A New Instrument
for
Atmospheric Aerosol Measurement

A thesis
submitted in partial fulfilment
of the requirements for the Degree
of
Doctor of Philosophy in Physics
in the
University of Canterbury
by
Gareth Thomas

Department of Physics and Astronomy
University of Canterbury
2003

Abstract

This thesis details the development of a novel optical particle counter, the Stratospheric Aerosol Composition and Loading Experiment (SPARCLE), for the measurement of atmospheric aerosols. The instrument has been designed with a view to application as a balloon borne, *in situ* instrument for measurement of stratospheric aerosol, as well as particles in the free troposphere. This study has produced a working, laboratory prototype of the instrument which has provided proof of concept results for the measurement system and associated numerical retrieval scheme.

The SPARCLE measurement system utilizes a laser beam to illuminate individual aerosol particles as they pass through the measurement volume. Separate sensors make high spatial resolution measurements of the scattering pattern and high temporal resolution measurements of the total light scattered across a large solid angle. These measurements are used to infer particle radius and complex refractive index by application of a numerical retrieval scheme that makes use of Mie scattering theory. This retrieval scheme integrates simulated annealing and the Levenberg-Marquardt technique. Application of this scheme to simulated SPARCLE measurements indicates that acceptable estimates of the particle state are produced in approximately 70% of cases. The retrieval scheme was also tested using data from the Dual Amplitude Weighted Nephelometer A aerosol spectrometer. The scheme proved unreliable when particle radius and both real and imaginary parts of the refractive index were retrieved. When only the size and real part of the refractive index were retrieved, the results were in good agreement with the original estimates.

Finally, the retrieval scheme was applied to measurements made by the SPARCLE prototype of monodisperse glycerol aerosols. It was found that, as with the Dual Amplitude Weighted Nephelometer A results, the full 3-

dimensional retrieval was unreliable, particularly in determining particle size. However, if the particle size and real part of the refractive index were retrieved, estimates to within ± 1 μm in radius and ± 0.05 in refractive index were possible for particles larger than ~ 5 μm in radius. For smaller particles refractive index estimates become unreliable, but radius estimates are still accurate, provided that *a priori* knowledge of the refractive index is reasonable.

Acknowledgements

During the course of this PhD I have been aided, encouraged and supported by a small army of individuals – without this help this work would never have reached completion.

Primary thanks must go to my line-up of supervisors. Dr Don Grainger provided the original idea, the first two years of direct supervision for this project and has maintained his involvement as much as possible from his new position at Oxford University. The form and content of this work owes much to his input, and I owe him many thanks. After taking the role of primary supervisor in 2000, Dr Darlene Heuff has provided much support. Her help in developing the retrieval method was of particularly valuable, thank you Darlene. I also owe a great deal of thanks to my third supervisor, Dr. Adrian McDonald. Over the past year he has provided countless pieces of practical advice and holds the speed record for providing thesis revisions. I would also like to acknowledge the support of my mentors, Dr William Tobin and Dr Mike Reid.cb

Projects of this nature are very heavily dependent on quality technical support and, in my case, this support has been superb and wide ranging. My thanks go to all of the support staff in the department, but the following people deserve special mention: Mr Geoff Graham, who constructed most of the data acquisition hardware and provided the programming for the microcontrollers. Mr Ross Ritchie, who designed much of the instrument electronics. And finally, Mr Ron Culley, who built the instrument body and pumps, as well as being heavily involved in their design.

I have also received a lot of help from people not directly involved in the project: Dr Roger Reeves was provided valuable advice on the optical design of the instrument and also allowed me to take over a substantial section of

his optical research laboratory for months on end. Thanks go to Prof Leon Phillips and Dr Chris Knox, from the Department of Chemistry, for the use of the VOAG and APS, and for help in getting the VOAG going. Dr Chris Price, from the Department of Mathematics and Statistics, provided help with the retrieval algorithm at a crucial time. Prof Terry Deshler, from the Department of Atmospheric Science at the University of Wyoming, provided advice and information on general aspects of *in situ* stratospheric aerosol measurement. He was also particularly helpful in aiding the design of the pumps used in the SPARCLE prototype. Thanks also to Prof Jack Baggaley who has provided advice and support on many occasion over the past few years.

The final group of people I would like to thank are those members of the Department of Physics and Astronomy who have had little to do with my PhD, but have provided such a pleasant work environment. There are too many names to list here, but cheers to the lot of you.

Finally, I would like to acknowledge the various funding bodies that have made this study possible. These include the University of Canterbury, which provided the funding for the project itself and provided me with a Doctoral Scholarship. The New Zealand Vice Chancellors Committee provided the funding which enabled me to travel to the Oxford/RAL Spring School in Quantitative Earth Observation in 2000. The Royal Society of New Zealand, the RSNZ Canterbury Branch, the American Geophysical Union and the Department of Physics and Astronomy have also provided funds to allow this work to be presented at the EGS/EUG/AGU Joint Assembly in Nice in April 2003.

Contents

Figures x

Tables xv

1 Introduction 1

2 Background information 5

2.1 Atmospheric aerosol 5

2.1.1 Stratospheric sulphate aerosol layer 5

2.1.2 Polar stratospheric clouds 15

2.2 Tropospheric aerosol 20

2.3 Previous atmospheric aerosol measurements and instruments 25

2.3.1 The Wyoming optical particle counter 25

2.3.2 Other instruments for *in situ* atmospheric aerosol measurement 28

2.3.3 Laboratory measurements of particle scattering 33

2.3.4 Remote measurements of atmospheric aerosol 34

3 Instrument design 43

3.1 Aerosol measurement techniques 44

3.1.1 Selection of the instrument type 47

3.2 Design details 48

3.3 Detection system 48

3.3.1 Light source 49

3.3.2 Photomultiplier 50

3.3.3 Photodiode array 54

3.4 Sample air delivery system 58

3.4.1 Flow measurement 59

3.4.2 Pumps 60

3.4.3 Numerical modelling of air flow 62

3.4.4 Defining the sample volume 73
3.5 Data recording systems 75

4 The retrieval scheme 79
4.1 The forward model 80
4.1.1 Quadrature of the scattering pattern 81
4.2 Constructing the measurement vector 85
4.3 The retrieval method 90
4.4 Retrieval using simulated SPARCLE measurements 93
4.4.1 2-Dimensional retrieval 96
4.4.2 3-Dimensional retrieval 109

5 Analysis of DAWN-A measurements 119
5.1 Normalized measurement vectors 123
5.2 Absolute measurement vectors 131
5.3 Summary of results from the DAWN-A retrievals 136

6 Instrument calibration and testing 139
6.1 Experimental setup 139
6.1.1 Aerosol generation and sizing equipment 139
6.1.2 Experiment configuration and measurement procedure 142
6.2 APS data 149
6.3 SPARCLE measurements 152
6.3.1 PMT data analysis 154
6.3.2 LDA data analysis 156
6.4 Summary 167

7 Analysis of SPARCLE measurements 169
7.1 2-Dimensional retrieval 170
7.2 3-Dimensional retrieval 190
7.3 Summary of retrieval results 210

8 Discussion, conclusions and future work 215
8.1 Instrument design 215
8.2 The retrieval scheme 218
8.2.1 Retrieval of simulated SPARCLE measurements 219
8.2.2 Retrieval of DAWN-A measurements 221
8.3 SPARCLE measurements 224

8.4 SPARCLE measurement retrieval 226
8.5 Summary 229
8.6 Future work 230
8.6.1 An issue raised in the review of this thesis 232

A Mie scattering theory 235
A.1 Analytical inversion of Mie scattering 239

B Retrieval theory 243
B.1 Introduction to the retrieval problem 243
B.2 Simulated annealing 247
B.3 Optimal estimation methods 249
B.3.1 Gauss-Newton iteration 251
B.3.2 The Levenberg-Marquardt method 252
B.4 Information content of the measurement 254

C Instrument hardware 259
C.1 Electronic components 259
C.1.1 LDA circuitry 259
C.1.2 PMT circuitry 268
C.2 Instrument body 275
C.3 Pumps 281

References 286

Figures

2.1 Stratospheric aerosol background size distribution. 8
2.2 Stratospheric aerosol volcanically enhanced size distribution. 9
2.3 In situ measured aerosol mixing ratio history over Laramie, Wyoming and Lauder, New Zealand. 11
2.4 SAGE II stratospheric aerosol climatology: latitudinal distribution. 12
2.5 Example tropospheric aerosol size distributions. 24
2.6 The layout of the Wyoming OPC. 26
2.7 Schematic of a particle levitation instrument. 34

3.1 The instrument layout with the elliptical mirror in place. 52
3.2 The model response curve for the photomultiplier signal with particle size. 53
3.3 The model response curve for the photomultiplier signal with particle refractive index. 54
3.4 Schematic of the instrument layout. 57
3.5 Schematic of the pipe system. 58
3.6 The Gilibrator 2 flowmeter. 59
3.7 The workings of a gear pump. 61
3.8 Flow rate with pump speed for the gear pumps. 62
3.9 Change in pump flow with time. 63
3.10 Model of the scattering chamber before meshing. 65
3.11 Gambit generated grid for numerical modelling of air flow. 66
3.12 The refined grid. 67
3.13 Velocity vectors and inlet flow paths. 69
3.14 Change in particle size in inlet tube. 70
3.15 Pressure variation in sample flow. 71
3.16 Velocity vectors and inlet flow paths for stratospheric conditions. 72
3.17 Schematic of the data collection system. 76

4.1 Quadrature of the scattering pattern. 82
4.2 The standard spherical polar coordinate system. 84
4.3 Information content and number of independent measurements for raw intensity measurements. 87
4.4 Information content and number of independent measurements for the processed measurement vector and non-absorbing particles. 88
4.5 Information content and number of independent measurements for the processed measurement vector and absorbing particles. 91
4.6 Information content and number of independent measurements for the processed measurement vector and absorbing particles. 92
4.7 Particle size distribution used to generate simulated measurements. 95
4.8 2-Dimensional retrieval state vs state. 97
4.9 2-Dimensional retrieval example cost functions. 99
4.10 Detail of Cost function. 100
4.11 χ_s^2 distribution for 2-dimensional retrievals. 103
4.12 Cost function distribution for 2-dimensional retrievals. 104
4.13 2-Dimensional retrieval size distributions. 105
4.14 2-Dimensional retrieval refractive index distributions. 106
4.15 2-Dimensional retrieval size and refractive index error distributions. 107
4.16 3-Dimensional retrieval state vs state. 110
4.17 χ_s^2 distribution for 3-dimensional retrievals. 113
4.18 Cost distribution for 3-dimensional retrievals. 114
4.19 3-Dimensional retrieval size distributions. 115
4.20 3-Dimensional retrieval real refractive index distributions. 116
4.21 3-Dimensional retrieval imaginary refractive index distributions. 117

5.1 DAWN-A retrieval results in the 2-dimensional case with normalized measurement vectors. 124
5.2 DAWN-A retrieval results in the 3-dimensional case with normalized measurement vectors. 125
5.3 Fitted scattering patterns with normalized DAWN-A measurement vectors. 127
5.4 The cost function surfaces for the 2-dimensional DAWN-A retrievals with normalized measurement vectors. 129
5.5 DAWN-A retrieval results in the 2-dimensional case with absolute measurement vectors. 132

5.6 DAWN-A retrieval results for the 3-dimensional case with absolute measurement vectors 133
5.7 Fitted scattering patterns with absolute DAWN-A measurement vectors. 134
5.8 χ^2 surfaces for the 2-dimensional DAWN-A retrievals with absolute measurement vectors. 137

