Optical Properties of Laboratory-Generated Polar Stratospheric Cloud Particles

A thesis submitted for the degree of Doctor
of Philosophy in the University of Oxford

Stephen Frank Bass,
Atmospheric, Oceanic and Planetary Physics.

March, 2003.

Contents

Abstract

Optical Properties of Laboratory-Generated Polar Stratospheric Cloud Particles

Stephen Frank Bass
Linacre College
Atmospheric, Oceanic and Planetary Physics

Thesis submitted for the degree of Doctor of Philosophy in the University of Oxford.

Hilary Term, 2003

Laboratory experiments were conducted to generate mimic polar stratospheric cloud aerosol particles in a chamber under stratospheric conditions. Infrared extinction spectra were measured at various resolutions with a Fourier transform spectrometer, permitting characterisation of both gas spectral lines and broad aerosol absorption features.

A novel band absorption model was coupled with a Mie extinction model in order to simultaneously retrieve wavenumber-dependent complex refractive indices and aerosol size parameters from low resolution spectra. A relationship was established to relate absorption band areas, produced in the retrieval of refractive indices, to chemical speciation of solutions using published binary thin-film and aerosol indices for calibration. This relationship was employed for determination of composition for binary and ternary aerosol refractive indices derived in this work, and was independently validated from gas concentrations calculated from high resolution spectra of gas absorption lines.

The first set of refractive indices derived from direct measurements of ternary solution aerosol are presented in this thesis. These indices are provided with comprehensive estimates of wavenumber-dependant uncertainties, and span a composition and temperature range of relevance to stratospheric aerosol. Comparison with published indices of directly-measured binary solutions, and ternary solutions calculated from combination of binary data, revealed findings of major significance for the use of these indices in remote sensing of stratospheric aerosol. It is found that certain sulphate absorption features exhibit consistent differences between thin-film and aerosol measurements, and that some sulphate features possess a pronounced temperature dependence; these findings support the rigorous efforts to ensure experimental conditions of relevance to this stratosphere in this work. Significant qualitative differences between absorption features are also detected at some compositions between published calculated indices of ternary solutions and those derived here. Critically, different sets of refractive indices exhibit pronounced variations in scaling. This impedes determination of PSC surface area density from remote spectral measurements.

Refractive indices calculated for this thesis advance scientific knowledge of the infrared characteristics of ternary solutions under stratospheric conditions, although further work is required to span the entire composition range of relevance to stratospheric ternary solution aerosol, and to resolve the issue of refractive index scaling.

Acknowledgments

This work was supported by the Natural Environment Research Council (NERC) grant ref. GT04/98/AS/304, and by the European Commission's Environment and Climate Programme under contract ENV4-CT97-0541 (POSTCODE). Laboratory experiments were carried out at the Molecular Spectroscopy Facility (MSF), part of the Rutherford Appleton Laboratory (RAL). A. Lambert (Atmospheric Chemistry Division, NCAR, Colorado) and R. G. Grainger (AOPP, Oxford, formerly University of Canterbury, Christchurch, New Zealand) are recognised as an original co-authors of the retrieval *engine* which formed the core of the retrievals used in this work. The support of RAL and the University of Canterbury, Christchurch, is acknowledged in respect to funding of a period of collaborative work with R. G. Grainger at the University of Canterbury, Christchurch, during April 2000.

Thanks to my supervisors, J. J. Remedios (Department of Physics and Astronomy, University of Leicester), R. G. Grainger (AOPP, Oxford), and D. A. Newnham (RAL) for help, encouragement, and for working unsociable hours. I also thank C. D. Rodgers (AOPP, Oxford) who was my supervisor for the first two years (handing over to R. G. Grainger on his arrival at Oxford), and J. Ballard (RAL) who was my supervisor for the first year (handing over to D. A. Newnham). Thanks to R. A. McPheat (RAL) and A. E. Heathfield (formerly at RAL) for helping me work through the steep learning curve during the initial laboratory experiments at the MSF.

Words cannot express my gratitude to Kathy, for her endless encouragement and understanding through testing times. And finally, my deepest gratitude to my parents, who gave me the means to achieve my full potential.

Acronyms

CFCs	ChloroFluoroCarbons
DLaTGS	Deuterium Lanthanum Triglycero Sulphide
DU	Dobson Unit
FFT	Fast Fourier Transform
FT	Fourier Transform
FTS	Fourier Transform Spectrometer
FTIR	Fourier Transform Infrared
FWHH	Full Width at Half Height
FWHM	Full Width at Half Maximum
ILS	Instrument Line Shape
MCT	Mercury Cadmium Telluride
MPD	Maximum Path Difference
MSF	Molecular Spectroscopy Facility, RAL
NAD	Nitric Acid Dihydrate
NAH	Nitric Acid Hydrate
NAM	Nitric Acid Monohydrate
NAT	Nitric Acid Trihydrate
OPD	Optical Path Difference
ppbv	parts per billion by volume
ppmv	parts per million by volume
PSC	Polar Stratospheric Cloud
RAL	Rutherford Appleton Laboratory
RFM	Reference Forward Model
sccm	standard cubic centimeters per minute
SAA	Sulphuric Acid Aerosol
SNR	Signal to Noise Ratio
STS	Supercooled Ternary Solution
TOMS	Total Ozone Mapping Spectrometer
VMR	Volume Mixing Ratio
ZPD	Zero Path Difference

Chapter 1

Ozone and Polar Stratospheric Clouds

The stratospheric ozone layer, which occurs as a result of oxygen-related chemistry in the atmosphere, exists in the lower stratosphere over an altitude range of 15 to 50 km. The altitude of peak ozone concentration decreases with increasing latitude. Ozone concentrations are variable (by approximately 10% – 1 σ) with season. Even at peak concentrations, the ozone mixing ratio is only about 10 ppmv (parts per million by volume). Although ozone is a trace atmospheric constituent, the ozone layer is of paramount important because it:

1. Moderates the flux of biologically damaging ultraviolet (UV) solar radiation by absorption;
2. Determines the temperature structure of the atmosphere because absorption of solar radiation in both the formation and destruction of ozone heats the stratosphere, causing the temperature to rise from the tropopause to about 50 km;
3. Affects surface temperature due to modulation of the solar flux reaching the lower atmosphere and surface.

All three factors are directly and indirectly of fundamental importance to the surface conditions that allow life to exist on the Earth's surface.

Molina and Rowland (1974) suggested that the decomposition products of man-made chlorofluorocarbons (CFCs) could potentially deplete the ozone layer. It was subsequently discovered in the mid-1980s that the ozone layer over Antarctica was being almost totally depleted during the early Austral spring period, leading to the *Ozone Hole* (Farman *et al.*, 1985). The annual ozone loss over Antarctica has resulted in a statistically significant global reduction in ozone in the southern hemisphere (Stolarski *et al.*, 1991). More recently, significant springtime ozone depletion has also been observed over the Arctic (Müller *et al.*, 1997). This loss of ozone is of great concern, since not only does it indicate an increase of biologically damaging solar UV flux at the surface (leading to reduced crop yields, higher rates of skin cancer etc.), but also a change in both the temperature of the lower stratosphere and of the surface (Houghton *et al.*, 2001).

The principle driver of current research in stratospheric ozone is the prediction of future ozone loss and recovery, and consequent changes in solar UV flux at the surface. Following international action, man-made emissions of ozone depleting substances have fallen since the late 1980s. However, stratospheric concentrations of chlorine have only recently stabilised, and stratospheric bromine is still increasing (Staehelin *et al.*, 2001). Recent modelling studies (Shindell *et al.*, 1998; Waibel *et al.*, 1999; Tabazadeh *et al.*, 2000b) have raised concerns that the ozone depletion may increase in severity, particularly over the Arctic, despite a predicted downturn in stratospheric concentrations of chlorine and bromine over coming decades. These predictions of greater ozone depletion arise from the fact that increasing concentrations of greenhouse gases are likely to further cool the stratosphere (Fels *et al.*, 1980; Mahfouf *et al.*, 1994), which is favourable to the polar ozone depletion processes described in this chapter.

1.1 The ozone layer

The ozone layer is naturally in a continuous state of destruction and re-generation as a result of oxygen and catalytic chemistry in the stratosphere. Additional inputs to the catalytic chemistry arise from man-made emissions through homogeneous (not involving a surface) gas phase reactions, and heterogeneous (surface initiated) gas phase chemistry in polar regions or volcanically perturbed conditions.

1.1.1 Oxygen-only chemistry

Ozone is formed in the atmosphere through dissociation of O_2 molecules by solar ultraviolet (UV) radiation followed by a three-body process which rapidly attaches released oxygen atoms to oxygen molecules:

$$\begin{aligned} \mathrm{O_2} + h\tilde{\nu}(< 242\mathrm{nm}) &\rightarrow \mathrm{O} + \mathrm{O} && (1.1) \\ \mathrm{O} + \mathrm{O_2} + \mathrm{M} &\rightarrow \mathrm{O_3} + \mathrm{M} && (1.2) \end{aligned}$$

where M can be any molecule. In classical Chapman theory, loss of ozone occurs by a number of processes, the most important being the dissociation of ozone by solar UV radiation:

$$\mathrm{O_3} + h\tilde{\nu} \rightarrow \mathrm{O_2} + \mathrm{O} \tag{1.3}$$

Ozone is also lost by reaction with excited oxygen atoms:

$$\mathrm{O(^3P)} + \mathrm{O_3} \rightarrow \mathrm{O_2} + \mathrm{O_2} \tag{1.4}$$

This leads to a quasi-steady-state balance of ozone in the upper atmosphere, with maximum concentration between 15 and 30 km altitude.[1]

1.1.2 Homogeneous chemistry: Catalytic reactions and reservoir species

It was recognised at the start of the 1950s that catalytic reactions of ozone were possible through the action of hydroxyl radicals (Bates and Nicolet, 1950). It was subsequently proposed in the 1970s that nitric oxides (Crutzen, 1970) and halogenated hydrocarbons (Molina and Rowland, 1974) could destroy ozone in the upper atmosphere. Ozone destruction can

---

[1] Although maximum O_3 production occurs at an altitude of about 40km.

take place through the following catalytic cycles, which can be simplified by:

$$\begin{aligned} \mathrm{X + O_3} &\rightarrow \mathrm{XO + O_2} && (1.5) \\ \mathrm{2(Cl + O_3)} &\rightarrow \mathrm{2(ClO + O_2)} && (1.6) \\ \mathrm{XO + O} &\rightarrow \mathrm{X + O_2} && (1.7) \end{aligned}$$

with the net result:

$$\mathrm{O + O_3 \rightarrow 2O_2}$$

where X could be nitric oxide (NO) or the hydroxyl radical (OH). In the case of chlorine, the Cl radical is liberated from ClO in the presence of sunlight:

$$\begin{aligned} \mathrm{ClO + ClO + M} &\rightarrow \mathrm{Cl_2O_2 + M} && (1.8) \\ \mathrm{Cl_2O_2} + h\nu &\rightarrow \mathrm{Cl + ClO_2} && (1.9) \\ \mathrm{ClO_2 + M} &\rightarrow \mathrm{Cl + O_2 + M} && (1.10) \end{aligned}$$

And BrO can act in a similar manner:

$$\begin{aligned} \mathrm{ClO + BrO} &\rightarrow \mathrm{BrO + Cl + O_2} && (1.11) \\ \mathrm{Cl + O_3} &\rightarrow \mathrm{ClO + O_2} && (1.12) \\ \mathrm{Br + O_3} &\rightarrow \mathrm{BrO + O_2} && (1.13) \end{aligned}$$

Both processes yield the net result:

$$\mathrm{2O_3 \rightarrow 3O_2}$$

The pivotal difference between these processes is that reactions 1.8 and 1.11 do not require atomic oxygen, unlike reaction 1.7. Thus, the ClO and BrO catalytic processes can take place efficiently in the lower stratosphere, where the concentration of photo-dissociated atomic oxygen for reaction 1.7 is limited.

It can be seen from reaction 1.9 that sunlight drives the daytime chlorine reactions through the production of Cl radicals. During nighttime, Cl_2O_2 (which is also thermally unstable) dominates over ClO, but daytime photolysis leads to a prevalence of ClO. Chlorine-catalysed depletion of ozone therefore requires sunlight, and is dominant in the colder winter regions of the stratosphere (UK Stratospheric Ozone Review Group, 1996).

Active forms of chlorine and bromine originate from predominantly man-made sources, particularly chlorofluorocarbons (CFCs). Although heavier than air, these molecules are transported to the stratosphere over a period of years by diffusion and mixing processes. Once in the stratosphere, they are photo-dissociated by solar UV radiation, where gas phase reactions involving chlorine monoxide and chlorine dioxide occur. Most of the chlorine in the stratosphere then becomes bound in the less reactive molecules of $ClONO_2$ and HCl, which are not directly involved in the catalytic loss of ozone. These are known as *reservoir species* (Solomon *et al.*, 1986); bromine counterparts also exist.

1.1.3 Summary

The catalytic homogeneous reactions detailed here are representative of *in situ* chemistry in a *clean* mid-latitude stratosphere; stratospheric aerosol perturbs chemistry from the homogeneous process scenario. Homogeneous ozone depletion is constrained since most chlorine and bromine is bound in less-reactive reservoir species, limiting the depletion of ozone in the ClO and BrO reactions. In order to explain the severe localised ozone depletion responsible for what became known as the Antarctic *ozone hole*, it was necessary to recognise the role of heterogeneous chemistry in the polar stratosphere.

1.2 The ozone hole

1.2.1 Discovery of the ozone hole

In 1985, Farman *et al.* (1985) reported considerable reductions in springtime column densities of ozone over two sites in Antarctica. They used ground based spectrophotometers to measure ozone in a vertical column at each site. It was found that October (Austral spring) daily low ozone column densities over Halley Bay (the most southerly site at 76°S) had been steadily decreasing since 1968 when values were greater than 300 Dobson units (DU), to below 200 DU by 1984.[2] This figure can be compared with the annual polar maximum in column density which occurs later in spring as ozone rich stratospheric air from lower latitudes mixes into polar regions (as the polar vortex breaks up); values of greater than 450 DU were recorded historically at Halley. The existence of low levels of ozone over the Antarctica region was later confirmed by re-analysis of satellite data.

Depletion of stratospheric ozone is observed to occur when the sun returns after polar night, and takes place over roughly six weeks. Between altitudes of 13-22 km almost all the ozone is destroyed (Hamill and Toon, 1991) annually over Antarctica, forming what is commonly termed the *ozone hole.* Figure 1.1 shows the ozone hole over Antarctica in 2001 as observed with the Total Ozone Mapping Spectrometer (TOMS; on board the Nimbus 7 satellite). Following the period of springtime ozone destruction, levels are seen to recover towards normal. However, since 1984 the magnitude of depletion has become progressively more severe, and the recovery less complete and more delayed. Indeed, the 1998-99 season at Halley Bay was one of the most severe so far with an ozone column density low of 120 DU in early October, and peak ozone values of 290 DU occurring in late December, which is over a month later and 25 % less than the historical maximum (BAS Ozone Bulletin July 98, 1999). Autumn values were still some 15 % below normal, so complete recovery of the ozone layer no longer appears to occur after the depletion phase. The annual depletion of ozone over Antarctica represents a loss of over 3 % of total global ozone (Hamill and Toon, 1991).

The region of ozone depletion over Antarctica corresponds closely to the position of the winter *polar vortex.* This is a region of low pressure which forms during the polar night in the stratosphere as air cools and descends. The Antarctic vortex is stable and

---

[2] A measurement in Dobson units (DU) expresses equivalent thickness of ozone at the ground; 100 DU represents a column of 1 mm when reduced to standard temperature and pressure throughout the layer.

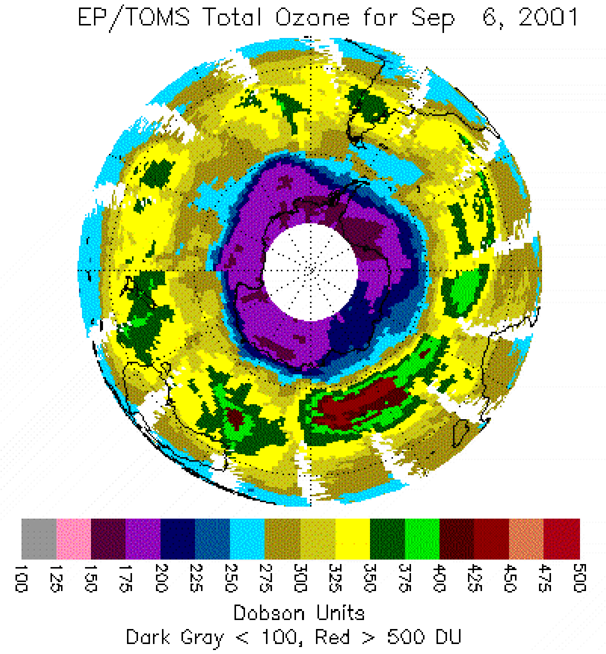

Figure 1.1: The ozone hole over Antarctica in 2001 from TOMS data (NASA).

lasts until spring. Air is essentially isolated inside the strong circulation and cools to very low temperatures, frequently below 185 K. In the Arctic, the polar vortex is weaker and tends to be disturbed more than once during the winter during stratospheric *warmings*. These are primarily caused by planetary waves of number one or two (Shindell *et al.*, 1998), which are related to the varied topography around the Arctic region. Hence, synoptic temperatures in the Arctic winter stratosphere tend to be about 10 K higher than over Antarctica. Ozone depletion has been observed during the last decade over the Arctic, particularly during the winters of 1994/95 and 1995/96 (UK Stratospheric Ozone Review Group, 1996; Müller *et al.*, 1997), and the 1999/00 season which was the coldest winter to that date (from over 40 years of stratospheric analysis) in the Arctic vortex. At around 18 km altitude over 60 % ozone loss occurred between January and March (Carslaw *et al.*, 1999b). Over limited areas, the magnitude of Arctic ozone depletion was similar to that experienced in the Austral spring of 1985. When the polar vortex breaks up during the spring, ozone poor air is mixed to lower latitudes, potentially contributing to a reduction in ozone column density over mid-latitudes.

The annual springtime depletion of stratospheric ozone in polar regions has been associated with downwards trends in midlatitude ozone concentrations. Figure 1.2 illustrates global depletion trends as deduced from TOMS (Stolarski *et al.*, 1991). It can be seen that trends are greatest towards polar regions, and near zero towards the equator. More recent analysis of stratospheric ozone trends (from ground-based and satellite observations) by Staehelin *et al.* (2001) has identified no significant trend over the tropics, but significant downward trends in the northern and southern midlatitudes of the order 2 to 4 % per decade over the period from 1970 to 1996 (with an acceleration in ozone depletion in the late 1980s). Large midlatitude trends are discerned by Staehelin *et al.* (2001) in the lower stratosphere, of -5.1 ±1.8 % per decade at 20 km and -7.4 ±2.0 % per decade at 15 km, where most of the stratospheric ozone resides.

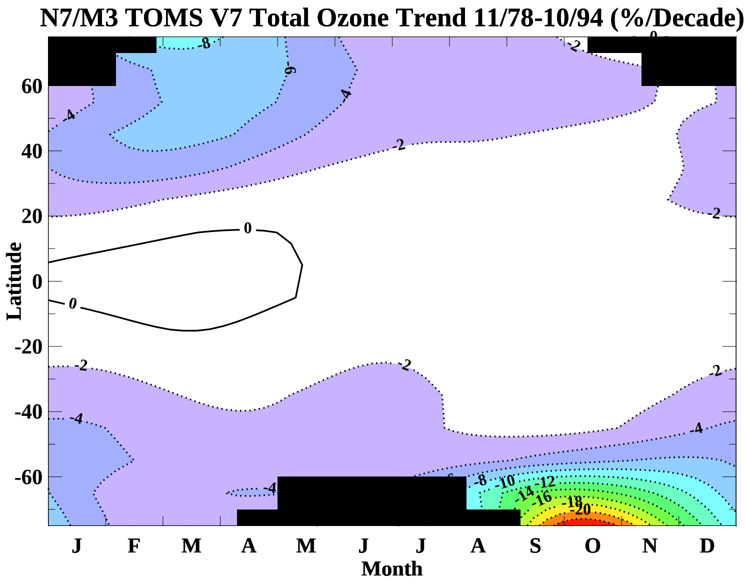

Figure 1.2: Ozone trends as a function of latitude and season calculated from TOMS data. The white area indicates where trends are not statistically different from zero at the 2σ level. The black area indicates polar night where no TOMS measurements were possible (Stolarski *et al.*, 1991).

1.2.2 Heterogeneous chemistry and PSCs

The cold temperatures that exist within the Antarctic polar vortex during winter, and the consequent formation of cloud particles, were pivotal to understanding the formation of the ozone hole over Antarctica. Cloud particles were known (Stanford, 1977; Stanford and Davis, 1974) to form from the condensation of volatiles, removing potentially HNO_3 and H_2O from gas phase at very low temperatures (required for condensation in the stratosphere).

Since the observed severe ozone depletion occurred in the only place in the stratosphere cold enough for PSCs to form widely during the polar winter, it was postulated (Solomon *et al.*, 1986) that chlorine was converted from inert reservoir species into active species through heterogeneous reactions on PSCs. The active chlorine (and other active halogen species) would then react catalytically with ozone in the presence of sunlight. The rates of heterogeneous reactions were determined in subsequent laboratory experiments (Molina *et al.*, 1987), and airborne measurements confirmed the presence of high concentrations of active halogens in areas of high ozone depletion (Brune *et al.*, 1989a,b). The activation of reservoir species by heterogeneous chemistry was found to account for approximately 95 % of catalysed ozone destruction (Anderson *et al.*, 1991), with ClO and BrO acting as the catalyst in 75 % and 20 % (respectively) of catalysed ozone loss in the lower stratosphere (UK Stratospheric Ozone Review Group, 1996).

Research also indicated that condensed HNO_3 was an important component of PSCs (Toon *et al.*, 1986; Crutzen and Arnold, 1986), and detection of HNO_3 in clouds followed (Fahey *et al.*, 1989; Pueschel *et al.*, 1989). This is significant since the removal of HNO_3 from the stratosphere delays the conversion of active species back into reservoir species, permitting much greater levels of ozone destruction, as detailed in terms of the chemistry in the following section.

1.2.3 Heterogeneous reactions on PSCs

The main heterogeneous reactions that occur on the surface of PSC particles are:

$$\mathrm{HCl(l) + ClONO_2(g) \rightarrow HNO_3(l) + Cl_2(g)} \tag{1.14}$$

$$\mathrm{ClONO_2(g) + H_2O(l) \rightarrow HNO_3(l) + HOCl(g)} \tag{1.15}$$

$$\mathrm{HCl(l) + HOCl(g) \rightarrow H_2O(l) + Cl_2(g)} \tag{1.16}$$

$$\mathrm{N_2O_5(g) + HCl(l) \rightarrow HNO_3(l) + ClONO(g)} \tag{1.17}$$

$$\mathrm{N_2O_5(g) + H_2O(l) \rightarrow 2HNO_3(l)} \tag{1.18}$$

where (g) and (l) indicate gas and liquid phases respectively; for solid PSCs change liquid to solid. These reactions occur very fast and convert the reservoir species into active forms (Cl_2 and HOCl) which can take part in the catalytic destruction of ozone when sunlight returns after polar night. Moreover, it can be seen that all reactions apart from reaction 1.16 produce HNO_3, which is absorbed into PSC particles. The direct effect of reactions 1.17 and 1.18, through the condensation of HNO_3 on PSCs, is to *denoxify* the stratosphere. This is very important for the ozone depletion process as the ClO dimer, in the presence of NO_2, would undergo a gas-phase reaction converting the chlorine back into a reservoir species (a similar process would take place with bromine):

$$\mathrm{ClO(g) + NO_2(g) + M(g) \rightarrow ClONO_2(g) + M(g)} \tag{1.19}$$

thus limiting the catalytic destruction potential of the chlorine. When present, PSCs can therefore act as a temporary sink of nitrogen (denoxifying the stratosphere) with the NO_x being returned when the cloud evaporates.

Nitrogen oxide can also be permanently removed from layers of the polar stratosphere through the process of *denitrification.* On formation of larger solid PSC particles, with high fall velocities, sedimentation (descent under gravity) from the stratosphere can occur within the lifetime of the particles. This has the effect of removing any HNO_3 in the particles from the stratosphere. In this case, NO_2 will not return to normal gas phase concentrations following evaporation of PSCs, slowing conversion of chlorine radicals back to reservoir species and prolonging the period during which ozone is catalytically destroyed when sunlight returns in spring. There is some uncertainty over the relative importance of denoxification and denitrification in the Antarctic. Portmann *et al.* (1996) argue that as temperatures remain cold enough for PSCs to remain after the spring equinox, the stored nitrogen content of PSCs is at least comparable to that lost through sedimentation of PSCs. Due to the warmer synoptic temperatures of the Arctic stratosphere in spring, this process is less likely to take place, and denitrification is probably critical for Arctic ozone depletion.

Over Antarctica, where temperatures rarely rise above the PSC threshold during polar winter and denitrification is extensive, active chlorine species can accumulate during the polar night in the presence of PSCs, leading to severe destruction of ozone when the sun returns. Chlorine and bromine radicals are then recycled in the catalytic depletion of ozone until the polar vortex breaks down and NO_x-rich air is mixed in from lower latitudes allowing conversion of the radicals back to reservoir species.

1.2.4 Summary

Figure 1.3 is a representation of the main stages in the process of chemical ozone depletion over polar regions from autumn through to spring. Note the feedback process in which a reduction in ozone heating of the stratosphere can lead to a colder, and thus more stable, polar vortex and longer-lived PSCs. The central research question is identification of the conditions, and resultant PSC types, that lead to denitrification. Therefore, it is of paramount importance that a thorough understanding of PSCs is obtained in order to predict future ozone depletion. This requires consideration of the following questions regarding PSCs:

- Which types of PSC are responsible for denitrification?
- What are the main types of PSC?
- What are the formation pathways for PSCs?

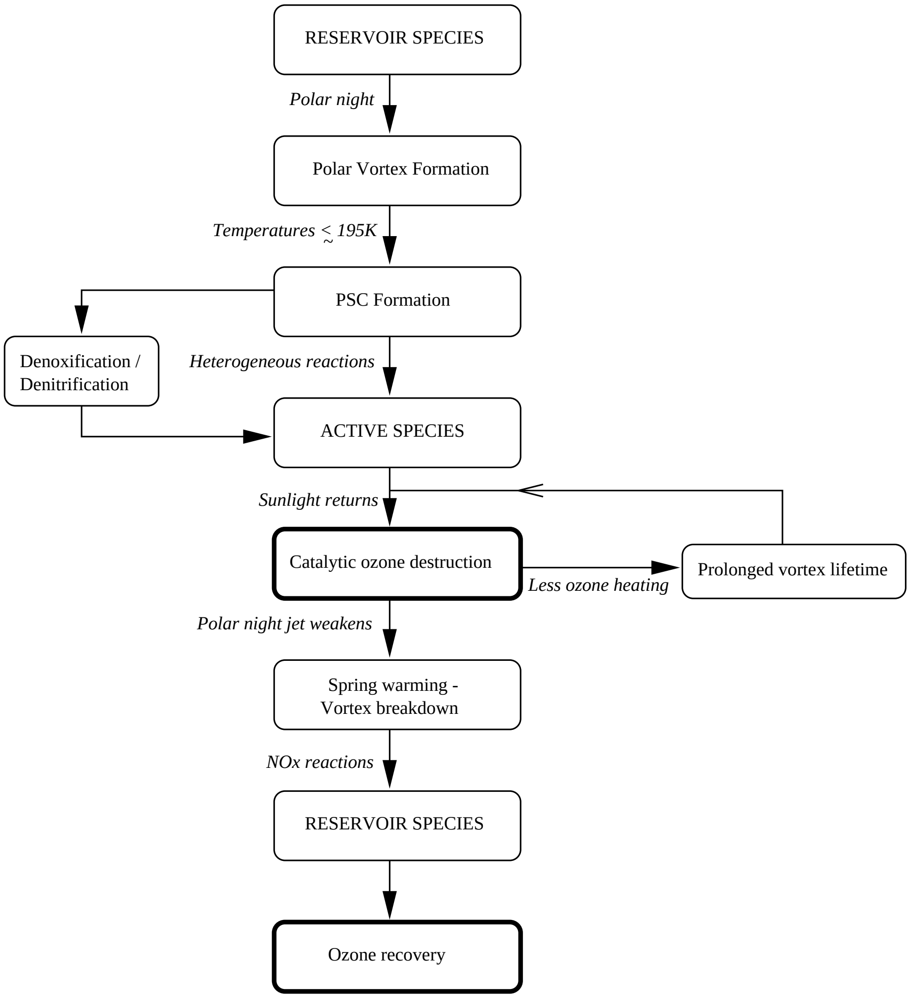

Figure 1.3: Schematic representation of stages in the ozone depletion process.

- What are the rates for heterogeneous activation on different types of PSC?

1.3 Polar Stratospheric Clouds

Polar stratospheric clouds were known to exist prior to the discovery of the ozone hole, and sporadic references can be found in scientific literature. Most notably *nacreous* or *mother-of-pearl* clouds, being visible to the naked eye, had been observed on many occasions from high latitudes. These PSCs are usually a type of lenticular cloud formed from mountain lee waves. From the ground, nacreous clouds are seen due to iridescence caused by diffraction

Figure 1.4: Nacreous clouds photographed over the U.K. from a train between Chesterfield and Sheffield during the evening of 16/02/96 (E. Barnett).

of incident light; naked-eye sightings from the U.K. have been documented, see figure 1.4, and studied through independent satellite observations (Hervig, 1999). Stanford and Davis (1974) published a study of stratospheric cloud reports over the period 1870-1972, and listed 156 dates when nacreous clouds were reported in the northern hemisphere, and 139 cases of Antarctic sightings. There are also many observations of a stratospheric haze – a further study by Stanford (1977) speculated upon the nature of this more extensive PSC veil or *stratospheric cist* as it was called at the time. Observations made during 1950-51 from Antarctica were analysed, and it was deduced that stratospheric cist consisted of H_2O ice clouds in the lower stratosphere, constraining the possible mixing ratio of water to 6-7 ppmv in the Antarctic stratosphere (Stanford, 1977).

1.3.1 Types of PSC as classified by lidar

At the end of the 1980s it became clear, notably through lidar observations during DC-8 flights (Browell *et al.*, 1990; Toon *et al.*, 1990), that there were several distinct types of PSC. From the lidar measurements of backscatter and depolarisation, two broad classes were defined: type I and type II PSCs. Backscatter is the radiation component scattered back to the lidar. Type II PSCs have a much stronger backscatter signal than type I, which implies a considerably larger scattering surface area. Relatively high depolarisation of backscattered signal indicates the presence of non-spherical crystals.[3] Now known to consist of essentially pure water ice crystals, type II particles are relatively large with radii between 10-100 μm (Hamill and Toon, 1991), and form below the ice frost point T_{ice}, which is around $\sim$188 K in typical conditions.

Type I PSCs have a much weaker lidar backscatter signal than type II, and therefore have smaller total surface area. Existing at temperatures above the ice frost point (up to $\sim$196 K) they are known to contain nitric acid, and have different properties to pure water. Type I PSCs can be further subdivided (Browell *et al.*, 1990; Toon *et al.*, 1990):

- **Type Ia** PSCs have low scattering and high depolarisation ratios in lidar returns. Low scattering is indicative of a small population of large non-spherical particles (r $>$ 1 μm). High depolarisation results from the non-spherical nature of these particles which suggests that they are solid, probably crystalline, forms of nitric acid hydrate (NAH, 1:x HNO_3/H_2O, where x is an integer). The most stable form of NAH under stratospheric conditions is thought to be nitric acid trihydrate (NAT, 1:3 HNO_3/H_2O) (Hanson and Mauersberger, 1988). If these solid particles gain excess frozen water, they would appear in lidar signals as type II PSC particles;

- **Type Ib** PSCs have a higher scattering return and low depolarisation ratios, indicating a larger population density (with a corresponding larger total surface area) of smaller particles (r = 0.01$\rightarrow$0.4 μm). The low depolarisation ratios suggest a spherical, and therefore liquid form, possibly consisting of supercooled ternary solutions (STS : $H_2SO_4/HNO_3/H_2O$);

- **Type Ic** PSCs have intermediate properties exhibiting high scattering (thus sim-

---

[3] Typically, lidars employ polarising beam splitters (such as a Glan-Foster prism) to separate the aligned and cross polarised components. The terms *aligned* and *cross* refer to the plane of polarisation of the linear polarised transmitted laser beam. Mathematically, if $S = (I, Q, U, V)$ denotes the Stokes vector of the backscattered light, the aligned channel *detects* the intensity $(I + Q)/2$, the *cross* channel $(I - Q)/2$. Thus, the (linear) depolarisation (ratio) can be identified as $\delta = (I - Q)/(I + Q)$ (G. Beyerle; personal communication).

ilar particle size and number density to type Ib) but intermediate depolarisation, indicating small non-spherical particles. These particles differ from type Ia in their potentially smaller physical size, and have been suggested as a separate classification in literature (Tabazadeh and Toon, 1996).

Finally, the presence of sulphate aerosol should be noted. Through photochemical reactions, sulphur dioxide gas, injected periodically into the stratosphere during large volcanic eruptions, is converted into sulphuric acid gas which then reacts with water to form aqueous sulphuric acid (H_2SO_4) droplets in a process known as heteromolecular nucleation. This results in a global sub-visible layer of sulphuric acid aerosol (SAA) (Hamill and Toon, 1991). This *background* SAA is detected in lidar returns (Beyerle *et al.*, 1994, 1997) and is important since it may be linked to the formation and development of PSCs, particularly in the case of type Ib PSCs. In addition, high levels of SAA may perturb ozone directly through heterogeneous reactions. The number density of this layer is highly variable as a result of irregular volcanic activity, and it is possible that the density of the SAA layer could become negligible during volcanically quiet periods.

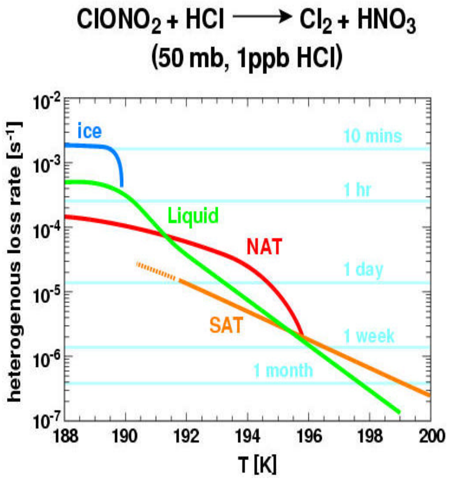

Figure 1.5: Reaction rate of $ClONO_2$ and HCl on different types of PSC particle. The horizontal lines show the lifetimes of reaction partners, calculated as the reciprocal of the loss rates (Carslaw *et al.*, 1999b).

1.3.2 Chlorine activation rates and PSC composition

The main driver of chlorine activation in the stratosphere is the total PSC surface area. Chlorine activation rate per unit area is a function of PSC type and temperature, as can be seen in figure 1.5 (Carslaw *et al.*, 1999b). It is apparent that sensitivity to temperature is greater than to the type of PSC particle. For example, cooling from a synoptic temperature of 194 K to 189 K in a mountain lee wave can lead to a shortening of the lifetime of these reservoir species from one day to less than an hour (in the case of liquid particles). Nonetheless chemical activation rates are mostly less than one day while meteorological conditions tend to force PSC existence of a few days or longer. Therefore, the dependence of chemical activation rates on PSC types (Ravishankara and Hanson, 1996) are likely to be most important in marginal Arctic winters with low chlorine activation and ozone destruction. The more important questions for the chemical effects of PSCs concern the prevalence and triggers for PSC existence and for denitrification:

1. How long are PSC particles present to cause chemical activation during polar winter and spring?
2. How long does chemical activation extend into spring?
3. What are the conditions that lead to extensive denitrification of the polar stratosphere, such as that observed over Antarctica?

The critical aspects of the questions, as regards PSCs, are therefore concerned with the mechanisms for formation and evolution of PSC composition.

1.3.3 Formation pathways

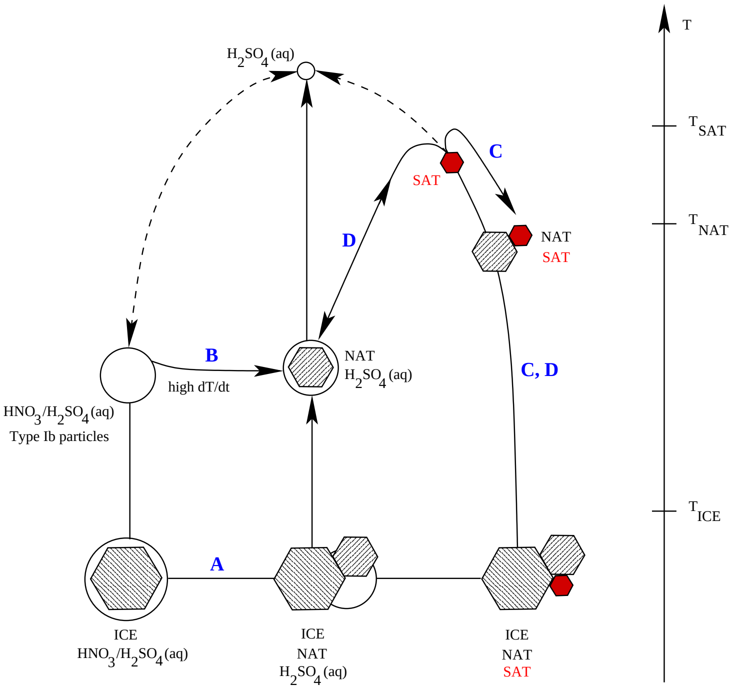

Figure 1.6: Formation pathways for PSCs (dashed lines indicate processes involved with liquid particles). The blue letters A to D refer to the scenarios described in the text. The temperature axis is not to scale (Carslaw *et al.*, 1999a).

A range of formation pathways for the principle types of PSC have been proposed, but the observational evidence to conclusively reveal true PSC development pathways has not yet been realised (Tolbert and Toon, 2001). In particular, there is considerable disagreement over formation mechanisms for solid, type Ia, PSCs. Although complicated, it is possible to distinguish four key aspects of PSC development:

1. The role of sulphate aerosol;
2. Temperature-controlled thermodynamics;
3. Nucleation mechanisms and rates (including cut-off temperatures);
4. Cooling rates yielding non-equilibrium conditions.

The most difficult aspects in modelling PSC formation are probably nucleation mechanisms and rates. In this brief overview, formation theories are reviewed in terms of PSC type.

Type Ib PSC. The formation of aqueous type Ib PSC is reasonably well understood (Carslaw *et al.*, 1997). As SAA is cooled, H_2O is absorbed from gas phase leading to dilution – concentrations are ~80 wt% at 230 K, typical of mid-latitudes, and ~45 wt% at 195 K in the polar vortex (Carslaw *et al.*, 1994). Condensation of HNO_3 in addition to more H_2O takes place upon further cooling, resulting in the transition from SAA to type Ib PSC (as indicated in indicated in figure 1.6). Type Ib PSCs therefore consist of liquid ternary (three component) solution particles of $H_2SO_4/HNO_3/H_2O$, which are often supercooled (hence the term supercooled ternary solution – STS) with respect to the main solid forms of NAH. Thus, STS particles exhibit a range of compositions under stratospheric conditions, ranging from almost binary H_2SO_4/H_2O solutions at higher temperatures, to almost binary HNO_3/H_2O at temperatures close to T_{ice}. Figure 1.7 illustrates this variation in composition of STS aerosol over the relevant temperature range under stratospheric conditions from the thermodynamic models of two groups (Carslaw *et al.*, 1995; Lin and Tabazadeh, 2001). It has been noted (Carslaw *et al.*, 1994) that STS solutions approach $\sim$ 1:4.5 HNO_3/H_2O towards T_{ice}, but this composition is highly dependant on HNO_3 gas partial pressure (Carslaw *et al.*, 1997).

This model of type Ib PSC growth from the background SAA layer may be somewhat simplistic. From *in situ* particle scattering measurements, Dye *et al.* (1990) found that

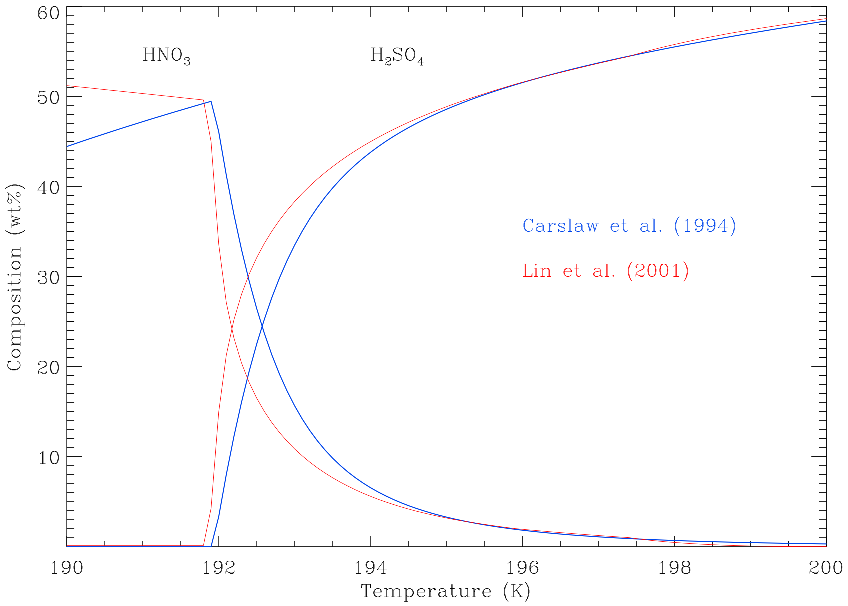

Figure 1.7: A temperature dependent ternary aerosol composition under typical stratospheric conditions of 10 ppbv total HNO_3 gas and 5 ppmv total H_2O gas at 55 mb; calculations produced from the thermodynamic models of Carslaw *et al.* (1995) and Lin and Tabazadeh (2001).

measurements of PSC number density were consistently an order of magnitude higher than measurements of condensation nuclei (all particles including SAA, soot, etc.) outside PSCs. This indicates that type Ib PSCs, as the most likely type of PSC to form upon initial cooling of the stratosphere, may nucleate on more than just SAA aerosol. However, homogeneous nucleation of type Ib PSC would require a supersaturation of at least 30 %, which would provide a considerable barrier to PSC formation which is not borne out from observations. Alternative sources of PSC formation nuclei include material from meteoric ablation, rocket exhaust, polluted tropospheric air entering through tropospheric folds, and cosmic rays.

The fate of liquid PSCs is important for understanding the nucleation and development of solid PSCs which are capable of growing to much larger sizes, at the expense of more numerous STS particles (since they have lower vapour pressures), and thereby causing significant denitrification of the polar stratosphere (due to their higher sedimentation velocities). Since STS particles grow from pre-existing sulphate aerosol, their relatively large number density restricts particle radii (due to the limited supply of volatiles) to the range 0.01-0.4 μm (Fahey *et al.*, 2001). Thus, STS particles have sedimentation velocities of less than a few metres per day, and are not thought to cause significant denitrification (Waibel *et al.*, 1999). The key questions regarding the fate of liquid PSCs are:

- What are the relative roles of solid particle nucleation within liquid STS aerosol (homogeneous), spontaneously in the stratosphere (homogeneous) or on other pre-existing solid particles (heterogeneous)?
- What temperature conditions and ternary compositions are necessary for solid particles to nucleate from STS aerosol?

Alternatively STS aerosol may simply evaporate, a process that can be detected from areas of *renitrification* (nitrogen enrichment) in the stratosphere.

Type II PSC. Excess water ice particles, or type II PSCs, could form through a number of different mechanisms leading to pure, coated, and co-existing ice particles (as indicated in figure 1.6). Laboratory and modelling studies have demonstrated that, in order to overcome the homogeneous nucleation barrier, cooling of at least 1-4 K below T_{ice} is required to freeze liquid STS aerosol (Carslaw *et al.*, 1998; Bertram *et al.*, 1996; Koop *et al.*, 1995). At temperatures below T_{ice}, STS aerosol becomes more dilute so that ice should precipitate readily (Koop *et al.*, 1997a). Alternatively, ice aerosol may form through heterogeneous nucleation on the largest NAT (or possibly solid sulphate) particles after the temperature drops below T_{ice} (Koop *et al.*, 1997b). Once frozen, rapid uptake of water vapour takes place due to the lower vapour pressure of H_2O above ice compared to liquid. This leads to the large observed sizes for type II PSCs. Growth is eventually arrested by dehydration of the stratosphere and sedimentation of the cloud particles, which occurs rapidly relative to other PSC types due to the large mass of type II particles.

Research into thermodynamic stability of liquid-solid mixtures has revealed that ice and/or NAH particles can co-exist in thermodynamic equilibrium with a liquid solution provided that the H_2SO_4 component remains in aqueous form (Koop *et al.*, 1997a; Iraci *et al.*, 1995); three states of co-existence are illustrated below T_{ice} in figure 1.6. Biermann *et al.* (1998) suggest that once a layer of NAT condenses on ice particles, the lifetime of these type II particles may be extended to temperatures above T_{ice}.

The role of solid sulphate hydrates. Laboratory studies have demonstrated that the most stable form of solid sulphuric acid hydrate under stratospheric conditions is sulphuric acid tetrahydrate (SAT, 1:4 H_2SO_4/H_2O) (Zhang *et al.*, 1993; Middlebrook *et al.*, 1993). Under stratospheric conditions, it is expected that SAT can exist at temperatures up to $T_{\text{SAT}} \sim 214$ K (Koop *et al.*, 1997a). However, evidence from other studies (Heathfield *et al.*, 1999; Koop *et al.*, 1997b, 1998) has demonstrated that it is very difficult to homogeneously freeze SAA even at temperatures well below T_{SAT}. Indeed, solid sulphuric acid hydrates have only been formed heterogeneously in laboratory experiments at stratospheric temperatures (Anthony *et al.*, 1997).

Assuming solid sulphate aerosol can be formed, it has been suggested (Dye *et al.*, 1992) that SAT particles could provide a surface for heterogeneous nucleation of NAH particles (Type Ia PSCs) at temperatures below T_{NAT} (route C in figure 1.6). Laboratory thin-film experiments (Hamill and Toon, 1991; Iraci *et al.*, 1995) have demonstrated that large supersaturation ratios (> 60) are required to condense NAT on a SAT layer. These conditions are unlikely to occur in synoptic scale processes in the stratosphere, although it is possible that transient cooling events (i.e. lee wave cooling) could lead to the required supersaturation (Iraci *et al.*, 1995). Iraci *et al.* (1995) conclude that "if sulphate aerosols freeze, NAH condensation is hindered". Zhang *et al.* (1996) found that once a NAT layer has formed on SAT and then evaporated, the SAT layer becomes pre-activated facilitating formation of a NAT layer upon subsequent cooling (route 4 in figure 1.6).[4] Finally, to date there is no conclusive evidence that SAT has been observed in the polar stratosphere,

---

[4] Temperatures only 2-3K below T_{NAT} are required to form a NAT coating once the SAT layer became pre-activated.

although some researchers have inferred the existence of solid sulphate particles from airborne and ground-based lidar observations (Gobbi and Adriani, 1993; Beyerle *et al.*, 2001). Additionally, Larsen *et al.* (1995) interpret a set of backscatter sonde profiles to infer the occurrence of SAT. Thus, although the existence of solid sulphuric acid hydrates in the stratosphere cannot be ruled out, both the laboratory (thermodynamic) and observational evidence do not support the existence of solid sulphate aerosol in the stratosphere.

Type Ia PSC. Type Ia PSCs are known to consist of crystalline NAH particles. Thermodynamically, the most stable form of NAH under stratospheric conditions is NAT (Hanson and Mauersberger, 1988), and during the course of this thesis the first conclusive observations of NAT particles in the polar stratosphere over the Arctic were obtained (Voigt *et al.*, 2000). The formation pathways and development of type Ia PSC is the least understood of the PSC types, which is a strong constraint on the accurate modelling of ozone depletion.

There is little consensus among researchers regarding the nucleation mechanism of type Ia PSCs from ternary solution aerosol (Remedios *et al.*, 2000; Zondlo *et al.*, 2000; Martin, 2000). Recent research has demonstrated that the effect on freezing temperature of HNO_3 in solution requires further characterisation from laboratory experiments (Tabazadeh *et al.*, 2000a). Homogeneous nucleation from supercooled liquid particles, at a temperature below $T_{\mathrm{NAT}} \sim 196$ K, appears unlikely since even small amounts of H_2SO_4 in solution has been observed to prevent NAT formation in the laboratory (Koop *et al.*, 1997b). Direct NAT particle nucleation from STS proceeds at such a slow rate that Tabazadeh *et al.* (2001) conclude that this process is "unimportant to denitrification" in a recent modelling study. Additionally, the barrier to nucleation in simple binary HNO_3/H_2O solutions is too high to facilitate nucleation of NAT on realistic timescales considering the large temperature variability in the Arctic (Prenni *et al.*, 1998), but slow cooling events in the Antarctic may lead to significant NAT formation if enough binary HNO_3/H_2O is available (Zondlo *et al.*, 2000) – this would require the sulphate component to have been removed from the ternary solution, probably through freezing, a process which is thermodynamically unlikely.

One possibility is that type Ia PSCs only nucleate after type II PSCs have formed through heterogeneous nucleation on a frozen particle (Disselkamp *et al.*, 1996; Biermann *et al.*, 1998); this process is shown as route A (heterogeneous NAT nucleation on ice) and route C (heterogeneous NAT nucleation on a solid sulphuric acid hydrate) in figure 1.6. In the case of route A, the supercooling (of up to 4 K below T_{ice}) required for homogeneous ice formation (see *type II PSC* section) implies that this process, although possible in the cooler Antarctic, cannot explain solid particle nucleation on a synoptic scale in the Arctic.

From observations of nacreous clouds (section 1.3) and other studies (Peter *et al.*, 1992; Tsias *et al.*, 1997; Voigt *et al.*, 2000), it has been established that processes in mountain lee waves are responsible for some solid PSC formation in both hemispheres. Adiabatic cooling (Gary, 1989) in mountain lee waves by as much as 12 K (Carslaw, 1998) has been inferred from observations. Adiabatic cooling increases with altitude and type II PSC ice clouds are formed from the freezing of practically all background aerosol between 22 and 25 km, at a number density of approximately 10 cm^{-3} (Deshler *et al.*, 1994; Carslaw *et al.*, 1998). It has been envisaged that HNO_3 may heterogeneously nucleate as NAT in the presence of ice aerosol during the warming phase (Zondlo *et al.*, 2000). Upon compressional warming in the lee wave, the water-ice evaporates leaving any heterogeneously formed NAT particles. Lidar observations suggest that NAT nucleates on approximately 2 % of ice particles (Carslaw *et al.*, 1998) in lee waves. Indeed, the first conclusive observations of NAT particles (Voigt *et al.*, 2000) were obtained from a balloon flight through a lee wave downstream of the Scandinavian mountains.

Although it is accepted that NAT is the most stable form of NAH in typical stratospheric conditions, laboratory experiments have demonstrated that other hydrates have a lower nucleation barrier. For example, Iraci *et al.* (1995) found that nitric acid dihydrate (NAD, 1:2 HNO_3/H_2O) condenses on SAT preferentially to NAT, and that homogeneous nucleation of NAH particles also favours NAD formation as the interfacial surface energy for NAD crystallisation from solution is much lower than NAT. However, limited availability of HNO_3 in the stratosphere may favour higher hydrates, especially in denitrified conditions, as has been noted by some researchers (Tabazadeh and Toon, 1996; Koop *et al.*, 1997b; Larsen *et al.*, 2000; Schreiner *et al.*, 1999). Some laboratory observations suggest that the solid nucleus that forms in aerosol particles consists of NAD (Bertram *et al.*, 2000; Salcedo *et al.*, 2001; Worsnop *et al.*, 1993). Generally, kinetically favourable solid phases nucleate first in particles, and then often (but not always) convert to the most stable form (in this case NAT) with time (Worsnop *et al.*, 1993). Hence, it is likely that NAD readily converts to NAT in the polar stratosphere (Tabazadeh *et al.*, 2001), and recent observations of NAT in the stratosphere (Voigt *et al.*, 2000) appear to support this.

It is possible that rapid changes in temperature could lead to non-equilibrium conditions between liquid aerosol and the vapour phase volatiles, promoting homogeneous nucleation (without ice) of the more kinetically favourable (faster nucleating) solid NAH particles than NAT (route B on figure 1.6) – these could eventually transform to NAT through the mechanism set out above. Tabazadeh *et al.* (1996) suggested that small mesoscale temperature fluctuations that occur continuously in the stratosphere expose all air masses to high cooling rates and convert liquid aerosol into solid forms of NAH aerosol, leading to a large number density of small, type Ic, PSC particles. These would form after an air parcel has spent a sufficiently long period at temperatures below T_{NAT}. It has also been postulated (Tsias *et al.*, 1997) that very rapid warming rates in lee waves could promote homogeneous nucleation of unstable NAD from non-equilibrium ternary droplets at temperatures above T_{ice}, between 190-194 K.

In a recent modelling study (Tabazadeh *et al.*, 2001), it has been proposed that homogeneous nucleation of NAD can take place in a “nucleation window” of approximately 2 K centred around 191 K in the polar stratosphere; formation of NAD is favoured in this temperature range due to a favourable composition of HNO_3/H_2O in the STS aerosol. Outside this temperature window, the NAD particle production rate is found to fall off rapidly. This mechanism has much greater effect in air masses which cool slowly since such air masses are exposed to the appropriate temperature range for a longer period of time. The result is a collar or a “polar freezing belt” around the polar vortex region in which

NAD formation takes place. Tabazadeh *et al.* (2001) conclude that when NAD formed in the polar freezing belt converts to NAT, the NAT number densities increase by an order of magnitude in the model compared with values of NAT nucleation directly from STS aerosol. However, Dhaniyala *et al.* (2002) are cautious about the possibility of homogeneous NAD formation since the Tabazadeh *et al.* (2001) study uses NAD nucleation rates near the high end of existing laboratory data.

1.3.4 Denitrification

It has been noted previously that ozone loss is significantly enhanced by denitrification, hence any reliable prediction of future ozone loss requires a detailed understanding of this process. Denitrification can only take place if a small number density of particles grow preferentially at the expense of STS aerosol and become relatively large (Dhaniyala *et al.*, 2002) – such particles would require lower vapour pressure than STS and would thus need to be of solid phase (Waibel *et al.*, 1999).

Satellite observations have revealed that the Antarctic stratosphere denitrifies before it dehydrates (Tabazadeh *et al.*, 2000b), implying that the stratosphere is not being denitrified *initially* by type II ice particles, but by type Ia NAH particles. This implication is supported by theoretical studies which suggest that sedimenting NAT particles can denitrify the stratosphere without ice particles playing a role (Salawitch *et al.*, 1989; Toon *et al.*, 1990). This has important implications for the Arctic stratosphere, where typical synoptic temperatures in winter are above T_{ice}, but are below T_{NAT} for long periods – if NAH particles are responsible for denitrification of the Antarctic stratosphere, then the Arctic may be close to the threshold of much greater denitrification, and by implication, greater ozone depletion. It is possible that denitrification takes place through the sedimentation of both ice particles, incorporating some nitric acid, and large NAH particles (Carslaw *et al.*, 1999a).

In situ observations of the Arctic stratosphere during the record cold winter season of 1999/2000 (Fahey *et al.*, 2001) revealed, for the first time, a very low number density, at about 10^{-4} cm^{-3}, of large (up to 24 μm) solid NAH, probably NAT, particles uniformly distributed in areas of the stratosphere with a temperature of less than T_{NAT}. These *NAT rocks* are ideal candidates for denitrification since, compared to total aerosol number density of approximately 10 cm^{-3}, they are sparsely distributed and can thus grow at the expense of the surrounding STS aerosol. Fahey *et al.* (2001) determined that the observed size distribution of NAT rocks was bimodal, with peak diameter distributions at 3.5 μm and 14.5 μm. A 14 μm diameter spherical particle has a sedimentation velocity of about 1.5 km per day at an altitude of 18 to 19 km (Peter *et al.*, 1992). Growth calculations were undertaken by Fahey *et al.* (2001) which showed that the largest diameter particles could attain before falling out of the stratosphere, and thus causing denitrification, would be about 24 μm – this assumes the particle was initially formed near the uppermost altitude of PSC existence. The lifetime of a NAT rock depends on the initial formation altitude, the vapour pressure of HNO_3 and H_2O along the trajectory of the particle, and the condition that temperature remains below T_{NAT}.

Sedimentation trajectory analysis was also undertaken by Fahey *et al.* (2001) which indicated that the NAT rocks initially nucleated in regions of temperature between 188-194 K. Although ice nucleation (in lee waves) could have led to the heterogeneous nucleation of NAH at these temperatures, the fact that these temperatures are at the upper bounds of ice existence indicates that the role of lee waves is still an open issue. The same trajectory analysis also showed that the largest particles, with greatest potential for denitrification, took about six days to fall through the stratosphere. This implies that, if air parcels in the Arctic polar vortex remain at temperatures below T_{NAT} for over six days, there is significant potential for denitrification – i.e. the Arctic climate could change to one of severe ozone depletion with a relatively small change in temperature. Indeed, by the end of March 2000, ozone depletion of around 70 % near 20 km was observed in the stratosphere within the polar vortex (Santee *et al.*, 2000; Sinnhuber *et al.*, 2000), and the significant amount of denitrification observed is thought to have contributed to this ozone loss (Sinnhuber *et al.*, 2000). The role of NAT rocks in Arctic denitrification is likely to be critical, and models that accurately represent the formation and development of large particles will be necessary to predict future ozone loss in the Arctic. Indeed, as suggested by Fahey *et al.* (2001), NAT rocks may be the principle cause of denitrification in the Arctic stratosphere. If such particles can grow for a number of days, then sedimentation is likely. The main outstanding questions regarding this issue are:

- What is the selective nucleation mechanism for NAT?
- Does selective nucleation of NAT occur all over the polar vortex?
- Is selective NAT nucleation a recent phenomenon for the Arctic?

The lee wave induced NAT particles observed by Voigt *et al.* (2000), at less than 2 μm in diameter, and of a number density of approximately 0.1-0.5 cm^{-3}, were too numerous to selectively grow to large enough sizes (due to limited volatiles in the stratosphere) to possess fast enough sedimentation velocities to cause significant denitrification of the polar stratosphere according to Tolbert and Toon (2001). A similar argument is also given by Tabazadeh *et al.* (2000b). However, a more recent modelling study (Dhaniyala *et al.*, 2002), which involves a detailed analysis of the growth and sedimentation of NAT particles formed in lee wave clouds, found that the "inhomogeneity in growth rate in and below the nucleation layer results in the temporal (and therefore geographic) redistribution of particles." However, this model was unable to conclusively account for NAT rocks, and Dhaniyala *et al.* conclude that a much more sophisticated analysis is required to establish whether lee waves are responsible for NAT rocks and thus the extensive denitrification of the Arctic stratosphere during extreme cold conditions.

1.3.5 Overview – the current state of knowledge

Laboratory studies, models, and field measurements have failed to enable researchers to reach a consensus regarding the formation pathways of large, solid, NAH particles that are responsible for denitrification of the polar stratosphere. Although models predict that liquid type Ib PSCs should be prevalent in the initial cooling phase of stratospheric winter, observations over the Arctic indicate the existence of solid particles within liquid aerosol clouds widely in both hemispheres at temperatures above T_{ice}, but below T_{NAT} during the early polar winter (Dye *et al.*, 1996, 1990; Hofmann and Deshler, 1991). The formation mechanisms responsible for type Ia nucleation, and their relative importance, have yet to be discerned from the theories raised in literature (Tolbert and Toon, 2001; Zondlo *et al.*, 2000). The central questions relating to type Ia PSC formation are:

- What is the role of solid sulphate aerosol?
- What is the relative importance of homogeneous and heterogeneous nucleation?
- What is the relative importance of synoptic-scale and lee wave processes?

In the final analysis, the fate of type Ib PSCs, liquid STS aerosol, could be important to understanding the formation of solid NAH, type Ia PSCs. In the absence of a generally-accepted model for solid particle nucleation, the development of the capacity to predict denitrification of the stratosphere, and therefore to model future ozone loss, will require detailed observations of the development of PSCs in the polar stratosphere.

1.4 PSCs and future ozone loss

Although the international community commenced phasing out CFCs and other ozone depleting substances from the late 1980s, as a result of the applications of these chemicals and their longevity in the atmosphere CFC levels have been slow to stabilise. Figure 1.8 shows projected lower stratospheric chemical loading for a number of important species (Shindell *et al.*, 1998) – it can be seen that CFC and Cl_y levels are projected to return to 1985 levels by around 2050. Modelling without climate forcing shows that the catalytic destruction of ozone should decrease with halogen loading – this is indeed the case regarding mid-latitude ozone destruction (Chipperfield and Jones, 1999), but PSC-initiated heterogeneous ozone destruction is far more dependant upon the spatial extent and persistence of PSCs.

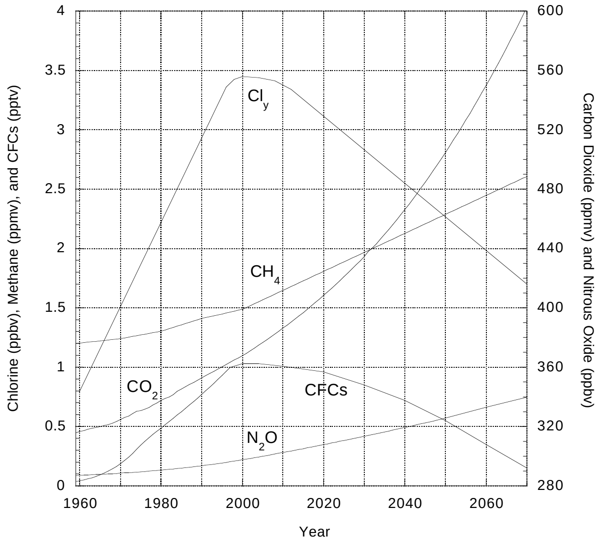

Figure 1.8: Projected trends for CO_2, N_2O, CH_4, CFCs and chlorine (Cl_y) (Shindell *et al.*, 1998). IPCC preferred scenario IS92a (Houghton *et al.*, 2001).

It has been known for some time that the stratosphere is likely to cool due to anticipated changes in greenhouse gas concentrations (Fels *et al.*, 1980). Rising carbon dioxide and methane levels, in addition to increasing stratospheric water vapour and decreasing ozone concentrations, are likely to lead to enhanced radiative cooling of the stratosphere (Fels *et al.*, 1980; Mahfouf *et al.*, 1994). The exact cause of the observed upward trend (Nedoluha *et al.*, 1998) in stratospheric water vapour is not known (Forster and Shine, 1999). Cooling trends have also been widely observed over Antarctica, principally due to decreasing ozone concentrations, and also during the last decades over the Arctic (Pawson and Naujokat, 1997). Both stratospheric cooling and increasing water vapour in the stratosphere are favourable for PSC formation. According to modelling by Tabazadeh *et al.* (2000b), the effect of a 1 K cooling in promoting denitrification is equivalent to an increase in water vapour mixing ratio of 1 ppmv.

The cooling and moistening of the polar stratosphere may increase ozone destruction by (Waibel *et al.*, 1999):

1. Enhancing heterogeneous chlorine activation rates on PSCs;

2. Increasing the frequency and duration of PSC episodes;

3. Intensifying denitrification.

Using a general circulation model which incorporated the first two points but not the third, Shindell *et al.* (1998) predicted that enhanced greenhouse-induced cooling of the Arctic stratosphere will result in more pronounced ozone loss in both hemispheres despite decreases in halogen loading. They calculate maximum ozone losses over the Arctic during the decade 2010 to 2019 as a result of increased PSC incidence, with more ozone destroyed over Greenland and Northern Europe in this period than anywhere within the Antarctic ozone hole. This increased incidence of PSCs results from calculated temperature decreases of 8-10 K over the Arctic, which is primarily a result of changes in stratospheric circulation (indicated by trends over the last decade (Zurek *et al.*, 1996)) due to 1 to 2 K of direct radiative cooling. However, other enhanced greenhouse models have shown less cooling of the lower stratosphere due to changes in stratospheric circulation (Graf *et al.*, 1998), or even negative feedback resulting in a warming of the Arctic polar vortex (Mahfouf *et al.*, 1994). Thus, the impact of dynamical feedback in controlling the northern polar vortex, and thus Arctic ozone depletion, remains an open issue.

Other research has demonstrated the importance of denitrification with regards to predicting the effects of climate change on ozone depletion. Waibel *et al.* (1999) conducted a study into denitrification caused by particle sedimentation and found that this mechanism could account for up to 35 % of ozone loss in the 1994-95 Arctic winter. The modelling of Waibel *et al.* revealed that denitrification increases Arctic cooling-induced ozone loss by more than a factor of 3, and that the Arctic stratosphere is currently at a threshold of denitrification. These findings were supported by Tabazadeh *et al.* (2000b) who calculated that severe denitrification could enhance future ozone loss by up to 30 %. Tabazadeh *et al.* conclude that the development of a stable, persistent and cold vortex is not required for Arctic denitrification to occur – i.e. enhanced radiative cooling (of only 1 K per decade), without associated dynamical changes, could be enough to trigger widespread denitrification.

1.5 Conclusions

The oxygen and catalytic chemistry involved in the production and destruction of the stratospheric ozone layer is now well established (section 1.1). Furthermore, the central role of PSCs in the activation of reservoir halogen species, which leads to severe polar ozone destruction, is also qualitatively understood (section 1.2). However, the nucleation and development of different types and compositions of PSC is complex, and is poorly characterised from research to-date. This is a major constraint with regards to modelling future ozone depletion, especially under PSC-marginal Arctic conditions, for two main reasons:

1. Since chemical composition and the phase of PSC particles determine the rate of halogen activation (Ravishankara and Hanson, 1996), total halogen activation in short-lived PSC events (where not all reservoir species are activated), and thus ozone depleting potential, is directly controlled by the type of PSC present;

2. Selective nucleation of a low number density of particles with lower surface vapour pressure than the surrounding ensemble of particles (i.e. nucleation of solid NAT particles within an STS cloud) can lead to the growth of few large particles (at the expense of the surrounding cloud), which are capable of falling out (sedimenting) of the stratosphere under gravity (Fahey *et al.*, 2001), causing permanent denitrification.

Thus, in order to better predict future ozone loss in a potentially colder stratospheric climate (section 1.4), a better understanding of solid particle nucleation mechanisms is required. This is particularly important with regards to the Arctic stratosphere, where synoptic temperatures are currently marginal for PSC formation. These processes are linked to the fate of liquid PSCs (section 1.3.3). Ideally, daily vortex observations of PSCs, with type and liquid composition determination, could be linked to stratospheric gas and temperature conditions to help better determine formation processes (including the relative importance of synoptic and lee wave processes).

Chapter 2

Optical Properties of PSCs

Since the importance of PSCs for ozone depletion was realised, a variety of techniques have been developed for PSC observation. These can be divided into two main categories as recognised in this chapter: (1) *in situ* measurements involving direct sampling of PSC particles, and (2) remote sensing observations, including remote detection of PSCs in the infrared.

Observation of PSCs in the infrared requires prior knowledge of the optical properties of PSCs. In particular, the spectral *fingerprint* of PSCs is characterised by the wavenumber-dependent refractive index. In this chapter, the remote detection of PSCs in the infrared is discussed, and an assessment of the impact of uncertainties in refractive index data is presented. The spectroscopic properties of PSCs are described in terms of Mie theory and the fundamental electromagnetic properties of liquids and solids. Finally, a review of published laboratory-generated optical data of PSCs provides an overview of current understanding.

2.1 *In situ* PSC observations

In situ measurements can provide detailed sampling of PSC properties. The main instruments used for *in situ* PSC observations are:

- Particle scattering probes;
- Backscatter sondes;
- Mass spectrometers.

2.1.1 Particle scattering probes

Particle scattering probes operate by measuring the intensity of scattered light at various angles from stratospheric aerosol particles; an appropriate model (e.g. a Mie model for spherical particles) can then be used to determine particle size. The operational principle is to entrain a stream of stratospheric air and aerosol for measurement while minimising the effect of the monitoring equipment on observations.

Some of the earliest particle measurements (Hofmann *et al.*, 1975) of the sulphate aerosol layer were conducted at Wyoming, USA, using a "dustsond". An optical particle counter measured scattered white light at 25 ± 5° from the forward scattering angle. This enabled measurement of particle density and radius, from balloon observations. An enhanced version of this instrument was later used to measure PSC properties from balloonborne measurements in the Arctic and Antarctic (Hofmann and Deshler, 1989). Further balloonborne measurements over the Antarctic (Hofmann and Deshler, 1991) confirmed that PSC aerosol exhibits a bimodal distribution in the cases observed. Hofmann and Deshler (1991) suggested that the large mode of particles, at number densities 3-4 orders of magnitude less than the smaller mode, may be the result of selective crystalline nitric acid trihydrate (NAT, 1:3 HNO_3/H_2O) nucleation from the sulphate layer. This was the first indication of NAT in the polar stratosphere.

