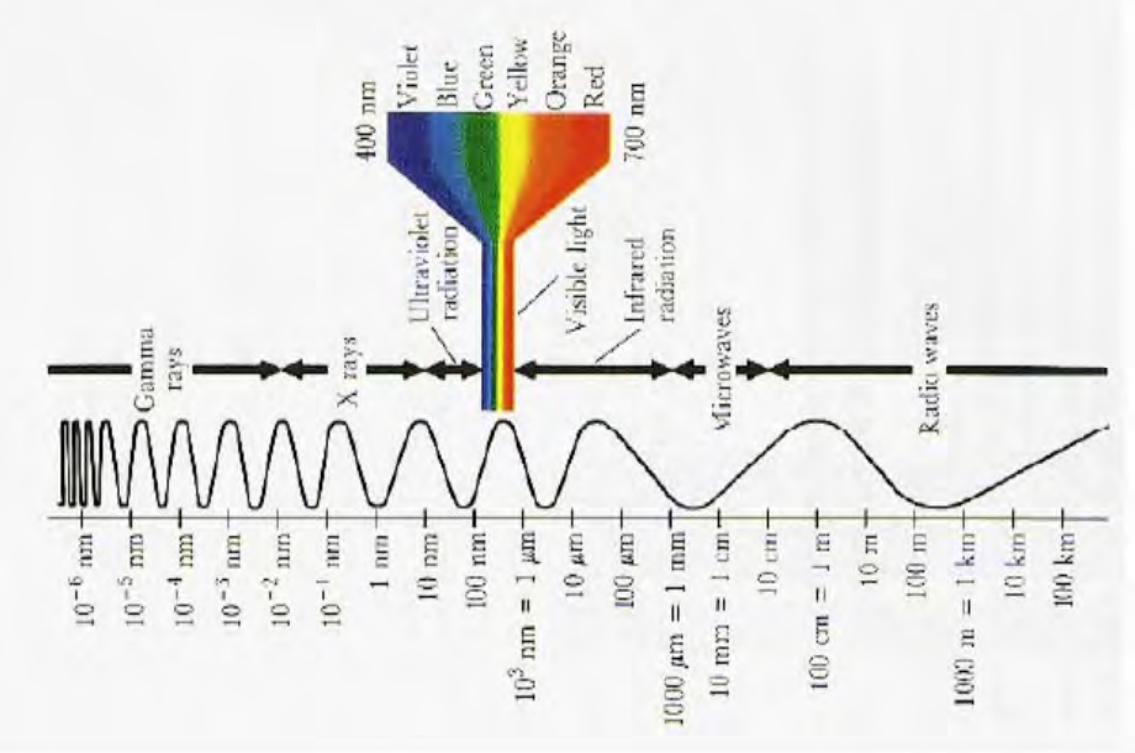

Figure 2.1: Electromagnetic spectrum.

InfraRed (NIR), Mid InfraRed (MIR) and Far InfraRed (FIR). IR spectrum information is presented in the form of spectrum intensity with wave-number (or wavelength) as the x-axis and the intensity as y-axis [23]. Fig. 2.2 show an example of the infrared spectrum which was measured with high resolution FTIR spectrometer in Addis Ababa.

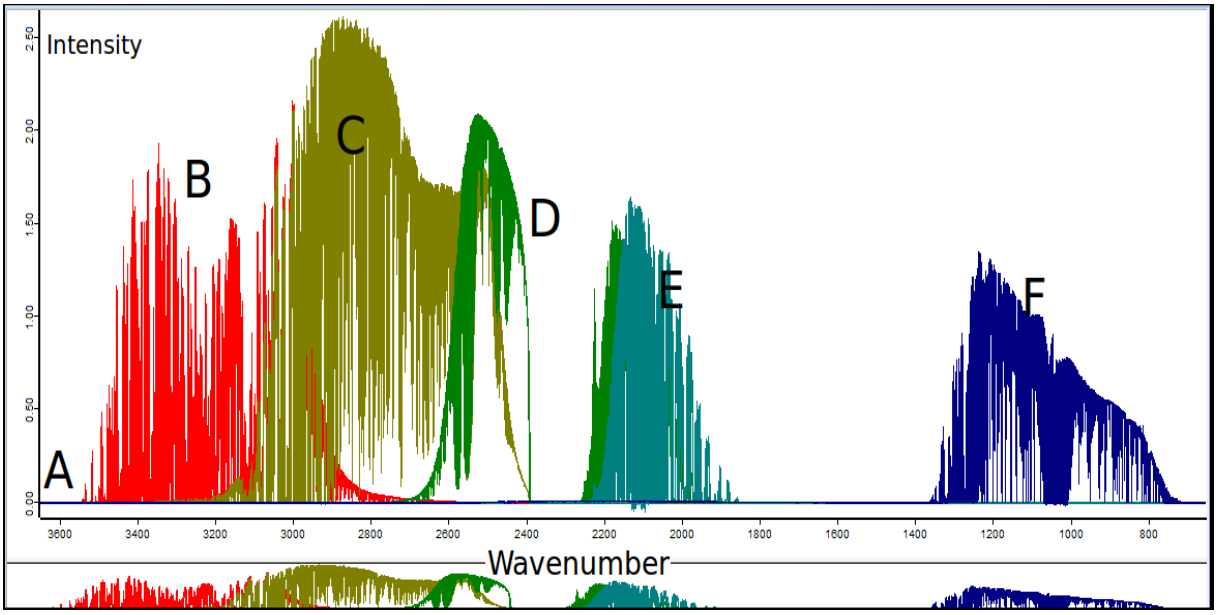

Figure 2.2: Infrared information spectrum measured with FTIR spectrometer in Addis Ababa.

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 [24] and Remedios [25].

2.1.1.1 Rotational Energy Levels

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 homo nuclear 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).

Rotational energy levels are quantized, this means after absorbing of energy the molecules in the ground rotational states will be raised to higher states, and therefore the energy levels can be derived by assuming rigid rotator molecule with masses m_1 and m_2 separated by a distance r_o and rotating about point S (Fig. 2.3). The moment of inertia, I, of the

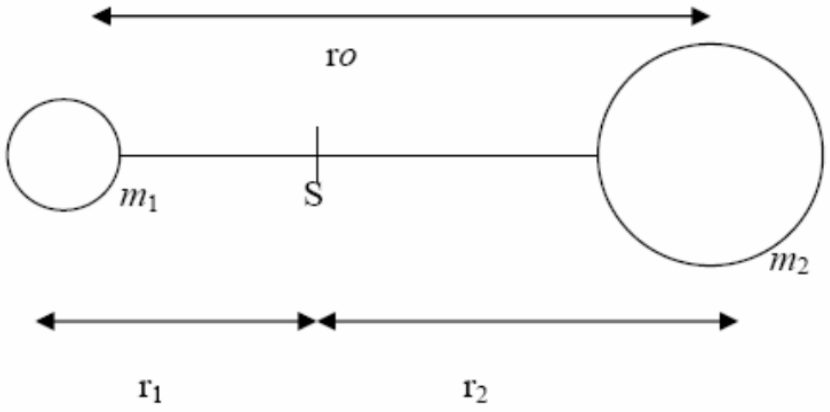

Figure 2.3: Rigid rotator.

molecule is defined as $I = \sum m_i r_i^2$, thus for a two atomic molecule I can be written as $I = m_1 r_1^2 + m_2 r_2^2$, using $r_o = r_1 + r_2$ and $m_1 r_1 = m_2 r_2$, therefore $r_1 = \frac{m_2 r_o}{m_1 + m_2}$ and $r_2 = \frac{m_1 r_o}{m_1 + m_2}$ leads to $I = \frac{m_1 m_2}{m_1 + m_2} r_0^2 = \mu r_0^2$ where $\mu = \frac{m_1 m_2}{m_1 + m_2}$ is the so called reduced mass. By Schrödinger equation the rotational energy levels for diatomic molecules are given as [23, 26, 27]

$$E_J = \frac{h^2}{8\pi^2 I} J(J+1) \tag{2.1.1}$$

where J = 0,1,2,3,... is rotational quantum number, h(6.626×10^{-34} J.s) is Plank's constant. The corresponding frequencies can be derived as $\nu = \Delta$ E/hc (c = 3×10^8 m/s, speed of light in the vacuum) therefore, $\varepsilon_J = \frac{E_J}{ch} = \frac{h}{8\pi^2 Ic} J(J+1) = BJ(J+1)$ (cm^{-1}), where $B = \frac{h}{8\pi^2 Ic}$ (cm^{-1}), rotational constant. 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 Energy Levels

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. In ethene there are 12 internal coordinates: 4 C-H stretching, 1 C-C stretching, 2 H-C-H bending, 2 CH_2 rocking, 2 CH_2 wagging, 1 twisting. Note that the H-C-C angles cannot be used as internal coordinates as the angles at each carbon atom cannot all increase at the same time. 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 homo-nuclear 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) of vibrational quantum number ν by [26, 37]

$$E_\nu = (\nu + \frac{1}{2})\hbar\varpi \qquad\qquad \varpi = \sqrt{\frac{k}{m}} \tag{2.1.2}$$

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 Hookes law (due to, e.g. nonlinear dipole moment variation with atomic displacement) the SHO transforms into an Anharmonic Oscillator with different vibrational selection rules ($\Delta\nu = \pm1, \pm2, \pm3, ...$). 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 $\Delta\nu = \pm1, \pm2, \pm3$ 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.

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 $\Delta\nu = \pm1, \pm2, \pm3, ...$ and $\Delta J = \pm1$.

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 Fig. 2.4 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 labeled as P_1, P_2, $P_3, ...$ and R_0, R_1, $R_2, ...$away from the band center, with the numbers signifying the rotational quantum number in the lower vibrational state (J''). P_0 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 approximation.

The discussion of rotational-vibrational spectra up to this point applies only to diatomic

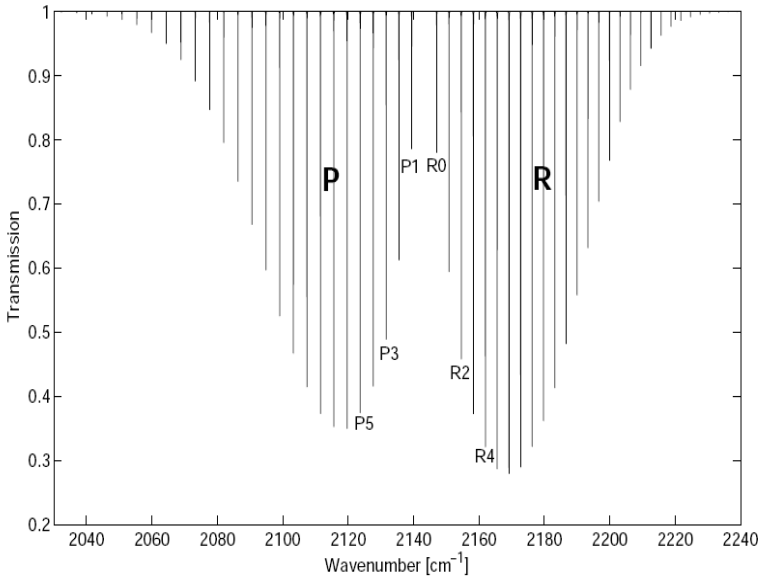

Figure 2.4: Rotational-vibrational transmission spectrum of CO showing its PR band structure.

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 (Fig. 2.5).

Finally, the spectra of molecules with more complicated geometries; symmetric tops (e.g.

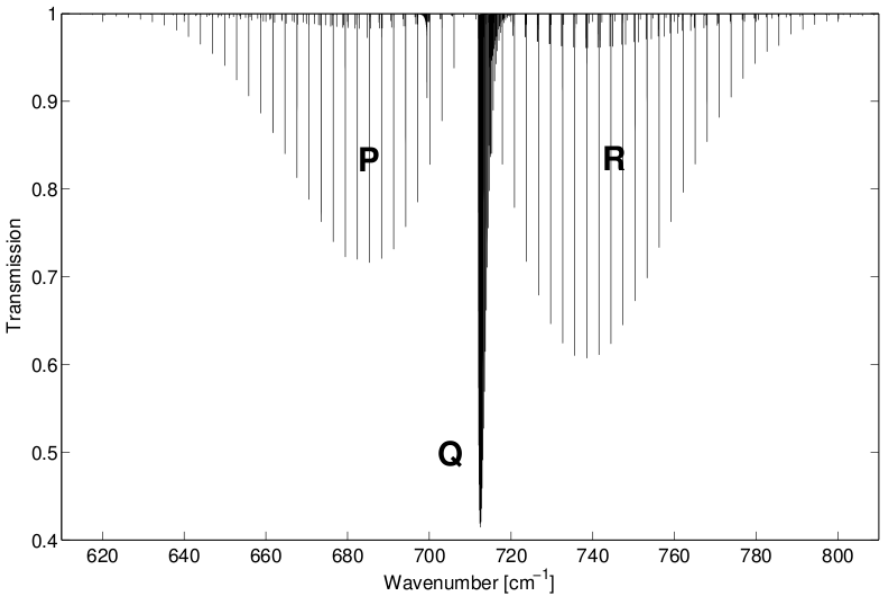

Figure 2.5: Rotational-vibrational transition spectrum of HCN showing its PQR band structure.

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.

2.1.2 Transition Line Strength and Shape

The strength S($\bar{\nu}_0$, T) [cm^{-1}/(molec.cm^{-2})] of an absorption line with a central frequency $\bar{\nu}_0$ (cm^{-1}) (distinguished from the vibrational quantum number by the bar) in a rotational-vibrational band and at temperature T (K) is given by

$$S(\bar{\nu}_0, T) = S(\bar{\nu}_0, T_0)\frac{Q(T_0)exp(-E''/kT)(1 - exp(-hc\bar{\nu}_0/kT))}{Q(T)exp(-E''/kT_0)(1 - exp(-hc\bar{\nu}_0/kT_0))} \tag{2.1.3}$$

where T_0 is a reference temperature of 296 K, Q is the partition function [$Q(T) \approx Q_v(T)Q_r(T)$], E'' is the lower state energy of the transition, h is the Planck's constant, c is the speed of light, and κ is the Boltzmann constant. The sensitivity of Eq. 2.1.3 to temperature is strongly influenced by E''. More details on the calculation of vibrational and rotational partition functions can be found in Zender [28], Fischer et al. [29] and Notholt et al. [30].

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. The absorption coefficient k($\bar{\nu}$) [1/(molec.cm^{-2})] that describes a broadened transition line is modeled as the product of the line strength with the broadening function f($\bar{\nu}$ - $\bar{\nu}_0$) (1/cm^{-1}), normalized as per Eq. 2.1.5:

$$k(\bar{\nu}) = S(\bar{\nu}_0, T)f(\bar{\nu} - \bar{\nu}_0) \tag{2.1.4}$$

$$\int_0^\infty f(\bar{\nu} - \bar{\nu}_0)\,\mathrm{d}\bar{\nu} = 1. \tag{2.1.5}$$

2.1.2.1 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. It is the result of the probability of spontaneous emission between an upper state and a lower energy state. This probability

limits the life-time in the upper state if the molecule is considered to be isolated from radiation fields and other particles. The life-time in a state and its energy are linked by the uncertainty principle ($\Delta t \Delta E \geq \hbar$). The intensity is given by the relationship:

$$I(\nu) \sim A(\nu)A^*(\nu), \tag{2.1.6}$$

$$I(\nu) = \frac{I_0}{(\nu - \nu_0 + \frac{1}{4\tau^2})}. \tag{2.1.7}$$

The profile thus obtained is a Lorentzian profile. The profile has to be normalized because changing the profile must not change the area under it. The normalized line profile is

$$\Phi(\nu - \nu_0) = kI(\nu) \tag{2.1.8}$$

where k is a constant of normalization to be determined using the following integral:

$$\int_0^\infty \Phi(\nu - \nu_0) d\nu = 1 \tag{2.1.9}$$

The integration gives $k = \Gamma/4\pi^2 I_0$ and the normalized line profile (line shape) due to the natural broadening is written as

$$\Phi_n(\nu - \nu_0) = \frac{a_n}{\pi[(\nu - \nu_0)^2 + a_n^2]} \tag{2.1.10}$$

where a_n is the natural line shape half width at half-maximum (HWHM) of the normalized Lorentzian profile. For a given transition (here $j \longrightarrow i$), it is possible to give an estimate of the natural width, provided one knows the corresponding life-times of these states:

$$\alpha(\nu) = \frac{1}{4\pi c}(\frac{1}{\tau_i} + \frac{1}{\tau_j}) \tag{2.1.11}$$

where τ_i and τ_j are the life-times of the lower and the upper energy states involved in the molecular transition respectively. The natural life time of the excited state of a molecule is of the order of 10^{-2} - 10^{-1} sec. This is much longer than the life time between collisions in a gas at all ordinary pressures and so in practice, therefore, the line shapes we are concerned with are dominated by collisional effects [21].