6.1 The TSI model 3450 VOAG. 140
6.2 Experimental layout. 143
6.3 Measured pump flows for all experimental measurements. 144
6.4 Laser power for all experimental measurements. 145
6.5 PMT background measurements. 147
6.6 APS data from run 12. 148
6.7 The average particle size distributions measured by the APS. 150
6.8 The amplitude distribution of a randomly sampled Gaussian. 155
6.9 The distribution of PMT signals for each measurement run. 157
6.10 Measured PMT signals compared to forward model prediction. 159
6.11 Raw LDA data. 160
6.12 LDA measured scattering pattern. 161
6.13 Processed LDA measurement. 162
6.14 Example variances from each measurement run. 165

7.1 Retrieved state cost function distribution for 2-dimensional retrieval run. 171
7.2 Retrieved state χ^2 distribution for 2-dimensional retrieval run. 172
7.3 Retrieved radius distributions from 2-dimensional retrieval run. 173
7.4 Retrieved distribution of the real part of the refractive index from 2-dimensional retrieval run. 175
7.5 Retrieved radii from 2-dimensional run. 177
7.6 Retrieved real refractive index from 2-dimensional run. 179
7.7 Retrieved radius distributions from the 2-dimensional run. 185
7.8 APS versus SPARCLE measured sizes in the 2-dimensional case. 188
7.9 Retrieved state cost function distribution for 3-dimensional retrieval run. 192
7.10 Retrieved radius distributions from the 3-dimensional retrieval run. 193
7.11 Retrieved distribution of the real part of the refractive index from 3-dimensional retrieval run. 195

7.12 Real refractive index distribution produced by the 2-dimensional retrieval using the best 500 measurements. 196
7.13 Retrieved distribution of the imaginary part of the refractive index from 3D run. 197
7.14 Retrieved radii from 3-dimensional run. 199
7.15 Retrieved real refractive index from 3-dimensional run. 201
7.16 Retrieved imaginary refractive index from 3-dimensional run. 203
7.17 Retrieved radius distributions from 3-dimensional run. 207

8.1 Comparison between original and corrected retrieval results. 234

A.1 Example phase functions. 238
A.2 Example phase functions. 239
A.3 Effects of measuring the phase function over a range of wavelengths and/or scattering angles. 240

C.1 The S5462-512Q CMOS Linear Image Sensor: Dimensions. 260
C.2 The S5462-512Q CMOS Linear Image Sensor: Timing. 262
C.3 Circuit diagram of the LDA signal amplifying circuit. 263
C.4 Circuit diagram of the LDA microcontroller circuit. 264
C.5 I/O timing diagram for reading LDA data. 266
C.6 The 9828 photomultiplier tube: dimensions and pin configuration. 270
C.7 The 9828 photomultiplier tube: dimensions of the QL30 stanless steel housing. 271
C.8 Analog to digital conversion circuit for PMT signal 272
C.9 Microcontroller circuit for PMT data collection 273
C.10 Instrument body specifications. 278
C.11 Instrument body specifications. 279
C.12 Instrument body specifications. 280
C.13 Isometric views of the large gear pump. 282
C.14 Dimensions of the large gear pump. 283
C.15 Small gear pump. 284
C.16 Dimensions of the small gear pump. 285

Tables

2.1 Satellite instrument web sites. 41

3.1 Properties of the Fluent standard air model. 65
3.2 Properties of the stratospheric air model. 71

4.1 Limits of allowed states. 93
4.2 Particle distribution used to generate simulated measurements. 95
4.3 States corresponding to the cost functions displayed in Figure (4.9). 98
4.4 2-Dimensional retrieval state distributions χ^2 test results. 109
4.5 Proportions of states where the retrieved refractive index lies within 1σ of the true state. 111
4.6 Comparison between the refractive index errors of absorbing and non-absorbing particles. 112
4.7 3-Dimensional retrieval state distributions χ^2 test results. 118

5.1 DAWN-A derived particle properties. 122

6.1 APS particle size bins. 142
6.2 Pump flow statistics for experimental measurements. 145
6.3 Statistical properties of the APS measured size distributions. 152
6.4 Nominal particle size and number of measurements for each measurement run. 154

7.1 χ^2 comparison between APS and SPARCLE size distributions in the 2-dimensional retrieval case. 183
7.2 Statistical properties of the size distributions given by the 2-dimensional retrieval. 184
7.3 Parameters of the Gaussian fits to the retrieved size distributions in the 2-dimensional case. 187

7.4 Values of the K-S statistic for comparisons between the 2-dimensional retrieval runs. 189
7.5 Statistical properties of the size distributions given by the 3-dimensional retrieval. 205
7.6 Parameters of the Gaussian fits to the retrieved size distributions in the 3-dimensional case. 206
7.7 Values of the K-S statistic for comparisons between the 3-dimensional retrieval runs. 209
7.8 Summary of 2-dimensional retrieval results for each run. 211
7.9 Summary of 3-dimensional retrieval results for each run. 212

C.1 LDA driver connector pins. 265
C.2 Pin connections for the LDA controller ISA bus interface. 269
C.3 PMT microcontroller instruction set. 274

Chapter 1

Introduction

The purpose of this study has been to investigate the feasibility of a new type of *in situ* aerosol photometer capable of determining not only the particle size distribution and number density of an aerosol population, but also the refractive index of the particles. Ultimately this instrument is to be used as part of a stratospheric balloon payload, and thus its design was constrained by the need to produce a small, light instrument capable of autonomous operation in widely varying ambient conditions.

Aerosols play an important role in both the chemical and radiative balance of the atmosphere from ground level to the mid-stratosphere. An overview of the current understanding of b atmospheric aerosols, as well as the techniques used to measure them is given in Chapter (2). In the troposphere, aerosols are one of the key elements controlling cloud formation and growth, they provide sinks for gaseous species and provide sites for heterogeneous chemistry involving atmospheric trace gases (Hobbs, 1993). Anthropogenic emissions form a major source of tropospheric aerosol (for example photochemical smog resulting from automobile exhaust fumes and the enhancement of upper tropospheric clouds along commercial aircraft corridors). However, due to the large spatial and temporal variations in both the sources of aerosol particles and atmospheric conditions, the properties of tropospheric aerosols (such as size distributions and composition), the health implications of exposure to them and their impact on the atmosphere as a whole are generally poorly understood.

Stratospheric aerosols are much more uniform in composition and more stable temporally than tropospheric aerosols and are thus better understood. The

stratosphere is populated by a layer of sub-micron sized sulphuric acid droplets with a background concentration of approximately 10 cm^{-3} (Turco *et al.*, 1982). The cold temperatures and low pressures of the stratosphere limit the rate of most gas phase reactions, making heterogeneous chemistry on aerosol surfaces particularly important. It is now believed that aerosols play an important role in controlling ozone chemistry throughout the stratosphere (Gleason *et al.*, 1993; Herman & Larko, 1994), but nowhere is this role more apparent than in the polar regions (Solomon *et al.*, 1986). The gravitational settling of polar stratospheric cloud particles and heterogeneous reactions which occur on their surfaces remove nitrogen from the polar stratosphere during the winter and release large quantities of reactive chlorine, and thus are a key element in the massive reductions of ozone seen in spring in the Antarctic stratosphere over the past 20 years. Although there has been much intensive study of polar stratospheric clouds in the past 15 years, much still needs to be understood about their formation, composition and structure.

There is a need for improved *in situ* instrumentation in almost every field of aerosol research. *In situ* measurements of aerosol are valuable in their own right, particularly in the lower troposphere where they are often the only techniques used, but are also required to aid in calibration, checking and constraining remote measurements. This is particularly true of stratospheric aerosol where the bulk of measurements come from satellite based radiometric measurements and ground based lidars. However, most in situ measurement techniques are unable to fully characterize the measured particles and thereby require assumptions to be made. For instance, almost all optically based instruments require assumptions to be made about the refractive index of the aerosol being measured to give size estimates. The technology now exists, both in sensors to make the necessary detailed measurements and in computing power to analyse the resulting data, to allow progress to be made in addressing many of the shortcomings of current aerosol measurement instruments.

The ultimate motivation for the development of this new instrument was to better characterize stratospheric aerosol, and thus the instrument has been christened the Stratospheric Aerosol Composition and Loading Experiment (SPARCLE). The two main measurement goals of the instrument are:

1. Size resolved measurements of particles near the peak of the fine mode of

the stratospheric aerosol size distribution ($0.05 - 0.10$ μm in radius).

2. Using angularly resolved measurements of scattering light, in conjunction with Mie scattering theory, to retrieve both particle size and refractive index.

The work presented here has concentrated on the latter of these two goals. The aim has been to determine whether such a measurement is practical and to develop measurement and data analysis techniques which provide the desired quantities in a laboratory setting, while still remaining compatible with application in a small autonomous instrument.

The instrument designed is an optical particle counter which makes measurements of single particle scattering produced as sample aerosol passes through a monochromatically illuminated measurement volume within the instrument. The scattered light is measured by two separate sensors, the first detecting light across a large solid angle and providing a single intensity value, the second providing a high resolution profile of the scattering pattern. The measurement system and instrument design is described fully in Chapter (3). It should be noted that the design of the instrument makes it applicable to a wide range of other aerosol measurement scenarios. Indeed this study suggests that the small size of the stratospheric background aerosol makes the measurement of their light scattering properties a very blunt tool for the determination of their refractive index. However, the ability to measure size and refractive index of spherical particles simultaneously has great potential in many areas, from the characterization of particulate pollution in urban areas, to studies of polar stratospheric clouds.

Of equal importance to the instrument itself is the retrieval scheme which converts the measurements of scattered light into the desired quantities of particle radius and refractive index. A new scheme which combines the properties of simulated annealing and the Levenberg-Marquardt optimal estimation techniques has been created to fulfil this role. The retrieval is described in Chapter (4), as are the results of the analysis of simulated measurements. Chapter (5) presents the results of applying a modified version of the retrieval to measurements made by the Dual Amplitude Weighted Nephelometer A multiangle aerosol spectrometer.

The experimental procedure, data analysis and retrieval results on laboratory measurements made using the SPARCLE prototype are presented in Chapter (6). Measurements of pseudo mono-disperse glycerol aerosols have been made using both SPARCLE and a commercial particle sizer, the Aerodynamic Particle Sizer from TSI. The use of mono-disperse aerosols of known composition and an independent particle size measurement has allowed the performance of the SPARCLE instrument and retrieval scheme to be well characterized.

Chapter 2

Background information

2.1 Atmospheric aerosol

2.1.1 Stratospheric sulphate aerosol layer

The original goal of this study was the development of a particle counter for making *in situ* measurements of stratospheric aerosol. Stratospheric aerosol can be divided into two main categories: (i) the stratospheric aerosol layer and (ii) polar stratospheric clouds. There are also trace amounts of carbon and meteoric dust, plus periodic injections of large quantities of volcanic ash.

Carbonaceous aerosol in the stratosphere mostly originates from commercial aircraft traffic and results from the incomplete combustion of jet fuel. The particles are generally highly non-spherical clusters of graphite fibres and have volume equivalent radii of ~ 0.07 μm and a total loading of approximately 1 kT. Remnants of the ablation of meteoroids in the mesosphere continuously descend into the stratosphere. These remnants are mostly particles of radii $\lesssim$ 10 nm made of oxides of iron and silicon. Large volcanic eruptions inject significant quantities of volcanic ash (of the order of kilotons) into the stratosphere. These particles are typically a few microns in size and settle out of the stratosphere within a few months. Grainger *et al.* (1998) provides further information on the sources of these trace species.

The term stratospheric aerosol layer is usually used to refer to a persistent layer of sub-micron sulphuric acid droplets which peak in concentration at about 20 km. Although the presence of aerosol particles in the stratosphere had long

been observed (e.g. the appearance of noctilucent clouds and observations of haze from aircraft), it was the work of Junge *et al.* (1961); Junge & Manson (1961) which provided the first understanding of their nature. Using a variety of balloon borne in situ instruments the particles were found to be composed of a sulphate solution and an approximate size distribution was obtained. They are now known to originate from injections of sulphur compounds directly into the stratosphere and by photodissociation of carbonyl sulphide (OCS).