Airborne measurements from NASA's ER-2, with a Forward Scattering Spectrometer Probe (FSSP) – another type of particle scattering probe, are reported by Dye *et al.* (1990). In addition to PSC particle number density and size, the FSSP also enabled measurement of condensation nuclei (CN) number density. From data analysis of a flight through the Arctic stratosphere on January 24, 1989, Dye *et al.* found that particle number densities within

PSCs were higher than CN concentrations, suggesting PSC nucleation "on more than just sulphuric acid particles." Dye *et al.* also found indications of a mode of larger particles near NAT saturation – suggesting selective nucleation of NAT particles and supporting the findings of Hofmann and Deshler (1991). Further analysis of the January 24 flight data (Dye *et al.*, 1992) confirmed that type I PSC remains liquid several degrees below the temperature of NAT stability. Modelling work revealed that the observations of Dye *et al.* are best explained if the type I PSCs are primarily composed of supercooled ternary solution (STS) aerosol (Carslaw *et al.*, 1994; Tabazadeh *et al.*, 1994).

Subsequent flights of the NASA ER-2 in the Antarctic stratosphere, on July 28, 1994, enabled Dye *et al.* to compare the FSSP results with those from the more advanced Multiangle Aerosol Spectrometer Probe (MASP). The analysis of Antarctic data (Dye *et al.*, 1996) supports findings from the earlier Arctic flight – most of the observed type I PSC consisted of liquid ternary solution droplets, although in some regions the composition was not clear and impactor results revealed some solid particles within the cloud, indicating selective nucleation of a larger mode of solid particles.

The refractive index of aerosol in the visible spectral region (m_∞ – see section 2.3.4) can be estimated with the MASP instrument by calculation of the ratio of light scattered in the forward and backward directions using Mie scattering theory (Baumgardner *et al.*, 1994); the purpose of this was to determine aerosol composition from the refractive index using the relations of Steele and Hamill (1981) . Comparison of refractive indices at 12821 cm^{-1} (780 nm), derived from the MASP flight data of July 28 (Baumgardner *et al.*, 1996), with theoretical indices of Steele and Hamill (1981) revealed lower values than expected. Baumgardner *et al.* suggest that this indicates that stratospheric aerosols are more hydrated than expected or that the relationship between refractive index and composition of Steele and Hamill is incorrect. However, this discrepancy may be due to the existence of non-spherical particles or absorption due to contamination (e.g. soot).

2.1.2 Backscatter sondes

The backscatter sondes described by Rosen and Kjome (1991), which are representative of this class of instrument, consist of a light source which emits light pulses every 7 seconds in a horizontal direction. Light scattered back to the balloon from aerosol within a few metres in front of the sonde is measured by two detectors, one measuring at 940 nm and the other at 480 nm. A *colour index* is calculated as the ratio of the aerosol backscatter at 940 / 480 nm – this is used to discriminate between types (including type I subdivisions) of PSC.

2.1.3 Mass spectrometers

Advanced mass spectrometers enable separate measurement of volatiles in vapour form and composition of PSC particles. An example is the Aerosol Composition Mass Spectrometer (ACMS), described by Schreiner *et al.* (1999) and Voigt *et al.* (2000), which uses an aerodynamic *lens* to focus particles into a narrow beam and separates particles from ambient gases. Particles are then evaporated in a small sphere and a quadruple mass spectrometer analyses the gas leaving the sphere. The traverse time is minimised (to less than 50 ms) to ensure minimal evaporation/condensation in the system. The ACMS therefore allows for separation of particles and ambient air, and for quantitative chemical analysis of condensed nitric acid and water in PSC particles (Schreiner *et al.*, 1999).

2.1.4 Summary

In situ measurements from single instruments can provide important indications of PSC characteristics and can be of use in supporting remote sensed data. However, the full potential of *in situ* measurements is realised when several instruments are combined to undertake simultaneous measurements of PSCs, such as the balloonborne combinations of particle scattering probes, backscatter sondes, a mass spectrometer and environment sensors reported by Schreiner *et al.* (1999) and Voigt *et al.* (2000). A subsequent flight in

January 2000 provided what is surprisingly the first definitive evidence of NAT in the polar stratosphere, at temperatures well above ice formation temperatures. *In situ* observations from the NASA ER-2 aircraft in the Arctic stratosphere between January and March 2000 with three instruments (two measuring gas and aerosol composition, and one measuring particle size and density), provided definitive evidence of the existence of low number densities of NAT particles widely dispersed across the Arctic stratosphere, suggesting the selective nucleation of NAT particles within STS clouds at temperatures above ice frost point (Fahey *et al.*, 2001).

In situ measurements have provided several keystone observations which have underpinned understanding of PSC formation and evolution in the stratosphere, these include:

- The first indication of HNO_3 in the stratosphere (Fahey *et al.*, 1989; Pueschel *et al.*, 1989);
- The first indication of NAT (Hofmann and Deshler, 1989, 1991);
- The first indication of STS particle formation 2-3 K below NAT temperature in the stratosphere, implying that STS is the dominant form of type I PSCs (Dye *et al.*, 1996; Schreiner *et al.*, 1999);
- The first indications that a separate mode of larger solid particles exist within a cloud of smaller STS aerosol (Hofmann and Deshler, 1991; Dye *et al.*, 1990, 1996);
- The first definitive identification of NAT in the polar stratosphere by Voigt *et al.* (2000);
- Substantial evidence of widespread selective nucleation of NAT at temperatures above the ice frost point (Fahey *et al.*, 2001).

In situ observations have provided fundamental insight into PSC particles in particular locations. The main limitation is the spatial and temporal coverage, given the necessity to base instruments on balloons and aircraft.

Further *in situ* measurements are essential to develop understanding of PSC processes and to resolve outstanding issues such as discrepancies between derived PSC refractive index at visible wavelengths (m_∞) and those predicted in theoretical studies (such as Steele and Hamill (1981) and Krieger *et al.* (2000b)). The MASP refractive index calculations indicated lower m_∞ values than expected from theory (Baumgardner *et al.*, 1996), while Larsen *et al.* (2000) found that higher m_∞ values than predicted from theory were required to explain observed backscatter from the Arctic balloon flight of January 25, 1998 (Schreiner *et al.*, 1999). Larsen *et al.* conclude by asserting that, in order to resolve this issue, further *in situ* measurements need to be combined with simultaneous measurements of optical properties of PSCs (i.e. remote sensing of PSCs).

2.2 Remote sensing of PSCs

Electromagnetic radiation provides a method for observing PSCs from a distance (remote sensing). The spatial coverage and resolution of remote sensing instruments is characterised by both the instrument type, and the instrument platform. Although restricted to a particular spatial region, ground based instruments can be physically larger and hence more advanced than flight-instruments. The principal ground based instrument types are spectrometers (including FTIR and SAOZ), which are passive and measure incoming radiation. Lidar, an active technique for probing the atmosphere, developed as a ground-based instrument, but mobile forms have been flown and satelliteborne lidar is planned. Lidar provides a considerably better vertical resolution than radiometers/spectrometers, but at the cost of reduced spatial coverage. Whereas airborne (balloon/aircraft) remote sensing observations are limited by flight duration, satellite instruments offer cost-effective continuous coverage of the atmosphere, and have potential to dramatically increase the spatial scale of observations compared with *in situ* methods. However, the vertical resolution of a satellite instrument field-of-view is limited to about 1 km at the current time.

Foreknowledge of certain PSC properties are necessary for analysis of remote observations, and simultaneous *in situ* measurements or key assumptions are commonly required to constrain analysis. Remote sensing is often limited by the difficulty in accounting for a large number of unknown parameters in analysis of spectral data.

2.2.1 Lidar observations

Lidar observations have played an important role in PSC observation. In the late 1980s, lidar observations during DC-8 flights (Browell *et al.*, 1990; Toon *et al.*, 1990), indicated that there were several distinct forms of PSC, and later analysis (Browell *et al.*, 1990; Höpfner *et al.*, 2002) provided the first PSC type classification (section 1.3.1).

The main operational components of a lidar instrument are the laser and detector. A collimated laser beam is directed through the atmosphere and the backscattered signal is recorded by the detector, at a vertical resolution of 75 m at best. The measured parameters are backscatter and depolarisation at the laser wavelength, typically 353, 532 or 1064 nm (28329, 18797 or 9399 cm^{-1}), i.e. in the visible spectral region. Ground based measurements are constrained by the requirement of an optically transparent troposphere and nighttime conditions. Instrumentation details and data reduction are described in detail elsewhere (Beyerle and Neuber, 1994). The products of the data analysis include measurements of aerosol volume density (aerosol volume within a volume of space), indication of visible refractive index at lidar wavelengths from aerosol scattering, and measurement of depolarisation. These data allow discrimination between liquid and solid phase, and estimation of aerosol particle volume and number density – for further information see section 1.3.1.

2.2.2 Satellite observations

The satellite instruments discussed here are limb sounders which observe tangentially to Earth's surface providing vertical profiles of the atmosphere (see section 2.2.3 for further details). Limb sounders are the principal satellite instruments employed for PSC observations. Nadir (vertically viewing) sounders have less sensitivity to PSCs owing to the relative pathways involved.

The earliest satellite observations of PSCs were those of the SAM II and SAGE I/II instruments in the 1980s. These instruments detected UV and visible radiation utilising scattered sunlight and direct observation of the sun (occultation) respectively. In the 1990s, a new generation of infrared **radiometers** observed PSCs in discrete spectral channels. The wide spectral range of radiometers is significant since these measurements are consequentially sensitive to different aspects of PSCs; absorption and thus PSC particle volume and composition in the infrared, and scattering and hence PSC mean size and distribution width in the near-infrared and visible. The surface area density of PSCs, an important parameter for modelling chemical activation, can be determined from spectral measurements but requires knowledge of PSC refractive index, particle shape, and size distribution (Grainger *et al.*, 1995; Massie *et al.*, 1998); particle size distribution and shape can be independently determined by *in situ* measurements or appropriate assumptions can be made for analysis of satellite data.

The Upper Atmosphere Research Satellite (UARS) contained several radiometers capable of providing PSC observations; these included the Improved Stratospheric and Mesospheric Sounder (ISAMS) and the Cryogenic Limb Array Etalon Spectrometer (CLAES) which performed coincident measurements between September 1991 and July 1992. Figure 2.1 shows channels from UARS instruments superimposed on a representative laboratory spectrum of PSC absorption. The Halogen Occultation Experiment (HALOE), also onboard UARS, observed PSCs infrequently due to the observation mode of this instrument and the orbital geometry of UARS, and could not provide coincident locations with ISAMS and CLAES. The capacity of satellite detectors to measure PSCs is illustrated in figure 2.2, which shows aerosol extinction data from the 12.1 μm channel of the ISAMS radiometer – a PSC to the north of the United Kingdom is indicated by high extinction values. Coincident CLAES measurements of gas phase HNO_3 were also used as an indicator of PSC activity.

Advances in technology have led the deployment of satelliteborne **spectrometers** for measurement of PSC properties. The key advantage of spectrometers over radiometers is the ability to sample a wide spectral range at high resolution. Although radiometer

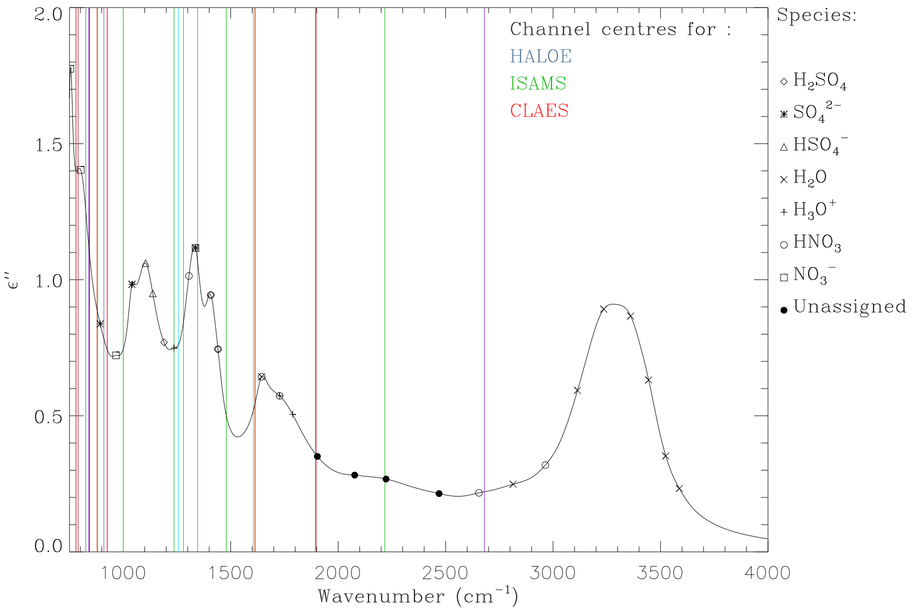

Figure 2.1: Retrieved band absorption and speciation (chapter 7) for a 196.7 K ternary solution aerosol spectrum in the imaginary dielectric constant, with channel centres for UARS radiometers indicated.

channels also cover a wide spectral range, individual instruments have limited coverage, the spectral channels are discrete and different wavelengths are not measured simultaneously. Thus, by recording extinction at many wavenumber points, spectrometers can potentially reduce uncertainties in PSC volume calculations (by providing coincident absorption and scattering measurements), and potentially capture the composition-dependant absorption fingerprint of PSCs in the infrared.

The ENVISAT satellite, launched in March 2002, contains the first collection of satelliteborne spectrometers capable of long-term measurements of PSCs. The MIPAS, SCIAMACHY and GOMOS instruments have considerable potential for coincident measurements of PSC over a wide spectral range; coverage with MIPAS extends from 685 to 2410 cm^{-1}, at a resolution of 0.035 cm^{-1} (apodised). Analysis of earlier measurements with the balloonborne MIPAS-B spectrometer, of similar configuration to the satelliteborne version, by Höpfner *et al.* (2002) provided indications of the wavenumber-dependant fingerprint of

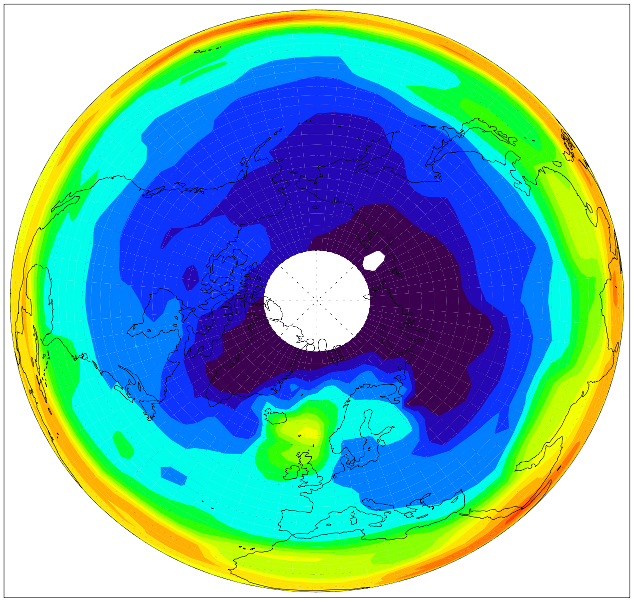

Figure 2.2: An image of ISAMS 12.1 μm extinction on January 9 1992 showing a PSC event north of the United Kingdom. Bright (yellow/orange/red) colours indicate high values and dark (blue) colours represent low values (Remedios *et al.*, 2000).

PSCs. Additionally, a mode of large particles was detected.

The Cryogenic Spectrometers and Telescopes for the Atmosphere (CRISTA) has been flown on the US Space Shuttle for use during missions. The CRISTA instrument is a limb-scanning instrument, which measures thermal emission of gases and aerosol in the 2500 to 141 cm^{-1} (4 to 71 μm) region, at a resolution of 2 cm^{-1}. The orbital geometry and instrument specification, for the CRISTA-2 flight during August 1997, resulted in an instrument field of view of 200 km along and 600 km across the satellite orbit track, with a vertical resolution of order 2 km. Measurements from CRISTA-2 have been analysed by Spang *et al.* (2001). The measurements contain more than a thousand infrared spectra of PSCs in the Antarctic polar vortex; the presence of PSCs was inferred from enhanced scattering of tropospheric radiation. Indications of PSC type were obtained from gas phase measurements of HNO_3, however lack of H_2O measurements from other instruments and uncertainties in the temperature structure of the atmosphere impeded definitive confirmation of PSC type by Spang *et al.*. However, in research for the EU MAPSCORE project (Remedios *et al.*, 2001), the appearance of a characteristic absorption feature in infrared CRISTA spectra (centred at 820 cm^{-1} and of 20 cm^{-1} width), as can be seen in figure 2.3, has been cited as tentative identification of NAT.

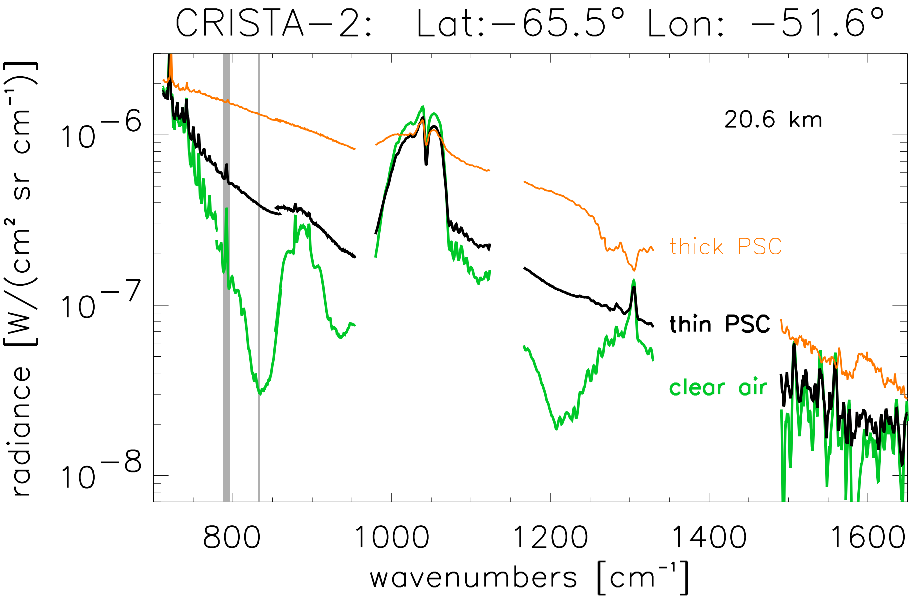

Figure 2.3: CRISTA-2 measurements for clean air and PSCs of varying optical thickness in the south polar vortex in August 1997 at 20.6 km altitude (Remedios *et al.*, 2001).

2.2.3 Limb sounding in the infrared

Limb sounding is a technique which offers high vertical resolution. The satelliteborne instrument observes the atmosphere tangentially, i.e. along a path directed from the satellite towards the *horizon*; this is illustrated schematically in figure 2.4 (Stephens, 1994). Radiation measured at the detector consists of signals originating from points along the limb path. In the infrared, this radiation is dominated by emission, due to the Planck function and spectroscopy of the atmosphere, and absorption. The spectroscopic behaviour is determined by the constituents (gases and aerosol particles) which interact with infrared radiation at the frequencies detected by the instrument. Although every point along the tangent path contributes to the observed signal, the field-of-view of aerosol particles is usually dominated by the region at the tangent point to Earth's surface.

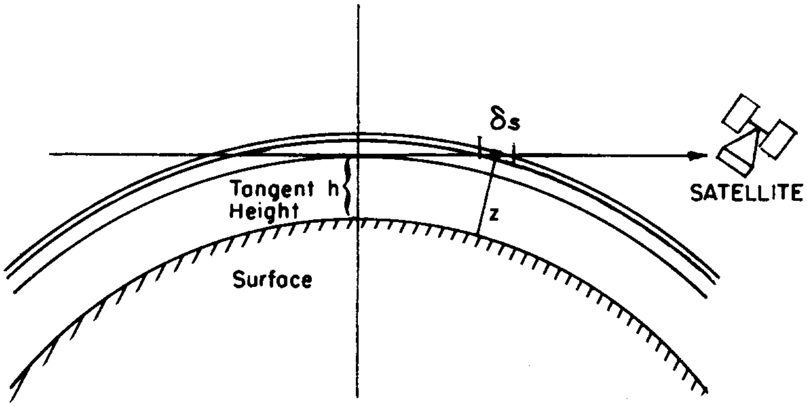

Figure 2.4: Limb viewing geometry, with tangent height h, and slab length s (Stephens, 1994).

Radiative transfer. In order to retrieve profiles of PSC properties from satellite measurements, it is necessary to first formulate a model of the observed signal. Such a model of the radiative transfer along the viewing path is known as the (case specific) forward model and is usually developed by representing the atmosphere as vertical layers. In the infrared, each layer of the atmosphere emits radiation which is absorbed by all of the layers above. The transmission $\tau(\tilde{\nu}, z)$, of a single tangent height z, therefore contains information on a number of vertical layers and a limb sounder normally observes a set of tangent heights. The forward representation suggests a number of retrieval methods including, for example, onion peeling in which the properties of the uppermost layer are retrieved from the uppermost tangent view, that of the layer below from that of the next lowest tangent view, and so on. More importantly, this demonstrates the key importance of the spectral properties of the PSCs and the dependence of the observed signal on other aspects of PSCs such as aerosol volume.

Consider a layer, of length s, which is a blackbody in local thermodynamic equilibrium. A beam of radiance, $L(\tilde{\nu})$ passing through the layer will undergo absorption while emission for the medium also takes place. The equation of radiation transfer may be expressed as:

$$\frac{dL(\tilde{\nu})}{\beta^{\text{ext}}(\tilde{\nu})ds} = -L(\tilde{\nu}) + B(T, \tilde{\nu}) \tag{2.1}$$

which is known as Schwartzchild's equation, where $B(T, \tilde{\nu})$ is the Planck function, and T the temperature of the layer. The volume extinction coefficient, $\beta^{\text{ext}}(\tilde{\nu})$, gives the cross sectional radiance removed from the beam per unit volume, by **both scattering and absorption**, and has dimensions of [L^{-1}]. Thus, $L(\tilde{\nu})\beta^{\text{ext}}(\tilde{\nu})ds$ is lost and $B(T, \tilde{\nu})\beta^{\text{ext}}(\tilde{\nu})ds$ is gained in the tangent path radiation.

In the case of aerosols, the physical thickness of each layer x is known, so a vertical transmission profile can be expressed in terms of the volume extinction coefficient $\beta^{\text{ext}}(\tilde{\nu}, z)$:

$$\tau(\tilde{\nu}, z) = \exp(-\beta^{\text{ext}}(\tilde{\nu}, z)x) \tag{2.2}$$

Now, the monochromatic optical path of a medium over length l is defined:

$$\tau_p(\tilde{\nu}, 0:l) = \int_0^l \beta^{\text{ext}}(\tilde{\nu})ds = \int_0^l k^{\text{ext}}\rho \, ds \tag{2.3}$$

where k^{ext} is the mass extinction coefficient, and ρ the medium density, such that $\beta^{\text{ext}}(\tilde{\nu}) = k^{\text{ext}}\rho$. By definition, $\partial\tau_p/\partial l = \beta^{\text{ext}}(\tilde{\nu})$. Equation 2.1 can be solved by integrating along the path l to yield:

$$L(\tilde{\nu}, l) = L(\tilde{\nu}, 0) \, \exp(-\tau_p(\tilde{\nu}, 0:l)) + \int_0^l B(T(s), \tilde{\nu}) \, \exp(-\tau_p(\tilde{\nu}, s:l))k^{\text{ext}}\rho \, ds \tag{2.4}$$

The first term represents attenuation of incoming radiation, which in the limb view case would be zero since the background is cold space. Thus, the second term, which represents the emission/absorption contribution from the medium along path 0 to l is applicable to the limb sounding case. This equation only accounts for radiation lost by scattering and does not include the effects of multiple scattering which can enhance the observed radiation. This is non-trivial to evaluate and is not the main subject of this thesis. It is important only to recognise for current purposes that multiple scattering could be more important when the number density is high enough for a significant number of photons to multiply scatter. In this case, however, the observed spectrum is usually *saturated* (i.e. $\tau(\tilde{\nu}) \to 0$), and the spectral sensitivity to PSC absorption is no longer observed over wide frequency ranges. Typically, the critical number density where saturation occurs is a function of mean particle size and number density, but a typical critical number density for liquid STS particles is 50 cm^{-3} (R. Spang; personal communication) – this number density would be lower for larger (i.e. solid) particles.

Sensitivity of retrieved PSC parameters to refractive index. In order to determine the impact of refractive index uncertainties on PSC remote sensing, it is necessary to establish the relationship between refractive index and the key PSC parameters retrieved from remote sensing, the PSC volume and surface area.

Assuming spherical aerosol particles, the volume extinction coefficient can be related to aerosol microphysical properties through:

$$\beta^{\mathrm{ext}}(\tilde{\nu}) = \pi \int_0^\infty r^2 Q^{\mathrm{ext}}(r, \tilde{\nu}, m(\tilde{\nu})) n(r)\, dr \tag{2.5}$$

where Q^{ext} is the extinction efficiency factor (from Mie theory), $n(r)dr$ denotes the number of particles with radii between r and $r + dr$, and $m(\tilde{\nu})$ is the complex refractive index of the aerosol, defined as $m(\tilde{\nu}) = n(\tilde{\nu}) + ik(\tilde{\nu})$. Spectral extinction due to PSC aerosol is governed by both the scattering properties and the absorption bands of aerosol. Absorption and scattering are directly related to the imaginary, $k(\tilde{\nu})$, and real, $n(\tilde{\nu})$, components of the refractive index respectively.

A sensitivity analysis was developed to calculate extinction spectra (using Mie theory) from refractive index data, assuming a homogeneous path. A size distribution typical of STS aerosol observed in the stratosphere (Dye *et al.*, 1992), and refractive indices of binary sulphuric acid thin-film (Palmer and Williams, 1975), were used.[1] It was found that, for wavenumbers of less than about 3500 cm^{-1}, the relationship between perturbations of $k(\tilde{\nu})$ and extinction is linear. For example, as illustrated in figure 2.5, a 1 % perturbation of $k(\tilde{\nu})$ leads to about 1 % perturbation in extinction in the same direction. However, above about 3500 cm^{-1}, extinction is less affected by perturbations in the imaginary index since $k(\tilde{\nu})$ tends towards zero. It was also found that, at wavelengths less than about 3500 cm^{-1}, variation of the real part of the refractive index, $n(\tilde{\nu})$, had little effect on observed extinction. At larger mean radii values, perturbation of $n(\tilde{\nu})$ was found to have greater effect at lower wavenumbers. In conclusion, it was determined that, for size distributions relevant to ternary solution PSCs, $k(\tilde{\nu})$ is linearly related to observed extinction and $n(\tilde{\nu})$ has considerably less influence over most of the infrared region.

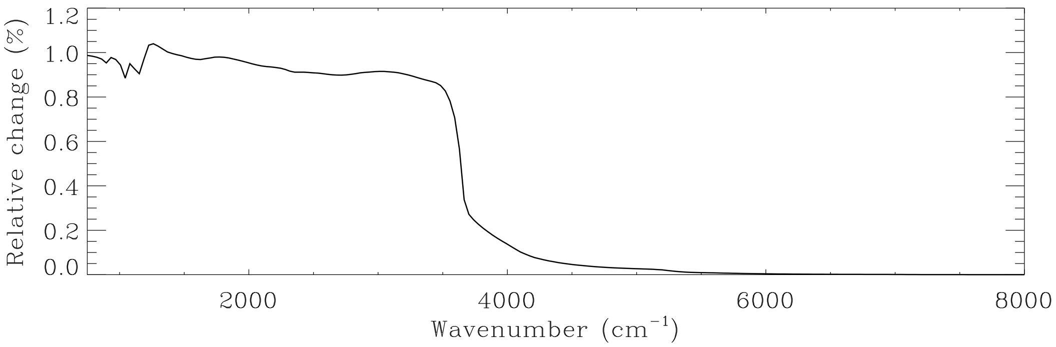

Figure 2.5: The effect of a 1 % increase in $k(\tilde{\nu})$ on extinction for sulphuric acid refractive indices, from a Mie model with a representative STS aerosol size distribution.

The relationship between measured extinction, aerosol volume density [L^3 L^{-3}] and aerosol surface area density [L^2 L^{-3}] was fully characterised by Grainger *et al.* (1995). First, the analytical solution of equation 2.5 shows that the volume density is linearly related to extinction. Secondly, analysis of an ensemble of 744 aerosol size distributions from *in situ* measurements of the Mt Pinatubo volcanic cloud (using balloonborne particle counters) supported this relationship. Finally, these measurements were employed to establish the

---

[1] The choice of indices employed is arbitrary since this is a sensitivity analysis.

relationship between aerosol volume density, V, and aerosol surface area, A, determined as:

$$A = 8.406\ V^{0.751} \tag{2.6}$$

where units of A are [μm^2 cm^{-3}], and V are [μm^3 cm^{-3}].

2.3 Spectroscopy of PSCs

Extinction in the mid-infrared spectral region from 750 to 4000 cm^{-1} (13.3 to 2.5 μm) is dominated by absorption, and in the near-infrared region from 4000 to 8000 cm^{-1} (2.5 to 1.25 μm) extinction primarily results from particle scattering. The typical laboratory-measured extinction spectrum in figure 2.8 exhibits a series of broad peaks below 4000 cm^{-1}, which are due to absorption. Towards the 4000 cm^{-1} limit, these bands diminish (as $k(\tilde{\nu})$ tends towards zero) and the upward curve in extinction above 4000 cm^{-1} results from scattering by aerosol particles. Hence, most of the information regarding $k(\tilde{\nu})$ is contained in the mid-infrared region, and most information pertaining to size distribution *and* $n(\tilde{\nu})$ is contained in the near-infrared spectral region. Despite the spectral delineation between scattering and absorption properties of aerosol, knowledge of *both* is required to accurately model the spectral regions near to 4000 cm^{-1} since some scattering may contribute to absorption below 4000 cm^{-1}, and some absorption may contribute above 4000 cm^{-1}; the amount of overlap depends upon the relative strength of absorption and scattering.

In order to determine refractive indices and aerosol size distribution parameters from laboratory spectra, a forward model was developed which evaluates an extinction spectrum from the refractive indices and scattering parameters of aerosol. This model was based on fundamental principles of aerosol spectroscopy expressed in a classical damped harmonic oscillator formulation for the absorption bands.

2.3.1 Fundamental principles of aerosol spectroscopy

In the following description, we assume homogeneous and spherical particles, with radii log-normally distributed (Kerker, 1969; Willeke and Baron, 1993), such that the number of particles, $n(r)$, whose radii lie between r and $r + dr$ is given by:

$$n(r) \quad = \quad \frac{N_0}{\sqrt{2\pi}}\frac{1}{\sigma}\frac{1}{r}\exp\left[-\frac{(\ln r - \ln r_m)^2}{2\sigma^2}\right] \tag{2.7}$$

where N_0 represents the particle concentration (usually in number per cm^3), r_m the median radius of the distribution and σ represents the spread of the particle size distribution.

Mie theory (Mie, 1908) can be used to calculate the extent to which a homogeneous sphere, such as a particle, extinguishes a beam of light through both scattering and absorption. Integrating over a distribution of particles gives the volume extinction coefficient as a function of $\tilde{\nu}$, see equation 2.5, where Q^{ext} is the extinction efficiency factor (from Mie theory) and $m(\tilde{\nu})$ is the complex refractive index of the aerosol. For a log-normal aerosol size distribution equation 2.5 becomes:

$$\beta^{\text{ext}}(\tilde{\nu}) = \frac{N_0}{\sigma}\sqrt{\frac{\pi}{2}}\int_0^\infty rQ^{\text{ext}}(r, \tilde{\nu}, m(\tilde{\nu}))\exp\left[-\frac{(\ln r - \ln r_m)^2}{2\sigma^2}\right] dr \tag{2.8}$$

Given N_0, the size distribution, and $m(\tilde{\nu})$, $\beta^{\text{ext}}(\tilde{\nu})$ can be calculated and, in the single scattering case, applied to equation 2.4 for limb sounding of PSCs.

The nature of the complex refractive index, $m(\tilde{\nu}) = n(\tilde{\nu}) + ik(\tilde{\nu})$, is investigated here in more detail. Equation 2.8 cannot be used to retrieve refractive index from a single extinction spectrum if one notes that for a measurement vector composed of N discrete wavenumbers there are $2N + 3$ unknowns; this is an under-constrained problem and cannot be solved. To solve this problem, Kramers-Kronig based techniques have been used by other researchers, where two measured spectra are used to increase the number of measured parameters to more than the number of unknowns, thereby yielding a solvable over-constrained problem. This technique is described in detail elsewhere (Clapp *et al.*, 1995). Basically, in the case of thin-films, the absorption of the sample is related to $k(\tilde{\nu})$ by a scaling factor, which takes into account the optical thickness of the film. Scattering from particles complicates the calculation of a scaling factor in the case of aerosol spectroscopy. Mie theory (Mie, 1908) can be used to model the scattering from, and absorption of, spherical droplets (a function of wavelength, particle size distribution and the refractive index). If an aerosol distribution has $r_{med} << 0.1$ μm, scattering is negligible at $\tilde{\nu} > 5000$ cm^{-1} (Richwine *et al.*, 1995). An extinction spectrum of aerosol particles of the small size regime is used as a first estimate of $k(\tilde{\nu})$. Another spectrum of aerosol of the same composition, but with larger size particles, can then be modelled using a Mie calculation based on this first guess of the imaginary index. An iterative fitting procedure can be undertaken to determine a self-consistent set of imaginary refractive indices. A subtractive Kramers-Kronig transform is finally applied for determination of the real part of the refractive index $n(\tilde{\nu})$ from the unscaled imaginary component, and thus the scaling factor for $k(\tilde{\nu})$ can be calculated.

Since the results from early settling experiments indicated it was not possible to reliably produce the necessary range of aerosol distributions while maintaining the same composition aerosol in this work (section 5.1.2), a different approach was adopted for this thesis which only requires measurement of a single spectrum for determination of refractive index.

2.3.2 Refractive index and absorption frequency

The elementary equations of wave motion can be used to prove the existence of electromagnetic waves propagating with velocity $v = c/\epsilon\mu$, where c is the speed of light in a vacuum, ϵ the dielectric constant (or permittivity) and μ the magnetic permeability of the medium. The dimensionless (real) refractive index, n, as the ratio of wave phase velocity in a vacuum to phase velocity in a medium, such that:

$$n = \frac{c}{v} = \left(\frac{\mu\epsilon}{\mu_0\epsilon_0}\right)^{\frac{1}{2}} = \left(\frac{\epsilon}{\epsilon_0}\right)^{\frac{1}{2}} \tag{2.9}$$

since $\mu = \mu_0$ over optical frequencies. Phase velocity and hence the refractive index of a medium are not constants with frequency; the frequency dependence of refractive index is known as dispersion.

Dispersion theory is a complex subject, at the highest level requiring a detailed quantum mechanical description. However, a simplified model of dispersion in a medium can be formulated by consideration of molecular structure. When a non-conductor is placed in an electric field, the distribution of electrons in the medium changes to produce an internal field which opposes the general field. This is known as polarisation, and arises for two main reasons:

1. **Polar** media consist of randomly distributed dipoles. When an external electric field $\mathbf{E}$ is applied, the dipoles become aligned parallel to this field direction. This is opposed by elastic forces in a solid, or the randomising action of thermal Brownian motion in a gas or liquid, such that an equilibrium condition arises;

2. Polarisation is induced in **non-polar** molecules by the alteration of electric charge distribution in the molecule by the applied electric field, resulting in an induced dipole moment.

A molecule is comprised of a number of atoms, which consist of positive nuclei surrounded by negatively charged electrons. The vector sum of all the dipole moments equals, to the first approximation, the polarisation vector $\mathbf{P}$. The motion of a nucleus and electrons resulting from application of an external electric or magnetic field has a quantum mechanical solution. However, classical wave theory can be applied, yielding essentially the same result. When the electrons are perturbed by an external field, a charge imbalance results leading to a quasi-elastic restoring force $Q = -q\mathbf{s}$, where $\mathbf{s}$ is the displacement vector.[2] The force exerted by the effective electric and magnetic field ($\mathbf{E}'$, $\mathbf{B}'$) on a charge e which has

---

[2] The nucleus will also be perturbed by the field, but being orders of magnitude heavier than an electron, its contribution can be neglected.

velocity $\mathbf{v}$ is given by the Lorentz law (based on Maxwell's fundamental equations):[3]

$$\mathbf{F} = e\left(\mathbf{E}' + \frac{1}{c}\mathbf{v} \wedge \mathbf{B}'\right) \tag{2.10}$$

If we assume that the electron velocity is small compared with the velocity of light in a vacuum, the magnetic field term can be ignored. Hence, the equation of motion of an electron of mass m is given by:

$$m\frac{\partial \mathbf{s}}{\partial t^2} + d\frac{\partial \mathbf{s}}{\partial t} + q\mathbf{s} = e\mathbf{E}' \tag{2.11}$$

where $d(\partial \mathbf{s}/\partial t)$ represents a damping force, which originates physically from the fact that vibrating electrons emit electromagnetic radiation, carrying energy away. If ω represents the angular frequency of the electric field, then $\mathbf{E}' = \mathbf{E}'_0 \exp(-i\omega t)$. Using $\mathbf{s} = \mathbf{s}_0 \exp(-i\omega t)$ as the solution to equation 2.11 yields the steady state solution:

$$\mathbf{s} = \frac{e\mathbf{E}'}{m(\omega_0^2 - \omega^2) - iwd} \tag{2.12}$$

where $\omega_0 = \sqrt{\frac{q}{m}}$ is the resonance or absorption frequency. It is clear that the inclusion of the damping term d leads to the complex nature of polarisation. This is of fundamental importance, since this imaginary term directly represents the absorption of the medium. This can be understood on a physical basis since, as damping corresponds to energy emission of the oscillating electrons, it similarly corresponds to absorption of incident radiation by electrons in the medium.

2.3.3 The Lorentz-Lorenz formula

The Lorentz-Lorenz formula relates Maxwell's equations to the atomic theory of matter, providing an expression linking the refractive index, m, with the dielectric constant, ϵ, and

---

[3] The effective electric field $\mathbf{E}'$ is used here as only non-polar molecules, which have no permanent dipole moment, are considered for simplification.

polarisability, α (Bertie *et al.*, 1994; Dignam, 1988):[4]

$$\frac{\epsilon - 1}{\epsilon + 2} = \frac{m^2 - 1}{m^2 + 2} = \frac{4}{3}\pi N\alpha \tag{2.13}$$

where N is the number of molecules per unit volume. The Lorentz law (equation 2.10) is utilised to relate the macroscopic electric field to the local field acting on each molecule in the liquid; for full treatment see Born and Wolf (1970). This model is only valid for isotropic media; although no liquid is isotropic at molecular distances, it is the mean orientation of the molecules per unit volume that is of interest, and to a good approximation we can assume that the *mean* polarisability is indeed isotropic. The electric moment $\mathbf{P}$ per unit volume is $\mathbf{P} = N\mathbf{p} = N\alpha\mathbf{E}'$. The polarisability expresses the polarisation $\mathbf{P}$ as a function of molecular density, and has units of volume.[5]

As demonstrated in equation 2.12, the polarisation becomes a complex quantity where (non-resonance) absorption occurs. Hence polarisability, the dielectric constant, and the refractive index all become complex quantities where the imaginary part represents the absorption of the oscillators. As this absorption is dispersive (frequency dependent), the Lorentz-Lorenz equation can be re-stated in wavenumber dependent complex form:

$$\frac{\epsilon(\tilde{\nu}) - 1}{\epsilon(\tilde{\nu}) + 2} = \frac{m(\tilde{\nu})^2 - 1}{m(\tilde{\nu})^2 + 2} = \frac{4}{3}\pi N\alpha(\tilde{\nu}) \tag{2.14}$$

The wavenumber dependent complex forms of refractive index, dielectric constant and polarisability can be written:

$$m(\tilde{\nu}) = n(\tilde{\nu}) + ik(\tilde{\nu}) \tag{2.15}$$

$$\epsilon(\tilde{\nu}) = \epsilon'(\tilde{\nu}) + i\epsilon''(\tilde{\nu}) \tag{2.16}$$

$$\alpha(\tilde{\nu}) = \alpha'(\tilde{\nu}) + i\alpha''(\tilde{\nu}) \tag{2.17}$$

Similarly, Maxwell's relation (equation 2.9) can be re-stated as $\epsilon(\tilde{\nu}) = m(\tilde{\nu})^2$. Since the

---

[4] This equation is also known as the Clausius-Mossotti equation.

[5] Since $\mathbf{p}$ and $\mathbf{E}'$ have dimensions $[el]$ and $[el^{-2}]$ respectively, where e is electron charge, and l length.

forward model retrieves parameters in dielectric space, it is useful to relate $\epsilon(\tilde{\nu})$ and $m(\tilde{\nu})$ using this relation:

$$
\begin{aligned}
\epsilon'(\tilde{\nu}) &= n(\tilde{\nu})^2 - k(\tilde{\nu})^2 && (2.18)\\
\epsilon''(\tilde{\nu}) &= 2n(\tilde{\nu})k(\tilde{\nu}) && (2.19)\\
n(\tilde{\nu}) &= \sqrt{(\sqrt{\epsilon'(\tilde{\nu})^2 + \epsilon''(\tilde{\nu})^2} + \epsilon'(\tilde{\nu}))/2} && (2.20)\\
k(\tilde{\nu}) &= \sqrt{(\sqrt{\epsilon'(\tilde{\nu})^2 + \epsilon''(\tilde{\nu})^2} - \epsilon'(\tilde{\nu}))/2} && (2.21)
\end{aligned}
$$

2.3.4 The classical damped harmonic oscillator

Dispersive absorption due to an oscillating molecule covers a finite wavenumber range and has a specific peak shape due to the fundamental modes of atomic and inter-atomic oscillation within the molecule; different molecular species absorb at different wavenumber centre positions, strengths, and band widths. Hence, the characteristics of absorption bands typical of aerosol spectra are determined by composition. The classical damped harmonic oscillator (CDHO) model provides a convenient way of expressing refractive index in terms of fundamental spectroscopic parameters, and is used in this thesis as part of the forward model to calculate refractive indices form aerosol extinction spectra. The CDHO model evaluates the dielectric spectrum from a set of n bands such that $j = 0, 1, 2, ...n$ with each band expressed by the central wavenumber of the oscillator $\tilde{\nu}_j$, width (damping constant) γ_j and band intensity S_j. The extra parameter $\epsilon'(\infty)$ denotes the *offset* value of the real part of the dielectric constant at zero frequency (infinite wavenumber) over the summation, such that $\epsilon'(\infty) = m_\infty^2$. The significance of m_∞ should be noted. Away from absorption bands, such as in the visible/UV spectral region, $k(\tilde{\nu}) = 0$ and m_∞ becomes equal to the real part of the refractive index, $n(\tilde{\nu})$. Thus, the value of m_∞ has a direct effect on scattering by aerosol in the visible/UV spectral regions (the other important parameters being particle number, size and shape parameters).

The CDHO model relates the complex *molar* polarisability $\alpha_m(\tilde{\nu})$ spectrum directly to molecular vibrational quantities through the following expressions (Bertie *et al.*, 1994):

$$\begin{aligned}
\alpha_m(\tilde{\nu}) &= \frac{\epsilon_\infty - 1}{\epsilon_\infty + 2}\frac{3N_A}{4\pi N} + \frac{N_A}{4\pi^2c^2}\sum_j \frac{\mu_j^2/3}{(\tilde{\nu}_j^2 - \tilde{\nu}^2 - i\tilde{\nu}\gamma_j)} && (2.22)\\
\alpha'_m(\tilde{\nu}) &= \frac{\epsilon_\infty - 1}{\epsilon_\infty + 2}\frac{3N_A}{4\pi N} + \frac{N_A}{4\pi^2c^2}\sum_j \frac{(\mu_j^2/3)(\tilde{\nu}_j^2 - \tilde{\nu}^2)}{(\tilde{\nu}_j^2 - \tilde{\nu})^2 + \gamma_j^2\tilde{\nu}^2} && (2.23)\\
\alpha''_m(\tilde{\nu}) &= \frac{N_A}{4\pi^2c^2}\sum_j \frac{(\mu_j^2/3)\gamma_j\tilde{\nu}}{(\tilde{\nu}_j^2 - \tilde{\nu}^2)^2 + \gamma_j^2\tilde{\nu}^2} && (2.24)
\end{aligned}$$

where N_A is Avogadro's number and μ_j is the dipole moment derivative with respect to the normal coordinate for the j^{th} band.[6] From the Lorentz-Lorenz equation, the complex molar polarisibility can be related to the complex dielectric constant:

$$\begin{aligned}
\alpha'_m(\tilde{\nu}) &= \frac{3V_m}{4\pi}\frac{(\epsilon'(\tilde{\nu}) - 1)(\epsilon'(\tilde{\nu}) + 2) + \epsilon''(\tilde{\nu})^2}{(\epsilon'(\tilde{\nu}) + 2)^2 + \epsilon''(\tilde{\nu})^2} && (2.25)\\
\alpha''_m(\tilde{\nu}) &= \frac{9V_m}{4\pi}\frac{\epsilon''(\tilde{\nu})}{(\epsilon'(\tilde{\nu}) + 2)^2 + \epsilon''(\tilde{\nu})^2} && (2.26)
\end{aligned}$$

In this expression, N_A/N is replaced by the molar volume V_m. The expression indicates that we cannot determine the molar polarisability from dielectric spectra in this analysis, since we have no information concerning V_m to calculate α_m through ϵ, or any knowledge of μ_j to calculate α_m directly by the CDHO model. These parameters cannot be distinguished from scattering due to aerosol in an extinction spectrum.

The CDHO expressions in complex dielectric constant space, as used in the forward model, can be obtained from the CDHO equations in complex polarisability space from the Lorentz-Lorenz relation for j bands (Bertie and Zhang, 1992):

$$\begin{aligned}
\epsilon'(\tilde{\nu}) &= \epsilon(\infty) + \sum_j \frac{\left(S_j\tilde{\nu}_j^2\right)\left(\tilde{\nu}_j^2 - \tilde{\nu}^2\right)}{\left(\tilde{\nu}_j^2 - \tilde{\nu}^2\right)^2 - \gamma_j^2\tilde{\nu}^2} && (2.27)\\
\epsilon''(\tilde{\nu}) &= \sum_j \frac{\left(S_j\tilde{\nu}_j^2\right)(\gamma_j\tilde{\nu})}{\left(\tilde{\nu}_j^2 - \tilde{\nu}^2\right)^2 - \gamma_j^2\tilde{\nu}^2} && (2.28)
\end{aligned}$$

---

[6] In order to account for the random orientation of molecules, μ_j is divided by 3.

where the dipole moment derivative μ_j is replaced by the band intensity S_j, since assuming mechanical and electrical harmony $S_j = (N/3\pi c^2)\mu_j^2 \ F_j^2$. The local field correction for a molecule in a liquid is F_j^2. The units of S_j are cm^{-2}.

2.3.5 Physical implications of the CDHO model

The molar polarisabilty (equation 2.22) is a direct representation of the molecular vibrational intensity of an oscillating species. Knowledge of this quantity would therefore directly relate (through an appropriate chemical model) to the concentrations of each oscillating species in a spectrum – the **speciation** or composition of a medium. In addition, if the CDHO expression for the dielectric constant is differentiated and equated to zero, the absorbance peak location is found to be located at a lower wavenumber than the oscillator centre $\tilde{\nu}_j$, such that the centre in $\epsilon''(\tilde{\nu})$ space is:

$$\tilde{\nu}_{\epsilon''_j} = \frac{-(\gamma_j^2 - 2\tilde{\nu}_j^2) + \sqrt{(\gamma_j^2 - 2\tilde{\nu}_j^2)^2 + 12\tilde{\nu}_j^4}}{6} \tag{2.29}$$

i.e. the band centre in dielectric space is a function of the true oscillator centre in polarisability space $\tilde{\nu}_j$ and the width γ_j. For example, a single band of arbitrary strength located at $\tilde{\nu}_j$ =1050 cm^{-1} has a centre in dielectric space $\tilde{\nu}_{\epsilon''_j}$ =1049.7 cm^{-1} where γ_j =50 cm^{-1}, and $\tilde{\nu}_{\epsilon''_j}$ =1045.2 cm^{-1} where γ_j =200 cm^{-1}. This raises the prospect of calculating the dipole moment derivative directly by using the CDHO model, which would yield a value for the molar polarisabilty. However, this requires the assumption that each band is sufficiently far apart that the dielectric constant due to oscillators other than the one of interest has no imaginary absorption ($\epsilon'' = 0$) and the real part is constant over the wavenumber region of the oscillator of interest. As will be seen later, this is far from the case; typical aerosol spectra are comprised of an overlapping arrangement of bands. Therefore we cannot calculate the most fundamental physical property of an oscillator, the complex molar polarisability, $\alpha(\tilde{\nu})$, in this analysis.

Fortunately, in a CDHO approach, it is not necessary to directly calculate molecular properties in order to determine refractive index data from measured extinction spectra. The imaginary components of $\alpha(\tilde{\nu})$, $\epsilon(\tilde{\nu})$ and $m(\tilde{\nu})$ all scale and are directly related to oscillator absorption, so a **relative calibration** can be performed to determine speciation in the absence of a direct determination of molecular polarisability.

2.3.6 Overview of the forward model

It is convenient at this point to review the underlying principles and advantages of using the CDHO model in the forward model for deriving refractive indices. Although this is the first use of a band model method for refractive index determination in this field, the CDHO model has previously been used for purposes of modelling dispersive absorption of solutions at room temperature. The inspiration for the approach here originates, in part, from the application of the CDHO model by Bertie and Zhang (1992) in simulating absorption of methanol and acetic acid for comparison with the more computationally expensive Kramers-Kronig method. This group published other papers on the subject, notably Bertie *et al.* (1994), which includes a detailed description of the determination of molecular band intensities from refractive index, dielectric constant and polarisibility spectra. Details of other applicable models for band structure in liquids are relatively sparse (Palamarev and Georgiev, 1994; Libnau *et al.*, 1994) since fundamental knowledge of the molecular structure of low-temperature solutions, even for pure water, is poor.

The CDHO model is the best representation available for the wavenumber dependent band shape of each oscillator in the aerosol measured for this thesis. The two major advantages of this model are that:

1. The band parameters contain fundamental information pertaining to the speciation (composition) of the aerosol;
2. A single spectrum can be characterised by relatively few variables (three per band) compared to the number of measurement points in a typical extinction spectrum.

The second point is of major significance to this study, since in a Kramers-Kronig analysis the retrieval state vector includes every wavenumber element plus the size parameters, leading to the necessary measurement of two spectra (with distinctly different size distributions) so that state vector elements do not outnumber measurement points. In the case of the band model retrieval, the state vector is now composed of N, r_m, S, m_∞ (section 2.5.7) as well as n sets of $S_j, \tilde{\nu}_j, \gamma_j$. Therefore, there will be more measurements than unknowns provided the number of measurements is greater than $4 + 3 \times n$. In this work, there are typically about 30 separate bands, yielding 94 state vector elements, compared with 1880 spectral points in a typical extinction spectrum. This permits determination of band parameters, and hence refractive index, from a single spectrum, and lowers computation time dramatically compared to a Kramers-Kronig transform.

2.3.7 Potential for infrared sensing of PSCs

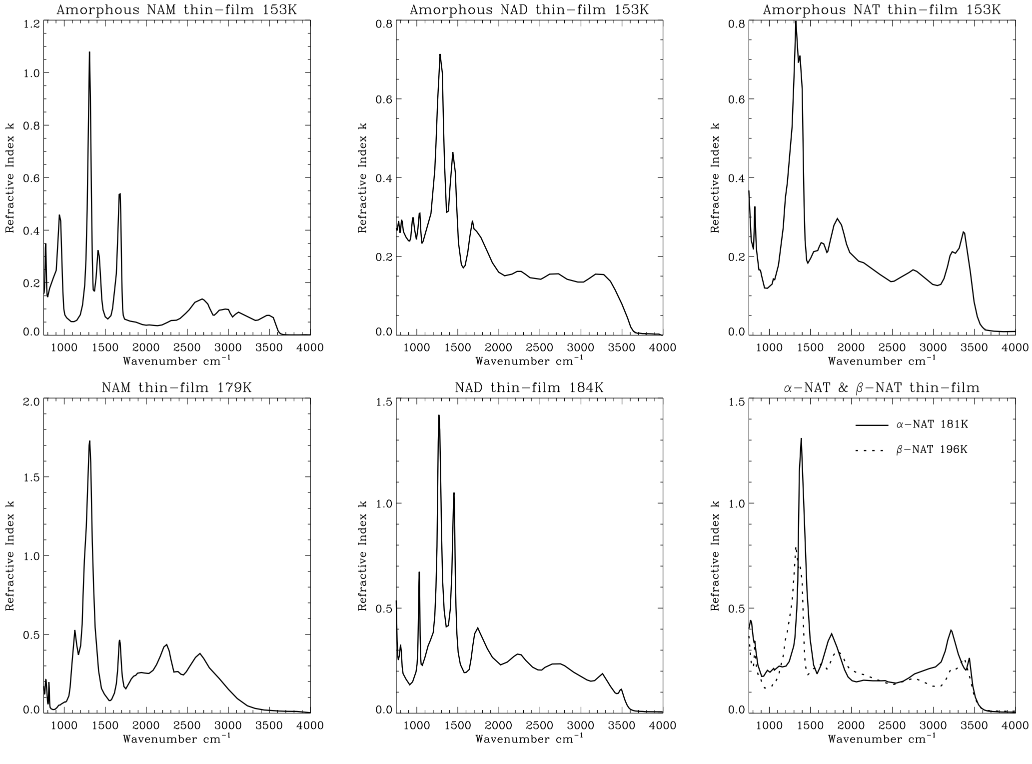

Figure 2.6: Imaginary refractive indices of PSC type thin-films (Toon *et al.*, 1994).

As identified in section 2.3.5, the absorption bands identified through the fitting of measured spectra theoretically relate to different chemical species in solution. Laboratory evidence has also revealed that different solid hydrates also have distinct absorption band *fingerprints*, as can be seen from the comparison of imaginary refractive index data reviewed by Toon *et al.* (1994), in figure 2.6.

Infrared spectroscopy therefore offers considerable scope for the detection and remote sensing of PSCs. Remote sensing of PSCs in the infrared has been undertaken with some success, as discussed in section 2.2.2, and the dependence of the volume and surface area of PSCs, retrieved from spectral extinction measurements, on the refractive index was established in section 2.2.3. Thus, the variation of infrared absorption with PSC type/composition suggests that both an applicable set of refractive index data should be used for PSC infrared sensing, and that the refractive index data should be an accurate representation of PSCs in the stratosphere. Having established the necessity of representative refractive index data of PSCs for the accurate measurement of PSC volume/surface area, and potentially determination of PSC type/composition, laboratory-generated refractive index data are reviewed and assessed in the following section.

2.4 Overview of laboratory measurements

Laboratory experiments have potential to provide information to characterise both the optical and the microphysical properties of PSCs. Knowledge of microphysical properties is very important for formulation of accurate microphysical models of PSC climatology, and thus major issues are highlighted here for completeness. However, since the primary aim of this work is the generation of refractive indices for ternary solution aerosol, the main focus of the following review of published data is laboratory measurements of PSC optical properties, principally refractive indices.

Two broad categories of laboratory experiments can be defined – aerosol particle experiments and thin-film experiments. Aerosol experiments involve generation of PSC type particle(s), usually in a settling chamber or a flowing system. Temperature, pressure and flow rates (hence particle residence times) of component gases and aerosol can be varied to achieve desired results. Thin-film experiments involve the deposition of layer(s) on a cold substrate (usually silicon) from vapour phase, and measurement of the properties of the mimic-PSC thin-film.

The main groups that have published data in this field are introduced in this section. None of the refractive index data has been published with associated wavenumber-dependant error estimates, although publications commonly give an estimate for wide spectral regions.

The University of North Carolina (UNC). This group published spectral properties and derived refractive indices for liquid binary solutions and solid forms of PSCs, typically over a spectral range of approximately 750-5000 cm^{-1}:

- Water ice 130-210 K (Clapp *et al.*, 1995);
- Nitric acid dihydrate 160-190 K (Niedziela *et al.*, 1998a);
- Nitric acid trihydrate 150-210 K (Richwine *et al.*, 1995);
- Binary H_2SO_4 aerosol 200-280 K (Niedziela *et al.*, 1998b, 1999);
- Binary HNO_3 aerosol 220 K (Norman *et al.*, 1999).

Max Planck Institute (MPI) and ETH Zürich. This group determined thermodynamic and spectral properties of thin-film PSCs and single particles. Refractive indices were measured as indicated below:

- The measurement of spectra for a wide composition/temperature range of binary H_2SO_4 and HNO_3 solutions between 500-5000 cm^{-1}, derivation of refractive indices, and formulation of a *ternary calculator* to combine binary indices to synthesise ternary indices (Biermann *et al.*, 2000);

- The freezing of bulk binary H_2SO_4 solutions (Koop *et al.*, 1995, 1997a,b);
- The freezing and melting points of single particles of binary H_2SO_4 solution at compositions greater than 20 wt% H_2SO_4 (Krieger *et al.*, 2000a);
- Temperature dependent visible refractive index (m_∞) values (and densities) for ternary solutions from measurements and theoretical models (Luo *et al.*, 1996; Krieger *et al.*, 2000b).

University of Colorado. Published analysis from thin-film experiments:

- Refractive indices of nitric-acid ices (Middlebrook *et al.*, 1994);
- Spectra of binary H_2SO_4 and HNO_3 and ternary solutions, with composition estimates from areas of absorption features (Anthony *et al.*, 1997);
- Refractive indices of binary H_2SO_4 solutions (Tisdale *et al.*, 1998).

University of Waterloo. Published analysis from aerosol experiments:

- The freezing point of binary H_2SO_4 solutions between approximately 19-35 wt% (Bertram *et al.*, 1996);
- Measurement of extinction spectra and determination of freezing point temperature (glass point) on cooling of binary HNO_3 solutions, and measurements of crystallisation into nitric acid hydrates (NAHs) upon subsequent warming (annealing), using solutions of initial composition 50-75 wt% (Bertram and Sloan, 1998b,a; Bertram *et al.*, 2000).

Other published refractive index data:

- Querry and Tyler (1980) derived refractive indices and absorption band properties for bulk solutions of HNO_3 over the spectral range 310-5000 cm^{-1}, at a single temperature of 300 K;

- Palmer and Williams (1975) calculated refractive index data sets from bulk measurements of a range of binary H_2SO_4 solutions over the spectral range 400-27778 cm^{-1}, at a single temperature of 300 K;

- Adams and Downing (1986) determined refractive indices from bulk measurements of a single ternary solution composition and a single composition solution of HNO_3 over the spectral range 500-5000 cm^{-1}, at a temperature of 298 K. This is the only known publication of refractive indices derived directly from a ternary solution measurement.

2.5 Laboratory measurements of liquid phase PSCs

Laboratory experiments to measure the spectral properties of liquid PSCs have been undertaken using both bulk (including thin-film) solutions and aerosol. Although liquid PSCs are likely to consist of ternary solutions of aqueous sulphuric and nitric acid, almost all experiments have also studied the properties of binary sulphuric acid and binary nitric acid. Experiments with binary solutions are both easier to characterise in terms of composition and real refractive index (as foreknowledge of m_∞ is required for refractive index derivation), and offer potential for use of mixing rules to model ternary refractive indices (Biermann *et al.*, 2000). The principal refractive index data and main experimental methods used in their production are now reviewed.

2.5.1 Binary sulphuric acid solutions

Sulphuric acid experiments have a relatively long history, primarily due to interest in the radiative forcing due to stratospheric sulphate aerosol on climate, and study of Venus' atmosphere. Early studies of optical constants of sulphuric acid were conducted with bulk solutions over a limited range of temperatures and compositions (Palmer and Williams, 1975; Pinkley and Williams, 1976). The refractive indices of Palmer and Williams (1975) have been widely used in atmospheric studies (Wooldridge *et al.*, 1995; Grainger *et al.*, 1995). The Palmer and Williams data show that refractive indices have a strong composition dependence, but are limited by the high-temperature (300 K) of measurements. As demonstrated by Grainger *et al.* (1995), although temperature effects can be accounted for through the Lorentz-Lorenz relationship, this correction is only valid away from absorption bands (i.e. where $k(\tilde{\nu}) = 0$). Through comparison of indices of the same strength sulphuric acid solutions from the 300 K Palmer and Williams (1975) data and the 250 K Pinkley and Williams (1976) data, Grainger *et al.* identify "significant differences" between the band maxima in $k(\tilde{\nu})$; these cannot be attributed to temperature differences alone since the Lorentz-Lorenz correction from 300 K to stratospheric temperatures changes absorption by 5 % or less – an order of magnitude less than the observed discrepancy. Grainger *et al.* concluded that further laboratory experiments are required to provide a consistent set of sulphuric acid refractive index data.

Recent studies of infrared properties of sulphuric acid have increased the temperature and composition range of measurements considerably (Niedziela *et al.*, 1998b; Bertram *et al.*, 1996; Tisdale *et al.*, 1998; Biermann *et al.*, 2000), revealing a strong temperature dependence of certain absorption features. The UNC group measured sulphuric acid in aerosol form (Niedziela *et al.*, 1999, 1998b) at temperatures from 200-300 K, with a range of compositions at each temperature. In a recent publication (Niedziela *et al.*, 1999), the UNC group compare their aerosol spectra with those produced in thin-film experiments by Tisdale *et al.* (1998). Niedziela *et al.* find that, despite the very different experimental approaches, the two data sets are on the whole in very good agreement. However, the thin-film indices of Tisdale *et al.* (1998) fail to reproduce the extinction profile observed in the aerosol spectra of Niedziela *et al.* around 1100 cm^{-1} – Niedziela *et al.* do not provide an explanation. Biermann *et al.* (2000) also compare their thin-film sulphuric indices with those of Niedziela *et al.* and Tisdale *et al.*, and note general agreement but also "remarkable differences" in all three data-sets in absorption around 1100 cm^{-1}. A comparison of UNC and Biermann *et al.* imaginary refractive indices can be seen in panels (a) and (b) of figure 2.7.

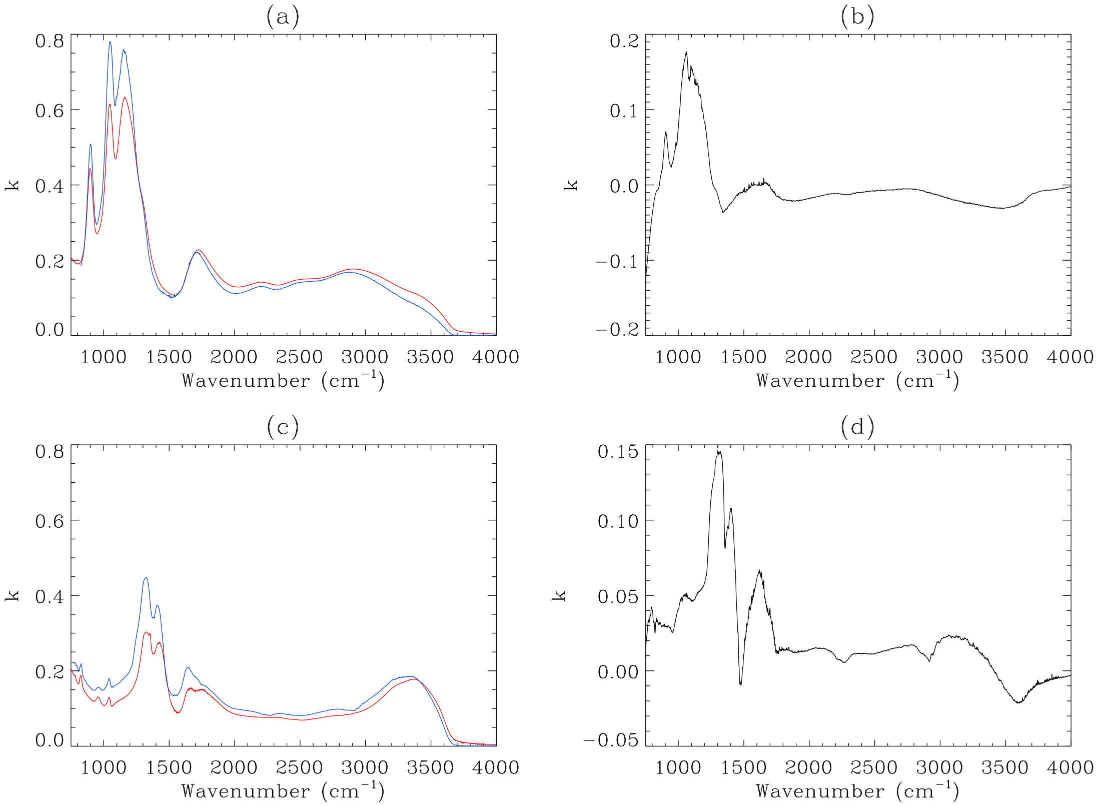

Figure 2.7: Comparison of imaginary refractive indices of Biermann *et al.* (red) and UNC refractive indices (blue). Panel (a) shows sulphate solutions of 75 wt%, 293 K (Biermann) and 76 wt% 280 K (UNC), with the difference UNC-Biermann in (b). Panel (c) shows nitrate solutions of 45 wt% 223 K (Biermann) and 45 wt% 220 K (UNC), with the difference plotted in (d).

2.5.2 Binary nitric acid solutions

Room temperature refractive indices have been determined from bulk solutions of nitric acid by several groups (Querry and Tyler, 1980; Boone *et al.*, 1980; Adams and Downing, 1986), but only recently have indices of low temperature nitric acid solutions been published. Data are now available from the 220 K aerosol experiments of the UNC group (Norman *et al.*, 1999), and the multi-temperature thin-film experiments of Biermann *et al.* (2000). Norman *et al.* compare their indices with bulk data published by Querry and Tyler (1980) and the amorphous solid films published by Toon *et al.* (1994), and identify significant differences in absorption bands. It is also evident through comparison of the Norman *et al.* (1999) and Biermann *et al.* (2000) indices for close temperature and composition conditions, see figure 2.7(c) and (d), that significant differences in absorption exist as with the sulphuric acid measurements. These differences are a key issue, and appear to result from a constant offset in absorption across the infrared spectral range rather than differences between specific features, as seen with sulphuric acid spectra – this may be due to differential scaling of the refractive indices arising from the various approaches used to derive these data. However, the main point here is that the observed temperature dependence of binary nitric acid is very weak (Biermann *et al.*, 2000), compared with binary sulphuric acid solutions.

Without detailed knowledge of the analysis conducted by Biermann *et al.* (2000) and Norman *et al.* (1999) to produce refractive indices, it is not possible to positively identify the origin of the observed differences in refractive index scaling. However, from the information given by the authors (and personal communication), possible causes for these discrepancies can be suggested, including: (1) unattributed effects resulting from different experimental approaches, and (2) variation resulting from different analytical techniques. Indeed, the different analytical methods used by these groups to derive refractive indices could result in scaling offsets in the imaginary refractive index. For calibration of thin-film thickness, Biermann *et al.* used room-temperature optical constants from tabulated literature sources (Palmer and Williams, 1975; Querry and Tyler, 1980). The fact that these data are room temperature and that differences between published experimental data are common may lead to problems with this assessment of film thickness, and to scaling errors in the refractive index (Biermann *et al.*, 2000). Moreover, scaling problems may also arise from the analysis conducted by Norman *et al.* (1999). Since Norman *et al.* could not produce the more dilute nitrate aerosol particles without freezing the particles at radii below about 0.1 μm, another method was used to produce a better approximation to a non-scattering spectrum in these cases for Kramers-Kronig analysis. A spectrum of the smallest liquid aerosol was modified by subtracting a theoretical scattering curve (a function of $\tilde{\nu}^4$) in order to produce the non-scattering component. This process may introduce refractive index scaling offsets through simplistic representation of size distribution dependent scattering, and could also be responsible for the anomalous zero $k(\tilde{\nu})$ values in some UNC nitrate aerosol data sets.

2.5.3 Ternary solutions

Very little data has been published relating to the optical properties of ternary solutions. The only set of refractive index data directly measured from a ternary solution is the room temperature bulk solution measurement of Adams and Downing (1986), with a composition of 75 % H_2SO_4, 10 % HNO_3, and 15 % H_2O.

Spectra, but not refractive indices, of ternary solutions have been measured in other laboratory studies (Anthony *et al.*, 1997; Biermann *et al.*, 2000). Anthony *et al.* (1997) measured infrared spectra of binary and ternary solution aerosol of compositions of relevance to the stratosphere. The main emphasis of the work of Anthony *et al.* was to measure freezing temperatures of ternary solution aerosol, and refractive indices of the aerosol were not determined.