2.1.2.2 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

$$f(\bar{\nu} - \bar{\nu}_0) = \frac{1}{\pi} \frac{\gamma_L}{(\bar{\nu} - \bar{\nu}_0 - \gamma_L)^2 + \gamma_L^2}, \tag{2.1.12}$$

where

$$\gamma_L(T, P) = \left(\frac{T_0}{T}\right)^{n_{air}} [\gamma_{air}(T_0, P_0)(p - P_s) + \gamma_{self}(T_0, P_0)P_s]. \tag{2.1.13}$$

The pressure-dependent shift of the line center frequency, $\delta_L = \delta_{air} p$, is introduced in Eq. 2.1.12 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 ($\text{cm}^{-1}.\text{atm}^{-1}$) 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. HWHM of the Lorentzian, γ_L (cm^{-1}), is calculated according to Eq. 2.1.13 given the different self and air-broadening coefficients γ_{air} and γ_{self} ($\text{cm}^{-1}.\text{atm}^{-1}$), 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}. Eq. 2.1.13 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 $\sim 25km$, and negligible at $\sim 40km$. 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.

2.1.2.3 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 Eq. 2.1.14, where M is the molecular mass and the HWHM is equal to $\sqrt{ln2}\gamma_D(cm^{-1})$.

$$f(\bar{\nu}-\bar{\nu}_0) = \frac{1}{\sqrt{\pi}\gamma_D} exp(\frac{-(\bar{\nu}-\bar{\nu}_0)^2}{\gamma_D^2}), \qquad \gamma_D(T) = \sqrt{\frac{2KT}{M}}\frac{\bar{\nu}_0}{c} \tag{2.1.14}$$

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 [31].

2.1.2.4 Voigt Line Shape

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

$$f(x,y) = \frac{1}{\gamma_D}\frac{y}{\pi^{3/2}}\int_{-\infty}^{\infty}\frac{exp(-t^2)}{(x-y)^2+y^2}\mathrm{d}t \tag{2.1.15}$$

where $x = (\bar{\nu} - \bar{\nu}_0)/\gamma_D$ and $y = \gamma_L/\gamma_D$. Eq. 2.1.15 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 (stands for Spectral FITting) [6] and PROFFIT (stands for PROFile FIT) [32] retrieval algorithm which are going to be described in detail in Chapter 3.

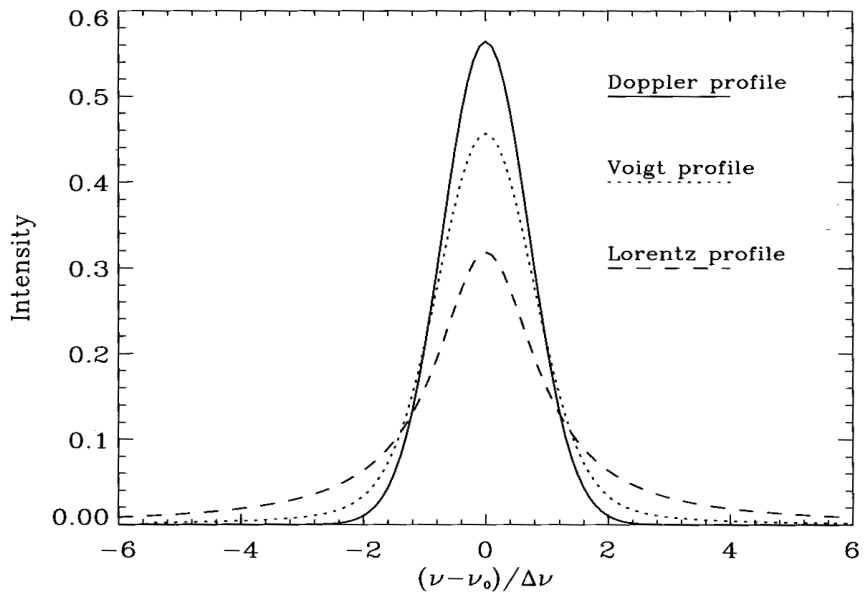

Figure 2.6: Comparison of the line shapes.

2.1.2.5 HITRAN Spectral Database

The HIgh resolution TRANsmission (HITRAN) database is a research-standard database containing reference temperature and pressure line strengths ($S(\bar{\nu}_0)$) 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, HITRAN 2000 and 2001, as well as HITRAN 2004 exist concurrently, and new, more accurate measurements are continuously being compiled for future updates of the database ([6], and references therein). 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 [33] and Stephens [34].

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

$$\frac{\mathrm{d}I_{\bar{\nu}}(\theta,\phi)}{ds} = \sigma_a[B_{\bar{\nu}}(T) - I_{\bar{\nu}}(\theta,\phi)] \tag{2.1.16}$$

where $I_{\bar{\nu}}$ is the monochromatic radiance ($W.cm^{-2}sr^{-1}$) 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 (cm^{-1}), and $B_{\bar{\nu}}(T)$ is the monochromatic blackbody emission at temperature T ($\mathrm{W.cm^{-2}sr^{-1}}$). Eq. 2.1.16 is a first order linear ordinary differential equation known in the remote sounding community as Schwarzchild's Equestion, which can be integrated to give

$$I_{\bar{\nu}}(0) = I_{\bar{\nu}}(TOA).exp\bigg(-\int_{TOA}^{0}\rho K_{\bar{\nu}}ds\bigg) + \int_{TOA}^{0}\rho K_{\bar{\nu}}B_{\bar{\nu}}(T).exp\bigg(-\int_{TOA}^{s}\rho K_{\bar{\nu}}ds''\bigg)ds \tag{2.1.17}$$

where ρ is the density of the absorbing gas ($\mathrm{molec.cm^{-3}}$), and the product $\rho K_{\bar{\nu}}$ 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 $\chi_{\bar{\nu}} = \int_{Z_1}^{Z_2} \rho K_{\bar{\nu}} dz$ then $d\chi_{\bar{\nu}} = \mu \rho K_{\bar{\nu}} ds$ and Eq. 2.1.17 simplifies to

$$I_{\bar{\nu}}(0) = I_{\bar{\nu}}(TOA).exp\left(\frac{-\chi^0_{\bar{\nu}}}{\mu}\right) + \int_0^{\bar{\nu}} B_{\bar{\nu}}(T).exp\left(-\frac{\chi^0_{\bar{\nu}} - \chi_{\bar{\nu}}}{\mu}\right)\frac{d\chi_{\bar{\nu}}}{\mu} \tag{2.1.18}$$

where $\chi^0_{\bar{\nu}}$ is the vertical optical depth from TOA to the surface. If we finally introduce the transmission between two vertical optical depths as $\tau_{\bar{\nu}}(\chi_1, \chi_2) = exp(-|\chi_2 - \chi_1|)$ then Eq. 2.1.18 is further simplified to

$$I_{\bar{\nu}}(0) = I_{\bar{\nu}}(TOA).\tau_{\bar{\nu}}(0, \chi^0_{\bar{\nu}})^{\frac{1}{\mu}} + \int_1^{\tau_{\bar{\nu}}(0,\chi^0_{\bar{\nu}})} B_{\bar{\nu}}(T).\tau_{\bar{\nu}}(0, \chi^0_{\bar{\nu}})^{\frac{1}{\mu}-1} d\tau_{\bar{\nu}} \tag{2.1.19}$$

When $\theta = 0$ then $\mu = 1$ and Eq. 2.1.19 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 PROFIT algorithm the atmospheric emission term is ignored in the forward model on grounds that it is negligible compared to the large solar radiance. A recent calculation by Notholt et al. [35] 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 Eq. 2.1.19. In practice these effects are taken into account by the auxiliary fscatm subroutine of the PROFFIT algorithm which performs detailed ray tracing calculations.

2.2 Fourier Transform Spectrometer

In the preceding section we describe 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 is described next.

2.2.1 The Ideal Fourier Transform Spectrometer

The FTS is based upon a Michelson interferometer, after its original designer Michelson [36]. Following Fig. 2.7 describes FTIR working principles 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 of the interferometer by means of a moving mirror. If the source radiation

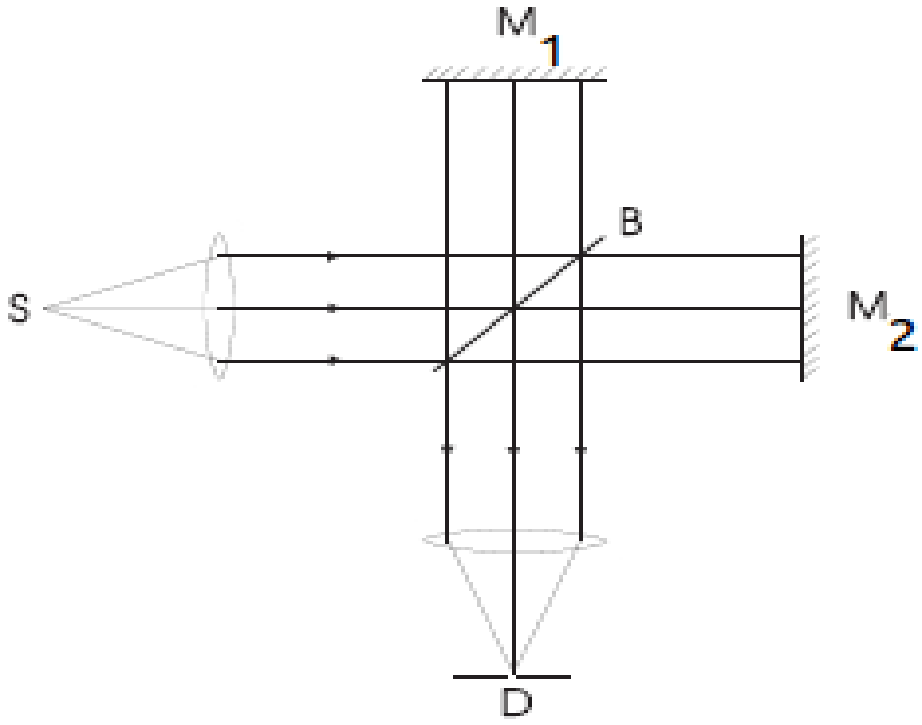

Figure 2.7: Michelson interferometer schematic after Chamberlain (1979), where S = point source, B = beamsplitter, M_1 = fixed mirror, M_2 = moving mirror, and D = detector.

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 ([6], and reference therein), 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 [37], for a monochromatic source at wavenumber $\bar{\nu}$ (wavelength $= 1/\bar{\nu}$), 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, $I(\bar{\nu})$:

$$I'(\delta) = \frac{1}{2}I(\bar{\nu})(1 + cos2\pi\frac{\delta}{\lambda}) = \frac{1}{2}I(\bar{\nu})(1 + cos2\pi\bar{\nu}\delta) \tag{2.2.1}$$

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

$$I(\delta) = \frac{1}{2}I(\bar{\nu})H(\bar{\nu})cos2\pi\bar{\nu}\delta \tag{2.2.2}$$

The factor $H(\bar{\nu})$ 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 $B(\bar{\nu}) = I(\bar{\nu})H(\bar{\nu})$ is introduced to represent the intensity of the source at wavenumber $\bar{\nu}$ 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 $\bar{\nu}$ is written as

$$I(\delta) = \frac{B(\bar{\nu})}{2}cos2\pi\bar{\nu}\delta \tag{2.2.3}$$

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

$$I(\delta) = \int_{-\infty}^{\infty} \frac{B(\bar{\nu})}{2}cos2\pi\bar{\nu}\delta d\bar{\nu} \tag{2.2.4}$$

Eq. 2.2.4 is one half of a cosine Fourier transform pair. The inverse transform is written as

$$B(\bar{\nu}) = 2\int_{-\infty}^{\infty} I(\delta)cos2\pi\bar{\nu}\delta d\delta \tag{2.2.5}$$

or equivalently (for $I(\delta)$ an even function) as

$$B(\bar{\nu}) = 4\int_0^\infty I(\delta) cos 2\pi\bar{\nu}\delta d\delta \tag{2.2.6}$$

Following the treatment of Remedios [25], we can derive Eq. 2.2.6 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 ($\langle\rangle = \frac{1}{T}\int_{\frac{-1}{T}}^{\frac{1}{T}} dt$) of the complex conjugate (*) product of the electric field component ([6], and reference therein):

$$I_0 = \langle \vec{E}(\vec{r},t)^*.\vec{E}(\vec{r},t)\rangle, \tag{2.2.7}$$

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

$$\vec{E}(\vec{r},t) = \int_{-\infty}^{\infty} \sqrt{B(\nu)} e^{2\pi i(\nu t + \vec{r}.\vec{K}(\nu))} d\nu, \tag{2.2.8}$$

where $\sqrt{B(\nu)}$ is the amplitude of frequency component $\nu(s^{-1})$, 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

$$\vec{E}_{out}(\vec{r},t) = \frac{\vec{E}(\vec{r},t)}{2} + \frac{\vec{E}(\vec{r},t+\tau)}{2} \tag{2.2.9}$$

and hence the intensity at the detector can be written as

$$I_{out} = \langle \left[\frac{\vec{E}(\vec{r},t)}{2} + \frac{\vec{E}(\vec{r},t+\tau)}{2}\right]^* . \left[\frac{\vec{E}(\vec{r},t)}{2} + \frac{\vec{E}(\vec{r},t+\tau)}{2}\right]\rangle. \tag{2.2.10}$$

After expanding Eq. 2.2.10, the intensity is written as

$$\begin{aligned}\langle I_{out}\rangle \quad &= \frac{1}{T}\int_{\frac{-1}{T}}^{\frac{1}{T}} \left[\frac{\vec{E}(\vec{r},t)}{2}^* . \frac{\vec{E}(\vec{r},t)}{2} + \frac{\vec{E}(\vec{r},t+\tau)}{2}^* . \frac{\vec{E}(\vec{r},t+\tau)}{2} + \right. \\ &\left. + \frac{\vec{E}(\vec{r},t)}{2}^* . \frac{\vec{E}(\vec{r},t+\tau)}{2} + \frac{\vec{E}(\vec{r},t+\tau)}{2}^* . \frac{\vec{E}(\vec{r},t)}{2}\right] dt \end{aligned} \tag{2.2.11}$$

which simplifies to

$$I_{out} = \frac{I_0}{2} + \frac{1}{T}\int_{\frac{-1}{T}}^{\frac{1}{T}} \left[\frac{\vec{E}(\vec{r},t)}{2}^{*} . \frac{\vec{E}(\vec{r},t+\tau)}{2} + \frac{\vec{E}(\vec{r},t+\tau)}{2}^{*} . \frac{\vec{E}(\vec{r},t)}{2}\right] dt \tag{2.2.12}$$

when we recognize that the first term in Eq. 2.2.11 is equal to the input beam intensity divided by 4, and that the second term is equal to the first term when the averaging time T is sufficiently greater than the phase delay τ . Finally, upon substituting Eq. 2.2.8 into Eq. 2.2.12 we write

$$I_{out} = \frac{I_0}{2} + \frac{1}{T}\int_{\frac{-1}{T}}^{\frac{1}{T}}\int_{-\infty}^{\infty} \left[\frac{B(\nu)}{4} e^{-2\pi i(\nu t + \vec{r}.\vec{k}) + 2\pi i(\nu(t+\tau)+\vec{r}.\vec{k})} + e^{-2\pi i(\nu(t+\tau)+\vec{r}.\vec{k}) + 2\pi i(\nu t + \vec{r}.\vec{k})}\right] d\nu dt \tag{2.2.13}$$

Expanding and simplifying the exponential terms reveals that the time variable vanishes (Eq. 2.2.14). 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 (Eq. 2.2.15). Eq. 2.2.16 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 :

$$I_{out} = \frac{I_0}{2} + \int_{-\infty}^{\infty} \frac{B(\nu)}{4}\left(e^{2\pi i\nu\tau} + e^{-2\pi i\nu\tau}\right) d\nu \tag{2.2.14}$$

$$= \frac{I_0}{2} + \int_{-\infty}^{\infty} \frac{B(\nu)}{4}(2cos2\pi\nu\tau) d\nu \tag{2.2.15}$$

$$= \frac{I_0}{2} + \int_{-\infty}^{\infty} \frac{B(\nu)}{2} cos2\pi\nu\tau d\nu \tag{2.2.16}$$

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

$$I(\tau) = \int_{-\infty}^{\infty} \frac{B(\nu)}{2} cos2\pi\nu\tau d\nu \tag{2.2.17}$$

Given $\tau = \delta/c$, $\nu = c\bar{\nu}$ 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 Eq. 2.2.17 in terms of δ and $\bar{\nu}$:

$$I(\delta) = \int_{-\infty}^{\infty} \frac{B(\bar{\nu})}{2} cos2\pi\bar{\nu}\delta d\bar{\nu} \tag{2.2.18}$$

which is identical to Eq. 2.2.4 resulting from the intuitive derivation after Griffiths and de Haseth [37]. The inverse cosine transform of Eq. 2.2.18 is again identical with Eq. 2.2.5. Since $I(\delta)$ is real, Eq.s 2.2.4 and 2.2.5 can also be written in terms of the full complex Fourier transform and the spectrum taken to be the real part of the transform:

$$I(\delta) = \int_{-\infty}^{\infty} B(\bar{\nu}) exp(i2\pi\bar{\nu}\delta) d\bar{\nu} \tag{2.2.19}$$

$$B(\bar{\nu}) = \int_{-\infty}^{\infty} I(\delta) exp(-i2\pi\bar{\nu}\delta) d\delta \tag{2.2.20}$$

where the multiplicative constants have been omitted. Eq. 2.2.20 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 the following sections.