OCS was first suggested as a source for stratospheric aerosol by Crutzen (1976). In the troposphere OCS is a relatively inert reservoir species of sulphur and has a constant concentration of approximately 500 ppt. Above the tropopause, however, it is photodissociated by UV light and also reacts with OH radicals and O atoms:

$$\mathrm{OCS} + h\nu \rightarrow \mathrm{CO} + \mathrm{S} \tag{2.1}$$

$$\mathrm{OCS} + \mathrm{O} \rightarrow \mathrm{CO} + \mathrm{SO} \tag{2.2}$$

$$\mathrm{OCS} + \mathrm{OH} \rightarrow \mathrm{CO_2} + \mathrm{HS} \tag{2.3}$$

These reactions reduce the concentration of OCS to approximately 10 ppt by 30 km. The products of reactions (2.1) – (2.3) are then oxidized to SO_2 by reactions with OH, O_3 and HO_2, for example:

$$\mathrm{S} + \mathrm{O_2} + \mathrm{M} \rightarrow \mathrm{SO} + \mathrm{O} \tag{2.4}$$

$$\mathrm{SO} + \mathrm{HO_2} \rightarrow \mathrm{SO_2} + \mathrm{OH} \tag{2.5}$$

$$\mathrm{SO} + \mathrm{O_3} \rightarrow \mathrm{SO_2} + \mathrm{O_2} \tag{2.6}$$

and the SO_2 further reacts with OH to form H_2SO_4 :

$$\mathrm{SO_2} + \mathrm{OH} + \mathrm{M} \rightarrow \mathrm{HOSO_2} + \mathrm{M} \tag{2.7}$$

$$\mathrm{HOSO_2} + \mathrm{O_2} \rightarrow \mathrm{HO_2} + \mathrm{SO_3} \tag{2.8}$$

$$\mathrm{SO_3} + \mathrm{H_2O} + \mathrm{M} \rightarrow \mathrm{H_2SO_4} + \mathrm{M} \tag{2.9}$$

Rosen (1971) was the first to show that stratospheric aerosol was most probably made up of concentrated sulphuric acid particles and found that a 75% solution of H_2SO_4 by weight provided the best fit to measurements. This is now known to be a typical composition, however the exact ratio is highly dependent

on ambient conditions (Yue & Deepak, 1981). Yue *et al.* (1994) used SAGE II water vapour profiles and National Meteorological Center collocated temperature profiles to produce a climatology of stratospheric aerosol composition. Their study assumed that the particles consisted entirely of a H_2SO_4/H_2O solution and extended from $-60°$S to $60°$N in latitude. Their results show a monotonic increase in H_2SO_4 weight percentage with altitude, from values below 60% near the tropical tropopause to over 85%. This increase can be attributed to increasing temperature and decreasing water vapour pressure with altitude. The H_2SO_4 weight percentage at a given altitude is also found to increase with increasing latitude, following the height variation of the tropopause. A seasonal variation in composition is also found at higher latitudes, with particles in the winter hemisphere tending to have lower concentrations. However, this work is based on indirect methods and there is a lack of direct measurements of the composition of stratospheric aerosol. Such measurements are required to provide calibration and constraints for remotely based determinations such as this.

An example size distribution of background aerosol (i.e. formed from OCS) measured by a balloon borne optical particle counter (the Wyoming OPC, see Section (2.3.1)) in 1999 is shown in Figure (2.1). The distribution can be described by a bimodal lognormal distribution with the dominant mode having a median radius of approximately 0.07 μm. The smallest radius measurement shown in this figure (that for $r \geq 10^{-2}$ μm) is made by a condensation nuclei counter, and thus represents a measurement of the total aerosol loading and is of the order of 10 cm^{-3}. It should also be noted that virtually no particles in the background aerosol layer have radii greater than 1 μm.

Large volcanic eruptions result in significant changes in the stratospheric aerosol because they introduce large quantities of sulphur compounds, particularly SO_2, directly into the stratosphere. For instance, McCormick & Veiga (1992) estimated that the Pinatubo eruption resulted in a increase of between 20 and 30 Mt in stratospheric aerosol mass. Initially, volcanic aerosol shows a highly bi-modal size distribution. The numerically dominant mode is similar in width to the background aerosol, however the median radius of the particles is larger. The secondary mode consists of considerably larger particles (radius ≈ 0.5 μm), with a number density similar to that of the aerosol before the eruption (Massie *et al.*, 1996; Deshler *et al.*, 1993). It is theorized that the secondary mode results from condensation of volcanic sulphuric acid

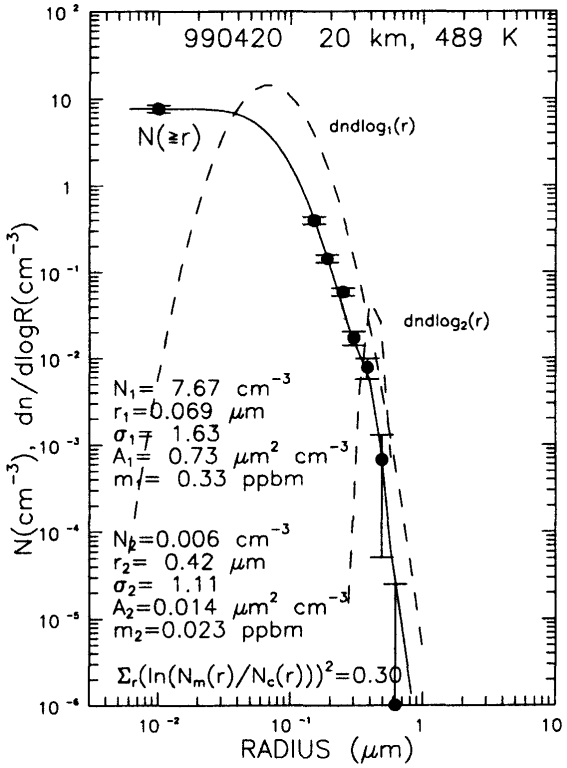


Figure 2.1: A stratospheric aerosol background size distribution at 20 km as measured by the Wyoming OPC and condensation nuclei counter. The measurement points are shown in cumulative form (i.e. all particles greater than the specified radius). Also shown is a bimodal, lognormal fit to the data, with the two lognormal distributions shown as a dashed lines and their associated cumulative concentration plotted as the solid line. The parameters of the fitted distributions, the associated surface areas and volumes as well as the root mean squared error for the fit are also given. From Deshler *et al.* (2003)

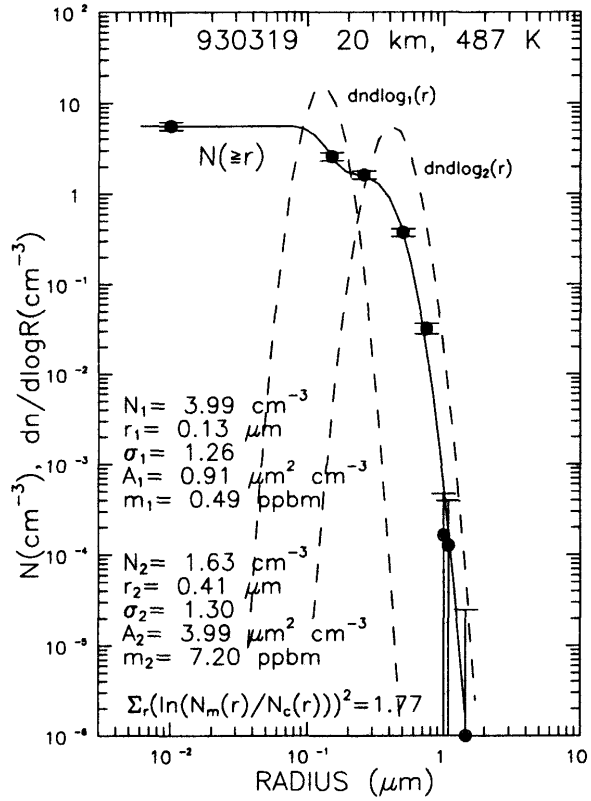


Figure 2.2: As with Figure (2.1) but for volcanically enhanced stratospheric aerosol. From Deshler *et al.* (2003)

vapour on existing aerosol particles, whereas the dominant mode is the result of homogeneous nucleation. Samples of volcanically enhanced aerosol collected by balloon borne impactors have shown that, due to the lack of condensation nuclei found, the nucleation of the particles is predominantly homogeneous (Deshler *et al.*, 1992a). Due to coagulation and gravitational settling the secondary mode then slowly disappears and the median size of both modes decrease.

Figure (2.2) shows a volcanically enhanced size distribution measured by the Wyoming OPC and condensation counter. The measurement was taken at the same altitude and at a similar time of year as that shown in Figure (2.1) but shows conditions less than two years after the eruption of Pinatubo. The differences are clear: firstly the secondary mode is much larger, although centred on the same radius as the secondary mode seen in the background aerosol. The primary mode is also of increased radius by a factor of approximately two. Taken together, these two changes result in an approximately 10 fold increase in the number of radius = 0.1 μm particles and an almost 100 fold increase in radius = 0.5 μm particles. Note that the total number of particles is approximately equal to that observed in the background aerosol, but that aerosol surface area and mass mixing ratio are increased by factors of 6.6 and 21.8, respectively. The increase in surface area is of particular importance because it is a key parameter in controlling heterogeneous chemistry.

Figure (2.3) displays the evolution of volcanic aerosol. Coincident with each major eruption there is a rapid increase in the aerosol mixing ratio, particularly after the Fuego, Mt. St. Helens, El Chichón and Pinatubo eruptions. Coincident with the increase in mixing ratio is a marked reduction in the ratio of radius $\geq$ 0.15 μm to radius $\geq$ 0.25 μm, indicating a shift in the size distribution to larger particles. After each eruption (with the exception of Mt. St. Helens) a slow, approximately exponential decay of the aerosol mixing ratios can be seen spanning several years, with the larger size bin decaying slightly faster than the smaller.

Global climatologies of stratospheric aerosol, such as that presented by Thomason *et al.* (1997) using SAGE II data, show that the aerosol is very evenly distributed zonally. Figure (2.4) shows a higher aerosol surface area at higher latitudes, with peak surface areas at 70°N/S being approximately twice

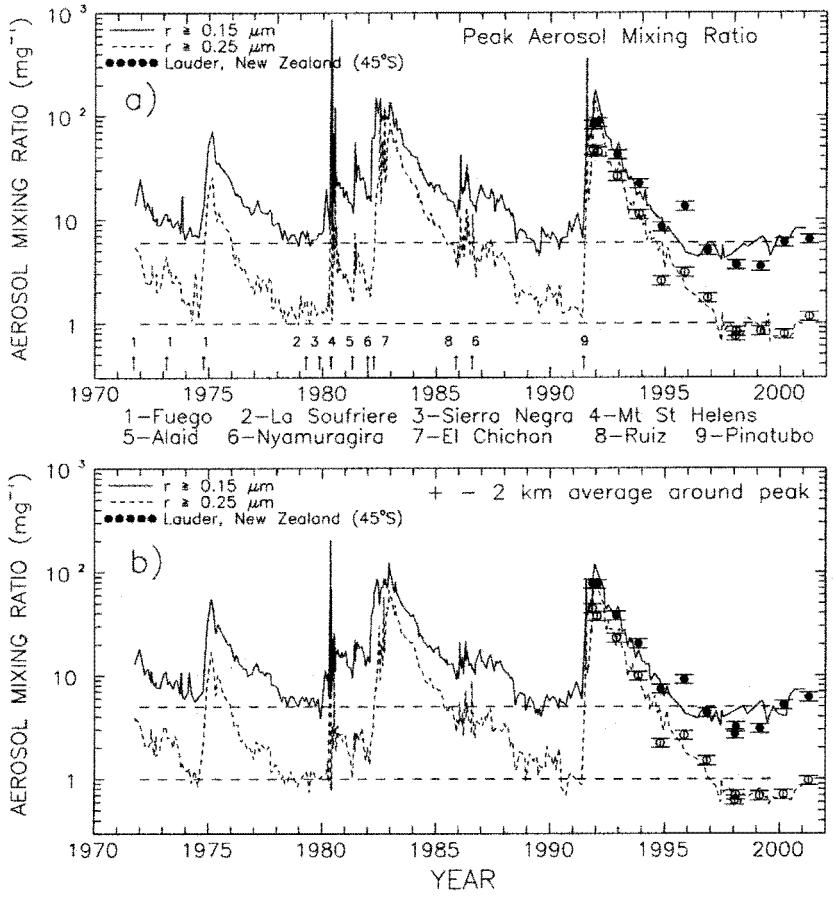


Figure 2.3: Thirty years of *in situ* OPC measured aerosol mixing ratios for particles with radii greater than 0.15 μm and greater than 0.25 μm measured at Laramie, Wyoming (41°N). Also shown is the 10 year record of measurements made using the same equipment at Lauder, New Zealand (45°S). The error bars on the Lauder data are also indicative of those on the Wyoming data and the horizontal dashed lines provide a reference. Plot (a) shows the peak mixing ratio measured and has the times of major volcanic eruptions indicated. Plot (b) shows an average mixing ratio taken over 4 km centred on the peak mixing ratio. From Deshler *et al.* (2003)