The only published source of refractive index data for low-temperature ternary solutions over a range of compositions (published after this work commenced) is that produced by Biermann *et al.* (2000); these indices are not directly derived from measured ternary solution spectra, but are instead calculated from binary thin-film solutions using a *ternary solution calculator.* Ternary solution refractive indices are evaluated from the *linear* summation of binary H_2SO_4 and HNO_3 solution refractive indices, weighted by weight percentage composition of the ternary solution. The composition and temperature of the binary indices used in the summation are selected from the Biermann *et al.* (2000) binary solution measurements for the closest match to the ternary composition and temperature. The ternary calculator relies on the key assumption that absorption (specifically the imaginary refractive index, $k(\tilde{\nu})$) of sulphate, nitrate and water ions in a ternary solution is very similar to that in the corresponding binary solution. This assumption that there are no infrared absorption features specific to the ternary system is supported by laboratory evidence (Anthony *et al.*, 1997; Koehler *et al.*, 1992).

2.5.4 Temperature dependence of sulphuric acid in ternary solutions

During the early experimental stages of this thesis, work was undertaken with project partners at the Rutherford Appleton Laboratory (RAL) within the framework of the European Union funded POSTCODE project (Remedios *et al.*, 2000). The main task involved comparing a composite spectrum of STS aerosol (McPheat *et al.*, 2002), measured in the early stages of this work, with optical data produced by the ternary calculator of Biermann *et al.* (2000). A single composite aerosol extinction spectrum, composed of two spectra measured with different detectors to extend the spectral range from 750 to 13000 cm^{-1}, was composed.[7] The temperature of the STS aerosol was identified as 182 ± 5 K; the temperature error improved in later experiments following the system characterisation undertaken for this thesis. Although this temperature is several degrees below T_{ice}, the STS aerosol remained fully liquid since no spectral features associated with freezing were detected; this can be accounted for by the fact that up to 4 K of supercooling below T_{ice}, and thus a temperature below approximately 184 K (within the error of the measured temperature), may be required to homogeneously freeze liquid aerosol to ice (section 1.3.3).

The Biermann *et al.* (2000) ternary solution refractive index calculator was employed to calculate refractive indices at the measurement temperature over a range of compositions. An adapted Mie forward model code was used to determine aerosol size distribution parameters from the composite spectrum using a log-normal particle distribution over the cell path-length. The size distribution was determined by fitting the spectral region above 4500 cm^{-1}, where absorption is negligible in this case. The refractive index data output by the Biermann *et al.* ternary calculator were used to generate simulated extinction spectra using the Mie code, which were then compared with the composite spectrum. The best fit to the composite spectrum was obtained by using indices calculated by the ternary calculator for a composition of 20.9 wt% H_2SO_4, 12.0 wt% HNO_3. The best fit extinction spectrum matches the composite spectrum to within 10 % between 1300 to 3400 cm^{-1}, and above

---

[7] Flow rates of SO_3/N_2 = 400 sccm, $H_2O/HNO_3/N_2$ = 200 sccm, and H_2O/N_2 = 400 sccm were used (1000 sccm total).

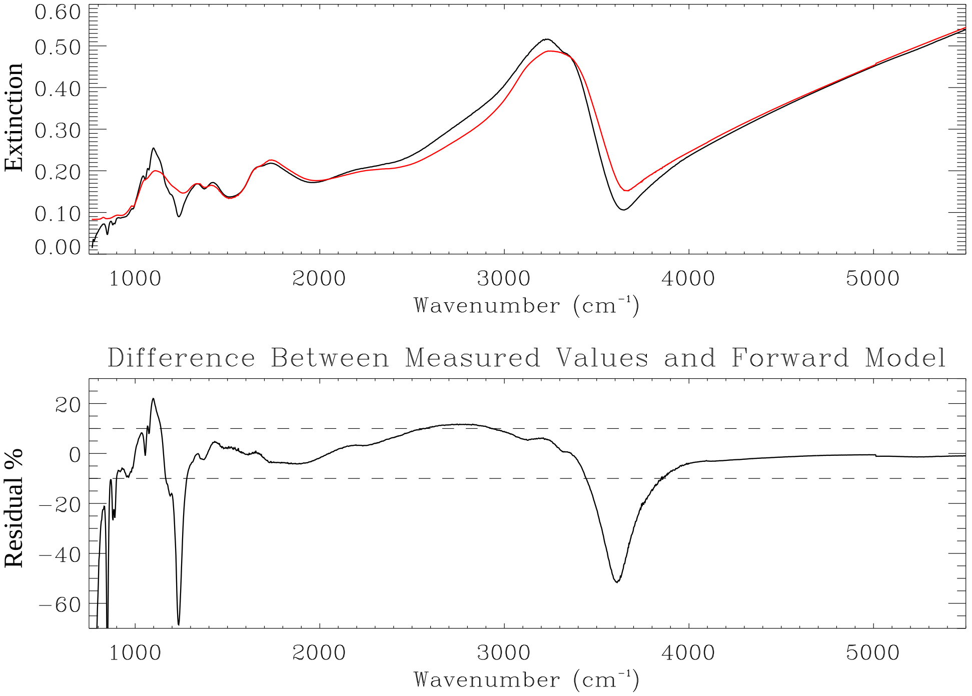

Figure 2.8: Best fit of Biermann *et al.* (2000) refractive indices to the laboratory STS aerosol composite spectrum at 182 K. Top panel shows relative extinction (black line is the composite spectrum, red line is the simulated spectrum), and lower panel shows the percentage residual. Particle density $N = 2.4 \times 10^7$ cm^{-3}, median radius $r = 0.18$ μm, and log-normal spread $\sigma = 1.99$. The composition of the best fit Biermann *et al.* indices is 20.9 wt% H_2SO_4, and 12.0 % HNO_3.

4000 cm^{-1}, as can be seen in figure 2.8. The divergence below 1300 cm^{-1}, and principally the differences in the sulphate feature around 1100 cm^{-1}, have been reported elsewhere (McPheat *et al.*, 2002; Bass *et al.*, 2000a,b) – similar differences have also been observed between other published data (Biermann *et al.*, 2000; Niedziela *et al.*, 1999).

As discussed in section 2.5.1, Niedziela *et al.* (1999) note differences around 1100 cm^{-1} between their aerosol data and thin-film indices of Tisdale *et al.* (1998), and suggested that fundamental differences exist between thin-film and aerosol refractive indices. However, in this case a check of the binary data used for the Biermann *et al.* ternary index calculation reveals that the nearest binary H_2SO_4 composition measurement is a 20 wt% H_2SO_4 solution at 253 K. This emphasises a limitation in the use of binary thin-film data for calculation of ternary indices – below freezing temperature, thin-film solutions heterogeneously freeze on the substrate preventing significant supercooling as may be observed in stratospheric aerosol, a concern for dilute solutions which have higher freezing temperatures. In ternary solutions where either the H_2SO_4 or HNO_3 component is relatively dilute, as is the case regarding H_2SO_4 for the composite spectrum and at low temperatures close to T_{ice} in the stratosphere (section 1.3.3), the corresponding binary measurement cannot be performed in thin-film form at temperatures as cold as the polar stratosphere. Additionally, Lorentz-Lorenz correction cannot account for the temperature dependence of absorption features (Grainger *et al.*, 1995), so accurate representation of temperature dependence of ternary solutions is impeded.

2.5.5 Ternary refractive indices from binary measurements

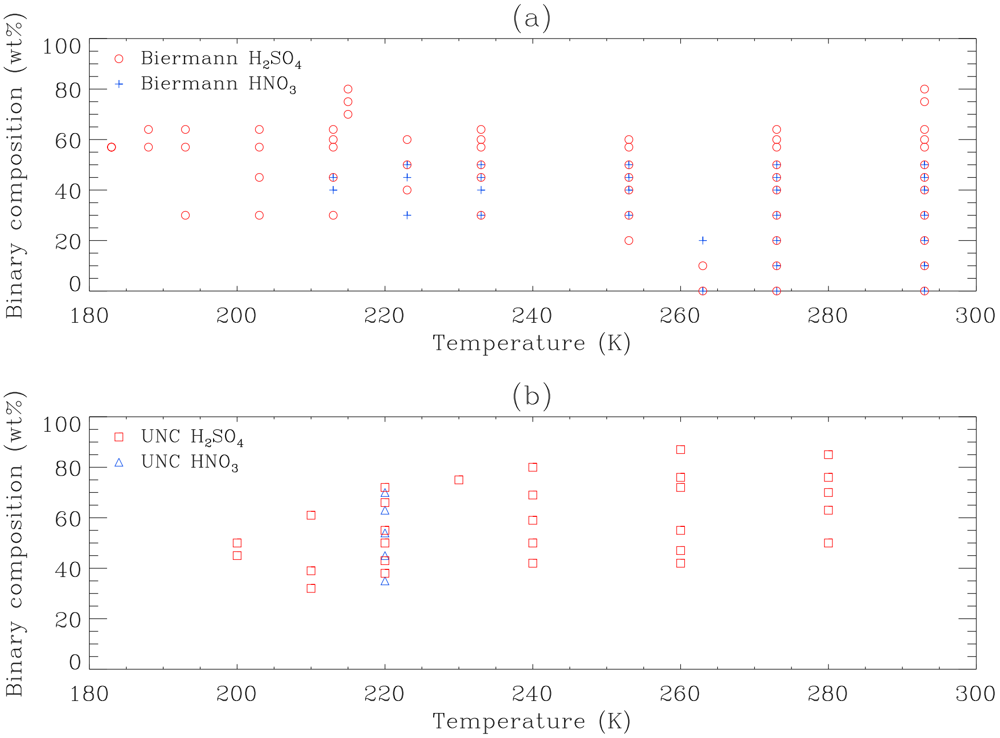

Figure 2.9: The composition and temperature of refractive indices in (a) the Biermann *et al.* (2000) data, and (b) the UNC data.

In addition to the clear relationship between infrared absorption bands and composition for both H_2SO_4 and HNO_3 features, the temperature dependence of H_2SO_4 has been identified in the ternary system and from published binary solution data. Conversely, HNO_3 absorption bands appear to exhibit a much weaker temperature dependence. Therefore, in order to characterise the infrared absorption features of sulphuric and nitric acids for use in remote sensing, it is important to undertake measurements at stratospherically-relevant compositions *and* temperatures (for H_2SO_4 solutions in particular). So far, the most comprehensive sets of binary solution refractive index data that fulfil these criteria are the binary H_2SO_4 and HNO_3 indices produced in the aerosol experiments of the UNC group (Niedziela *et al.*, 1998b, 1999; Norman *et al.*, 1999), and the binary H_2SO_4 and HNO_3 indices measured in the thin-film experiments of Biermann *et al.* (2000). The range of compositions and temperatures covered by these data is represented in figure 2.9. These refractive indices are used later in this thesis for the determination of composition from retrieved band parameters (section 7.1).

Biermann *et al.* (2000) showed that thin-film binary refractive indices could be combined in their ternary calculator to reproduce direct measurements of ternary solution refractive indices, although some discrepancies are evident in this comparison. The critical questions regarding use of the Biermann *et al.* ternary calculator are:

1. Do the thin-film data reproduce the refractive indices of aerosol in the stratosphere?
2. Does it cover an appropriate composition and temperature range to reproduce data of relevance to ternary solution PSCs?
3. What are the errors in the binary refractive index data *and* the what are the combined errors in the ternary data from (1) binary indices, and (2) and model uncertainties?

In response to the first question, it is not straightforward to assess whether the differences observed between published data result from experimental approach or thin-film/aerosol differences. The differences in the temperature dependent sulphuric acid feature around

1100 cm^{-1} between data sets has been noted (section 2.5.1). Biermann *et al.* (2000) attribute this variation to unquantified experimental differences. Detailed analysis of published low-temperature binary H_2SO_4 refractive indices for both thin-film and aerosol forms, conducted for band parameter calibration against composition later in this thesis (section 7.1), revealed a consistent difference between sets of published data in sulphuric acid absorption bands around 1100 cm^{-1} between thin-film and aerosol measurements. From the information provided by the authors, it is difficult to assess the likely cause, although it is speculated here that this may be a result of differences between spectroscopic behaviour of thin-film and aerosol solutions of H_2SO_4.

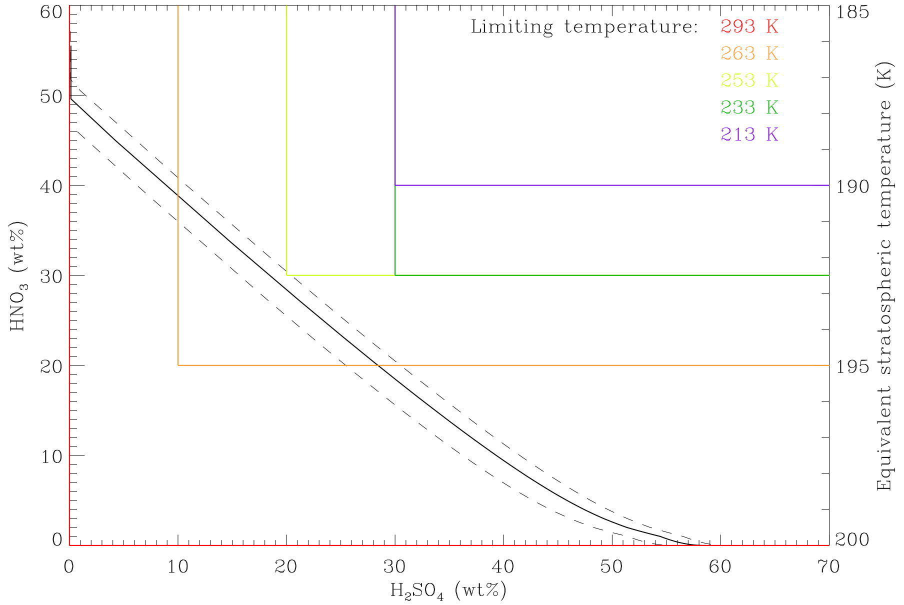

Figure 2.10: Results from the thermodynamic model of Lin and Tabazadeh (2001) showing the range of ternary solution compositions in the stratosphere; the **equivalent temperature in the stratosphere** used for this calculation is shown on the right axis. The thick black line represents typical stratospheric HNO_3 and H_2O vapour concentrations of 5 ppmv and 10 ppbv respectively, and the dashed lines represent likely variation of ± 50 % in H_2O and ± 20 % in HNO_3. The regime *inside* (to the top right hand side of the lines) the coloured boxes indicates the composition range measured by Biermann *et al.*, i.e. the lines represent the most dilute measurements at each temperature.

The second question can be investigated by comparing the range of compositions measured by Biermann *et al.*, figure 2.9(a), with the expected range of compositions in the stratosphere, represented earlier in figure 1.7, but reproduced in a different form here to compare against the measurements of Biermann *et al.* in figure 2.10. It can be seen that the binary refractive index data used for the Biermann *et al.* ternary calculator encompasses the entire range of expected compositions for ternary solutions in the stratosphere at 293 K. However, the critical point is that at the lower end of the Biermann *et al.* temperature range, i.e. at temperatures approaching those relevant to ternary aerosol system in the stratosphere (see right hand axis in figure 2.10), the measured compositions are too concentrated to be of relevance to the stratosphere. This is only a problem if solutions are found to exhibit a strong temperature dependence. Although the temperature dependence of HNO_3 solutions is weak, a stronger temperature dependence in some absorption features of binary H_2SO_4 solutions (and H_2SO_4 features in ternary solution) has been identified. It is not known whether the temperature dependence of absorption features in ternary solution differ from those in binary solution. This limitation of the Biermann *et al.* data may explain the discrepancies observed between the Biermann *et al.* ternary calculator indices and spectra produced in this and related work (section 2.5.4), and may constrain the applicability of Biermann *et al.* ternary calculator in analysis of ternary solution PSCs in the stratosphere. Thus, assuming the temperature dependence of HNO_3 is negligible, the Biermann *et al.* ternary calculator may only accurately represent ternary refractive indices with H_2SO_4 components above 30 wt%, the most dilute H_2SO_4 concentration measured at 193 K.

The lack of low-temperature measurements in the Biermann *et al.* data is probably due to experimental limitations of thin-film experiments. It is not possible to achieve the supercooling required for the stratosphere for dilute acid concentrations of relevance; at stratospheric temperatures the substrate surface initiates (heterogeneous) freezing of the solution below frost temperature. Despite the apparent advantage of aerosol experiments in this respect, the UNC data simply does not contain H_2SO_4 measurements in this low composition/low temperature regime. The only other ternary solution refractive index data available for comparison with the ternary calculator of Biermann *et al.* are the bulk solution indices of a 75 % H_2SO_4, 10 % HNO_3, and 15 % H_2O solution, directly measured by Adams and Downing (1986) at 298 K. However, this composition is outside the range of validity (total acid concentration must be less than 80 wt%) of the Biermann *et al.* ternary calculator, so it is not relevant to compare these data.

Finally, wavenumber-dependant errors for the binary refractive indices of Biermann *et al.* (2000) were not published. Without these uncertainties, assessment of whether the observed differences between output of the ternary calculator and the selected ternary measurements used for testing the ternary calculator by Biermann *et al.* cannot be attributed to either errors in the binary data, or uncertainties resulting from use of the ternary calculator.

2.5.6 Determination of composition

One of the most challenging aspects of temperature-dependant experimental studies appears to be the accurate determination of liquid composition (Niedziela *et al.*, 1999). In thin-film studies composition is commonly indirectly determined from characterisation of the vapour deposition process (Toon *et al.*, 1994; Iraci *et al.*, 1998), or film analysis using temperature programmed desorption methods (Koehler *et al.*, 1992). In the aerosol experiments of the UNC group, high resolution measurements of gas lines were undertaken with a tuneable diode laser to characterise vapour pressure of H_2O (Niedziela *et al.*, 1999) or H_2O and HNO_3 (Norman *et al.*, 1999). With knowledge of the temperature, only the vapour pressure of H_2O need be known to calculate the composition of binary H_2SO_4 or HNO_3 aerosol from appropriate thermodynamic theory to a quoted accuracy of ±1 % (Niedziela *et al.*, 1999), assuming the aerosol is in thermodynamic equilibrium. However, the vapour pressure of HNO_3 was also measured in the nitric acid aerosol experiment of Norman *et al.* (1999) as a double-check on the accuracy of composition determination and to verify HNO_3 gas measurement in advance of a planned ternary aerosol measurement where both H_2O and HNO_3 vapour pressures are required to determine composition.[8]

---

[8] Equilibrium concentrations of H_2SO_4 gas are so low that this gas can not be practically measured in laboratory aerosol experiments.

An alternative approach to aerosol composition determination was adopted by Anthony *et al.* (1997). The areas below selected bands from measured absorption spectra, representative of H_2SO_4, HNO_3 and H_2O, were measured by drawing a line across the base of bands and assessing the absorption band area. The ratio of a single sulphate band area and two nitrate band areas against a water band area were calibrated against the composition of binary species, and then used to assess composition of ternary aerosol. This method is less accurate than the vapour-based technique reported by Niedziela *et al.* (1999), with a quoted composition accuracy of ± 3 wt% for HNO_3 and H_2SO_4, and ± 6 wt% for H_2O.

2.5.7 Experiments to determine the real refractive index for liquid PSCs

Laboratory experiments (Luo *et al.*, 1996; Krieger *et al.*, 2000b) have demonstrated that the refractive index at high wavenumbers (above approximately 4000 cm^{-1}), m_∞ by definition, is a function of solution composition and temperature. The importance of m_∞ is that, in addition to the size distribution of aerosol (or thickness of film), it defines the extinction of the medium in the scattering dominated spectral region above 4000 cm^{-1} (towards visible wavelengths). Krieger *et al.* (2000b) present an empirical fit to measured thin-film binary data for purposes of modelling the real refractive index at visible wavelengths for the ternary solution system over a range of compositions and temperatures; temperature-dependant variation of the refractive index is modelled by the Lorentz-Lorenz relationship. Although the model of Krieger *et al.* (2000b) is in agreement with previous laboratory work (Luo *et al.*, 1996), some field studies have problems relating theoretical values to observations of the stratosphere as discussed previously in section 2.1.4.

2.5.8 Band parameters compiled from literature

Table 2.1: Table of published band parameters.

Molecule	Band	$\tilde{\nu}(\mathrm{cm}^{-1})_{\mathrm{max}}$	$\gamma(\mathrm{cm}^{-1})$	$k(\tilde{\nu})_{\mathrm{max}}$
$_{\mathbf{B}}H_2SO_4$ >70wt%		564		
$_{\mathbf{Q}}H_2O$	ν_L	580	500	0.443
$_{\mathbf{B}}SO_4{}^{2-}$		593		
$_{\mathbf{B}}HSO_4{}^{-}$		613		
$_{\mathbf{Q}}HNO_3$	$\nu_6(A')$	691 ± 2	26 ± 6	0.256
$_{\mathbf{Q}}NO_3{}^{-}$	$\nu_4(E)$	730 ± 7		
$_{\mathbf{Q}}HNO_3$	$\nu_8(A'')$	778 ± 2	15 ± 2	0.259
$_{\mathbf{Q}}NO_3{}^{-}$	$\nu_2(A_2)$	820 ± 3	35 ± 10	0.198
$_{\mathbf{B}}SO_4{}^{2-}$		885		
$_{\mathbf{B}}H_2SO_4$ >70wt%		905		
$_{\mathbf{Q}}HNO_3$	$\nu_5(A')$	949 ± 2	49 ± 4	0.361
$_{\mathbf{B}}H_2SO_4$ >70wt%		965		
$_{\mathbf{B}}SO_4{}^{2-}$		1030		
$_{\mathbf{B}}SO_4{}^{2-}$		1050		
$_{\mathbf{B}}HSO_4{}^{-}$		1104		
$_{\mathbf{B}}H_2SO_4$ >70wt%		1170		
$_{\mathbf{Q}}H_3O^+$	$\nu_2(A_1)$	1210 ± 40		
$_{\mathbf{Q}}HNO_3$	$\nu_4(A')$	1304 ± 2	50 ± 5	0.546
$_{\mathbf{B}}SO_4{}^{2-}$		1341		
$_{\mathbf{Q}}NO_3{}^{-}$	$\nu_3(E)$	1350 ± 2	157 ± 10	0.099
$_{\mathbf{B}}H_2SO_4$ >70wt%		1370		
$_{\mathbf{Q}}HNO_3$	$\nu_3(A')$	1429 ± 2	98 ± 6	0.294
$_{\mathbf{Q}}H_2O$	ν_2	1640	110	0.137
$_{\mathbf{Q}}HNO_3$	$\nu_2(A')$	1672 ± 4	47 ± 8	0.292
$_{\mathbf{Q}}H_3O^+$	$\nu_4(E)$	1742 ± 15	180 ± 20	0.148
$_{\mathbf{Q}}HNO_3$	$2\nu_4$	2633 ± 20		0.082
$_{\mathbf{Q}}HNO_3$	$2\nu_3$	2935 ± 20		0.071
$_{\mathbf{Q}}H_2O$	ν_1,ν_3,$2\nu_2$	3395	390	0.297

$_{\mathbf{Q}}$After Querry and Tyler (1980).
$_{\mathbf{B}}$After Biermann *et al.* (2000).

To assist in formulation of a physically-realistic *a priori* and first guess set of band parameters for the CDHO model, an extensive literature survey was conducted to identify physically realistic band parameters for liquid solutions of nitric and sulphuric acid. In spite of numerous publications concerning thin-film and (fewer) aerosol measurements, little data exists on the fundamental physical properties of individual oscillators from spectra. Indeed, the most comprehensive set of band parameters dates back to Querry and Tyler (1980), although these measurements were limited to room temperature (300 K) bulk solutions of HNO_3. Querry and Tyler (1980) present the central frequencies, half-widths and strengths of infrared absorption bands due to vibrational modes of liquid water and intra-molecular vibrational modes of NO_3^-, H_3O^+ and HNO_3. The Querry and Tyler band strength S_b is evaluated from $S_b = \frac{1}{\rho_0} \int_{band} k(\tilde{\nu})\ d\tilde{\nu}$ where ρ_0 is the molecular density [molecules/cm^3]; this is essentially the area under a band corrected for density in the imaginary refractive index. Thus, S_b values are not directly interchangeable with band intensities S_j, as used in the CDHO model, which are not a representation of area under a peak but maximum intensity in imaginary dielectric space. Moreover, the refractive index location of band centres will differ slightly from $\tilde{\nu}_j$ positions in dielectric space, as demonstrated previously (equation 2.29); since these differences are small, they can be neglected for purposes of general comparison.

The Querry and Tyler (1980) bands are presented in order of ascending wavenumber in table 2.1. The band column in table 2.1 details the modes of vibration of the listed species as explained by Querry and Tyler. Additionally, a compilation of bands composed by Biermann *et al.* (2000) from other literature sources for liquid sulphuric acid are also included in table 2.1. Only central positions are listed for these bands, no other parameters or details concerning mode of vibration are given, but this completes the list of main absorbing species for sulphuric and nitric acid solutions.

2.6 Laboratory measurements of solid phase PSCs

Although the main focus of this thesis is to determine the optical properties of liquid STS aerosol, a brief review of published laboratory studies of solid PSCs is presented here to provide a comprehensive context.

2.6.1 Crystalline solids

The main forms of solid PSC (section 1.3.3) are likely to be comprised of pure water ice, NAT, and the more unstable nitric acid dihydrate (NAD, 1:2 HNO_3/H_2O). Refractive indices for pure water ice have been characterised over a range of temperatures, and data appears to be in broad agreement between groups (Downing and Williams, 1975; Kou *et al.*, 1993; Clapp *et al.*, 1995). However, a recent review of spectral properties of NAT and NAD films and aerosol reveals that there may be several spectroscopically different forms of these hydrates (Tisdale *et al.*, 1999).

Two spectroscopically-distinct types of NAT have been reported in literature. In thin-film experiments, below a temperature of 185 K in the initial cooling cycle, α-NAT forms. Upon annealing (warming) above 188 K, or through vapour deposition at temperatures above 183 K, α-NAT transforms to more stable β-NAT (Koehler *et al.*, 1992). Koehler *et al.* hypothesise that during the NAT conversion from α to β forms, the waters of hydration change from a mixture of random hydrogen bonds to a more regular array of hydrogen bonds. The refractive indices of both forms of NAT, in addition to NAD and nitric acid monohydrate (NAM, 1:1 HNO_3/H_2O), and amorphous NAM, NAD and NAT, were first measured in the thin-film experiments of Toon *et al.* (1994), and are illustrated in figure 2.6. In aerosol experiments, NAT has been produced by cooling liquid aerosol below glass temperature and annealing the resulting amorphous solid to induce crystal formation (Richwine *et al.*, 1995; Bertram and Sloan, 1998b). At the low glass point temperatures (around 150-160 K) required for solid particle nucleation in aerosol experiments, the α form of NAT precipitates from the amorphous solid, and the characteristic features of β-NAT have not been observed in aerosol experiments. However, it is expected that α-NAT can convert to β-NAT under stratospheric conditions (Koehler *et al.*, 1992), so the thin-film optical constants of β-NAT (Toon *et al.*, 1994) may be more appropriate for remote detection of NAT aerosol in the stratosphere (Tisdale *et al.*, 1999).

It has been proposed that heterogeneous nucleation on a substrate in thin-film ex-

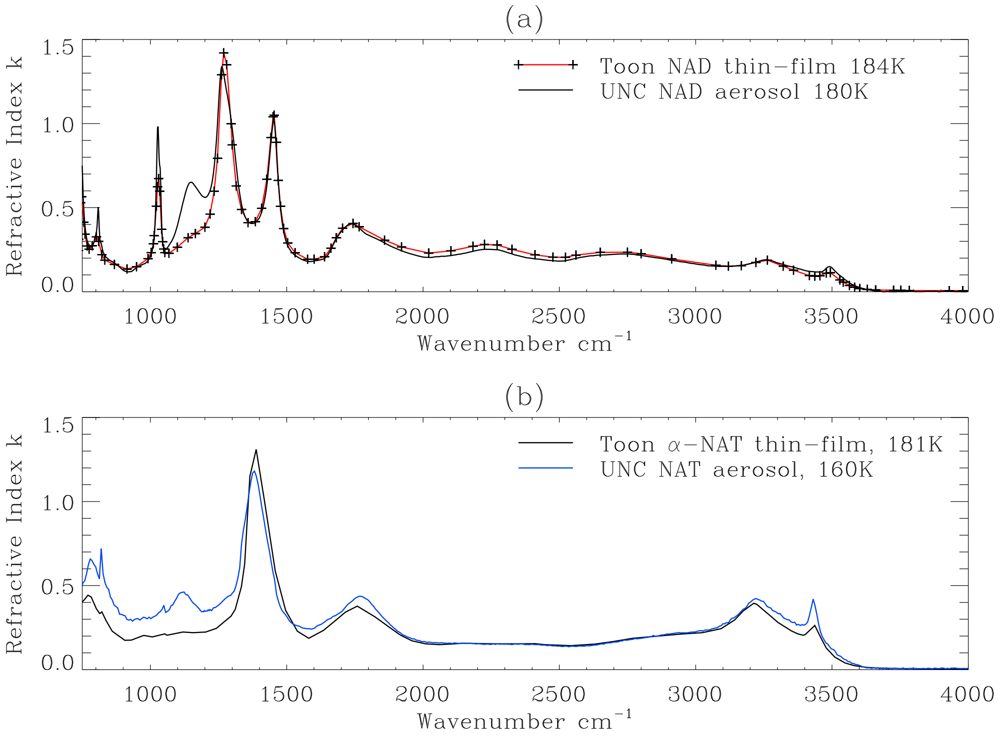

Figure 2.11: (a) NAD thin-film (Toon *et al.*, 1994) and aerosol (Niedziela *et al.*, 1998a) imaginary index comparison, data points have been plotted for the film index to show the resolution of this data, which is much lower than the aerosol data. (b) A comparison of NAT thin-film (Toon *et al.*, 1994) and aerosol (Richwine *et al.*, 1995) indices.

periments may lead to different spectral characteristics for crystalline forms resulting from interactions with the substrate, complicating the application of thin-film derived refractive indices for application in analyses of stratospheric PSC particles (Anthony *et al.*, 1997; Richwine *et al.*, 1995; Niedziela *et al.*, 1998a; Bertram and Sloan, 1998a). Niedziela *et al.* (1998a) argue that thin-film spectra from other experiments are different to each other and different to UNC aerosol spectra for crystalline forms. Figure 2.11 illustrates a comparison of the UNC aerosol indices (Richwine *et al.*, 1995; Niedziela *et al.*, 1998a) with thin-film indices for NAD and NAT (Toon *et al.*, 1994), at similar temperatures. Although the film indices are low resolution, clear differences can be seen in the residuals, which are not simply a result of intensity (or scaling) differences. In their review of optical properties of

NAH, Tisdale *et al.* (1999) suggest that optical constants measured using thin-films and aerosol are both suitable for remote infrared observation of PSCs, asserting that the main variables that influence spectral characteristics of NAT and NAD are formation conditions and temperatures. Tisdale *et al.* (1999) also suggest that NAT could be birefringent, with its optical signature highly dependant on crystallographic orientation; in this case no single set of refractive indices would be truly reliable for PSC detection.

2.6.2 Amorphous solids

By definition, amorphous solids have frozen without an organised crystalline lattice form; homogeneous precipitation takes place when cooled below nucleation temperature or *glass point* (Bertram and Sloan, 1998a). Bertram and Sloan observed that liquid aerosol particles of 1:2 HNO_3/H_2O stoichiometry remained liquid in a cooling cycle down to their glass temperature at around 160 K, when they solidified as amorphous 1:2 HNO_3/H_2O particles. Upon annealing the aerosol gradually crystallised into NAD, until all particles become crystalline NAD when the temperature exceeded 168 K. The same group also undertook the same experiment with 1:3 HNO_3/H_2O droplets (Bertram and Sloan, 1998b), and found that similar temperatures were required to solidify the droplets to amorphous form and anneal them to produce NAT.

Bertram and Sloan conclude that, under stratospheric conditions, homogeneous nucleation into amorphous form occurs on time-scales of one hour to one day at temperatures of 170-175 K, at stratospheric pressures. These temperatures are too low for these processes to be important in the polar stratosphere for NAD and NAT formation.

2.6.3 Summary

This review has highlighted significant differences in refractive indices of the same forms of solid PSC between data published from different research groups. There is some evidence that this may be caused by different experimental methods, i.e. through interaction with thin-film substrates, or by dependence on formation conditions. Indeed, as concluded by Tisdale *et al.* (1999), it is clear that further laboratory studies, conducted under stratospherically-relevant conditions, are required to accurately characterise the optical signatures of NAD and NAT for remote detection of PSCs in the stratosphere.

2.7 Summary: The case for infrared refractive index measurements of PSC particles

The infrared spectroscopy of PSC is characterised by broad-band absorption features which arise from the chemical speciation of the medium. Laboratory studies have revealed that different types of PSC have distinct absorption characteristics (Toon *et al.*, 1994). Thus, the configuration and relative strengths of the absorption bands have potential use in identification of PSC type and liquid composition from remote sensing in the infrared. Spectral extinction in the infrared originates from a combination of absorption and scattering, which can be modelled in the stratosphere with knowledge of the wavelength-dependant complex refractive index and some assumptions (or *in situ* measurements) of the size distribution of PSCs, to yield PSC volume and, more importantly for analysis of chlorine activation, surface area of PSCs. Knowledge of the radiative properties and stratospheric loading of aerosol is also necessary to model the effect of aerosol on radiative forcing, and effect on climate.

Refractive index data can be calculated from laboratory spectroscopy of PSCs, although both the spectral analysis and experimental procedure are non-trivial. Although a wide range of solid forms of PSC and liquid compositions have been measured, this review has identified the main problems with published refractive index data:

1. No refractive indices have been derived from **direct** spectral measurements of the liquid ternary system over a range of stratospherically-relevant compositions and temperatures in either thin-film or aerosol form;

2. Comprehensive wavenumber-dependent uncertainties have **not** been provided with

any published set of refractive indices;

3. Significant discrepancies exist between refractive index data of the same types of PSC for liquid and solid forms of PSC published by different research groups.

Knowledge of the uncertainty of derived refractive index data is necessary to assess resultant uncertainty in the measurement of PSC properties through remote sensing (i.e. aerosol volume/surface area), and to determine the accuracy of combining binary indices to produce ternary solution refractive indices.

An understanding of the underlying causes of the discrepancies between published refractive indices for the same forms of PSC is necessary to improve future experiments, and to provide consistent stratospherically-relevant refractive index data. Explanations proposed in literature include:

- Differences in solid forms have been attributed to fundamental differences between thin-film and aerosol experiments (i.e. thin-films may interact with the substrate), and dependence on formation conditions/crystallographic orientation of crystals;
- Differences in liquid measurements, primarily the temperature-dependant sulphate features around 1100 cm^{-1}, remain unexplained – Biermann *et al.* (2000) state that the "differences in data sets are reflecting uncorrelated problems of methods used."

In the case of liquid PSCs, if the explanation of Biermann *et al.* (2000) stands (second point above), then determination of refractive indices representative of stratospheric PSCs can only be determined from measurements conducted under stratospheric conditions. Indeed, as identified in this chapter, it is not possible to supercool liquid thin-film solutions, of binary compositions of relevance to the stratosphere, to PSC-temperatures without freezing the film; thus any temperature dependence of absorption features of supercooled PSCs cannot be characterised with thin-films. This implies that, in order to determine indices of relevance to the stratosphere, it is necessary to derive refractive indices from spectra of ternary solution **aerosol** of stratospherically-relevant compositions at temperatures of the polar stratosphere (i.e. significantly supercooled).

In order to ensure that the data produced from this thesis address the main gaps in scientific knowledge by providing a set of refractive indices relevant to stratospheric aerosol, it is important to:

- Measure ternary solution aerosol at cold temperatures and compositions of relevance to the stratosphere;
- Establish the accuracy of derived refractive index data for ternary aerosol;
- Characterise the temperature dependence of absorption features in ternary solution aerosol.

Chapter 3

Fourier Transform Spectrometry

Fourier transform (FT) spectroscopy is utilised for the generation of spectra for this work; as such, an overview of this technique is necessary but is also important for an understanding of instrument effects on spectra. Advantages over classical techniques are fundamental and distinguish this as the foremost technique for modern spectroscopy, particularly in the infrared region. Many published works exist that explain the techniques and applications of FT spectroscopy (Gronholz and Herres, 1985; Beer, 1992; Banwell, 1983; James, 1995; Chamberlain, 1979), and so this discussion will be restricted to the main elements of FT techniques.

3.1 Introduction

With a *conventional* dispersive spectrometer, such as a scanning prism or grating spectrometer, the voltage output through a detector at each discrete frequency point is directly related to radiation at that frequency. Fourier transform spectrometers are more subtle since the detector does not measure the output as a function of frequency. Rather, the interference pattern from re-combined beams of radiation which have travelled different distances (path difference) is determined. The detector signal is thus a measurement of beam interference from which the spectrum can be extracted using the mathematical

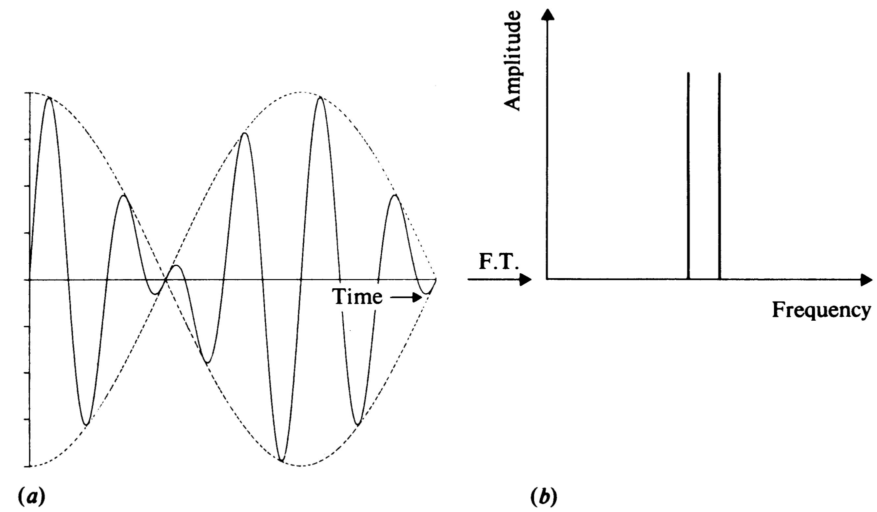

Figure 3.1: The Fourier transform of (a) the summed sine waves in the time domain to (b) their corresponding amplitudes in the frequency domain (Banwell, 1983).

technique of the Fourier transform. The detector measures all spectral frequencies simultaneously at each discrete measured path difference point between the beams to construct the interferogram. Increasing path difference implies increasing orders of interference and therefore greater measurement resolution. Figures 3.1 and 3.2 help explain what is meant by this concept. First, consider a single, infinitely thin spectral line (Dirac delta function) emitting radiation at a precise frequency, ν. In the frequency domain it appears as a single line, but in the time domain is appears as a simple sine wave, oscillating at frequency ν. Next, consider two emission lines at different frequencies, in this case the detector records the summation of the two sine waves in the time domain (figure 3.1). The detector signal also shows an overall oscillation at a frequency which is the difference between the two frequencies - the beat frequency. The combined sine waves can be mathematically resolved into frequency components; knowledge of the time domain signal at four points will give the solution of four simultaneous equations. As progressively more sine waves are added at discrete frequencies this process becomes more complex, with more equations to solve, but the principle remains the same. In reality spectral lines are broadened to some extent, but again can be represented by summations of sine waves about a central frequency, or essentially a *package* of waves which have infinitely small differences in frequency about the central peak. Figure 3.2 shows the time domain beat oscillation signatures of (a) a narrow emission line and (b) a broad line at the same frequencies. It is clear that the broader the line, the faster the signal decays in the time domain. The time domain signal decays since, whilst all the sine waves are in phase initially (the centre burst), as time elapses the package of many different frequencies will become out of phase, so that their product tends to zero with time. The same is true of the variation of the interferogram; radiation combines constructively at zero path difference (ZPD), and decays towards increasing path difference. This illustrates the equivalence of the time and path difference variables. Following this argument to its conclusion, a white source would have a single central peak that has no beats in the time domain, since it would consist of superimposed sine waves over the entire frequency range.

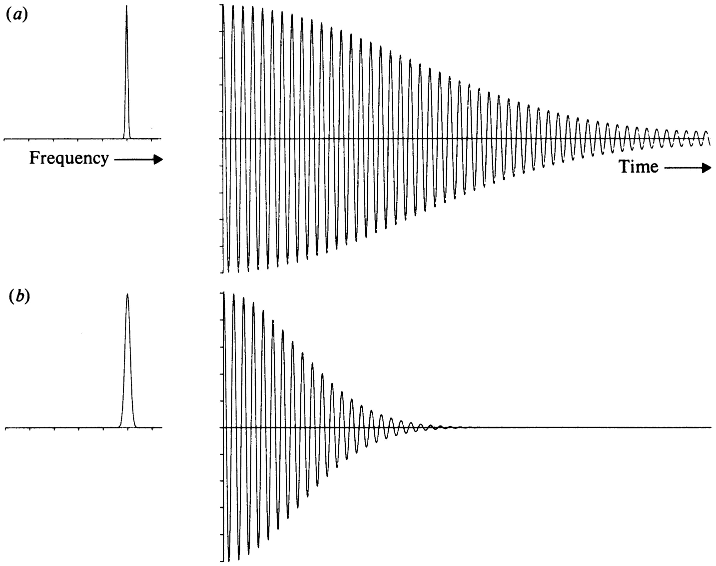

Figure 3.2: Time domain signal as a spectral line varies in position and width (Banwell, 1983).

Converting a white source interferogram signal from time domain to frequency domain using simultaneous equations would be very time consuming, but theoretically possible. However, the process of converting spectral information from the time domain to the frequency domain can be performed in a much more straightforward way by application of the fully reversible Fourier transform.[1] This mathematical transform is basically an integration of the complex interferogram, to produce the spectrum intensity with frequency from the real part of the transform. In the case of a measured interferogram (section 3.2), the time variable is equivalent to mirror travel distance, x, so the conjugate variables $\tilde{\nu}$ and x are used.[2] The interferogram function $I_f(x)$ can be recovered from the frequency spectrum $S(\tilde{\nu})$ by inversion. The inversion has the same form as the initial transform, and in this case since $S(\tilde{\nu})$ is real and $I_f(x)$ is symmetric about the centreburst, then $I_f(x)$ is the FT of $S(\tilde{\nu})$ and vice versa, forming a *Fourier pair*.[3]

A revolutionary development in the history of Fourier spectrometry was the development of the Cooley-Tukey algorithm in the 1960s. Now known as the Fast Fourier Transform (FFT) (Brigham, 1974; Griffiths and de Haseth, 1986) this technique permits calculation of a discrete (summation rather than continuous integral) transform in $2N\log_2 N$ operations rather than $2N^2$, which results in a substantial reduction in computing time for a Fourier transform, making the FT technique viable on all modern computing platforms.[4]

3.2 The interferometer

The optical configurations of the Bruker spectrometers used in this thesis are essentially a variant of the Michelson interferometer, using a collimated source rather than an extended un-collimated source.[5] A basic schematic showing the layout of an idealised Michelson interferometer can be seen in figure 3.3. Radiation is emitted from a stable internal source, through an aperture (annulus) of area A_s, and transformed into a collimated parallel beam

---

[1] Named after Jean Baptiste Fourier who developed the method in the early 1800s.

[2] $S(\tilde{\nu})$ is the Fourier transform of $I_f(x)$.

[3] Symbolically, $S(\tilde{\nu}) \rightleftharpoons I_f(x)$, where only one integral has a negative exponent; the choice is arbitrary but has to be adhered to for consistency in following calculations.

[4] For $N = 10^6$, this yields a speed increase of 50000, transforming days of computing time into seconds.

[5] Strictly, this is a Twyman-Green interferometer.

by a collimating lens (focal length f_c).[6] The beam is then incident on a beamsplitter (potassium bromide, KBr, in this case) which ideally transmits and reflects 50 % of the incident radiation. The reflected half of the radiation is directed to a fixed mirror M_1 and the transmitted half to the scanning mirror M_2, which can be moved to distance x from the beamsplitter. Both reflected beams recombine to form a single beam at the beamsplitter where 50 % is transmitted towards the output port and the reflected half is sent back to the input port as an inevitable consequence of the design. The beams are coherent, and therefore interfere upon recombination, having an optical path difference (OPD) of $2x$. After passing through the aerosol cell, the collimated beam is focused on a detector, where the interferogram, which is the amplitude $I_f(x)$ of the recombined beams as a function of the moving mirror displacement, is measured. The interferogram can the be manipulated and the FFT undertaken on a lab computer in order to obtain a spectrum in the frequency domain.

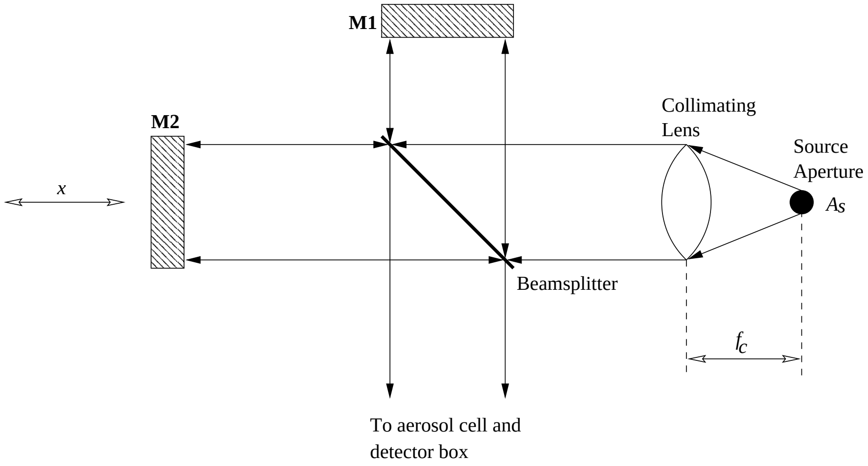

Figure 3.3: An idealised representation of the interferometer component of an FT spectrometer.

---

[6] For infrared measurements, a 150 W silicon carbide globar source is used.

3.3 Advantages of Fourier transform infrared spectroscopy

3.3.1 The Connes advantage

All motor drives are subject to slight inaccuracies, and the drive on the moving mirror is no exception. To circumvent this problem, Fourier transform spectrometer (FTS) instruments employ the interference pattern of a monochromatic laser source for calibration; early instruments used a CO_2 laser, most now use a red He-Ne laser source, as with the Bruker instruments.[7] The recombined waves from such a monochromatic laser source have a regular interference pattern, interfering constructively where OPD is an integer multiple of λ_c, the laser wavelength.[8] Now, the relation between the interferogram and x is $I_f(x) = S(\tilde{\nu}_c) \times \cos(2\pi\tilde{\nu}_c x)$, where wavenumber $\tilde{\nu} = 1/\lambda$ [cm^{-1}], and $S(\tilde{\nu}_c)$ is the strength of the monochromatic laser line at $\tilde{\nu}_c$. This permits very precise tracking of the moving mirror. For spectra presented here, data acquisition was triggered when the laser signal equals zero, i.e. twice per laser wavelength (undersampling; sample spacing divisor of 1).[9,10] Thus, OPD can be simply deduced from the laser wavelength by counting the number of fringes from ZPD.[11] This gives FTS instruments an integrated wavenumber calibration of better than 0.01 cm^{-1} (Gronholz and Herres, 1985), depending on the accuracy to which the laser wavelength is calibrated.

3.3.2 The Jacquinot or throughput advantage

In classical spectrometers radiation is commonly brought to focus on a slit, which is then imaged on a detector. The narrower the slit, the better the resolving power of the instrument since a smaller spread of frequencies falls on the detector at any point in time. However, at high resolution the radiation passing through the necessarily narrow slit will be limited.

---

[7] For the Bruker IFS 66v/S the He-Ne laser (not temperature stabilised) $\tilde{\nu}_c = 15798.08223\ \mathrm{cm}^{-1}$.

[8] $2 \times x = n \times \lambda_c$ $(n = 0, 1, 2, ...)$.

[9] Higher undersampling spacing divisors can be used to reduce interferogram sampling points when a lower wavenumber ceiling is required.

[10] For spectral measurements at wavenumbers above the laser wavenumber, electronic sample triggering (oversampling; sample spacing divisor < 1) between fringes is required.

[11] $\mathrm{OPD} = 2x = p_{igm}/2 \times \lambda_c$; p_{igm} = no of interferogram points, λ_c = laser wavelength.

In the case of infrared radiation, where sources are not particularly intense and thermal noise (section 3.5.5) can become a problem, the use of high-gain amplifiers to enhance these small signals can lead to *noisy* signals; a compromise between resolution and noise is necessary, limiting the useful resolution of classical instruments in the infrared. In Fourier spectrometry, parallel beams are used throughout the system, until the light is brought to focus on a detector; a slit is not required and a large fraction of the source energy is imaged on the detector. Thus, the signal does not require as much amplification, drastically improving the signal to noise ratio (SNR) and overcoming the resolution-limiting noise constraint of classical instruments; as a result, FT instruments are far superior to classical spectrometers for infrared spectroscopy.[12]

3.3.3 The Fellgett or multiplex advantage

One of the great benefits of Fourier transform spectrometers is the fact that the measured frequency range is simultaneously recorded. In classical instruments each frequency is examined consecutively until a required spectral window is *swept out*. This can be a time-consuming operation, especially when the required information is only a few spectral lines. Conversely, the measurement time of a FTS is the time taken to move the movable mirror over a set distance for a required resolution. Since the mirror can be moved quickly, it is possible to record consecutive medium resolution spectra at a sub-second rate; the individual interferograms can be stored enabling Fourier transformation at a later time as required. Combined with the throughput advantage, it becomes clear that FT spectrometers can be used to attain much higher resolutions, especially in the infrared, than classical spectrometers.

Other advantages include the fact that the resolution of FT spectrometers is nominally constant over the entire spectral range (ignoring the effects of beam divergence, section 3.4.3), whereas resolving power of a prism or grating instrument depends on the

---

[12] This is why FT spectrometers were initially developed for the low-energy far infrared region.

angle between the output frequency resolved beam and the input radiation beam, which varies with frequency and thus resolution becomes poor at the edges of the measured spectral window. Finally, modern computing power enables rapid co-addition (averaging to reduce random noise), manipulation and transformation of interferograms (section 3.5).

3.4 Practical limitations of an interferometer

3.4.1 Resolution

Consider the interferogram of two spectral lines - the closer the frequency of the lines, the further apart their beat oscillation will be in the interferogram; (a) and (b) in figure 3.2 illustrate this. For lines separated by $\delta\tilde{\nu}$ in the spectrum, the corresponding beat oscillation in the interferogram will occur at $f/\delta\tilde{\nu}$. Hence, the limiting resolution of the FT instrument is inversely proportional to the maximum optical path difference (MPD) of the scanning mirror, such that $resolution = f/\text{MPD}$, where f depends on the definition of resolution and other instrument effects (such as instrument line shape). Under the Rayleigh criterion, applicable to dispersive spectrometers where the ILS is a sinc^2 function, $f = 1$, or if the full width at half maximum (FWHM) is used then $f = 0.603$; the Bruker definition of resolution is 0.9/MPD. The fundamental resolution limit is imposed by the overlap of two lines, i.e. the ability to distinguish two lines.

3.4.2 Instrument line shape (ILS)

The finite MPD of an interferometer modifies the shape of measured spectral features. Effectively, truncation of the interferogram at MPD is the same as multiplying an infinite length (ideal) interferogram with a *boxcar* function $B(x)$ where $B(x) = 1$ for $0 \leq x \leq \text{MPD}$, otherwise $B(x) = 0$ (for a single sided interferogram). Transformation of this function into the frequency domain yields:

$$J(\tilde{\nu}) = \sin(2\pi\tilde{\nu}\ \text{MPD})/(2\pi\tilde{\nu}\ \text{MPD}) = \text{sinc}(2\pi\tilde{\nu}\ \text{MPD}) \tag{3.1}$$

This normalised *sinc* function (see figure 5.4) is convolved with the underlying ideal spectrum in the frequency domain; it is multiplied by the ideal spectrum at every frequency, effectively as a running average. Symbolically, the result of the *convolution theorem* is:

$$I_f(x) \times B(x) \rightleftharpoons \mathbf{FT} \rightleftharpoons S(\tilde{\nu}) \otimes J(\tilde{\nu}) \tag{3.2}$$

Convolution with $J(\tilde{\nu})$ conserves the total power of the spectrum since the sinc function is normalised (has unit area) but changes the appearance of spectral lines; the first sidelobe minima have amplitudes 22 % of the central maximum. The ILS therefore smears the output, resulting in loss of resolution, and the sidelobes can create some spectral artefacts, for example transmission in excess of 100% in the sidelobes of a line. The process of apodisation has been developed to smooth out spectra to produce something more like traditional spectra; the interferogram is multiplied about ZPD, to maintain symmetry about line centres, by an empirical function designed to cancel out the sidelobes of the ILS.[13] Various apodisation functions exist, but all result in further increase in width of the effective instrument lineshape, and hence loss of information. It is desirable to achieve a compromise between sidelobe reduction and loss of resolution for different situations; Norton and Beer (1976) tested about 1150 functions to arrive at three preferred formulations. The apodisation functions used for this analysis are detailed in section 3.5.2.

3.4.3 Beam divergence

An idealised interferometer has a point light source. In reality, all sources have a finite size and hence beams in the spectrometer will not be exactly parallel. The finite aperture, A_s, through which the beam passes, introduces oblique rays in the parallel beam resulting in an asymmetric line shape, the effects of which can be deconstructed into two components: (1) symmetric broadening of the lineshape, and (2) wavenumber dependent shift of central wavenumber. A practical limit is imposed on the maximum measured wavenumber as a

---

[13] The term apodise originates from Greek, literally meaning *feet removal.*

function of the aperture diameter since:

$$\tilde{\nu}_{\mathrm{max}} = \frac{8f^2}{\mathrm{OPD}\ .\ d^2} \tag{3.3}$$

where d = aperture diameter, and f = collimator focal length. For example, to achieve a higher resolution, a longer MPD is required and thus the beam diverges further due to the longer path distance; in order to maintain constructive interference along the whole path progressively smaller apertures (with smaller angles of beam divergence) are required. This imposes fundamental limits on maximum resolution of FT spectrometers; at higher resolutions signal decreases with reducing aperture until decreasing SNR prevents useful measurement. However, the Bruker IFS 66v/S is limited by MPD, not SNR, over the spectral region of measurements presented in this thesis.

The wavenumber shift that results from beam divergence was first determined by J. Connes, and its derivation is described in detail elsewhere (Chamberlain, 1979). The wavenumber scale is shifted by $\delta\tilde{\nu}$ as a function of the apparent (unshifted) wavenumber of the measurement, $\tilde{\nu}'$, such that:

$$\delta\tilde{\nu} = -\frac{A_s}{4\pi f_c^2}\tilde{\nu}' \tag{3.4}$$

where A_s is the source aperture area, and f_c is the collimator focal length. Results for the gas retrieval in section 5.2.3 show that this shift is seen clearly in our data, and is not negligible since the compact nature of the Bruker IFS 66v/S requires a short collimator focal length, resulting in a larger shift (especially at low resolution where a larger aperture can be used). For example, $\delta\tilde{\nu} \approx -0.5$ cm^{-1} at 7000 cm^{-1} with the low resolution configuration used for aerosol measurements

3.5 Data processing

The software supplied by Bruker performs the necessary data manipulation to produce the required spectra.[14] As can be seen from the example measured interferogram in figure 3.4, the raw interferogram of a broad polychromatic source has a strong peak around ZPD $x = 0$, but much weaker *wings* which contain most of the spectral information. Hence an analogue-to-digital converter (ADC) of high dynamic range is required to maximise information content. In the Bruker setup a 16-bit ADC converter is used, with the first bit indicating sign and the remaining 15 bits yielding decimal ADC counts up to $2^{15} = 32768$. Higher dynamic range, equivalent to 19-bit ADC, is achieved using a gain of 8 (2^3) outside the interferogram centerburst region. A variable gain is also applied to the entire interferogram in order to maximise signal for each detector and (field stop) aperture setting. The spectrometer and detector-box rotating mirror are fully controlled from a dedicated lab computer.

The proprietary software undertakes a number of steps for analysis of data. Raw single-sided interferograms are acquired and stored initially stored in memory.[15] After each scan, interferograms are co-added (averaged); the user initially specifies the number of scans required along with other measurement parameters. Co-adding interferograms prior to transformation effectively reduces random noise according to standard statistical sampling theory, i.e. $\mathrm{SNR} \propto \sqrt{\text{no. scans}}$. The procedure is then to apply a zero filling apodisation function, FFT the interferogram, apply phase correction, save the data and then re-commence measurement after a set period or on user command. A description of each main stage in data analysis follows.

---

[14] The Windows NT version of OPUS on the Bruker IFS 66v/S.

[15] Both sides about ZPD are not required since the interferogram should be symmetrical.

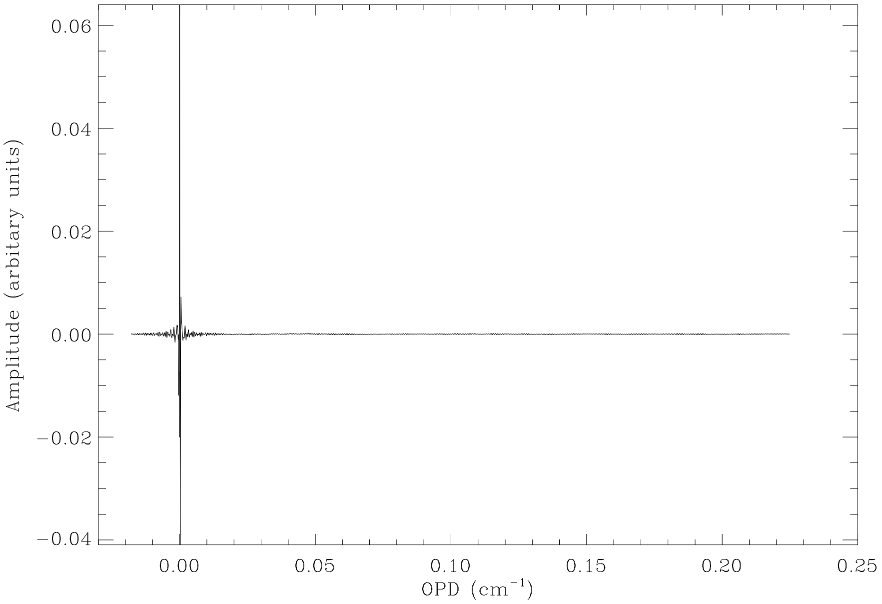

Figure 3.4: A low resolution DLaTGS interferogram. Although this is a single-sided measurement, 516 data points are measured to the left of ZPD for phase correction purposes.

3.5.1 Zero filling

If the centre of a spectral line lies between two sample points in the wavenumber grid problems can occur; in the worst case, when a component lies halfway between sample points, false signal reduction of up to 36 % can occur (Gronholz and Herres, 1985). Transmission peaks in affected spectra appear badly clipped. This is known as the *picket-fence* effect, but it is not so pronounced in cases where spectral lines are of similar or greater width than the wavenumber grid. The solution to this problem is to include zeros at the end of the interferogram, increasing the effective MPD but not introducing any extra information. Zero filling the spectrum effectively interpolates the spectrum to a finer wavenumber grid. In the worst cases of picket-fencing, a zero filling factor (ZFF) of up to 8 is recommended, and the minimum recommendation is a ZFF of 2 (Gronholz and Herres, 1985). For all spectra presented here, extra zero filling of at least ZFF=2 was applied, interpolating the spectra to a grid with half or less the wavenumber spacing of the measurement resolution;

to fulfil the FFT criteria for the number of interferogram points, additional zero filling was automatically added (section 3.5.2). Since H_2O gas lines are typically more narrow than the highest measurement resolution used for this work, zero filling is advantageous in this case, but is not necessary for broad aerosol or HNO_3 gas spectral measurements. Moreover, the gas retrieval forward model accounts for gridding effects to a certain extent in the gas phase spectrum. The negative aspect of zero filling is the introduction of further cross-binning of random errors, or cross-talk, so that random error at each wavenumber point is no longer independent. The ILS and apodisation also contribute to cross-talk, making it less straightforward to deal with error in the retrieval of gas phase components (section 5.2.4).

3.5.2 Apodisation function

The need for deliberate application of apodisation functions has already been explained in section 3.4.2. The functions used by OPUS on data presented here were Blackmann-Harris 3-term for all the low resolution DLaTGS measurements, and Norton-Beer Strong for high resolution measurements.

3.5.3 Fast Fourier transform (FFT)

A price of using the Cooley-Tukey algorithm is that the number of points in the interferogram has to be a power of 2. Resolution is pre-set by the user, fixing MPD and sample points since there are two interferogram sampling points per laser wavelength (section 3.3.1). In order to fulfil the FFT points criteria, zero filling is automatically applied to increase the number of interferogram points, prior to application of specified zero padding (ZFF=2).

3.5.4 Phase correction

The purpose of phase correction is to determine the required amplitude spectrum $S(\tilde{\nu})$ from the complex output $C(\tilde{\nu})$ of the FT process; artefacts arising from electronic, sampling and instrument effects result in this complex spectrum rather than the amplitude spectrum.

The main causes of phase error are ZPD not coinciding with a sample point, the one-sided interferogram (section 3.5), and intrinsic asymmetry of the interferogram due to non-ideal instrument factors or electronic filtering.[16] *Mertz* phase correction was used in this study (Born and Wolf, 1970).

3.5.5 Noise

Total signal to noise ratio (SNR) can be expressed as (Beer, 1992):

$$\mathrm{SNR} = \frac{\text{number of signal photoelectrons}}{\sqrt{\text{total number of carriers from all sources}}} \tag{3.5}$$

A comprehensive model would take account of temperature, transmittance, emittance, reflectance and scattering from all surfaces from source to detector. However, here we need consider two main sources - noise from photon statistics (Poisson statistics), and system noise which incorporates noise including that from the setup, electronics and thermal emission. Photon noise can considered to have a standard deviation of $\sqrt{N}$ where N is the number of photons, and is therefore dependent on optical throughput and spectral filter width. The relative random error falls with increasing radiation intensity (thus, high absorption leads to high standard error). Modern detectors are usually limited by photon statistics, as other sources can be minimised, but in the infrared regime significant system noise is generated by thermal emission from the detector. Thermal noise can be considered as wavenumber dependent, decreasing towards the visible region and having an absolute (rather than relative) magnitude. Use of liquid nitrogen coolant with the MCT detector significantly reduces thermal noise (photon noise then dominates).

As previously stated (section 3.5), SNR increases proportionally with the square root of integration time (number of averaged scans) with random errors. Unfortunately, the OPUS software does not return an estimate of variance at each wavenumber from the co-added scans used to produce spectra, since each scan is added into a running interferogram

---

[16] A few hundred points are measured the other side of ZPD for phase correction; physically this represents the offset of the fixed mirror from the beamsplitter.

average and not stored for later analysis. Although later test measurements yield an estimate of thermal noise, experimental conditions cannot be exactly reproduced, and photon noise remains unknown, since the exact absorbance spectrum of aerosol is highly variable so noise measurements can only be performed during experiments. Two main techniques have been developed to overcome this limitation and estimate noise from spectra; local comparison with the forward model for the gas line retrievals (section 5.2.4) and a global next-neighbour analysis with low resolution spectra aerosol retrievals (section 5.4). Moreover, other sources of non-random error, such as interference channelling from the optics and cross-talk of random errors create further problems (section 5.2.4).

Chapter 4

Experimental Configuration

The main objective of laboratory experiments conducted for this thesis was to generate mimic PSC aerosol particles, principally supercooled ternary solution (STS) aerosol, for determination of accurate refractive index data. The experiments were designed in order to address the key gaps in scientific knowledge identified at the end of chapter 2. Thus, the aims of the experiments were:

- To measure ternary solution aerosol at cold temperatures and compositions of relevance to the polar stratosphere;
- To complement the range of conditions covered by Biermann *et al.* (2000) in terms of composition and temperature, and to investigate use of the Biermann *et al.* thin-film data for calculation of ternary solution indices of relevance to the stratosphere;
- To fully characterise the experimental system to determine PSC development and assess the similarity of cell conditions to those in the stratosphere;
- To characterise the composition of liquid aerosol in the cell;
- To provide comprehensive uncertainties for all aspects of the experiment system.

Experiments were conducted at the Molecular Spectroscopy Facility (MSF) at the Rutherford Appleton Laboratory (RAL), where expertise and equipment had been developed through prior experiments to measure spectra of sulphate aerosol with a small cell, detailed by Heathfield *et al.* (1999). Experiments conducted for this thesis evolved from a similar methodology and setup to that described by Heathfield *et al.*, to a considerably more complex system involving use of a larger cell.

The experiments described in this thesis are segregated into two distinct sub-sets – the **main experiments** and the **settling experiments**. The main experiments provide the bulk of the experimental data used for the production of refractive indices in this work. In the earlier settling experiments, in addition to measurements of stable aerosol, observations of aerosol falling under gravity were undertaken in order to assess possible variation in aerosol size distribution. The earlier settling experiments did not have the advanced automated temperature monitoring system or the facility to measure high resolution spectra with a separate detector (section 4.1.5) of the main experiments.

This chapter includes a description of the equipment used, followed by an overview of the experiment methodology, including the main experimental principles. System characterisation, including temperature calibration and aerosol generation, is described in the next chapter.

4.1 Description of equipment

In this section the laboratory equipment used in the aerosol experiments is detailed, along with the configuration of this equipment for the experiments presented in this thesis.

4.1.1 The large aerosol cell

For the experiments of relevance to this thesis, the larger of two aerosol cells available at the RAL was employed. The large cell, described in detail by McPheat *et al.* (2001), provided several key advantages over the small cell as used by Heathfield *et al.* (1999):

- An aerosol residence time of order 30 minutes, rather than 10-30 seconds in the small cell under typical experimental conditions;

- The capacity to conduct aerosol settling experiments due to the alignment and much greater vertical height of the large cell;
- Less interaction with the cell walls due to an internal volume to wall area ratio of approximately eight times higher than the small cell.

The large cell is optically interfaced to the Fourier transform spectrometer (FTS) via evacuated stainless-steel chambers and tubing. The stainless-steel inner vessel of the large cell is 1.0 m in height and 0.3 m in diameter, with an inner volume of approximately 75 litres. As illustrated in figure 4.1, the cell was mounted vertically with gas/aerosol injected at the top and evacuated from the base plate. Four view-ports, each at 90° to each other, are located 0.27 m above the base plate. Calculations show that under typical conditions of 200 Torr cell pressure and 190 K temperature, the gas residence time with a total flow rate of 1000 standard cubic centimetres per minute (sccm) is approximately 24 minutes.[1,2] This calculation neglects factors such as the effect of gravity on particles and turbulence, and assumes particles travel with no resistance to the gas flow. Anthony *et al.* (1997) noted that, under similar experimental conditions to those here, complete mixing of ternary solution aerosol should be achieved within 5 minutes; i.e. if two separate binary populations of H_2SO_4 and HNO_3 aerosol were initially generated, these would form a homogeneous ternary solution aerosol within this time period through vapour transfer. Thus, the longer residence time of the large cell indicates that the aerosol measured by the FTS at the base of the cell should be fully mixed and in thermodynamic equilibrium.

A major advantage of the large cell design is that the height and vertical orientation of the cell allows for settling experiments. From Stoke's law, it can be calculated that particles of radius 1.2 μm (a size representative of the particles produced in these experiments) take

---

[1] Using $t = \frac{P \times 1.33 \times 273.15 \times v}{1015 \times T \times f}$ where P = cell pressure[Torr], v = cell volume[cm^3], T = temperature[K], f = flow rate[sccm], t = residence time[min].

[2] 1 Torr $\equiv$ 1.333 mb.

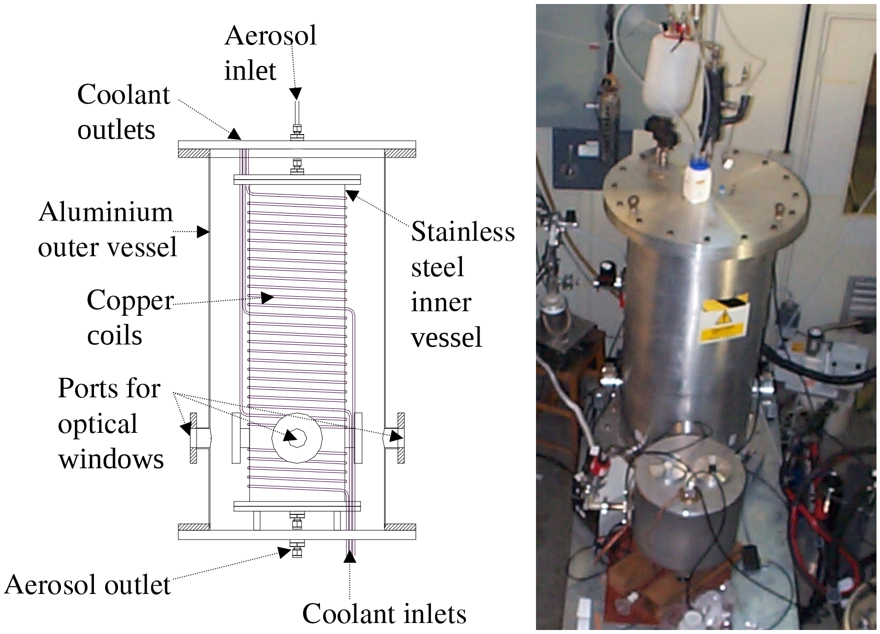

Figure 4.1: Schematic of the large aerosol cell (left; R.G. Williams, RAL), and (right) image of the large cell and detector box (in the foreground).

approximately 58 minutes to gravitationally settle 1 m at 190 K, neglecting turbulence.[3,4] Thus, if flow into and out of the cell is shut down, and the cell contents are isolated, particles would settle out at different rates, with smaller particles taking longer to descend. Two competing effects should then act to vary particle size compared with the flow situation. The larger particles in the distribution settle out faster, leaving smaller particles behind and a reduction of mean radius with time. At sufficiently high particle densities, coalescence of particles to form larger particles may become important with time. Analysis of the results of settling experiments, with retrieved aerosol size distribution, is described with the characterisation of the experimental system in section 5.1.2.