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:

$$B(\bar{\nu}) = 2L\frac{sin(2\pi(\bar{\nu} - \bar{\nu}_o)L)}{2\pi(\bar{\nu} - \bar{\nu}_o)L} = 2Lsinc_\pi(2(\nu - \nu_0)L) \tag{2.2.21}$$

$sinc_\pi(x) = sin(\pi x)/(\pi x)$ is an alternative definition of the sinc function whose integral between $\pm\infty$ is 1. The first zeros of the sinc function (Fig. 2.8) 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 sidelobes and maintaining the line width.

Based on Fig. 2.8 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 $0.004cm^{-1}$ 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. 0.66×0.004 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.

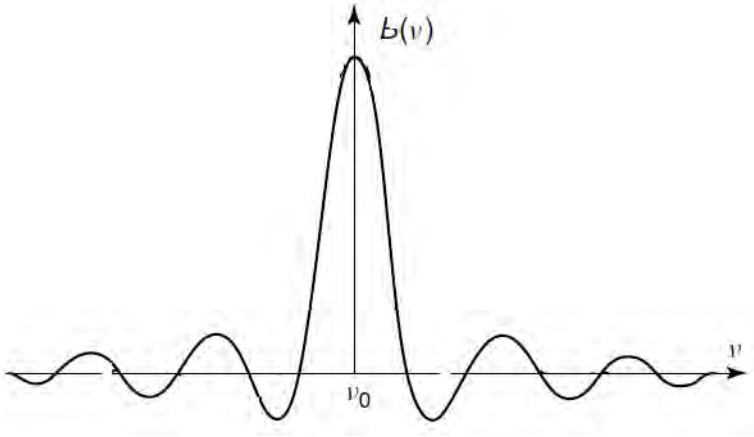

Figure 2.8: Instrumental Line Shape with Boxcar apodization function.

2.2.3.2 Finite Field-of-View (FOV)

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 Fig. 2.7 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 ([6], and reference therein).

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 $\text{dA}^*(\text{m}^2)$ (inclined at an

angle ϕ to the optical axis), a distance r(m) away from this source is

$$d\Phi = \frac{B}{r^2}(cos\theta ds)(cos\phi dA^*) \tag{2.2.22}$$

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 $d\Omega$, 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 [28]. We now write the radiant flux as

$$d\Phi = BdAd\Omega \tag{2.2.23}$$

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 etendue') as

$$E \equiv (d\Phi/B)_{source} = dAd\Omega = dA'd\Omega' = dA''d\Omega'' = ... = (d\Phi/B)_{detector} \tag{2.2.24}$$

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 ([6], and reference therein), 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

$$f_c^2 = \frac{A_s}{\Omega_s} = \frac{A_c}{\Omega_c} \tag{2.2.25}$$

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

$$E = \Omega_c A_s = \Omega_s A_c \tag{2.2.26}$$

which shows that a practical FTS must have a finite-sized entrance aperture for a nonzero 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, $dS_{\bar{\nu}0}$, produced by the off-axis rays of monochromatic radiation at $\bar{\nu}_0$ (and assuming that α is small such that $cos\alpha \approx 1 - \alpha^2/2 = 1 - \Omega/2\pi$) is written as

$$dS_{\bar{\nu}_0}(\delta) = S_{\bar{\nu}_0}\left[1 + cos(2\pi\bar{\nu}_0\delta(1 - \frac{\Omega}{2\pi}))\right]\frac{d\Omega}{\Omega_s} \tag{2.2.27}$$

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 $S_{\bar{\nu}_0}$, 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 Eq. 2.2.26:

$$dS_{\bar{\nu}_0} = I_{\bar{\nu}_0} dE = \frac{S_{\bar{\nu}_0}}{E} d\Omega_c A_s = \frac{S_{\bar{\nu}_0}}{\Omega_s A_c} d\Omega_s A_c = \frac{S_{\bar{\nu}_0} d\Omega_s}{\Omega_s} \tag{2.2.28}$$

where $d\Omega_s$ corresponds to $d\Omega$ in Eq. 2.2.27 (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 Eq. 2.2.27 over the solid angle, and consists of a constant and a variable term :

$$S_{\bar{\nu}_0}(\delta) = \int_0^{\Omega_s} S_{\bar{\nu}_0}\left[1 + cos(2\pi\bar{\nu}_0\delta(1 - \frac{\Omega}{2\pi}))\right]\frac{d\Omega}{\Omega_s} \tag{2.2.29}$$

$$= S_{\bar{\nu}_0} + S_{\bar{\nu}_0} sinc_\pi(\frac{\bar{\nu}_0 \delta \Omega_s}{2\pi}) cos\left[2\pi\bar{\nu}_0\delta(1 - \frac{\Omega_s}{4\pi})\right] \tag{2.2.30}$$

The variable term is altered in two ways in comparison to Eq. 2.2.29. The first effect of a finite FOV is to shift the spectral feature at wavenumber $\bar{\nu}_0$ to the lower wavenumber $\bar{\nu}_0(1 - \Omega_s/4\pi)$. 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 $\bar{\nu}_0 = \Omega_s/2$, 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 [38]. 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:

$$d = 2f_c\sqrt{\frac{2}{\bar{\nu}_0 L}} = f_c\sqrt{\frac{8}{R}} \tag{2.2.31}$$

where R is the (wavenumber-dependent) resolving power of the FTS, defined as $R = \bar{\nu}_0/\Delta\bar{\nu} = \bar{\nu}_0 L$ [38].

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 $(N\Delta\delta)^{-1} = \Delta\bar{\nu}$, or alternatively that $\Delta\delta\Delta\bar{\nu} = 1/N$, and we can write the

DFT pair as

$$F(k\Delta\bar{\nu}) = \Delta\delta \sum_{j=0}^{N-1} f(j\Delta\delta)e^{-i2\pi Kj/N} \tag{2.2.32}$$

$$f(j\Delta\delta) = \Delta\bar{\nu} \sum_{k=0}^{N-1} F(k\Delta\bar{\nu})e^{i2\pi Kj/N} \tag{2.2.33}$$

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 ([6], and reference therein). 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 ([6], and references therein). 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 n^{th} 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, e.g. twice per laser wavelength leads to a maximum detectable frequency of 15 800 cm^{-1}. However, in a method developed

by Brault ([6], and reference therein) 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 [25], in the presence of all three error types, we write the interferogram as

$$I(\delta) = [1 + \epsilon_2(\delta)] \int_{-\infty}^{+\infty} B(\bar{\nu}) exp[i2\pi\bar{\nu}(\delta + \epsilon_3(\bar{\nu}))] d\bar{\nu} + \epsilon_1 \tag{2.2.34}$$

$$= [1 + \epsilon_2(\delta)] \int_{-\infty}^{+\infty} B(\bar{\nu}) exp(i2\pi\bar{\nu}\delta) exp(i\phi(\bar{\nu})) d\bar{\nu} + \epsilon_1 \tag{2.2.35}$$

where $\phi(\bar{\nu}) = 2\pi\bar{\nu}\epsilon_3(\bar{\nu})$. 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 $exp(i\phi(\bar{\nu}))$ in Eq. 2.2.35, 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 $exp(i\phi(\bar{\nu}))$. The basic choice of recording a single-sided rather than a double-sided interferogram fundamentally destroys the interferograms 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 [38].

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 Eq. 2.2.35 can be written as

$$B'(\bar{\nu}) = B(\bar{\nu})exp(i\phi(\bar{\nu})) = B(\bar{\nu})[cos(\phi(\bar{\nu})) + isin(\phi(\bar{\nu}))] = B'_R(\bar{\nu}) + B'_I(\bar{\nu}) \qquad (2.2.36)$$

One could recover a real spectrum of the same magnitude as $B(\bar{\nu})$ by simply taking the modulus of the complex spectrum, $B'(\bar{\nu})$; however, this operation reduces the signal-to-noise ratio of the real spectrum because both the real and the imaginary components of

$B'(\bar{\nu})$ contain an equivalent amount of noise, which is carried into the real spectrum by the modulus operation [38]. Therefore, to recover the true spectrum, we find the phase angle $\phi(\bar{\nu})$, given by

$$\Phi(\bar{\nu}) = arctan\left(\frac{B'_I(\bar{\nu}}{B'_R(\bar{\nu}}\right) \tag{2.2.37}$$

and multiply the complex spectrum by $exp(i\phi(\bar{\nu}))$ (here implicitly assuming that intensity errors, $\epsilon_2(\bar{\nu})$, 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, ref. in [6]). 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. ref. in [6]), 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 $exp(i\phi(\bar{\nu}))$. 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 $B(\bar{\nu})$. 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 [37].

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 ([6], and reference therein), while a detailed discussion of noise is found in Griffiths and de Haseth [37], Chamberlain ([6], and reference therein) and Davis et al. [38].

2.2.5.1 Fellgett (Multiplex) Advantage

Historically, the Fellgett [39] advantage 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:

$$SNR_{FTS} \propto T^{1/2} \tag{2.2.38}$$

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:

$$SNR_G \propto (T/M)^{1/2} \tag{2.2.39}$$

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

$$\frac{SNR_{FTS}}{SNR_G} \propto M^{1/2} \tag{2.2.40}$$

Since modern grating spectrometer 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

$$SNR \propto \sqrt{(T/M)I(\bar{\nu})}. \tag{2.2.41}$$

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 (Eq. 2.2.26) without a loss of resolution. This is achieved by

increasing the area and the focal length of the collimating optics. Bell ([6], and reference therein) 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

$$\frac{E_{FTS}}{E_G} \simeq 2\pi \frac{f}{l} \tag{2.2.42}$$

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 spectrometer, 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 $1000cm^{-1}$ 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 Eq. 2.2.34, 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 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. [38] is to write the noise as being proportional to Ik, 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

$$SNR_\delta = I(0)/\epsilon_\delta \tag{2.2.43}$$

where $I(0)$ is the signal intensity at the interferogram centerburst. By the Fourier transform equal area theorem, $I(0) = \Delta\bar{\nu}N(\bar{B})$, where N now refers to the number of points in the interferogram and $\bar{B}$ is the mean signal strength in the spectral band. Using this relation, Davis et al. [38] show that the interferogram SNR is related to the spectrum SNR by

$$SNR_{\bar{\nu}} = \frac{B(\bar{\nu})}{\epsilon_{\bar{\nu}}} = \sqrt{\frac{2}{N}}\frac{B(\bar{\nu})}{\bar{B}}SNR_\delta \tag{2.2.44}$$

where $\epsilon_{\bar{\nu}}$ is the uniformly distributed spectrum noise. Eq. 2.2.44 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 discussion in the previous sections 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 [37], 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 $\sim$ 20 minutes or $\sim$ 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 FTIR Spectrometer Set up in Addis Ababa

The high-resolution FTIR Spectrometer, IFS 120M, manufactured by Bruker Optics Company is installed in Addis Ababa, at Arat Killo Campus of Addis Ababa University(AAU) in May, 2009 in collaboration with the Karlsruhe Institute of Meteorology and Climate Research (IMK), currently called Karsruhe Institute of Technology (KIT), Karlsruhe, Germany. This instrument is the first of its kind in Africa. The location of the observation site is closer to geoobservatory, which is found in science faculty of AAU. The container (housing) of the instrument is first set in a level made of concrete by clearing bushes and trees near by the container in order to prevent shadings. For optimum spectrometer performance and long-term reliability, the instrument operates at stable temperature of 20^0C, which is maintained by air conditioner.

Figure 2.9: The FTIR measurement site in Addis Ababa.

2.3.1 FTIR Components

2.3.1.1 Solar Tracker

The need for solar tracker is to measure direct solar spectra. The solar tracker in use for Addis Ababa station is not interfaced with the instrument and computer. It is a tripod with reflecting mirror facing the sun, moving with the speed of the sun. The tracker has a motor to provide a constant power so that the incoming solar radiation is reflected from the plane mirror of the tracker to another plane mirror inside the container which in turn reflects the beam to interferometer passing through aperture of size of interest.

The stability of the solar radiation (being remain focused on the aperture) is about an hour, which is sufficient time to take measurements in five filters for ten scan times. Actually, the stability can be improved by aligning the elevation and azimuth of the tripod, directly to polaris (north star) during clear sky night, or by developing a software to automate tracking.

2.3.1.2 Optical Devices

The optical components of the FTIR are the most sensitive parts of the instrument. It consists of paraboloids (parabolic mirrors) used to focus the beam on to aperture and to transform the beam into parallel rays; aperture is used to limit the field of view and to limit the intensity of incoming radiation as per the choice of the operator. Note that the smaller the aperture size, the lesser self apodization (interference of light emerging from the aperture due to finite aperture size); Filters are used to limit the wavenumber range as per our interest. Note that filters are responsible for channelling, and any leakage of the filter can be seen on baseline of the measured spectrum. The spectral coverage of these filters have overlaps in order to guarantee the possibility of retrieval using lines in one extreme of the filter coverage. Our IFS 120M instrument has two sets of filter wheels each has seven separate filters. The system is also equipped with six narrow-band optical interference filters that are widely used in the INDACC InfraRed Working Group (IRWG). Table 2.1 summarizes trace gases commonly measured with this filter set, both detectors and the KBr beamsplitter. The beam splitter is a core of the interferometer and it is very

Table 2.1: Spectral Coverage of the IFS 120M instrument at Addis Ababa site

Filter	Wavenumber (cm^{-1})	Wavelength (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	H_2O, C_2H_2, HCN ...
3	2400-3100	3.3-4.1	O_3, HCl, N_2O, CH_4, 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, $ClONO_2$, HNO_3, N_2O, CH_4, C_2H_2, C_2H_4, CFCs ...

expensive and highly sensitive to the environment. The role of beam splitter is to split the incoming beam into two parts: the one transmitted goes to movable mirror (also called Retroreflectror) and the reflected one goes to fixed mirror, as shown in Fig. 2.10. The

two beams are then recombined at the beam splitter after moving different path lengths. Then the recombined beam goes to a detector. There are two beam splitters in our FTIR room, CaF_2 and KBr. KBr is highly hygroscopic and it is recommended not to use during rainy season and put it in a dry, warm places. It is kept in a beam splitter storage box to protect it from exposure to humidity. Fig. 2.10 shows our IFS 120M instrument optical setup.

Figure 2.10: The IFS 120M instrument optical set up in Addis Ababa.

2.3.1.3 Detectors

Eventhough detectors are part of the optical devices, we better introduce it independently. Detector is very important component of FTIR. It has analogues function as eyes for human beings. We use two detectors to cover the mid infrared (MIR) region . The two detectors are made of mercuric cadmium telluride (HgCdTe or MCT) and indium antimonide (InAs) alloys. The InAs detector covers a spectral range of (1,850-9,600) cm^{-1} and the MCT covers (750 - 5000) cm^{-1} while both effectively operate at a temperature $\leq 70K$ attained by filling the detector dewar by liquid nitrogen (LN_2). In our IFS 120M instrument, we usually use InAs detector to cover spectral range from (1800-4200) cm^{-1} using five filters, and MCT for (780 -2000) cm^{-1}. Both types of detectors operated in Photo-voltaic principle while the InAs is a linear detector. That is, both the MCT and

InAs elements absorb IR photons, and as a result electrons are promoted from the valence band to conduction band, and give rise to an electrical current. The current is then a direct measure of infrared (IR) intensity which increases with the number of IR photons. The disadvantage of MCT over InAs is its easy saturation, which results in non-linearity of detector response for incoming radiation.

2.3.1.4 Scanner

The moving mirror is the only moving part in FTIR spectrometer. The mirror, driving cables and two driving units constitute the main part of FTIR scanner. The two driving units of the scanner are 1 DC motor and 1 linear drive unit. The two driving units are also called the outer and inner scanners. The outer scanner ensures the gliding movement of scanner, and the main function of the inner scanner is to compensate the fluctuations of the outer scanner, same purpose as fine tuning in radio. The position of the moving mirror is precisely controlled by stabilized HeNe laser signal (631.8 nm), so that the optical path difference can be measured accurately. The laser signal is also used to determine the data sampling positions. In our IFS 120M we get more than 23000 data points from 0 to 4000 wavenumber intervals. If the scanner moves with uniform speed, the sinusoidal signal is generated as a result of modulation of HeNe laser, and this generated signal reaches the detector (either InAs or MCT). The intensity of the interfering beams (after reflection from the fixed and moving retroreflectors and recombined at the beam splitter) is measured as a function of optical path difference (OPD), and the result is an interference pattern, called interferogram. Note that, the beam splitter, retroreflectors, and detector constitute the main components of interferometer. The angle of incidence of the beam to the interferometer is 30^0. This narrow angle of incidence reduces polarization effects and uses the beam splitter more efficiently.