those found in the tropical stratosphere. The peak altitude of the aerosol surface area also varies with latitude: at the equator the highest densities are found at approximately 20 km, while in the polar regions they occur between 10 and 15 km. The meridional distribution is mainly influenced by the variation in tropopause height with latitude and by dynamical effects. For instance the strong maxima in aerosol surface area found near the poles can be attributed to downward transportation during the winter and spring (Thomason *et al.*, 1997). The stratospheric aerosol layer also shows an annual variation which is linked with changes in the tropopause height (Hofmann *et al.*, 1975; Hofmann & Rosen, 1981; Barnes & Hofmann, 2001). Vertical profiles show that annual variation is strongest near the tropopause and declines with altitude.

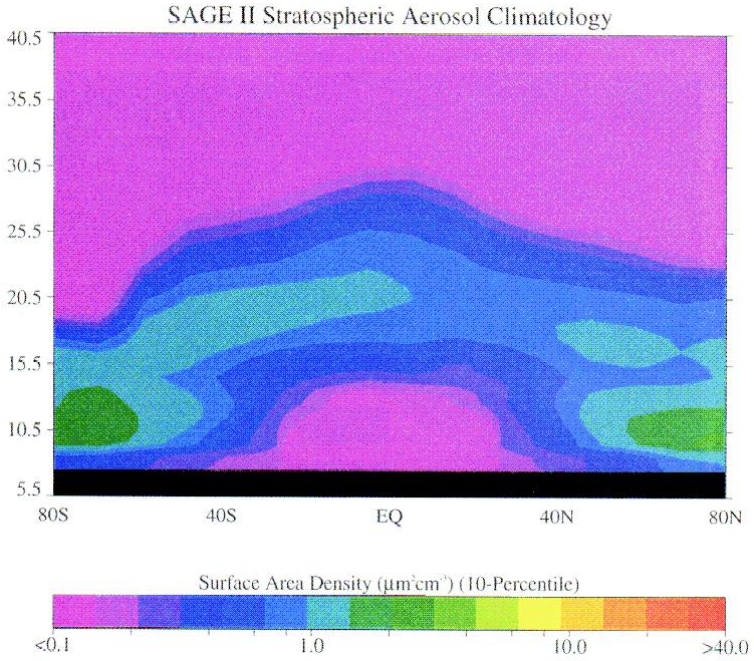


Figure 2.4: The 10-percentile latitude – altitude distribution of stratospheric aerosol surface area density as measured by SAGE II between 1985 and 1994. The use of the 10-percentile value rather than the mean or median value reduces the effects of the El Chichón and Pinatubo eruptions. The y-axis is height in km. From Thomason *et al.* (1997)

Figure (2.3) shows that the past 30 years have been characterized by a series of large volcanic eruptions. This has resulted in stratospheric aerosol being dominated by volcanic enhanced aerosol for the majority of this period.

The best characterisation of stratospheric aerosol prior to this period were the measurements made by Junge *et al.* (1961); Junge & Manson (1961) in the late fifties. Other early measurements produced little information on the character of stratospheric aerosol and first lidar measurement campaigns were started at approximately the same time as the Wyoming OPC measurements (Fiocco & Grams, 1964). The measurements of Junge were made in a volcanically quiet period, but were not numerous enough to give a precise estimate of the background, OCS generated, aerosol level.

Attempts have been made to determine this background level and to measure any long term change using measurements made during the past thirty years (Rosen & Hofmann, 1980; Hofmann, 1990; Thomason *et al.*, 1997). The analyses were performed using measurements made in 1974, 1979 and 1989, under the assumption that these periods represent background aerosol levels, and showed an increase of $\sim 5\%$ in aerosol mass over this period. However the low levels seen in the period extending from the decay of the Pinatubo aerosol add weight to the suggestion made by Thomason *et al.* (1997) that these periods, particularly 1989, were still volcanically enhanced (Hayashida & Horikawa, 2001). It is only since the end of the Pinatubo influence in the late 1990s that the background aerosol levels have been able to be characterized. For instance Barnes & Hofmann (2001) have made measurements of the background levels over Hawaii using lidar. They have found a correlation between the stratospheric aerosol loading and changes in the upward transport of material across the tropopause related to the phase of the quasi-biennial oscillation (QBO).

The large changes in stratospheric aerosol produced by the eruptions of the past 30 years, particularly Pinatubo, have enabled much to be learnt about the nature of the aerosol's effect on the radiative balance and chemistry of the stratosphere. An overview of the important processes influenced by aerosol properties, with emphasis on the effects of the Pinatubo eruption, can be found in McCormick *et al.* (1995). Increased aerosol loading in the stratosphere results in an increase in the albedo of the Earth, resulting in a net cooling of the atmosphere. The stratosphere experiences a heating effect however, due to increased absorption of infrared radiation emitted by the Earth (i.e. an increased greenhouse effect). The size of the particles plays a role in determining how strong each of these effects are, as the absorption of infrared radiation increases

with particle size. Lacis *et al.* (1992) show that if the effective radius of the stratospheric aerosol was to exceed ~ 2 μm there would be a net warming effect on the atmosphere.

Stratospheric aerosol also plays an indirect role in the radiative balance of the atmosphere. A major sink of stratospheric aerosol is transport into the troposphere. Once the aerosol particles enter the troposphere they offer ideal sites for condensation of cloud droplets. An increase in the influx of condensation nuclei into the upper troposphere will result in smaller average cloud particle sizes, increasing the albedo of the cloud, and thus introducing another cooling influence (Sassen, 1992; Twomey *et al.*, 1984).

Study of the effects of the Pinatubo and El Chichón aerosol on ozone has revealed that heterogeneous chemistry plays an important role in the chemistry of the stratosphere. Measurements made by Total Ozone Mapping Spectrometers (TOMS) on board the Nimbus 7 and Meteor 3 satellites after the eruption of Pinatubo showed a gradual reduction in total ozone across the entire globe (Gleason *et al.*, 1993; Herman & Larko, 1994). The reduction continued for approximately two years after the eruption and by mid 1992 the global average was 4% below the lowest level ever measured by TOMS in the 11 year record prior to the eruption. The reduction was also found to be greater in the Northern hemisphere. These observations were supported by ground based Dobson spectrometer measurements from 22 separate stations, which also showed significant drops below the previous trend.

The observation of large reductions in stratospheric ozone levels in conjunction with the enhancement of aerosol surface area, along with evidence of highly perturbed levels of NO_2 and HNO_3 (Johnston *et al.*, 1992; Kawa *et al.*, 1993), strongly suggests that heterogeneous chemistry plays an important role in the chemistry of thestratosphere. Clear evidence of this had only been observed in polar stratospheric clouds (PSCs) prior to the Pinatubo eruption. The main contribution that sulphate aerosols make to the chemistry of the stratosphere is believed to be in converting reactive nitrogen to the reservoir species HNO_3. Substantial drops (by more than 50% in some cases) in NO_2 levels were seen by various instruments (e.g. Johnston *et al.*, 1992; Kawa *et al.*, 1993; Hofmann & Solomon, 1989) following the eruptions of both Pinatubo and El Chichón. Simul-

taneous measurements of NO, NO_2 and HNO_3 (Webster *et al.*, 1994) showed an anti correlation between the NO_x and HNO_3 linked with aerosol loading. Chemical modelling showed that the most likely explanations for these results was a repartitioning of active nitrogen through the reactions:

$$N_2O_5 + H_2O \rightarrow 2HNO_3 \quad (2.10)$$

$$ClONO_2 + H_2O \rightarrow HNO_3 + HOCl, \quad (2.11)$$

where H_2O is in the liquid phase at the surface of an aerosol particle and the products are in the gas phase. As these are heterogeneous reactions, their rates will be greatly increased by an increased aerosol surface area, causing a repartitioning to HNO_3 from NO_x. This results in accelerated ozone destruction, as the various nitrogen oxides (NO, NO_2 and NO_3) are principal sinks of reactive Cl via reactions such as:

$$ClO + NO_2 + M \rightarrow ClONO_2 + M \quad (2.12)$$

2.1.2 Polar stratospheric clouds

Polar stratospheric clouds (PSCs) constitute the second common form of stratospheric aerosol. As the name indicates they are substantial clouds of particles which form in the polar stratosphere during winter and early spring. They play a key role in the chemistry of the polar stratosphere and the creation of the ozone hole.

To understand the formation of PSCs and the chemistry associated with them one must first understand the special conditions that occur in the polar stratosphere. At the most basic level, large scale zonal transport in the middle atmosphere consists of motion from low latitudes towards the poles. In the polar regions the air cools and sinks, and then moves back towards the equator at lower altitudes. A result of this is that large quantities of ozone enriched air, along with molecular reservoirs of chlorine, are transported to the polar regions. This results in the stratosphere at both poles having elevated levels of ozone for much of the year, since the low levels of UV radiation lower the amount of atomic oxygen available for ozone destruction.

In the polar winter the lack of solar heating results in the polar air cooling, sinking and forming a very stable westerly circulation. This phenomenon

is know as the polar vortex. The Antarctic vortex is generally much more stable than the Arctic one, due to the higher planetary wave activity caused by surrounding land masses in the Northern hemisphere. This in turn results in lower temperatures in the Antarctic winter stratosphere and a longer lived vortex. In both hemispheres, the vortex virtually isolates the air in the polar stratosphere from the surrounding atmosphere, and makes it a closed chemical system for the duration of the winter and early spring. It is not until the return of solar heating that the vortex breaks up and the polar air can mix with the free atmosphere, usually in November for the Antarctic vortex.

The lack of solar heating or mixing with external air leads to extremely low temperatures (approaching 180 K) within the polar vortex. As the temperature within the vortices drops the sulphate particles making up the stratospheric aerosol layer absorb more water. By 195 K they contain $\sim$ 40% H_2SO_4 as opposed to the 75 – 80% found at 220 K (Toon *et al.*, 1986). At temperatures below approximately 205 K the sulphate particles also take up nitric acid vapour. This process produces particles which are much larger than the original sulphate aerosols and whose composition is dominated by HNO_3 (Molina *et al.*, 1993; Tabazadeh *et al.*, 1994; Carslaw *et al.*, 1994). These particles are referred to in the literature as type I PSC particles and are subdivided into type Ia or Ib depending on the phase of the HNO_3. Type Ia particles consist of solid hydrates of nitric acid, either nitric acid trihydrate (NAT) or nitric acid dihydrate (NAD). Type Ib particles are made up of super-cooled ternary solutions of HNO_3, H_2O and H_2SO_4.