---

[3] Using $v = \frac{\frac{4}{3}g.a^2(\rho_{sol}-\rho_{gas})}{6\times\left(\frac{\eta}{\sqrt{T_{room}}}\right)\times\sqrt{T}}$ where v = particle velocity[m/s], g = acceleration due to gravity[m/s^2], a = particle radius[m], ρ_{sol} = solution density[kg/m^3] for ~ 40 wt% H_2SO_4 solution, ρ_{gas} = gas density[negligible], η = gas viscosity[N s/m^2], T_{room} = room temperature[K], T = cell temperature[K].

[4] Gas viscosity is independent of gas pressure except at very low and very high pressure (Kaye and Laby, 1973).

4.1.2 The aerosol generation system

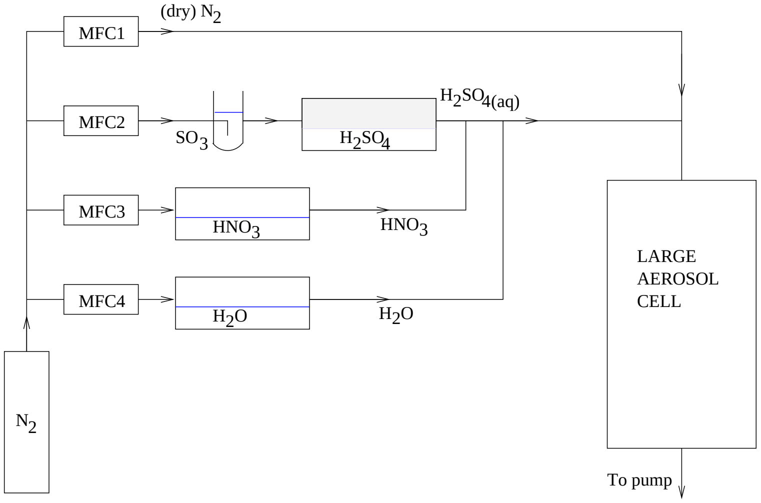

Figure 4.2: Schematic of the aerosol generation system.

In order to create ternary solution aerosol in the large cell, the aerosol generation system was designed operate through a process analogous to that which takes place in the stratosphere. As in the stratosphere, the initial aerosol formed in the experimental system is sulphuric acid, produced in this case at room temperature. Additional HNO_3 and H_2O vapour are added at controlled rates. Upon entry into the cell the sulphate aerosol composition should change with cooling due to condensation of volatiles according to thermodynamic theory, with the condensation of HNO_3 vapour resulting in the formation of ternary solution aerosol. It follows from simple thermodynamic theory that, in addition to internal cell temperature, variation of relative flow rates should control the composition of the aerosol.

Figure 4.2 shows a basic schematic of the aerosol generation system as used in the main experiments, and a more detailed representation of the (slightly different) system employed for the settling experiments can be seen in figure 4.3. Up to four separate mass flow controllers (MFCs; MKS 1179A with Kalrez seals) were used to channel high-purity inert N_2 buffer gas through the aerosol generation system with maximum flow rates of

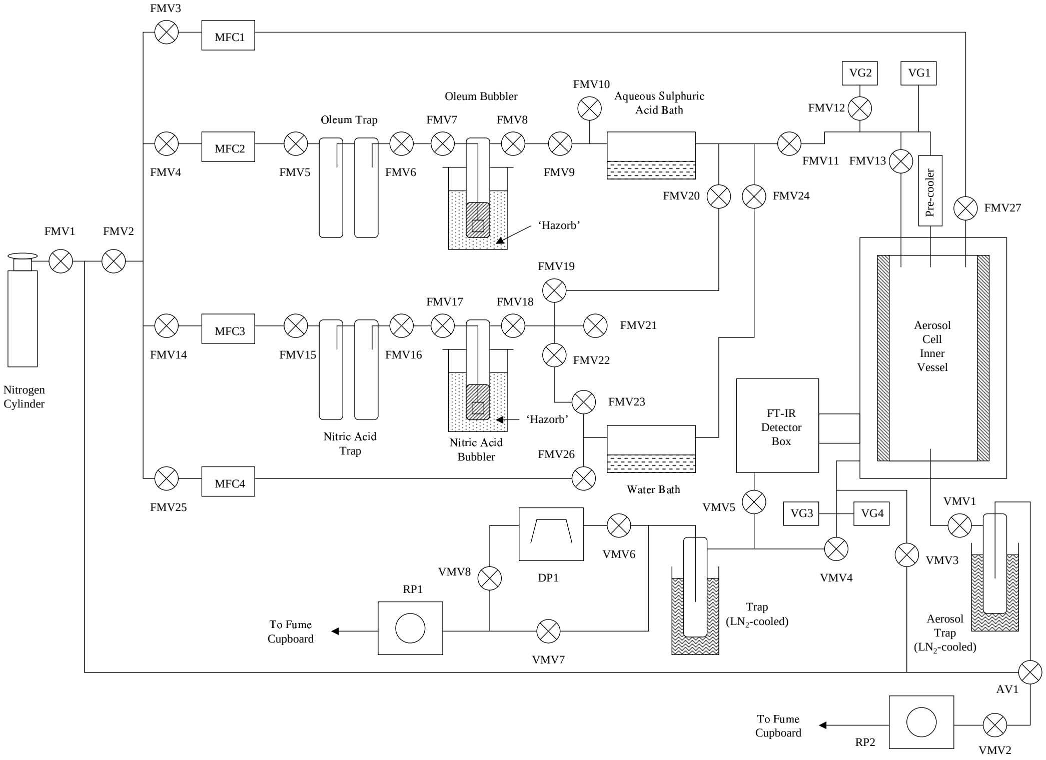

Figure 4.3: Detailed schematic of experimental setup. Valves on the generation side are marked FMV, on the vacuum pump side are marked VMV, the rotary pumps are marked RP, the diffusion pump DP, and the pressure gauges VG (D.A. Newnham, RAL).

2000 sccm (dry N_2 only) and 500 sccm. The separate streams are combined prior to entry into the cell through the top plate, and evacuated through the base plate, maintaining a constant system pressure and flow through the cell (thus a pressure gradient exists across the system). For all the experiments detailed in this thesis, the internal cell pressure was maintained at 200 Torr in a constant flow situation. A 1000 mb full-scale pressure gauge (MKS 750 Baratron) provided pressure readings which were used by an automatic system to control the variable aperture valve (MKS 248; AV1 in figure 4.3) and ballast the evacuation rotary pump with N_2, hence maintaining a constant cell pressure. This system successfully controlled the pressure to within ± 5 Torr of the set-point.

Acid resistant materials were used throughout the aerosol generation system. The bubblers, baths and traps were made of borosilicate glass and were connected in the system by flexible perfluoroalkoxy (PFA) tubing with PFA vacuum seals and polytetrafluoroethylene (PTFE) valves. This provided a gas-tight system that could be used at low pressure without significant leaks of air into the system. Only the PTFE valves between the oleum bubbler and the sulphuric bath showed signs (blackening) of significant chemical attack; this was due to the highly reactive nature of SO_3 gas. These components were periodically replaced to ensure pressure integrity of the system. Double-tubed glass traps were situated (figure 4.3) between the MFCs and the bubblers to contain acid and protect the MFCs from acid in the event of any excess pressure build up due to a blockage (by ice, for example) in the aerosol generation system. Other precautionary safety measures included placing the bubblers in vessels containing absorbent materials and permanently mounting the gas/aerosol generation equipment behind Perspex screens.

The generation and modification of aerosol particles is described in the remainder of this section.

Generation of sulphate aerosol. Aqueous sulphuric acid aerosol was generated at room temperature prior to entering the cell. Nitrogen buffer gas was initially directed through an oleum (fuming sulphuric acid, 27% to 33% excess SO_3, Fluka) bubbler.[5] This liberates a small quantity of SO_3 gas into the flow. The SO_3/N_2 flow is directed above the surface of the sulphuric acid bath, causing rapid homogeneous nucleation of sulphuric acid aerosol.[6] The sulphate aerosol produced is clearly visible to the naked eye above the sulphuric acid bath, as can be seen in figure 4.4.

The process of sulphuric acid aerosol generation is not well characterised (Lovejoy and Hanson, 1995), but probably involves gas phase reaction of SO_3 with H_2O to produce gas phase H_2SO_4, which in the presence of H_2O nucleates aqueous H_2SO_4 and grows further by condensation of H_2SO_4, SO_3, and H_2O from the gas phase. Other proposed mechanisms

---

[5] To ensure that removal of SO_3 from the oleum did not impact the experiments, a fresh sample of a relatively large quantity (approximately 50 cm^3) of oleum was used for each experiment period.

[6] Previous experiments at the MSF had determined that this was indeed homogeneous nucleation by injecting varying amounts of condensation nuclei into the stream.

Figure 4.4: Aerosol forming as the SO_3/N_2 stream flows across the surface of the sulphuric acid (conditioning) bath. The filling port can also be seen at the top of the glass vessel.

involve clusters of SO_3/H_2O, but are not well understood. What is clear is that the composition of the sulphate aerosol equilibrates to the composition of sulphuric acid in the bath since the partial pressure of water vapour is controlled by the sulphuric acid bath.

In the case where only the H_2SO_4 aerosol/N_2 flow is directed into the cell at room temperature, one would expect the measured binary sulphuric acid aerosol to be of the same composition as the H_2SO_4 solution in the bath; this was confirmed from test experiments. Upon cooling of the large cell, the H_2SO_4 aerosol can be expected to become more dilute once in the cell as H_2O vapour condenses on the aerosol in accordance with thermodynamic theory. The dilution of H_2SO_4 aerosol should be limited by the availability of H_2O vapour. This was indeed identified as the case by Heathfield *et al.* (1999), who used the same method for H_2SO_4 aerosol generation. The composition of H_2SO_4 aerosol in the large cell is thus a function of cell temperature and the total quantity of additional volatiles directed into the cell.

Analysis of spectra recorded for this work showed that the aerosol in the large cell exhibited a range of radius values from about 0.3 to 1.0 μm, and a monomodal log-normal spread (σ) of between 1.0 and 1.5 (1.0 is monodispersed, i.e. all particles in the distribution are of the same size). This range of distributions is caused by the various experimental conditions employed; the radius in particular is a function of the condensation of volatiles on aerosol in the cell.

Control of ternary aerosol composition. The sulphate aerosol was combined with other gas streams to dry out the cell contents (by adding pure N_2), or introduce additional H_2O or/and HNO_3 vapour. When sulphate aerosol is cooled upon entering the cell, volatiles condense on the aerosol according to thermodynamics in a process analogous to the formation of PSCs in the stratosphere. However, the condensation process is ultimately limited by the availability of volatiles in the cell. Therefore, the composition of aerosol is controlled by flow rates of vapours and aerosol in addition to temperature. Additional H_2O vapour is provided by entraining saturation vapour pressure (SVP) of H_2O into the N_2 gas flow above a pure water bath (BDH AnalaR grade water). Similarly, HNO_3 vapour (and H_2O vapour) is entrained at SVP from a 69 wt% HNO_3 solution (BDH AnalaR). A nitrate bubbler was used in the settling experiments, but in the main experiments this was exchanged with a bath to prevent a separate aerosol of binary nitric acid being generated from the bubbler – no perceptible change of aerosol spectra were observed at the same flow rates, indicating that the only place where aerosol is generated was above the H_2SO_4 bath in all experiments.

Control of number density. The number density of aerosol in the large cell is directly related to the number density above the H_2SO_4 bath, assuming the only place that aerosol is nucleated in the system is above this bath. The flow rate above the H_2SO_4 bath was maintained at 400 sccm in all experiments in order to ensure that the number density above this bath remained an experimental constant. The number density in the cell is directly related to the flow rate above the H_2SO_4 bath as a ratio against the *total* flow rate into the cell; aerosol is simply dispersed into the total flow, and aerosol number density inside the cell can only remain the same as above bath or fall when mixed with other gas streams (in a constant flow situation). From measurements of number density of aerosol in the cell, a mean number density of 1.4^6 cm^{-3} was deduced for the aerosol above the H_2SO_4 bath in these experiments. This figure provides a lower estimate since it includes loss of aerosol particles in the system, including loss to tubing walls prior to entry into the cell. Analysis has revealed that coagulation and loss to the cell walls appears minimal, but aerosol loss in the external tubing is difficult to assess since the only indication of the number density above the bath in this analysis is from measurement of aerosol in the large cell (after loss due to tubing has taken place). Fortunately, this is not of major concern to these experiments since number density was independently determined for each measurement, and the loss to tubing would be expected to remain relatively constant (due the constant flow rate above the H_2SO_4 bath).

Overview of alternative methods for aerosol generation. Similar techniques have been used by other groups for aerosol generation. Lovejoy and Hanson (1995) entrained gaseous SO_3 from solid fuming SO_3 into a buffer gas prior to homogeneous nucleation of H_2SO_4 aerosol in the presence of a sulphuric acid solution. Niedziela *et al.* (1999) used a similar method for H_2SO_4 aerosol generation, heating a solution of H_2SO_4 to provide high vapour pressures – upon passing into a cooled cell, supersaturation led to homogeneous nucleation and condensation. These methods represent variations on the theme of homogeneous nucleation from supersaturation (or *effective* supersaturation in cases where SO_3 gas is used). Mechanical generation of aerosol particles from bulk liquids has been used in fewer cases. A syringe pump with a constant atomizer (TSI) was used by Anthony *et al.* (1997, 1995) to generate 96 wt% polydispersed sulphate aerosol at room temperature. This stream was then combined with gaseous H_2O and HNO_3 to dilute the aerosol and produce ternary solutions if desired. Norman *et al.* (1999) generated larger HNO_3 aerosol particles by flowing buffer gas through a glass frit containing room temperature HNO_3 solution; such techniques of mechanical dispersion are described in detail elsewhere (Ulevicius *et al.*, 1997). Mechanical aerosol generation often suffers from problems with the high viscosity of concentrated sulphuric acid, leading to a very broad size distribution that is difficult to model.

Summary. The method of aerosol generation for the experiments of this thesis is unique; ternary solution aerosol is produced through condensation on sulphate aerosol which has been produced by homogeneous nucleation. The only other group to have produced ternary solution aerosol, Anthony *et al.* (1997), used a mechanical technique to generate sulphate aerosol, adding H_2O and HNO_3 vapour and cooling the aerosol to produce ternary aerosol as with experiments for this thesis. However, the main problem with the approach of Anthony *et al.* (1997) is that the wide size distribution of aerosol particles produced through mechanical particle generation is difficult to characterise, impeding determination of refractive indices. Indeed, Anthony *et al.* (1997) did not produce refractive index data from extinction spectra of ternary solution aerosol. The main disadvantage of the aerosol generation system of this thesis is the inability to produce markedly different size aerosol distributions of the same composition aerosol. Settling experiments did not result in a significant change in aerosol particle radius (section 5.1.2). This problem was overcome in this work by application of the CDHO band model allowing the retrieval of refractive indices from single aerosol extinction spectra.

4.1.3 The cooling system

The cooling system is represented in figure 4.5. The large cell is double jacketed with copper coils containing the ethanol coolant encircling, and soldered to, the inner vessel. Ethanol was pumped around the circuit and through two stainless steel coils immersed in Dewars containing liquid nitrogen. The liquid nitrogen level in the Dewars was controlled by an automated system (Eurotherm 818P) which used the temperature reading from a single temperature sensor (platinum resistance thermometer – PRT; Fisher Rosemount, PT-100 grade B) attached to the coils as a set-point for control of a cryogenic solenoid valve to keep the temperature of the ethanol constant. Once cooling to the set point temperature was complete, this system provided stable temperature control to within 1 K (of set-point temperature) for most experiments; the temperature monitoring system and calibration is described in detail later in section 5.3. The thermal inertia of the system meant it took about three hours to cool the cell from room temperature to a stable temperature of 200 K. External cooling pipework was lagged to minimise thermal resistance. It was found that the cell interior could be cooled to temperatures around 190 K before the pure ethanol coolant froze. Although the freezing temperature of pure ethanol is 159 K, the higher cooling limit is imposed by the increasing viscosity of ethanol with cooling. As the ethanol cools below about 190 K the pump is unable to circulate the increasingly viscous ethanol through the liquid N_2 cooled section fast enough to prevent freezing and resultant blockage in the cooling system.

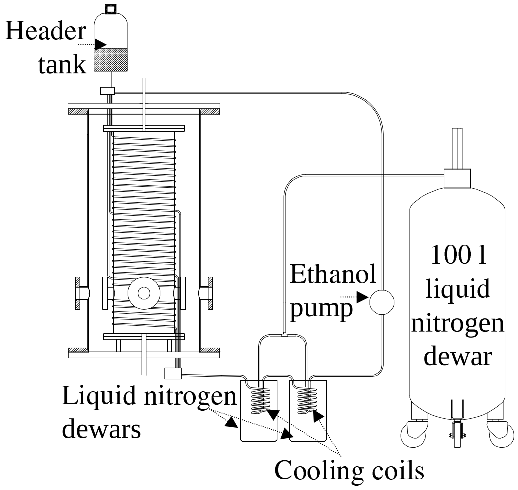

Figure 4.5: Large cell cooling system (R.G. Williams, RAL).

4.1.4 The Fourier Transform Spectrometer

The Fourier transform spectrometer is central to this work; a description of Fourier transform spectrometry was provided in chapter 3. For the experiments described here, a Bruker

IFS 66v/S Fourier transform spectrometer was used. Shown in figure 4.6, this spectrometer provided a maximum resolution of 0.12 cm^{-1}, allowing characterisation of both gas-phase and aerosol components. The Bruker IFS 66v/S spectrometer scanner uses air bearings, so that the minimum pressure achievable inside the spectrometer is approximately 2 mb; high purity dry nitrogen was used as the buffer gas since nitrogen has no absorption lines in the spectral regions measured, and the partial pressures of H_2O and CO_2 were thus minimised. In this work, *high resolution* measurements refer to spectra recorded at 0.12 cm^{-1} or 0.4 cm^{-1} resolution, and *low resolution* refers to a resolution of 4.0 cm^{-1}.

Figure 4.6: The Bruker IFS 66v/S spectrometer, sited inside the air-conditioned spectrometer room at the MSF, connected by an evacuated tube to the large aerosol cell visible through the window.

4.1.5 Optics

All optical paths external to the FTS and the aerosol cell were evacuated during experiments to ensure minimal contamination of spectra by gas absorption due to CO_2 and H_2O vapour, and to prevent condensation of volatiles on window surfaces. Likewise, the detector box and cell outer were also maintained under high vacuum through the experiments. All optical windows were wedged at an angle of approximately 1° to the incident radiation to minimise reflections interfering with the main beam; this effect is known as channelling, and is manifest in spectra by regular baseline oscillations. The main optical components are detailed below.

Light source. A silicon carbide globar (150 W) light source was used in the spectrometer for all experiments. This source provides high stability (less that 1 % drift in intensity of output radiation over 24 hours) over the spectral region of interest (approximately 750-8000 cm^{-1}).

Beamsplitter. A germanium coated potassium bromide (KBr/Ge) beamsplitter, with a spectral range of 550 to 13000 cm^{-1}, was mounted within the FTS. This wide spectral range prevented the need for beamsplitter changes, allowing simultaneous measurement of both the infrared absorption and near-infrared scattering region.

Fourier transform spectrometer output beam. The FTS beam was collimated into a parallel beam of approximate diameter 25 mm by a 90° off-axis paraboloid mirror (focal length 121 mm) before leaving the spectrometer. It was then directed through evacuated tubing, through the cell volume (of internal path 434 mm), and finally to a detector box as shown in figure 4.7. Care was taken to ensure minimal divergence/convergence over the entire optical path between the spectrometer and detector. However, the beam was slightly divergent and was clipped by the minimum aperture of the cell output window to an approximate diameter of 35 mm, with a slight loss of beam intensity.

Detectors. Low resolution (4 cm^{-1}) spectra for characterisation of broad band aerosol features were measured with a room temperature deuterium lanthanum triglycero sulphide (DLaTGS) detector. The electrical response of this detector is very linear to photon detection rate over the 350 to 8000 cm^{-1} spectral range. The DLaTGS detector provides accurate extinction measurements due to this inherent linearity, has a broad spectral response, and adequate signal to noise ratio (SNR) in low resolution spectra. In early experiments, such as those conducted to produce the composite spectrum for the comparison with Biermann *et al.* (2000) data (section 2.5.4), a liquid nitrogen cooled indium antimonide (InSb) detector was used to measure from 1800 to 13000 cm^{-1}. The range was not adequate for determination of the absorption features of aerosol as well as the scattering component.

For the measurement of high resolution spectra for characterisation of gas absorption lines, a broadband mercury cadmium telluride (MCT) detector was used; this is cooled by liquid nitrogen to reduce thermal electron noise. This detector also has wide spectral coverage from 550 to 7000 cm^{-1}, but has higher SNR than the DLaTGS detector despite the requirement to use small beam apertures in the FTS to achieve high resolution. Tests revealed a factor of 9 improvement in SNR with the MCT compared with the DLaTGS detector per unit time at the highest resolution of the Bruker IFS 66v/S (0.12 cm^{-1}). A well known problem with MCT detectors at high integrated radiation power is non-linearity of electrical output with respect to photon incidence (Chamberlain, 1979); this non-linear response with intensity leads to spurious features in the spectra after Fourier transform. However, the spectra measured with the MCT detector used in this study showed no perceivable zero offset, indicating good linearity even at maximum incident power.

The detector box. A remotely controlled (via a stepper-motor) switching mirror allowed the FTS beam to be steered on one of two detectors mounted in the box at a time. A gold-coated 90° off-axis paraboloid mirror (focal length 43 mm) focussed the parallel beam of light to a sharp image on the surface of the detector element. Unfortunately, the switching mirror was not operational in the settling experiments when only a single DLaTGS detector was used for both high and low resolution measurements.

Windows. Optically transparent windows were used to segregate the evacuated tubing from the spectrometer and to isolate the cell contents, as shown in figure 4.7. Potassium bromide (KBr) windows, which are optically transparent from 400-40000 cm^{-1}, were sited external to the aerosol cell. The two ports on the inner cell were fitted with barium fluoride (BaF_2) windows (47 mm diameter, 4-5 mm thick). Barium fluoride provides relatively good transmission through most of the region of interest and has good resistance (compared to other candidate materials) to the acids in the cell volume. The negative aspects are that BaF_2 has a minimum wavenumber cut-off at 750 cm^{-1}, and transmission at wavenumbers near this limit was found to exhibit a pronounced temperature dependence, as characterised in section 5.4.2.

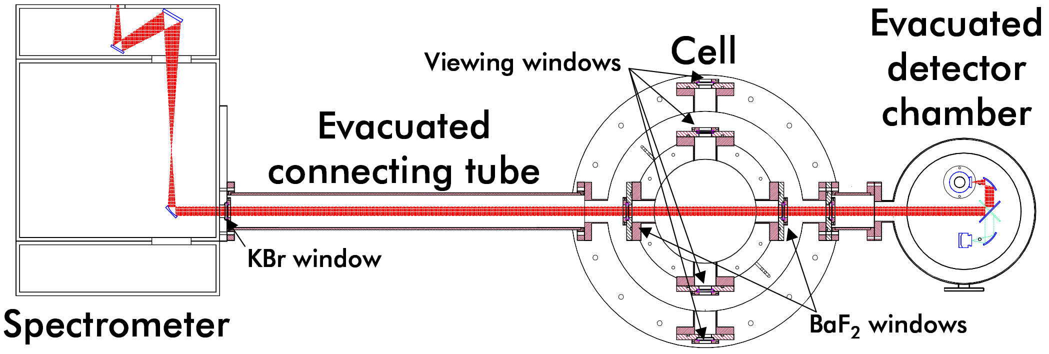

Figure 4.7: Schematic of the optical configuration of the large cell in plan view; the light beam is illustrated in red, mirrors/detectors in blue. Note the two detectors and rotating mirror (main experiments) in the detector box. The unassigned outer windows are also KBr (R.G. Williams, RAL).

The combination of optics described here yielded spectral coverage sufficient to characterise the absorption and scattering characteristics of aerosol with low resolution DLaTGS measurements over the spectral region from 750 to 8000 cm^{-1}. Additionally, high resolution MCT measurements were sufficient to characterise H_2O, HNO_3, CO_2 and CO gas-phase absorption lines.

4.1.6 Experimental performance issues

Located outside the air-conditioned spectrometer room in the main experiments, the gas-handling system in particular may be sensitive to the large temperature variation that can potentially occur in the open laboratory. Typically, temperature extremes varied from around 5°C in winter to 30°C in summer; this would affect SVP entrained over the baths. Therefore, laboratory temperatures in the vicinity of the gas handling system were logged. Fortunately, the typical range of temperature over a single day was found to be relatively small, producing no observable effects on the aerosol composition.

Early experiments involving HNO_3 at near-room temperature led to contamination of the BaF_2 cell windows in the form of a permanent (not removable by leaving the cell under vacuum), optically absorbing, film on the surface which most likely formed from reaction with HNO_3 vapour. Since vapour pressures of HNO_3 increase with temperature, the greatest risk of such contamination is correspondingly at the highest experimental temperatures. However, addition of sulphate aerosol, as required for generation of aerosol in the large cell, resulted in reduced HNO_3 vapour concentrations in the cell, preventing most window contamination below room temperature (section 5.4.2). Therefore, in the experimental procedure employed, the N_2/HNO_3 flow was initiated last (after the sulphate aerosol) to prevent excess quantities of HNO_3 vapour in the cell.

Another potential problem with room temperature experiments was lack of turbulent mixing in the cell. The experiments rely on a certain degree of turbulence inside the aerosol cell to thoroughly mix the contents. In the first of the main experiments at room temperature a definite column of aerosol was visible by eye through the unused viewports (at 90° to the FTS beam). The aerosol appeared to descend through the centre of the cell with minimal mixing. However, this was not observed at lower temperatures or in any later experiments in the following days, probably because the cell temperatures did not usually equilibrate completely, resulting in enough turbulent mixing through convection to ensure the aerosol particles were thoroughly mixed and uniformly distributed.

Gas absorption lines due to CO_2 in the cell were observed in spectra attained on a number of days in the main experiments. The cause was traced to a contaminated N_2 supply cylinder, and the effect on aerosol composition would have been minimal, as explained in section 5.3.2.

In all experiments the vacuum in the outer jacket deteriorated during cooling, typically from 4×10^{-3} mb to about 7.4×10^{-2} mb at low temperature. It was determined that build up of contaminant CO_2 and H_2O vapour from air in the outer jacket had negligible spectral effects; condensation in a liquid nitrogen cooled trap most likely depleted these volatiles to insignificant concentrations.

4.2 Methodology

The main measurements of aerosol spectra for this thesis were conducted in two distinct regimes: (1) a steady-state flow situation in which the flow rates of all gas/aerosol streams into the cell were not changed during each day of experiments, or (2) settling experiments where, at the end of a day of experiments the cell contents were isolated to allow aerosol to settle under gravity (section 5.1.2). The differing methodology of these experiments is described following an overview of general experimental methodology.

Background spectra. Background spectra were measured on every day of experiments. The measurements of the empty cell provide the system spectral response comprising the detector sensitivity, the modulation efficiency of the interferometer, the spectral signature of the light source, and the transmission of cell windows. This is necessary since uncalibrated spectral intensity (or radiance since the two are proportional) is the quantity recorded by the detectors. The extinction spectrum of aerosol can then be calculated from $E_m(\tilde{\nu}) = 1 - T(\tilde{\nu}) = 1 - I_m(\tilde{\nu})/I_b(\tilde{\nu})$, where T is transmission, $I_m(\tilde{\nu})$ is the measurement radiance, and $I_b(\tilde{\nu})$ is the background radiance. Initial background measurements were recorded with the cell either at experiment temperature or at the first temperature of a cooling cycle, and usually filled with N_2 buffer gas to 200 Torr pressure. Final backgrounds were also measured at the end of each experiment day to determine any change in detector output that may have occurred and whether any contamination (i.e. acid corrosion) of the BaF_2 window surfaces had taken place. For low resolution DLaTGS aerosol spectra in the main experiments (excluding the earlier settling experiments), the temperature dependence of the BaF_2 windows was characterised (from a calibration experiment) to enable interpolation to a corrected background at any temperature from the initial and final background measurements of the experiment day (section 5.4.2).

Aerosol equilibration. For all experiments an internal gas pressure of 200 Torr was maintained by buffering the gas/aerosol flows with inert pure nitrogen. Although the internal pressure is relatively high compared with that experienced at relevant stratospheric altitudes, it promotes rapid thermal transfer and thus cooling of the internal gas to the wall temperature, and also provides sufficient buffering of equilibrium vapour pressures (of H_2O and HNO_3), as measured along the beam path, from colder parts of the cell (such as the liquid nitrogen trap, next to VMV1 in figure 4.3). For example, Niedziela *et al.* (1999) found that in a cell of approximate length 2.5 metres and diameter 10 cm, buffer gas pressures (helium) above 100 Torr were required to isolate the vapour pressures (of H_2O and HNO_3) in the measurement section of their aerosol cell from cooler parts of their cell.

As gas and aerosol flows through the cell, the cell walls are *conditioned* by the aerosol. Three distinct mechanisms for conditioning of the cell walls can be defined – adsorption, condensation and coalescence/sticking. If the cell walls are *dry* and have been under vacuum for some time, they will have a high capacity to absorb gas phase volatiles – H_2O molecules become attached to empty microscopic cavities in the steel surface; the process of adsorption. The loss of gas phase molecules through this process may be significant, but is short-lived as cavity sites diminish. Condensation of volatiles on cell walls and coalescence/sticking of aerosol to the cell walls would have the same result of coating the walls in aqueous acidic solutions of the same composition as the aerosol.

Visual inspection of the interior walls of the inner vessel following experiments revealed a general coating and pooling of acidic aqueous solutions, emphasising the extent of wall conditioning in the experiments even after extended (sometimes weekend) periods of placing the cell under vacuum. When a different gas/aerosol mixture enters the cell, or the flow rates are changed, it therefore takes a significant period of time for the aerosol inside the cell to *equilibrate* (by vapour exchange) with material on the walls, and for the stream passing through the cell to completely replace the former contents (longer than the residence time, due to turbulent mixing and diffusion). This process will take a longer period of time if the cell walls have been previously conditioned with a radically different composition solution. It was typically found that periods longer than four hours at room temperature were required for initial conditioning of the large cell walls at the start of each experimental day. Low resolution spectra were measured during this process; when no variation in extinction spectra could be detected between successive aerosol measurements it was assumed that the aerosol flow through the cell had reached *steady-state* equilibrium. This test was used throughout the experiments to ensure stability. Although variation of aerosol composition after a change in temperature is thermodynamic and should not depend on the residence time, the large relative quantity of liquid on the cell walls delays equilibration of aerosol when temperature is varied.

4.2.1 Settling experiments

With one exception, the settling experiments started with a cooling of the cell to approximately 230 K for initial low resolution (4.0 cm^{-1}) backgrounds at 200 Torr N_2 pressure with dry N_2 flowing at 1000 sccm (20 minutes residence time). The gas/aerosol flow in to the 200 Torr N_2 atmosphere was then initiated. After stabilisation of the aerosol, indicated by successive stable low resolution spectra, a high resolution (0.12 cm^{-1}) DLaTGS spectrum was recorded to attempt to measure gas-phase absorption lines. Following the 20 minute duration of the high resolution DLaTGS measurement (absence of a switching mirror in the detector box prevented use of an MCT detector), further low resolution measurements were undertaken to ensure aerosol stability. If the aerosol spectra were stable, the cell was then cooled to 220 K and the process repeated. Finally, the cell was cooled to 190 K (25 minute residence time), no high resolution measurements were made (since equilibrium H_2O and HNO_3 vapour pressures are too low to be detectable at this temperature) but, after aerosol stabilisation, the cell contents were isolated to allow settling under gravity. It was typically possible to measure the spectral signature of settling aerosol for approximately one hour before absorbance by aerosol became negligible. These experiments were performed on seven separate experimental days, with six different flow-rate combinations (table 4.1). The analysis of the settling experiments is described in section 5.1.2.

Table 4.1: Main experiment configurations, in reverse chronological order.

Date	△Flow rate (sccm)					Bath (wt%)	◇Temps (K)	
(2000)	$wHNO_3$	$dHNO_3$	SO_3	wN_2	Total	H_2SO_4	Max	Min
Main experiments								
⊙01/12		451	400	500	1351	74	286	190
⊙30/11		450	400	500	1350	74	286	190
29/11		242	400	234	876	74	283	190
28/11		424	400	206	1030	74	280	190
⊘24/11		11	400	9	420	74	285	190
⊘23/11		43	400	48	491	74	285	190
⊘22/11			400	205	605	74	285	190
⊘21/11		170	400	78	648	74	292	190
03/11			400		400	74	286	260
Settling experiments								
29/03	250		400	350	1000	60	230	190
28/03		200	400	400	1000	60	230	190
27/03		275	400	325	1000	60	230	190
24/03		250	400	350	1000	60	230	190
23/03		300	400	300	1000	60	230	190
22/03	200†		400	400	1000	60	225	190
21/03	200†		400	400	1000	60	230	190

◇	Temperatures are only approximate experiment targets.
†	Low level of HNO_3 solution in bubbler.
△	Maximum 500 sccm each.
⊙	Aerosol froze at lowest temperatures.
⊘	CO_2 contamination; allowed accurate temperature retrieval.
w / d prefix	Flow passes over the water bath (w), or straight into the cell (d).

4.2.2 Main experiments

The main experiments were designed to systematically obtain aerosol spectra simultaneously with data that could be used to provide the best information on aerosol compositions. Two specific improvements were therefore made to the procedure adopted in the settling experiments in order to improve information on composition of the PSC particles. Firstly, a switching mirror was installed in the detector box so that high resolution spectra could be obtained with an MCT detector and low resolution spectra could be rapidly obtained with the DLaTGS detector. Secondly, the first measurements of aerosol on each day were undertaken at a higher temperature (approximately 270 K as opposed to 230 K). This provided a greater range of temperatures over which both H_2O and HNO_3 vapour pressures could be determined from high resolution spectra. Two types of high resolution spectra were measured: one at 0.12 cm^{-1} resolution for H_2O absorption, and another at 0.4 cm^{-1} for improved characterisation of the main HNO_3 gas absorption feature.

Two other important changes were made in the main experiments. No settling experiments were undertaken since analysis had revealed that no significant changes in aerosol mean size occurred during the particle sedimentation process (section 5.1.2). The other change was that initial background spectra were recorded with the cell evacuated, rather than pressurised to 200 Torr dry N_2, since this was shown to cause no discernable errors in the calculated extinction. The cell was then filled to 10 Torr pressure with dry N_2 (to prevent volatiles condensing on the windows), and subsequently pressurised to 200 Torr operational pressure with a gas/aerosol flow as to be used on the experiment day. This approach was found to decrease the aerosol conditioning time as the aerosol was not initially dehydrated by 200 Torr of dry N_2 in the cell.

The experimental procedure employed is illustrated in the flow diagram in figure 4.8. The flow rates used during each day are detailed in table 4.1, along with the appropriate date of each experiment – experiment dates are useful for comparing with other results obtained from analysis. As can be seen, the main experiments comprised nine separate

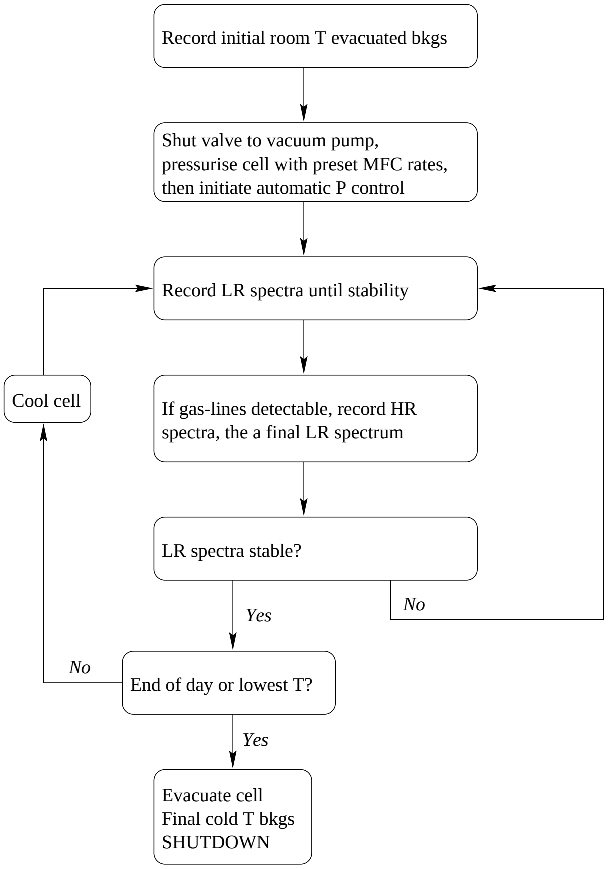

Figure 4.8: Flow diagram of typical experimental procedure during the main experiments. Acronyms and abbreviations: HR=High Resolution, LR=Low Resolution, T=Temperature, P=Pressure, bkgs=background spectra, and MFC=Mass Flow Controller.

experiment days, with eight distinct flow rate combinations.

4.3 Summary

The experiments used for the main analysis presented in this thesis were conducted in two distinct sessions:

1. **The settling experiments** focussed on measurements to characterise size distribution changes with gravitational settling of aerosol;
2. **The main experiments** produced data for characterisation of aerosol evolution and

refractive index derivation.

Aerosol particles were successfully generated in the cell from room temperature to approximately 195 K. Sixteen experiments were performed delivering a compact set of data for analysis of aerosol size distribution and refractive indices. The successful generation of ternary solution aerosol at low temperatures is confirmed through analysis presented in the following chapter.

Chapter 5

System Characterisation

The experiments described in chapter 4 were carefully designed to provide not only the aerosol spectra from which refractive indices can be derived, but also the requisite information to determine the physical and chemical states of the aerosol/gas mixture in the cell, which includes:

1. The type of aerosol in the cell;
2. The behaviour of aerosol in the cell;
3. Partial pressures of volatiles in the cell;
4. Mean gas temperature along the spectrometer beam path.

The gas partial pressures (3), and gas temperatures (4), are derived through the retrieval of information from gas absorption lines in an implementation of the optimal estimation technique (see Appendix), henceforth termed the *gas retrieval.* The same analysis also provides information on the wavenumber shift due to beam divergence in the Fourier transform spectrometer (FTS).

It has been suggested that discrepancies between refractive index data published by different research groups of the same types of PSC may be caused by different experimental conditions. However, the mechanisms by which these influences affect optical properties have not been resolved in literature. Therefore one of the more important aspects of the experiments conducted for this work is to characterise the experimental system and quantify knowledge of system parameters.

5.1 Aerosol characterisation

The purpose of this section is to provide an overview of the characterisation of aerosol evolution in the aerosol generation system (section 4.1.2) for experiments undertaken for this thesis.

5.1.1 STS aerosol in the large cell

The main experimental aim of this work is the generation and measurement of ternary solution aerosol. It is thus of prime importance to validate the existence of ternary solution aerosol in the cell. The arguments presented here can be segregated in two categories: (1) independent evidence of the existence of ternary solution aerosol, and (2) evidence from comparison with published data.

Independent proof. Measurements of the infrared extinction spectra of liquid sulphuric acid aerosol have been undertaken previously at RAL (Heathfield *et al.*, 1999), and were repeated in these experiments. As a result, the absorption bands of sulphate aerosol have been well observed as a function of composition and temperature. When the N_2 buffer gas was passed above the HNO_3 solution bath and directed into the cooled cell, the aerosol extinction spectrum changed distinctly – a new set of absorption features appeared repeatedly superimposed on the pure sulphate spectrum. Clearly the aerosol were no longer binary solutions of sulphuric acid, and ternary solution would be expected to be formed. However two alternative scenarios to the production of ternary aerosol are: (1) the observed changes in aerosol spectra may not be due to condensation of HNO_3 on the sulphate aerosol but some unknown volatile instead, or (2) that the HNO_3/N_2 may directly introduce or lead to separate nucleation of binary nitrate aerosol, which co-exists with a separate population of binary sulphate aerosol.

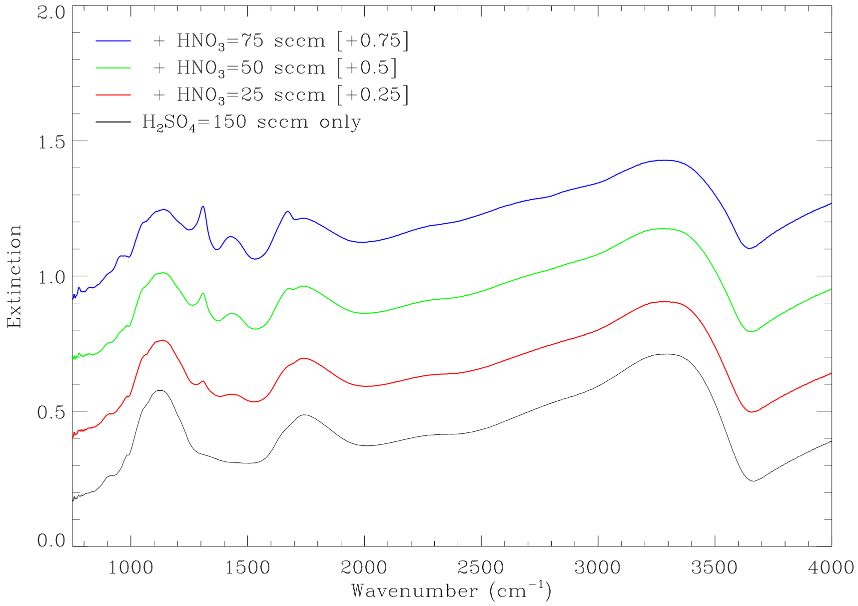

Figure 5.1: Aerosol extinction spectra recorded at 200 ±5 K illustrating the effect of adding HNO_3/N_2 at various rates with sulphate aerosol.

The new absorption features become more pronounced as the HNO_3/N_2 flow rate is increased, see figure 5.1, indicating that the change in composition is directly related to the condensation of HNO_3. Furthermore, measured HNO_3 vapour pressures were much lower when the HNO_3/N_2 stream was combined with the sulphate aerosol. This strongly indicates the condensation of HNO_3 into the existing liquid aerosol.

If a separate stream of binary nitric acid aerosol was created, it should not be detectable from absorption spectra alone since, according to published works (Anthony *et al.*, 1997; Koehler *et al.*, 1992; Biermann *et al.*, 2000), there are no infrared absorption features specific to the ternary system (that do not exist in binary aqueous H_2SO_4 and HNO_3 spectra). The retrieval of STS aerosol size distribution typically shows that the distribution is monomodal, and almost monodispersed (section 5.1.2) – i.e. the aerosol particles are all of the same size. It would be unlikely that two separate populations of binary nitrate and binary sulphate aerosol would be of the same size – in this case one would expect a bimodal distribution. Moreover, test experiments were conducted where only the HNO_3/N_2 stream was directed into the cell at the usual flow rates. In all cases binary nitrate aerosol were not observed. These findings indicate the second scenario is highly unlikely, and that a single population of ternary solution aerosol is present. In any case, if separate binary aerosol populations were initially formed on entry into the cell, these should equilibrate through vapour transfer to form a homogeneous ternary solution aerosol by the time the aerosol reaches the viewports (previously discussed in section 4.1.1).

Evidence of production of ternary solution aerosol from comparison with published data. Although little data for directly-measured ternary solution aerosol exists, a comparison can be undertaken against the thin-film refractive index data produced by the Biermann *et al.* (2000) ternary calculator, described previously in section 2.5.3. Such a comparison, reviewed in section 2.5.4, revealed that the bands in the ternary solution aerosol spectra measured in this thesis agree across most of the spectral range with those produced by Biermann *et al.*. Although significant differences in the spectra exist in some spectral regions, absorption band centre positions are qualitatively consistent. The broad agreement is also repeated in quantitative comparison with the ternary solution aerosol spectra published by Anthony *et al.* (1997). Finally, the presence of aerosol absorption bands in spectra measured for this thesis coincide with binary nitric acid and sulphuric bands compiled from literature (section 2.5.8). The general agreement with published ternary solution data is a strong indication of the existence of ternary solution aerosol in the cell.

Summary. The evidence that ternary solution aerosol was produced in the experiments of this thesis rests on five main points:

1. The spectra are qualitatively very similar to published aerosol spectra and thin-film indices;

2. Binary nitric acid aerosol are not produced from the HNO_3/N_2 stream alone – only

gas phase HNO_3 is measured in the cell;

3. Upon addition of the HNO_3/N_2 stream with cooled sulphate aerosol, most gas phase HNO_3 is removed and aerosol features change significantly – indicating HNO_3 is taken up by sulphate aerosol producing ternary aerosol;

4. Further evidence that mixed binary aerosol populations do not co-exist is provided by the monomodal size distribution observed in the analysis;

5. The long residence time in the large cell should ensure aerosol is fully equilibrated – as noted by Anthony *et al.* (1997), aerosol of a single composition would form through vapour exchange of any co-existing binary aerosol (section 4.1.1) in this time.

5.1.2 Settling experiments: Aerosol size distribution

Kramers-Kronig based techniques have been widely used in previous studies to derive refractive index data from laboratory spectra of mimic PSC particles and thin-films. In the case of aerosol spectra, this requires two spectra of the same composition: one of small particles of median radius $r << 0.1$ μm, that can be approximated to non-scattering, and one of larger particles of $r >> 0.1$ μm (section 2.3.1). In order to assess whether settling experiments with the large cell resulted in sufficient particle size variation for a Kramers-Kronig analysis, a particle scattering retrieval was used to analyse settling spectra of STS aerosol, over the spectral range 4500 to 7000 cm^{-1} (1.42 to 2.2 μm), from a typical settling experiment (21/03/2000). This retrieval was a simplified version of the refractive index retrieval, using Mie theory to calculate radiation scattered out of the beam by spherical particles and the absorption of radiation by the particles. In this case absorption was neglected – only the effect of scattering due to log-normally distributed particles was modelled. This is valid since above about 4500 cm^{-1} the imaginary part of the refractive index, $k(\tilde{\nu})$, which defines the absorbing properties of a medium, is very close to and tends towards zero with increasing wavenumber. The real part of the refractive index, $n(\tilde{\nu})$, tends towards a constant value with increasing wavenumber above about 4500 cm^{-1}, giving a constant refractive index value approaching m_∞, the zero frequency (infinite wavenumber) refractive index offset. In addition to the size distribution of aerosol, m_∞ defines the visible scattering of the sample.

Spectra were recorded in immediate succession after the flow had been stopped and the cell contents isolated. This yielded a 62 minute series of good quality settling spectra separated by 2 minutes during settling of the STS aerosol at 190 K. The results of the Mie analysis for the scattering properties of each spectrum are plotted in figure 5.2. In each spectrum, 129 measurement points were fitted between 4500 to 7000 cm^{-1}, and the following parameters were set to constant values of $m_\infty = 1.5$ and $k(\tilde{\nu}) = 0$.

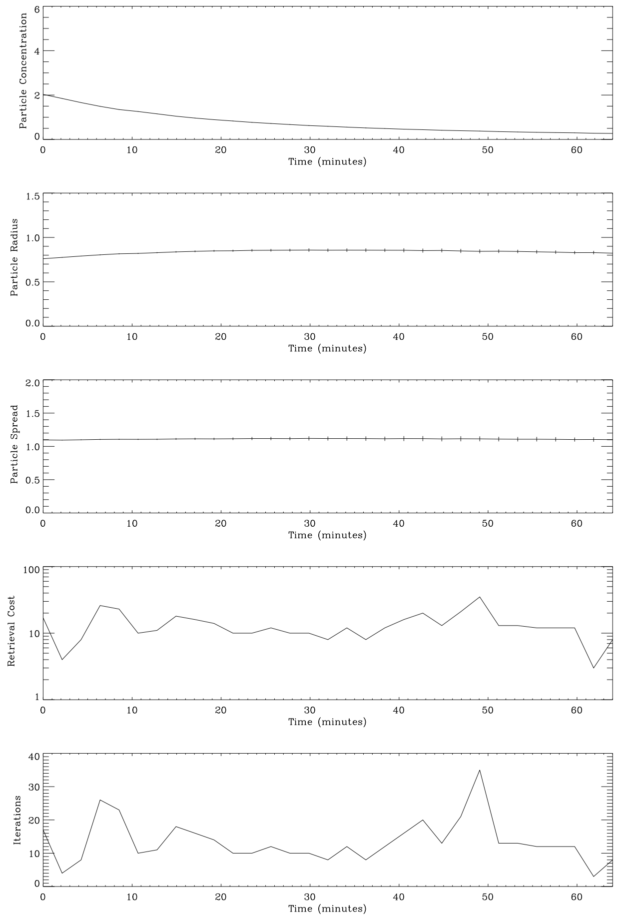

Figure 5.2: Fit of N [10^6 cm^{-3}], r [μm] and σ for 129 points with m_∞ fixed at 1.5.

The results show that the median radius of the distribution, r, and the spread, σ, do not change significantly outside the range of the very small retrieval error bars, and do not have a trend within the error bars, during the course of settling. Only the number density of the aerosol falls, as expected from the decrease in extinction that was observed. The main cause of the fall in aerosol number density is probably loss to the cell walls. The mean particle size distribution parameters (N, r, and σ) retrieved were 0.765×10^6 cm^{-3}, 0.836 μm and 1.11. As σ has a value close to 1 this is nearly a monodispersed distribution, i.e. all the particles are nearly of the same size. Using Stoke's law (section 4.1.1) the calculated time for a particle of 0.863 μm diameter to fall 1 m is about 112 minutes. Monodispersed particles would all fall at the same speed, and if the coagulation rate is low the particle size would not change, as observed. The flow rate prior to settling forces the particles through the cell much faster (only 24 minutes at 190 K, section 4.1.1) than they fall under gravity, so cutting off this flow would also cause the number density of aerosol in the cell to diminish with time.

Thus, it is concluded that the settling experiments did not significantly vary the particle size, and that refractive index data cannot be derived from these results using the Kramers-Kronig technique.

5.1.3 Aerosol evolution model

A simple aerosol evolution model was used to investigate the evolution of aerosol particles in the large cell. This utilises thermodynamic models (Carslaw *et al.*, 1995; Lin and Tabazadeh, 2001) to characterise the temperature and vapour-dependent nature of aerosol composition and gas/liquid partitioning inside the cell. The main aim of the exercise was to provide a basic understanding of the characteristics of aerosol development in an enclosed system and to identify the controlling factors, not to precisely predict the development of aerosol in the large cell which would have required a considerably more advanced microphysical model.

Model description. The central principle of the aerosol evolution model is the conservation of the total volatile inventory, partitioned between liquid and gas phase, in the enclosed system of the large cell. The inputs to the cell are the sulphate aerosol, and the gas volatiles entrained above the water, sulphuric and nitric acid baths. The temperature dependant aerosol droplet evolution in the cell is modelled by relating the condensation/evaporation of volatiles from the aerosol to determine droplet composition and radius where equilibrium is reached. Equilibrium vapour pressures above aerosol droplets are determined from the thermodynamic model of Lin and Tabazadeh (2001). The problem is solved through an iterative minimisation process; the droplet mass and therefore volume can be calculated with knowledge of composition (since m_a is constant) which is constrained by the limited availability of volatiles inside the cell and equilibrium vapour pressures. The main assumptions of the model can be summarised:

- Saturated vapour pressure (SVP) of H_2O (over the water and sulphuric acid baths) and H_2O and HNO_3 (above the nitric acid bath) is entrained into the N_2 buffer gas above these baths;
- The total equivalent pressure (quantity) of H_2O and HNO_3 is conserved in the system;
- Aerosol composition is in thermodynamic equilibrium with vapour phase H_2O and HNO_3;
- The total mass of sulphuric acid (m_a) per droplet is conserved – i.e. no appreciable coagulation of particles takes place (confirmed by the settling experiments);
- The number density of particles generated over the H_2SO_4 bath remains constant (at the deduced mean value of 1.4^6 cm^{-3});
- Total SVP and m_a entrained into the flows remains constant during each day of experiments (since variation of ambient laboratory temperature is small);
- Aerosol droplets are all of the same physical size (monodispersed distribution).

Experimental parameters are constrained to those observed in the experimental system where known. For example, the retrieved number density (from the refractive index retrieval) varies between each day of experiments due to the different flow rate combinations used, but the number density of aerosol generated above the H_2SO_4 bath is stable since a flow rate of 400 sccm SO_3 / N_2 is used in all experiments presented here. The number density of aerosol above the sulphuric acid bath is calculated from the number density of aerosol observed inside by assuming simple dispersion of aerosol relative to flow rates, and the mean calculated experiment value of 1.4^6 cm^{-3} (section 4.1.2) is used in the model; loss of particles to tubing prior to entry in to the cell is inherently taken into account. The mass of sulphuric acid per droplet, $m_a = 1.17\times10^{-12}$ g, was determined from a calibration experiment. Since the flow rate above the sulphuric acid bath was not varied, it is assumed that m_a remains constant and also that the number density over the sulphuric acid bath remains constant (this is a simplification since the number density, and thus m_a, varies slightly between experiments probably due to variations in external laboratory conditions). The initial radius and volume of aerosol above the sulphuric acid bath was calculated using the mass of acid, assuming particles equilibrate to the bath composition. The Kelvin effect, which acts to increase vapour pressure above small droplets, was found to be negligible in the aerosol radius range of the experiments if a monodispersed size distribution is assumed, as used in the model.

Flow rates across each bath relative to the total flow rate were used to evaluate the vapour pressure dispersed into the cell volume. The thermodynamic model of Carslaw *et al.* (1995) was used to determine SVP of HNO_3 and H_2O above the 69 wt% HNO_3 bath (employed in all experiments), and SVP of H_2O vapour above the pure water and sulphuric acid baths. The calculated SVP pressures of H_2O and HNO_3 were used to evaluate the corresponding vapour pressures dispersed (according to relative flow rates) into the cell. Similarly, it was straightforward to evaluate the reduction of aerosol number density from above the sulphuric acid bath into the cell volume as a ratio of flow rates.

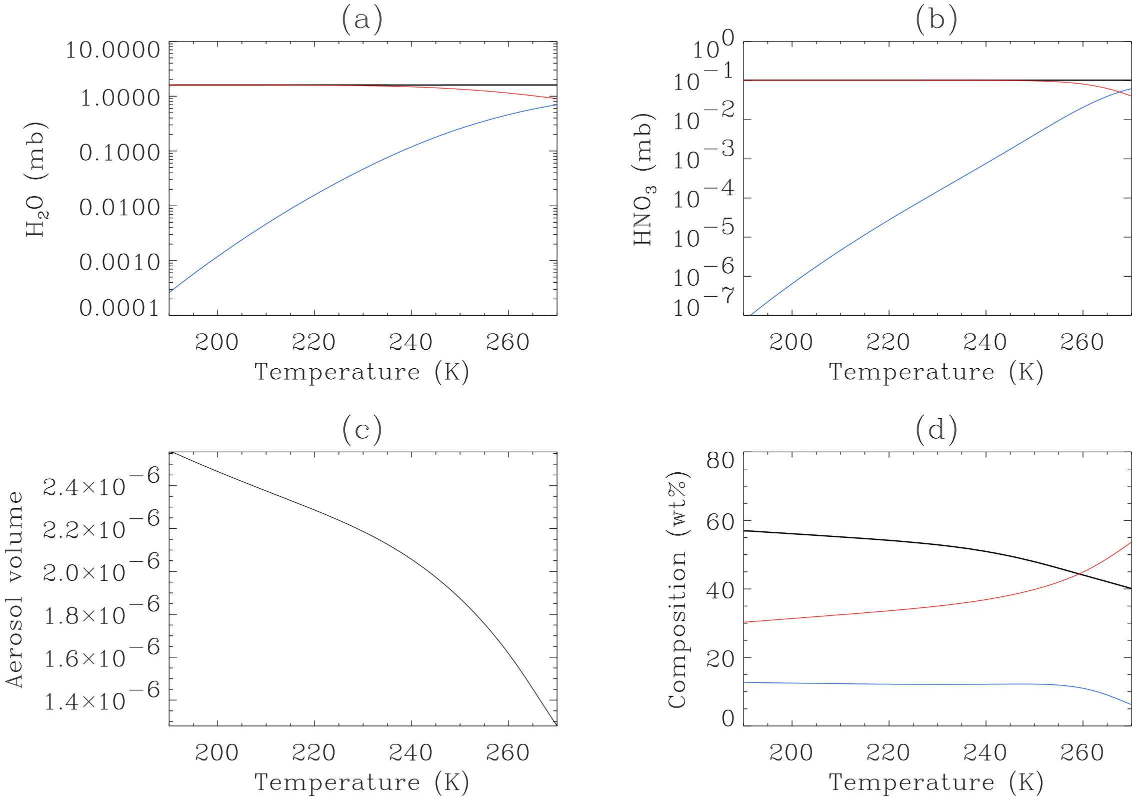

Figure 5.3: Output from the aerosol evolution model, flow rates set to those of 23/11/00; SO_3/N_2 = 400 sccm, HNO_3/N_2 = 43 sccm, and H_2O/N_2 = 48 sccm. Panels (a) and (b) show partitioning for H_2O and HNO_3 respectively, total equivalent pressure is shown by the thick black line, the gas phase component is the blue line and the liquid component the red line. Aerosol volume density (dimensionless – m^3/m^3) is shown in panel (c), and the composition is displayed in panel (d) where the thick black line indicates total wt% H_2O, and the red and blue lines are wt% H_2SO_4 and HNO_3 respectively.

Results. Output from the aerosol evolution model is presented in figure 5.3. It can be seen that variation of composition with temperature occurs in a different temperature range to that in the stratosphere due to the high number density of aerosol in the cell (necessary to observe aerosol extinction) and consequent high initial vapour pressures of H_2O and HNO_3. The greatest change in composition occurs above about 240 K, and further cooling does not strongly influence composition. The main implication is that the composition of aerosol in the large cell at low (sub-240 K) temperatures is a strong function of flow rates but not temperature. This is an important finding since the achieved temperature accuracy of ±2.13 K in the main experiments would have translated into a high composition uncertainty in a stratospheric-like thermodynamic system.

The aerosol evolution model results presented in figure 5.3 are calculated with the flow rate configuration used during the main experiments on 23/11/00. Compared with the analysed results for this experiment day (figure 8.2), the model over-estimates the condensation of volatiles yielding a larger aerosol volume than retrieved. This indicates that, in the experiments, volatiles are *lost* through an unknown mechanism. The most likely cause is that gases are not being transferred at full SVP into the N_2 flow (possibly due to evaporative cooling above the baths), or that volatiles are being lost to the walls (through adsorption, absorption, condensation, or loss to the liquid N_2 trap). The former process implies that faster flow rates would be less efficient at entraining volatiles, and the latter process would be difficult to model. Loss factors for the gaseous volatiles were introduced into model calculations but it proved very difficult to match observations without adjustment of m_a; although this parameter was deduced from a test experiment with a 400 sccm SO_3/ N_2 flow, the assumption that m_a is constant from day-to-day and the accuracy of the deduced value are uncertain.

Summary. The aerosol evolution model provided a useful insight into the fundamentals of the aerosol system, indicating that:

- Temperature-dependant variation of aerosol composition in the cell takes place at much warmer temperatures than the stratosphere due to the high number density of aerosol in the cell and limited volume;

- The important factors for determination of aerosol composition in the cell at low temperatures are the relative flow rates of aerosol, volatiles, and buffer gas – not precise knowledge of temperature.

The composition of aerosol in the model did not significantly change below 240 K, indicating that composition determined at this temperature could potentially relate to aerosol composition at polar stratospheric temperatures, assuming no change in flow rates. A more detailed model, incorporating improved representation of aerosol microphysics in the large cell, could be employed to provide a more realistic representation of aerosol evolution, but this was not necessary to achieve the aims of this thesis.

5.2 Retrieval of information from gas spectral lines

The absorption of infrared radiation by gas molecules appears as clusters of closely spaced, sharp lines known as bands. The appearance of these bands depends on fundamental spectroscopic parameters:

- **Line position.** The location of the line centre, $\tilde{\nu}_c$, is determined by the gas, the mean velocity (Doppler shift) and potentially pressure (pressure shift);
- **Line strength.** The line intensity or strength, related to the concentration dependent magnitude of absorption at line centre, is dependent on the energy level of the transition (the transition probability), and temperature;
- **Line shape.** The line shape is related to pressure broadening and temperature.

The determination of fundamental parameters from spectra of gas line absorption is an example of an *inverse problem* which can be solved using a retrieval process based on optimal estimation. The gas retrieval also required formulation of an appropriate forward model to simulate high resolution gas absorption spectra measured with the FTS. The parameters that define gas absorption, principally temperature (T), total pressure (P), and volume mixing ratio (VMR) of the absorbing gas species, are then varied iteratively to obtain a *best fit* to the measured high resolution spectrum of gas absorption.

This work represents the first instance of the use of high resolution Fourier transform infrared (FTIR) spectra for identification of ternary aerosol composition. Elsewhere, tuneable diode laser (TDL) spectrometry has been used to identify H_2O and HNO_3 gas components in separate binary sulphate and nitrate aerosol experiments (Niedziela *et al.*, 1999, 1998b; Norman *et al.*, 1999). Use of high resolution FTIR spectroscopy to measure gas lines has the distinct advantage that spectral coverage is far wider than with a TDL system, and the wavenumber accuracy is considerably better.[1] Hence, different species can be resolved in the measured spectrum and retrieval windows can be selected for optimal information content, avoiding saturated lines, and leading to less dependence on spectroscopic errors for a single line. The disadvantage of the Bruker FTS compared with a TDL, with regards to gas characterisation, is that SNR is comparatively lower. A TDL can thus resolve gas lines of lower intensity and can be used to measure vapour pressures of H_2O and HNO_3 at lower temperatures where equilibrium vapour pressures become too low to be measured with a FTS. Indeed, the FTS cannot be used to measure equilibrium vapour pressures at temperatures below approximately 220 K (this temperature depends on aerosol composition).

5.2.1 The forward model

The forward model of the gas retrieval simulates high resolution spectra of gas absorption measured with the FTS. The state vector consists of the unknown parameters which define the measured spectrum:

$$\mathbf{x} = (\ \ln P,\ T,\ \ln \mathrm{VMR},\ C_i\tilde{\nu}^i,\ ...,\ C_n\tilde{\nu}^n\ ) \tag{5.1}$$

where C_i is the i^{th} order polynomial coefficient used to fit what is defined here as the *local background*, and $i = 0$ to n where n is the order of polynomial fit. In the retrieval process, the elements of the state vector are independently varied through an iterative procedure until differences between the measurement vector, $\mathbf{y}$, and the forward model, $\mathbf{F}(\mathbf{x})$, are minimised, i.e. a best fit is achieved (see Appendix for a description of the retrieval process). The measurement vector $\mathbf{y}$ consists of all the raw measured radiance values at each wavenumber point sampled by the spectrometer.[2] The forward model returns

---

[1] Fourier transform spectrometers are commonly used for TDL wavenumber calibration.

[2] Units of radiance are arbitrary since the beam is not radiometrically calibrated.

derivatives or Jacobians (weighting functions) relating the sensitivity of the forward model, $\mathbf{F}(\mathbf{x})$, to perturbations in each of the state vector elements. This forms the weighting matrix, $\mathbf{K}$, which is evaluated from small perturbations to each state vector element. The weighting matrix is used in the retrieval to direct the iterative variation of state vector elements towards the optimal solution.

Local background modelling. Transmission spectra cannot be calculated for gas lines as for the aerosol case, because of the presence of broad band aerosol absorption features which are not modelled in the gas retrieval. The approach adopted here is to calculate what is termed in this thesis as the *local background* by fitting the region between gas lines with an i^{th} order polynomial function. The ratio of the measured radiance spectrum against the fitted polynomial function effectively yields a gas transmission spectrum, since the function accounts for (but does not characterise) the aerosol absorption, the detector response (including electronics), spectral emission from the source, and absorption in the optical system. The process is also able to account for build up of contaminants on window surfaces, temperature dependence of window transmission, and icing on the liquid N_2 cooled MCT detector. In the forward model, the simulated gas transmission spectrum is multiplied by the polynomial function to synthesise the measured radiance spectra. Diagnostics returned from the retrieval show that the polynomial parameters were retrieved independently from gas absorption parameters, i.e., the gaps between the gas lines provide sufficient information to account for the local background. No errors due to channelling were observed above the spectral noise in the high resolution measurements.

Radiative transfer model. The radiative transfer model used to characterise gaseous absorption in the forward model is the Oxford Reference Forward Model (RFM), a general line-by-line FORTRAN code which has been developed over a number of years to provide a reliable and versatile tool for modelling spectra principally for atmospheric modelling (Dudhia, 1997). The large cell gas path is a straightforward case to model; the RFM is well-suited to calculations along a single homogeneous path. The gas retrieval program formulates the necessary inputs to the RFM including details of gas concentration, provision of the HITRAN-96 database (Rothman *et al.*, 1998) of spectral parameters, and other information such as resolution, spectral range and absorbing gas species. The Jacobians returned by the RFM are calculated using 1 % perturbations in P and VMR, which are retrieved in (natural) log space.[3] A perturbation of 1 K is used for T. All gases are modelled with a Voigt lineshape; i.e. pressure (collision) induced line-mixing is neglected. No significant differences resulted for a test case CO_2 retrieval which used sub-Lorentzian line wings.

The spectral range used for RFM calculations is wider than the measured spectral window by the width of the modelled instrument line shape (ILS) in the spectral domain (approximately 5 cm^{-1}). This avoids edge-effects when this function is convolved with the RFM output. An RFM wavenumber grid exactly 60 or 121 times finer than the measurement sample grid is used for 0.12 cm^{-1} and 0.4 cm^{-1} resolution spectra respectively; this permits accurate convolution of the ILS and the RFM spectrum, prior to degradation to the measurement grid.

Processing of FTS data. A modelled instrument lineshape (ILS) function was used to represent the multiplication of the instrument box-car ILS function (due to finite MPD) and an applied apodisation function in interferogram space (section 3.4.2). The apodisation function used for all the high resolution gas line spectra is of the Norton Beer Strong formulation (Norton and Beer, 1976). Blackman-Harris 3-term apodisation was used for low resolution aerosol spectra. The modelled ILS is calculated over a frequency width of approximately 5 cm^{-1} as this is an ample range to characterise the dampened side lobes in the model; in reality the function would be of infinite width. Figure 5.4 shows the modelled ILS for resolutions of 0.4 cm^{-1} (HNO_3 measurements) and 0.12 cm^{-1}; the unapodised sinc

---

[3] This avoids numerical precision errors since P and VMR can vary over many orders of magnitude in different situations.

function for the ILS at 0.12 cm^{-1} is also shown to demonstrate the effect of apodisation on reducing side lobes. The convolved simulated spectrum is calculated on the measurement sample grid, after the wavenumber shift induced by beam divergence (section 5.2.3) is taken into account.