2.3.1.5 Internal Source

The internal source in FTIR is used to provide IR radiation to the spectrometer. Our instrument uses globar (glow bar), which is a peace of ceramic usually made of silicon carbide rod. When electric current passes through it, the resistance of the device makes it heat up and give off IR radiation. The globar may heat above 1400 K, so that we can get more IR radiation, which in turn results in higher throughput and lower noise in our FTIR spectra. The globar compartment is supplied with a closed-loop cooling system, where the cooling unit accommodate about half-a-litter of water for this purpose. The cooling unit must be filled with distilled or deionized water only; and it is recommended to change the water at least once every three months. The globar is mainly used for the calibration of the FTIR by means of a black body, and also to align optical device to get maximum radiation at the detector. In order to determine the instrumental line shape (ILS), which tells us the status of the instrument, we take black body radiation with globar source and the radiation passes through a low-pressure gas cell filled with HBr, which has sharp absorption lines.

2.3.1.6 Electronics Unit

The electronics unit of an FTIR spectrometer is used to convert optical signals into electrical signals. The intensity of IR photons reaching the detector is converted into electrical current by the electronics unit and send to monitor (PC). The electronics unit is also used to calibrate the IR wavenumber with respect to the laser wavenumber. Some of the jobs carried out by electronics units is the analog-to-digital conversion (ADC), phase correction, pre-amplification of the raw data (Interferogram) and automatic component recognition (ACR). The IFS 120M instrument is controlled by a PC data system, provided the Microsoft windows^{TM} operating system and the optical user software (OPUS^{TM}), which is a spectrometer software, installed on PC to take measurement, and further analysis of

the spectra. The Internet connection allows to control status of the spectrometer components via Internet. The interference between the two secondary beams (Interferogram) is digitized and sent to PC where the Fourier transformation (FT) takes place using OPUS software to get the final infrared spectrum for analysis.

2.3.2 FTS Characterization

The alignment and performance of an FTS is characterized by its ILS. 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. Bernardo and Griffith 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 FTS has been monitored on a regular basis by means of calibrated low-pressure gas cell measurements [3]. The cells used to derive the instrumental line shape are HCl and HBr cells. Fig. 2.11 shows the cells used to derive instrumental line shape. HBr gas cell is used to check for the well being of the our instrument. Using the globar

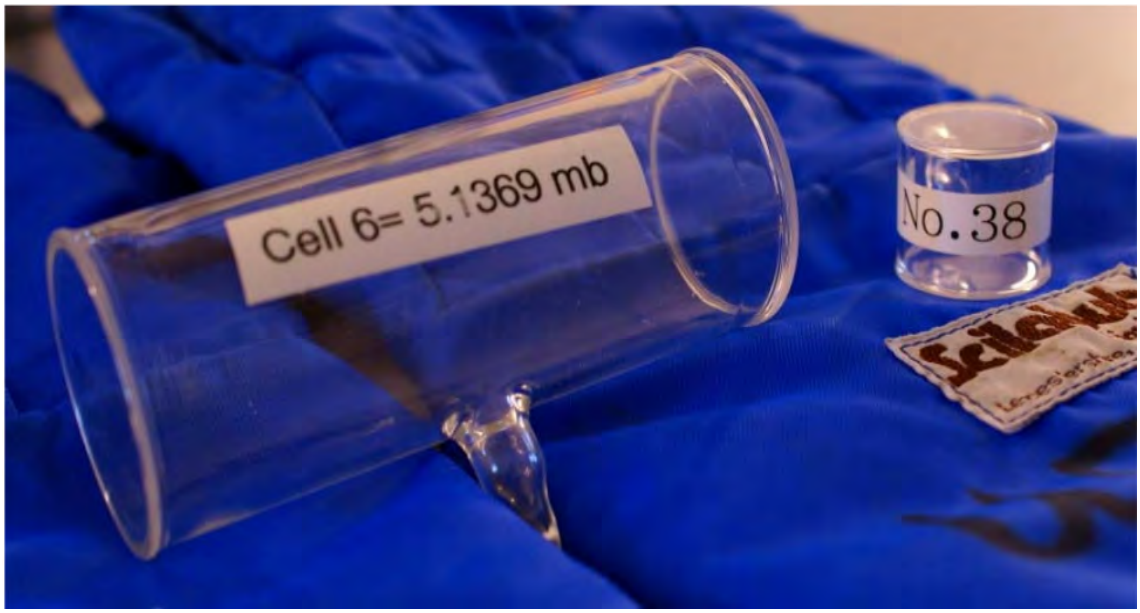

Figure 2.11: Cells used to derive ILS.

as a source of IR radiation, up to 100 scans are coadded to get spectra of both with and without the presence of HBr cell. The HBr spectra is rationed with the background to find the absorptance of HBr. LINEFIT software uses this result as an input to calculate modulation efficiency and phase error.

2.3.2.1 LINEFIT

LINEFIT is a software that was developed by Hase to calculate and retrieve the instrumental line shape (ILS) using gas cell measurements. The description of retrieval procedure is that the user has to choose between either simple parameter set with a number of parameters $2 + 2n_w$, or extended parameter set with a number of parameters $40 + 2n_w$, n_w denotes as spectral windows. Simple set and extended set are coupled by adjustable regularization constraint. The simple set consists of a constant phase error and the modulation efficiency versus ideal instrument, which is supposed to decline linearly with the optical path difference. In the case of the simple parameter set, the problem is well-posed and does not require additional constraints and the improved values f^{i+1} are

$$f^{i+1} = f^i + (A^{iT}A^i)^{-1}A^{iT}\Delta L^i \tag{2.3.1}$$

where

$f^i \Rightarrow$ Actual parameters, $\Delta L^i \Longrightarrow$ The residual between measured and calculated spectrum, $A^i \Longrightarrow$ Jacobian matrix and T $\Longrightarrow$ Denotes for transpose.

After a couple of iterations, the parameter values that minimize $(\Delta L)^2$ are found. The extended set consists of 20 values of both modulation efficiency versus ideal instrument and phase error at positions equidistant in the interferogram domain. In this case the problem is ill-posed and additional regularization constraints are needed. These constraints can be written as $B_m\mathrm{f} = 0$ and $B_p\mathrm{f} = 0$.

B_m and B_p are matrices that connect adjacent components in the subspaces spanned by the modulation and phase parameters, the subscripts m and p denote modulation and

phase respectively. Now

$$f^{i+1} = (A^{iT}A^i + \Upsilon_m^2 B_m^T B_m + \Upsilon_p^2 B_p^T B_P)^{-1} A^{iT} x(\Delta L^i + A^i f^i) \tag{2.3.2}$$

where Υ_m and Υ_p are constants and so-called regularization parameters that weight the smoothing conditions, and the functional, $(\Delta L)^2 + \Upsilon_m^2 (B_m f)^2 + \Upsilon_p^2 (B_p f)^2$is minimized. An ideal instrument has instrumental line shape (ILS), this ILS is affected by modulation loss due to self-apodization of the interferogram taking into account symmetric and a finite field of view. ILS is brodened due to a modulation loss and asymmetric due to a phase error. ILS is asymmetric/symmetric, and is broadened/weakened due to the difference between ideal and real instruments, such as the assumptions of finite entrance aperture size or finite optical path difference [23].

Chapter 3

Trace Gas Retrievals

In the previous chapter we have described how solar absorption spectra arise and how they are recorded with a Fourier transform infrared spectrometer. Furthermore, we have 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 the measured spectra, PROFFIT allows to choose between two regularization methods: the Optimal Estimation (OE) and the Tikhonov-Phillips (TP) Methods, 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 direct 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 [40] 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 ~ 41 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, x_a, of the n-dimensional state vector, x (vertical profile of the trace gas of interest) is

combined with the m-dimensional measurement vector, y (high-resolution measurements of a spectral absorption feature) as a weighted mean. The (nonlinear) forward model F(x, b) relates the state vector to the measurement vector, given some model parameters, b

$$y = F(x, b) + \epsilon = F(x_a, b) + K(x - x_a) + \epsilon \tag{3.1.1}$$

where ϵ is the random spectral measurement noise and $K = \partial F(x, b)/\partial x$, i.e. the linearization of F(x, b) about some reference state, here taken to be x_a. $K_{ij} = \partial F_i(x, b)/\partial x_j$ is an element of an $m \times n$ matrix of derivatives of a forward model element with respect to a state vector element. 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 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 $n \backsimeq 41$ 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 K, respectively. When $p < n$, the row space of K does not span all of state space, and the column space of 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 K, as it contributes nothing to a measurement. According to Rodgers [40] "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 K by performing a Singular Value Decomposition (SVD) of K

$$K = U\Lambda V^T \tag{3.1.2}$$

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

columns of 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 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 Eq. 3.1.1 to include a linearization about a set of model parameters b_a we write

$$y - F(x_a, b_a) = \frac{\partial F}{\partial x}(x - x_a) + \frac{\partial F}{\partial b}(b - b_a) + \epsilon = K(x - x_a) + K_b(b - b_a) + \epsilon \quad (3.1.3)$$

where K represents the sensitivity of the forward model to the true state and K_b represents the sensitivity of the forward model to its parameters. 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 Aldona and Rodgers [6, 40], we also define an inverse model, 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, $\hat{x}$

$$\hat{x} = I(y, b). \quad (3.1.4)$$

We linearize this inverse model about the measurement expected given x_a what Rodgers [40] refers to as an a priori measurement, y_a and the same set of model parameters

$$\hat{x} - I(y_a, b_a) = \frac{\partial I}{\partial y}(y - y_a) + \frac{\partial I}{\partial b}(b - b_a) \equiv G(y - y_a) + G_b(b - b_b). \quad (3.1.5)$$

The rows of the $n \times m$ gain matrix, 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, T:

$$\hat{x} = I(F(x,b),b) = T(x,b), \tag{3.1.6}$$

which must also be linearized about the a priori state

$$\hat{x} - T(x_a,b) = \frac{\partial T}{\partial x}(x - x_a) = \frac{\partial T}{\partial F}\frac{\partial F}{\partial x}(x - x_a) = GK(x - x_a). \tag{3.1.7}$$

If properly designed, a retrieval method will be unbiased and return the a priori when the input measurements are simulated from the a priori ($T(x_a,b) = x_a$), allowing us to finally write

$$\hat{x} - x_a = GK(x - x_a) \equiv A(x - x_a). \tag{3.1.8}$$

The rows of the $n \times n$ averaging kernel matrix, A, give the sensitivity of the retrieved state to the true state, by definition,

$$A = GK = \frac{\partial I}{\partial y}\frac{\partial F}{\partial x} = \frac{\partial \hat{x}}{\partial x}, \tag{3.1.9}$$

and are also used to calculate the sensitivity of the retrieval to the a priori, which follows from Eq. 3.1.8:

$$\frac{\partial \hat{x}}{\partial x_a} = I_n - A. \tag{3.1.10}$$

In order to find the form of G and the best estimate of the state, $\hat{x}$, we must now consider the details of the error statistics of the a priori and the measurement, since these errors will weigh the relative contribution of prior and measurement information to the solution. The average and variance of a scalar, y, are defined as

$$\bar{y} = \int yP(y)dy, \qquad \sigma^2 = \int (y - \bar{y})^2 P(y)dy \tag{3.1.11}$$

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

$$P(y) = \frac{1}{(2\pi)^{1/2}\sigma} exp\left[\frac{-(y-\bar{y})^2}{2\sigma^2}\right]. \tag{3.1.12}$$

When y is a vector quantity we write the vector PDF, P(y), as

$$P(y) = \frac{1}{(2\pi)^{m/2}|S_\epsilon|^{1/2}} exp\left[\frac{-1}{2}(y-\bar{y})^T {S_\epsilon}^{-1}(y-\bar{y})\right], \tag{3.1.13}$$

where S_ϵ is the matrix covariance of an ensemble of measurements (y_k) about the average measurement (y), or the measurement noise covariance matrix, whose elements are defined as

$$S_\epsilon(i,j) = \varepsilon[(y_i - \bar{y}_i)(y_j - \bar{y}_j)] \neq 0. \tag{3.1.14}$$

In Eq. 3.1.14, ε represents the expectation value operator ($\varepsilon(q) = \int qp(q)dq$), 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 $y_i - \bar{y}_i$ the i^{th} random measurement error ϵ_i. Switching to vector notation, we also define the matrix covariance of an ensemble of true states, x_i, about the a priori, i.e. the a priori covariance matrix, S_a, as

$$S_a = \varepsilon[(x - x_a)(x - x_a)^T]. \tag{3.1.15}$$

According to Rodgers [40], "it is less realistic, but convenient, to describe prior knowledge of x by a Gaussian PDF"

$$P(x) = \frac{1}{(2\pi)^{n/2}|S_a|^{1/2}} exp\left[\frac{-1}{2}(x-\bar{x})^T {S_a}^{-1}(x-\bar{x})\right]. \tag{3.1.16}$$

Although we do not possess knowledge of the true state, starting from Eq. 3.1.13 we can write the conditional probability of the measurement given a certain true state as

$$P(y|x) = \frac{1}{(2\pi)^{m/2}|S_\epsilon|^{1/2}} exp\left[\frac{-1}{2}(y-Kx)^T {S_\epsilon}^{-1}(y-Kx)\right]. \tag{3.1.17}$$

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

$$P(y|x) = \frac{P(x,y)}{P(x)}, \qquad P(x|y) = \frac{P(x,y)}{P(y)}, \tag{3.1.18}$$

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

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)}. \tag{3.1.19}$$

If we substitute Eq.s 3.1.13, 3.1.16 and 3.1.17 into Eq. 3.1.19 and take the logarithm of the resulting expression we can finally write

$$-2lnP(x|y) = (y - Kx)^T S_\epsilon^{-1}(y - Kx) + (x - x_a)^T S_a^{-1}(x - x_a) + c_1, \tag{3.1.20}$$

where P(y), which does not depend on x, has been incorporated into the constant c_1, along with the normalizing factors of $P(x|y)$ and P(x). Eq. 3.1.20 is quadratic in x, and hence it is possible to write it as

$$-2lnP(x|y) = (x - \hat{x})^T \hat{S}^{-1}(x - \hat{x}) + c_2. \tag{3.1.21}$$

By comparing the terms quadratic in x between Eq.s 3.1.20 and 3.1.21 Rodgers [40] shows that the covariance, $\hat{S}$, of the expected value, $\hat{x}$, is given by

$$\hat{S} = (K^T S_\epsilon^{-1} K + S_a^{-1})^{-1}, \tag{3.1.22}$$

while the expected value itself is given by equating terms that are linear in x^T

$$\hat{x} = (S_a^{-1} + K^T S_\epsilon^{-1} K)^{-1}(S_a^{-1} x_a + K^T S_\epsilon^{-1} y) \tag{3.1.23}$$

$$= x_a + (S_a^{-1} + K^T S_\epsilon^{-1} K)^{-1} K^T S_\epsilon^{-1}(y - K x_a) \tag{3.1.24}$$

$$= x_a + S_a K^T (K S_a K^T + S_\epsilon)^{-1}(y - K x_a). \tag{3.1.25}$$

Eq. 3.1.23 shows that the expected value of the posterior PDF is indeed the weighted sum of measurements and prior information. Eq.s 3.1.24 and 3.1.25 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 $\hat{x}$. Furthermore, starting from Eq. 3.1.24 it is trivial to show that G must be equal to

$$G = (S_a^{-1} + K^T S_\epsilon^{-1})^{-1} K^T S_\epsilon^{-1} = S_a K^T (K S_a K^T + S_\epsilon)^{-1}, \tag{3.1.26}$$

thus allowing us to calculate A from Eq. 3.1.9. Finally, we note that the solution to our inverse problem is the posterior distribution, $P(x|y)$, of which $\hat{x}$ is the expected value and $\hat{S}$ is the covariance.

3.1.2 Optimal vs Suboptimal Approaches

While the Bayesian approach provides us with a complete posterior PDF of the state vector, $P(x|y)$, we need to select only one solution logically, it should be one that optimizes something. Eq.s 3.1.23-3.1.25 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, $\hat{S}$, 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 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 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.