The second broad category of PSC particle (referred to as type II) is predominantly made up of water ice. These particles only form at temperatures below the frost point, which is approximately 190 K in the polar stratosphere due to the very low water vapour pressure. Their growth process is thought to be condensation of water on pre-existing type I particles and they can grow large enough to undergo rapid gravitational settling, thus permanently removing the nitrogen and water they contain from the stratosphere.

PSC particles show a wide range of shapes. Liquid particles (type Ib) are highly spherical, but the solid particles (types Ia and II) show a wide range of crystalline shapes. There are very few measurements of PSC particle structure,

however Goodman *et al.* (1997) used the NASA Ames wire impactor (AWI) to collect two samples of PSC particles. In both cases the PSC sampled was believed to be predominantly type Ib with some type II particles present. The ice particles observed were mostly classified as hexagonal plates ($\sim 50-60\%$ by number) or columns ($\sim 25-28\%$) with the remainder being pentagonal or triagonal crystals. The average radii observed for the particles ranged between approximately 6 and 10 μm and the average length of the particles was approximately 21 μm. Laser backscatter depolarisation measurements from lidar observations suggest that type Ia NAT particles are slightly aspherical although their structure varies greatly with the temperature history of the particles (Poole & McCormick, 1988). Measurements made by Hofmann & Deshler (1991) using optical particle counters found median radii for type I PSC particles in the range 1.5 to 3.5 μm.

The mechanisms by which these particles form are highly dependent on both atmospheric conditions and the rate of change of these conditions. The rate at which the sulphate aerosol takes up HNO_3 is very dependent on the temperature (Bertram *et al.*, 2000). If a particle is rapidly cooled, by passing through an orographically induced gravity wave for instance, the proportion of HNO_3 contained within the resulting particle will be much higher than in undisturbed conditions. The freezing point and crystalline form of nitric acid ices is, in turn, highly dependent on the concentration of the solution it is condensing from, as well as the temperature. Finally, these dynamics are also considerably different in aerosol particles than in the bulk material. Thus, determining what type of PSC will form, even with knowledge of stratospheric conditions, is not yet possible. This is a considerable gap in knowledge since these particles play a vital role in the chemistry of the polar stratosphere. Attempts to determine the composition and phase of PSC particles using combinations of remote and *in situ* measurements have produced ambiguous results which do not agree well with model and theoretical predictions (Deshler *et al.*, 2000; Larsen *et al.*, 2000). Because their chemical properties vary strongly with changing composition it is important that measurements distinguishing one type from another are made to aid in realizing the reliable prediction of PSC properties.

In 1985 results were published from a British Antarctic Survey campaign of total column ozone measurements made at Halley Bay in Antarctica (Farman *et al.*, 1985). The measurements showed a massive annual decrease in

stratospheric ozone during spring (September to October). These results were then confirmed by reanalysis of data from Total Ozone Mapping Spectrometer (TOMS) and Solar Backscatter Ultraviolet (SBUV) instruments (both on board the Nimbus 7 satellite). Subsequent ozone height profile measurements revealed that the ozone depletion was occurring between 10 – 20 km (the lower stratosphere) and was almost complete within this region (Hofmann *et al.*, 1987). The known mechanisms for ozone destruction involving the ClO_x catalytic cycle were unable to explain these observations. The usual ClO_x cycle attributed with stratospheric ozone destruction requires large quantities of photodissociated oxygen, which is not available in the polar stratosphere. Furthermore, ozone loss due to gas phase reactions involving halogens and chlorofluorocarbons (CFCs) would be most efficient at ~ 40 km, not the 10 – 20 km observed.

The now-accepted mechanism describing the destruction of ozone in the polar stratosphere was proposed by Solomon *et al.* (1986). The initial step in the process is the absorption of gaseous HCl by PSC particles. HCl is usually a long lived reservoir species of chlorine in the stratosphere and its solubility in normal stratospheric aerosol particles is low. At temperatures below 200 K PSC particles absorb water however, and their reduced acid concentration makes the absorption of HCl efficient. The presence of an ice surface allows the reaction of HCl with another chlorine reservoir species, $ClONO_2$, to take place at a much higher rate than it can in the gas phase (Crutzen & Arnold, 1986; Molina, 1991; Webster *et al.*, 1993):

$$HCl(s) + ClONO_2 \rightarrow Cl_2 + HNO_3(s) \tag{2.13}$$

where (s) denotes the substance is present on the surface of a PSC particle. Similarly $ClONO_2$ can react with water ice on a particle's surface:

$$ClONO_2 + H_2O \rightarrow HOCl + HNO_3(s) \tag{2.14}$$

which can undergo a further reaction,

$$HOCl + HCl(s) \rightarrow Cl_2 + H_2O \tag{2.15}$$

Finally, in the presence of N_2O_5 the following reaction can occur (McElroy *et al.*, 1986a):

$$N_2O_5 + HCl(s) \rightarrow ClNO_2 + HNO_3(s) \tag{2.16}$$

Once HNO_3 is within the PSC particles it is trapped there until the particle re-evaporates and may even be removed from the stratosphere entirely by the settling of larger particles. All of the chlorine products of the above reactions (i.e. Cl_2, HOCl which has not undergone reaction (2.15) and $ClNO_2$) will remain in the gas phase until the return of sunlight in spring. In the presence of sunlight all of these species rapidly photodisociate, releasing Cl atoms.

The above scheme is based around the release of reactive chlorine. The increase in stratospheric chlorine due to the release of man-made chlorofluorocarbons into the atmosphere is the root cause of the ozone hole. However the chlorine catalytic cycle is not the only one which plays a role in ozone destruction: for example there are schemes involving bromine, as described by McElroy *et al.* (1986b). It is, however, the role that PSCs play in the chlorine cycle that is of greatest importance.

PSCs persist well into spring and thus virtually all of the reactive nitrogen (in the form of HNO_3) in the polar spring stratosphere remains trapped in particulate form. The almost complete lack of reactive nitrogen means that the reactive forms of chlorine released by the heterogeneous PSC reactions cannot be converted back into reservoir species. The free Cl atoms then begin to destroy ozone via the reaction (McElroy *et al.*, 1986a,b; Molina & Molina, 1987; Webster *et al.*, 1993):

$$\mathrm{Cl} + \mathrm{O_3} \rightarrow \mathrm{ClO} + \mathrm{O_2} \tag{2.17}$$

Due to the lack of atomic oxygen in the polar stratosphere, the usual catalytic chlorine cycle, which is closed by the reaction

$$\mathrm{ClO} + \mathrm{O} \rightarrow \mathrm{Cl} + \mathrm{O_2}, \tag{2.18}$$

cannot proceed. The concentration of ClO therefore rises, until another catalytic becomes efficient (Molina & Molina, 1987):

$$\mathrm{ClO} + \mathrm{ClO} + \mathrm{M} \rightarrow \mathrm{ClOOCl} + \mathrm{M} \tag{2.19}$$

$$\mathrm{ClOOCl} + h\nu \rightarrow \mathrm{Cl} + \mathrm{ClOO} \tag{2.20}$$

$$\mathrm{ClOO} + \mathrm{M} \rightarrow \mathrm{Cl} + \mathrm{O_2} + \mathrm{M} \tag{2.21}$$

$$2\,[\mathrm{Cl} + \mathrm{O_3}] \rightarrow 2\,[\mathrm{ClO} + \mathrm{O_2}] \tag{2.22}$$

$$\text{Net}: \quad 2O_3 + h\nu \rightarrow 3O_2$$

2.2 Tropospheric aerosol

Although the stated aim of this project was initially to design an instrument suitable for *in situ* measurement of stratospheric aerosol, it is likely that the measurement of tropospheric aerosol offers as much, if not more, scope for useful application of the type of instrument created. The much wider range of sizes and compositions of tropospheric aerosol, coupled with the obvious reduction in the complexity of getting the instrument *in situ* to begin with, make it likely that the first atmospheric measurements made by SPARCLE instruments will be of tropospheric aerosol. In particular, the larger size and wide range of composition (and hence refractive index) of many tropospheric aerosol populations lend themselves to measurement by the SPARCLE system.

As noted above aerosol populations in the troposphere are more varied than in the stratosphere and show large temporal variations and dependence on local conditions. Seinfeld & Pandis (1998) and Hobbs (1993) both give a good summary of the properties of tropospheric aerosol and details of the chemistry associated with major populations.

Particulate matter near the ground is generated by a vast number of processes and results in a very wide variety of particles. Natural sources include:

- Marine salt aerosols from the evaporation of sea spray Fairall *et al.* (1983). These particles are fairly confined to the marine environment and will consist mostly of approximately cubic salt crystals and clusters thereof. Their production and distribution varies very strongly with ambient sea and air conditions.

- Wind blown dust and soil. This is particularly prevalent in arid or desert regions (d'Almeida & Schütz, 1983), but makes up a significant portion of aerosols seen in heavily vegitated regions (Artaxo & Maenhaut, 1988) and can be transported thousands of kilometres by the wind (d'Almeida, 1986). These particles will have a range of irregular shapes depending on the size and material of the particle and on the process by which the material became pulverised.

- Plant material. This includes pollen, plant fragments and oils/waxes released by plants, particularly when damaged (Artaxo & Maenhaut, 1988; Schnell & Tan-Schnell, 1982). Pollen grains have a multitude of shapes, most of which are based around fairly regular polygons. Plant fragments tend to be either single long fibres or conglomerations thereof. The shape of wax particles will largely depend on their formation process, while liquid phase oils will form spherical droplets. As with particle shape, the spatial distribution of plant material is dependant on its type. Generally, plant related aerosols are locally concentrated. However, wind borne pollen (for instance) can travel many hundreds of kilometres.

- Microbial life. Fungal spores, bacteria and viruses are all known to be present in the air. Very little is known about the amount and distribution of microbiological aerosol.

- Condensation of the products of oxidation and other chemical reaction products forming particles of nitrate, sulphate and ammonium solutions or organic compounds. Particles of this nature make up the so-called nucleation mode found in tropospheric aerosol (see below). They are almost universally sub-micron spherical droplets and are ubiquitous throughout much of the troposphere.

- Cloud processes. Clouds form both a sink (through precipitation) and source of aerosol particles. Cloud particles are efficient scavengers of existing aerosols and are sinks for gas phase compounds such as nitrates and sulphates. If a cloud droplet evaporates it will leave behind a particle made up of this collected material. This process has been suggested as the source of much of the uniform background aerosol found in the troposphere, particularly the larger particles (Hobbs, 1993). Particles formed from this process will generally be approximately spherical clusters of crystalline fragments.

Anthropogenic sources, which usually dominate the aerosol population in urban areas, include:

- Combustion products, both primary and secondary (Seiler & Crutzen, 1980; Spurny, 2000). These result from the burning of anything from petrol to rubbish and include organic compounds, sulphates and nitrates, soot and fine ash, metal oxides and the products of secondary photochemical reactions (such as photo-chemical smog). Solid particles produced in this

fashion will generally be highly non-spherical clusters of material. Soot particles, for instance, are generally made up of clumps of graphite chains. Liquid particles that form from secondary reactions of sulphates and nitrates will be spherical, but in highly polluted areas may well contain inclusions of solid particles. The distribution of this type of aerosol depends largely on where it is generated. Urban pollution (from heating and automobiles) is usually reasonably confined to a specific location as dilution occurs rapidly away from the source.

- Commercial aircraft traffic. This is really a sub-set of the previous category, but is unique enough to warrant separate attention. The natural background aerosol loading in the upper troposphere is very low, so the large increase in the amount of aircraft traffic in this region in the last 30 years has the potential to have a large effect. The exhaust from jet aircraft affects the particulate loading of the upper troposphere in two ways. Firstly it introduces carbon soot particles directly into this region; secondly it raises the partial pressures of water and sulphate compounds. The introduction of carbonaceous aerosol into the upper troposphere provides extra nucleation sites for secondary particles. Also, if amounts were substantial enough, it could have an impact on the radiative balance of the region, due to the high absorption of carbon across a wide range of wavelengths. The introduction of water vapour is responsible for the enhancement of cirrus clouds (Ström & Ohlsson, 1998; Lawson *et al.* , 1998) which in turn can also affect the radiative balance of the region.