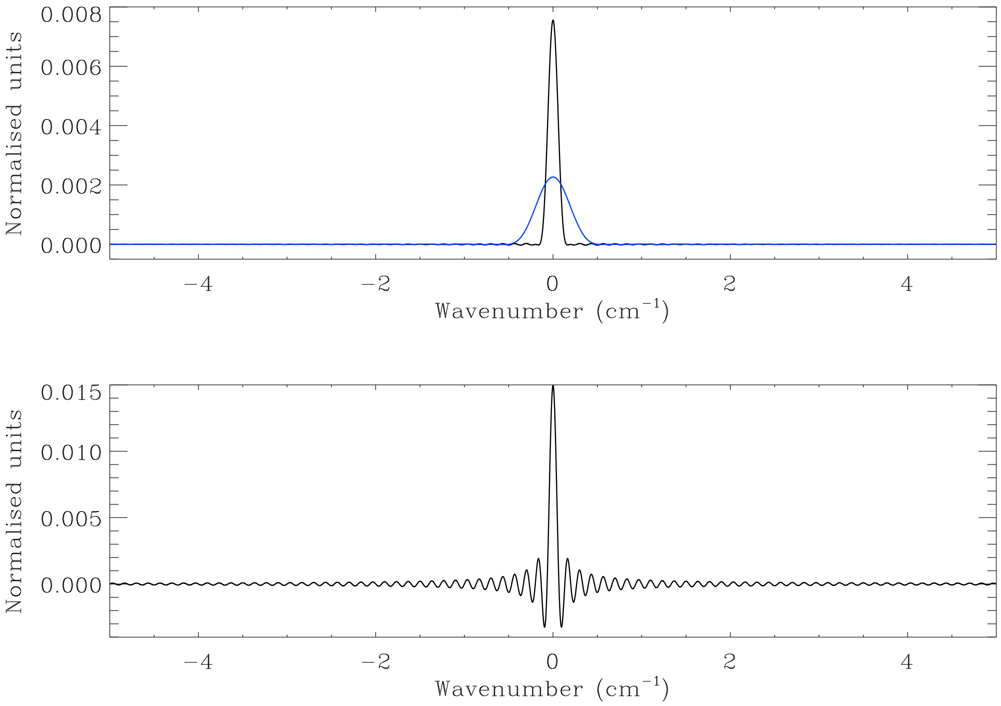

Figure 5.4: The top panel shows the modelled apodised ILS function in spectral space for a resolution of 0.12 cm^{-1} (black) and 0.4 cm^{-1} (blue). The unapodised sinc function for 0.12 cm^{-1} resolution is plotted in the lower panel.

5.2.2 Selection of spectral windows

The number of points in the measurement vector is limited by computational factors to a spectral width of less than about 200 cm^{-1} at 0.12 cm^{-1} resolution (a greater number of points slows processing considerably). The principal selection criteria for spectral windows are that:

- The **local background** has to be smooth enough to be characterised by a simple polynomial function over the entire spectral range of aerosol absorption, fortunately the case over most of the spectrum;

- The selected region is located to maximise **information content** of absorbing species. If gas absorption covers a small region there is no information advantage in modelling a larger spectral region. Regions of line saturation or very weak gas lines were avoided;
- The absorber should be represented over the selected spectral window in the HITRAN-96 **spectral database**. For example, HNO_3 is not represented above 1770 cm^{-1} (Rothman *et al.*, 1998).

It was found that spectral lines which in the convolved modelled spectrum appeared to have non-saturated transmissions of around 40 % or more, were sometimes saturated in the RFM spectrum. This effect was screened for in the retrieval model to avoid the use of saturated lines, which were found to introduce spurious results. Spectral windows were carefully selected to ensure measurement of H_2O lines in at least one window over the range of experimental conditions.

Figure 5.5 shows the spectral regions that were used for the retrieval of the main gaseous species against a plot of a high resolution 0.12 cm^{-1} MCT initial background indicating representative maximum detector signal (the H_2O gas lines result from out-gassing from condensates on the cell walls); colours used here form the convention throughout this thesis unless otherwise stated.

5.2.3 Determination of wavenumber shift

As described in section 3.4.3, the effect of beam divergence or self-apodisation on the spectrum is to shift the wavenumber grid of the measurements by $\delta\tilde{\nu}$. However, shifting of the wavenumber grid can also be introduced through phase correction on a single-sided interferogram. Hence, initially it was necessary to determine the wavenumber shift of the measured spectra independently. A wavenumber shift variable was included in the state vector and retrieved; the *a priori* and first guess shifts were set to zero. The wavenumber

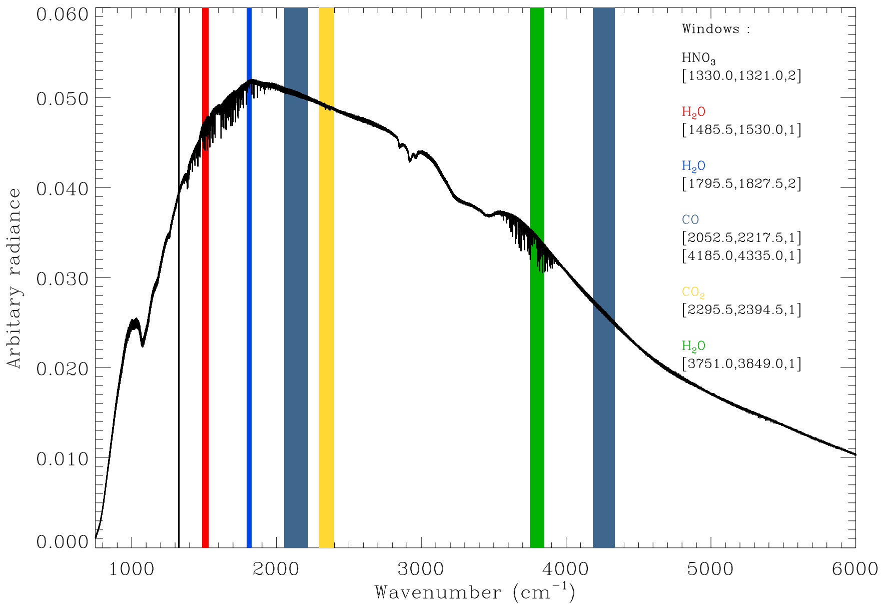

Figure 5.5: Spectral windows used for gas retrievals against a typical initial room temperature evacuated background. Annotation: [$\tilde{\nu}_1$, $\tilde{\nu}_2$, order of polynomial function].

shift was added to the RFM wavenumber limits to give an accurate representation of a sub-measurement resolution grid shift, and the RFM was executed again with a small wavenumber perturbation to produce the $\delta\tilde{\nu}$ Jacobian.

The retrieval results, see figure 5.6(a), revealed negative wavenumber shifts (measurements relative to RFM output) which increased with the wavenumber of the spectral window; retrieved shifts were consistent between separate spectra in the same spectral regions and between different gas species. Thus, database inaccuracies in HITRAN-96 can be ruled out as the cause of shifting. Accuracy of retrieved shift diminished in cases of low H_2O gas pressure, so a sample of measurements above a limiting threshold were used for comparison. Figure 5.6a shows wavenumber shifts for an averaged sample of H_2O, CO_2 and CO retrievals, with additional water retrievals above 5000 cm^{-1} from a representative single spectrum. It can be seen that the retrieved shifts match up very well with the theoretical shift due to beam divergence, $\delta\tilde{\nu} \;=\; -(A_s/4\pi f_c^2)\tilde{\nu}'$, where f_c is the collimator focal length

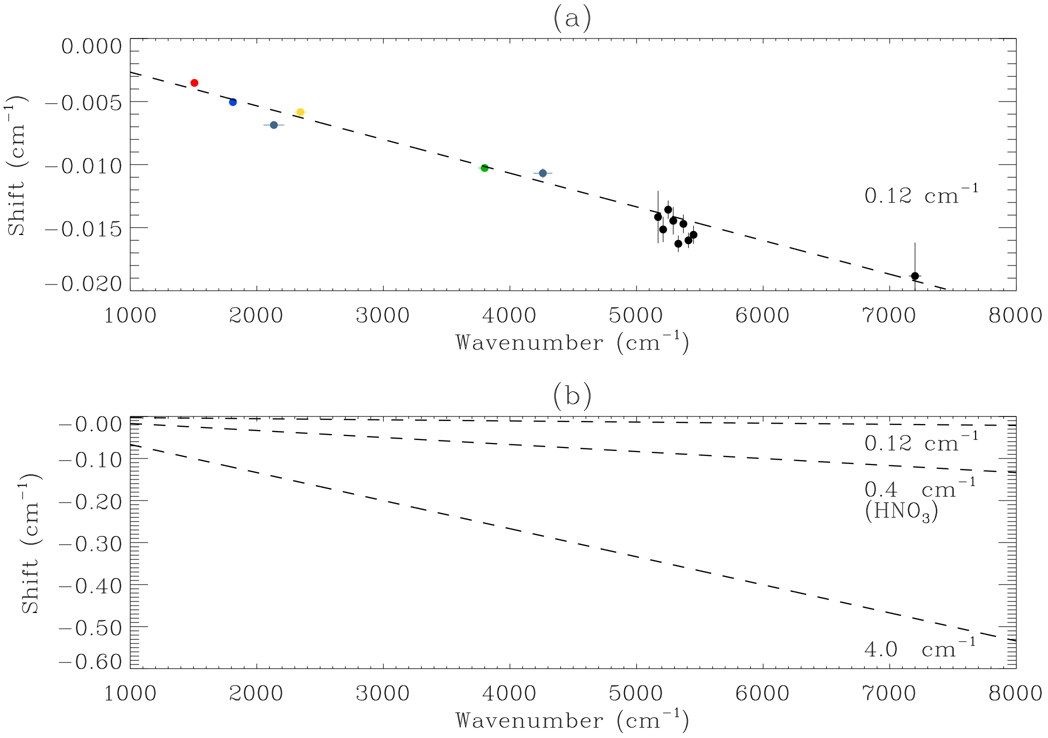

Figure 5.6: (a) Retrieved $\delta\tilde{\nu}$ shifts for averaged H_2O, CO_2, and CO (red/green/blue, yellow, dark blue), and calculated $\delta\tilde{\nu}$ line (dashed) due to beam divergence. (b) Calculated $\delta\tilde{\nu}$ at each resolution.

of 153 mm for the Bruker IFS 66v/S FTS. This is an important result, since although f_c remains fixed, the source aperture area, A_s, is increased at lower resolution resulting in a greater wavenumber shift, as illustrated in figure 5.6(b). For example, low resolution (4 cm^{-1}) aerosol measurements exhibit a shift of around -0.5 cm^{-1} at 7000 cm^{-1}. In all subsequent retrievals presented in this thesis, this effect was screened in the model by the direct application of the theoretical expression for $\delta\tilde{\nu}$.

5.2.4 Error characterisation in high resolution spectra

The main types of noise inherent with Fourier transform spectrometers were introduced previously in section 3.5.5. The approach taken here assumes that absolute (additive Gaussian) thermal/detector noise dominates over relative photon noise. Essentially, when the forward model error is negligible compared with the measurement error, the measured radiance $y(\tilde{\nu}) = \mathbf{F}(\tilde{\nu}) + \epsilon$, where $\mathbf{F}(\tilde{\nu})$ is the forward model and ϵ the error term, so the error can be ascertained from the standard deviation of the residual $\epsilon = \mathbf{F}(\tilde{\nu}) - y(\tilde{\nu})$. The retrieval is thus executed twice; initially with an approximate absolute error term, then with an absolute error estimated from the residual of the first fit. This **double-pass noise estimation** improves the fit and yields accurate retrieval errors in other parameters. Although detector noise is wavenumber dependent, use of a single noise value in this case is justified since the spectral windows are relatively narrow. Where radiance varies in magnitude across a spectral window, comparison of the residual standard deviation at high and low signal showed only small variations.

It is important to recognise that point-to-point noise variation will not be entirely random when using a FTS. Through convolution of apodisation function and ILS, zero filling of the interferogram, and phase correction, a certain amount of correlation or cross talk between neighbouring points on the wavenumber scale occurs. Hence, the $\sqrt{N}$ dependence is not strictly true, but this is not of major concern since the relation is not used to calculate noise from co-added scans. Use of local backgrounds in the gas retrieval renders spectra more susceptible to channelling. However, since little evidence for channelling can be seen in most high resolution spectra, no systematic errors are included in the gas retrieval. The error derived in this way may also incorporate systematic errors arising from inaccuracies in fitting, e.g. spectroscopic errors in the HITRAN database.

5.2.5 Implementation

The gas retrieval program was used to retrieve gas parameters for H_2O, HNO_3, CO_2 and CO over regions indicated previously in figure 5.5. The applicable spectrometer parameters were carefully determined from the FTS experiment log files. The *a priori* state parameters were set according to the species being retrieved, see table 5.1. *A priori* state pressure was set with a relatively large uncertainty of 20 % due to concerns over the accuracy of the pressure probe, which may have been affected by corrosion from acid. Polynomial coefficients for the background fit also had large *a priori* uncertainties and were assigned a value of unity for the constant term with zero for the other coefficients. First guess values in the state vector were automatically set to *a priori* values, except for the polynomial coefficients, which were calculated from a fit to outlier points at the ends of the measured spectral window. Table 5.1 lists the *a priori*/first guess values used in the retrieval program. Limits were also placed on some state vector variables in order to prevent the retrieval from converging on a solution at unrealistic or non-physical values; i.e. $0.001 < P < 1000$ mb, $0.001 <$ VMR $<10^6$ ppmv, and $100 < T < 350$ K.

Table 5.1: Gas retrieval setup.

Variable		ln P (mb)	T (K)	ln VMR (ppmv)	Polynomial coefficients		
					C_0	C_1	C_2
◇*A priori*	$\mathbf{x}_\mathrm{a}$	⊙measured	220	⊔variable	1.0	0	0
State vector variance	$\sqrt{\mathbf{S}_a}$	50 %	100	10^5 %	1.0	1.0	1.0

◇ Apart from polynomial coefficients, these values are used as first guess state vector (x_0).
⊙ Measured value (from driver table).
⊔ Values used: 100 ppmv for HNO_3/CO_2, 20000 ppmv for H_2O, 30000 ppmv for CO.

The structure of the gas retrieval, an IDL program *ret_gas*, is represented by the flow diagram in figure 5.7. Numbers in square brackets represent IDL functions and procedures called by this program as assigned in figure 5.8, which illustrates modules called from the forward model. The forward model operates as explained in section 5.2.1; the relevant parameters are written to the RFM input files [F1, F2], the RFM executed and the resulting high resolution spectrum read-in, wavenumber shifted, convolved with the ILS, and spectrally degraded to the measurement resolution.

5.2.6 Validation

Extensive validation of the gas retrieval program was conducted in various steps:

- **Validation against synthetic spectra.** Forward model simulated spectra were used as input for the retrieval to ensure consistency;
- **Retrieval diagnostics.** Diagnostic matrices output from the retrieval, such as corre-

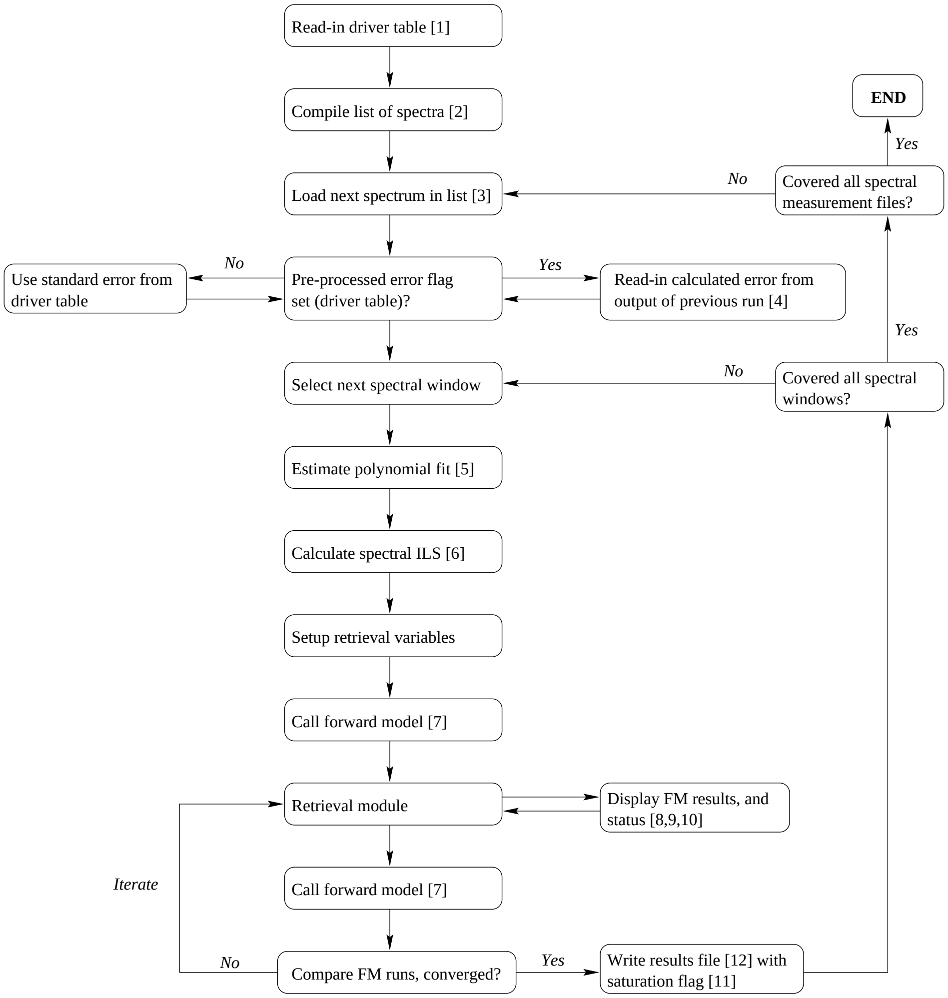

Figure 5.7: Main elements of the gas retrieval, *ret_gas*. Numbers in square brackets refer to program modules, see figure 5.8.

lation matrices and averaging kernels, were checked to ensure correct implementation and the independence of the retrieved variables;

- **Self-consistency.** The retrieval of H_2O parameters from different spectral windows in the same spectrum provided a measure of the retrieval consistency with real spectra;
- **Calibration from retrieved values.** Retrieved H_2O partial pressure values were compared with expected values from calibration experiments.

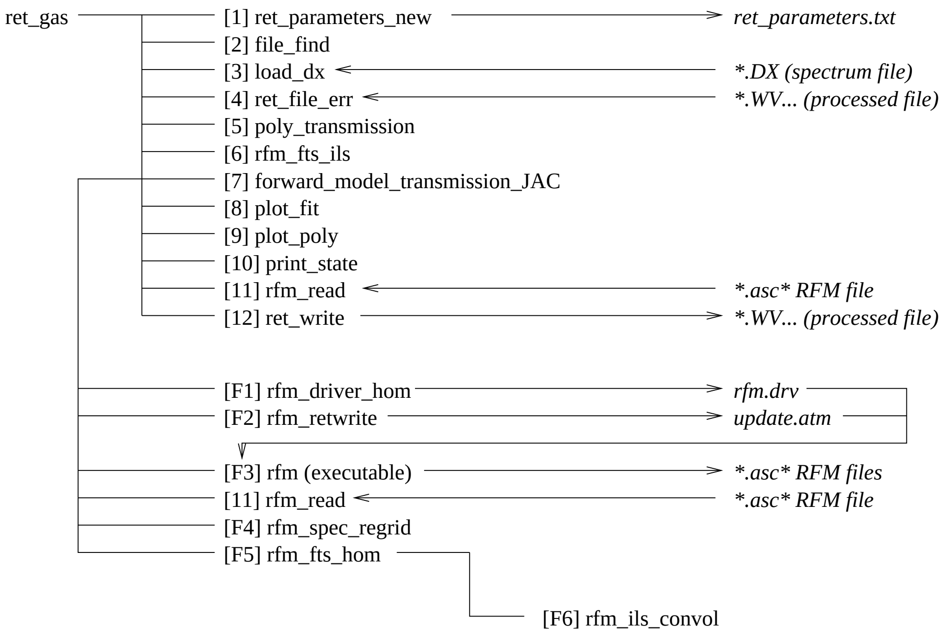

Figure 5.8: Main IDL functions and procedures of the gas retrieval (modules numbered in square brackets). Arrows link input and output files (in italics).

Validation against synthetic spectra. A fundamental method for testing a retrieval is to produce a simulated spectrum with the forward model and then input the spectrum into the retrieval. The retrieved parameters should match those used to produce the synthetic spectrum to within the retrieval error bars. For the test, a typical polynomial background was extracted from a measurement retrieval and used to produce the synthetic spectrum. A retrieval was then performed. Using the simulated parameters as first guess in the retrieval produced zero-residuals, as expected.[4] In addition, starting the retrieval from the standard *a priori*/first guess values resulted in rapid convergence to correct values. Next, typical levels of normally-distributed (Gaussian) additive random noise of 2×10^{-4} (units of uncalibrated radiance) were included in synthetic H_2O spectra over each spectral window. The converged retrieval fits over each spectral region can be seen in figure 5.9; at convergence each retrieval had the expected residual standard deviation of approximately 2×10^{-4}. It

---

[4] Slight residuals at the 10^{-9} level were probably due to double precision errors; this is well below the measurement noise level of $\sim 10^{-4}$ (uncalibrated radiance).

can be seen that the retrieved gas parameters agree within the retrieval uncertainties – this was also true for the polynomial coefficients (not shown). This same analysis was conducted for other species, validating the gas retrieval.

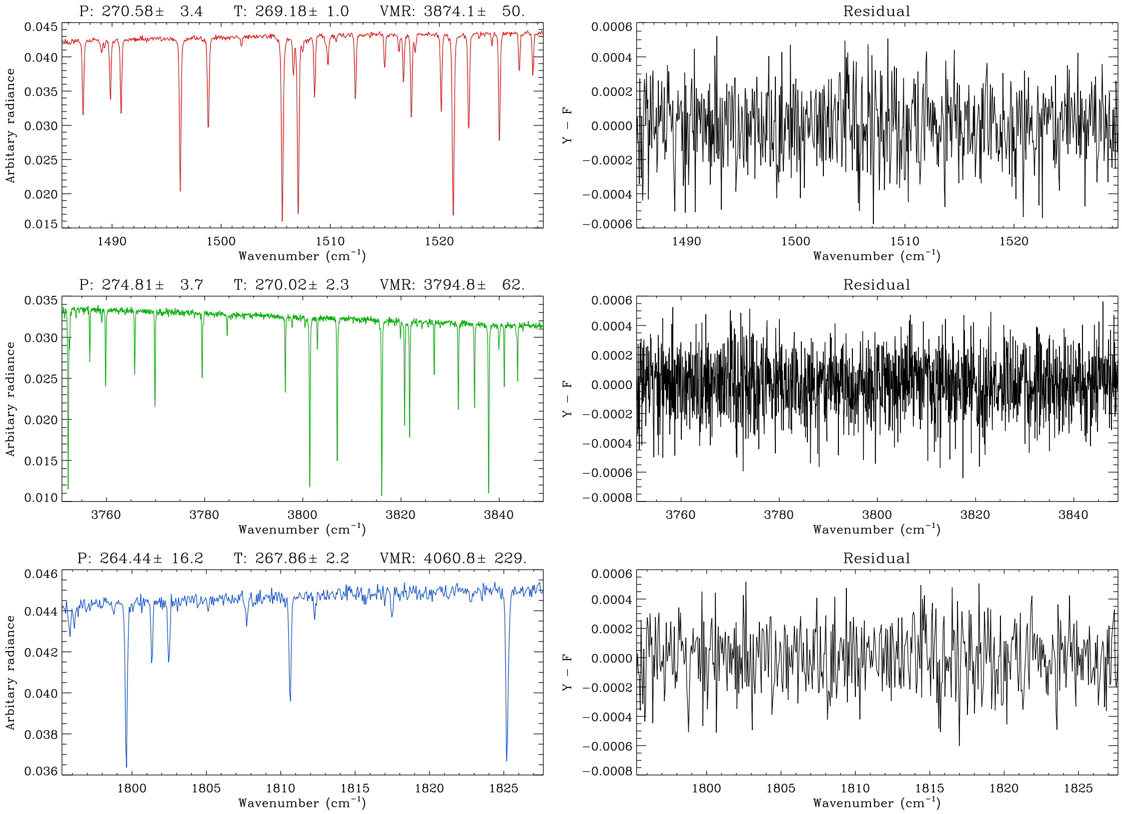

Figure 5.9: Converged H_2O retrieval of a synthetic spectrum over each spectral window. The synthetic spectrum has $P = 270.8$ mb, $T = 270.0$ K, VMR $= 3880.0$ ppmv, and $\sigma = 2{\times}10^{-4}$. The left panels show the synthetic spectrum over each spectral window (colour coded) with retrieved parameters and uncertainties (units as above), with the corresponding fit residual to the right.

Retrieval diagnostics. The output from the retrieval of synthetic spectra provides a good test to ascertain the robustness of the retrieval method. The correlation matrix is the retrieval error covariance scaled by its diagonal elements, essentially $\mathbf{C}(i,j) = \hat{\mathbf{S}}(i,j)/[\hat{\mathbf{S}}(i,i) \times \hat{\mathbf{S}}(j,j)]^{\frac{1}{2}}$, where $\hat{\mathbf{S}}$ is the retrieval covariance matrix. For the first spectral window in figure 5.9 the correlation matrix is:

$$\mathbf{C} = \begin{pmatrix} 1.0000 & 0.4795 & -0.9443 & 0.1009 & 0.1007 & -0.1004 \\ 0.4795 & 1.0000 & -0.2668 & -0.0406 & 0.0403 & -0.0400 \\ -0.9443 & -0.2668 & 1.0000 & 0.0891 & -0.0899 & 0.0888 \\ -0.1009 & -0.0406 & 0.0891 & 1.0000 & -1.0000 & 1.0000 \\ 0.1007 & 0.0403 & -0.0890 & -1.0000 & 1.0000 & -1.0000 \\ -0.1004 & -0.0400 & 0.0889 & 1.0000 & -1.0000 & 1.0000 \end{pmatrix} \tag{5.2}$$

with elements ordered as in equation 5.1, i.e. ($\ln P$, T, $\ln$ VMR, $C_i\tilde{\nu}^i$, ..., $C_n\tilde{\nu}^n$). The correlation matrix is a measure of the normalised sensitivity of the covariance of each state parameter to all parameters, hence the unity values in the diagonal. It can be seen that the errors in P and VMR are strongly anti-correlated with each other, and that there is some correlation between the P and T uncertainties. Moreover, there are strong correlations between the uncertainties in the polynomial coefficients; this is not of major concern since the errors in these parameters have no further impact in the analysis. The dependence of the gas parameter errors indicate that there is some degree of *smearing* of errors in parameter space between these elements. If errors are combined to produce the error in partial pressure, anti-correlation between these uncertainties act to reduce partial pressure error considerably from quadrature error. Of some concern was that retrieved pressure varied considerably in spectra taken on the same days, where there is a constant measured pressure. Fixing the pressure in control-runs illustrated that the retrieved gas VMR then changes to compensate, yielding the same partial pressure compared with the free-retrieval, demonstrating the robustness of partial pressure retrieval in all situations.

The averaging kernel provides some indication of the independent retrieval of parameters. The averaging kernel matrix is a measure of the sensitivity of the retrieved state to the true state $\mathbf{A} = \partial\hat{\mathrm{x}}/\partial\mathrm{x}$. In the case of the synthetic H_2O retrievals the kernels are essentially delta functions in the diagonal, indicating that the values of retrieved parameters do not show pronounced inter-dependence. This is important, since it confirms **the independent retrieval of parameters**. As an example the averaging kernel, with off-diagonals scaled to the *a posteriori* (retrieved) covariance, from the first spectral window in the synthetic

H_2O spectrum fit (figure 5.9) is:

$$\mathbf{A} = \begin{pmatrix} 0.99 & -3.99\times10^{-5} & -2.84\times10^{-7} & 1.33\times10^{-3} & -1.49\times10^{-4} & -7.45\times10^{-5} \\ -3.04\times10^{-4} & 1.00 & 1.52\times10^{-7} & 6.44\times10^{-4} & 1.56\times10^{-4} & 7.82\times10^{-5} \\ 6.01\times10^{-4} & 2.53\times10^{-5} & 1.00 & -1.17\times10^{-3} & 1.45\times10^{-4} & 7.28\times10^{-5} \\ 6.55\times10^{-5} & 7.89\times10^{-6} & -1.94\times10^{-8} & 0.99 & -7.77\times10^{-6} & -2.75\times10^{-6} \\ -6.53\times10^{-5} & -7.87\times10^{-6} & 1.93\times10^{-8} & 1.39\times10^{-2} & 1.00 & 2.71\times10^{-6} \\ 6.51\times10^{-5} & 7.85\times10^{-6} & -1.92\times10^{-8} & -1.39\times10^{-2} & -7.63\times10^{-6} & 1.00 \end{pmatrix} \tag{5.3}$$

where the small off-diagonals indicate the independence of the retrieved parameters. The highest gas parameter off-diagonal is the $\partial\mathrm{VMR}/\partial P$ term, indicating poor distinction between P and VMR.

Self consistency. Retrievals over the three H_2O spectral windows on measured spectra were analysed to ascertain consistency of the operational retrieval. Figure 5.10 shows a single typical example (with unsaturated lines in all windows). The partial pressure, the key parameter of interest, is seen to be consistent between retrieved spectral windows.

Retrieved temperature varies by over 4 K between windows, but the retrieved error in all cases is of order 1 K. Where H_2O vapour concentration is lower (at lower temperatures) this disparity becomes even more obvious (as weaker H_2O lines are lost in the noise). Figure 5.11 illustrates this; retrieved temperatures and uncertainties from all high resolution (0.12 cm^{-1}) spectra from a single day are plotted, along with partial pressures. The cause of this problem is likely due to poor characterisation of the temperature dependence of H_2O vapour in the HITRAN-96 spectral database. Uncertainties are not included in the edition of the HITRAN spectral database used here, thus limitations in knowledge of spectral parameters are not conveyed in the uncertainties output from the retrieval, which should therefore be considered as a lower limit. It was found that the temperature retrieval with CO and CO_2 was much more consistent, permitting use of these species for temperature calibration (section 5.3.2).

Calibration from retrieved values. Retrieved partial pressures of H_2O during two test experiments are used here to further validate the gas retrieval, since expected pressures were

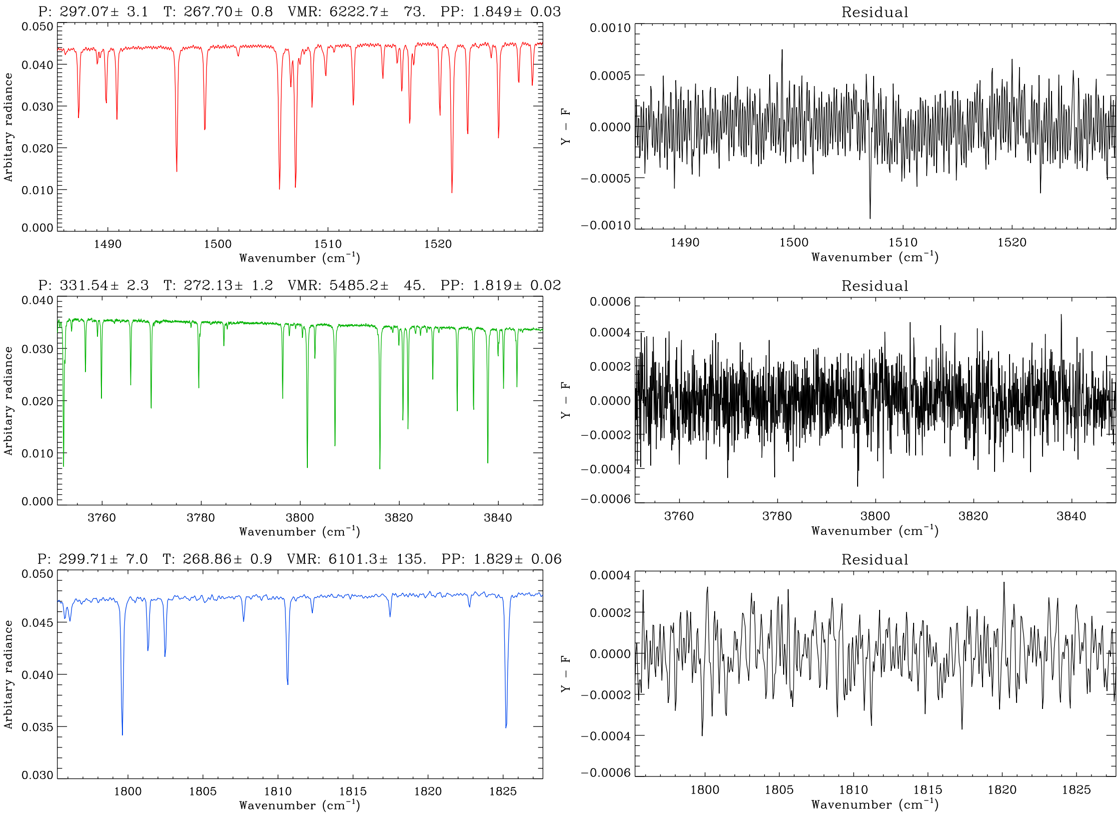

Figure 5.10: Converged H_2O retrieval of a measured spectrum (28/11/00) over each spectral window. The left panels show the measured spectrum (colour coded) with retrieved parameters, uncertainties and partial pressure (PP); units are [mb], [K], [ppmv], and [mb] respectively. Associated residual is displayed to the right.

retrieved. The first test was conducted by connecting the water bath directly to the large aerosol cell, setting the N_2 flow over the bath to a flow rate of 100 sccm, and achieving a steady-state cell pressure of approximately 50 Torr. The retrieved H_2O partial pressure was 21.03 ± 0.76 mb, corresponding to the saturation vapour pressure (SVP) of water at 291.5 ± 0.6 K. The measured calibrated temperature was 291.2 ± 2.1 K inside the cell and the laboratory temperature near the bath was 293.8 K; this corresponds to an SVP of 24.3 mb at bath temperature, and 20.7 ± 2.8 mb at calibrated cell temperature. Since H_2O partial pressure can only decrease through condensation after the bath, the retrieved H_2O partial pressure lies well within the error bars for the expected SVP. The retrieval of an

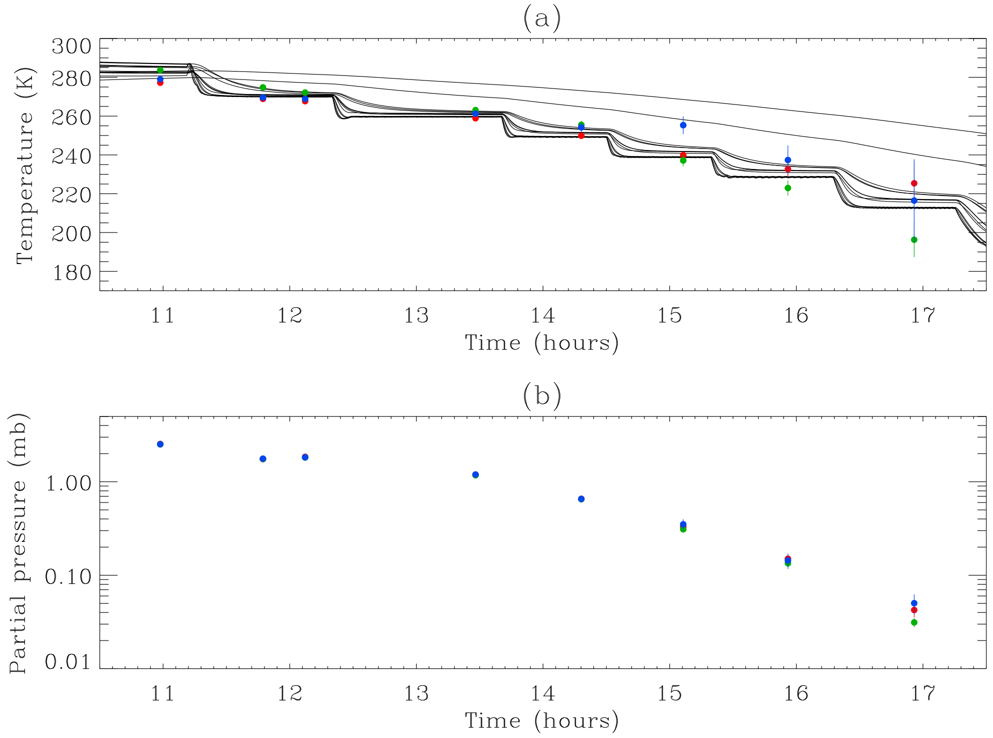

Figure 5.11: H_2O vapour gas retrievals from 28/11/00. Panel (a) shows retrieved temperature from each retrieval; also plotted are the PRT temperatures for comparison. The retrieved partial pressure is shown in panel (b). The standard colour scheme is used for each spectral window, and error bars, although not visible in many cases, are plotted.

expected SVP pressure is a further validation of the gas retrieval. With the flow rate over the water bath increased to 400 sccm (the maximum possible rate is 500 sccm) the retrieved H_2O vapour value fell to 16.2 ± 0.73 mb. Later experiments using a similar setup conducted at RAL (R. A. McPheat, personal communication), revealed that the evaporation in baths could result in significant cooling which becomes more pronounced at higher flow rates, explaining the reduction of SVP pressures observed at higher flow rates.

The second test was performed through retrieval of the water vapour component from measurements of pure sulphate aerosol flowing into the large cell at room temperature. A thermodynamic model based on that of Steele and Hamill (1981) was used to calculate the expected composition of sulphuric acid, using the retrieved H_2O vapour pressure. The sulphuric acid bath used was of 74.4wt% concentration, as for the main experiments. The calculated sulphuric acid composition was 77.0 ± 0.35wt% using CO_2 retrieved temperature and error combined with the H_2O vapour pressure error. The fact that the calculated droplet composition is slightly higher than the bath is probably due to mixing errors resulting from the manual bath preparation process; this could be a significant source of systematic uncertainty for this calibration. Since this difference is small, these results further validate the gas retrieval.

5.2.7 Summary

The main achievements of the gas retrieval can be summarised:

- The forward model has been extensively validated and found to provide an excellent representation of measured spectra to within measurement noise;
- Partial pressures of H_2O and HNO_3 are retrieved to an accuracy typically of better than 4 % in each spectral window – retrievals between the different H_2O spectral windows are combined to reduce uncertainties further;
- The position of line centres permitted characterisation of the wavelength shift due to beam divergence in the FTS.

The main limitation of this model is that the HITRAN-96 spectral database does not contain uncertainties; hence the uncertainties returned from the gas retrieval represent a lower-limit to the total error.

5.3 Temperature monitoring and calibration

Knowledge of the temperature of the aerosol/gas in the cell is of utmost importance in the analysis of particle evolution, composition and determination of thermodynamic properties (i.e. freezing points). The temperature monitoring system is described here in order to complement the description of temperature calibration. The system as used in the main experiments consisted of 14 platinum resistance thermometers (PRTs; Fisher Rosemount, PT-100 grade B) mounted external to, but in thermal contact with the body of the inner cell. With current techniques, PRTs cannot be mounted inside the cell since the solder joints are attacked by acid. During the settling experiments temperatures were monitored from four selected PRTs manually during experiments. By the time of the main experiments, an automated PRT logging system had been installed. Calibration of the PRT readings was undertaken from analysis of the main experiments. The gas retrieval was key to the temperature calibration, providing an independent assessment of the temperature along the beam path through retrievals of CO_2 gas lines in measured spectra.

5.3.1 Temperature monitoring system

The PRT numbering and spatial placement is shown schematically in figure 5.12; PRTs 11,12 and 13 are mounted inside sealed tubes suspended from the top plate of the cell. The operation of PRTs is based on the temperature-dependant variation of resistance. In both the manual and automated logging of PRTs, resistance was indirectly determined by passing a constant current through each PRT wired in series (two wires) and monitoring the voltage across each PRT (two wires). This four-wire approach has the distinct advantage of eliminating the resistance of the connecting wires. As long as the constant current is accurately known, it is straightforward to calculate PRT resistances from the voltage measurements using Ohm's Law. All four wires from each PRT were connected via electrical vacuum *feed-throughs* to the exterior wiring to permit PRTs to be removed from the circuit in case of malfunction. All external wiring and connections were shielded to minimise electromagnetic interference.

During the settling experiments readings of four PRTs were manually logged during each FTS measurement with a switching box and a digital multi-meter; a stable 1 mA constant current source was used. Constant current was monitored to an accuracy of ± 0.001 mA and the voltage to ± 0.01 mV. Thus, the main source of error in the temperature analysis is the uncertainty of the constant current which at 190 K translates to ± 0.2 K, and the sparse coverage of the cell temperature with only four PRTs. All temperatures quoted for settling experiments are calculated from an average of the monitored PRTs, which were distributed across the cell to give a representative mean temperature. At a stable 190 K cell temperature, the individual PRTs gave temperature values that varied by up to 5 K across the cell, and hence an estimated temperature uncertainty of ± 5 K is adopted for the settling experiments.

During the main experiments a new proprietary (National Instruments Corporation) temperature logging system was installed to automatically log PRTs. Voltage measurements were digitised through a 24 bit ADC card in a dedicated PC system, giving temperature values of accuracy ± 0.1 K. Electrical filtering for 50 Hz noise was automatically applied, and the calculated temperature reading from all 14 PRTs logged to file with a time-stamp at 5 second intervals.

Figure 5.12 illustrates the wide distribution of temperatures recorded by the PRTs throughout an experiment. It can be appreciated at this stage that it is not possible to ascertain the temperature of the gas and aerosol inside the cell from these data without calibration. According to specifications, individual PRTs have an accuracy of better than ± 0.5 K with comparatively negligible systematic variation between units. Hence, the variations in temperature between PRTs reflect the real distribution of temperature on the outside of the cell, and in the internal probe tubes (PRTs 11,12,13). A real-time plot of PRT temperature data was displayed on a dedicated logging computer, which was utilised during experiments to help determine when the temperature had stabilised in conjunction with checking stability of successive aerosol spectra. It can be seen that the coldest group of PRTs stabilise relatively rapidly during each cooling phase, and that the two other groups of PSCs take longer to cool, and approach stability at higher temperatures. Being relatively thermally isolated from the cooling coils, the bottom and top plates are at considerably higher temperatures.

Aerosol spectra still varied once the coldest group of PRTs had stabilised, indicating that the temperature of the aerosol/gas is not directly linked to these PRT readings.

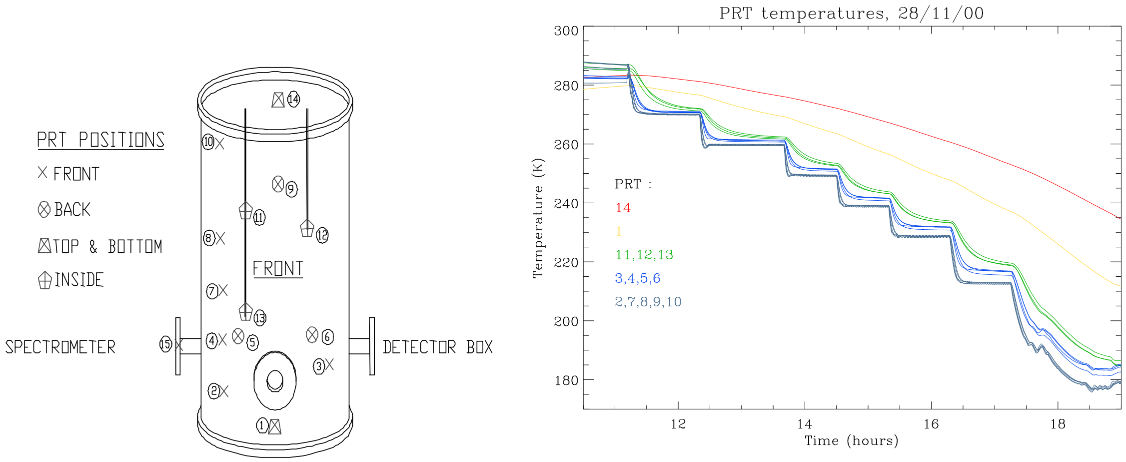

Figure 5.12: Final large cell PRT numbering and locations (RAL) (left), and PRT temperature distribution on a typical experimental day (right).

Theoretically, the mean temperature of the aerosol/gas along the FTS beam path should lie somewhere between the coldest readings and the internal probe tube PRT temperatures (the green group in figure 5.12). Since they are thermally isolated from the cooling system by the much warmer top plate (PRT 14), the internal PRTs can only be cooled to lower temperatures by heat conduction from the aerosol/gas inside the cell (radiation from the cell walls would play a smaller role), with some steady-state temperature being attained that is higher than the cell contents. However, the mean internal temperature of the aerosol/gas along the FTS beam path would depend on a complex interplay of parameters, including the temperature and flow rate of the incoming stream from the laboratory, the internal pressure, the temperature distribution across the cell, and cooling history and turbulence within the cell (it is assumed that the wide temperature distribution results in considerable turbulent mixing, facilitating the stabilisation of aerosol and temperature). Due to the limited information available, it was not considered feasible to model the resulting temperature along the beam path from PRT temperatures alone. Only retrieved gas temperatures provide the necessary calibration of PRT measurements.

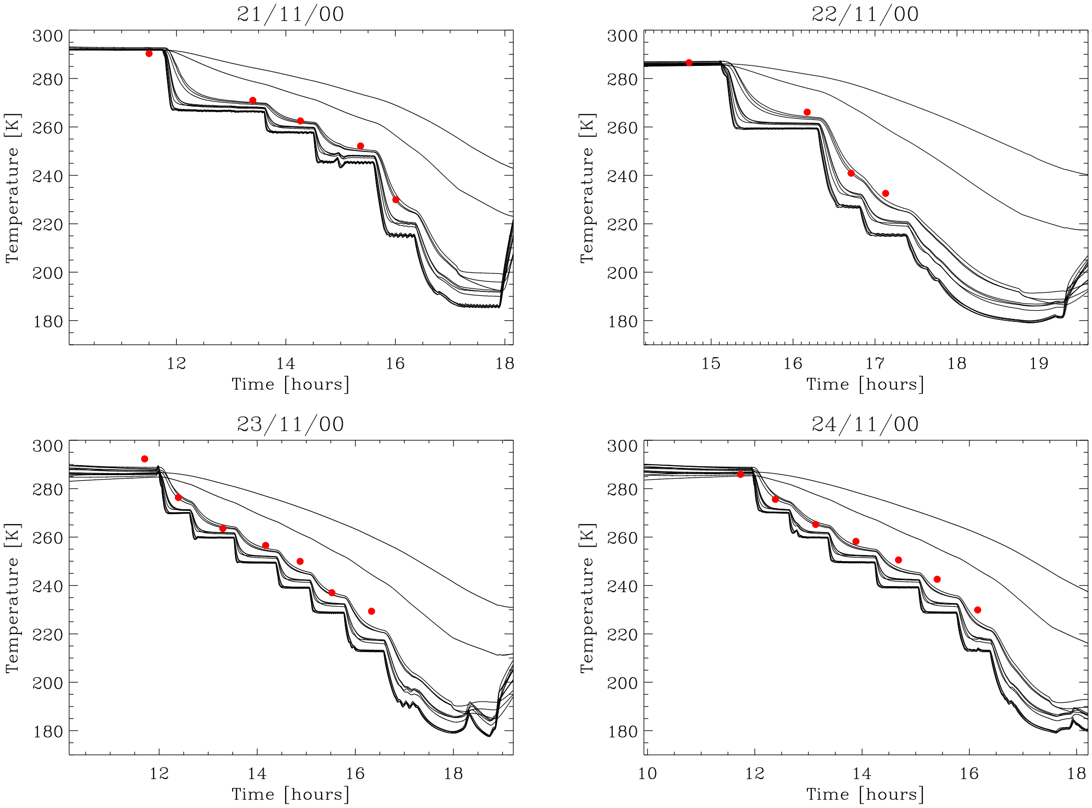

Figure 5.13: Retrieved CO_2 gas temperatures (red, error bars are not visible on this scale) and range of PRT temperatures on days of CO_2 contamination.

5.3.2 Calibration

Good information on temperature could be retrieved from two sets of high resolution data:

1. Four days of spectra contaminated by CO_2;
2. A low temperature measurement of CO in the cell.

As explained section 5.2.6, H_2O and HNO_3 gas measurements could not be employed for temperature measurement due to systematic errors. Contamination with CO_2, indicated by a CO_2 absorption band around 2350 cm^{-1} in the recorded spectra, was measured on four days (21/11/00 to 24/11/00) during the main experiments. Since the temperature dependence of CO_2 is well characterised in HITRAN-96, this contamination enabled the retrieval of mean gas temperatures along the FTS beam path with a standard deviation of less than 1 K, from high resolution (0.12 cm^{-1}) spectra. Figure 5.13 shows the retrieved CO_2 temperatures plotted against PRT measurements for each day – the retrieved temperatures lie close to the temperatures of the internal PRTs, or slightly above; unexpected since the internal PRTs are cooled by only the gas and should be warmer. However, radiative cooling to the colder walls may reduce PRT readings, or more likely the gas along the FTS beam path, being close to the base of the cell, is marginally warmed by the base plate and the warmer viewport recesses. The retrieved CO_2 gas temperatures were also compared with each of the internal PRTs and the sum of these PRTs. It was found that, for the combination of all four days, the gas temperatures most closely match PRT 13, which can be understood since this is the closest internal PRT to the viewports.

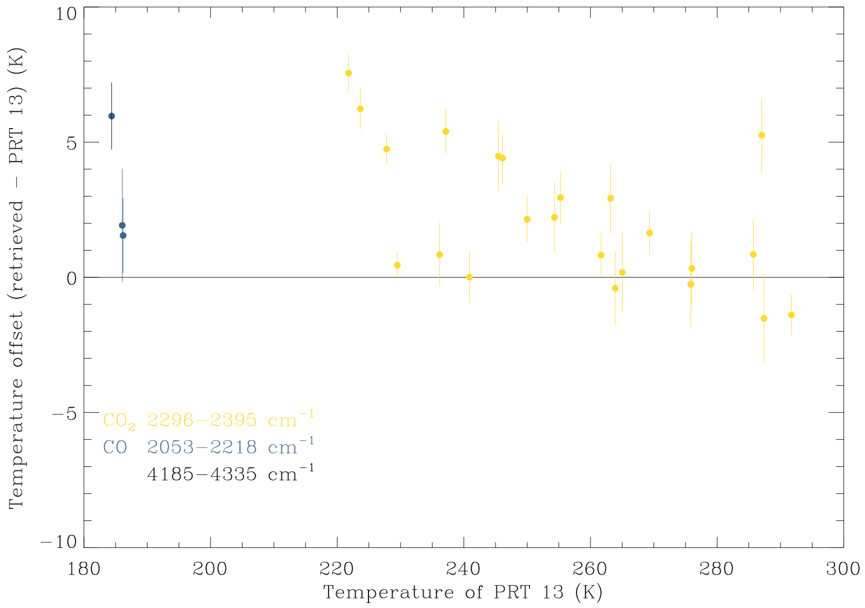

Figure 5.14: Retrieved CO and CO_2 gas temperatures for all measurement days relative to PRT 13. The standard spectral window colour scheme is used here.

A plot of all the retrieved CO_2 and CO temperatures against PRT 13 can be seen in figure 5.14. Although the CO temperatures show good agreement with PRT 13, these values were not included in a calibration due to the non-standard nature of the CO test experiment. The optimal calibration of PRT 13 to the gas temperature along the FTS beam path was obtained by using the temperatures determined from the CO_2 gas retrievals. A distinct trend can be seen in the CO_2 temperatures, relative to the temperature of PRT 13, in figure 5.14. This trend is probably due to the increasing temperature gradients across the cell at lower temperatures, and can be best fitted by a weighted straight line through the CO_2 temperature offsets, such that $T_{path} = a + b(\text{PRT 13}) \pm \sigma$. Figure 5.15(a) illustrates the best fit line through the offsets, and 5.15(b) shows the straight line calibration plotted with the retrieved CO_2 temperatures and uncertainties, which demonstrates that the CO_2 temperatures show a random distribution about the line. However, since the straight line fit is only an approximation to the mean gas temperature along the FTS path (the CO_2 temperatures), the error of the straight line fit is **not** an indication of the error in real temperature. The definitive estimate of uncertainty is the distribution of retrieved temperatures about the line, such that $\sigma = 2.13$ K for all temperatures quoted for the main experiments of this thesis (unless otherwise stated).

5.3.3 Summary

It is fortunate that contamination by CO_2 gas permitted independent temperature retrieval on several experiment days, thereby enabling temperature calibration of ± 2.13 K.

New methods to improve temperature characterisation in future work include:

- Mounting PRTs **inside** the inner cell, preferably close to the FTS beam path. Following calibration, these PRT should yield a much better estimate of real gas temperature. New shielding technology would be necessary for PRTs to be placed in this highly acidic environment;
- Use of a temperature marker, such as a small quantity of CO_2 or CO in the flow, to allow accurate temperature retrieval.

Carbon monoxide is the preferable choice as a temperature marker, since CO_2 may freeze at

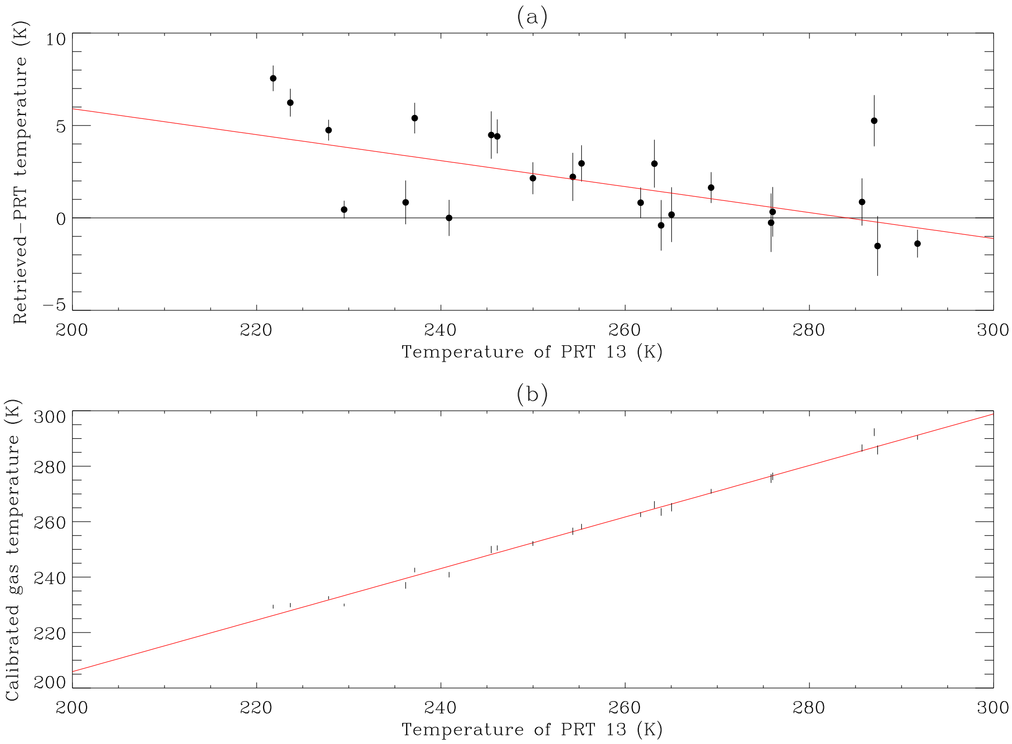

Figure 5.15: CO_2 temperatures retrieved from 21/11/00 to 24/11/00. The temperature offset from PRT 13 can be seen in panel (a), with the weighted straight line fit (red) through the retrieved temperatures. Panel (b) shows the fit through the CO_2 temperatures with associated errors.

temperatures lower than approximately 217 K. The main problem here is that the marker would only be of use in high resolution measurements, not the majority of low resolution aerosol spectra. Use of another spectrometer (FTS, TDL, etc.) could allow simultaneous high resolution measurements to track temperature.

5.4 Error characterisation in low resolution spectra

The gas retrievals from high resolution (0.12/0.4 cm^{-1}) spectra were performed over relatively small spectral windows, where it could be justifiably assumed that the measurement uncertainty remains constant and that the forward model is a very good representation of reality; this enabled the measurement of uncertainty from the residual between the forward model fit and the measured spectrum for high resolution measurements. However, a much wider spectral range is covered in the band model retrieval of refractive indices from low resolution (4 cm^{-1}) spectra, and uncertainties increase considerably at the margins of spectral coverage. Thus, an independent detailed assessment of uncertainties was necessary for low resolution aerosol extinction spectra, as described in this section.

5.4.1 Wavenumber-dependant uncertainties in radiance spectra

The first stage in the determination of errors in low resolution extinction spectra was the assessment of random noise in raw radiance spectra. Although each low resolution spectrum consisted of 100 co-added spectrometer scans, from which it would be straightforward to determine the variance at every wavenumber point, this was not possible due to limitations of the hardware and software employed. Instead, smooth sections of raw radiance spectra were identified and fit with polynomial functions to calculate the standard deviation of points about the function as an estimate of noise. This method was used to assess random noise from 0 to 10000 cm^{-1}.

As previously described (section 5.2.6), there are two main categories of spectral noise, absolute (additive – thermal/amplifier noise) and relative (multiplicative – photon statistics) noise. These can be distinguished where we have two radiance spectra of different strengths, i.e. a background and measurement spectrum, from:

$$\begin{aligned} \epsilon_{\text{background}} &= \epsilon_{\text{absolute}} + (\epsilon_{\text{relative}} \times I_{\text{background}}) \\ \epsilon_{\text{measurement}} &= \epsilon_{\text{absolute}} + (\epsilon_{\text{relative}} \times I_{\text{measurement}}) \end{aligned}$$

where ϵ indicates the relevant error terms, and I the measured radiance spectrum. The calculated noise over identical spectral regions was found to vary little between spectra of varying radiance intensity, such that the calculated relative noise was of order 10^{-5}. Thus, the main component of noise at **all radiance intensities**, down to zero signal level, was absolute (additive Gaussian) in nature and independent of radiance intensity.

Since the relative component of noise was negligible compared with absolute noise, an accurate estimate of noise was obtained from background spectra only, measured at the start of each day of measurements in the main experiments. It was found that noise varied little between days. Thus, the mean of noise estimates for all backgrounds from the main experiments was used to evaluate wavenumber-dependant noise in radiance spectra across the entire spectral range. This parameterisation was used for the weighted polynomial fitting of the variation of BaF_2 window transmission with temperature, and the evaluation of wavenumber-dependant uncertainty in low resolution extinction spectra.

5.4.2 Temperature dependence of barium fluoride windows

It was clear by the main experiments that a component of the optical system exhibits significant temperature dependence since it was observed that extinction would become very low (even negative) near to the infrared window cut-off around 750 cm^{-1} towards the end of experiments, after the cell had been cooled from the temperature of the initial backgrounds. Comparison of initial (room temperature) and final (cold) background spectra revealed a considerable increase in transmission near the 750 cm^{-1} cut-off with reducing temperature. Figure 5.16 shows an example of initial and final low resolution DLaTGS spectra where transmission increases by over 50 % around 750 cm^{-1}. Similarly, high resolution MCT spectra exhibited the same temperature-dependent characteristics. Investigation into the properties of the barium fluoride (BaF_2) windows on the aerosol cell revealed that there is a known "temperature influence on infrared cut-off."[5] Although BaF_2 has been used by other groups (Niedziela *et al.*, 1999; Norman *et al.*, 1999), no references to the temperature dependence of BaF_2 in similar experiments have been identified. This effect is characterised in this work, and a correction mechanism is formulated along with an assessment of uncertainties.

In order to characterise the temperature dependence of the BaF_2 windows, low resolu-

---

[5] From the *SPECAC Catalogue*, 2000.

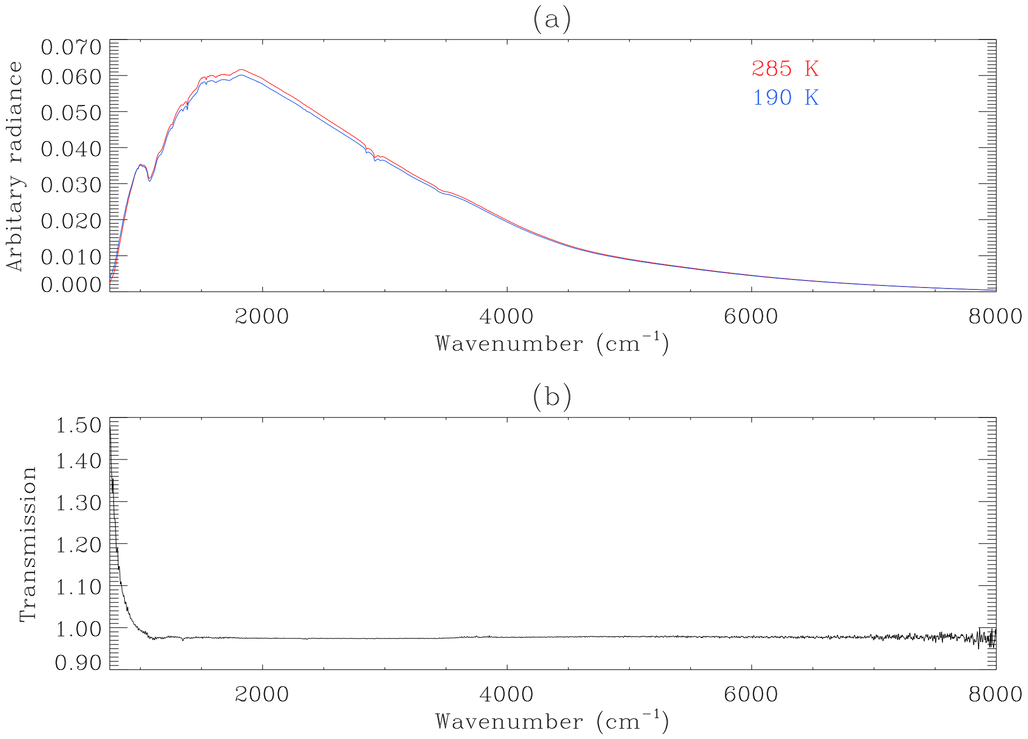

Figure 5.16: Low resolution DLaTGS background spectra from 28/11/00. The initial (warm, 200 Torr N_2) and final (cold, evacuated) backgrounds can be seen in panel (a). A ratio of the backgrounds (b) reveals a significant increase in transmission at low temperature near the infrared cut-off. This is typical of all days where cooling took place.

tion DLaTGS background spectra were measured in a test experiment over a representative temperature range (approximately 290-190 K), with the cell evacuated. It was found that, at all wavelengths, measured radiance (uncalibrated units) and calibrated temperature exhibit a smooth relationship. Ideally, temperature readings of the BaF_2 windows would be expected to yield the most realistic relationship, but window PRTs were not monitored in the experimental configurations used. Figure 5.17 illustrates that variation of radiance with calibrated temperature is a smooth function during the cooling phase at a fixed wavenumber. It was found that the variation of radiance with cooling is best represented by a second order polynomial at all wavenumber points. The polynomial coefficients from the weighted (by the radiance uncertainties) calibration fit at each wavenumber point (from

0-10000 cm^{-1}) were stored during the calibration process, along with the full covariance matrix for determination of errors.

For each measurement day backgrounds are measured at two temperature points, room temperature and the coldest point. Temperature correction uses the polynomial coefficients from the background calibration and fits to the initial and final background radiance. In figure 5.17 the polynomial fit (blue curve) does not coincide with the selected initial and final background radiance points. However, if a straight line joining the initial and final background radiance differences from the polynomial is subtracted from the polynomial fit, the resulting corrected polynomial function joins these points (yellow curve), such that:

$$\mathrm{poly\_rad}(T) = a_1 + (a_2 \times T) + (a_3 \times T^2) \tag{5.4}$$

$$d_2(T) = \frac{(\mathrm{poly\_rad}(T_{ini}) - \mathrm{rad}_{ini}) - (\mathrm{poly\_rad}(T_{fin}) - \mathrm{rad}_{fin})}{T_{ini} - T_{fin}} \tag{5.5}$$

$$d_1(T) = (\mathrm{poly\_rad}(T_{ini}) - \mathrm{rad}_{ini}) - d_2 \times T_{ini} \tag{5.6}$$

$$\mathrm{corrected\_rad}(T) = (a_1 - d_1) + (a_2 - d_2) \times T + (a_3 \times T^2) \tag{5.7}$$

where poly_rad is the radiance from the calibration polynomial fit, $d_2(T)$ is the difference gradient and $d_1(T)$ the difference constant between the fit and the local background radiance, rad_{ini} and rad_{fin}. The final corrected radiance at a set wavenumber element for a given temperature is corrected_rad. The coefficients d_1 and d_2 represent a straight line fit through the differences. This process is used to fit initial and final background radiance for each measurement day. The radiance for any selected calibrated temperature is then straightforward to calculate from the corrected second order polynomial function.

By evaluating this interpolation between initial and final backgrounds according to the polynomial fit, the temperature dependence of the $\mathrm{BaF_2}$ window transmission is accurately represented with respect to the background calibration experiments. Limitations of this method include:

- The possibility that the calibration differs fundamentally from normal experiments since the cell was evacuated (affecting thermal equilibration of the windows) in the

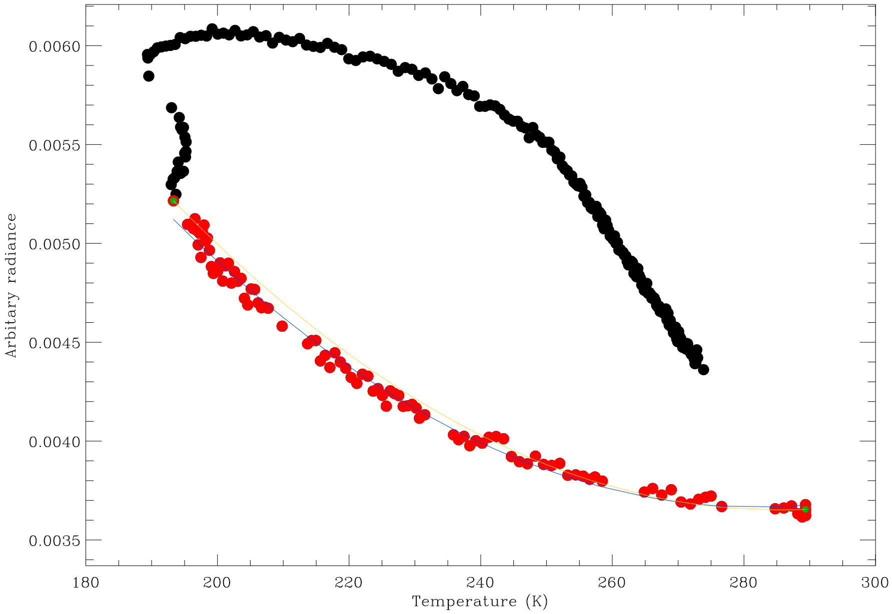

Figure 5.17: Variation of radiance at 771 cm^{-1} with calibrated temperature. Spectra measured during background calibration. Red points were measured during cooling, and black points during CO measurements and overnight warming. The blue curve represents a second-order polynomial fit to the calibration (red) points, and the yellow line is the adjusted fit to the initial and final background points (green).

test experiment;

- Poor characterisation of contamination – excess HNO_3 gas can attack the BaF_2 windows, resulting in anomalous absorption not fully modelled by smooth interpolation between initial and final backgrounds;
- The lack of a more direct measure of window temperatures imposes limitations on accuracy.

This work represents the first known calibration of the temperature dependence of transmission of BaF_2 windows. Figure 5.18 shows a series of stable low resolution aerosol spectra recorded during on a single day, with and without background correction. It can be seen that there is a significant difference below 1000 cm^{-1} between these cases. Comparison with the work of Niedziela *et al.* (1999) reveals that spectra published by this group may

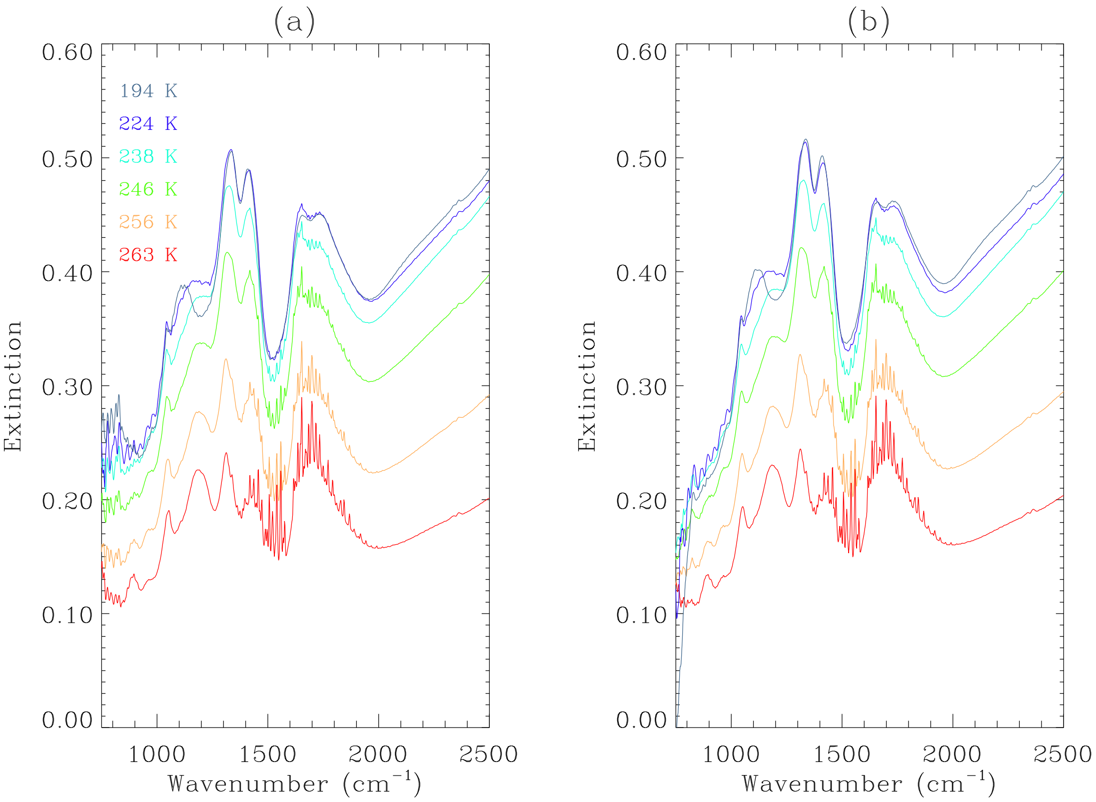

Figure 5.18: Extinction spectra of ternary solution aerosol measured on 28/11/00 (a) with corrected backgrounds, and (b) with initial background.

have been affected by BaF_2 temperature dependence.