3.1.3 Nonlinear Problems

In deriving the optimal solution and its error covariance (Eq.s 3.1.22 and 3.1.23-3.1.25) 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 ($\bigtriangledown$) of Eq. 3.1.20 with respect to x and equating it to zero in order to find $\hat{x}$:

$$\bigtriangledown_x(-2lnP(x|y)) = -(\bigtriangledown_x F(x))^T S_\epsilon^{-1}(y - F(x)) + S_a^{-1}(x - x_a) = 0. \tag{3.1.27}$$

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

$$-(K(\hat{x}))^T S_\epsilon^{-1}(y - F(\hat{x})) + S_a^{-1}(\hat{x} - x_a) = 0, \tag{3.1.28}$$

which is an implicit equation for $\hat{x}$ that must be solved numerically. The right hand side of Eq. 3.1.27 is the gradient of the cost function, g(x), which is again differentiated in the vector-equivalent of Newton's method

$$x_{i+1} = x_i - (\bigtriangledown_x g(x_i))^{-1} g(x_i), \tag{3.1.29}$$

where the second derivative of the cost function, $\bigtriangledown_x g$, follows from Eq. 3.1.27:

$$\bigtriangledown_x g = S_a^{-1} + K^T S_\epsilon^{-1} K - (\bigtriangledown_x K^T) S_\epsilon^{-1}(y - F(x)). \tag{3.1.30}$$

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 Eq. 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 Eq.s 3.1.23-3.1.25

$$x_{i+1} = x_i + (S_a^{-1} + K_i^T S_\epsilon^{-1} K_i)^{-1}\{K_i^T S_\epsilon^{-1}(y - y_i)) - S_a^{-1}(x_i - x_a)\} \tag{3.1.31}$$

$$= x_i + (S_a^{-1} + K_i^T S_\epsilon^{-1} K_i)^{-1} K_i^T S_\epsilon^{-1}\{(y - y_i) + K_i(x_i - x_a)\} \tag{3.1.32}$$

$$= x_a + S_a K_i^T (K_i S_a K_i^T + S_\epsilon)^{-1}\{(y - y_i) + K_i(x_i - x_a)\}, \tag{3.1.33}$$

where K_i is taken to mean K evaluated at x_i and $y_i = F(x_i)$. Neglecting the $\bigtriangledown_x K^T$ term in Eq. 3.1.30 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 $\hat{x}$ is found

$$\hat{S} = (\hat{K}^T S_\epsilon^{-1} \hat{K} + S_a^{-1})^{-1}, \tag{3.1.34}$$

and is the inverse of the Hessian (Eq. 3.1.30) when we neglect the $\bigtriangledown_x K^T$ term ($\hat{K}$ means "K evaluated at $\hat{x}$ "). The averaging kernel matrix is also obtained in an analogous manner as $\hat{A} = \hat{G}\hat{K}$.

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 analyze how different sources of error propagate to the retrieved solution. Both sections follow the general approach of Rodgers [40], 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 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

$$\tilde{x} = S_a^{-1/2} x \qquad \tilde{y} = S_\epsilon^{-1/2} y \qquad \tilde{\epsilon} = S_\epsilon^{-1/2} \epsilon, \tag{3.2.1}$$

then

$$\tilde{y} = S_\epsilon^{-1/2} K S_a^{1/2} \tilde{x} + \tilde{\epsilon} \equiv \tilde{K}\tilde{x} + \tilde{\epsilon}. \tag{3.2.2}$$

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

$$S_{\tilde{\epsilon}} = \varepsilon(\tilde{\epsilon}\tilde{\epsilon}^T) = S_\epsilon^{-1/2} \varepsilon(\tilde{\epsilon}\tilde{\epsilon}^T) S_\epsilon^{-1/2} = S_\epsilon^{-1/2} S_\epsilon S_\epsilon^{-1/2} = I_m. \tag{3.2.3}$$

For this reason the above transformation of y is referred to as pre-whitening because it turns noise into white noise. From Eq. 3.1.36, we have that the prior covariance of the pre-whitened measurement vector is given by

$$S_{\tilde{y}} = \varepsilon(\tilde{y}\tilde{y}^T) = \varepsilon[(\tilde{K}\tilde{X} + \tilde{\epsilon})(\tilde{K}\tilde{X} + \tilde{\epsilon})^T] = \tilde{K}\tilde{K}^T + I_m \tag{3.2.4}$$

According to Eq. 3.1.37, I_m is the component of the covariance due to measurement noise, and hence $\tilde{K}\tilde{K}^T$ must be the component of the covariance due to the variability of the

state. However, $\tilde{K}\tilde{K}^T$ is not normally diagonal, and one final transformation is necessary before we can compare it to I_m. Using the SVD of $\tilde{K} = U\Lambda V^T$, we define

$$y' = U^T\tilde{y}, \qquad x' = V^T\tilde{x}, \qquad \epsilon' = U^T\tilde{\epsilon}, \tag{3.2.5}$$

to give a fully diagonal forward model equation

$$y' = \Lambda x' + \epsilon'. \tag{3.2.6}$$

The covariances of ϵ' and x' remain equal to unity (now of dimension I_p), because both U^TU and V^TV are equal to I_p, and finally, we can see that the covariance of y', $S_{y'}$, is equal to the diagonal matrix $\Lambda^2 + I_p$. The diagonal elements of Λ^2 can now be compared to the diagonal elements of the transformed measurement noise covariance (I_p). Elements of y' which vary more than the noise correspond to $\lambda_i^2 \backsimeq 1$, where λ_i represent the non-zero singular values of $\tilde{K}$. 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 $S_\epsilon^{-1/2} K S_a^{1/2}$ which are greater than about unity,"[40].

In order to make the concept of information content more precise, Rodgers [40] defines the degrees of freedom for signal, d_s, and the degrees of freedom for noise, d_n as follows:

$$d_s = \varepsilon[(\tilde{x} - x_a)^T S_a^{-1}(\tilde{x} - x_a)], \tag{3.2.7}$$

$$d_n = \varepsilon[(y - K\tilde{x})^T S_\epsilon^{-1}(y - K\tilde{x})]. \tag{3.2.8}$$

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

$$\chi^2 = (y - Kx)^T S_\epsilon^{-1}(y - Kx) + (x - x_a)^T S_a^{-1}(x - x_a), \tag{3.2.9}$$

with $y - Kx = \epsilon$. 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 x and y, or by the transformed variables x' and y':

$$d_s = tr[(S_a^{-1} + K^T S_\epsilon^{-1} K)^{-1} K^T S_\epsilon^{-1} K] = tr[\Lambda^2(\Lambda^2 + I_m)^{-1}] = \sum_{i=1}^{m} \frac{\lambda_i^2}{1+\lambda_i^2} \qquad (3.2.10)$$

$$d_n = tr[(K^T S_\epsilon^{-1} K + S_a^{-1})^{-1} S_a^{-1}] + m - n = tr[(\Lambda^2 + I_m)^{-1}] = \sum_{i=1}^{m} \frac{1}{1+\lambda_i^2}, \qquad (3.2.11)$$

where $tr(A)$ represents the mathematical trace operator. Note that Λ is now an $m \times m$ matrix (because it now also includes the $m-p$ singular values of $\tilde{K}$ that are equal to zero), such that $d_s + d_n = m$ in both formulations. Furthermore, we note from Eq. 3.2.10 that $d_s = tr(A)$, which hints at a relationship between the eigenvalues and eigenvectors of A and the singular values and singular vectors of $\tilde{K}$. If we consider the eigenvalue decomposition of $A = R\Delta L^T$, where R and L contain the right and left eigenvectors, respectively, and Δ is the diagonal matrix of the eigenvalues of A, then it can be shown that $R = S_a^{1/2} V, L = S_a^{-1/2} V$, and $\Delta = \Lambda^2(\Lambda^2 + I_p)^{-1}$. Since the eigenvalues of 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 A that are of order 1.

In characterizing the information content of the retrievals, we take the trace of 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 INDACC; 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 the retrieval characterization for a suite of trace gases are discussed in chapter 4.

3.2.1.1 Other Definitions of Information Content

In the linear Gaussian case, the inverse covariance matrix $\hat{S}^{-1}$ 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 $(S_a^{-1} + K^T S_\epsilon^{-1} K)$.

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 [40], 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

$$H = \frac{1}{2} ln|\hat{S}^{-1} S_a| = -\frac{1}{2} ln|I_n - A| = \frac{1}{2} \sum_{i=1}^{m} ln(1 + \lambda_i^2). \tag{3.2.12}$$

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 [40] 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 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 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 [40], 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 the retrievals.

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

$$\hat{x} = x_a + A(x - x_a). \tag{3.2.13}$$

If a measurement system resolves each element of the true state vector with perfect sensitivity, 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 A are not unity, representing imperfect sensitivity to the true state at all heights, and off diagonal elements are present in all rows of A, representing the inability to independently resolve all elements of the true state vector.

Since $A_{ij} = \dfrac{\partial \tilde{x}_i}{\partial x_j}$, we can see that the i^{th} row of 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 A are referred to as smoothing functions, or averaging kernels, or resolution functions. On the other hand, the j^{th} column of 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 $\sim 1-4$ 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 [40], 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.

3.3 Retrieval Error Analysis

We resume the error analysis postponed in Section 3.1.1 and rewrite Eq. 3.1.6 as

$$\hat{x} = I(F(x,b) + \Delta f(x,b,b') + \epsilon, b) = T(x,b), \tag{3.3.1}$$

where we have now distinguished between the forward function f (x, b, b') and the forward model F(x, 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. b$'$ represents parameters of the forward function that have been ignored in constructing the approximate forward model (where F(x, b) is the best-known empirical approximation, b are unknown). It is assumed that F(x, b) ≃ f (x, b, b') and that Δ f = f (x, b, b')- F(x, b). ϵ is again the random measurement noise. Replacing the forward model in Eq. 3.3.1 by a linearization about x_a and b_a (Eq. 3.1.3), we have

$$\hat{x} = I[F(x_a,b_a) + K(x-x_a) + K_b(b-b_a) + \Delta f(x,b,b') + \epsilon, b_a], \tag{3.3.2}$$

which upon linearization about y_a finally yields

$$\hat{x} = I[F(x_a, b_a), b_a] + \frac{\partial I}{\partial F}[K(x - x_a) + K_b(b - b_a) + \Delta f(x, b, b') + \epsilon] \tag{3.3.3}$$

$$= x_a + GK(x - x_a) + GK_b(b - b_a) + G\Delta f(x, b, b') + G\epsilon \tag{3.3.4}$$

$$= x_a + A(x - x_a) + G\epsilon_y, \tag{3.3.5}$$

where we assume that the retrieval method does not introduce a bias $I[F(x_a, b_a), b_a] = x_a)$ and that ϵ_y now represents the total error due to forward model parameter errors, forward model error and random measurement noise ($\epsilon_y = K_b(b - b_a) + \Delta f(x, b, b') + \epsilon$). Eq. 3.3.4 can be rearranged as

$$\begin{aligned} \hat{x} - x = (A - I_n)(x - x_a) \quad & (smoothing\ error) \\ +GK_b(b - b_a) \ & (model\ parameter\ error) \\ +G\Delta f(x, b, b') \quad & (forward\ model\ error) \\ +G\epsilon \qquad\qquad & (retrieval\ noise) \end{aligned} \tag{3.3.6}$$

where $\hat{x} - x$ gives the error in $\hat{x}$, and can be grouped into the smoothing error and the retrieval error ($G\epsilon_y$), in turn comprised of the model parameter error, forward model error, and retrieval noise. We note that Rodgers [40] 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 Eq. 3.3.6, we obtain the error covariance matrices

$$S_s = (A - I)S_a(A - I)^T \tag{3.3.7}$$

$$S_f = GK_bS_bK_b^TG^T \tag{3.3.8}$$

$$S_m = GS_\epsilon G^T, \tag{3.3.9}$$

where $S_b = \varepsilon[(b - b_a)(b - b_a)^T]$, and where we have neglected the forward model error due to the complications in evaluating it without knowledge of x and b. S_s is the covariance

of the error caused by the smoothing of the true state by the averaging kernel, S_f is the covariance of the error due to inaccuracies in forward model parameters, and S_m is the covariance of the error due to random measurement noise propagating to the retrieval. Starting from the definition of S_s, S_m, and G, it is straightforward to show that $S_s + S_m = \hat{S}$, i.e. the total error (apart from model parameter error) is identical to the covariance of $P(x|y)$, as shown in Eq. 3.1.22. While S_m is purely random and S_f can contain both random (e.g. temperature, solar zenith angle) and systematic (ILS, spectroscopic parameters) components, S_s is a good example of an error that is systematic on short time scales, but random on long time scales. Rodgers [40] makes a critical point about S_s, i.e. that its estimate is only correct insofar as S_a represents the covariance of a real ensemble of atmospheric states, and 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

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 S_a, S_s, and S_m, and $\hat{S} = S_s + S_m$ on the same plot. 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, S, by solving its eigenvalue equation ($Sl_i = \lambda_i l_i$) for its eigenvalues (λ_i) and eigenvectors (l_i). Because a covariance matrix is symmetric, S can be decomposed into

$$S = \sum\nolimits_i \lambda_i l_i^T = \sum_i e_i e_i^T, \tag{3.3.10}$$

where the orthogonal vectors $e_i = \lambda_i^{\frac{1}{2}} l_i$ are what Rodgers [40] refers to as error patterns. This is useful because the error in the state vector, ϵ_x, can be written as the sum of the error patterns multiplied by a random factor a_i of unit variance

$$\epsilon_x = \sum_{i=1}^{n} a_i e_i \tag{3.3.11}$$

and because the retrieval contains only the a priori plus these error pattern shapes [40]. For more readers are advised to refer Aldona and Rodgers [6, 40].

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 S_ϵ. After the retrieval is performed, the contribution of the parameter in question to the retrieval error covariance is calculated as $GK_bS_bK_b^TG^T$. In the second option, the model parameter is not retrieved, but the retrieval is performed with an effective error covariance $K_bS_bK_b^T + S_\epsilon$. 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 x and b together in order to obtain the correct weighting functions [40]. In the analysis of the 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 are neither retrieved nor accounted for by an effective error covariance during a retrieval

(e.g. $K_b S_b K_b^T + S_\epsilon$, where S_T is a 38 by 38 temperature error covariance). However, their error contribution is estimated (post-retrieval) after calculating the $m \times 38$ temperature weighting function matrix by perturbation methods. The error analysis of the retrievals is presented for HCN and C_2H_6 in chapter 4.

3.4 Retrieval Algorithms

The above OEM formulation of Aldona and Rodgers [6, 40] is implemented semi-empirically in the SFIT-2 (the line-by-line nonlinear least-squares spectral fitting software) algorithm developed at NASA Langley and the New Zealand Institute for Water and Atmosphere (NIWA) [6]. 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 [40]. SFIT-2 has been successfully compared to another OEM retrieval algorithm (PROFFIT-9) developed by Hase et al. [32]. The retrieval code we used in this thesis is PROFFIT-9 developed by Hase. The main steps in the retrieval can be divided into two parts generally: first, the forward calculation (PROFFWD) which provides synthetic and measured spectrum and derivatives for error estimation. In this part, the software performs a radiative transfer calculation for a given set of relevant atmospheric (T, p, VMRs) and auxiliary (SZA, ILS) input quantities for the spectral microwindows selected for analysis. The derivative of the spectrum with respect to the target and auxiliary quantities are allowed to vary in the process of searching the best estimate of the atmospheric state; second, the inversion model (PROFFIT), which processes the PROFFWD model output and suggests improved solution, restarts forward model with updated variables, cycles until convergence is reached. In short, PROFFIT determines a best estimate of the observed atmospheric state by improving fit quality to recorded spectrum iteratively. PROFFIT allows to choose between two regularization methods: the Optimal Estimation (OE) and the Tikhonov-Phillips (TP) Methods. We have used the TP

method, which additionally enables to impose two kinds of constraints absolute regularization or slope regularization of a start profile and the retrieved profile. Using the TP approach, PROFFIT allows to perform the whole inversion procedure on a logarithmic scale. This avoids negative VMR values, and we employ it for all results presented here. It is possible to invert various profiles simultaneously by PROFFIT. Additionally, it allows a simple scaling of a given start profile (climatological profile), which is often done for weak interfering gases.

Chapter 4

Results and Discussion

In the following sections we provide an analysis of the observations of HCN and C_2H_6 recorded at Addis Ababa, Ethiopia with high spectral resolution Fourier transform infrared (FTIR) spectrometer during May 14, 2009 - February 07, 2011.

4.1 Spectral Signatures of HCN and C_2H_6 in MIR

The data analysis is based on the fact that trace gases absorb light at certain wavelengths, which are characteristics of their chemical bonds, leading to vibrational or rotational motions of the molecules. Information on the concentrations of the atmospheric species in the air volume is based on the amount of light they absorbe [2].

4.1.1 Microwindow Selection

Deriving information about the vertical distribution of trace gases out of high resolution FTIR spectra is possible because of the pressure broadening of the absorption lines, leading to an altitude dependence of the line shapes. While the line centers provide information about the higher altitudes of the distribution, the wings of a line provide information about the lower altitudes. Therefore the information content of the retrieval will strongly depend on the choice of the absorption lines. In retrieving a measured spectrum data microwindow selection is the first and an important task. Microwindows are set of narrow

Table 4.1: Microwindows and interfering species.

Target Gas	Microwindow (cm^{-1})	Main interfering species
HCN	3268.00-3268.38	H_2O, CO_2, N_2O
	3331.40-3331.80	
C_2H_6	2976.60-2977.10	H_2O, CH_4, O_3
	2996.70-2997.10	H_2O, CH_4, O_3
	3000.10-3000.60	CH_4, O_3

spectral interval that are selected as those that contain the best information on the target parameters and are less affected by systematic error, such as uncertain spectroscopic data, interference of non-target species, NLTE (Non Local Thermodynamic Equilibrium) and line mixing effects . For the retrieval we have used different microwindows in the spectral range of 3268.00 to 3331.80 cm^{-1} for hydrogen cyanide and from 2976.60 to 3000.60 cm^{-1} for ethane. The optimum microwindows are selected by a software called ALFIP (Automatic Line FInding Program) and INTDIF (INTerfering gases based on DIFferences) [41]. ALFIP calculates the spectra and finds the most suitable lines and optimized microwindow size and INTDIF uses the output of ALFIP and calculates the interfering gases; and the outcome is a number of optimized microwindows. An optimum microwindow is a spectral interval consisting of grid points selected by minimizing the retrieval error of the target quantity. We used 2 microwindows for the retrieval of HCN and 3 for C_2H_6, as shown in Table 4.1.