- Mechanical processes. Mechanical processes such as car tyre and brake wear, and industrial processes, produce coarse aerosol made up of minerals, metal compounds and rubber etc. Particles of this nature will have a vast array of shapes and sizes and will usually only be found close to their source.

- Enhancement of natural sources. Agricultural activity, such as ploughing and harvesting can greatly enhance the quantities of plant and soil aerosol present.

As well as fairly continuous sources there are also short-term events, such as volcanic eruptions and forest fires, which introduce large quantities of particulate matter into the atmosphere. Although the perturbations introduced

by these events are often dramatic they are usually short lived in the troposphere.

The particles in the free troposphere (above about 5 km) are much more homogenised. As mentioned above, cloud processes have been postulated as a major source of these particles. Clouds are estimated to produce $\sim 3000\ \mathrm{Tg\,yr^{-1}}$ of aerosol (Hobbs, 1993), as opposed to $\sim$ 1300 $\mathrm{Tg\,yr^{-1}}$ for homogeneous nucleation and settling of stratospheric particles.

Tropospheric aerosol is often described in terms of surface area or volume distributions and is usually best described by a tri-modal lognormal distribution (Hobbs, 1993). The three modes are referred to as the nucleation mode (radius $\lesssim 0.05\ \mu$m), the accumulation mode ($0.05 \lesssim$ radius $\lesssim 0.5\ \mu$m) and the coarse mode (radius $\gtrsim 0.5\ \mu$m). These boundaries are only approximate and depend on which aerosol population is being examined, for instance the coarse mode starts at a radius of 0.3 μm for marine aerosol. The distribution of the particles among these three modes also varies strongly with different situations, but in general most particles occur in the smallest mode, while the aerosol surface area and volume distributions are dominated by the larger modes. Figure (2.5) shows example size distributions for various populations of tropospheric aerosol. The urban, rural, polar, desert dust storm and remote continental distributions should be viewed as examples of the aerosol population near the ground in the appropriate locations. The background aerosol, in contrast, represents the aerosol distribution found in the free troposphere.

Clearly, most of the populations shown in Figure (2.5) can be viewed as enhancements on the relatively uniform background aerosol. The exceptions to this are the polar and remote continental aerosol. The polar aerosol population shown here is compiled from measurements from both the Arctic and Antarctic. The polar regions are relatively free of aerosol sources, so this population can be considered to be aged in character. Given this, we can see that the smallest particles (those in the nucleation mode) have the shortest lifespan, probably due to processing by clouds and to coagulation. Interestingly, the largest particles show no departure from the free troposphere background level. Since these particles can also be expected to have a short lifetime (due to gravitational settling) this suggests that these particles are probably being produced in the polar troposphere. The remote continental size distribution also shows a

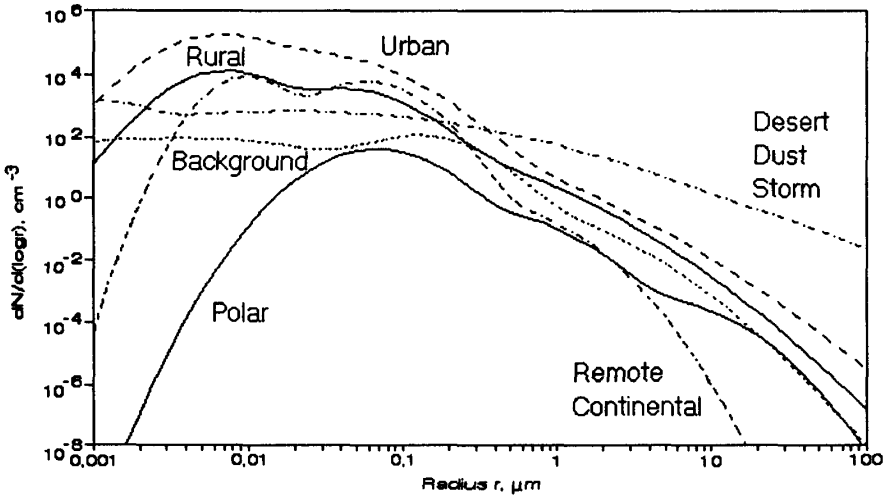


Figure 2.5: Example tropospheric aerosol size distributions compiled from a wide range of measurements. After Hobbs (1993).

dearth of small particles, but unlike the polar population it shows a lack of large particles. There is a general lack of measurements in remote continental sites and this distribution is based on one set of measurements made in Siberia (Koutsemogii, 1992). However, it does suggest the scale of variation in the background level, since if large particles were present in the free troposphere above a measurement site their high settling speed would make it likely that they would be observed at ground level.

The impact of human activity can clearly be seen in Figure (2.5) with both rural and urban aerosol populations showing a strong enhancement over the background level. In particular the urban environment sees an increase in particles in the nucleation and accumulation modes of $\sim$ 3 orders of magnitude. This points to the large number of particles produced by nucleation of combustion products. It should be noted that urban aerosol population is one of the most variable and this distribution can only be taken as indicative. The desert dust storm population shows the very uniform size range of particles that originate from the suspension of eroded soil and rock. Although this distribution is characteristic of the particles produced at the source, such material can be transported thousands of kilometres once airborne. Because of the large number of big particles produced by desert dust storms these events have a major effect on aerosol surface area, volume and mass. A major tropospheric aerosol

distribution not included in Figure (2.5) is the marine, which shows very large variations with both sea and atmosphere.

There is a need for further measurements of tropospheric aerosol in almost every area. For example most measurements monitoring urban aerosol are currently based around the determination of PM_{10} (total particulate matter of diameter less than 10 μm) concentrations and collection of particles using filters and impactors. The ability to measure particle concentration, size distribution and refractive index within one instrument would enable a much more complete picture of aerosol populations to be built. This in turn would simplify the task of identifying source and transport mechanisms, as well as aiding the characterisation of chemical processes which may be occuring within the particles, and potential adverse effects of the aerosol.

Hobbs (1993) cites the lack of instrumentation designed for measuring atmospheric aerosol properties as reason for the current uncertainty on the sources, sinks and properties of tropospheric aerosol as a whole. An instrument such as the one discussed in this study has the potential to go some way towards filling this gap.

2.3 Previous atmospheric aerosol measurements and instruments

2.3.1 The Wyoming optical particle counter

The instrument produced in this study owes a lot of its conception and some features of its design to the series of balloon borne optical particle counters operated first by the University of Minnesota and then by the University of Wyoming. This series of instruments, referred to here as the Wyoming OPC, was first introduced in 1963 (Rosen, 1964) and the design has been in use continuously since then with only relatively minor modifications.

In its original configuration the OPC consisted of an incandescent lamp focused onto a sample volume through which ambient air was pumped. Light scattered by particles as they passed through the sample volume was collected

by a microscope set off-axis to the light beam from the lamp. The microscope focused the sample volume onto the photocathode of a photomultiplier tube (PMT) and the output of this was passed through a two channel pulse height discriminator, providing some particle sizing ability. In this configuration measurements of the total number density of particles with radius$>$ 0.28 μm and with radius $>$ 0.38 μm were obtained with an uncertainty of $\sim$ 10% or better.

Later a second microscope/PMT combination was added (Rosen, 1967, 1968). The second detector viewed the same range of scattering angles as the first and was included to avoid false counts due to cosmic ray strikes in the PMTs and to improve the signal to noise ratio. A schematic of the instrument layout in the new configuration is shown in Figure (2.6). This diagram can be taken as indicative of all Wyoming OPC instruments post 1967. In this configuration the instrument's detectors were centred at an angle of 23° from forward scatter and subtended an angle of 30°. The system still used a two level pulse height discriminator, but the size ranges were then radius $>$ 0.15 μm and radius $>$ 0.25 μm. The sample flow used in this counter was 1 l/min.

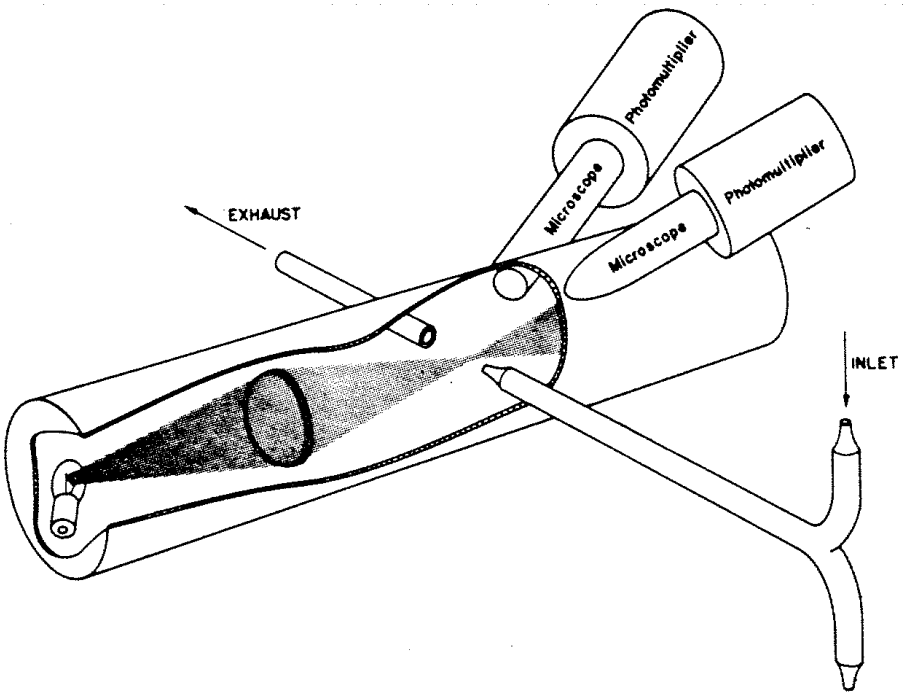


Figure 2.6: A schematic of the Wyoming OPC with the dual detector arrangement. After Rosen (1967).

The next modification made to the OPC was the addition of a condensation nuclei (CN) counter (Rosen & Hofmann, 1977). This type of instrument passes the sample air through a region containing some super-saturated vapour (glycol in the case of the Wyoming instrument). Any particles contained in the sample flow act as condensation nuclei for the vapour and grow many orders of magnitude in size. The enlarged particles are then detected by the OPC in the usual fashion. The use of the CN counter provided a third size bin of all particles greater than approximately 0.01 μm in radius.

Extensive calibration work in the 1970s (Pinnick *et al.*, 1973; Pinnick & Hofmann, 1973) showed that the instrument's response with respect to particle size became multi-valued for particles in the radius range 0.5 to 2 μm. It was also found that the instrument's response showed large variations with refractive index for particles larger than $\sim$ 0.2 μm in radius[1]. The combination of these effects was found to result in large departures in the counting efficiency from unity for particles larger than $\sim$ 0.7 μm. It should be remembered, however, that stratospheric aerosol particles are generally smaller than this.

During the early 1970s regular flights of the OPC were begun from Laramie, with roughly monthly launches from 1972 to the present day. This impressive data set has been used to determine stability and long term trends in the stratospheric aerosol layer (Hofmann *et al.*, 1975; Hofmann, 1990), the impact of various volcanic eruptions on the number of particles in the stratosphere (Hofmann & Rosen, 1977; Rosen & Hofmann, 1980; Deshler *et al.*, 1992a, 1993) as well as measurements of tropospheric aerosol profiles (Hofmann, 1993).