5.4.3 Evaluation of uncertainties in extinction spectra

In the correction of BaF_2 temperature dependence, both the second order polynomial background calibration fit and the corrective straight line fit have associated errors. The covariance matrices can be added (assuming the covariances are independent) prior to matrix multiplication with the derivative of the corrected polynomial function and its transpose to yield the final error in the calculated radiance point, i.e. $\sigma^2 = \frac{\partial \mathbf{V}}{\partial c}(\mathbf{S}_\epsilon)\frac{\partial \mathbf{V}}{\partial c}^T$, where $\mathbf{V}$ is the expression for the corrected polynomial, c the polynomial coefficients, and $\mathbf{S}_\epsilon$ the combined covariance matrix of the fit. Figure 5.19 illustrates how the errors from the polynomial background fit and the corrective straight line fit combine to produce the total error at a single wavenumber value over a range of temperatures exceeding the background temperatures from this day (plotted). The uncertainty of the second order polynomial fit is low compared with the corrective straight line error, due to the subtraction of two measurement points (absolute noise level) before fitting in the latter case. This results in a higher than absolute noise level error around the temperatures of the backgrounds. However, this does not include the true error in the background correction model, which would be greatest in the middle range where the assumption that the calibration day scales the same as other days would lead to the greatest uncertainties. Errors are combined at every wavenumber to produce an error spectrum for the background. Finally, the calibrated temperature uncertainty (± 2.13 K) is used to calculate the correlating uncertainty in background radiance over this range. It is assumed this error is small enough to be symmetric, although it dominates at about ten times the level of the corrected measurement noise.

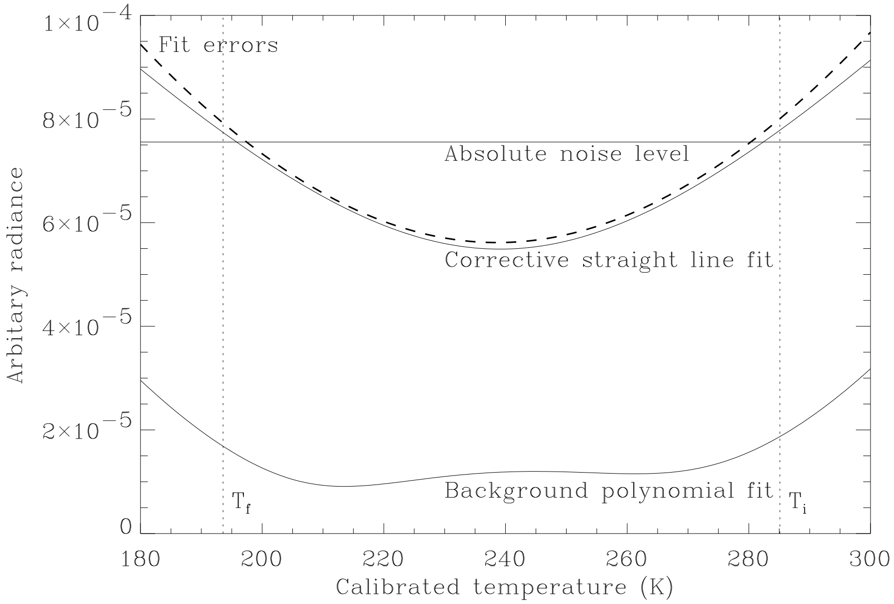

Figure 5.19: Propagation of radiance uncertainty in background spectra at a single wavenumber (869 cm^{-1}) on 22/11/00. The initial and final background temperatures are indicated by the vertical lines $\mathrm{T_i}$ and $\mathrm{T_f}$ respectively. The absolute noise level is the estimated noise at this wavenumber.

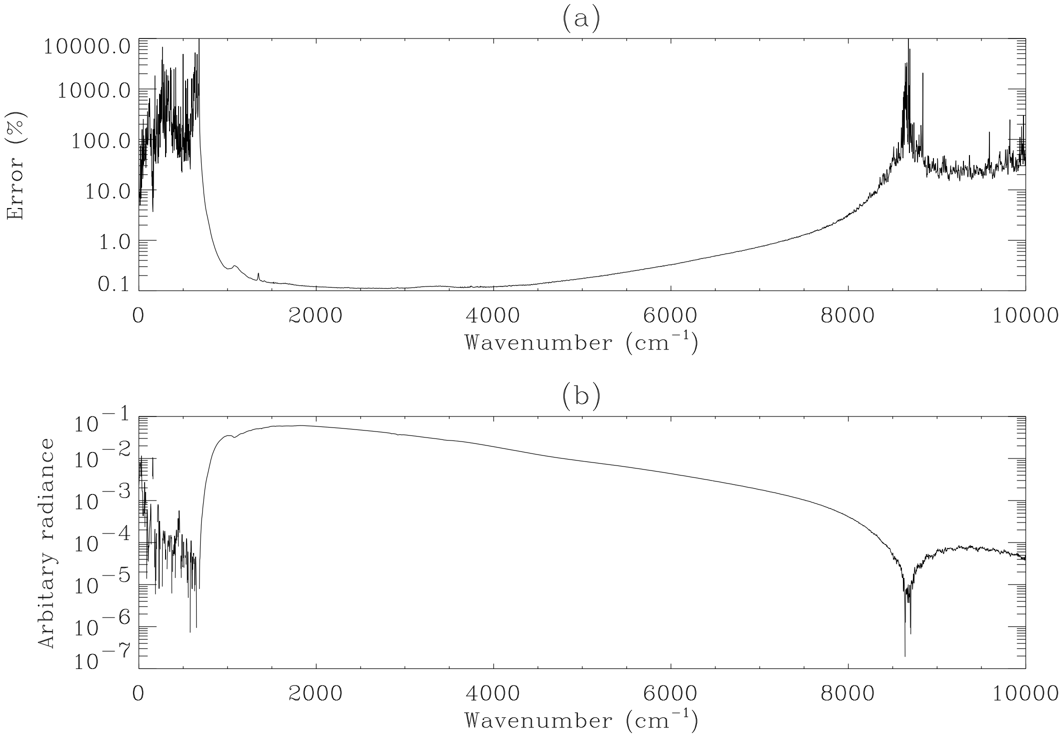

Figure 5.20: Panel (a) shows the calculated uncertainty as a percentage of a typical aerosol measurement (from 22/11/00), and the corresponding corrected background can be seen in (b).

Since the measurement errors are the calculated absolute noise values, it is a simple matter to combine these with the background errors. It is assumed these errors are completely independent. Figure 5.20(a) shows the magnitude of the calculated error as a relative percentage of a typical sulphate aerosol spectrum, measured on 22/11/00. Increased uncertainty around 750 cm^{-1} and 1300 cm^{-1} is due to the window temperature dependence and change in window contamination respectively. The corrected background for this measurement can be seen in figure 5.20(b); the very low signal around 8700 cm^{-1} is possibly due to electronic filtering in the FT instrument, or detector non-linearity (since this is outside the detector range – see section 4.1.5). It is important not to confuse measurement uncertainty here with information content, since outside the spectral range of the detector and the optics spurious signals can be confused with useful data. Therefore, it is concluded that the main experiment setup provides good quality spectral coverage over the range 750-8000 cm^{-1}.

5.5 Summary

Characterisation of the experiment system has encompassed:

- Identification of STS aerosol in the cell at a range of temperatures;
- The first use, in this field, of a single FTS instrument to characterise broad band features of aerosol **and** gas components for measurement of partial pressures and calibration of cell temperature measurements;
- Characterisation, and modelling, of the temperature dependence of the transmission of BaF_2 windows;
- Comprehensive calibration of uncertainties in the experiment system enabling calculation of uncertainties in the spectral data (and potentially refractive indices).

Chapter 6

Refractive Index Determination

The derivation of wavelength-dependant complex refractive indices is key to achieving the objectives of this thesis. It was revealed in the characterisation of the experiment system that it was not possible to produce data suitable for derivation of refractive indices using *traditional* Kramers-Kronig based techniques. The novel approach adopted here is based on the classic damped harmonic oscillator (CDHO) model (Bertie and Zhang, 1992), which enables the retrieval of refractive indices from a single measured extinction spectrum. The fundamental principals of modelling absorption spectra using the CDHO model were discussed in detail in section 2.3.4. In addition to fitting the absorption features of aerosol spectra with symmetrical absorption bands, the *forward model* of the refractive index retrieval fits the scattering of aerosol in order to accurately model the measured aerosol extinction spectra. Thus, the products of the retrieval are the CDHO band parameters, which can be used to represent the wavenumber-dependant refractive index and to determine the speciation of solutions (after calibration – section 7.1), and aerosol particle size distribution parameters which can be employed to evaluate related parameters such as aerosol volume and gravitational settling velocities.

6.1 The forward model

The theoretical basis for the forward model of the refractive index retrieval was detailed previously in section 2.3. The forward model evaluates the complex dielectric spectrum from the CDHO model band parameters of oscillator centre $\tilde{\nu}_j$, width (damping constant) γ_j and intensity S_j for a set of j bands at each wavenumber element. The complex refractive index can be calculated from the dielectric constant using equations 2.20 and 2.21. A Mie scattering model is used to simulate extinction of homogeneous spherical droplets through scattering and absorption using this refractive index, see equation 2.5. A log-normal size distribution of aerosol particles is assumed, see equation 2.7; this has been widely accepted as the definitive distribution to use in similar aerosol experiments (Willeke and Baron, 1993). The state vector elements describing this distribution, the number of particles per cm^3, N, the median radius, r, and the width of the distribution, σ, are the first elements in the state vector. The offset value of the (real) refractive index at infinite wavenumber (zero frequency), m_∞, is retrieved as the fourth element in the state vector, followed by j sets of band parameters. Modified versions of this scheme were used in order to test and validate aspects of the retrieval. The wavenumber shift due to beam divergence was accounted for independently of the forward model (section 5.2.3).

Calculation of the Jacobian (or weighting function), $\mathbf{K}$, was undertaken using a numerical perturbation method (rather than analytical derivative calculation) for each state vector element. Since this involves calling the Mie code to calculate the extinction at every wavenumber for each state vector element this can be a computationally expensive process.[1] An analytical calculation of $\mathbf{K}$ for each state vector element raises the prospect of a much faster retrieval, but remains elusive as this is a non-trivial derivation from Mie formulae.

---

[1] Typically taking over 30 minutes for Jacobian calculation for all state vector elements in a 30-band case from 750-8000 cm^{-1}.

6.1.1 Multiple scattering

The assumption is made in the forward model that radiation is not scattered back into the beam from either single or multiple scattered photons, and that no thermal emission occurs – this assumption was tested by modelling the scattering of aerosol in the cell. The contribution of scattered photons to the measured signal was assessed using a Monte Carlo simulation with the applicable geometry. Previous analysis demonstrated that scattering of light into the beam was negligible in the case of a smaller cell used at RAL (Heathfield *et al.*, 1999). More recently calculations were undertaken for the large cell (R.G. Grainger & A. Brian; personal communication) which produced a similar result. The main findings are that, in spite of the greater internal area of the large cell and the consequent higher probability for photons to multiple scatter around the cell, the probability that photons scatter back into the beam with a small enough angle to the transmitted beam (the acceptance angle) to reach the detector surface is negligible, thus the enhancement of the observed radiation from multiple scattering is small at realistic optical depths.[2]

6.1.2 Formulation of a generic set of band parameters

The first stage of analysis was to determine an initial estimate of the log-normal monomodal size parameters N, r and σ, by fitting the region of spectra where extinction is dominated by scattering of aerosol particles (4500 to 7000 cm^{-1}). The next stage was to fit CDHO band parameters to model absorption below approximately 4500 cm^{-1}, where $k(\tilde{\nu}) \neq 0$.

Initial use of an *a priori* and first guess set of band parameters compiled from literature sources (detailed in table 2.1, section 2.5.8) resulted in a generally poor fit to a typical representative extinction spectrum of supercooled ternary solution (STS) aerosol. User-intervention was thus required in the initial stages of analysis to generate a *generic set* of band parameters to fit measured aerosol extinction spectra more closely. This process involved calculating peak wavenumbers of measured absorption bands (from extinction

---

[2] Most photons will be absorbed rather than reflected by the acid-coated stainless steel cell walls.

spectra), and gradually improving the fit by using the residual between the measured and modelled spectrum as a guide. Bands were combined from binary sulphate spectra and binary nitrate spectra (from literature) in order to obtain a complete set of band parameters applicable in the binary and ternary aerosol cases. Only physically representative bands were included in this analysis. The generic band set is detailed in table 6.1.

Table 6.1: Generic band set parameters.

Molecule	$\tilde{\nu}_j[\mathrm{cm}^{-1}]$	$\gamma_j[\mathrm{cm}^{-1}]$	$S_j[\mathrm{cm}^{-2}]$
H_2SO_4	573.3	0.1	129512.9
$NO_3{}^-$	743.8	98.0	42282.6
$NO_3{}^-$	822.7	106.2	22424.2
$SO_4{}^{2-}$	896.6	96.3	13706.2
H_2SO_4	961.3	62.2	3203.4
$SO_4{}^{2-}$	1029.7	128.7	11109.0
$SO_4{}^{2-}$	1047.7	38.7	4901.9
$HSO_4{}^-$	1119.4	104.6	15665.4
$HSO_4{}^-$	1149.6	76.1	43311.2
H_2SO_4	1188.0	90.4	24364.0
H_3O^+	1229.9	63.6	7648.5
HNO_3	1299.8	86.7	15537.8
$NO_3{}^-$ ($SO_4{}^{2-}$)	1343.9	78.7	29010.3
HNO_3 (H_2SO_4)	1402.2	84.0	24297.3
HNO_3 (H_2SO_4)	1445.9	105.5	22008.0
H_2O (HNO_3)	1640.3	83.3	35879.5
H_3O^+ (HNO_3)	1716.1	268.7	91824.6
H_3O^+	1792.2	178.2	17085.9
Unassigned	1912.4	291.3	48254.0
Unassigned	2069.6	210.4	16870.6
Unassigned	2219.5	430.8	94071.3
Unassigned	2464.6	379.6	38443.9
HNO_3	2653.3	192.2	8242.9
H_2O	2824.4	266.4	30155.6
HNO_3	2969.5	176.5	23376.1
H_2O	3113.7	187.0	68486.2
H_2O	3240.4	185.5	170255.0
H_2O	3356.0	152.4	150793.6
H_2O	3444.2	131.1	134657.7
H_2O	3524.9	90.9	62833.4
H_2O	3595.3	80.0	39971.5

The generic band set was used as an *a priori* state vector for all refractive index retrievals reported in this thesis, although there is some band redundancy in dilute and binary solutions. For example, nitrate bands are not present in a binary sulphate spectrum and vice versa – this is not important since the non-relevant bands tend towards zero band area in the retrievals, indicating that the generic band set is a robust *a priori* state for all binary and ternary solution retrievals conducted here. More generally, it should be noted that the retrieval is not strongly constrained to the *a priori* state values for the generic band set (section 6.1.7). As such the *a priori* state should be considered to represent a *basis* for the retrieved band parameters within the retrieval constraints specified in section 6.1.7. The exact band parameters in the generic band set are not crucial to the outcome of the retrieval.

6.1.3 Forward model error

A converged fit to a cold sulphate aerosol extinction spectrum as measured in the final laboratory session can be seen in figure 6.1. The limitations of the forward model in the region near 3700 cm^{-1} are evident from the normalised residual between the forward model and the measured spectrum.[3] The primary problem is *ringing* that reaches a maximum around 3650 cm^{-1}, which corresponds to about 20 % of the measured extinction. The normalised residual is so large in this region that the retrieval does not calculate such a good fit over the remainder of the spectrum. Another problem can be seen with CO_2 contamination around 2350 cm^{-1}, the source of which has been described previously (section 5.3.2).

The multiple bands located around 3400 cm^{-1} are due to molecular vibrations in H_2O molecules (Querry and Tyler, 1980). There are three intra-molecular vibrations, the v_1 symmetric and v_3 antisymmetric O-H stretching modes, and the v_2 O-H-O bending mode. The v_2 band is located around 1640 cm^{-1}. The v_1 and v_3 stretching modes produce overlapping bands, possibly in Fermi resonance with the first overtone of the v_2 band around 3280 cm^{-1}. Moreover, the asymmetric O-H stretching modes lead to asymmetric

---

[3] The normalised residual is the straight residual between the measurements and the forward model, scaled by the measurement uncertainty, i.e. $\frac{\mathrm{Y-F}}{\delta \mathrm{Y}}$.

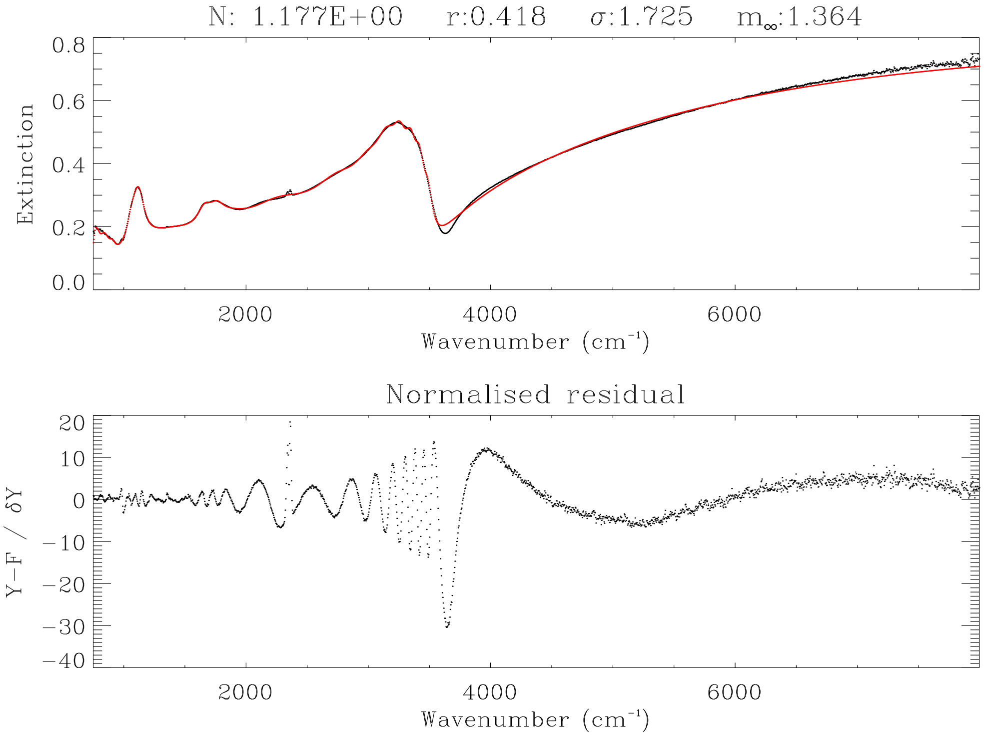

Figure 6.1: Converged refractive index retrieval fit to a 198.9 K sulphate aerosol spectrum from 22/11/00, no forward model error included. The red line in the first panel is the forward model, black is the measurement. N = particle density [10^6 cm^{-3}], r = median radius [μm], σ = spread, and m_∞ is refractive index anchor point.

band shapes (Palamarev and Georgiev, 1994; Libnau *et al.*, 1994), which cannot be directly modelled by the CDHO model (which produces a symmetric band shape).

In the fit shown in figure 6.1, the solution consists of a total of seven OH bands from 2786-3487 cm^{-1}. A larger number of separate symmetric bands are used here to represent a smaller number of asymmetric bands. However, this method does not produce a perfect fit to the asymmetric bands, as can be seen from the non-physical structure or ringing around the peak of the OH band at the centre of each narrow band. Two main approaches were taken to improve the fit to the OH bands: (1) modifying the bandshape, and (2) varying the number of symmetric bands fit to the OH region. In the first case, a number of measures were implemented, including the introduction of a band asymmetry term into the retrieval, but with no significant improvement of the retrieval. In the second approach, the number of symmetric bands in the OH region were increased to 40, which produced an excellent fit for a highly-acidic binary sulphuric acid measurement (of about 77 wt% composition), but did not appreciably improve the fit to the OH region for ternary solution measurements. Coupled with an additional computational (time) penalty of using such a high number of bands, this approach was found not to improve the retrievals. As a result of this analysis, it was found that the OH region was best represented by an intermediate number of symmetrical bands – in the generic band set (section 6.1.2) nine bands were employed to model the region from 2500 to 3600 cm^{-1}.

Two methods were investigated to mitigate the effects of the poor fit to the OH cut-off region: (1) removing the 2500 to 3600 cm^{-1} wavenumber region from the retrieval, and (2) introducing forward model error into the retrieval. The first approach was found to have the undesirable side-effect of leading the retrieval to vary bands inside the blanked out region to high strengths in order to help the fit to the surrounding regions; this resulted in many cases in high absorbance values at high wavenumbers where k should tend towards zero. This high wavenumber absorbance, defined here as *k leakage*, would then result in the retrieval of incorrect size parameters. In the second approach, forward model error was introduced in the retrieval by selectively increasing the measurement error over problem regions (A. Lambert; personal communication). Although the true forward model error can only be estimated, the measurement error was increased by a factor of ×40 from 3500 to 5500 cm^{-1} in all retrievals. This method reduces sensitivity to the normalised residual over this region, yielding a compromise between k leakage and improving the fit across other regions.

Forward model errors of appropriate factors were also included to cover the CO_2 contamination region and regions of strong H_2O lines in specific instances at room temperature (i.e. H_2O lines can be seen in the room temperature extinction spectrum in figure 6.1). Preliminary investigation was undertaken to assess the feasibility of modelling the CO_2, H_2O, and HNO_3 gas lines in low resolution DLaTGS spectra, but with so few sampling points

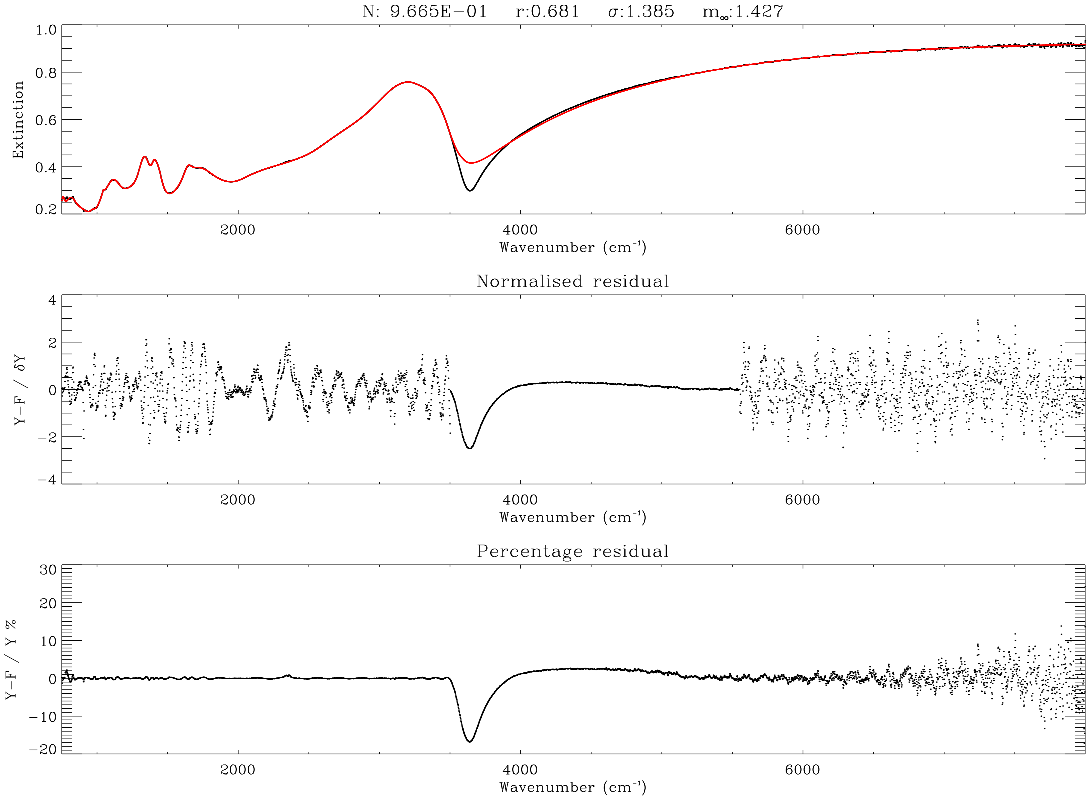

Figure 6.2: Converged refractive index retrieval fit to a 196.7 K ternary aerosol spectrum from 29/11/00, forward model error included as in the final retrieval configuration. The red line in the first panel is the forward model, black is the measurement, and the third panel shows the total percentage error. N = particle density [10^6 cm^{-3}], r = median radius [μm], σ = spread, and m_∞ is refractive index anchor point.

this would be unreliable. This is not of major concern as gas lines are not visible above background noise in laboratory spectra at stratospheric temperatures since the equilibrium partial pressures reduce considerably with temperature, and the optical path in the cell is short. Figure 6.2 shows the converged fit to a cold ternary solution aerosol spectrum, using forward model error; although the fit from 3500 to 5500 cm^{-1} remains poor, the fit over the remainder of the spectrum is improved. Indeed, the third panel in figure 6.2 shows that the percentage error of the fit is below 1% through most of the spectrum. This is a far better fit to spectra measured for this thesis than was achieved with the refractive index data of Biermann *et al.* (2000) (section 2.5.4).

The key points to note from this analysis are that:

1. The quality of fit between the forward model and measured extinction in the spectral regions of interest is good (to within 1 % extinction), although larger than the measurement error of spectral extinction;
2. The quality of fit is conveyed in the errors in refractive index produced from the retrieval.

6.1.4 Validation of the method

In the first instance the refractive index retrieval can be considered to be self-validating since the retrieved refractive index can be used to accurately model the measured extinction through the forward model. However, further inspection reveals that the solution in band space is not unique, which raises important questions pertaining to the independence of retrieved parameters. In order to explain this, further investigation of the nature of the solution is required. A converged fit to an aerosol spectrum is comprised of a series of 31 overlapping bands (as used for the majority of the refractive index retrievals conducted in this work). Bands are separated by varying distances, and some are comprised of doublets or even multiple bands (such as the OH stretch) which in a selection of cases directly represents reality, and in other cases is an attempt to better model an asymmetric band shape. Of interest in this discussion is the uniqueness of the solution, or rather the independence of the retrieved parameters in the state vector.

Figure 6.3 shows the band components of the converged solution for the ternary solution spectrum in figure 6.2. The bands are displayed in dielectric space since the band parameters are retrieved as dielectric quantities. From visual inspection of figure 6.3 it can be hypothesised that a different configuration of bands can yield the same aggregate solution within the fit uncertainty; implying different band solutions can produce the same refractive indices. The consequence of this would be that information is shared (or smeared) between the state vector elements, implying that the band parameters are not unique for a

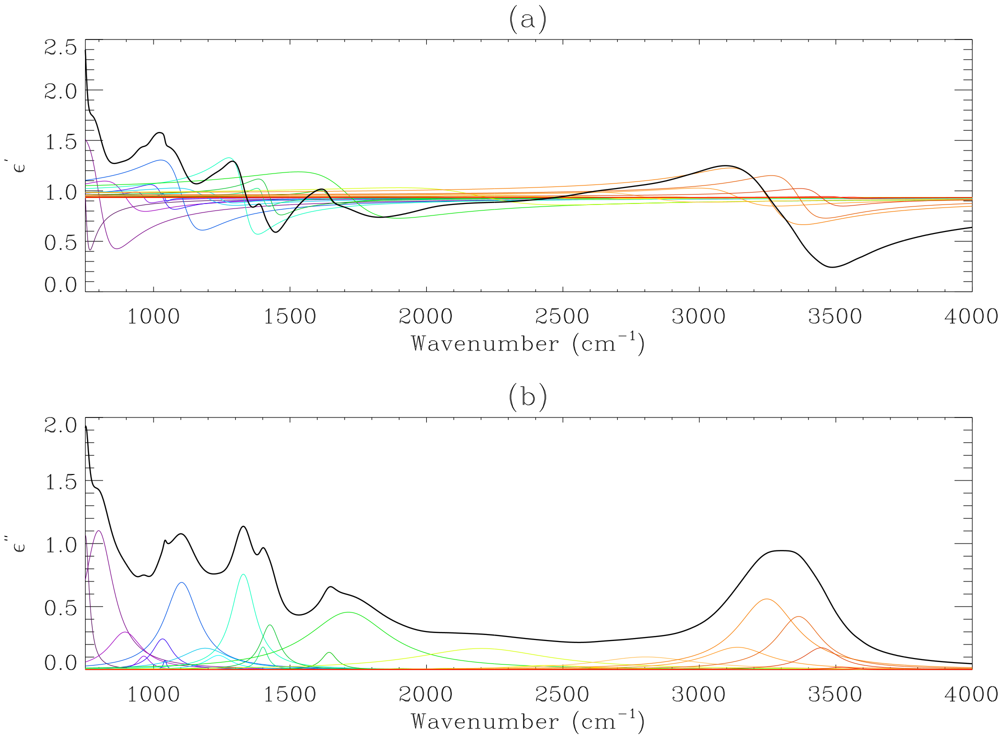

Figure 6.3: Individual bands in dielectric space comprising the converged solution refractive index retrieval fit to a 196.7 K ternary aerosol spectrum from 29/11/00, forward model error included. Band colours range from blue at the lowest wavenumbers to red at higher wavenumbers. Panel (a) shows the real component, and (b) the imaginary component of the complex dielectric constant. The thick line shows the aggregate dielectric constant, forming the solution.

typical multi-band refractive index data set. This hypothesis is confirmed by analysing the averaging kernel $\mathbf{A}$ scaled against the retrieved covariance matrix $\hat{\mathbf{S}}$, and the correlation matrix as previously defined and explained in section 5.2.6.

The averaging kernel. The state vector consists of N, r, σ, m_∞, followed by j sets of band parameters $\tilde{\nu}_j$, γ_j and S_j. The scaled averaging kernel, a measure of the sensitivity of the retrieved state to the true state $\mathbf{A} = \partial\hat{\mathrm{x}}/\partial\mathrm{x}$, is shown as a surface in figure 6.4. In this representation the matrix elements are arranged so that the sensitivity of the retrieved state element to each true state element in turn is calculated from left to right across the surface. For example, in the line closest to the front, from left to right, the elements consist of $\partial\hat{N}/\partial N$, $\partial\hat{N}/\partial r$, $\partial\hat{N}/\partial\sigma$, $\partial\hat{N}/\partial m_\infty$, $\partial\hat{N}/\partial\tilde{\nu}_0$, $\partial\hat{N}/\partial\gamma_0$, $\partial\hat{N}/\partial\mathbf{S}_0$, and so forth. The next line back contains the next retrieved state vector element in the numerator, with the same true state denominators from left to right. Hence, the physical meaning of the averaging kernel as plotted in figure 6.4 can be interpreted as the sensitivity of the retrieved state vector element on the front axis with respect to the true state vector elements along the left axis.

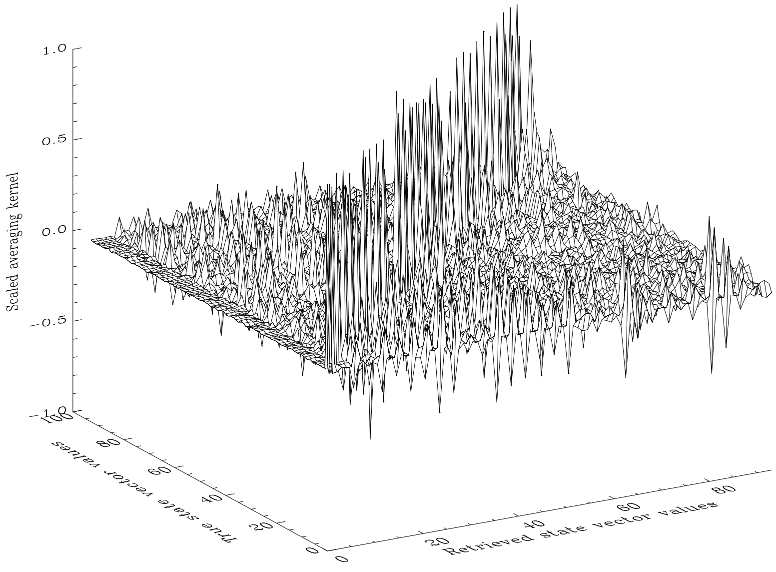

Figure 6.4: Averaging kernel scaled against the retrieved covariance from the fit to a 196.7 K ternary aerosol spectrum from 29/11/00.

Ideally the scaled averaging kernel in figure 6.4 would be flat (zero) except for unity values along the diagonal, indicating the retrieved values are truly independent. The scaled (against the retrieved covariance) off-diagonals indicate a smearing of information quite widely across band parameters. The highest off-diagonals are located close to the centre diagonal indicating that most smearing of information occurs between adjacent bands and the position, width and strengths of individual bands. The strip of near zero off-diagonals for the four elements along the left-hand side indicates that the retrieved size parameters are not particularly sensitive to the other state vector elements. However, the significant values in the first four strips at the front indicate that the retrieved size parameters and m_∞ have some dependence on the true state vector elements for the band parameters. Thus, the values of the retrieved size parameters and m_∞ are likely to be modified to some extent by the absorbance bands. For clarity, a graph of the diagonal terms segregated into components is displayed in figure 6.5. It can be clearly seen that the size parameters and m_∞ are relatively independent, whilst there is wide variation amongst the band parameters. Generally, the band positions show greater independence than the band widths, while the retrieved band strength values show least independence.

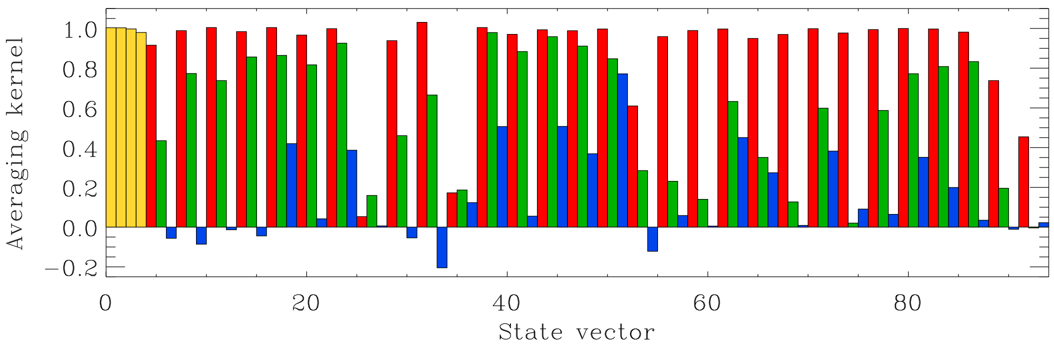

Figure 6.5: Graph showing diagonal elements from the averaging kernel of the converged solution to a 196.7 K ternary aerosol spectrum from 29/11/00. The yellow boxes indicate the size parameters N, r, σ and m_∞ respectively, followed by j bands with the red boxes indicating band position, $\tilde{\nu}_j$, green the band width, γ_j, and blue the band intensity, S_j.

Correlation matrix. Also of interest in the validation is the correlation matrix, which is the covariance matrix normalised against diagonal terms, which therefore always equal unity. The correlation matrix is symmetric about the diagonal since the correlation between two error terms will be equivalent regardless of the sense of the calculation. The correlation matrix for the same 29/11/00 spectrum as discussed previously can be seen in figure 6.6. Widely distributed non-zero off-diagonals indicate significant correlation between error terms, which is in agreement with the smearing of information between the retrieved parameters already noted from the averaging kernel. These correlations are used to evaluate the uncertainties associated with finding the ratio of multiple band areas in the determination of composition from band areas (section 7.1.1).

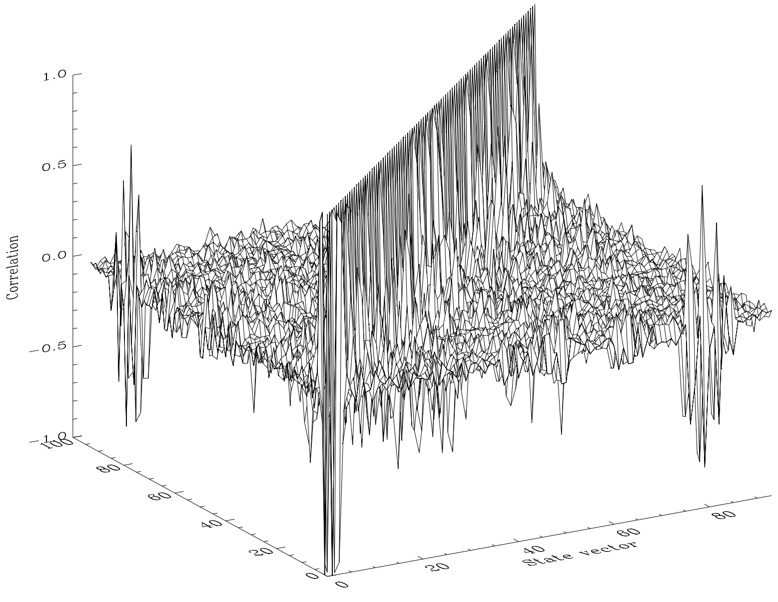

Figure 6.6: Correlation matrix of the fit to a 196.7 K ternary aerosol spectrum from 29/11/00.

A feature of the correlation matrix in figure 6.6 is the high correlation between the last elements of the state vector and N, r, σ and m_∞. Further investigation reveals that the maximum correlation (positive) and anti-correlations (negative) are of approximate magnitude 0.8, and occur most strongly as anti-correlations to band position of the higher bands with N, σ and m_∞ and correlation with r. These high correlations suggest a problem with anomalous absorption (non-zero k) extending to high wavenumbers from the summation of the tail ends of all the band absorptions in the imaginary index (k leakage), particularly those with the high correlation values – indicating that this absorption is affecting retrieval of the size parameters and m_∞. Figure 6.7 shows the $k(\tilde{\nu})$ contribution from all bands at high wavenumber from the fit to the same ternary spectrum as displayed in figure 6.3, using the same colour scheme for comparison.

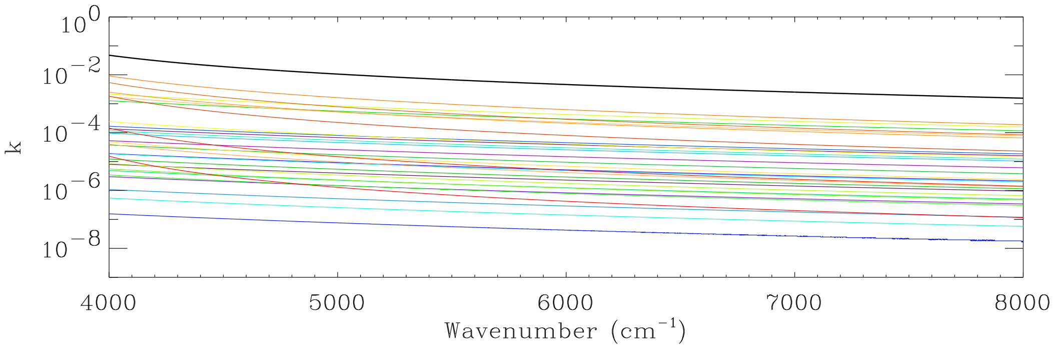

Figure 6.7: Individual bands in imaginary dielectric space comprising the converged solution refractive index retrieval fit to a 196.7 K ternary aerosol spectrum from 29/11/00. The more red the colour, the higher the centre $\tilde{\nu}$ of the band. The thick line shows the aggregate dielectric constant; the non-zero value of $k(\tilde{\nu})$ illustrates the concept of k leakage.

Other published sets of refractive index data (Biermann *et al.*, 2000; Niedziela *et al.*, 1998b, 1999) for sulphuric and nitric acid solutions give k values much closer to zero above 4000 cm^{-1}. It may well be the case that some absorption occurs at high wavenumbers, but this cannot be verified due to lack of validated data. To ascertain the effect of this anomalous absorption on size parameters and m_∞, the size parameters from two retrievals were compared where the value of k was set to zero above 4200 cm^{-1} in one case; the results are listed in table 6.2. The aerosol volume density for this table was evaluated from the moments of the log-normal distribution using:

$$V = \frac{4}{3}\pi N r^3 e^{\left(\frac{9\ \ln(\sigma)^2}{2}\right)} \tag{6.1}$$

This volume density calculation is dimensionless since this expression evaluates the volume [m^3] of droplets per unit volume [m^{-3}]. It can be seen that there are significant variations in the individual size parameters and the volume; the retrieved volume is 0.8 % higher when k is fixed to zero, relatively large compared with the smaller retrieval uncertainties. This indicates that k leakage is causing the retrieval to underestimate the aerosol volume by 0.8 %. Although this difference is outside the retrieval error, its small magnitude indicates that volume retrieval is more robust than individual size parameters.

Table 6.2: Effect of k leakage.

Param.	N ($\times10^6$ cm^{-3})		r (μm)		σ		V		%
Free k	0.603	$\pm$0.006	0.571	$\pm$0.003	1.312	$\pm$0.002	6.548×10^{-7}	$\pm4.93\times10^{-10}$	$\pm$0.08
$k=0$	0.527	$\pm$0.006	0.603	$\pm$0.003	1.293	$\pm$0.002	6.496×10^{-7}	$\pm4.96\times10^{-10}$	$\pm$0.08
Diff.	0.0762		-0.0316		0.0187		5.191×10^{-9}		0.8

Fits to simulated spectra. Finally, the retrieval was tested by fitting to a synthetic extinction spectrum; the forward model output from the retrieval fit to a ternary solution aerosol spectrum from 29/11/00 (as shown in figure 6.2) was used as the input to the retrieval. The retrieval was then executed from the *a priori* state and first guess values of the generic band set (section 6.1.2) that was used for the initial retrieval fit to the measured spectrum. No noise was added to the synthetic spectrum, due to problems with applying the correct apodisation of noise at such low resolution, and the associated measurement error in the retrieval was set to a nominal value. As with the gas retrieval (section 5.2.6), retrieval of a synthetic spectrum with no noise should be ideally expected to yield zero residual. The retrieved size parameters and volume are compared in table 6.3; the retrieved errors on the synthetic spectrum are irrelevant since there is nominal measurement noise, but the retrieved values agree to within the uncertainty of the parameters initially retrieved from measured data.

Table 6.3: Comparison of retrieved size parameters from synthetic retrieval.

Param.	N ($\times10^6$ cm^{-3})		r (μm)		σ		V		%
Synth.	0.967	$\pm$0.035	0.681	$\pm$0.013	1.385	$\pm$0.010	2.061×10^{-6}	$\pm3.87\times10^{-9}$	$\pm$0.19
Ret.	0.907	N/A	0.705	N/A	1.378	N/A	2.115×10^{-6}	N/A	N/A
Diff.	0.0591		-0.0239		0.0067		-5.45×10^{-8}		2.6

The percentage residual (figure 6.8) shows minor ringing in the extinction spectrum of the fit to a synthetic spectrum, reaching a maximum around the OH band cut-off. The reason for the high residual in this region is due to the fact that the band parameters are a non-unique solution for the refractive index data (as previously discussed); highest residuals occur where band spacing is closest which results in the most information smearing (the least unique band solution). The residual of the synthetic fit is the forward model error in extinction due to the non-unique nature of the retrieved band parameters – this is distinct from the forward model error due to poor fitting of asymmetric band parameters. The latter source of forward model error dominates; compare the percentage residual in figure 6.2 with that of the synthetic retrieval in figure 6.8. This demonstrates that, although there is a perceptible error in refractive indices due to the non-unique nature of the band solution, the magnitude of this error is considerably less than that due to other sources. The uncertainties in band parameters, due to non-unique band parameters in the converged solutions, are accounted for in the final error in retrieved refractive index data.

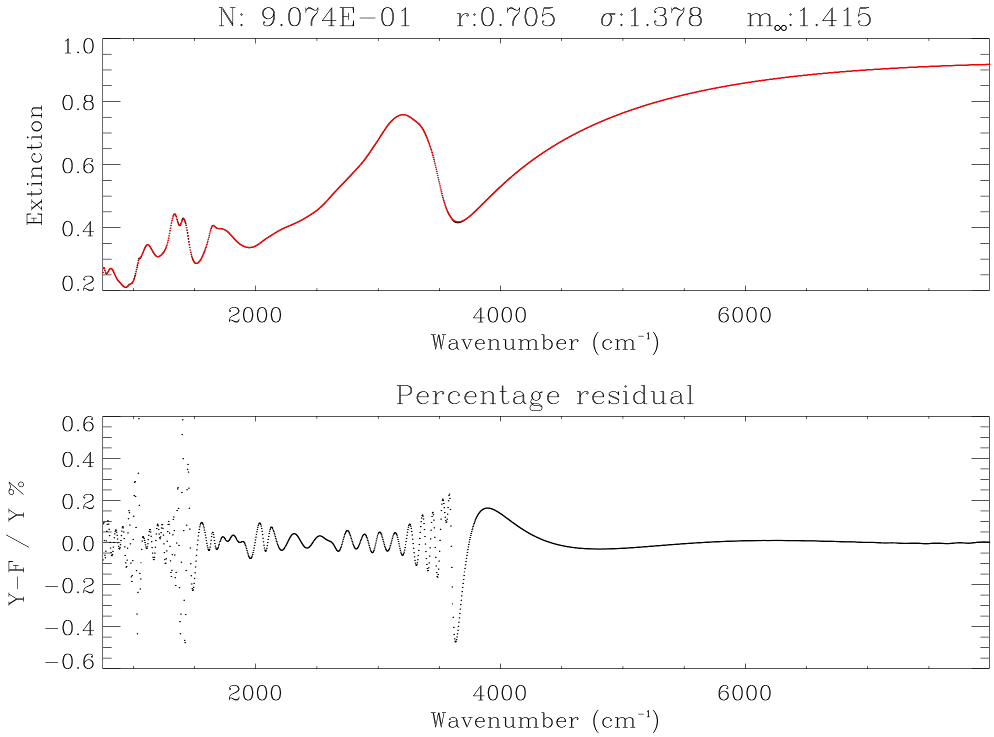

Figure 6.8: Converged refractive index retrieval fit to a synthetic extinction spectrum (the forward model fit to a cold ternary solution spectrum for 29/11/00). The red line in the first panel is the forward model, black is the measurement. N = particle density [10^6 cm^{-3}], r = median radius [μm], σ = spread, and m_∞ is refractive index anchor point.

6.1.5 Uncertainty in retrieved refractive index

The covariance matrix of the state vector solution can be used to evaluate the error in the refractive index. A matrix $\mathbf{e} = \partial\epsilon(\tilde{\nu})/\partial\mathbf{X}$, is evaluated for all state vector elements using analytically evaluated partial derivatives for the real and imaginary parts of the dielectric constant at every wavenumber. The covariance matrix relating the real and imaginary parts of the dielectric constant at a single wavenumber can then be calculated using $\hat{\mathbf{S}}_\epsilon = \mathbf{e} \times \hat{\mathbf{S}} \times \mathbf{e}^T$. Similarly, it is then straightforward to relate $\hat{\mathbf{S}}_{\epsilon(\tilde{\nu})}$ to the partial derivatives of the equations that relate the refractive index to the dielectric constant (equations 2.20 and 2.21), to produce values of the retrieval uncertainty in the real and imaginary parts of the refractive index.

Typical refractive index uncertainty is represented in figure 6.9 for a ternary aerosol retrieval; the maximum retrieval error is at the lowest wavenumber (750 cm^{-1}). The higher uncertainty below 1000 cm^{-1} is due to the propagation of high uncertainty from the BaF_2 window temperature dependence correction process. A secondary maximum around 3650 cm^{-1} is due to the poor fit of the bands model to the OH cut-off. The refractive index percentage error over most of the spectrum is less than 1 %, validating this method for the accurate retrieval of refractive indices. Although the band parameters are not unique, the resultant error in refractive index is small over most of the measured spectral region and the aim of retrieving refractive indices is achieved.

6.1.6 Relationship of composition to retrieved real refractive index

Published theoretical studies and experiments (Luo *et al.*, 1996; Krieger *et al.*, 2000b) (section 2.5.7) have demonstrated that the refractive index towards infinite wavenumber (zero frequency offset), defined as m_∞ in this work, is a function of solution composition and temperature. A single value of m_∞ can arise from a range of possible composition and temperature combinations, in the ternary solution case (Krieger *et al.*, 2000b). However, retrieval of an independent value of m_∞ is of interest in validating composition determina-

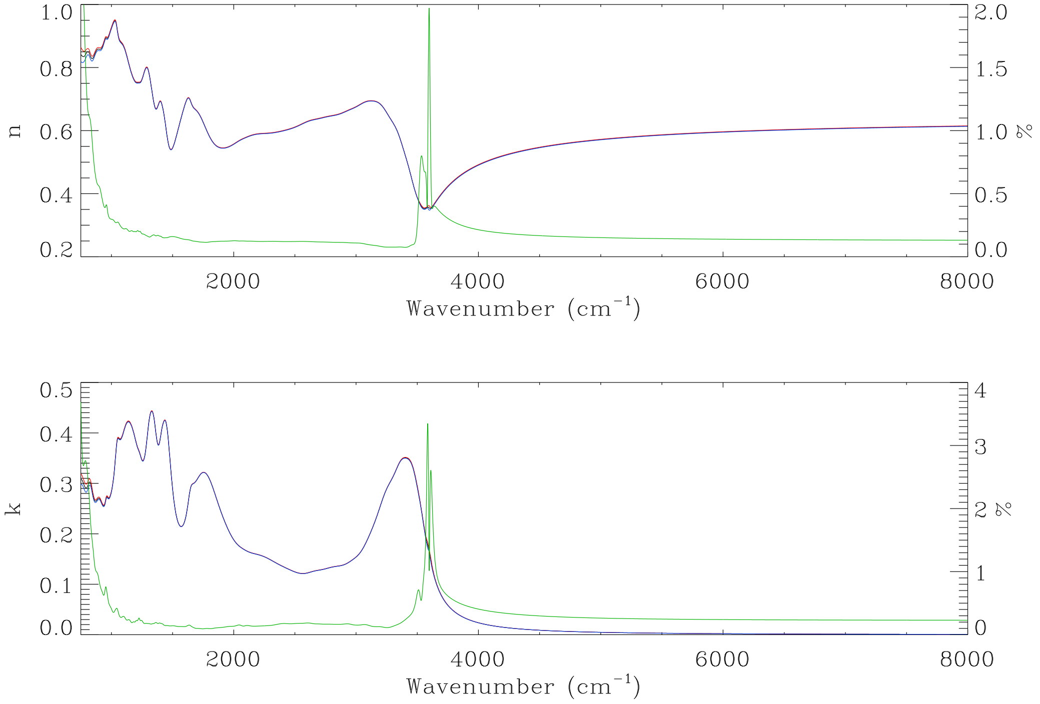

Figure 6.9: Refractive indices from a converged fit to a 208.4 K ternary solution aerosol spectrum from 21/11/00. Refractive index is represented by the black line, with the upper and lower uncertainty limits represented by the red and blue lines respectively; the error is so small it is barely distinguishable at this scale. The green line shows the error as a percentage of refractive index (axis on the right).

tion of the aerosol in this work. It was found last section that when k was forced to zero in a test run (table 6.2) the retrieved value of m_∞ changed from 1.417 to 1.413 ($k = 0$); although a larger magnitude change than the retrieval error of 0.0005, this stability indicates some independence from size parameters. However, in the synthetic retrieval (table 6.3) the value of m_∞ varied considerably from 1.427 in the synthetic spectrum to 1.415 in the synthetic retrieval. Since this spectrum contains wide band features which also leads to a higher complex refractive index error, there could be some effect from the wide side lobes of these bands on the value of m_∞. Certainly, m_∞ is not exempt from the smearing of information in this spectrum as indicated by the averaging kernel in figure 6.4.

The composition determined from analysis of retrieved band parameters (section 7.1)

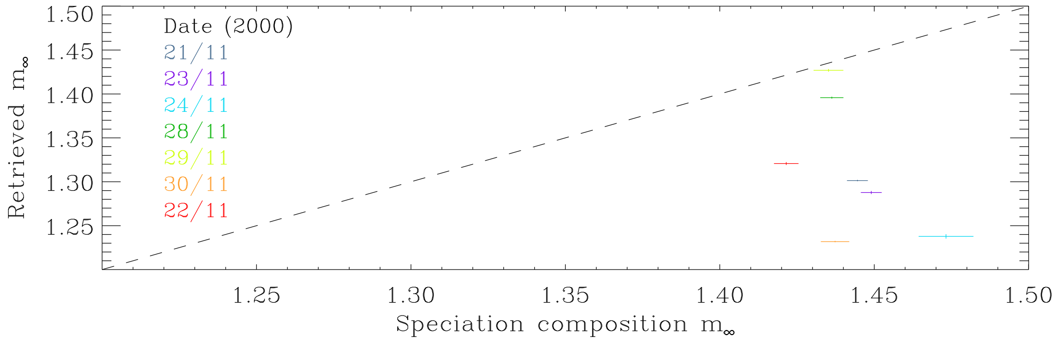

Figure 6.10: Values of m_∞ retrieved through the refractive index retrieval (y-axis) plotted against m_∞ values calculated from composition determined from analysis of retrieved band parameters and temperature using the model of Krieger *et al.* (2000b). Error bars are plotted for both quantities and the dashed line, $y = x$, represents optimal agreement of data sets. Measurements dates, from the main experiments, are colour coded.

is used here to calculate values of m_∞ from the model of Krieger *et al.* (2000b) for the coldest stable aerosol measurements from the main experiments of this work, error bars are evaluated from the uncertainty in composition; for the relevant compositions and uncertainties compare the dates in figure 6.10 with the results in table 8.1. The predicted m_∞ values are compared with the retrieved values of m_∞ (with error bars) in figure 6.10; it can be seen that for all but one measurement day in the main experiments (29/11/00), the m_∞ values are not in agreement (do not lie on the line $y = x$). This discrepancy may result from the retrieval of m_∞ not being truly independent, and thus prone to information smearing from other parameters in the refractive index retrieval (such as k leakage), as previously discussed. Fixing the value of m_∞ in retrievals to values predicted by the Krieger *et al.* (2000b) model forced most retrievals into an unrealistic solution in state space, so this parameter was left loosely constrained (see next section).

Analysis of *in situ* measurements of PSCs, discussed in section 2.1.4, suggest that observations are not in good agreement with m_∞ values predicted from theory. As such, the differences between the retrieved m_∞ values and those predicted by Krieger *et al.* (2000b) do not indicate that the method used here is incorrect. Indeed, some of the composition and temperature combinations measured for this work and compared here were outside the stated range of validity of the Krieger *et al.* (2000b) model. These findings strengthen the conclusion of section 2.1.4 – it is clear that further work, involving *in situ*, laboratory and theoretical studies, is required to produce a validated relationship between m_∞ and composition/temperature for liquid stratospheric aerosols.

6.1.7 Typical retrieval configuration

For most of the refractive index retrievals, the same retrieval configuration was used and is summarised in this section. The generic bandset (section 6.1.2) was used as the *a priori* state, with large uncertainties for the size parameters and m_∞, since foreknowledge of these parameters is minimal. For each day of measurements, stable low resolution DLaTGS measurements were selected for retrieval, since analysis of all measured DLaTGS spectra would not have been possible in the timescale of this work. Spectra for each day were analysed sequentially from the coldest to the warmest stable measurement. The generic band set was used as first guess for the band parameters of the coldest spectrum, and for subsequent warmer spectra the converged solution to the previous spectrum was used a first guess; it was advantageous to retrieve the coldest spectra first, since they do not contain significant gas lines which can affect band structure and de-stabilise the retrieval. The *a priori* error constraints and boundaries are listed in table 6.4.

Table 6.4: Refractive index retrieval configuration.

<table>
<tr><th>Parameter</th><th colspan="2">N ($\times10^6$ cm^{-3})</th><th colspan="2">r (μm)</th><th colspan="2">σ</th><th colspan="2">m_∞</th><th colspan="2">$\tilde{\nu}_j$ (cm^{-1})</th><th colspan="2">γ_j (cm^{-1})</th><th colspan="2">S_j (cm^{-2})</th></tr>
<tr><td>*A priori*</td><td colspan="2">1.3</td><td colspan="2">1.0</td><td colspan="2">1.2</td><td colspan="2">1.4</td><td colspan="6">Generic bands</td></tr>
<tr><td>*A priori* error</td><td colspan="2">2</td><td colspan="2">2</td><td colspan="2">2</td><td colspan="2">1</td><td colspan="2">100</td><td colspan="2">1000</td><td colspan="2">1×10^6</td></tr>
<tr><td>Boundaries</td><td>0.001</td><td>1000</td><td>0.001</td><td>2</td><td>1.01</td><td>3</td><td>1.1</td><td>3</td><td>500</td><td>5000</td><td>10</td><td>1000</td><td>0.1</td><td>1×10^7</td></tr>
</table>

Following extensive tests, it was determined that the Jacobian (weighting function) for each state vector element did not need to be calculated every iteration; variation was sufficiently small to allow Jacobian calculation every five iterations resulting in considerable computation time saving.[4] The retrieval could only converge on an iteration if the Jacobian for each state vector element had been calculated in that iteration to ensure the correct solution and uncertainty. The step size, γ, was varied by the retrieval so that an increase (division) factor of 2 and a decrease (multiplicative) factor of 10 was used in order to prevent the retrieval increasing step size too far causing the retrieval to overshoot the minimum on the χ^2 surface; changing the increase factor from 10 to 2 was found to reduce the number of iterations required for the convergence criteria to be met. The convergence parameter, ϵ, was set to 10^{-6}, so that the retrieval converged if the relative change in each state vector element, and the total retrieval cost, was less than this value for ϵ. An upper ceiling of 500 iterations was applied, but all retrievals of measurement spectra converged within this limit. Retrievals were conducted from 750 to 8000 cm^{-1} where measurement noise is reasonable (figure 5.20), with a sampling divisor of 2 – i.e. only alternate sample points were used because of computing limitations.

6.2 Summary

The refractive index retrieval represents the first known use of the CDHO model for retrieval of refractive indices of laboratory generated mimic PSC aerosol particles. This innovative approach was necessary to produce refractive index data for this thesis from single extinction spectra. The other main advantages of this approach include the:

- High accuracy of refractive index retrieval over spectral regions of interest to satellite observations (figure 6.9) – better than 1 % in the imaginary refractive index for most wavenumbers;

- Much faster retrieval of refractive index compared with a Kramers-Kronig analysis, due to the fact that the spectrum can be represented by far fewer state vector elements;

---

[4] Without Jacobian calculation, an iteration typically lasted about one minute, compared with over 30 minutes including Jacobian calculation.

- Lack of a requirement to extend the wavenumber range to infinite limits as with Kramers-Kronig analysis – the resulting edge truncation seen in some sets of data produced using Kramers-Kronig analysis (Biermann *et al.*, 2000) is thus not a problem with the refractive index retrieval used here.

The main problems encountered with the refractive index retrieval were:

- The fit to a particular extinction spectrum is not a unique solution in terms of band parameters. This leads to loss of independence of retrieved variables, but this is included in the comprehensive error analysis;
- The requirement for first guess parameters close to the solution necessitates considerable user-intervention in the initial stages of the retrieval process;
- Significant errors when modelling the asymmetric OH stretch region, and the need to allow for gas absorption in some cases, necessitated the inclusion of forward model error in some wavenumber regions – these were outside the range of principal interest for remote sensing so the resultant increase in refractive index error in these regions is not a major concern.

6.3 Future development

The primary constraint on future development of the refractive index retrieval is the limited understanding of fundamental properties of species in supercooled liquid solutions. Both the Lorentz-Lorenz relation and the CDHO model represent approximations. The existing model could be adapted for characterisation of non-spherical (crystalline) particles by replacing the Mie code with an applicable model, for example the *T-matrix* theoretical model described in literature (Zakharova and Mishchenko, 2001; Baran *et al.*, 2001; Doicu and Wriedt, 2001). Other methods for characterisation of scattering from solid particles also exist (Nebeker *et al.*, 2001). Implementation of such a method is outside the aims of this thesis, but would be of use for future studies (e.g. the EU MAPSCORE project).

Chapter 7

Aerosol Composition

One of the most difficult aspects of laboratory experiments to determine the optical properties of liquid PSCs is the measurement of composition. The analysis detailed in this chapter is concerned with the determination of composition of aerosol measured in the large cell. Since the system is enclosed it was not possible to directly sample aerosol and measure composition *in situ*, and the lack of a mass spectrometer meant that the aerosol composition within the cooled low-pressure environment of the large cell could only be determined from remote measurements.

The approach adopted here for determination of composition is based on the use of the classic damped harmonic oscillator (CDHO) band parameters, which are produced implicitly as the product of the retrieval of refractive indices described in the previous chapter. As previously discussed (section 2.3.5), the strengths of absorption bands are linked to the chemical speciation of the solution. This relationship is examined in more detail in this chapter, and a best-fit calibration between band parameters and chemical speciation is obtained by retrieving band parameters from published refractive index data of binary H_2SO_4 and HNO_3 solutions. This calibration is then used in order to determine the composition of binary and ternary solutions measured for this thesis. The limited data published for ternary solution refractive indices impedes rigorous validation of this method for ternary solution composition determination, although comparison against the limited data available provides some encouraging results.

In order to validate the determination of composition from retrieved band parameters in this work, the high resolution measurements of gas absorption lines were analysed (section 5.2) to determine partial pressures of H_2O and HNO_3. At a given temperature, these vapour pressures can be used to calculate the composition of ternary solution aerosol with an appropriate thermodynamic model, assuming the system is in thermodynamic equilibrium along the measurement path of the optical beam. The low vapour pressures at the coldest temperatures (equilibrium vapour pressure falls with temperature), below approximately 220 K for H_2O and 200 K for HNO_3 gas, could not be measured due to detection limitations of the Fourier transform spectrometer (FTS). Thus, high resolution measurements cannot be used to determine composition at the lowest temperatures equivalent to those in the stratosphere. Only vapour pressure measurements at temperatures above approximately 220 K are used to validate compositions determined from band parameters. The gas vapour pressures of H_2O and HNO_3 are found to be in agreement with the equivalent vapour pressures, calculated with a thermodynamic model, of the compositions determined from band parameters. This validates composition determination from band parameters for a large number of points.

7.1 Determination of speciation from band parameters

As indicated previously (section 2.3.5), each absorption band in the infrared spectrum is related to a particular chemical species in solution. In this section, a relationship between band parameters and composition is established, and calibrated using published refractive index data for binary H_2SO_4 and HNO_3 solutions.

7.1.1 The relationship between speciation and dielectric band parameters

In solution, molecules of sulphuric acid and nitric acid dissociate into ionic species according to:

$$\mathrm{H_2SO_4 + H_2O \rightleftharpoons H_3O^+ + HSO_4{}^-} \tag{7.1}$$

$$\mathrm{HSO_4{}^- + H_2O \rightleftharpoons H_3O^+ + SO_4{}^{2-}} \tag{7.2}$$

$$\mathrm{HNO_3 + H_2O \rightleftharpoons H_3O^+ + NO_3{}^-} \tag{7.3}$$

For purposes of composition evaluation, we are interested in the absorption bands that correspond to these species in spectra. Now, in sulphuric acid, the total number of *sulphate* moles, where molar quantities are denoted by square brackets, is given by $[N_s] = [\mathrm{H_2SO_4}] + [\mathrm{HSO_4{}^-}] + [\mathrm{SO_4{}^{2-}}]$, the total number of *nitrate* and *water* moles are given by $[N_n] = [\mathrm{HNO_3}] + [\mathrm{NO_3{}^-}]$ and $[N_w] = [\mathrm{H_2O}] + [\mathrm{H_3O^+}]$ respectively.

As previously discussed is section 2.3.5, the imaginary parts of the polarisability spectrum, $\alpha''(\tilde{\nu})$, the dielectric constant, $\epsilon''(\tilde{\nu})$, and the refractive index $k(\tilde{\nu})$ all relate to the absorption of the medium, characterised by the band parameters of the CDHO model. As reported by Bertie *et al.* (1994), the areas under the absorbing bands are related to the vibrational intensities of the corresponding chemical species, and therefore the molar quantity in solution. Bertie *et al.* found that the $\alpha''(\tilde{\nu})$ spectrum contains the most accurate information regarding vibrational intensities. Although $\epsilon''(\tilde{\nu})$ and $k(\tilde{\nu})$ spectra yield high accuracy of vibrational intensity in some cases, "anomalous dispersion in the real dielectric constant" (Bertie *et al.*, 1994) introduces an uncertainty that increases with band strength. It is not possible to retrieve polarisability spectra in this work due to constraints arising from lack of knowledge of fundamental physical parameters of solutions (section 2.3.5). Band areas in dielectric spectra are therefore calculated from the imaginary part of complex dielectric spectra, $\epsilon''(\tilde{\nu})$, instead. The uncertainties that arise in such determinations as alluded to by Bertie *et al.*, are quantified in the quality of fit to measured spectra in the refractive index retrieval. These uncertainties are carried over into the error analysis of the composition determination from band parameters.

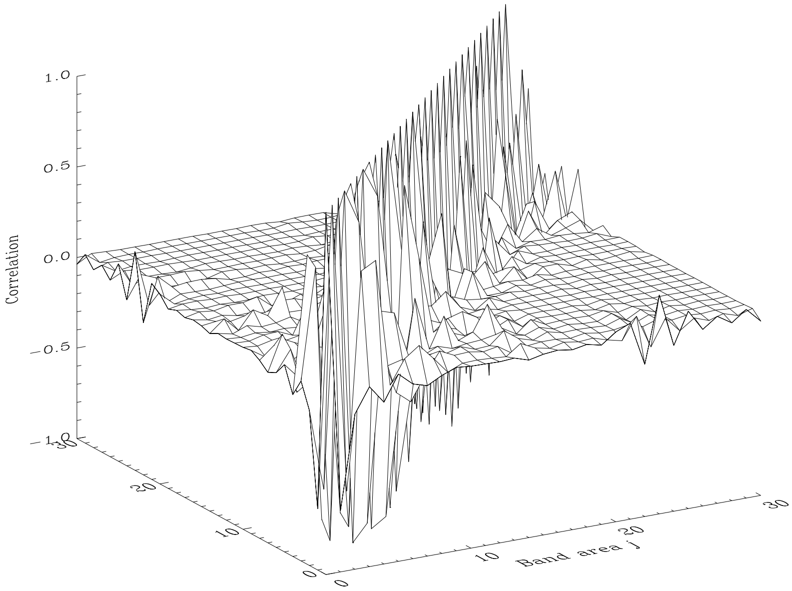

Figure 7.1: Correlation matrix of band areas calculated from the covariance matrix of retrieved band parameters from a Biermann *et al.* (2000) 57 wt% H_2SO_4 solution at 293 K.

There are several factors to consider when relating dielectric band areas to speciation. First, it is important to clarify the physical justification of using the band model with regards to whether each band truly represents a single oscillating species. It has already been demonstrated, in the retrieval of refractive indices from measured extinction spectra, that this is not that case for certain spectral regions (section 6.1.3), such as the OH stretch of H_2O molecules where an asymmetric feature is approximated by several overlapping symmetric bands in the model. This can be accounted for by recognising that each band represents part of the asymmetric H_2O oscillator and including all the overlapping band areas into any calibration involving H_2O molecules. Moreover, the degree of independence of retrieved band areas is important in the assessment of composition. Band areas are evaluated analytically in this analysis; the integral of $\epsilon''(\tilde{\nu})$ (equation 2.28) over all wavenumbers yields the band area, A:

$$A = \int_{-\infty}^{\infty} \epsilon''(\tilde{\nu})\ d\tilde{\nu} = \frac{\pi S_j \gamma_j}{2\left(\gamma_j^2 \tilde{\nu}_j^2 - \frac{\gamma_j^4}{4}\right)^{\frac{1}{2}}} \tag{7.4}$$

The correlation matrix of all band areas is shown in figure 7.1. Note that the band correlations are strongly aligned near the diagonal, indicating that the smearing of information is small and mostly concentrated between adjacent bands; hence summation of adjacent bands leads to the combination of correlated uncertainties, reducing the final uncertainty in the summed band areas which is carried forward into the calculation of composition uncertainty.