4.2 Observations of HCN and C_2H_6 from FTIR

In ground-based Fourier transform infrared spectroscopy, information about the vertical distribution of atmospheric trace gases is derived from well-resolved Lorentz-broadened solar absorption line shapes. The observation geometry of ground-based measurements, the maximum achievable spectral resolution and measurement SNR limits the vertical resolution to at best 5-10 km [42].

4.2.1 Error Analysis

PROFFIT allows error estimation. We categorized the errors into: systematic error and random (or statistical) error. As seen in Fig. 4.1, the errors in HCN and C_2H_6 VMRs vary as a function of altitude. The errors for HCN decreases with altitude while increases for C_2H_6. The estimated error values for some layers are summarized in Tables 4.2 - 4.5. In both cases random error is found to be greater than the systematic error for each layers. The statistical part of the error corresponds to the error provided by the fitting algorithm [43]. Most of the error has come from the baseline and the lowest error is from the LOS. The statistical error of the total column is 38.8% for HCN and 17.7% for C_2H_6. The total column is characterized by systematic error of 37.3% for HCN and 12% for C_2H_6. Here the total random as well as systematic error of C_2H_6 is less than that of Senten [44] which was 75.70% and 19.13% respectively.

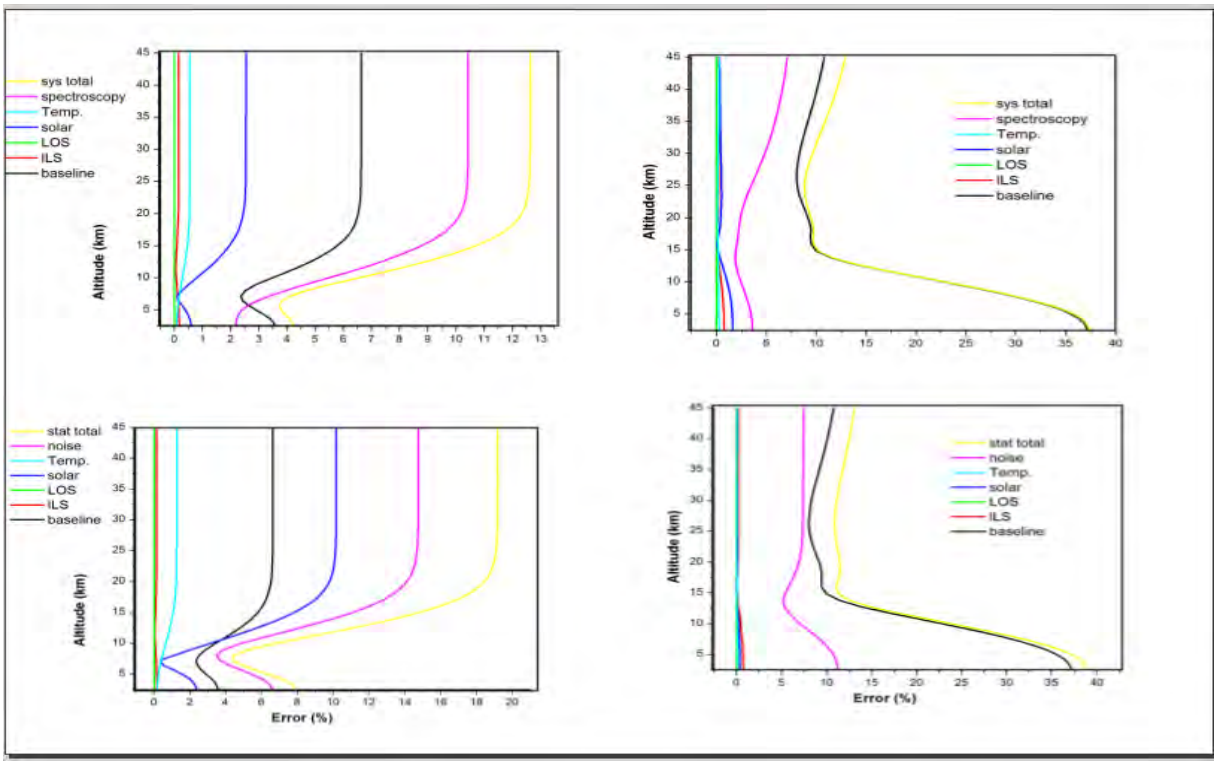

Figure 4.1: Altitude variation of systematic error and statistical error of HCN (right panel) and C_2H_6 (left panel).

Table 4.2: Systematic error budget for the retrieval of C_2H_6 VMR profile (%)

Height	Baseline	ILS	LOS	Solar	Temperature	Spectroscopy	Total sys.
2.45	3.57637	0.17122	0.00159	0.59285	0.08207	2.1881	4.23875
5.62	2.65284	0.15003	0.00155	0.28342	0.12638	2.59771	3.7282
9.78	3.35624	0.07128	0.00164	0.74656	0.28024	5.31886	6.34002
14.84	3.35624	0.07128	0.00164	0.74656	0.28024	5.31886	6.34002
18.31	6.41555	0.13136	0.00203	2.35934	0.52078	9.90135	12.04484

Table 4.3: Statistical error budget for the retrieval of C_2H_6 VMR profile (%)

Height	Baseline	ILS	LOS	Solar	Temperature	Noise	Total sta.
2.45	3.57637	0.17122	0.01428	2.37138	0.19156	6.60852	7.88424
5.62	2.65284	0.15003	0.01399	1.13349	0.29495	4.89176	5.68952
9.78	3.35624	0.07128	0.01473	2.98653	0.65379	4.69214	6.52831
14.84	5.75141	0.08868	0.01713	7.7357	1.0697	10.95286	14.62933
18.31	6.41555	0.13136	0.01831	9.43797	1.21495	13.52167	17.73692

Table 4.4: Systematic error budget for the retrieval of HCN VMR profile (%)

Height	Baseline	ILS	LOS	Solar	Temperature	Spectroscopy	Total sys.
2.45	37.1254	0.78499	0.01859	1.64795	0.26253	3.56997	37.32907
6.38	33.19744	0.70501	0.01859	1.51776	0.25775	3.24112	33.40128
10.72	20.27699	0.42595	0.01856	0.94969	0.23607	2.23855	20.42679
19.54	9.4427	0.12286	0.01869	0.33753	0.13772	2.22702	9.70718
23.44	8.29618	0.22134	0.01889	0.52777	0.1428	3.03098	8.85153

Table 4.5: Statistical error budget for the retrieval of HCN VMR profile (%)

Height	Baseline	ILS	LOS	Solar	Temperature	Noise	Total sta.
2.45	37.1254	0.78499	0.00207	0.41199	0.11251	11.23945	38.78382
6.38	33.19744	0.70501	0.00207	0.37944	0.11045	9.98544	34.68259
10.72	20.27699	0.42595	0.00206	0.23739	0.10119	6.42171	21.27473
19.54	9.27446	0.16103	0.00208	0.10745	0.05529	6.83848	11.52331
23.44	8.29618	0.22134	0.0021	0.13189	0.0612	7.26483	11.03136

Retrieval of HCN and C_2H_6 are performed following retrieval algorithm described in previous chapter (section 3.4). Fig.s 4.2 and 4.3 show the spectral fit between the measured and calculated spectra of these microwindows in Table 4.1. The figures show the measured and calculated spectra. Besides, the target species (HCN and C_2H_6) and the interfering species of Table 4.1 identified by INTDIF are shown. The difference between the measured and simulated spectra (the residuals in %) are shown in the upper panel of the figure for HCN while at the right panel for C_2H_6.

From the figures, the maximum percentage error for the spectral is less than 0.1% for

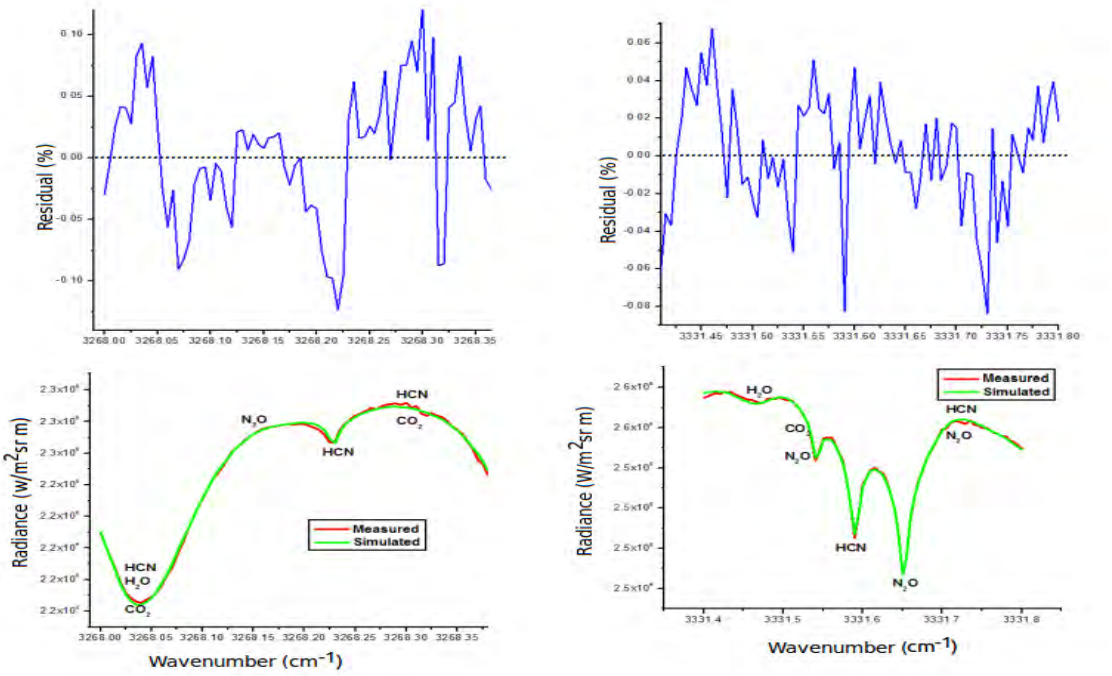

Figure 4.2: Spectral radiance of HCN.

HCN and 3.5% for C_2H_6. If we compare the residuals of the first microwindows of the gases with that of the results of Zhao et al.[1] which has a maximum residuals of 1% for both, C_2H_6 is comparable while our result for HCN is one tenth. The noise in the figures is dominantly random except some systematic features near the peak positions of the target and interfering species. These microwindows were also used by Walsh [45] and

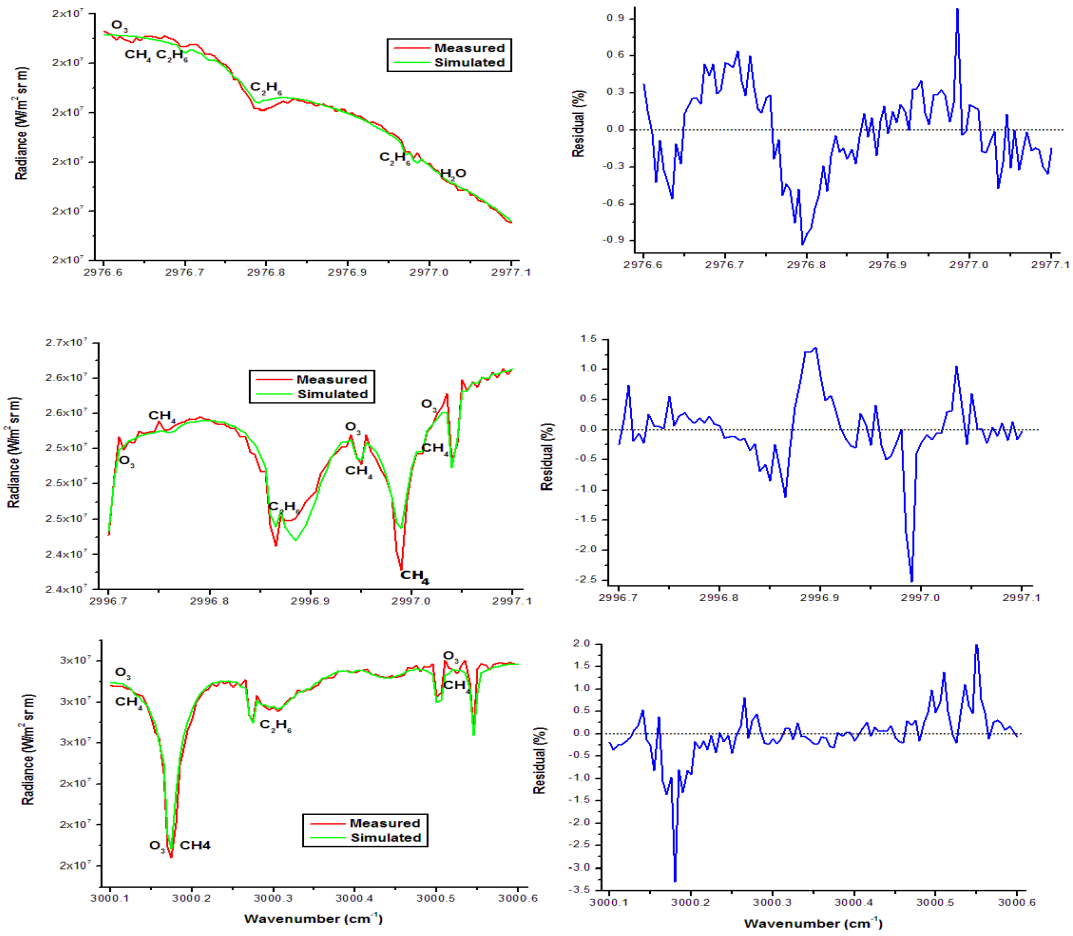

Figure 4.3: Spectral radiance of C_2H_6.

the fittings as well as the residuals are similar with ours. Hence all the results indicate that the misalignment is not much to cause significant error in our result.

4.2.2 The Retrieved Profiles of HCN and C_2H_6

Under this subsection the results of the retrieved profiles and the averaging kernels of the selected gases are presented. The time variation of C_2H_6 VMR at an altitude of 5.62 km measured on December 22, 2010 is also part of our discussion. Fig. 4.4 shows contour

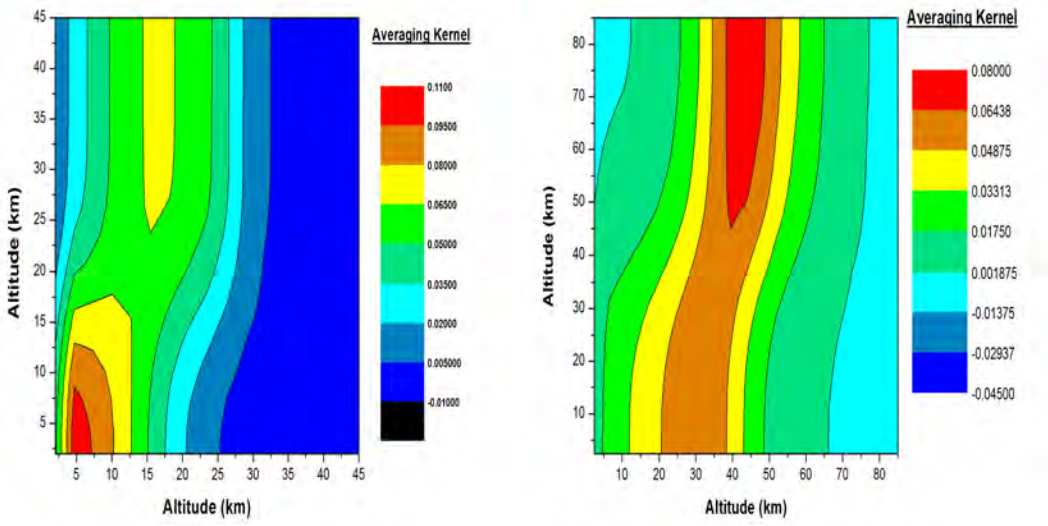

Figure 4.4: Averaging kernels of HCN (right panel) and C_2H_6 (left panel).

plots of the averaging kernel matrices of HCN (right panel) and C_2H_6 (left panel) which is the measure of the sensitivity of the retrieved state to the true state as given by Eq. 3.1.9. The plot represents a 34 by 34 (a 34 grid layers of 2.45 -45 km) matrix for C_2H_6 and a 41 by 41 (41 grid layers of 2.45 - 85 km) for HCN. C_2H_6 is more sensitive up to 8 km (which shows C_2H_6 is a tropospheric gas) while HCN more sensitive above 45 km. HCN is upper tropospheric and lower stratospheric gas of tracer of tropospheric biomass burning. The study by Glatthor et al.[46] assumed that HCN to be well mixed in the troposphere, slowly decreasing with altitude and lost by reaction with OH and O(^{1}D) in the stratosphere as well as by photolysis in the upper stratosphere, led to an estimated life time of 2.5 years. From the same study model calculations by Li et al. [47], who assumed biomass burning as the only source of HCN and ocean uptake as the only sink could explain the HCN distribution measured in the TRACE-P mission and led to an estimated tropospheric lifetime of 5.3 months. The lifetime of stratospheric HCN, where the main sink is reaction with OH, is still assumed to be several years. C_2H_6 is the second most abundant organic trace gas in the troposphere [19]. According to Rudolph [19] it has approximately two sources of equal importance, namely biomass burning and natural gas

losses. Xiao et al. [48] state fossil fuel production as the major source of C_2H_6, followed by biofuel and biomass burning. The tropospheric lifetime of C_2H_6 has been estimated at a few months by Rudolph and Ehhalt ([47], and reference therein) and at 2 months by Hough [49]. The main tropospheric loss process is reaction with OH, whereas the major stratospheric sink is reaction with Cl [50].