In the mid 1980s measurements of polar stratospheric clouds were made with the OPC. For these measurements a third pulse height discriminator level was added to detect large particles (radius $\geq$ 1 μm) (Hofmann *et al.*, 1989). Although these measurements were able to determine that the size distribution of the particles within the clouds was bi-modal, with a secondary mode of particles larger than approximately 0.3 μm, the instrument was barely able to detect the low PSC particle concentrations encountered ($< 10^{-2}$ cm^{-3}). Furthermore, the multivalued region of the instrument's size response curve occurred over the precise size

[1]These effects are due to the complex nature of the scattering pattern produced by spherical particles and its dependence the particle refractive index (see Chapter 3).

range of interest for type I PSC particles ($0.3 \lesssim$ radius $\lesssim 1$ μm). To overcome these problems three modifications were made to the instrument (Hofmann & Deshler, 1991):

1. The sample flow rate was increased from 1 l/min to 12 l/min.

2. The detectors were moved to be centred at scattering angles of 40°. This greatly reduced the dependence of instrument response on particle refractive index and made the size determination essentially single valued in the region of interest at the expense of reducing the signal strength for the larger particle sizes by a factor of approximately two.

3. A seven channel pulse height discriminator was added, providing measurements of particle concentrations for radii greater than 0.15, 0.25, 0.5, 1.0, 2.0, 3.0 and 5.0 μm.

This form of the instrument, especially when used in conjunction with the CN counter described above, allows the height-resolved measurement of the particle size distribution (see, Figures (2.1) and (2.2)). This form of the instrument remains in use today. The limitations of the instrument lie in its inability to determine particle refractive index and/or shape. Calibration is based on calculations and measurements of the instrument response to spherical particles of known size and refractive index. For most non-absorbing atmospheric aerosol (with a real part of the refractive index $1.33 < \mathrm{n} < 1.50$) this does not introduce a sizing error of more than $\sim 5\%$. However, if absorbing particles and/or non-spherical particles are encountered considerably larger sizing errors will result. Hofmann (1993) quotes a sizing error of 250% for radius= 1.0 μm particle with a complex refractive index of $\mathrm{m} = 1.67 - 0.26i$ for instance. Another limitation of the instrument is its inability to make size-resolved measurements at or below the peak of the condensation mode (see Figure (2.1)). Ideally the measurements need to be made to a radius of ~ 0.05 μm to fully characterize this mode. It should be noted, however, that measurement to this size range is probably beyond the capabilities of current technology to achieve in a small *in situ* instrument.

2.3.2 Other instruments for *in situ* atmospheric aerosol measurement

Although the Wyoming dust sonde is the most important instrument currently used for *in situ* measurement of atmospheric aerosol particles from the view

point of this project (being a balloon-borne optical particle counter for use in the stratosphere), it is by no means the only such instrument. There is a wide range of both balloon-borne and aircraft-mounted systems in use, particularly for tropospheric aerosol measurement. This section outlines some of the instruments which can be used to make *in situ* measurements of stratospheric and free tropospheric aerosol. Measurement techniques for boundary layer measurements are not discussed.

Sugita *et al.* (1999) present the design and initial measurements made by a balloon-borne optical particle counter for stratospheric aerosol measurement. The design owes much to the Wyoming OPC. As with the Wyoming instrument the system uses a halogen lamp as a light source and uses two symmetrically mounted photomultiplier tubes, which view the sample volume through microscope optics. These detectors are centred at a scattering angle of 60° and the system uses a sample air flow rate of 73 mls^{-1}. As with the Wyoming OPC the system uses a seven channel pulse height discriminator, allowing the measurement of a cumulative size distribution. The discriminator is set to provide size bins for particle radii larger than 0.17, 0.20, 0.25, 0.30, 0.40, 0.50 and 0.75 μm and is calibrated to spheres with refractive index $n = 1.45 - 0i$.

Another significant balloon borne aerosol instrument is the backscatter sonde (Rosen & Kjome, 1991). This instrument uses a strobe light projecting from the side of the balloon payload to make measurements of aerosol backscatter. The scattered signal is detected by two PMTs, each of which is behind a different optical filter, to produce information in two wavelength bands. The instrument produces similar results to a lidar (i.e. backscatter ratio and/or depolarization) and has a sample volume which extends only a few meters from the balloon package. Because of its relative simplicity and low cost it has been used in conjunction with more complicated instrument platforms, either to initially determine the presence of an aerosol layer (Schreiner *et al.*, 1999), or as an additional detector on a large gondola. These same features also make the backscatter sonde ideal for ongoing monitoring of aerosol height profiles, such as the campaign described by Liley *et al.* (2001).

The backscatter method is also used by the Laser Backscatter Sonde (LABS) instruments (Adriani *et al.*, 1999). These instruments use two diode lasers of

different wavelengths and two avalanche photodiodes equipped with narrow band filters to make dual wavelength backscatter measurements. The system can also make depolarization measurements by using only one laser wavelength and adding polarizing filters to the detectors.

Arnold *et al.* (1998) describe a balloon-borne, in situ, mass spectrometer for measuring the sulphuric acid content of stratospheric aerosol. This system vaporizes the aerosol particles contained within the sample air stream and then measures the gaseous sulphuric acid using a quadrupole mass spectrometer. Comparison to particle volume measurements from Wyoming OPC and condensation nuclei measurements showed good agreement for a sulphuric acid concentration of $\sim 70\%$ in stratospheric particles.

The airborne polar nephelometer (Gayet *et al.*, 1997a,b) is an aircraft mounted instrument that measures cloud particles by collecting the light scattered by multiple particles as they pass through a laser beam. The scattered light is collected by a parabolic mirror and reflected onto an array of 33 photo diodes, allowing measurement of the scattering pattern produced. The instrument is sensitive to particles of a few microns up to a millimetre in radius, the exact range being determined by the concentration of the particles and the sampling rate used. Oshchepkov *et al.* (2000) and Gayet *et al.* (1998) present results of an inversion method for retrieving particle size distributions for mixed phase ice/water clouds using polar nephelometer data.

Several commercially produced aerosol instruments are available, some of which are suitable for use on board an aircraft for *in situ* atmospheric work – most notably the Forward-Scattering Spectrometer Probe (FSSP) series (Dye & Baumgardner, 1984; Baumgardner *et al.*, 1985) made by Particle Measuring Systems (PMS) and the instruments produced by TSI. See Willeke & Baron (1993) for a description of most of the major comercial instruments. These instruments are widely used in measurement campaigns and have generally been very extensively characterized. An example of this is the SUCCESS (Subsonic Assessment: Cloud and Contrail Effects Special Study) programme (Heymsfield *et al.*, 1998b,a; Anderson *et al.*, 1998a,b), for studying the creation of contrails, which used TSI condensation nuclei counters.

The Multi Angle Aerosol Spectrometer Probe (MASP) (Baumgardner *et al.*, 1995) is an aircraft mounted single particle counter which attempts to measure not only the particle size but also the refractive index. The method used to do this is to compare the light scattered in the forward direction (scattering angles 30° to 55°) with that scattered in the backward direction (120° to 145°). The method used to determine particle composition from this instrument is to compare the forward to back scattered ratio (FBSR) with that expected for various types of aerosol particle, such as sulphuric acid, soot and water droplets.

In an attempt to improve on the performance of the commercial FSSP probes and the MASP, a new sensor, the Cloud Aerosol and Precipitation Spectrometer (CAPS), has been developed by Baumgardner *et al.* (2001). The probe contains the Cloud and Aerosol Spectrometer (CAS) which combines the measurement methods used in both the FSSP probes and MASP. The layout of the CAS sensor is similar to that of the FSSP probes but it measures both forward-scattered (scattering angles 4–13°) and backscattered radiation. The second major sensor contained in the CAPS probe is based on the earlier PMS Two-Dimensional Optical Imaging Probe (2D-OAP). This sensor uses a 64 element linear diode array to record a 1-bit profile of the shadow produced by particles as they pass through a laser beam. The probe also includes a heated wire liquid water content sensor (the LWCD) and a Pitot tube to provide accurate measurements of the air speed at the probe. The CAPS can make measurements of particles between 0.35 μm and 1.55 mm in diameter at particle densities of up to approximately $13,000$ ml^{-1}. The CAS has 40 programmable size bins, which allows the measurement of the complex particle size distributions often found in clouds. The main advantage of the measurement of both forward and backscattered radiation by the CAS is the elimination of the multi-valued size response curves seen in many particle counters, as well as offering the potential for limited refractive index determination.

Other aerosol sensors for use on board aircraft include the Focused Cavity Aerosol Spectrometer II (FCAS II) (Jonsson *et al.*, 1995) and the Nucleation-Mode Aerosol Size Spectrometer (N-MASS), both of which are operated by the Department of Engineering at the University of Denver (Mullen, 2002). The FCAS II and its predecessor, the FCAS, are particle counters which measure particles in the 70 μm to 1 mm size range. The FCAS systems are based around

a helium-neon laser and photo diode detectors, and use focusing mirrors to collect scattered light over a large solid angle. The N-MASS is a much more complicated instrument which uses 5 parallel condensation nucleus counters to measure particles in the 4 to 60 nm size range. The condensation counters are calibrated to detect particles above different lower size limits, thereby giving some information on the size distribution. These instruments, as well as the MASP, particle impactors and airborne and ground based lidars have recently been used in a large study to validate the SAGE III satellite instrument: the SAGE III Ozone Loss and Validation Experiment (SOLVE). The SOLVE and SOLVE II web sites (NASA, 2000, 2002) provide further information on these missions.

As well as instruments that measure aerosol properties directly, there are also a number that make *in situ* measurements from which aerosol properties can be derived. These instruments are usually spectrometers of some kind, designed to measure the concentration of trace gases such as O_3, nitrogen and chlorine compounds. Instruments of this type usually use the Sun, Moon or stars as their light source and hence have very large sample volumes (i.e. they measure the total extinction along the line of sight to the Sun). In this sense they are remote instruments, even though they make their measurements *in situ*. AMON (Absorption par Minoritaires Ozone et NOx) is one such instrument (Naudet *et al.*, 1994). This instrument is a grating spectrometer with a CCD detector which observes stars through a 20 cm telescope, allowing the retrieval of the concentration of various gas species and aerosol optical depth.

A German instrument, MIPAS-B (Michelson Interferometer for Passive Atmospheric Sounding) (Fischer & Oelhaf, 1996; Stowasser *et al.*, 1999) is a balloon-borne limb emission Fourier transform spectrometer. This instrument was a development stage in the MIPAS satellite instrument and is capable of making concentration measurements of a very wide range of gas species as well as the aerosol optical depth. Recent results from MIPAS-B have shown a strong signal from scattered tropospheric radiation in measured radiances of PSC clouds (Höpfner *et al.*, 2002). This has implications for retrievals of PSC properties from limb viewing satellite instruments which operate in the infrared. A French/German balloon-borne spectrometer is described in Jarisch *et al.* (1997), this one being a limb sounding cryogenically cooled IR spectrometer.

This instrument was the precursor to the Cryogenic Infrared Spectrometers and Telescopes for the Atmosphere (CRISTA) instrument which is to be used on the Space Shuttle (Offermann *et al.*, 1999).

The Jet Propulsion Laboratory in California has also produced a series of balloon-borne spectrometers, the latest being the MkIV Interferometer (Toon, 1991; Sen *et al.*, 1998; Toon *et al.*, 1999). These instruments are solar absorption Fourier transform infrared (FTIR) spectrometers, and as with the European instruments are capable of measuring the concentration of a wide range of gases simultaneously.

2.3.3 Laboratory measurements of particle scattering

Although this work aims to produce an instrument suitable for in situ measurement of atmospheric aerosol, the measurement technique employed is in many ways more closely related to some laboratory based instruments than any previous *in situ* instrument. A historical overview of light-scattering aerosol instruments is given by Kerker (1997).

The detailed measurement of aerosol scattering patterns has traditionally been performed using particle levitation instruments (Blau *et al.*, 1970; Phillips & Wyatt, 1972; Pluchino *et al.*, 1980; Pluchino, 1987). These instruments use electrostatic forces to suspend a single aerosol in a well defined sample volume. The particle can then be imaged and its scattering pattern measured by a movable detector. This technique provides the most detailed and complete characterization of a free particle's scattering properties possible. Uses for these instruments include testing the theoretical descriptions of light scattering, especially for non-spherical particles, and measuring phase changes of particles.