Summed areas of selected bands are thus assumed to yield quantities proportional to $[N_s]$, $[N_n]$ and $[N_w]$. The ratio of the summed band areas representative of sulphate or nitrate, is taken against a summed band area of molecular water, to produce an **area ratio**, A_s for sulphate and A_n for nitrate, which is proportional to mole ratio $R_s = [\mathrm{N_s}]/[\mathrm{N_w}]$ or $R_n = [\mathrm{N_n}]/[\mathrm{N_w}]$. Using the band parameters retrieved from the dielectric band retrievals of known compositions of published binary Biermann *et al.* (2000) and UNC refractive indices, the relationship between band area ratio and molar ratio is characterised here by a best fit empirical function.

Selection of bands for composition calibration. The objective of this work is to select bands which allow composition to be determined from band area. In the approach adopted here, only certain bands of particular species are employed to represent total sulphate or nitrate in solution. Although the equilibrium balance of different species that comprise total sulphuric or nitric acid may change with composition, and individual species may have different relationships to band area, these factors are implicit in the empirical calibration of the selected bands against composition so are not of concern for purposes of composition determination.

It is of more importance that the selected bands have minimal temperature depen-

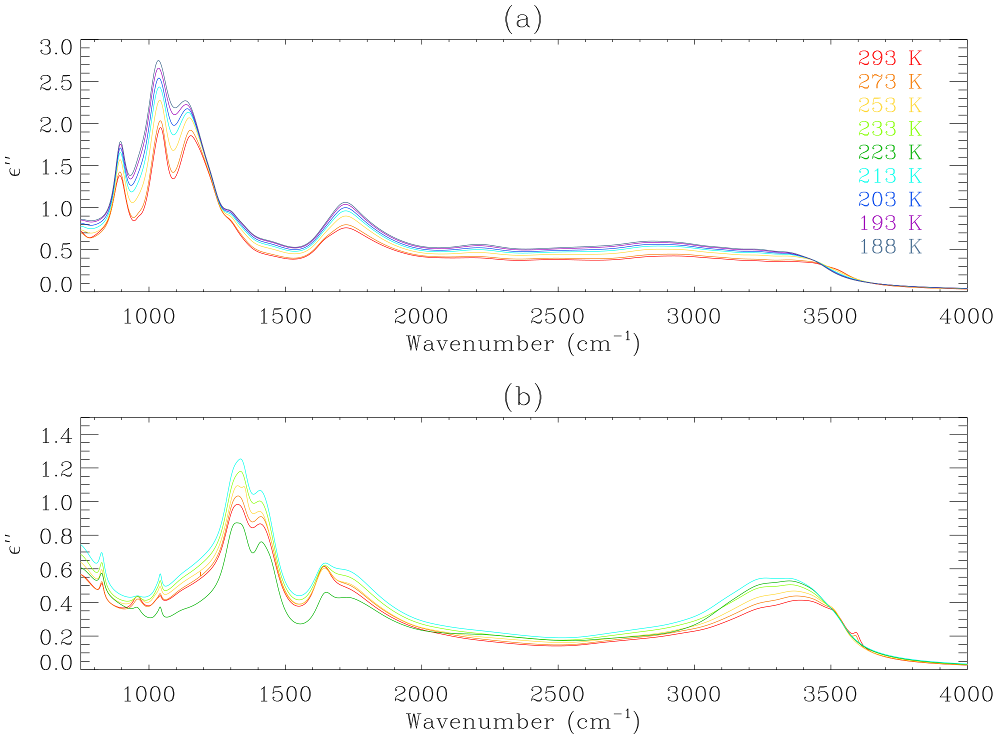

Figure 7.2: Retrieved imaginary dielectric constant spectra of Biermann *et al.* (2000) refractive indices over a range of temperatures. Panel (a) shows 64 wt% binary H_2SO_4 solutions, and 40 wt% HNO_3 solutions can be seen in panel (b).

dence in the calibration of band area against composition. Figure 7.2 illustrates the variation of imaginary dielectric band absorption of Biermann *et al.* (2000) H_2SO_4 and HNO_3 thin-films with temperature for constant composition solutions. Although the temperature dependence of published liquid indices was discussed previously in section 2.5, two points of interest for composition determination can be drawn from the figure presented here: (1) the variation of scaling of imaginary indices with temperature, and (2) the disproportionate (to scaling) variation of the sulphate feature around 1100 cm^{-1} with temperature. The first point is not of major concern with regards to composition determination from band areas since scaling effects are nullified by the ratio calculations of the composition determination. However, the temperature dependence of some sulphate features is important for composition determination, but is circumvented in this work through careful band selection.

Binary solution composition determination. Weight fraction composition for binary solutions can be evaluated using the molecular weight of water (M_w), sulphuric acid (M_s), and nitric acid (M_n):

$$W_{2s} = \frac{[N_s]M_s}{[N_s]M_s+[N_w]M_w} \tag{7.5}$$

$$W_{2n} = \frac{[N_n]M_n}{[N_n]M_n+[N_w]M_w} \tag{7.6}$$

where W_{2s} and W_{2n} are the weight fractions of sulphuric acid and nitric acid in binary solutions. If these relations are divided through by $[N_w]$, then weight fraction can be related to mole ratio, such that in the case of x species:

$$W_{2x} = \frac{R_x M_x}{R_x M_x + M_w} \tag{7.7}$$

$$R_x = \frac{M_w W_{2x}}{M_x(1 - W_{2x})} \tag{7.8}$$

Hence, it is straightforward to relate the compositions from the published binary refractive index data sets to molar ratios for the evaluation of the composition/band area relationship.

Ternary solution composition determination. Similarly, the composition of ternary solutions can be evaluated from:

$$W_{3x} = \frac{[N_x]M_x}{[N_n]M_n + [N_s]M_s + [N_w]M_w} \tag{7.9}$$

$$= \frac{R_x M_x}{R_n M_n + R_s M_s + M_w} \tag{7.10}$$

where x denotes water, sulphuric acid or nitric acid. The second expression, obtained by dividing through by [H_2O], allows the weight fraction of any component to be calculated in a ternary solution using the mole ratio values R_s and R_n which can be obtained from A_s and A_n from the calibration lines calculated for the binary indices from literature.

Implicit in this relationship is the assumption that speciation does not change in ternary solutions; i.e. the bands selected (representative of specific species) for binary

H_2SO_4 and HNO_3 composition calibration are not affected by additional species in the ternary system. This is implied by the composition calibration adopted by Anthony *et al.* (1997), by the mixing rule used by Biermann *et al.* (2000) in the calculation of ternary refractive indices from binary data, and is supported by the fact that no features specific to the ternary system have been identified in literature (Anthony *et al.*, 1997; Koehler *et al.*, 1992).

Characterisation of the relationship between speciation and dielectric band area. The optimal empirically-determined relationship between the band area ratio and mole ratio for both H_2SO_4 and HNO_3 solutions was found to be linear in terms of the log of mole ratio; a straight line fit, $y = mx + c$, to log mole ratio provided the optimal characterisation:

$$A_x = (m \ln R_x) + c \tag{7.11}$$

for species x where $\ln R_x$ is the fixed variable in a weighted (by retrieved error) straight line fit to the area ratios A_x. When plotted on a linear mole ratio scale this straight line fit thus appears as an exponential function. This log-linear relationship does not have any implications for the arithmetic addition of band areas since the total integrated band area, in imaginary dielectric space, is the linear quantity in the best fit function. A wide range of other relationships were assessed, including linear fits between A_x and R_x (and wt % composition) but, over the range of compositions used here, the adopted relationship provided the best empirical fit to the data points. Comprehensive composition uncertainties were evaluated from all known sources of error, including the quality of fit to published data and the uncertainties in the retrieved band areas.

7.1.2 Use of published binary data for calibration

In order to establish the relationship between dielectric band parameters and composition it is necessary to have knowledge of band parameters for a wide range of concentrations of binary solutions in addition to independent knowledge of solution concentrations. Due to the availability of published refractive index data of binary H_2SO_4 and HNO_3 solutions, and the experimental difficulties of direct determination of aerosol composition and limitations on experimental time in this work, published data for binary solutions were used for the calibration of band parameters against acid concentration for this analysis.

The dielectric band parameters were retrieved directly from published refractive index data of binary H_2SO_4 and HNO_3 solutions using a simplified version of the refractive index retrieval technique described in chapter 6. Essentially, the main change is expressed in the measurement vector which is composed in this case of the real and imaginary parts of the dielectric constant (derived simply from published refractive index data), rather than the measured extinction spectrum. Since Mie calculations are not undertaken in the forward model this retrieval is considerably faster than the refractive index retrieval. In order to maintain consistency with the refractive index retrieval, the same retrieval configuration was used as detailed in table 6.4 for applicable parameters.

The *generic band set* (table 6.1), developed for retrieval of band parameters from binary and ternary solution spectra in chapter 6, was employed here for the retrieval of band parameters from published data. The problem with high residuals, experienced in the refractive index retrieval (section 6.1.3), was encountered over the OH cut-off region once again; hence the same forward model error of ×40 was used over this region (3500-5500 cm^{-1}) as for the refractive index retrieval.

Refractive index data for sulphuric acid and nitric acid thin-films published by Biermann *et al.* (2000) were used for dielectric band retrieval in addition to refractive index data produced from experiments with sulphate and nitrate aerosol by the group based at the University of North Carolina (UNC) (Niedziela *et al.*, 1999, 1998b; Norman *et al.*, 1999). Neither group published refractive indices for directly measured ternary solutions, and the *ternary calculator* of Biermann *et al.* (2000) was not used for the calibration – only the directly measured binary solutions were utilised. These data were reviewed in greater detail previously in section 2.5.

7.1.3 Assessment of methodology

Anthony *et al.* (1997) used a similar technique to that used here for determination of composition. However, the important distinction between the Anthony *et al.* method and that used here is that band areas are assessed from aerosol spectra in absorbance space (a log quantity), instead of dielectric space, by drawing a line between the endpoints of a band and calculating the area of absorption above or below that line. However, the accuracy of the Anthony *et al.* composition determination is limited by three main factors: (1) the scattering due to aerosol particles is neglected when assessing absorption band areas – this may introduce a size distribution dependency that has not been assessed, (2) the influence of adjacent absorption bands on the calculated area of the selected bands is not considered, and (3) the act of "drawing a line" under bands could introduce subjective errors. It is likely that the accuracy of this method could be improved if these factors are quantified.

Three separate regions for calculation of peak areas were selected by Anthony *et al.*; a sulphate region from 990-1260 cm^{-1}, a nitrate region from 1375-1515 cm^{-1}, and the integrated area of the OH band from 2700-3025 cm^{-1}. The ratio of the sulphate area to the OH area and likewise the nitrate area to the OH area was found to have a linear relationship to mole ratio. This relationship does not contradict the speciation/band area calibration in this work, since Anthony *et al.* use a different quantity, absorbance (optical depth), which is a log quantity, instead of the imaginary dielectric constant which is more fundamentally related to vibrational intensity. An accuracy of ± 3 wt% for HNO_3 and H_2SO_4, and ± 6 wt% for H_2O calibration in ternary solution aerosol is estimated. Although this band area/mole ratio calibration was undertaken from room temperature measurements, test cases of known-composition low temperature measurements were not found to deviate significantly from the calibration line. Anthony *et al.* therefore conclude that the effect of temperature on composition calibration is negligible, and the room temperature calibration was used for determination of composition at lower temperatures (approximately 205 K).

Use of retrieved dielectric band areas should provide a more accurate calibration against composition since it negates the two main problems of the approach of Anthony *et al.* (1997). First, the scattering component is no longer a factor since we are dealing with fundamental derived optical constants. Secondly, the fact that the CDHO model retrieves individual bands with specific speciation allows more selective integration of absorption areas pertaining to particular species of interest, reducing and quantifying (through calculated uncertainties) the contribution to absorbance over a region from adjacent bands. Furthermore, as an analytical solution, this method is inherently more robust (not least in the error determination) than using an *ad hoc* area calculation similar to that employed by Anthony *et al.* (1997). Tests were conducted to compare the Anthony *et al.* (1997) band/composition relationship with that formulated for this work; all bands were integrated over specific wavenumber limits in dielectric space. The resulting relationship between band areas and mole ratio did not approach the accuracy of the analytical determination of band areas used in this work.

7.1.4 Sulphate solution calibration

It has been determined (section 2.5.1) that H_2SO_4 solutions exhibit a pronounced temperature dependence in absorbance around 1100 cm^{-1}. This was also one of the principal regions of difference between aerosol spectra measured for the POSTCODE project and the thin-film Biermann *et al.* (2000) data (section 2.5.4). Two particular possibilities can be considered:

- Thin-film measurements could show less temperature dependence than aerosol measurements. Indeed, as noted by Niedziela *et al.* (1999), the thin-film data of Tisdale *et al.* (1998) does not show such a marked temperature dependence as the UNC data;
- The binary thin-film measurements used for the Biermann *et al.* (2000) ternary calculator do not include low temperature measurements of dilute H_2SO_4 solutions since such solutions freeze on the substrate.[1]

---

[1] Hence use of a relatively strong acid concentration in figure 7.2(a).

In any case, for purposes of composition determination from band areas in this work the temperature dependant sulphuric acid features are not included in the calibration.

The temperature dependence of H_2SO_4 solutions (section 2.5.4) is likely to originate from equilibrium shifts from the sulphate (HSO_4^-) to the bisulphate (SO_4^{2-}) ions in solution, causing a pronounced intensification in the bisulphate features at 1030 cm^{-1} and 1050 cm^{-1} with temperature reduction, readily observable in figure 7.2(a). Variation in total intensity of the spectra with temperature is also apparent, which may be due to uncertainties in calculation of temperature dependent film thickness; this is not of major concern for the band area/speciation calibration since the act of calculation of the ratio between sulphate and water bands accounts for these variations.

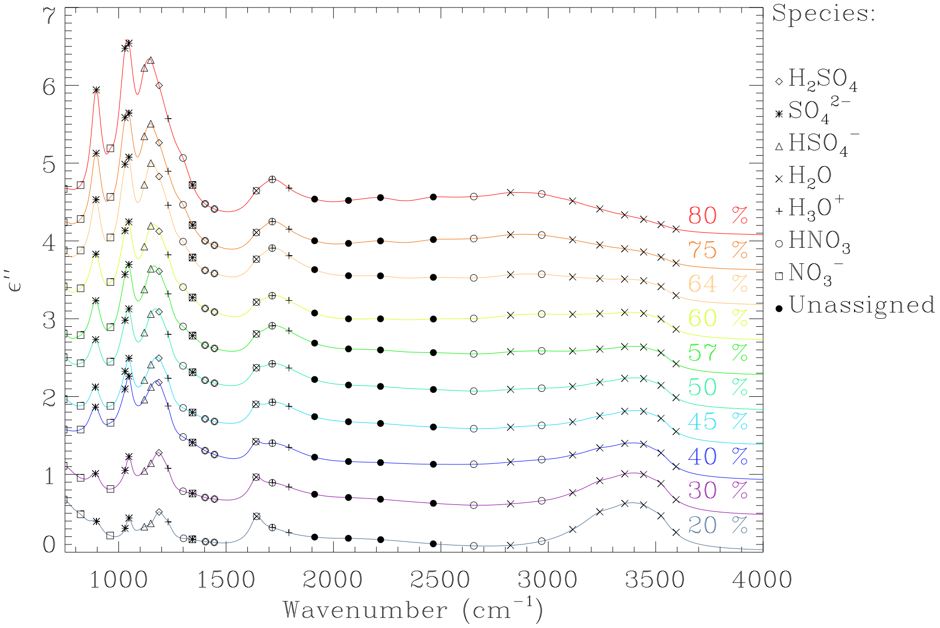

Figure 7.3: Retrieved imaginary dielectric spectra for a range of binary sulphate compositions from Biermann *et al.* (2000) thin-film refractive indices at 293 K. Band centres, $\tilde{\nu}_j$, are denoted by the appropriate species symbol given in key to the right and compositions are listed in the plot. Ambiguous assignments are indicated by overlaid symbols.

The speciation of imaginary dielectric spectra of a range of binary H_2SO_4 solutions is illustrated by the Biermann *et al.* (2000) data presented in figure 7.3 (note that nitric acid bands are included since they exist in the *generic band set*, but the strength of these bands tend towards zero in binary H_2SO_4 band retrievals indicating the applicability of this complete band set to binary data). Calibration of band areas against mole ratio was undertaken by selectively calculating the ratio of summed sulphuric acid band areas to H_2O bands in the OH stretch. The ratios were then related to (log) mole ratio to determine the most linear relationship. It was found that the optimal calibration for H_2O used the summation of all bands from 3000-3500 cm^{-1}; although the peak location of this region shifts to higher wavenumbers with increased dilution, the main area is captured by the overlapping bands in this region. Various bands were found to produce a linear relationship against H_2SO_4 composition when ratioed against the H_2O region at room temperature. However, as can be seen in figure 7.4, when the single $SO_4{}^{2-}$ band (a) at 900 cm^{-1} and the summed $SO_4{}^{2-}$ bands (b) at 1030 cm^{-1} and 1050 cm^{-1} are used for the ratio, pronounced temperature dependence can be seen in the calibration for a 64 wt% solution in the Biermann *et al.* data. The only calibration that does not exhibit temperature dependence, see panel (c) in figure 7.4, uses two $HSO_4{}^-$ bands summed with the molecular H_2SO_4 band around 1200 cm^{-1}. These bands are the best empirical representation of total moles of sulphate, N_s, in the calibration against mole ratio, R_s, and are thus the optimal band configuration for calibration of temperature-independent H_2SO_4 mole ratio.

As discussed previously (section 2.5.1), significant differences exist between the Biermann *et al.* and UNC sulphuric acid refractive indices. Band area/speciation calibration yields incompatible relationships between these datasets, as can be seen in figure 7.5 from the fit to each data set. The band ratio uncertainties from the retrieval appear large in figure 7.5, especially at higher compositions. This is indicative of uncertainty in the retrieved band parameters of the two $SO_4{}^-$ and the molecular H_2SO_4 bands from around 1200 cm^{-1} used in the ratio calculation; errors grow with increasing band strength at higher concentrations of H_2SO_4. It should also be noted that these errors are only an upper estimate since values of the uncertainty in both sets of refractive index data were not published, but the relative magnitudes of errors should be in proportion allowing a representative weighted

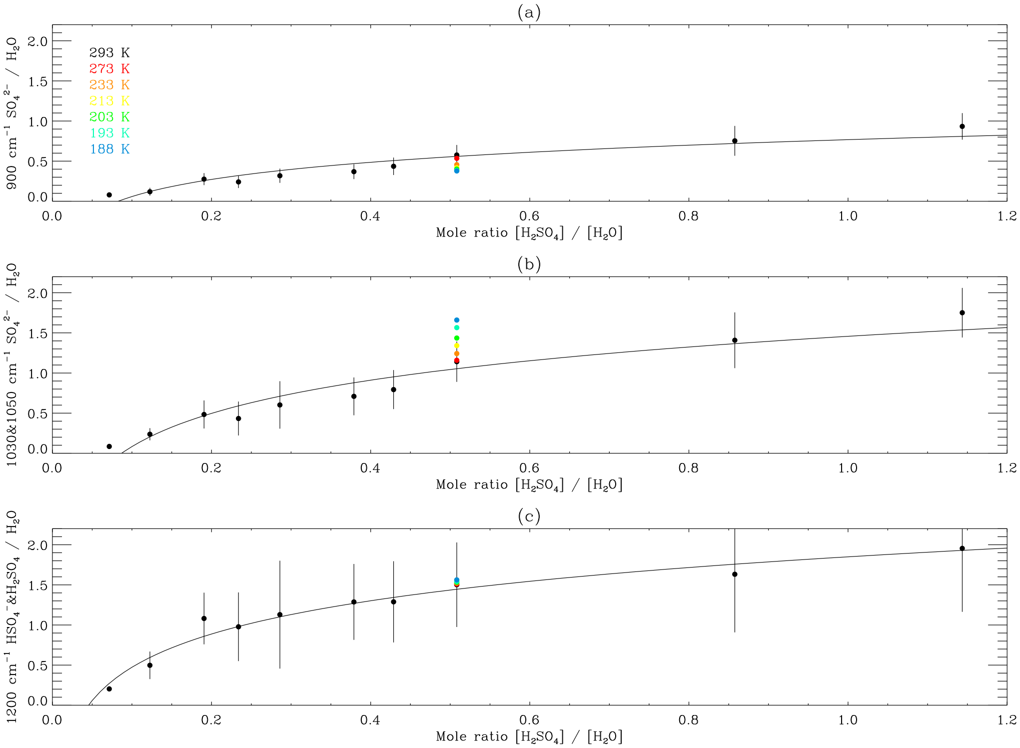

Figure 7.4: Ratio of Biermann *et al.* imaginary dielectric band areas plotted against H_2SO_4 mole ratio; different bands were used for each calibration as described in the text. The straight line fits (log mole ratio) through the 293 K data are displayed along with the band ratios for 64 wt% H_2SO_4 at varying temperatures (in colour and without error bars for clarity).

fit.

There were two main reasons for selecting the UNC calibration line over the Biermann *et al.* calibration for aerosol measurements. First, UNC indices agree much more closely with indices derived in this work, possibly since both data sets are derived from aerosol measurements. Secondly, the UNC calibration agrees within uncertainty with the measured sulphate aerosol composition from the test experiments of 03/11/00 (section 5.2.6) – a composition of 77.0 ± 0.35 wt% was determined using CO_2 retrieved temperature and retrieved H_2O vapour pressure. Relating the band area ratio from the corresponding measured aerosol spectrum to the UNC calibration for a binary sulphuric acid solution yielded a composition of 78.0 ± 4.4 wt% (red point in figure 7.5), in agreement with the inde-

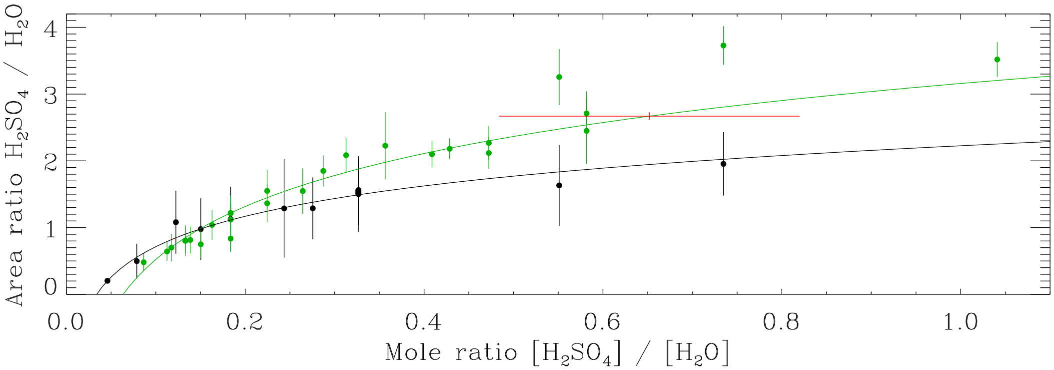

Figure 7.5: Biermann *et al.* sulphate ratio points (black) at 293 K with weighted straight line fit (log mole ratio), and all UNC sulphate ratio points (green) with retrieved error and weighted straight line fit. The red point is the measured composition and error from 03/11/00.

pendently determined value; a corresponding fit to the Biermann *et al.* data produced a composition of 91.5 ± 3.2 wt%, a mismatch clearly outside the uncertainty. Table 7.1 shows compositions determined from band areas using the UNC R_s calibration for indices derived from spectra measured in binary sulphate aerosol experiments on 22/11/00. It can be seen that the uncertainties for most measurements are below 3 wt% and that composition falls with temperature, as is expected. Uncertainty in the mole ratio, as calculated from the band area ratio calibration, is evaluated by independent combination of the weighted straight line covariance matrix with the retrieved uncertainty of the measurement band area ratio; this uncertainty is then propagated through to the final estimated error in wt % composition.

Table 7.1: Binary H_2SO_4 compositions for indices derived from 22/11/00 spectra; UNC R_s calibration line used.

Temperature (K)	H_2SO_4 (wt%)	
198.9	37.2	±2.5
208.3	38.4	±2.4
229.6	43.1	±2.2
285.7	47.3	±2.2

7.1.5 Nitrate solution calibration

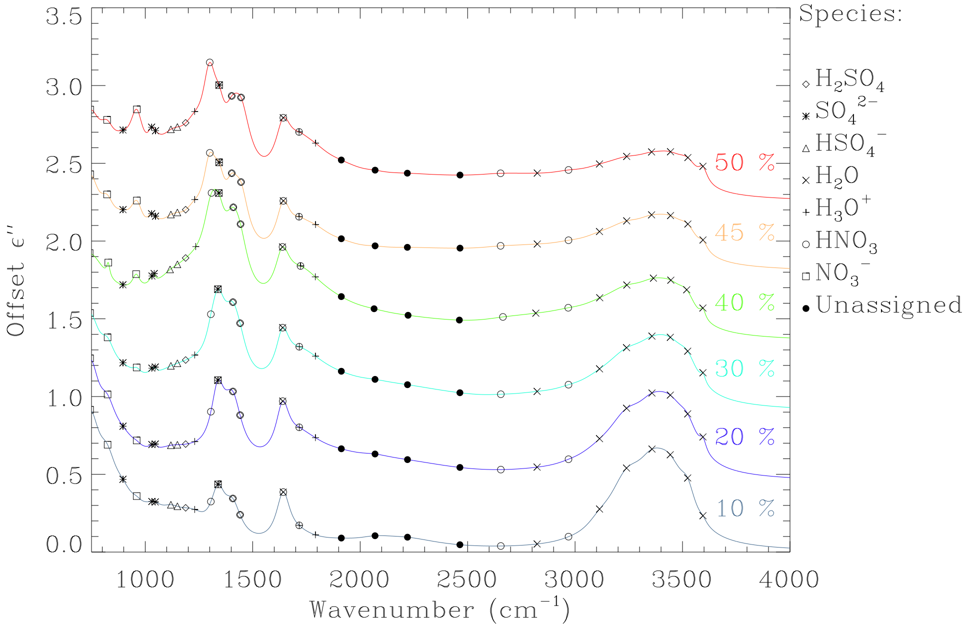

Figure 7.6: Retrieved imaginary dielectric spectra for a range of binary nitrate compositions from Biermann *et al.* (2000) thin-film refractive indices at 293 K. Band centres, $\tilde{\nu}_j$, are denoted by the appropriate species symbol given in key to the right, and compositions are listed in the plot. Ambiguous assignments are indicated by overlaid symbols.

The speciation of imaginary dielectric spectra of a range of binary HNO_3 solutions is illustrated by the Biermann *et al.* (2000) data presented in figure 7.6. The four bands in the spectral region defined by Anthony *et al.* (1997) as the nitrate region, from 1375-1515 cm^{-1} a single $NO_3{}^-$ band, and three molecular HNO_3 bands, were tested separately and in various summations to find the most linear relationship of the ratio against H_2O with (log) mole ratio and the minimal temperature dependence.

As discussed previously in section 2.5.2, binary HNO_3 solutions appear to exhibit a much less pronounced temperature dependence than H_2SO_4 solutions. However, various band area combinations were assessed to ensure minimal temperature dependence of the calibration developed here. Slight temperature dependence was exhibited by various combinations, but the ratio of all four summed integrated band areas against composition was the most linear and least temperature dependent. This is probably because temperature dissociation of nitric acid shifts the equilibrium between molecular HNO_3 and NO_3^-; since both molecules are captured in this four band calibration, wavenumber shifts are balanced through the summation of all areas and this is a good representation of the total moles of nitrate, N_n, in solution. This band selection also has the notable advantage that the Biermann *et al.* (2000) data are in good agreement with the UNC data (unlike the H_2SO_4 calibration).

The indices shown in figure 7.2(b), for 40 wt% HNO_3 over a range of temperatures were used to verify temperature dependence. With the optimal four band calibration, band ratios from all temperatures agreed with the main calibration line except an outlier at 223 K. It can be clearly seen in figure 7.2(b) that the spectrum of 40 wt% HNO_3 at 223 K is inconsistent with the trend of the other data. One other 223 K measurement, of 45 wt% HNO_3, exhibits a similar offset, hence both 223 K indices were excluded from final HNO_3 calibration.

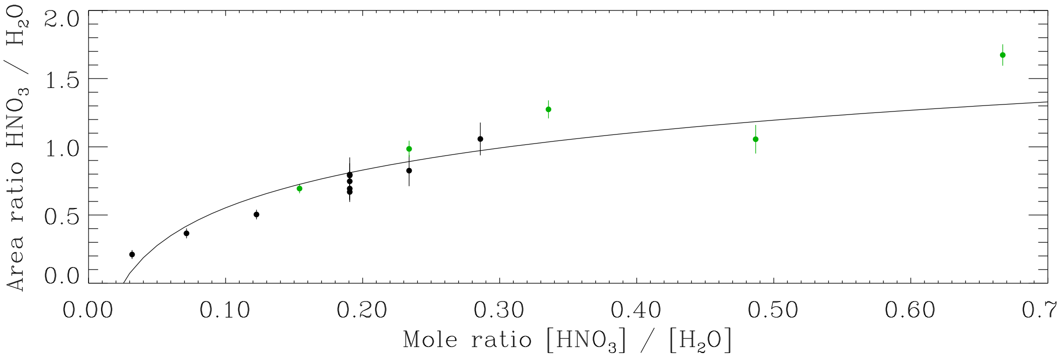

Figure 7.7: Biermann *et al.* nitrate ratio points (black) at all temperatures (except 223 K), and all UNC nitrate ratio points (green) with retrieved errors and weighted straight line fit (log mole ratio) to all points.

The nitrate band ratio composition calibration against all UNC and selected Biermann *et al.* (2000) data is displayed in figure 7.7. It is clear that the error on the ratio values is considerably less than the sulphate calibration, which allows more accurate calibration relative to the lower number of sample points.

7.1.6 Ternary solution calibration

To assess whether the calibration between band areas and composition reliably yields concentrations of the H_2SO_4 and HNO_3 components in ternary solutions, within uncertainties determined in this analysis, compositions were determined for the limited published data, measurements from this work, and data produced from the Biermann *et al.* (2000) ternary calculator.

The only known directly measured refractive indices for a ternary solution are those published for a 75 wt% H_2SO_4, 10% HNO_3, 15% H_2O thin-film system at 298 K by Adams and Downing (1986); uncertainties are not provided with these data. Both the Biermann *et al.* and UNC R_s calibration lines were used to determine R_s from the area ratio values retrieved from this data set. Using the UNC R_s calibration gave a composition of 58.3 ± 4.1 wt% H_2SO_4, 18.9 ± 5.5 wt% HNO_3 and 22.8 ± 2.9 wt% H_2O for the ternary system. The Biermann *et al.* calibration of R_s yielded a composition of 76.7 ± 3.5 wt% H_2SO_4, 10.5 ± 6.4 wt% HNO_3 and 12.8 ± 3.4 wt% H_2O. These results are a further demonstration that the spectral signature of sulphuric acid is different between thin-films and aerosol, with the thin-film Biermann *et al.* (2000) R_s calibration providing a good estimate of the Adams and Downing data here, and the UNC R_s calibration agreeing with independent measurements of binary H_2SO_4 aerosol composition in this work (section 7.1.4).

Ternary compositions calculated using band areas (and the UNC R_s calibration) for indices derived from spectra measured in this work on 29/11/00 are displayed in table 7.2. The high uncertainties on the 263.9 K and the 285.8 K measurements are most likely due to interference of strong gas vapour lines and the low intensity/high uncertainty of the nitrate bands at high temperature, but otherwise the errors are encouragingly small and the values believable.

As a result of the lack of refractive index data produced directly from ternary solution measurements, the ternary calculator of Biermann *et al.* (2000) was used to produce

Table 7.2: Ternary solution compositions for indices derived from 29/11/00 spectra; UNC R_s calibration line used.

Temperature (K)	HNO_3 (wt%)		H_2SO_4 (wt%)		H_2O (wt%)	
196.6	13.0	±2.5	29.6	±0.7	57.5	±2.1
209.7	13.6	±138.0	30.4	±26.9	56.1	±111.1
226.0	16.8	±2.3	31.3	±0.8	51.9	±1.7
247.1	16.4	±2.2	33.3	±0.7	50.3	±1.6
263.9	13.7	±>1000	38.0	±>1000	48.3	±>1000
273.3	14.7	±4.4	58.3	±1.6	27.0	±2.8
285.8	4.5	±>1000	77.0	±>1000	18.5	±>1000

representative refractive indices for testing. Due to the difference in R_s calibration lines of the published data, the Biermann *et al.* calibration for R_s was used in this process. A series of dielectric band retrievals were conducted using the Biermann *et al.* ternary calculator results for fixed 20 wt% and 40 wt% H_2SO_4 and HNO_3 compositions of 40, 30, 20, 10 and 0 wt%, all at 293 K. Results of ternary composition calculation are shown in table 7.3 for the fixed 40 wt% H_2SO_4 case. Since retrieved uncertainties for band parameters, and thus band areas, are high in several cases, the resultant error in the composition determined from the band area calibration is also high. Hence, although the compositions determined from the band area calibration appears to differ considerably with those entered in to the Biermann *et al.* ternary calculator, this comparison is inconclusive at best. This may result from the fact that some of the comparison points are outside the measured composition/temperature regime of the binary indices measured by Biermann *et al.* (2000) for the ternary calculator (section 2.5.5).

Table 7.3: Ternary solution compositions from band areas from the Biermann *et al.* (2000) ternary calculator, using fixed 40 wt% H_2SO_4 and variable HNO_3 composition at 293 K.

Ternary calculator	Composition from band areas			
HNO_3 (wt%)	HNO_3 (wt%)		H_2SO_4 (wt%)	
0	11.5	±13.1	43.1	±3.9
10	11.5	±8.7	28.7	±5.3
20	13.6	±12.6	25.9	±5.4
30	19.6	±15.4	21.4	±6.3
40	23.3	±9.9	24.7	±5.6

7.1.7 Summary

This work has exploited the relationship between absorption bands obtained through refractive index retrieval and composition. Retrieved band areas were empirically related to mole ratio of nitric acid and sulphuric acid to water in the solutions, allowing weight fraction composition to be calculated. The main assumptions in the composition determination were:

- That retrieved band parameters directly relate to the physical speciation in a solution;
- The area of the imaginary constant under specific bands is proportional to the number of moles of that species in solution;
- Speciation does not change between binary and ternary solutions;
- The areas under bands selected for composition analysis vary in proportion to total moles of H_2SO_4, N_s, and total moles of HNO_3, N_n, in solution, and have negligible temperature dependence.

The primary findings of the composition analysis from speciation can be summarised:

- Widely separated retrieved band areas are independent, though some information is shared between adjacent bands – this is the main source of error propagated through to the area ratio uncertainty;
- The optimal empirical relationship between the band area ratio A_x and the mole ratio R_x for species x is a straight line fit in natural log space of mole ratio;
- Careful band selection for composition derivation from speciation is required to avoid temperature dependent effects in H_2SO_4 calibration around 1100 cm^{-1} – this contradicts the findings of Anthony *et al.* (1997);
- Sulphate features around 1200 cm^{-1} exhibit pronounced differences between thin-film and aerosol spectra – the correct calibration line must be used to ensure a realistic

composition is returned. The Biermann *et al.* (2000) N_s calibration correctly yielded the Adams and Downing (1986) thin-film ternary composition, while the UNC N_s calibration agreed with an independently determined binary H_2SO_4 aerosol test measurement.

The bands selected to represent the total moles of sulphate, nitrate, and water in this calibration do not necessarily contain a balanced representation of the species that comprise total sulphate, nitrate or water.[2] However, the empirical relationship developed between band areas and molar ratio implicitly accounts for any change of speciation within the sulphate, nitrate and water systems with variation in composition.

Higher retrieval errors in the sulphate area ratio values are offset somewhat in the calibration by relatively large number of calibration points compared with the nitrate calibration. However, both the sulphate and nitrate calibration lines could be improved by more measurement points (from published data). Furthermore, the absence of errors for published refractive index data **and** for published compositions leads to some uncertainty over the error in the weighted calibration line. It was also found that the calculated error on composition is highly dependent on the quality of the refractive index retrieval, which is particularly poor in a minority of cases where, for example, significant gas absorption lines are present in published data at high temperatures.

In spite of these problems, the fact that the ternary calibration returns the composition of the ternary indices of Adams and Downing (1986), and that the composition of binary aerosol measured in this work is correctly identified represents tentative validation of the determination of composition from band areas. Fundamental and apparently universal differences between thin-film and aerosol liquid sulphuric acid spectra are identified in absorption features around 1200 cm^{-1}. The origin of these differences is unknown, but may be a result of chemical interactions of sulphuric acid solutions with silicon substrates used in thin-film experiments. Further independent validation of the calibration developed here

---

[2] For example, in the sulphate case, it was necessary to omit the temperature dependant ${SO_4}^{2-}$ bands around 1100 cm^{-1}.

is desirable. This is achieved in this work through measurement of H_2O and HNO_3 vapour pressures from high resolution spectra of gas lines, as detailed in the following section.

7.2 Validation of aerosol composition determination through vapour pressure measurements

7.2.1 Methodology

The partial pressures of H_2O and HNO_3 determined through the gas retrieval (section 5.2) are used here to validate the composition determined from band parameters obtained in the refractive index retrieval. Unique values for the partial pressure of HNO_3 and H_2O in the presence of any ternary solution composition can be evaluated over the relevant temperature range of 180 to 270 K using the thermodynamic model of Lin and Tabazadeh (2001), assuming thermodynamic equilibrium. Aerosol compositions determined from retrieved band parameters are entered in to the Lin and Tabazadeh thermodynamic model to evaluate what is termed here as **speciation partial pressures** – the equivalent HNO_3 and H_2O partial pressures above aerosol particles. The Kelvin effect, which acts to increase vapour pressure above small particles, was found to be negligible for the size range of aerosol in the large cell and is ignored in this analysis. Uncertainties in composition (determined from band parameters) and temperature values were propagated through the analysis to produce equivalent uncertainties for the speciation partial pressures.

7.2.2 Validation

Partial pressures retrieved from high resolution gas spectra recorded immediately before the low resolution aerosol measurements used for composition determination from band parameters, were compared with the speciation partial pressures for all measurements on all days in the main experiments where an accurate ($<$ 10 wt% error) composition was determined from band parameters and both HNO_3 and H_2O partial pressures could be measured with the FTS (i.e. generally at temperatures above 220 K, as equilibrium vapour

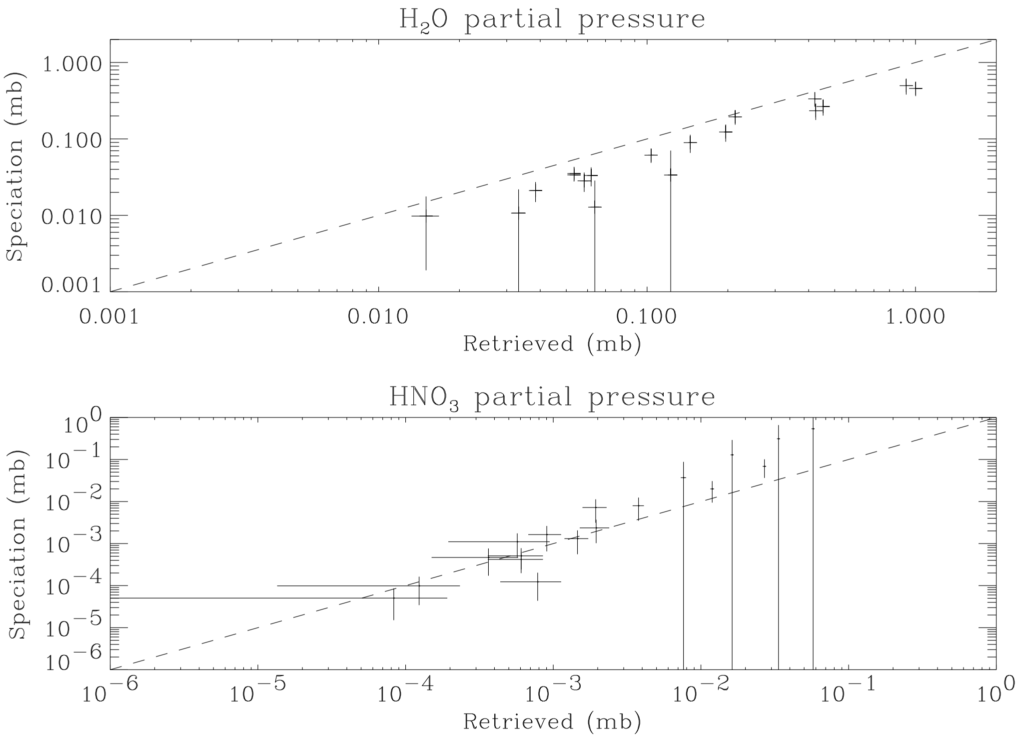

Figure 7.8: Partial pressure of H_2O and HNO_3 from the gas retrieval plotted against values determined by calculating equivalent partial pressure from composition determined from band parameters using the thermodynamic model of Lin and Tabazadeh (2001). The dashed line represents $y = x$.

pressure falls with temperature). Figure 7.8 shows the retrieved gas partial pressures plotted against the speciation partial pressures. If both data sets are in agreement the points should lie along the straight line $y = x$ within the error bars. This is the case for HNO_3, but H_2O partial pressures agree less well. As discussed previously in section 1.3.3, published thermodynamic models do exhibit some differences. The alternative thermodynamic model of Carslaw *et al.* (1995) was also used to calculate the equivalent partial pressures from the composition determined from band parameters (the speciation partial pressure). The resultant change in the speciation partial pressures brought the H_2O partial pressures to much better agreement, with HNO_3 partial pressures now agreeing less well. Therefore, it can be concluded that the composition measurement from retrieved band parameters has been validated, to within the constraints of the published thermodynamic models of the ternary system, by independent gas phase measurements of H_2O and HNO_3.

Chapter 8

Overview of Results

The results of this thesis are, to some extent, integrated with the previous chapters which focussed on characterisation of the experimental system and determination of aerosol composition. In this chapter, the relevant results are collated to present a complete overview of the development of aerosol in the large cell, and representative low-temperature refractive indices are presented. The relevance of this analysis to the stratospheric system is discussed, and indices are compared with published data to establish the contribution to scientific knowledge of this work. Finally, the loop is closed with stratospheric detection of PSCs, and differences between data-sets are discussed in terms of the volume and surface area retrievals of PSCs from remote sensing in the infrared.

8.1 Temperature dependence of sulphate features in ternary solution aerosol

Temperature dependence of the ${SO_4}^{2-}$ bands near 1100 cm^{-1} was discussed in the context of the determination of composition from speciation in section 7.1.4. Imaginary refractive indices over a range of temperatures for the ternary solution measured on 21/03/00 (during the settling experiments) are plotted in figure 8.1; an increase in absorption can be clearly seen at lower temperature – this is not due to a change of composition, but results from the temperature dependence of chemical speciation of H_2SO_4 solutions. The output from the Biermann *et al.* (2000) ternary calculator composition match (12.0 wt% HNO_3, 20.9 wt% H_2SO_4) at 190 K, as calculated for the POSTCODE analysis (section 2.5.4), is plotted for comparison. Due to the fact that the nearest H_2SO_4 solution near this composition in the measured Biermann *et al.* binary H_2SO_4 data base was measured at 253 K, the temperature dependence of this feature does not appear to be correctly reproduced by the Biermann *et al.* thin-film data.

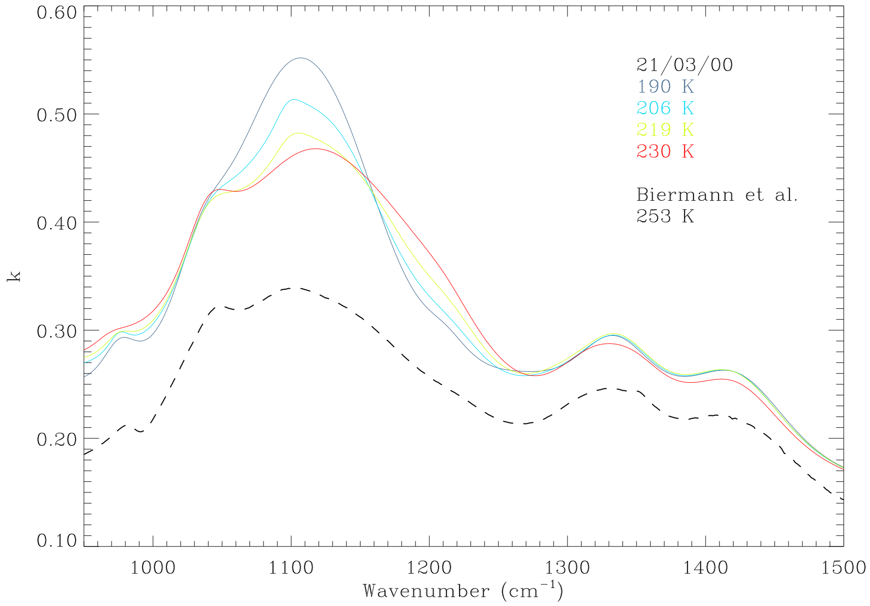

Figure 8.1: Temperature dependence of $SO_4{}^{2-}$ bands near 1100 cm^{-1} in the imaginary refractive index from indices derived on 21/03/00 (settling experiments). Output from the Biermann *et al.* (2000) ternary solution calculator for the closest match is overplotted; the limitations of the parent data base limit the closet H_2SO_4 measurement to 253 K, explaining the poor qualitative fit to the 1100 cm^{-1} region at low temperature.

8.2 Parameterisation of aerosol in the large cell

Figure 8.2 shows a series of results compiled for the **stable** ternary solution aerosol measurements from the main experiments, for six out of the seven experiment days (two were very similar). Compositions and uncertainties have been calculated from the band area/speciation calibration as described and validated in chapter 7. It can be seen that the errors bars for some compositions are very large – this is due to high errors propagating through from the retrieved band parameters. The underlying causes of such high uncertainties include spectral interference from gas lines, and poor fitting of specific regions due to the non-unique nature of the band parameters. Cases where a high degree of spectral interference from gas components was experienced (near room temperature) were omitted in figure 8.2. The size distribution, retrieved during derivation of refractive index data, is utilised to calculate aerosol volume density in the cell, and is plotted for each day; associated volume uncertainties are too small (generally less than 1 %) to be visible at the scale used.

The results presented from ternary experiments in figure 8.2, and from the binary H_2SO_4 measurements in figure 8.3, reveal the change of aerosol composition with cooling. Observed aerosol evolution is broadly in agreement with the simple aerosol evolution model of the large cell, described previously in section 5.1.3. Thus, the condensation of H_2O and HNO_3 vapour takes place as temperatures are reduced (maintaining thermodynamic equilibrium), leading to the observed variation in composition and increase in aerosol volume with reduction of temperature. As revealed in the output of the aerosol evolution model (section 5.1.3), the high number density of aerosol in the cell, compared to the stratosphere, leads to condensation of volatiles and a subsequent temperature-dependant variation of aerosol at a much higher temperature, generally above 240 K, than the stratospheric regime (between about 185 to 200 K). Below about 240 K, composition varies only slightly with temperature in the cell. However, as observed from the 21/11/00 experiment, aerosol volume can increase significantly as temperatures are reduced below about 250 K, despite a relatively small corresponding change in aerosol composition. Again, this is in agreement with the results of the aerosol evolution model (see figure 5.3), and can be explained by condensation of gas phase volatiles that co-exist in thermodynamic equilibrium with the aerosol. The pronounced decrease in retrieved volume below 209.1 K on 30/11/00 was due

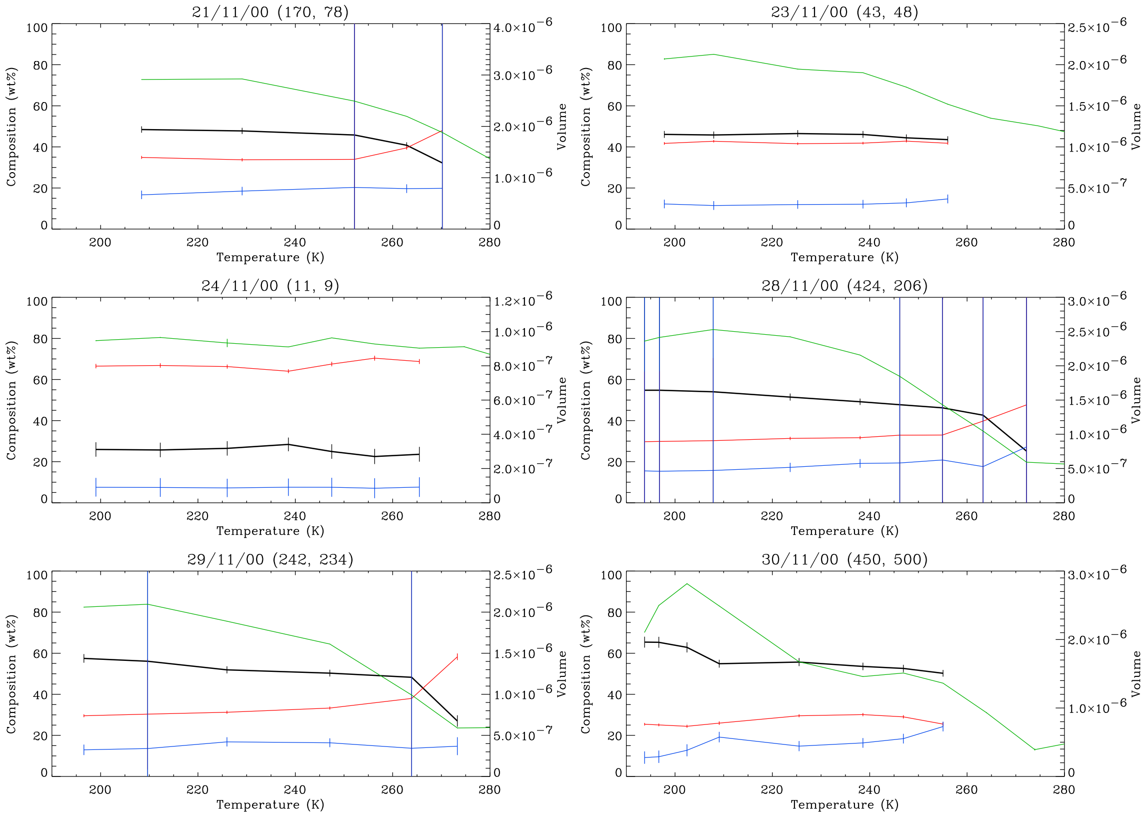

Figure 8.2: Results from analysis of stable ternary solution aerosol spectra measured during the main experiments. The date is given in the title with the flow rates [sccm] of HNO_3/N_2 and H_2O/N_2 respectively listed in parentheses; all experiments were conducted with a SO_3/N_2 flow rate of 400 sccm. The thick black line is the total water composition, and the red and blue lines are H_2SO_4 and HNO_3 compositions [wt%] respectively. The green line is the aerosol volume density (dimensionless; scale on the right axis).

to the freezing of aerosol, as discussed later (section 8.4); the representative cold liquid ternary solution aerosol measurement for this day was thus recorded at 209.1 K.

8.3 Refractive indices of liquid binary and ternary aerosol

Refractive index spectra have been retrieved through the band model refractive index retrieval for all of the coldest stable spectra measured in the settling and main experiments, and for all stable spectra from the main experiments. Prior to comparison with published data and investigation of relevance to the stratospheric system, results are presented to

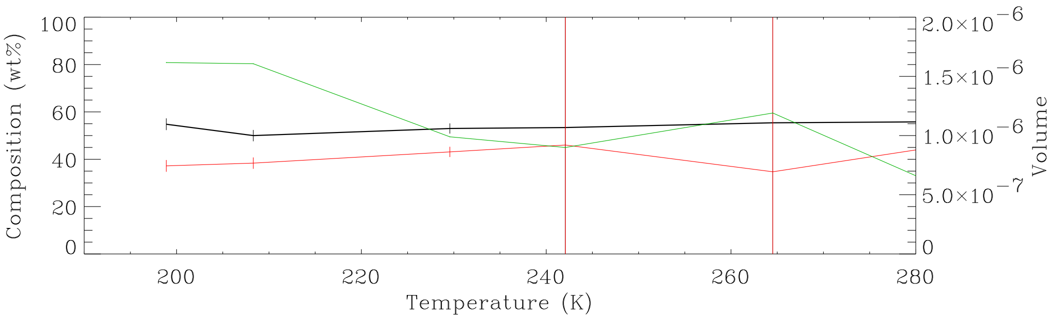

Figure 8.3: Results from analysis of stable binary H_2SO_4 aerosol spectra measured during the main experiments on 22/11/00. Flow rates were 400 and 205 sccm for SO_3/N_2 and H_2O/N_2 respectively. The thick black line is the total water composition, and the red line is the H_2SO_4 composition [wt%]. The green line is the aerosol volume density (dimensionless; scale on the right axis).

illustrate variation of absorption features over a composition range representative of that spanned by these experiments, and uncertainties in derived index data are displayed.

Retrieved refractive index data for four compositions are displayed in figure 8.4; these were derived from cold stable spectra from separate measurement days over a range of stratospherically-relevant compositions as indicated in the annotation. Uncertainties are small for most of the spectral range, instilling high confidence in these data. Slightly higher uncertainties can be seen towards the infrared cut-off for all the indices from the main experiments, due to the errors inherent in the correction of BaF_2 window temperature dependence. This process was not undertaken during the measurement from the settling experiments (24/03/00), hence the slight disagreement between these indices below 1000 cm^{-1} and the smaller errors; this should be taken into account in use of these data.[1] The upper spectral range is truncated at 3500 cm^{-1} due to the problems with fitting the OH-cut off. However, the range 750-3500 cm^{-1} contains absorption features of the major chemical species, and can be used to characterise composition and type of aerosol.

From the imaginary refractive indices presented in figure 8.4, a clear transition of absorbance features can be seen with change in composition in the liquid supercooled aerosol.

---

[1] Since initial background measurements are recorded at 230 K in the settling experiments, the effects of BaF_2 window temperature dependence are not as acute as experienced in the main experiments.

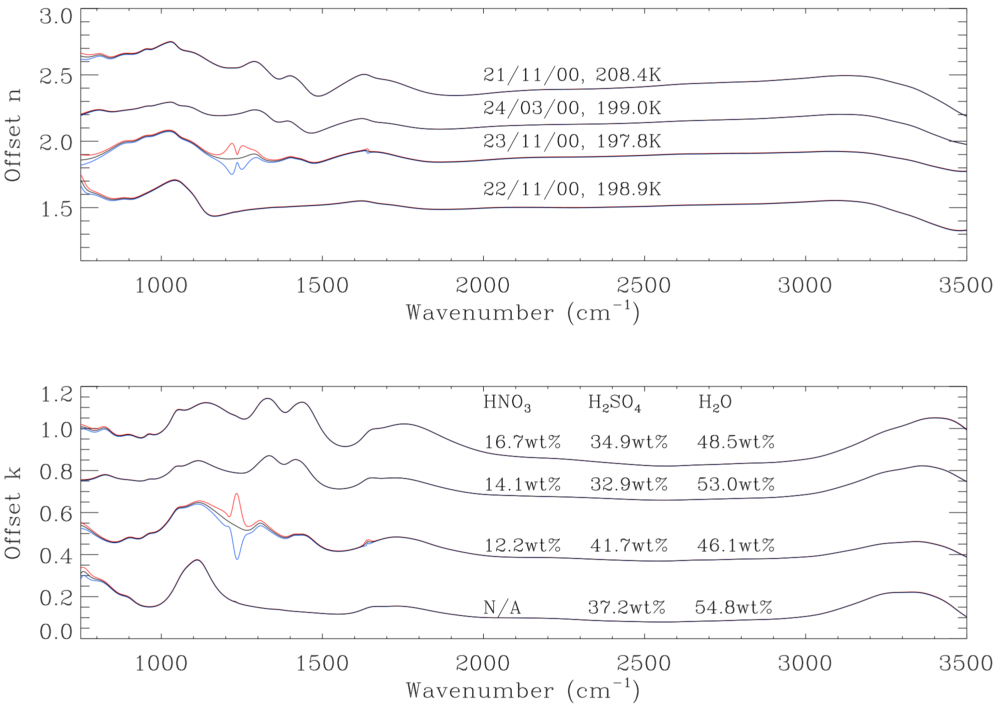

Figure 8.4: Refractive indices derived from the coldest stable aerosol spectra on the experiment days as indicated in the annotation, along with the calibrated temperature and composition of the measurements (see table 8.1 for uncertainties in these values). The real and imaginary parts of the refractive index are displayed in the upper and lower panels respectively, and are offset for clarity. The black lines represent the retrieved index, with the red and blue lines indicating the upper and lower uncertainties respectively.

The binary H_2SO_4 aerosol indices are a useful indicator of sulphate only features, and the other three indices show the effect of increasing HNO_3 composition on nitrate features. The slight variation in measured temperatures is unlikely to contribute significantly to differences in absorption features.

8.4 Solid aerosol

As noted from the aerosol composition and volume data for 30/11/00 (figure 8.2), this experiment produced very different spectral features compared with those previously observed. It was recognised that these newly observed absorbance features were very similar to those of water ice by visual comparison with the aerosol measurements of Clapp *et al.* (1995), of the UNC group. Therefore, it was inferred that aerosol was freezing into pure water ice. This experiment configuration was repeated the next day, 01/12/00, over a wider range of temperatures in order to identify the freezing point. Extinction over the temperature range 207.9-199.8 K is plotted in figure 8.5 to highlight the transition from a typical liquid ternary solution aerosol to solid aerosol; the freezing point was estimated at around 203 K from the observed spectral transition.

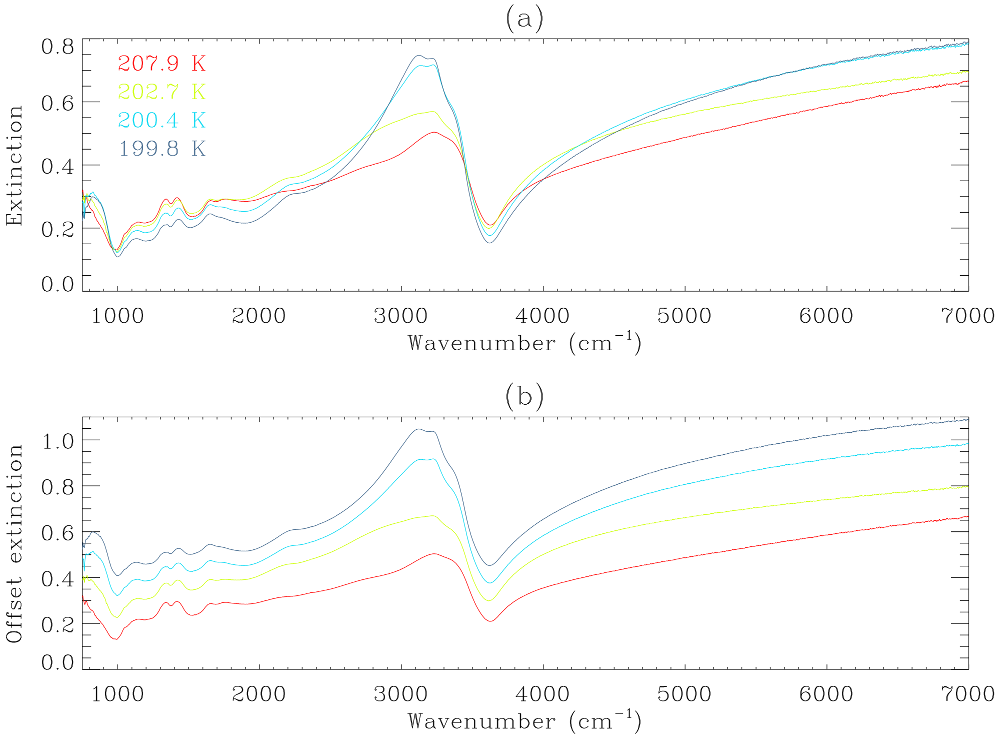

Figure 8.5: A series of stable extinction spectra recorded during cooling on 01/12/00. Gradual transition to a spectrum more resembling that of water ice can be seen as temperatures cool. Spectra are offset in panel (b) for clarity.

At each stable temperature stage plotted in figure 8.5 extinction spectra were observed to stabilise with time, even though aerosol freezing appears to have been initiated. Although the vapour pressure over pure H_2O ice is much lower than over a solution of pure liquid H_2O, the H_2O vapour pressure is also lower over ternary solutions. At the temperature and composition calculated for the coldest measurement from 30/11/00 in table 8.1, and including uncertainties from calibrated temperature and composition speciation, the H_2O vapour pressure over pure H_2O ice is $(6.30 \pm 1.90)\times10^{-3}$ mb and over the ternary solution the H_2O vapour pressure is $(4.13 \pm 1.37)\times10^{-3}$ mb. These calculations were undertaken using the empirical relation for water ice from Marti and Mauersberger (1993) and the thermodynamic model of Lin and Tabazadeh (2001) – the Kelvin effect is ignored since it is negligible for particles of sizes observed here. The H_2O vapour pressures agree within uncertainties, indicating it may be thermodynamically possible for ternary solution aerosol and pure ice aerosol to coexist inside the cell under conditions immediately prior to the observed partial freezing of aerosol.

Refractive indices derived from the coldest spectrum plotted in figure 8.5 are compared with the pure H_2O ice aerosol refractive indices of Clapp *et al.* (1995) in figure 8.6. It can be seen that even after several hours of evolution, the spectral signature of STS remains in our spectra superimposed on the signature of pure H_2O ice; we appear to have a mixture of pure H_2O ice and liquid STS aerosol in equilibrium, in agreement with the thermodynamic calculations. As the cell temperature is reduced, the equilibrium balance shifts towards water ice, as can be seen in figure 8.5. However, the aerosol density was observed to fall with time at the lowest temperatures achievable in the cell. The experiment of 01/12/00 was conducted until the aerosol disappeared completely from the cell. This could be explained if the freezing of aerosol coincided, or led to, freezing of condensates on the cell wall – this could preferentially scavenge volatiles from the cell atmosphere at low temperature due to a change in balance of volatile vapour pressures. It was determined that condensates had not solidified on the cell windows by observing the changing extinction spectra when the cell contents were evacuated at the end of the 30/11/00 experiment; the ice features disappeared at the same rate as the STS features.

The possibility that some form of crystalline hydrate had formed was investigated, but discounted since none of the sharp absorption features common to nitric and sulphuric acid hydrates were observed; the remote possibility remains that a high hydrate (i.e. above

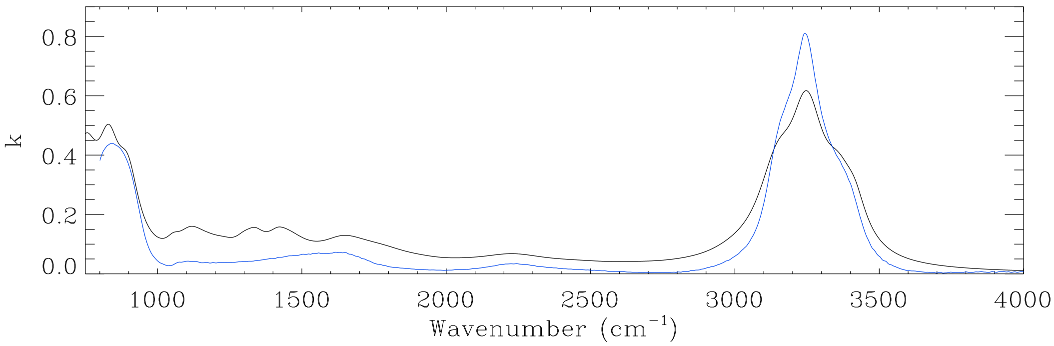

Figure 8.6: Comparison of imaginary refractive indices from the 199.8 K measurement (black) on 01/12/00 compared with indices of pure ice aerosol at 200 K (blue) as measured by Clapp *et al.* (1995).

1:3 HNO_3/H_2O) of nitric acid formed, for which there is no previously published data for identification.

It is probable that aerosol froze in these experiments due to the relatively dilute nature of the initial solution. Although the total H_2O composition was higher on 29/11/00 and similar on 28/11/00 (table 8.1), the H_2SO_4 component was lowest during the experiments where freezing occurred; this concurs with thin-film experiments (Biermann *et al.*, 2000) where the freezing point is observed to be warmer for stronger H_2SO_4 compositions. Further experiments are required to investigate the composition dependent degree of supercooling required to nucleate solid particles from ternary compositions of stratospheric relevance.

8.5 Relevance of results to the remote sensing of PSCs

The nature of the temperature dependent composition regime of ternary solution aerosol (type Ib PSCs) in the stratosphere was discussed previously in section 1.3.3; figure 1.7 shows the transition from almost binary H_2SO_4 aerosol to near binary HNO_3 aerosol with a few degrees of cooling around 193 K. The same range of compositions for stratospheric ternary aerosol from the thermodynamic model of Lin and Tabazadeh (2001) are plotted in figure 8.7, indicating the effect of the likely stratospheric variation of HNO_3 and H_2O

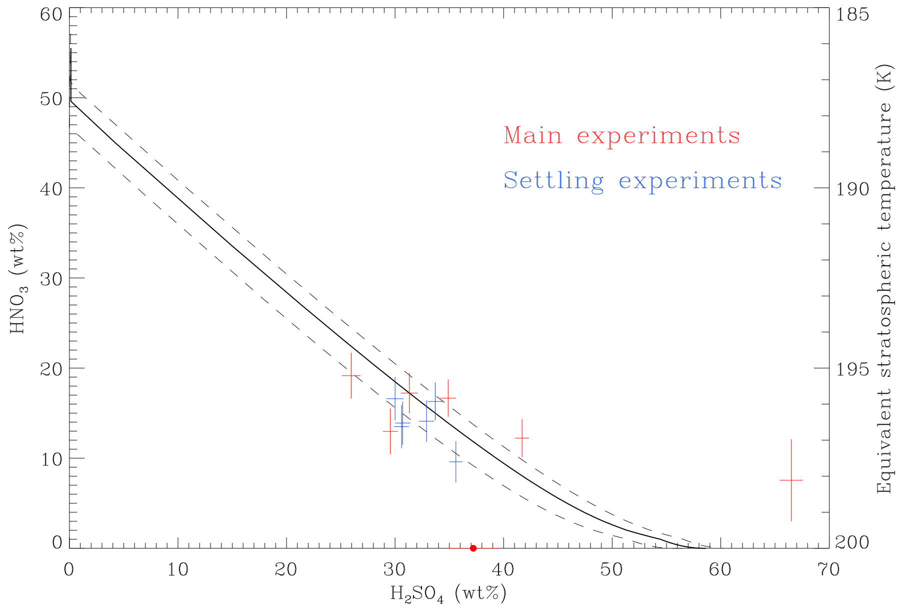

Figure 8.7: Results from the thermodynamic model of Lin and Tabazadeh (2001) showing the range of ternary solution compositions in the stratosphere; the **equivalent temperature in the stratosphere** used for this calculation is shown on the right axis. The thick black line represents typical stratospheric HNO_3 and H_2O vapour concentrations of 5 ppmv and 10 ppbv respectively, and the dashed lines represent likely variation of ± 50 % in H_2O and ± 20 % in HNO_3. The compositions and uncertainties for the coldest stable aerosol measurements from the settling and main experiments are plotted.

partial pressures (Carslaw *et al.*, 1997). Compositions (and uncertainties) for the coldest stable measurements from each experiment day are also indicated in figure 8.7, and are listed chronologically in table 8.1. It can be seen that most measurements lie within the mid-range of STS compositions expected in the stratosphere. Temperatures of the laboratory measurements are universally higher than equilibrium temperature of particles of similar composition in the stratosphere. This difference is of order 10 K, but represents a considerable improvement over previously published data from thin-films and binary aerosol at higher temperatures. The range of sub-20 wt% HNO_3 composition ternary solution compositions is therefore well characterised. The outlier compositions measured on 22/11/00 and 24/11/00 are of use in terms of understanding the composition dependent features of aerosol and characterising indices for possible compositions of unstable aerosol in the stratosphere.

Table 8.1: Stable compositions and temperatures of the coldest liquid aerosol measurements measured during the settling and main experiments. These data are plotted in figure 8.7.

Date (2000)	Temperature (K)		HNO_3 wt %		H_2SO_4 wt %		H_2O wt %	
Main experiments								
30/11∅	209.1	±2.1	19.2	±2.6	26.0	±0.9	54.9	±1.9
29/11	196.6	±2.1	13.0	±2.5	29.6	±0.7	57.5	±2.1
28/11†	223.7	±2.1	17.2	±2.3	31.3	±0.8	51.5	±1.7
28/11	193.8	±2.1	15.5	±155	29.7	±34	54.7	±121
24/11	199.0	±2.1	7.6	±4.6	66.5	±1.1	25.9	±3.5
23/11	197.8	±2.1	12.2	±2.1	41.7	±0.6	46.1	±1.7
22/11	198.9	±2.1	N/A	N/A	37.2	±2.5	54.8	±2.5
21/11	208.4	±2.1	16.7	±2.1	34.9	±0.7	48.5	±1.6
Settling experiments								
29/3	200	±10	13.5	±2.4	30.6	±0.7	55.9	±2.0
28/3	200	±10	13.9	±2.4	30.7	±0.7	55.4	±2.0
27/3	200	±10	16.6	±2.4	30.0	±0.8	53.4	±1.8
24/3	200	±10	14.1	±2.3	32.9	±0.7	53.0	±1.8
23/3	200	±10	16.3	±2.1	33.7	±0.7	50.0	±1.6
21/3	200	±10	9.6	±2.3	35.6	±0.6	54.8	±1.9

† High uncertainties in composition at the lowest temperature, however the composition is consistent and believable within the errors of the higher temperature measurements. The higher temperature composition was plotted in figure 8.7.