Fig. 4.5 is a plots of retrieved profiles of HCN (right panel) and C_2H_6 (left panel). From

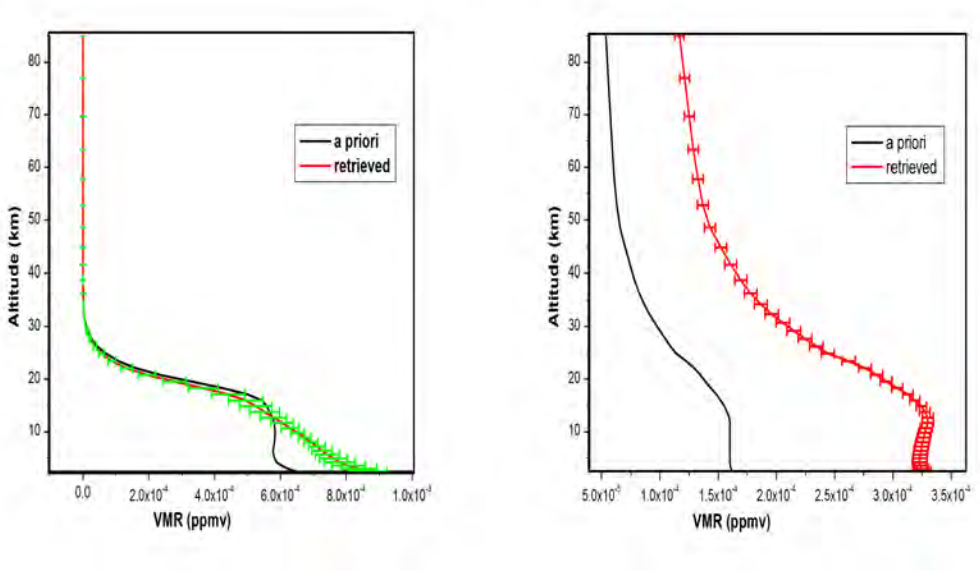

Figure 4.5: Retrieved profiles of HCN (right panel) and C_2H_6 (left panel).

this figure we observe that C_2H_6 shows a minor slope from 4×10^{-4} ppmv at nearly 13 km to 8.5×10^{-4} ppmv at the observation site, 2.45 km while HCN is around 3.75 ppmv at the same altitude. The profiles of both species show that the VMR rapidly decreasing with altitude. This shows both of these trace gases are concentrated at the lower atmosphere, troposphere. The standard deviations of the retrieved profiles are shown in the same figure as error bar having a magnitude between 4.51×10^{-6} at the observation site to 2.574×10^{-6} at 85 km for HCN while for C_2H_6, it is between 8.221×10^{-5} ppmv and 2.583×10^{-16} ppmv at respective altitudes. Here the error for HCN is greater than that of C_2H_6, but their degrees of freedom (DOFs) are also different (2 for HCN and 1.5 for C_2H_6). This indicates that there are about 2 independent pieces of information for HCN while 1.5 for C_2H_6 from

the measurement. The error can be minimized by increasing of the constraint but this decreases the degrees of freedom. The time variation of C_2H_6 VMR for a particular day

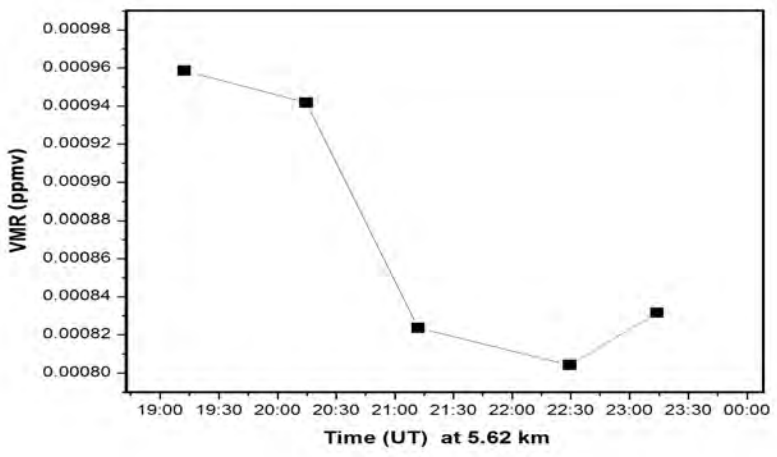

Figure 4.6: Variation of C_2H_6 VMR with time.

is shown in Fig. 4.6. The measurement was taken on December 22, 2010 starting from morning 4:00 a.m to 8:00 p.m local time. It is shown that the VMR of C_2H_6 decreases with time from late morning to just noon and tends to increases right after noon. From this result and Eq. 1.1.9 we found that in addition to reactions with OH, C_2H_6 also varies due to photolysis. Transportation of airmass is an other means for its variation (sec. 4.3).

4.2.3 Tropospheric Time Series of HCN and C_2H_6

The data at hand and considered for this study is limited to short measurement periods, however we tried to see the time series of HCN and C_2H_6 for 106 and 118 measurement days with a total of 154 and 181 measured spectra respectively since the first date of continuous operation of our FTIR spectrometer. We presented the time series of the total column amounts and vertical profiles of HCN and C_2H_6 from May 2009 to February 2011 and the time series of total column HCN and C_2H_6 for January 2010, which has better number of measurement days. Since measurements are only possible on sunny days with

clear sky, there are periods without measurement due to bad weather, restricting the number of observation days. The limited number of measurements within this period is due to unfordable natural phenomena of the rainy season (from middle of June to late September) and unexpected cloud cover over Addis Ababa which severely limited the observing capability of our FTS in addition to technical problems of the instrument. Generally, the tropospheric lifetimes of these trace gases depend on the seasons and latitudes with global averages of approximately 1 and between 2-4 months or several years for C_2H_6 and HCN respectively. Their seasonal cycles are mainly controlled by changes in biomass burning

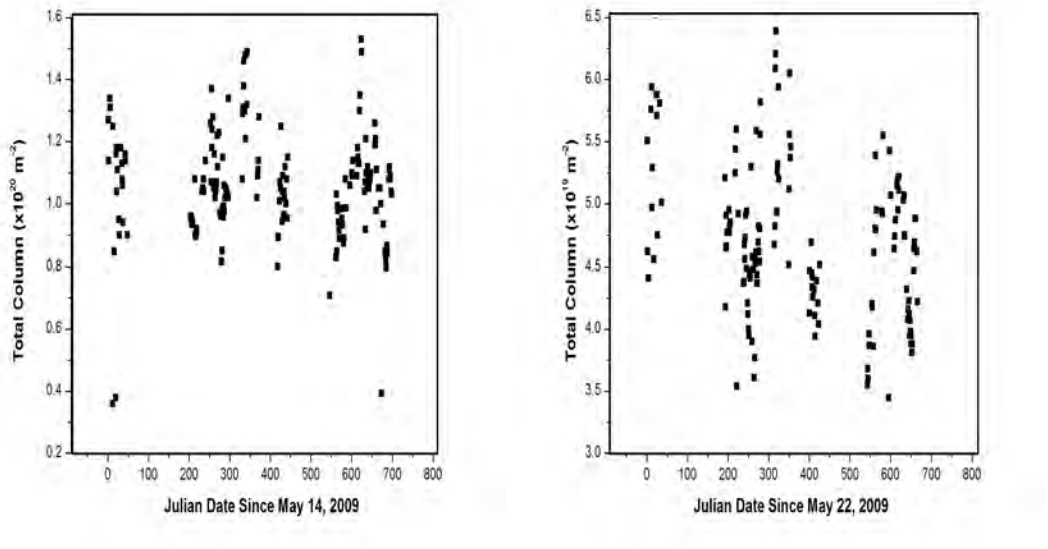

Figure 4.7: Time series of the total columns of HCN (right panel) and C_2H_6 (left panel).

emissions and OH concentrations due to the influence of solar flux. However, estimates of the roles of HCN in atmospheric chemistry, biogeochemistry, global sources, sinks and the lifetime are still uncertain due to insufficient data [2]. Fig. 4.7 shows the time series of the total columns of HCN and C_2H_6 (in molecules/m^2) observed during the measurement period from May, 2009 to February, 2011. The blank space in between is when no measurement was taken. The total column amount of HCN observed during this period ranges from 9.30×10^{18} to 6.39×10^{19} while that of C_2H_6 from 3.60×10^{19} to 1.70×10^{20} molecules/m^2. The total column amount of both HCN and C_2H_6 is maximum in March

which is reasonable to say that their source is really biomass burning. Because in March there is biomass burning due to agricultural activity. C_2H_6 also has a maximum value in December which is not due to biomass burning but rather unexpected noise during measurement (like in Fig. 2.2 specially filters E and F). Such measurement was also done over different countries (like Bremen and Japan) and the seasonal variations of tropospheric total column amounts of HCN are found to be from 3×10^{20} to 7.5×10^{20}, $6.57\pm0.84\times10^{19}$ and for C_2H_6 from 5×10^{21} to 3.5×10^{22} molecules/m^2 respectively [1, 2]. From these results we have observed that the total column amount of our study over Addis Ababa is almost comparable with that over those countries. This indicates that there is much biomass burning sources over Addis Ababa. Of course in Ethiopia agriculture is means of income for most (>85% of the people) which is the major cause of biomass burning. So high biomass burning results of such type are expected. The contour plot in Fig. 4.8 shows the

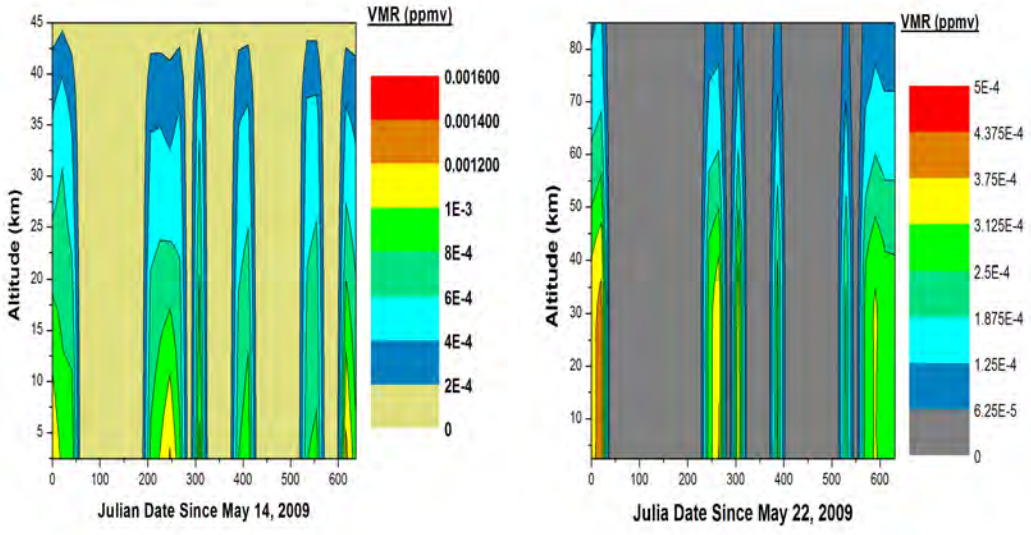

Figure 4.8: Time series of the vertical profiles of HCN (right panel) and C_2H_6 (left panel).

time series of the vertical profiles of HCN (right panel) and C_2H_6 (left panel). The dark yellow color for C_2H_6 and dark gray for HCN indicates the non-measurement days and/or zero VMR values. As indicated in this figure the mixing ratio reaches up to 1.6×10^{-3} ppmv for C_2H_6 and 5×10^{-4} ppmv for HCN and steadily decreases in the stratosphere for

both. The maximum VMR for both gases is observed in March. For January month we

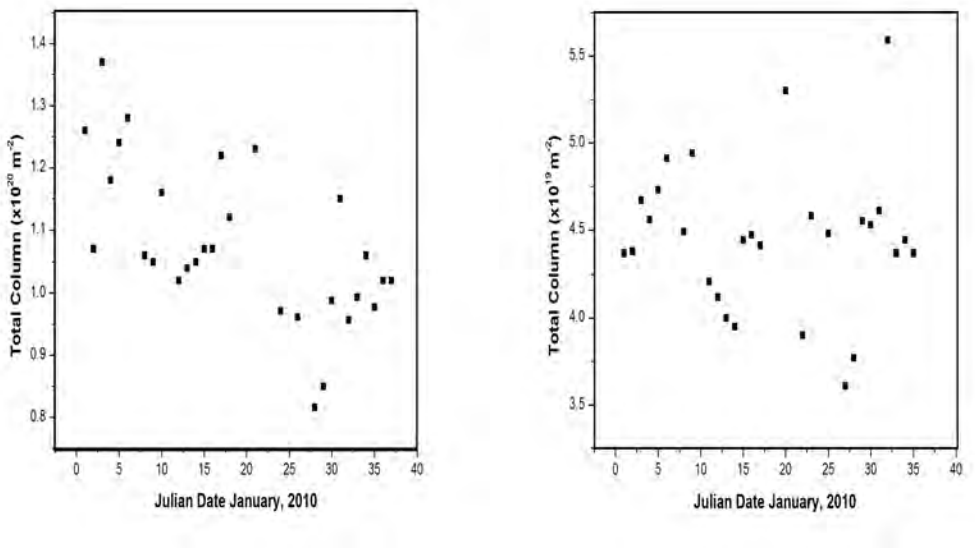

Figure 4.9: Time series of total columns of HCN (right panel) and C_2H_6 (left panel) for January, 2010.

have a better number (20 days) of measurements than other months. For this month we presented the time series of the total columns as well as vertical profiles for both species. Time series of total columns of HCN (right panel) and C_2H_6 (left panel) for January, 2010 are presented in Fig. 4.9. From this figure one can observe that total column amount of HCN is between 3.62×10^{19} and 5.60×10^{19} molecules/m^2 and that of C_2H_6 is between 8.19×10^{19} and 1.37×10^{20}. Except for two measurements for HCN the amount decreases with time. This shows that biomass burning at the beginning of the month is greater than that of the end of the month. Fig. 4.10 shows the contour plots of vertical profiles of HCN (right panel) and C_2H_6 (left panel) for January, 2010. Most C_2H_6 is found below 7 km while HCN reaches up to 35 km. The minimum and maximum of HCN VMR is 8×10^{-5}ppmv and 3×10^{-4}ppmv respectively while that of C_2H_6 is 1.75×10^{-4}ppmv and 1.23×10^{-3}ppmv.

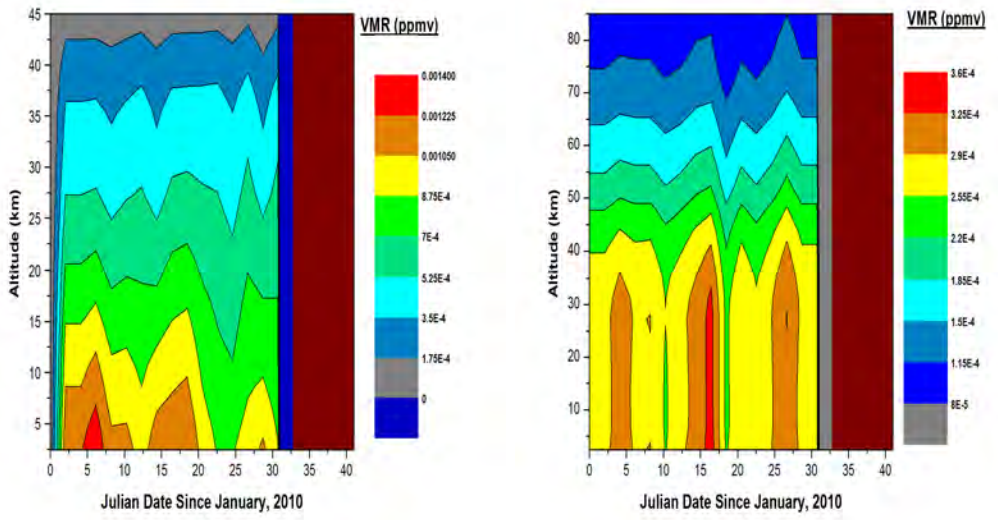

Figure 4.10: Vertical profiles of HCN (right panel) and C_2H_6 (left panel) for January, 2010.