∅ High temperature used since the aerosol froze below this temperature.

In terms of the remote sensing of ternary solution PSCs, the compositions at the cold end of the ternary solution range, i.e. with more concentrated HNO_3 and weaker H_2SO_4, would probably be of most interest since these are the most likely liquid PSCs to be detected because the volume of STS particles increases considerably towards the ice nucleation temperature (Beyerle *et al.*, 1997). In order to establish whether the refractive index data derived in this work represent an advance, it is useful to compare data against those produced from the only source of published low-temperature ternary solution data, the Biermann *et al.* (2000) ternary refractive index calculator. The composition/temperature range of binary solution data used by the ternary calculator was reviewed previously; see figure 2.10 which is of the same format as figure 8.7. Given the temperature dependence of sulphuric acid established from literature in section 2.5.4, and now from the results of this thesis in section 8.1, it appears that in terms of the characterisation of H_2SO_4 features, the Biermann *et al.* data are limited to H_2SO_4 compositions above 30 wt% at stratospheric temperatures. Hence, in terms of the high-HNO_3 / low-H_2SO_4 concentration range most of interest for remote sensing of STS aerosol, this work may represent a slight advantage over the ternary calculator of Biermann *et al.*. However, both datasets are not particularly close to this regime, but these data are complementary in characterising the mid-range of ternary aerosol compositions found in the stratosphere.

8.6 Comparison of refractive indices with published data

8.6.1 Comparison of ternary refractive indices

Figure 8.8 shows a comparison of the coldest stable ternary solution aerosol imaginary indices from (a)29/11/00 and (b)23/11/00, and indices from the Biermann *et al.* (2000) ternary calculator for the same compositions and temperatures (see table 8.1 for these parameters). Indices are displayed from 750 to 3500 cm^{-1}, the valid range for refractive indices derived in this work. Significant differences are apparent. First, there appears to be a scaling difference across the infrared – it is difficult to attribute this to either method used for refractive index calculation. However, as previously noted in section 2.5.2, the Biermann *et al.* and UNC group indices also exhibit significant scaling differences which, as previously speculated, could be due to uncertainties in the thin-film thickness calculation of Biermann *et al.* (2000).

Secondly, it can be seen that while the general absorption features of the $k(\tilde{\nu})$ index appear to qualitatively agree in figure 8.8(b), significant differences exist in the comparison in figure 8.8(a). These ternary solution measurements were selected for comparison since they represent the extremes of the composition range measured at sub-200 K, as can be seen by inspection of figure 8.7 and table 8.1. As concluded in the review of the Biermann *et al.* data in section 2.5.5, these calculated data may not be applicable at stratospheric

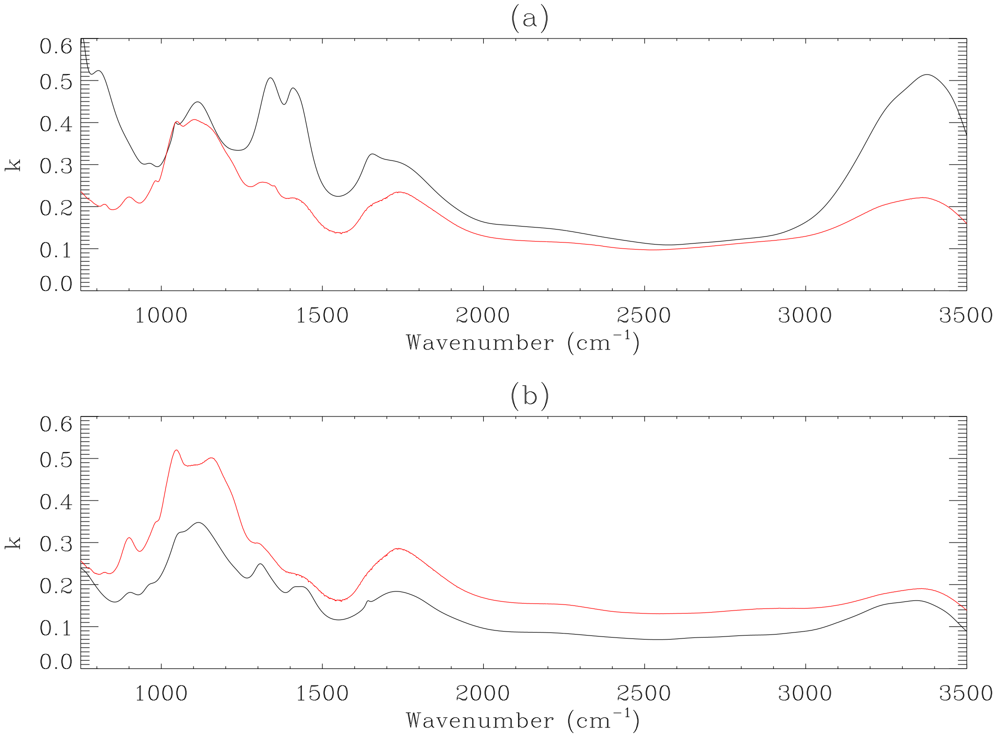

Figure 8.8: Panel (a) shows ternary (13.0 ±2.5 wt% HNO_3, 29.6 ±0.7 wt% H_2SO_4) solution imaginary refractive indices (black) from 29/11/00 (196.6 K), plotted against indices for the same composition/temperature from the Biermann *et al.* ternary calculator. Similarly, panel (b) shows ternary (12.2 ±2.1 wt% HNO_3, 41.7 ±0.6 wt% H_2SO_4) solution imaginary refractive indices (black) from 23/11/00 (197.8 K), plotted against indices for the same composition/temperature from the Biermann *et al.* ternary calculator.

temperatures in ternary solutions where the H_2SO_4 component is less than 30 wt%. The comparison in figure 8.8(a) is slightly below this threshold, while the comparison in figure 8.8(b), having a much higher H_2SO_4 concentration, is safely within the bounds of the Biermann *et al.* binary index data regime. Without further measurements at other points in the STS composition range, it is difficult to ascertain whether the data produced here is an improved representation of STS indices representative of colder conditions in the stratosphere (and thus of more interest in remote sensing). The differences in absorption features are pronounced at the low-H_2SO_4 range of ternary concentrations, indicating that the temperature limitations of the Biermann *et al.* data set (section 2.5.5) may compromise the application of these data to the ternary system in the stratosphere.

8.6.2 Comparison of binary sulphuric acid refractive indices

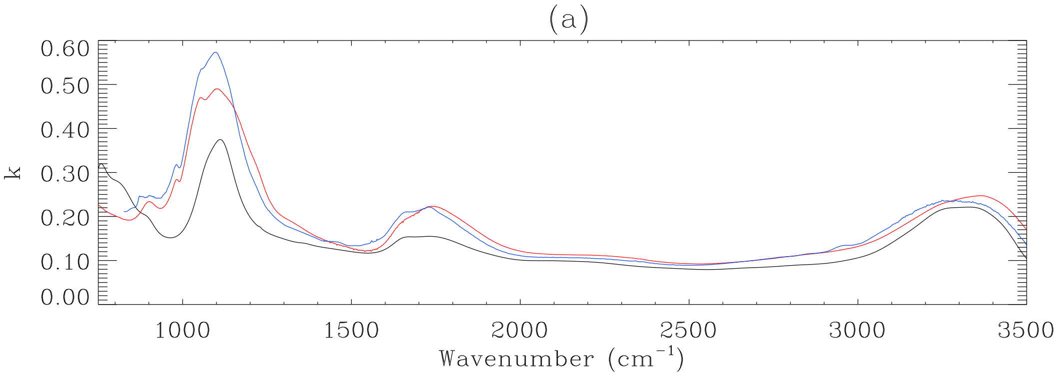

Figure 8.9: The retrieved 37.2 ±2.5 wt% (198.9 K) binary H_2SO_4 aerosol imaginary refractive index from 22/11/00 (black line), compared with a Biermann *et al.* (2000) thin-film calculation for the same temperature and composition (red) and aerosol refractive indices of binary H_2SO_4 (blue) of 39 wt% composition (220 K) from Niedziela *et al.* (1999).

Binary H_2SO_4 refractive index data published by Niedziela *et al.* (1999) and Biermann *et al.* (2000) are compared here with H_2SO_4 refractive indices measured in this work. As can be seen from these data in figure 8.9, there is a considerable scaling difference between all three sets of data. The disparities between absorption features of thin-film and aerosol H_2SO_4 indices, and the implications for calculation of composition from dielectric band areas, have already been noted in section 7.1.4. It is thus noteworthy that the spectral shape of the two aerosol indices, those measured for this work and the data of Niedziela *et al.* (1999), are qualitatively the most similar. This observation helps justify the use of the UNC indices for calibration of composition against band areas.

8.6.3 Conclusions

Further measurements of refractive indices of ternary solutions at the high-HNO_3 / low-H_2SO_4 (low temperature) range of the expected stratospheric composition regime for STS aerosol are necessary to provide data representative of STS aerosol most likely to be observed in the stratosphere. This work has characterised the mid-range of expected STS compositions, and has been found to qualitatively complement some of the Biermann *et al.*

(2000) ternary refractive index data. Towards the high-HNO_3 / low-H_2SO_4 composition range, the agreement between absorption features in the $k(\tilde{\nu})$ indices produced here and those of the Biermann *et al.* ternary calculator is poor. This may be a result of unknown uncertainties in this work, or possible problems with the Biermann *et al.* ternary calculator which may result from:

1. The limited temperature range of binary solution data used in the Biermann *et al.* ternary calculator which does not capture the identified temperature-dependence of H_2SO_4 in ternary solutions;
2. Universal differences in H_2SO_4 absorption features between thin-film and aerosol measurements (section 7.1.4);
3. Problems with the linear wt% based mixing rule used by Biermann *et al.* (2000) to calculate ternary indices from binary indices (more specific information is necessary to characterise this).

In order to ascertain relevance of both sets of data to the stratosphere further aerosol measurements of ternary solution refractive indices, at stratospheric temperatures, are required over a wider composition range.

With respect to the remote sensing of PSCs, the fingerprint of ternary solution absorption (represented by the $k(\tilde{\nu})$ index) has potential use in identification of PSC type and possibly liquid composition. However, this requires wide spectral coverage of PSCs in the infrared, something that has only recently become possible on a continuous basis from satellite observations with the launch of the ENVISAT platform (section 2.2.2). Previous attempts to identify PSC type/composition from remote sensed data have been problematical, due to the large number of unknowns in the analysis, although solid nitric acid trihydrate (NAT) has been tentatively identified in the Antarctic stratosphere from infrared absorption features (Remedios *et al.*, 2001). Thus, precise characterisation of the absorption features of STS, while of importance in the determination of PSC volume and surface area, is less important in terms of PSC identification at this stage.

The critical point that arises from the comparison of available data is the lack of general agreement of scaling of the $k(\tilde{\nu})$ index between data sets. This has been identified in this chapter from three-way comparison between data produced here and the binary H_2SO_4 indices of Biermann *et al.* (2000) and Niedziela *et al.* (1999), and separate comparisons with ternary calculator indices of Biermann *et al.* (2000). As established previously in section 2.2.3, the relationship between $k(\tilde{\nu})$ and PSC volume density, as calculated from remote measurement of PSC extinction, is approximately linear across most of the infrared (up to about 3500 cm^{-1}) for size distributions typical of STS in the stratosphere. The PSC surface area density is directly related to volume density as expressed by equation 2.6. Thus, as extreme examples where both scaling and absorption differences combine in phase to produce large differences, the maximum difference between binary H_2SO_4 $k(\tilde{\nu})$ indices produced here and by the UNC group (figure 8.9) of about 50 % at 1100 cm^{-1}, would lead to a 50 % difference in calculated PSC volume density, and similarly the observed difference of over 100 % against the calculated ternary indices of Biermann *et al.* in figure 8.8(a) would lead to a calculated PSC volume density difference of this magnitude.

Thus, it is clear that the scaling offsets of $k(\tilde{\nu})$ between different sets of data is the key issue limiting the applicability of laboratory data to the remote sensing of PSCs in the infrared. The calculation of refractive indices for this thesis incorporated all known experimental uncertainties to produce the final uncertainty in wavenumber-dependant refractive index. This uncertainty is significantly smaller than the differences between indices compared here, see the results presented previously in figure 8.4, indicating that refractive indices produced for this work most closely represent ternary solution aerosol at stratospheric temperatures. However, it is possible that some degree of dependence between retrieved size distribution and absorption band parameters may affect the scaling of the $k(\tilde{\nu})$ index determined in this work (section 6.1.4). It is suggested here that future experiments using this equipment at RAL should include a device for independent measurement of particle size distribution, such as a particle counter. This additional information would either validate or correct the current refractive index scaling.

Chapter 9

Summary and Future Work

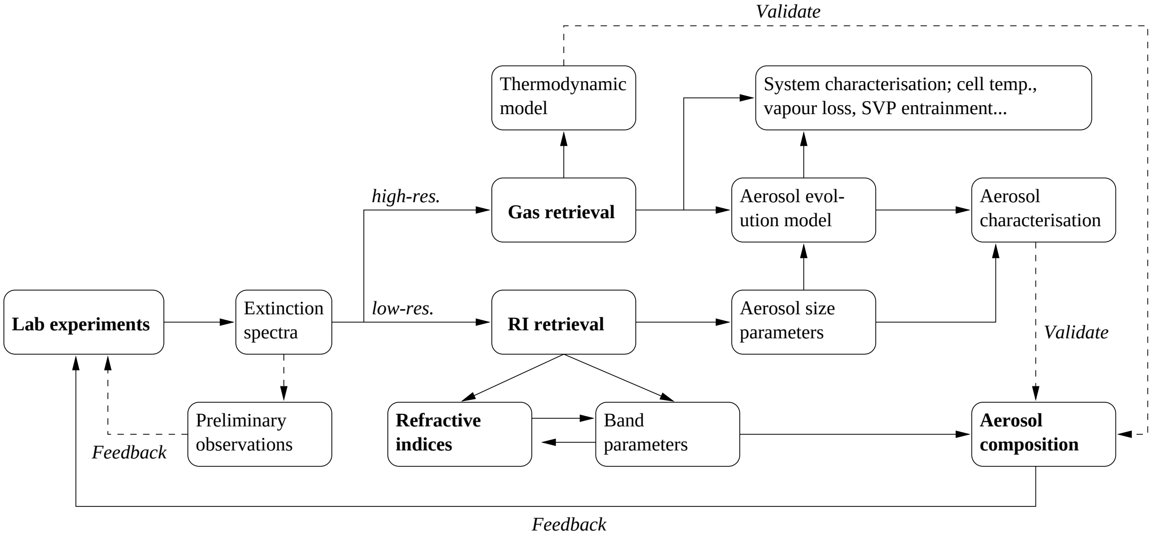

Figure 9.1: Schematic showing the main components of this work and the links between them; the dashed lines indicate secondary links. Major components are highlighted in bold text.

A schematic overview of work conducted for this thesis is presented in figure 9.1; the main numerical analysis was carried out in the gas and refractive index retrievals using the spectral measurements as indicated. The results were then employed for further analysis in order to characterise aerosol development and composition, and to characterise aspects of the system such as internal cell temperature, and aerosol residence time. Comprehensive validation was undertaken in the main areas of analysis, and where possible uncertainties were quantified in order to produce representative errors in key data produced for this thesis, principally wavenumber-dependant refractive indices.

9.1 Thesis Summary

This work yielded refractive indices of STS aerosol and mixed phase ternary solution and ice aerosol to a quantified high level of accuracy over the spectral range of 750 to 3500 cm^{-1}. The main achievements of this thesis include:

1. The first determination of refractive indices from direct-measurements of supercooled ternary solution aerosol;
2. The measurement of a range of ternary solution compositions of relevance to the stratospheric system;
3. Use of the novel classical damped harmonic oscillator (CDHO) band-absorption model which enabled derivation of refractive indices from a single extinction spectrum;
4. Band parameters produced through the CDHO model in determination of refractive indices were related to chemical speciation of binary and ternary solutions to evaluate aerosol composition (validated by gas-phase measurements);
5. Comprehensive determination of uncertainties in measurements and analytical techniques provided a rigorous assessment of wavenumber-dependant uncertainties in the derived refractive index data (a first for this work);
6. Experimental performance was comprehensively evaluated – including temperature measurement, wavenumber shift due to spectrometer beam divergence, and full characterisation of the temperature dependant transmission of the cell windows.

Comprehensive determination of wavenumber-dependant uncertainties in refractive indices potentially enables the accuracy of PSC surface area calculation from remote sensed data to be determined. Moreover, novel analytical techniques developed in this work, particularly use of the CDHO model to evaluate refractive indices from a single extinction spectrum, have considerable potential for analysis of other laboratory-generated data not only for study of PSCs but also for wider use.

Analysis of the results of this thesis and comparison against published work produced several key findings:

1. Refractive index data derived in this work and separate sets of published indices exhibit large variations in the scaling of the refractive index;
2. Comparison of ternary solution indices derived here and those produced by the Biermann *et al.* (2000) ternary calculator revealed considerable qualitative differences in infrared absorption features for low-temperature low-H_2SO_4 (sub 30 wt%) solutions;
3. Aqueous sulphuric acid exhibits significant temperature dependence in absorption around 1100 cm^{-1} in ternary and binary solutions;
4. Thin-film and aerosol solutions possess consistent and characteristic differences in sulphuric acid absorption features around 1200 cm^{-1} – otherwise, thin-film and aerosol absorption features are qualitatively in agreement for binary and ternary solutions;
5. Due to the high density of aerosol in the large cell, temperature-dependant variation of ternary solution aerosol composition occurs at a much higher temperature than in the stratosphere – thus, the ±2.13 K temperature uncertainty in the large cell is irrelevant for composition measurement at stratospheric temperatures.

The first point is of critical importance for calculation of PSC surface area from remote measurements of spectral extinction; evaluation of the rate of chemical activation in the stratosphere is constrained by poor knowledge of refractive index scaling. Additionally, the potential for PSC type/composition determination from broad spectral measurement of absorption features in the infrared is limited by poor knowledge of refractive indices of ternary solutions. For this work measurements were undertaken at temperatures much closer to those of the stratosphere than other major published data, and in aerosol form in order to ensure that the effects listed in points three and four do not compromise suitability of the measured refractive index data for the remote sensing of PSCs.

As detailed in the second point above, qualitative differences in infrared absorption features are observed for comparisons of some ternary solution refractive indices against those from the only other published source of ternary solution indices for compositions of relevance to the stratosphere, the ternary solution calculator of Biermann *et al.* (2000). Some of these differences can be attributed to temperature-dependent H_2SO_4 absorption around 1100 cm^{-1}. Agreement between these data is best for H_2SO_4 compositions above 30 wt%, where the data are broadly complementary (scaling differences neglected). Although it is possible that the Biermann *et al.* (2000) binary data (used for the ternary calculator) are temperature-limited to the regime above 30 wt% H_2SO_4, to test this hypothesis further comparisons over a greater range of ternary compositions than measured for this thesis are necessary.

The main limitations of this work include:

1. A relatively narrow range of ternary solution compositions were measured in the centre of the range expected for ternary solution aerosol in the stratosphere;
2. The high-HNO_3 / low-H_2SO_4 end of the expected stratospheric range of ternary solutions would have potentially been of more interest for the remote sensing of PSCs;
3. It is not possible to resolve the observed discrepancies in scaling of refractive index data between separate sets of published data and with this work due to possible systematic scaling errors introduced through k leakage in this work.

The main technique adopted for variation of ternary solution composition at low temperature in experiments for this work was variation of relative flow rates of gas streams into the large cell. Although this resulted in H_2SO_4 variation from approximately 25 to 65 wt%, and HNO_3 variation from approximately 5 to 20 wt% in ternary solutions, this range is only a sub-set of that expected through temperature-dependant variation of ternary solution PSCs in the stratosphere. Indeed, the largest ternary solution aerosol particles in the stratosphere would be expected near to the ice frost point following condensation of volatiles to form the high-HNO_3 / low-H_2SO_4 end of the stratospheric ternary composition range. Since these large particles would have the largest volume density, this is the most likely ternary composition to be observed from remote sensed data. It is desirable to measure such compositions since they have not been measured in this or other publications.

Finally, accurate determination of binary/ternary solution refractive indices has potential to resolve the uncertainties introduced by the observed scaling differences between published data. However, indices produced in this work again do not agree well with the scaling of any other major data set. Furthermore, the possibility of some dependence between size parameters and the imaginary refractive index (k leakage) introduces some uncertainty in scaling of refractive indices produced in this work. This is one of the critical outstanding issues that should be addressed by further development of experiments in future work.

9.2 Recommendations for future work

The principle recommendations for future work arising from this analysis comprise:

1. Independently measure aerosol size parameters to resolve the refractive index scaling issue;
2. Increase the range of ternary solution compositions measured – focus on the high-HNO_3 / low-H_2SO_4 end of the expected stratospheric composition range;
3. Measure a wide composition range of binary sulphate and nitrate solutions, while independently measuring composition, for purposes of developing a composition/band area calibration independent of limited published data.

In the same order, general experimental recommendations to achieve these suggestions involve:

1. Use a particle counter or measure spectra further into the visible spectral region;

2. Decrease the number density/composition of sulphuric acid aerosol, and increase the strength of the nitric acid bath;

3. Use a range of sulphuric acid bath compositions and develop new techniques to nucleate binary nitrate aerosol – determine composition from H_2O vapour pressure as derived from high resolution spectra.

The measures indicated in the first recommendation could prove difficult to implement due to potential problems including window contamination (due to higher HNO_3 gas pressures) and injection of enough HNO_3 gas to sufficiently modify the sulphate aerosol; new techniques may be necessary. A range of detailed recommendations have been compiled from this work, and are detailed for completeness below.

9.2.1 Detailed recommendations

Experiment direction:

- Further extensive temperature validation could reduce the uncertainty of the calibrated temperature. Gas *markers* such as CO_2 or CO, which have well known temperature dependencies, could be used for validation of PRTs during test experiments (or main experiments); CO is better for low temperatures due to its lower freezing point. However, temperature can only be assessed from high resolution measurements. Of greater value would be improved calibration of PRTs. Experiments to ensure that the gas temperature is indeed a good representation of the temperature along the cell path (by placing PRTs in the path), and to place PRTs permanently *inside* the cell during experiments using some kind of protective coating to prevent acid attacking the sensors, should reduce the temperature uncertainty;

- Due to the lack of crystalline solid PSC refractive index data derived from aerosol spectra, measurements of such particles would be of great use to the community – the differences between thin-film and aerosol crystalline spectra have been highlighted, justifying new aerosol measurements in the laboratory. Previous generation of solid particles could be used as a blueprint for future experiments;

- Measurements of depolarisation of a plane-polarised radiation source through the cell volume would indicate whether aerosol particles are non-spherical (i.e. solid) if pronounced depolarisation were observed;
- Vapour transfer tests could be conducted to confirm the independence of vapour pressure as measured along the beam path to the coldest points in the cell. The findings of Niedziela *et al.* (1999) are used here to support the assumption that buffer gas pressure is high enough to isolate colder areas of the cell, so that the retrieved vapour pressure is in equilibrium with particles observed along the beam path – this is supported by the composition validation using retrieved volatile partial pressures;
- Lower temperatures are required to simulate the coldest parts (such as in lee waves) of the polar stratosphere. A different coolant would be required to attain temperatures below 190 K since ethanol freezes in the current configuration below this cell temperature.

Further practical experiment suggestions:

- Cell window material could be changed (from BaF_2) to (a) reduce the cut-off wavenumber for greater spectral coverage into the infrared, and (b) use a material with transmission properties that are less sensitive to temperature variation. The same properties of high resistance to HNO_3 and H_2SO_4 would be required of any replacement material;
- Separate PRTs could be used to directly log the temperature of cell windows for improved modelling of any temperature-dependent transmission effects;
- A radioactive source could be introduced to seed for nucleation of ice particles in a cloud of supercooled liquid aerosol or a supersaturated gas medium (to simulate

nucleation by cosmic rays in the stratosphere);

- A protected fish-lens camera, inserted into the top of the cell (i.e. an adapted small web-cam) could be used to image the cell contents to address if/where any ice forms on the sides of the cell when solid aerosol is formed and to assess the mixing of aerosol. A coloured *tracer* could be fed into the gas/aerosol flow to help determine the degree of turbulent mixing, and confirm residence times;
- Shock cooling by a pre-cooler placed before the cell could be used to freeze liquid aerosol, or/and to homogeneously nucleate solid aerosol from supersaturated gas phase;
- Automatic recording of flow-rates, errors in co-added spectrometers scans, and scan duration would assist analysis and determination of spectral uncertainties.

Appendix: Retrieval Theory

Spectra of aerosol and gas lines are characterised by fundamental physical parameters. Determination of these quantities is an central part of the analysis presented in this thesis. These parameters cannot be determined directly from measured spectra, which are complex functions of the physical parameters (and hence indirect measurements). Such functions relate, for example, how the spectral signature of gas lines are modified by pressure, temperature and gas concentration. In deriving these data from spectra we are presented with an *inverse problem*. The retrieval programs used for this analysis extract information from spectra by comparing measurements with the output from a *forward model*, reducing the differences iteratively to a minimum by varying the physical parameters represented in the model. The forward model is essentially a representation of the complex functions that relate spectra to physical parameters. Formulation of a representative forward model requires a good understanding of the physics of the problem.

Structure of a generic retrieval program. The quantification of information and associated uncertainties contained within measured spectra can be undertaken by a retrieval program by varying a state vector, $\mathbf{x}$, of variables that are used to represent the measurements, $\mathbf{y}$, by means of the forward model, $\mathbf{F}(\mathbf{x})$. An optimal estimate of the values of the state vector is found; this processes is described in depth elsewhere (Rodgers, 2000). A flow-chart representation of a standardised retrieval program can be seen in figure 9.2.

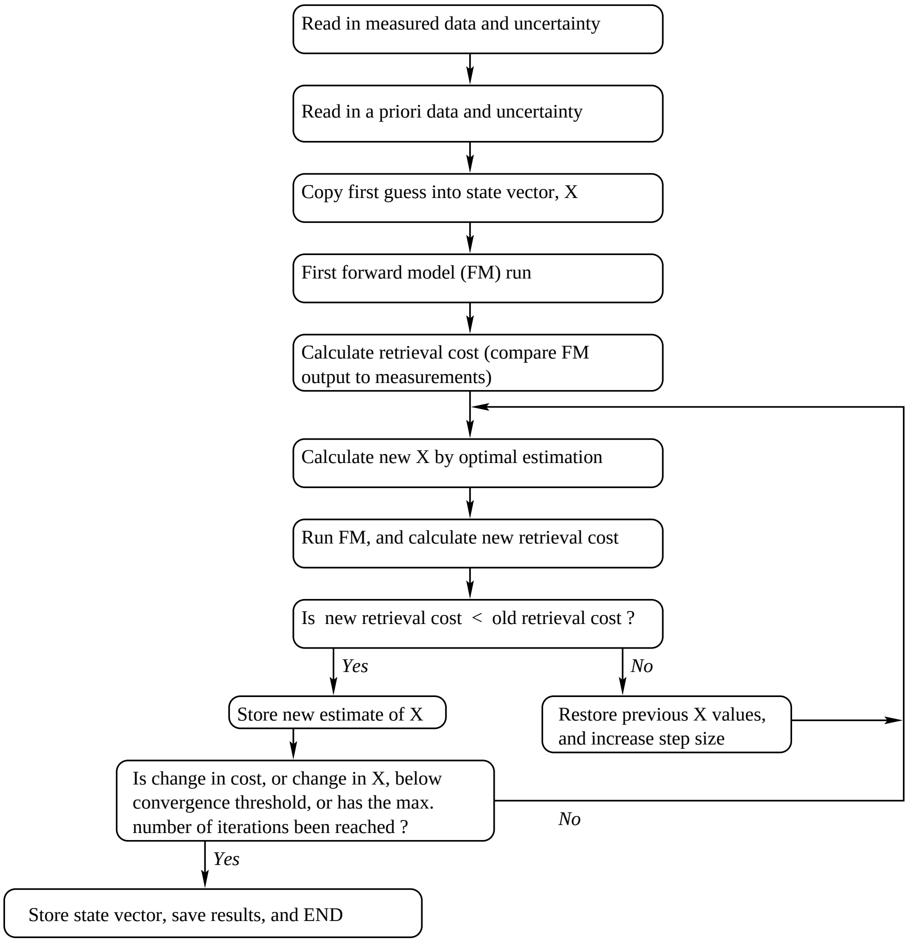

Figure 9.2: Representation of a generic retrieval program.

Effectively an $(i+1)^{\mathrm{th}}$ estimate of the solution is found from the i^{th} estimate using :

$$\begin{aligned} \mathrm{x}_{i+1} &= \mathrm{x}_i + \left[(1+\gamma)\mathbf{S}_a^{-1} + \mathbf{K}_i^T\mathbf{S}_\epsilon^{-1}\mathbf{K}_i\right]^{-1} \\ &\quad \left[\mathbf{K}_i^T\mathbf{S}_\epsilon^{-1}\left(\mathrm{y} - \mathbf{F}(\mathrm{x}_i)\right) - \mathbf{S}_a^{-1}\left(\mathrm{x}_i - \mathrm{x}_a\right)\right] \end{aligned}$$

In this equation $\mathbf{x}_a$ represents the *a priori* state vector, $\mathbf{S}_a$ the covariance matrix of the *a priori*, $\mathbf{S}_\epsilon$ the covariance matrix of the measurements, $\mathbf{K}$ the weighting matrix and γ the step size. Iteration using this equation simultaneously minimises the χ^2 difference between the state vector and its *a priori* and the χ^2 difference between the measurements and the output of the forward model, which are combined to produce the total *retrieval cost.* Iterations continue until a pre-set convergence criterion is met, usually when the retrieval cost has fallen below a threshold. Caution must be taken where a number of minima are present in the χ^2 surface. Figure 9.3 illustrates a simple one-dimensional case where it would be possible for the retrieval to *fall* into one of five local minima rather that the (lowest) global minimum should the first guess start far outside the slope of the global minimum in state space. It is clear that the real situation will be conceptually more intricate since the χ^2 surface will have multiple dimensions corresponding to the number of elements in the state vector, but the principles are the same. Hence, it may become necessary for the user to *prompt* the retrieval by adjusting the first guess to avoid an incorrect solution. The real solution in cases where the forward model poorly represents true physics may not be the lowest point in the χ^2 surface.

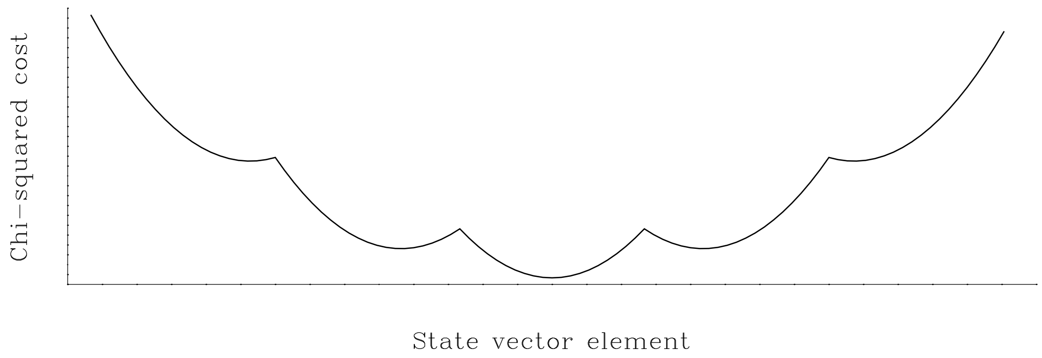

Figure 9.3: Illustration of local and global χ^2 minima in a 1-D state vector case.

The forward model represents the physics of the problem by describing the transformation between the state vector and the measurements. Since all models are a simplification of reality, forward model errors propagate from: (1) numerical approximation (e.g. poor gas characterisation, and use of finite-resolution reference forward model (RFM) spectra in gas model), (2) parametric errors (e.g. incorrect temperature), and (3) physical errors (over-simplified modelling of processes).

Bibliography

Adams, R. W. and Downing, H. D. (1986). Infrared optical constants of a ternary system of 75% H_2SO_4, 10% HNO_3, and 15% H_2O. *J. Opt. Soc. Am. A*, **3**(1), 22–28.

Anderson, J. G., Toohey, D. W., and Brune, W. H. (1991). Free-radicals within the Antarctic vortex – The role of CFCs in Antarctic ozone loss. *Science*, **251**(4989), 39–46.

Anthony, S. E., Tisdale, R. T., Disselkamp, R. S., Tolbert, M. A., and Wilson, J. C. (1995). FTIR studies of low temperature sulfuric acid aerosols. *Geophysical Research Letters*, **22**(9), 1105–1108.

Anthony, S. E., Onasch, T. B., Tisdale, R. T., Disselkamp, R. S., and Tolbert, M. A. (1997). Laboratory studies of ternary $H_2SO_4/HNO_3/H_2O$ particles: Implications for polar stratospheric cloud formation. *Journal of Geophysical Research*, **102**(D9), 10777–10784.

Banwell, C. N. (1983). *Fundamentals of molecular spectroscopy*. McGraw-Hill, third edition.

Baran, A. J., Francis, P. N., *et al.* (2001). A study of the absorption and extinction properties of hexagonal ice columns and plates in random and preferred orientation, using exact T-matrix theory and aircraft observations of cirrus. *Journal of Quantitative Spectroscopy & Radiative Transfer*, **70**, 505–518.

BAS Ozone Bulletin July 98 (1999). BAS ozone bulletin 07/98 issued 1999 May 13.
Web site: http://www.nerc-bas.ac.uk/public/icd/jds/ozone/bulletins/bas0798.html.

Bass, S. F., Grainger, R. G., McPheat, R. A., Newnham, D. A., and Remedios, J. J. (2000a). Spectroscopic studies of laboratory generated polar stratospheric cloud particles. In *Proceedings of the International Radiation Symposium, Russia 2000*. International Radiation Commission.

Bass, S. F., Grainger, R. G., Remedios, J. J., Newnham, D. A., and Lambert, A. (2000b). Refractive indices of laboratory generated supercooled ternary solution aerosol. In *Proceedings of the Quadrennial Ozone Symposium, Sapporo 2000*, pages 303–304. International Ozone Commission, International Association for Meteorology and Atmospheric Sciences.

Bates, D. R. and Nicolet, M. (1950). The photochemistry of atmospheric water vapor. *Journal of Geophysical Research*, **55**(301).

Baumgardner, D., Dye, J. E., Gandrud, B., Rogers, D., Weaver, K., *et al.* (1994). The multiangle aerosol spectrometer probe: A new instrument for airborne particle research. In *Proceedings of the Ninth Symposium on Meteorological Observations and Instrumentation*, pages 434–439, Charlotte, N.C. American Meteorological Society.

Baumgardner, D., Dye, J. E., Gandrud, B., Barr, K., Kelly, K., and Chan, K. R. (1996). Refractive indices of aerosols in the upper troposphere and lower stratosphere. *Geophysical Research Letters*, **23**(7), 749–752.

Beer, R. (1992). *Remote Sensing by Fourier Transform Spectrometry*. Number 120 in Chemical Analysis. John Wiley & Sons.

Bertie, J. E. and Zhang, S. L. (1992). Infrared intensities of liquids. IX. the Kramers-Kronig transform, and its approximation by the finite Hilbert transform via fast Fourier transforms. *Can. J. Chem.*, **70**, 520–531.

Bertie, J. E., Zhang, S. L., and Keefe, C. D. (1994). Infrared intensities of liquids XVI. Accurate determination of molecular band intensities from infrared refractive index data and dielectric constant spectra. *Journal of Molecular Science*, **324**, 157–176.

Bertram, A. K. and Sloan, J. J. (1998a). Temperature-dependant nucleation rate constants and freezing behavior of submicron nitric acid dihydrate aerosol particles under stratospheric conditions. *Journal of Geophysical Research*, **103**(D3), 3553–3561.

Bertram, A. K. and Sloan, J. J. (1998b). The nucleation rate constants and freezing mechanism of nitric acid trihydrate aerosol under stroatospheric conditions. *Journal of Geophysical Research*, **103**(D11), 13261–13265.

Bertram, A. K., Patterson, D. D., and Sloan, J. J. (1996). Mechanisms and temperatures for the freezing of sulfuric acid aerosols measured by FTIR extinction spectroscopy. *Journal of Physical Chemistry*, **100**, 2376–2383.

Bertram, A. K., Dickens, D. B., and Sloan, J. J. (2000). Supercooling of type 1 polar stratospheric clouds: The freezing of submicron nitric acid aerosols having HNO_3 mol fractions less than 0.5. *Journal of Geophysical Research*, **105**(D7), 9283–9290.

Beyerle, G. and Neuber, R. (1994). The stratospheric aerosol content above Spitsbergen during winter 1991/92. *Geophysical Research Letters*, **21**(13), 1291–1294.

Beyerle, G., Neuber, R., Schrems, O., Wittrock, F., and Knudsen, B. (1994). Multiwavelength lidar measurements of stratospheric aerosols above Spitsbergen during winter 1992/93. *Geophysical Research Letters*, **21**(1), 57–60.

Beyerle, G., Luo, B., Neuber, R., Peter, T., and McDermid, I. S. (1997). Temperature dependence of ternary solution particle volumes as observed by lidar in the Arctic stratosphere during winter 1992/1993. *Journal of Geophysical Research*, **102**(D3), 3603–3609.

Beyerle, G., Deckelmann, H., Neuber, R., Rosen, J. M., Reimer, E., and Schoeberl, M. R. (2001). Occurrence of solid particles in the winter polar stratosphere above the nitric acid trihydrate coexistence temperature inferred from ground-based polarization lidar observations at ny-alesund, spitsbergen. *Journal of Geophysical Research*, **106**(D3), 2979–2992.

Biermann, U. M., Crowley, J. N., Huthwelker, T., Moortgat, G. K., Crutzen, P. J., and Peter, T. (1998). FTIR studies on lifetime prolongation of stratospheric ice particles due to NAT coating. *Geophysical Research Letters*, **25**(21), 3939–3942.

Biermann, U. M., Luo, B. P., and Peter, T. (2000). Absorption spectra and optical constants of binary and ternary solutions of H_2SO_4, HNO_3, and H_2O in the mid latitude at atmospheric temperatures. *Journal of Physical Chemistry A*, **104**(4), 783–793.

Boone, T. L., Fuller, K. A., and Downing, H. D. (1980). Infrared optical constants of aqueous solutions of nitric acid. *Journal of Physical Chemistry*, **84**, 2666–2667.

Born, M. and Wolf, E. (1970). *Principles of Optics*. Pergamon, Oxford.

Brigham, E. O. (1974). *The Fast Fourier Transform*. Prentice Hall, New York.

Browell, E. V., Butler, C. F., Ismail, S., Robinette, P. A., Carter, A. F., *et al.* (1990). Airborne lidar observations in the wintertime Arctic stratosphere: Polar stratospheric clouds. *Geophysical Research Letters*, **17**(4), 385–388.

Brune, W. H., Anderson, J. G., and Chan, K. R. (1989a). Insitu observations of ClO in the Antarctic – ER-2 aircraft results from 54-degrees-S to 72-degrees-S latitude. *Journal of Geophysical Research - Atmospheres*, **94**(D14), 16649–16663.

Brune, W. H., Anderson, J. G., and Chan, K. R. (1989b). Insitu observations of BrO over Antarctica – ER-2 aircraft results from 54-degrees-S to 72-degrees-S latitude. *Journal of Geophysical Research - Atmospheres*, **94**(D14), 16639–16647.

Carslaw, K. S. (1998). Increased stratospheric ozone depletion due to mountain-induced atmospheric waves. *Nature*, **391**, 675–678.

Carslaw, K. S., Luo, B. P., Clegg, S. L., Peter, T., Brimblecombe, P., and Crutzen, P. J. (1994). Stratospheric aerosol growth and HNO_3 gas phase depletion from coupled HNO_3 and water uptake by liquid particles. *Geophysical Research Letters*, **21**(23), 2479–2482.

Carslaw, K. S., Clegg, S. L., and Brimblecombe, P. (1995). A thermodynamic model of the system HCl-HNO_3-H_2SO_4-H_2O, including solubilities of HBr, from < 200 to 328 K. *Journal of Physical Chemistry*, **99**(29), 11557–11574.

Carslaw, K. S., Peter, T., and Clegg, S. L. (1997). Modeling the composition of liquid stratospheric aerosols. *Reviews of Geophysics*, **35**(2), 125–154.

Carslaw, K. S., Wirth, M., Tsias, A., Luo, B. P., Dornbrack, A., *et al.* (1998). Particle microphysics and chemistry in remotely observed mountain polar stratospheric clouds. *Journal of Geophysical Research*, **103**(D5), 5785–5796.

Carslaw, K. S., Peter, T., Bacmeister, J. T., and Eckermann, S. D. (1999a). Widespread solid particle formation by mountain waves in the Arctic stratosphere. *Journal of Geophysical Research*, **104**(D1), 1827–1836.

Carslaw, K. S., Volkert, H., Haynes, P., Harris, N. R. P., Larsen, N., Amanatidis, G., and Peter, T. (1999b). The European workshop on mesoscale processes in the stratosphere – Overview and outcomes, air pollution research report 69. *European Commission*, pages 1–6.

Chamberlain, J. (1979). *The Principles of Interferometric Spectroscopy*. John Wiley & Sons.

Chipperfield, M. P. and Jones, R. L. (1999). Relative influences of atmospheric chemistry and transport on Arctic ozone trends. *Nature*, **400**, 551–554.

Clapp, M. L., Worsnop, D. R., and Miller, R. E. (1995). Frequency dependant optical constants of water ice obtained directly from aerosol extinction spectra. *Journal of Physical Chemistry*, **99**, 6317–6326.

Crutzen, P. J. (1970). The influence of nitrogen oxides on the atmospheric ozone content. *Quarterly Journal of the Royal Meteorological Society*, **96**, 320–325.

Crutzen, P. J. and Arnold, F. (1986). Nitric-acid cloud formation in the cold Antarctic stratosphere – A major cause for the springtime ozone hole. *Nature*, **324**(6098), 651–655.

Deshler, T., Peter, T., Müller, R., and Crutzen, P. (1994). The lifetime of leewave-induced ice particles in the Arctic stratosphere: 1. Balloonborne observations. *Geophysical Research Letters*, **21**, 1327–1330.

Dhaniyala, S., McKinney, K. A., and Wennberg, P. O. (2002). Lee-wave clouds and denitrification of the polar stratosphere. *Geophysical Research Letters*, **29**(9).

Dignam, M. J. (1988). Fourier-transform polarization spectroscopy. *Applied Spectroscopy Reviews*, **24**(1–2), 99–135.

Disselkamp, R. S., Anthony, S. E., Prenni, A. J., Onasch, T. B., and Tolbert, M. A. (1996). Crystallization kinetics of nitric acid dihydrate aerosols. *Journal of Physical Chemistry*, **100**(21), 9127–9137.

Doicu, A. and Wriedt, T. (2001). t-matrix method for electromagnetic scattering from scatterers with complex structure. *Journal of Quantitative Spectroscopy & Radiative Transfer*, **70**, 663–673.

Downing, H. D. and Williams, D. (1975). Optical constants of water in the infrared. *Journal of Geophysical Research*, **80**, 1656–1661.

Dudhia, A. (1997). Reference Forward Model - Software User's Manual. Technical Report PO-MA-OXF-GS-0003, AOPP, Oxford University. Also to be found at: http://www.atm.ox.ac.uk/RFM/.

Dye, J. E., Gandrud, B. W., Baumgardner, D., Chan, K. R., Ferry, G. V., *et al.* (1990). Observed particle evolution in the polar stratospheric cloud of January 24, 1989. *Geophysical Research Letters*, **17**(4), 413–416.

Dye, J. E., Baumgardner, D., Gandrud, B. W., Kawa, S. R., Kelly, K. K., *et al.* (1992). Particle size distributions in Arctic polar stratospheric clouds, growth and freezing of sulfuric acid droplets and implications for cloud formation. *Journal of Geophysical Research*, **97**, 8015–8034.

Dye, J. E., Baumgardner, D., Gandrud, B. W., Drdla, K., Barr, K., *et al.* (1996). In-situ observations of an Antarctic polar stratospheric cloud: Similarities with Arctic observations. *Geophysical Research Letters*, **23**(15), 1913–1916.

Fahey, D. W., Murphy, D. M., Kelly, K. K., Ko, M. K. W., Proffitt, M. H., *et al.* (1989). Measurements of nitric-oxide and total reactive nitrogen in the Antarctic stratosphere – observations and chemical implications. *Journal of Geophysical Research*, **94**(D14), 16665–16681.

Fahey, D. W., Gao, R. S., Carslaw, K. S., Kettleborough, J., Popp, P. J., *et al.* (2001). The detection of large HNO_3-containing particles in the winter Arctic stratosphere. *Science*, **291**, 1026–1031.

Farman, J. C., Gardiner, B., and Shanklin, J. D. (1985). Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction. *Nature*, **315**, 207–210.

Fels, S. B., Mahlman, J. D., Schwarzkopf, M. D., and Sinclair, R. W. (1980). Stratospheric sensitivity to perturbations in ozone and carbon dioxide: radiative and dynamical response. *J. Atmos. Sci.*, **37**, 2265–2297.

Forster, P. M. D. and Shine, K. P. (1999). Stratospheric water vapour changes as a possible contributor to observed stratospheric cooling. *Geophysical Research Letters*, **26**(21), 3309–3312.

Gary, B. (1989). Observational results using the microwave temperature profiler during the airborne antarctic ozone experiment. *Journal of Geophysical Research*, **94**, 11223–11231.

Gobbi, G. P. and Adriani, A. (1993). Mechanisms of formation of stratospheric clouds observed during the Antarctic late winter of 1992. *Geophysical Research Letters*, **20**(14), 1427–1430.

Graf, H. F., Kirchner, I., and Perlwitz, J. (1998). Changing lower stratospheric circulation: The role of ozone and greenhouse gases. *Journal of Geophysical Research*, **103**(D10), 11251–11261.

Grainger, R. G., Lambert, A., Rodgers, C. D., Taylor, F. W., and Deshler, T. (1995). Stratospheric aerosol effective radius, surface area and volume estimated from infrared measurements. *Journal of Geophysical Research*, **100**(D8), 16507–16518.

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

Gronholz, J. and Herres, W. (1985). Understanding FT-IR data processing. Combined reprint of Dr. Alfred Huethig Publishers (obtainable from Bruker, Karlsruhe), Instruments and Computers, Vols. 1 (1984) and 3 (1985).

Hamill, P. and Toon, O. B. (1991). Polar stratospheric clouds and the ozone hole. *Physics Today*, pages 34–42.

Hanson, D. and Mauersberger, K. (1988). Laboratory studies of the nitric-acid trihydrate – Implications for the south polar stratosphere. *Geophysical Research Letters*, **15**(8), 855–858.

Heathfield, A. E., Newnham, D. A., Ballard, J., Grainger, R. G., and Lambert, A. (1999). Infrared and visible Fourier-transform spectra of sulphuric-acid-water aerosols at 230 and 294 K. *Applied Optics*, **38**(30), 6408–6420.

Hervig, M. (1999). Stratospheric clouds over England. *Geophysical Research Letters*, **26**(8), 1137–1140.

Hofmann, D. J. and Deshler, T. (1989). Comparison of stratospheric clouds in the Antarctic and the Arctic. *Geophysical Research Letters*, **16**(12), 1429–1432.

Hofmann, D. J. and Deshler, T. (1991). Stratospheric cloud observations during formation of the Antarctic ozone hole in 1989. *Journal of Geophysical Research*, **96**(D2), 2897–2912.

Hofmann, D. J., Rosen, J. M., Pepin, T. J., and Pinnick, R. G. (1975). Stratospheric aerosol measurements I: Time variations at northern midlatitudes. *J. Atmos. Sci.*, **32**.

Höpfner, M., Oelhaf, H., Wetzel, G., Friedl-Vallon, F., Kleinert, A., *et al.* (2002). Evidence of scattering of tropospheric radiation by PSCs in mid-IR limb emission spectra: MIPAS-B observations and KOPRA simulations. *Geophysical Research Letters*. 10.1029/2001GL014443.

Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., der Linden, P. J. V., *et al.*, editors (2001). *IPCC 2001: Climate change 2001: The scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Iraci, L. T., Middlebrook, A. M., and Tolbert, M. A. (1995). Laboratory studies of the formation of polar stratospheric clouds: Nitric acid condensation on thin sulfuric acid films. *Journal of Geophysical Research*, **100**(D10), 20969–20977.

Iraci, L. T., Fortin, T. J., and Tolbert, M. A. (1998). Dissolution of sulfuric acid tetrahydrate at low temperatures and subsequent growth of nitric acid trihydrate. *Journal of Geophysical Research*, **103**(D7), 8491–8498.

James, J. F. (1995). *A student's guide to Fourier transforms: with applications in physics and engineering*. Cambridge University Press.

Kaye, G. W. C. and Laby, T. H. (1973). *Tables of physical and chemical constants, 14th edition*. Longman.

Kerker, M. (1969). *The Scattering of Light and other Electromagnetic Radiation*. Academic, New York.

Koehler, B. G., Middlebrook, A. M., and Tolbert, M. A. (1992). Characterization of model polar stratospheric cloud films using Fourier Transform Infrared Spectroscopy and temperature programmed desorption. *Journal of Geophysical Research*, **97**(D8), 8065–8074.

Koop, T., Biermann, U. M., Raber, W., Luo, B. P., Crutzen, P. J., and Peter, T. (1995). Do stratospheric aerosol droplets freeze above the ice frost point? *Geophysical Research Letters*, **22**(8), 917–920.

Koop, T., Carslaw, K. S., and Peter, T. (1997a). Thermodynamic stability and phase transitions of PSC particles. *Geophysical Research Letters*, **24**(17), 2199–2202.

Koop, T., Luo, B. P., Biermann, U. M., Crutzeen, P. J., and Peter, T. (1997b). Freezing of $HNO_3/H_2SO_4/H_2O$ solutions at stratospheric temperatures: Nucleation statistics and experiments. *Journal of Physical Chemistry*, **101**(6), 1117–1133.

Koop, T., Ng, H. P., Molina, L. T., and Molina, M. J. (1998). A new optical technique to study aerosol phase transitions: The nucleation of ice from H_2SO_4 aerosols. *Journal of Physical Chemistry A*, **102**, 8924–8931.

Kou, L., Labrie, D., and Chylek, P. (1993). Refractive indices of water and ice in the 0.65 to 2.5 micron range. *Applied Optics*, **32**, 3531–3540.

Krieger, U. K., Colberg, C. A., Weers, U., Koop, T., and Peter, T. (2000a). Supercooling of single H_2SO_4/H_2O aerosols to 158 K: No evidence for the occurrence of the octrahydrate. *Geophysical Research Letters*.

Krieger, U. K., Mössinger, J. C., Luo, B. P., Weers, U., and Peter, T. (2000b). Measurement of the refractive indices of H_2SO_4-HNO_3-H_2O solutions to stratospheric temperatures. *Applied Optics*, **39**(21), 3691–3703.

Larsen, N., Rosen, J. M., Kjome, N. T., and Knudsen, B. (1995). Deliquescence and freezing of stratospheric aerosol observed by balloonborne backscattersondes. *Geophysical Research Letters*, **22**(10), 1233–1236.

Larsen, N., Mikkelsen, S., Knudsen, B. M., Schreiner, J., Voigt, C., *et al.* (2000). Comparison of chemical and optical in situ measurements of polar stratospheric cloud particles. *Journal of Geophysical Research*, **105**(D1), 1491–1502.

Libnau, F. O., Kvalheim, O. M., Christy, A. A., and Toft, J. (1994). Spectra of water in the near- and mid-infrared region. *Vibrational Spectroscopy*, **7**, 243–254.

Lin, J. and Tabazadeh, A. (2001). A parameterization of an aerosol physical chemistry model for the $NH_3/H_2SO_4/HNO_3/H_2O$ system at cold temperatures. *Journal of Geophysical Research*, **106**(D5), 4815–4829.

Lovejoy, E. R. and Hanson, D. R. (1995). Measurement of the kinetics of reactive uptake by submicron sulfuric acid particles. *Journal of Physical Chemistry*, **99**, 2080–2087.

Luo, B. P., Krieger, U. K., and Peter, T. (1996). Densities and refractive indices of $H_2SO_4/HNO_3/H_2O$ solutions to stratospheric temperatures. *Geophysical Research Letters*, **23**(25), 3707–3710.

Mahfouf, J. F., Cariolle, D., Royer, J. F., Geleyn, J. F., and Timbal, B. (1994). Response of the Mto-France climate change model to changes in CO_2 and sea surface temperature. *Clim. Dyn.*, **9**, 345–362.

Marti, J. and Mauersberger, K. (1993). A survey and new measurements of ice vapor-pressure at temperatures between 170 and 250 k. *Geophysical Research Letters*, **20**(5), 363–366.

Martin, S. T. (2000). Phase transitions of aqueous atmospheric particles. *Chem. Rev.*, **100**, 3403–3453.

Massie, S. T., Baumgardner, D., and Dye, J. E. (1998). Estimation of polar stratospheric cloud volume and area densities from UARS, stratospheric aerosol measurement II, and polar ozone and aerosol measurement II extinction data. *Journal of Geophysical Research*, **103**(D5), 5773–5783.

McPheat, R. A., Newnham, D. A., Williams, R. G., and Ballard, J. (2001). A large-volume, coolable spectroscopic cell for aerosol studies. *Applied Optics*, **40**(36).

McPheat, R. A., Bass, S. F., Newnham, D. A., Ballard, J., and Remedios, J. J. (2002). Comparison of aerosol and thin film spectra of supercooled ternary solution aerosol. *Journal of Geophysical Research*, **107**(D19). art. no. 4371.

Middlebrook, A. M., Iraci, L. T., McNeill, L. S., Koehler, B. G., Wilson, M. A., *et al.* (1993). Fourier transform-infrared studies of thin H_2SO_4/H_2O films: formation, water-uptake, and solid-liquid phase-changes. *Journal of Geophysical Research*, **98**(D11), 20473–20481.

Middlebrook, A. M., Berland, B. S., George, S. M., Tolbert, M. A., and Toon, O. B. (1994). Real refractive indices of infrared-chearaterized nitric-acid/ice films: Implications for optical measurements of polar stratospheric clouds. *Journal of Geophysical Research*, **99**(D12), 25655–25666.

Mie, G. (1908). Beitrage zur optik truber medien speziell kolloidaler metallosungen. *Ann. Phys.*, **25**, 377.

Molina, M. J. and Rowland, F. S. (1974). Stratospheric sink for chlorofluoromethanes: chlorine atom catalyzed destruction of ozone. *Nature*, **249**, 810–812.

Molina, M. J., Tso, T. L., Molina, L. T., and Wang, F. C. Y. (1987). Antarctic stratospheric chemistry of chlorine nitrate, hydrogen-chloride and ice-release of active chlorine. *Science*, **238**(4831), 1253–1257.

Müller, R., Crutzen, P., Grooß, J., Brühl, C., Russell, J. M., *et al.* (1997). Severe chemical ozone loss in the Arctic during the winter of 1995-96. *Nature*, **389**, 709–712.

Nebeker, B. M., J. L. de la Peña, and Hirleman, E. D. (2001). Comparisons of the discrete-dipole approximation and modified double interaction model methods to predict light scattering from small features on surfactes. *Journal of Quantitative Spectroscopy & Radiative Transfer*, **70**, 741–759.

Nedoluha, G. E., Bevilacqua, R. M., Gomez, R. M., Siskind, D. E., Hicks, B. C., *et al.* (1998). Increases in middle atmospheric water vapor as observed by the Halogen Occultation Experiment and the ground-based Water Vapor Millimeter-wave Spectrometer from 1991 to 1997. *Journal of Geophysical Research*, **103**(D3), 3531–3543.

Niedziela, R. F., Miller, R. E., and Worsnop, D. R. (1998a). Temperature and frequency-dependant optical constants for Nitric Acid Dihydrate from aerosol spectroscopy. *Journal of Physical Chemistry A*, **102**(32), 6477–6484.

Niedziela, R. F., Norman, M. L., Miller, R. E., and Worsnop, D. R. (1998b). Temperature- and composition-dependant infrared optical constants for sulfuric acid. *Geophysical Research Letters*, **25**(24), 4477–4480.

Niedziela, R. F., Norman, M. L., DeForest, C. L., Miller, R. E., and Worsnop, D. R. (1999). A temperature- and composition- dependent study of H_2SO_4 aerosol optical constants using fourier transform and tunable diode laser infrared spectroscopy. *Journal of Physical Chemistry A*, **103**(40), 8030–8040.

Norman, M. L., Qian, J., Miller, R. E., and Worsnop, D. R. (1999). Infrared complex refractive indices of supercooled liquid HNO_3/H_2O aerosols. *Journal of Geophysical Research*, **104**(D23), 30571–30584.

Norton, R. H. and Beer, R. (1976). New apodizing functions for Fourier spectrometry. *J. Opt. Soc. Am.*, **66**(3), 259–264.

Palamarev, H. and Georgiev, G. (1994). Statistical distribution of hydrogen-bonded OH oscillators in HOD on the basis of infrared spectra. *Vibrational Spectroscopy*, **7**, 255–264.

Palmer, K. F. and Williams, D. (1975). Optical constants of sulfuric acid: application to the clouds of Venus? *Appl. Opt.*, **14**, 208–219.

Pawson, S. and Naujokat, B. (1997). Trends in daily wintertime temperatures in the northern stratosphere. *Geophysical Research Letters*, **24**, 575–578.

Peter, T., Müller, M., Drdla, K., Petzoldt, K., and Reimer, E. (1992). A microphysical box model for EASOE: Preliminary results for the January/February 1990 PSC event over Kiruna. *Ber. Bunsenges. Phys. Chem.*, **96**, 362–367.

Pinkley, L. W. and Williams, D. (1976). The infrared optical constants of sulfuric acid at 250 K. *J. Opt. Soc. Am.*, **66**, 122–124.

Portmann, R. W., Solomon, S., Garcia, R. R., Thomason, L. W., Poole, L. R., and McCormick, M. P. (1996). Role of aerosol variations in anthropogenic ozone depletion in the polar regions. *Journal of Geophysical Research*, **101**(D17), 22991–23006.

Prenni, A. J., Onasch, T. B., Tisdale, R. T., Siefert, R. L., and Tolbert, M. A. (1998). Composition-dependent freezing nucleation rates for HNO_3/H_2O aerosols resembling gravity-wave-perturbed stratospheric particles. *Journal of Geophysical Research*, **103**(D21), 28439–28450.

Pueschel, R. F., Snetsinger, K. G., Goodman, J. K., Toon, O. B., Ferry, G. V., *et al.* (1989). Condensed nitrate, sulfate, and chlorine in Antarctic stratospheric aerosols. *Journal of Geophysical Research*, **94**(D9), 11271–11284.

Querry, M. R. and Tyler, I. L. (1980). Reflectance and complex refractive indices in the infrared for aqueous solutions of nitric acid. *J. Chem. Phys.*, **72**(4), 2495–2499.

Ravishankara, A. R. and Hanson, D. R. (1996). Differences in the reactivity of Type I polar stratospheric clouds depending on their phase. *Journal of Geophysical Research*, **101**(D2), 3885–3890.

Remedios, J. J. *et al.* (2000). Final report of the postcode project. *European Environment and Sustainable Development Programme*.

Remedios, J. J. *et al.* (2001). Mapping of polar stratospheric clouds and ozone levels relevant to the region of europe (mapscore): Annual report 2001. Technical report, European Union: EVK2-CT-2000-00077.

Richwine, L. J., Clapp, M. L., Miller, R. E., and Worsnop, D. R. (1995). Complex refractive indices in the infrared of nitric acid trihydrate aerosols. *Geophysical Research Letters*, **22**(19), 2625–2628.

Rodgers, C. D. (2000). *Inverse Methods for Atmospheric Sounding: Theory and Practice*. Number 2 in Atmospheric, Oceanic and Planetary Physics. World Scientific Publishing Company.

Rosen, J. M. and Kjome, N. T. (1991). Backscattersonde: A new instrument for atmospheric aerosol research. *Applied Optics*, **30**, 1552–1561.

Rothman, L. S., Rinsland, C. P., Goldman, A., Massie, S. T., Edwards, D. P., *et al.* (1998). The HITRAN molecular spectroscopic database and HAWKS (HITRAN atmospheric workstation): 1996 edition. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **60**, 665–710.

Salawitch, R. J., Gobbi, G. P., Wofsy, S. C., and McElroy, M. B. (1989). Denitrification in the Antarctic stratosphere. *Nature*, **339**(6225), 525–527.

Salcedo, D., Molina, L. T., and Molina, M. J. (2001). Homogeneous freezing of concentrated aqueous nitric acid solutions at polar stratospheric temperatures. *Journal of Physical Chemistry A*, **105**(9), 1433–1439.

Santee, M. L., Manney, G. L., Livesey, N. J., and Waters, J. W. (2000). UARS Microwave Limb Sounder observations of denitrification and ozone loss in the 2000 Arctic late winter. *Geophysical Research Letters*, **27**(19), 3213–3216.

Schreiner, J., Voigt, C., Kohlmann, A., Arnold, F., Mauersberger, K., and Larsen, N. (1999). Chemical analysis of polar stratospheric cloud particles. *Science*, **283**, 968–970.

Shindell, D. T., Rind, D., and Lonergan, P. (1998). Increased polar stratospheric ozone losses and delayed eventual recovery owing to increasing greenhouse-gas concentrations. *Nature*, **392**, 589–592.

Sinnhuber, B. M., Chipperfield, M. P., Davies, S., Burrows, J. P., Eichmann, K. U., *et al.* (2000). Large loss of total ozone during the Arctic winter of 1999/2000. *Geophysical Research Letters*, **27**(21), 3473–3476.

Solomon, S., Garcia, R. R., Rowland, F. S., and Wuebbles, D. J. (1986). On the depletion of Antarctic ozone. *Nature*, **321**(6072), 755–758.

Spang, R., Riese, M., and Offermann, D. (2001). CRISTA-2 observations of the south polar vortex in winter 1997: A new dataset for polar process studies. *Geophysical Research Letters*, **28**(16), 3159–3162.

Staehelin, J., Harris, N. R. P., Appenzeller, C., and Eberhard, J. (2001). Ozone trends: A review. *Reviews of Geophysics*, **39**(2), 231–290.

Stanford, J. L. (1977). On the nature of persistent stratospheric clouds in the Antarctic. *Tellus*, **29**, 530–534.

Stanford, J. L. and Davis, J. S. (1974). A century of stratospheric cloud reports: 1870-1972. *Bulletin American Meteorological Society*, **55**(3), 213–219.

Steele, H. M. and Hamill, P. (1981). Effects of temperature and humidity on the growth and optical properties of sulphuric acid-water droplets in the stratosphere. *Journal of Aerosol Science*, **12**(6), 517–528.

Stephens, G. L. (1994). *Remote Sounding of the Lower Atmosphere*. Oxford University Press.

Stolarski, R. S., Bloomfield, P., McPeters, R. D., and Herman, J. R. (1991). Total ozone trends deduced from Nimbus 7 TOMS data. *Geophysical Research Letters*, **18**(6), 1015–1018.

Tabazadeh, A. and Toon, O. B. (1996). The presence of metastable HNO_3/H_2O solid phases in the stratosphere inferred from ER-2 data. *Journal of Geophysical Research*, **101**(D4), 9071–9078.

Tabazadeh, A., Turco, R. P., Drdla, K., Jacobson, M. Z., and Toon, O. B. (1994). A study of Type I polar stratospheric cloud formation. *Geophysical Research Letters*, **21**(15), 1619–1622.

Tabazadeh, A., Toon, O. B., Gary, B. L., Bacmeister, J. T., and Schoeberl, M. R. (1996). Observational constraints on the formation of Type Ia polar stratospheric clouds. *Geophysical Research Letters*, **23**(16), 2109–2112.

Tabazadeh, A., Martin, S. T., and Lin, J. (2000a). The effect of particle size and nitric acid uptake on the homogeneous freezing of aqueous sulfuric acid particles. *Geophysical Research Letters*, **27**(8), 1111–1114.

Tabazadeh, A., Santee, M. L., Danilin, M. Y., Pumphrey, H. C., Newman, P. A., *et al.* (2000b). Quantifying denitrification and its effect on ozone recovery. *Science*, **288**, 1407–1411.

Tabazadeh, A., Jensen, E. J., Toon, O. B., Drdla, K., and Schoeberl, M. R. (2001). Role of the stratospheric polar freezing belt in denitrification. *Science*, **291**, 2591–2594.

Tisdale, R. T., Glandorf, D. L., Tolbert, M. A., and Toon, O. B. (1998). Infrared optical constants of low-temperature H_2SO_4 solutions representative of stratospheric sulfate aerosols. *Journal of Geophysical Research*, **103**(D9), 25353–25370.

Tisdale, R. T., Prenni, A. J., Iraci, L. T., Tolbert, M. A., and Toon, O. B. (1999). Variation of the infrared spectra of nitric acid hydrates with formation conditions: Impact on PSC identification. *Geophysical Research Letters*, **26**(6), 707–710.

Tolbert, M. A. and Toon, O. B. (2001). Solving the psc mystery. *Science*, **292**, 61–63.

Toon, O. B., Hamill, P., Turco, R. P., and Pinto, J. (1986). Condensation of HNO_3 and HCL in the winter polar stratospheres. *Geophysical Research Letters*, **13**(12), 1284–1287.

Toon, O. B., Browell, E. V., Kinne, S., and Jordan, J. (1990). An analysis of lidar observations of polar stratospheric clouds. *Geophysical Research Letters*, **17**(4), 393–396.

Toon, O. B., Tolbert, M. A., Koehler, B. G., Middlebrook, A. M., and Jordan, J. (1994). Infrared optical constants of H_2O ice, amorphous nitric acid solutions, and nitric acid hydrates. *Journal of Geophysical Research*, **99**(D12), 25631–25654.

Tsias, A., Prenni, A. J., *et al.* (1997). Freezing of polar stratospheric clouds in orographically induced strong warming events. *Geophysical Research Letters*, **24**(18), 2303–2306.

UK Stratospheric Ozone Review Group (1996). Stratospheric Ozone 1996. *Department of the Environment*.

Ulevicius, V. K., Grinshpun, S. A., *et al.* (1997). Aerosolization of paricles form a bubbling liquid: Characteristics and generator developement. *Aerosol Science Technology*, **26**, 175–190.

Voigt, C., Schreiner, J., Kohlmann, A., Zink, P., Mauersberger, K., *et al.* (2000). Nitric acid trihydrate (NAT) in polar stratospheric clouds. *Science*, **290**, 1756–1758.

Waibel, A. E., Peter, T., Carslaw, K. S., Oelhaf, H., Wetzel, G., *et al.* (1999). Arctic ozone loss due to denitrification. *Science*, **283**, 2064–2069.

Willeke, K. and Baron, P. A., editors (1993). *Aerosol Measurement: Principles Techniques and Applications*. John Wiley & Sons.

Wooldridge, P. J., Zhang, R., and Molina, M. J. (1995). Phase-equilibria of H_2SO_4, HNO_3, and HCL hydrates and the composition of polar stratospheric clouds. *Journal of Geophysical Research*, **100**(D1), 1389–1396.

Worsnop, D. R., Fox, L. E., Zahniser, M. S., and Wofsy, S. C. (1993). Vapor-pressures of solid hydrates of nitric-acid – Implications for polar stratospheric clouds. *Science*, **259**(5091), 71–74.

Zakharova, N. T. and Mishchenko, M. I. (2001). Scattering by randomly oriented thin ice disks with moderate equivalent-sphere size parameters. *Journal of Quantitative Spectroscopy & Radiative Transfer*, **70**, 465–471.

Zhang, R., Wooldridge, P. J., Abbatt, J. P. D., and Molina, M. J. (1993). Physical chemistry of the H_2SO_4/H_2O binary system at low temperatures: Stratospheric implications. *Journal of Physical Chemistry*, **97**(28), 7351–7358.

Zhang, R., Leu, M., and Molina, M. J. (1996). Formation of polar stratospheric clouds on preactivated background aerosols. *Geophysical Research Letters*, **23**(13), 1669–1672.

Zondlo, M. A., Hudson, P. K., Prenni, A. J., and Tolbert, M. A. (2000). Chemistry and microphysics of polar stratospheric clouds and cirrus clouds. *Annu. Rev. Phys. Chem.*, **51**, 473–499.

Zurek, R. W., Manney, G. L., Miller, A. J., Gelman, M. E., and Nagatani, R. M. (1996). Interannual variability of the north polar vortex in the lower stratosphere during the UARS mission. *Geophysical Research Letters*, **23**, 289–292.