4.3 Backward Trajectory Analysis and Fire Map

The Air Resources Laboratory's HYbrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model is a complete system for computing both simple air parcel trajectories and complex dispersion and deposition simulations. The model calculation method is a hybrid between the Lagrangian approach, which uses a moving frame of reference as the air parcels move from their initial location, and the Eulerian approach, which uses a fixed three-dimensional grid as a frame of reference. In the model, advection and diffusion calculations are made in a Lagrangian framework following the transport of the air parcel, while pollutant concentrations are calculated on a fixed grid [51].

After we determine the trajectory of the air parcel for the measurement days of which maximum total column amount is observed, knowing the situation (whether there was really fire for those days or not) of countries where the trajectory is passing over, is important. This can be done using an online spot plot from world fire atlas web site [52]. Fig. 4.11 shows plots of Air Resource Laboratory and World Fire Atlas from 8 to 12 March, 2010. Even if we did not include here, but using the same technique ([51] and [52])

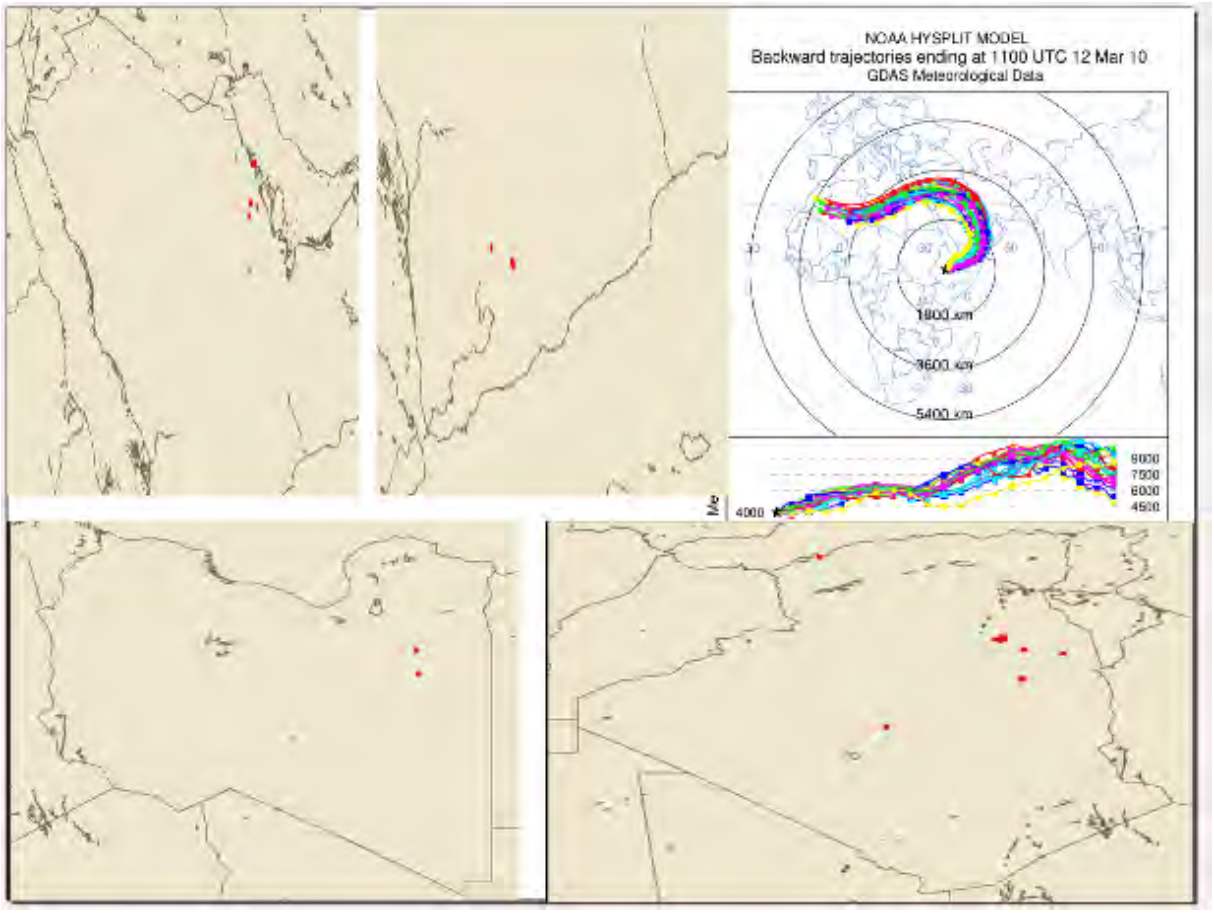

Figure 4.11: Biomass burning areas and back trajectory of the air mass during the period March 08 to March 12, 2010.

one could find a trajectory that passes through Mali, Niger, Chad, Sudan and Ethiopia and there is a fire over these countries during those days, from 8 to 12 March, 2010. From this figure we observe that the four countries (Yemen, Saudi Arebia, Liba and Algeria) are the possible sources of the gases due to biomass burning.

Chapter 5

Conclusion

Air pollution may be defined as a situation in which substances that result from natural and/or anthropogenic activities present at concentrations sufficiently high above their normal ambient levels to produce a measurable effect on humans, animals, vegetation, or materials. Biomass burning is a global phenomenon with impacts that affect global atmospheric composition and chemistry. The major sources of tropospheric HCN and C_2H_6 is biomass burning and reactions with hydroxyl radicals (OH) are primary sinks of C_2H_6, while ocean uptake and reactions with $O(^1D)$ are part of the suggested sinks of HCN. C_2H_6 has a life time of approximately 1 month while HCN is from 2 to 4 months or several years (up to 2.5 years in the stratosphere).

Spectroscopic measurements of tropospheric HCN and C_2H_6 are made using ground-based FTIR spectrometer over Addis Ababa (9.01^0 N, 38.76^0 E and 2.45 km asl), Ethiopia. From the measured spectrum we retrieved the VMRs and column amounts for each gas and the respective specrtal residual was also calculated. The spectral fittings of these windows were good. The maximum residual error of HCN is less than 0.1% and C_2H_6 is 3.5%. Under the error budget analysis statistical error (a total of 38.78% for HCN and 17.74% for C_2H_6) is larger than the corresponding systematic (a total of 37.33% for HCN and 12.04% for C_2H_6) error while the smallest for both is LOS. A baseline error, which arises if the two beams in a doublebeam spectrometer are not fully equivalent in transmitted power nor are corrected to be so, is the dominant error budget for both types of errors

of HCN while for C_2H_6 systematic error, spectroscopy and baseline interchangeably and statistical error dominated the selected altitudes. Even if the error for HCN is larger than that of C_2H_6 it has 2 degrees of freedom and for C_2H_6 we get 1.5 degrees of freedom. By changing the constraint during retrieval we can have different values of errors and DOFs. The retrieved profiles for both HCN and C_2H_6 decrease with altitude and the averaging kernel of HCN is found to be more sensitive beyond troposphere, but C_2H_6 is at the troposphere. This shows that the sensitivity of the retrieved state to the true state is at the troposphere for C_2H_6 while above this atmospheric layer for HCN. We tried to see the time series of HCN and C_2H_6 for 106 and 118 measurement days with a total of 154 and 181 measured spectra respectively since the first date of continuous operation of our FTIR spectrometer. The total column amount of HCN observed during this period ranges from 9.83×10^{18} to 6.39×10^{19} while for C_2H_6 from 3.60×10^{19} to 1.70×10^{20} molecules/m^2. As a general observation the total column amounts of these species are maximum during March. This helps us to conclude that biomass burning is their source. Of course it was better if there was a continuous measurement to see the seasonal variation of the gases. But it was impossible due to cloud covers for measurement times as well as technical problems of the instrument. We tried to see if biomass burning tracers like HCN and C_2H_6 are concentrated over Addis Ababa. If so air pollution is likely due to a change of ozone concentration in the troposphere as well as their poisonous nature.

Bibliography

[1] Zhao Y. et al. (2002). Spectroscopic measurements of tropospheric CO, C_2H_6, C_2H_2, and HCN in northern Japan, Journal of geophysical research, Vol. 107, No. D18, 4343, doi:10.1029/2001JD000748.

[2] Salau R. (2009). Tropospheric trace gases at Bremen measured with FTIR spectrometry, Journal of Environmental Monitoring, DOI: 10.1039/b903198g.

[3] Ministry of Environment, Part III. National Reporting Guidelines for CDS-14/15 Thematic Areas A. Atmosphere/Air Pollution Government Focal point(s): Responding ministry/office(s).

[4] Martens P. et al.(2003). Climate Change: An Integrated Perspective, Advances in global change research Vol.1, Kluwer academic publishers New York, Boston, Dordrecht, London, Moscow.

[5] Kasischke S. and Penner E. (2002). Improving Global Estimates of Atmospheric Emissions from Biomass Burning, Joint GOFC/GOLD Fire and IGBP-IGAC/BIBEX Workshop.

[6] Aldona W. (2006). First Trace Gas Measurements Using Fourier Transform Infrared Solar Absorption Spectroscopy at the University of Toronto Atmospheric Observatory, PhD Dissertation, Department of Physics, University of Toronto.

[7] Boubel W. et al. (1992). Fundamentals of Air Pollution, 3^{rd} Edition.

[8] Curtis L. (2002). Biomass Burning: Wood, Leaves, Grass, Forests, Crops and Trash, Burning Issues Special Edition Nov. 11 [additions noted with "ed."].

[9] Seinfeld J. and Pandis S. (2006). Atmospheric Chemistry and Physics; From Air pollution to climate change, A Wiley-Interscience Publication, 2^{nd} Edition.

[10] Wamukonya N. (2009). Africa Environment Outlook 2. Our Environment, Our Wealth, United Nations Environment Programme.

[11] Wondwossen T. Welfare analysis of controlling indoor air pollution, (the case of urban Ethiopia).

[12] Jacobson Z. (2005). Fundamentals of Atmospheric Modeling Second Edition, Stanford University.

[13] Mari H. et al. (2008). Tracing biomass burning plumes from the Southern Hemisphere during the AMMA 2006 wet season experiment, Atmos. Chem. Phys., 8, 3951-3961.

[14] Williams E. et al. (2010). The influence of biomass burning on tropospheric composition over the tropical Atlantic Ocean and Equatorial Africa during the West African monsoon in 2006, Atmos. Chem. Phys. Discuss., 10, 7507-7552.

[15] Seboka Y. et al. (2009). Biomass energy for cement production: opportunities in Ethiopia, United Nations Development Programme.

[16] Simeonova P. et al. (2004). Hydrogen Cyanide and Cyanides: Human Health Aspects, Concise International Chemical Assessment Document 61.

[17] Lary J. (2004). Atmospheric pseudohalogen chemistry, Atmos. Chem. Phys. Discuss., 4, 5381-5405, European Geosciences Union.

[18] Lupu A. et al. (2009). Hydrogen cyanide in the upper troposphere: GEM-AQ simulation and comparison with ACE-FTS observations, Atmos. Chem. Phys. Discuss., 9, 2165-2194.

[19] Rudolph J.(1995). The tropospheric distribution and budget of ethane, Journal of geophysical research, Vol.100, No. D6, 11,369-11,381.

[20] Hafemeister W. (1991). Physics and nuclear arms today.

[21] Gezahegn S. (2010). Observation of Atmospheric Carbon Monoxide by Ground-Based FTIR Spectrometer over Addis Ababa, Ethiopia, MSc. thesis, Physics Department, Addis

Ababa university.

[22] Wang X. (2005). The Economic Impact of Global Climate and Tropospheric Ozone on World Agricultural Production, MSc. thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

[23] Al-Hazaimay S. (2008). Impact of Instrumental Parameters on atmospheric Trace Gases Retrievals using FTIR Spectrometry, MSc. Thesis Institute of Environmental Physics, University of Bremen.

[24] Banwell C. and McCash M. (1994). Fundamentals of Molecular Spectroscopy, 4^{th} Edition, McGraw-Hill Book Company, Toronto.

[25] Remedios J. (1990). Spectroscopy for Remote Sounding of the Atmosphere, Ph.D. thesis, University of Oxford.

[26] Liou K. (2002). An Introduction to Atmospheric Radiation, 2^{nd} Edition, International Geophysical Series, Vol.84, Academic press, New York.

[27] Atkins P. and Friedman R. (2005). Molecular Quantum Mechanics, Fourth Edition, Oxford university press.

[28] Zender C. (2005). Radiative Transfer in the Earth System, Department of Earth System Science, University of California.

[29] Fischer J. et al. (203). Total internal partition sums for molecular species in the 2000 edition of the HITRAN database, J. Quant. Spectrosc. Radiat Transfer, 82, 401-412.

[30] Notholt J. et al. (1995). Total Column Densities of Tropospheric and Stratospheric Trace Gases in the Undisturbed Arctic Summer Atmosphere, J. Atmos. Chem., 20 (3), 311-332.

[31] Barret B. et al. (2005). Coheur, eLine narrowing effect on the retrieval of HF and HCl vertical profiles from ground-based FTIR measurements, J. Quant. Spectrosc., 95, 499-519.

[32] Hase F. et al. (2007).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. Workshop

on spectra analysis using PROFFIT (V9.5) and PROFFWD(V1.7).

[33] Kidder S. and VonderHaar H. (1995). Satellite Meteorology, Academic Press, Toronto.

[34] Stephens G. (1994). Remote Sensing of the Lower Atmosphere: An Introduction, Oxford University Press, New York.

[35] Notholt J. et al. (2006). Spectral line finding program for atmospheric remote sensing using full radiation transfer, J. Quant. Spectrosc. Radiat. Transfer, 97, 112-125.

[36] Michelson A. (1892). On the Application of Interference Methods to Spectroscopic Measurements, Phil. Mag., 34, 280.

[37] Griffiths P. and J. A. de Haseth (1986). Fourier Transform Infrared Spectrometry, John Wiley and Sons, Toronto.

[38] Davis S. et al. (2001). Fourier Transform Spectrometry, Academic Press, New York.

[39] Fellgett P. (1958). A propos de la théorie du spectromètre interférentiel multiplex, J. Phys. Radium, 19, 187-191.

[40] Rodgers, C.D.(2000). Inverse Methods for Atmospheric Sounding theory and practice, Series on Atmospheric, Oceanic and Planetary physics, Vol.2.

[41] Milkessa G. (2010). Stratospheric Aerosol Climatology From Sage II and Observation of its Deriver, Carbonyl Sulfide, Over Equatorial Africa, MSc. thesis, Physics Department, Addis Ababa university.

[42] Wiacek A. et al. (2006). Ground-Based Solar Absorption FTIR Spectroscopy: Characterization of Retrievals and First Results from a Novel Optical Design Instrument at a New NDACC Complementary Station, Journal of atmospheric and oceanic technology Vol.24.

[43] Senten C. et al. (2011). Ethane, ethyne and carbon monoxide concentrations in the upper troposphere and lower stratosphere from ACE and GEOS-Chem: a comparison study, Atmos. Chem. Phys., 8, 3483-3508.

[44] Senten C. et al. (2008). Technical Note: New ground-based FTIR measurements at Ile de La Reunion: observations, error analysis, and comparisons with independent data,

Atmos. Chem. Phys. Discuss., 8, 827-891.

[45] Walsh P. (2009). Measurements and modelling of emissions from biomass burning in Australia, PhD. thesis, Department of Chemistry, University of Wollongong.

[46] Glatthor N. et al. (2009). Source classification of upper tropospheric pollution by MIPAS HCN and C_2H_6 global distributions, Atmos. Chem. Phys. Discuss., 9, 16197-16232.

[47] Li Q. et al. (2003). A global three-dimensional model analysis of the atmospheric budgets of HCN and CH3 CN: Constraints from aircraft and ground measurements, J. Geophys.Res., 108(D21), 8827, doi:10.1029/2002JD003075.

[48] Xiao et al. (2008). Global budget of ethane and regional constraints on US sources, J. Geophys. Res., 113, D21306, doi:10.1029/2007JD009415.

[49] Hough A. (1991). Development of a Two-Dimensional Global Tropospheric Model: Model Chemistry, J. Geophys. Res., 96(D4), 7325-7362.

[50] Aikin A. (1982). Atmospheric Chemistry of Ethane and Ethylene, J. Geophys. Res., 87(C4), 3105-3118.

[51] http://ready.arl.noaa.gov

[52] http://wfaa-dat.esrin.esa.int

Declaration

This thesis is my original work, has not been presented for a degree in any other University and that all the sources of material used for the thesis have been dully acknowledged.

Name: AMBACHEW ABEJE ALEMU

Signature:————————

Place and time of submission: Addis Ababa University, June 2011

This thesis has been submitted for examination with my approval as University advisor.

Name: DR. GIZAW MENGISTU

Signature:————————