The improvement in detector technology, particularly the advent of high-speed CCD cameras, has led to the development of instruments which image the scattering patterns produced by particles passing though an illuminated sample volume. These instruments are particularly useful in making measurements of non-spherical particles, or particles with inclusions (Secker *et al.*, 2000). Kaye *et al.* (1997) present a method of distinguishing crocidolite and chrysotile asbestos fibres from each other and background particles in real time. The

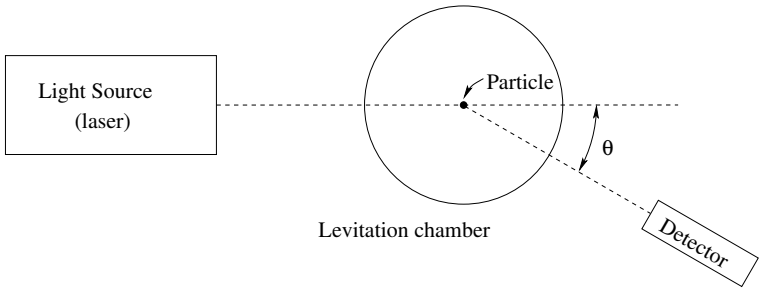


Figure 2.7: A schematic of a simple particle levitation instrument. The detector is free to move around the suspended particle at a fixed distance to record the scattering as a function of angle.

system uses an imaging instrument and a neural network to 'recognize' the distinctive scattering pattern produced by the asbestos fibres.

The final category of instrument used to make angularly resolved measurements of scattered light consists of those which use an array of individual detectors to measure the scattering pattern at fixed angles. Two such instruments are the Dual Amplitude Weighted Nephelometer A (DAWN-A) and the Fibre Aerosol Analyzer (FAA). Both instruments use optical fibre feed photomultiplier tubes to simultaneously measure the scattered intensity at different angles. DAWN-A is designed to make measurements of general ambient aerosol and is more fully described in Chapter (5). The FAA is designed to distinguish between cylindrical fibres and other aerosol particles. It is capable of measuring spherical particles with diameters ranging from a few tenths of a microns up to a few microns, and fibres with diameters in the same range and lengths from a few microns to several tens of microns. The measurement also allows for the determination of the length and diameter of the fibres (Barthel *et al.*, 1998). The eight detectors are located at a constant scattering angle and cover an azimuth angle range of approximately 20° (see Chapter (3) for the definition of these terms).

2.3.4 Remote measurements of atmospheric aerosol

So far, the emphasis of this review has been on *in situ* and direct measurements of the properties of atmospheric aerosol particles – this section discusses remote

measurement techniques. The two major classes of instruments used to make remote measurements of the stratosphere are lidars and satellite instruments.

Lidar

Lidar (an acronym for Light Detection and Ranging) is the optical equivalent of pulsed radar. Lidars work by emitting short pulses of light, usually from a laser, and detecting the backscattered signal. As with radar the time delay between the transmission of the pulse and the detection of the backscattered signal gives the range to the target. When used to probe the atmosphere each lidar pulse returns a continuous profile of backscattered radiation as the photons interact with air molecules and particles along the beam path. This signal is discretized into a series of range bins, often by use of a mechanically shuttered detector. The resulting measurement is a series of intensities corresponding to the backscatter from the atmosphere at increasing ranges. Clark & Hester (1995) and Hornak (2002) provide an overview of lidar techniques and technology.

The exact method used in measuring the return signal varies with the application of the lidar. The simplest measurement is of elastic scatter off air molecules and aerosol particles. Such a measurement can be related to the sum of the Rayleigh (from air molecules) scattering and aerosol (Mie) extinction cross sections. By using assumptions about the extinction and scattering cross sections it is possible to separate the aerosol and molecular components. Once this is done the measurement yields an aerosol optical depth profile and an atmospheric density profile, which can then be used to determine atmospheric temperatures, using the ideal gas law and the hydrostatic equation. Because of the relatively high signal strength this method is often used for measurements of the middle atmosphere.

A variation on this technique is High Spectral Resolution Lidar (HSRL). These systems use two measurement systems, the first of which is the standard arrangement described above. The second uses a very narrow band absorbing filter at the laser wavelength and measures only those backscattered photons influenced by thermal Doppler broadening. As the much larger mass of aerosol particles mean they have a much lower thermal velocity than the surrounding air molecules, this signal is due to molecular scattering only. HSRLs thus make

separate measurements of the molecular and aerosol extinction and scattering coefficients and so remove the need for the assumptions used in single channel Rayleigh lidar. Holz (2002) suggests that the application of these assumptions can result in large errors in the retrieved parameters.

Many lidars also use measurements of the polarization of the returned signal to determine sphericity of aerosol particles. Lasers produce linearly polarized radiation, and the use of polarization filters on the detectors enables the determination of the depolarization ratio, defined as

$$\delta = \frac{I_\perp}{I_\parallel} \tag{2.23}$$

where $I_\perp$ and $I_\parallel$ are the backscattered intensities perpendicular and parallel to the original polarization. Spherical scattering targets will not produce any rotation of the polarization and hence will result in $\delta = 0$. Scattering off elongated particles rotates the polarization plane however. The degree of this rotation depends on the shape of the particle and its orientation with respect to the incoming polarization. Thus, scattering from a distribution of non-spherical aerosol particles will produce a non-zero depolarization ratio, $0 < \delta < 1$. This technique is often used in determining the phase of upper tropospheric clouds (cirrus) and polar stratospheric clouds.

Other techniques employed in lidar measurements include employing a Fabry-Perot interferometer on the returned signal to enable the measurement of a Doppler shift related to the radial speed of the target. An example of this type of instrument is the High Resolution Doppler Lidar (HRDL) operated by the National Oceanic and Atmospheric Administration Environmental Technology Laboratory (NOAA ETL) (Wulfmeyer *et al.*, 1998; Brewer, 2002). Differential Absorption Lidars (DIAL) use two or more laser wavelengths to measure concentration of different species from the relative absorption of the different wavelengths. This technique is commonly used to measure stratospheric ozone concentrations.

As well as using the elastic scattering of atmospheric constituents, lidar techniques can also make use of inelastic Raman scattering. Excitation of the vibrational and rotational modes of various molecules allows a multitude of

measurements to be made, such as nitrogen densities (leading to temperature profiles), water vapour content and ozone concentration (Philbrick & Mulik, 2000). Because the return signals are at specific wavelengths away from the emission wavelengths of the lasers used, these instruments offer measurements which are essentially unaffected by aerosol scattering. Thus they are very useful for measuring the concentration of trace species (particularly water vapour) in the aerosol-laden lower troposphere. The low signal strength of Raman scattering limits the range of these instruments to a few kilometres however.

The advent of compact, high efficiency lasers, as well as the improvements in electronic and computer technology, has enabled the development of small robust lidars suitable for operation in remote areas or on-board aircraft or spacecraft. Examples of portable lidars suitable for aerosol measurements are the Depolarization and Backscatter Unattended Lidar (DABUL), the TEA CO2 Doppler Lidar, the Airborne Excimer Ozone DIAL and the Airborne Aerosol Lidar (ABAEL); all of which are operated by NOAA (Brewer, 2002). The first space based lidar system was the Lidar In-space Technology Experiment (LITE). This system was a technology development and validation experiment which flew on the space shuttle mission STS-64 in 1994 (Winker & Powell, 1998). The Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALISTO) satellite is a follow up to the LITE instrument. It is to be launched in 2004 and will investigate the role that clouds and aerosol play in the Earth's radiation budget (Armistead & Cannon, 2003).

Lidars are a powerful tool for aerosol measurement. Their ability to provide a high resolution aerosol extinction profile as well as a measurement of the phase of the particles is unique. Lidar measurements are particularly powerful when made in conjunction with OPC measurements (Adriani *et al.*, 1992; Deshler *et al.*, 1992b; Jäger *et al.*, 1995). By comparing lidar extinction profiles and particle counter size/number density measurements a refractive index can be inferred for the particles. Also, measurements of PSC phase by lidar can be related to particle size estimates from the *in situ* instrument, providing insight into PSC formation. Lidars also lend themselves equally to the measurement of long term stratospheric aerosol trends (Barnes & Hofmann, 2001), as well as the rapid evolution of aerosol properties which can occur in the troposphere.

Satellite instruments

In the past three decades there have been many instruments launched on satellites designed to measure stratospheric composition. Most of these instruments are primarily designed for measuring trace gas concentrations. However, since they all use radiometric methods to make their measurements, the contribution of aerosol to the signal must be taken into account, and results in a measurement of aerosol extinction.

The first space based instrument capable of making measurements of stratospheric aerosol was a hand-held photometer used by the astronauts on the Apollo-Soyuz space mission in 1975. The Atmospheric Trace Molecule Spectroscopy (ATMOS) project followed on from this and led to the development of a Fourier transform infrared spectrometer for use on board the Space Shuttle. The instrument has been used on four separate missions: Space Lab 3 and ATLAS 1 – 3 (Gunson *et al.*, 1996). The spectral resolution of this instrument is higher than can be reasonably be achieved with autonomous satellite based instruments but, due to the prohibitive cost of Shuttle missions, its use is limited to short term measurement campaigns. It has been proposed to place a version of the ATMOS instrument on board the International Space Station, enabling much more continuous measurement, but as yet no firm plans are in place.

By the early eighties the automated Stratospheric Aerosol Measurement instrument (SAM) II (McCormick *et al.*, 1979, 1981a,b) and Stratospheric Aerosol and Gas Experiment (SAGE) (McCormick *et al.*, 1982) were providing global measurements. Both of these instruments made measurements using the solar occultation technique (i.e. they observed the extinction of sunlight as it passed behind the limb of the Earth), SAGE operating over the tropics and mid latitudes and SAM over the poles.

The longest running stratospheric aerosol satellite instrument is SAGE II. This instrument was a replacement for SAGE and was launched in 1984 on board the Earth Radiation Budget Satellite (ERBS). SAGE II makes measurements in seven bands centred on wavelengths of 1020, 940, 600, 525, 453, 448 and 385 nm. In 2001 a 3rd generation instrument was launched (SAGE III) on board a Meteor-3M spacecraft. This instrument uses a CCD based spectrometer for continuous spectral coverage between 290 and 1030 nm, as well as a separate

aerosol channel at 1550 nm. Improvements in sensor technology allows SAGE III to make measurements from lunar as well as solar occultation. Although the SAGE instruments are primarily for stratospheric measurement their measurements can extend well into the troposphere when aerosol loading is low and the light path is free of cloud.

In September 1991 the Upper Atmosphere Research Satellite (UARS) was launched carrying 10 instruments and enabled the measurement of chemical species, aerosols, winds and temperature. The satellite carries three instruments for aerosol measurement, the Cryogenic Limb Array Etalon Spectrometer (CLAES), the Halogen Occultation Experiment (HALOE)and the Improved Stratospheric And Mesospheric Sounder (ISAMS). CLAES and ISAMS both measured infrared emission from the Earth's limb and were only operational for a relatively short period. CLAES was shut down at the beginning of May 1993 when it's supply of coolant ran out. ISAMS ceased operation in July 1992 because of a mechanical failure. HALOE is a solar occultation instrument similar to the SAGE instruments and is still operational.

Another series of solar photometers specifically designed for aerosol observation are the Polar Ozone and Aerosol Measurement (POAM) instruments. The first operational version was POAM II which was launched in 1993 on board the French SPOT 3 satellite. It made extinction measurements in nine narrow band channels between 350 and 1060 nm in the latitude ranges 55 to 71° North and 63 to 88° South. Due to the failure of the SPOT 3 satellite in 1996 a replacement, SPOT 4, was launched in 1998. This satellite, which is in an almost identical orbit to SPOT 3, carries the POAM III instrument. POAM III is very similar to it's predecessor, although the nine wavelength bands have slightly altered, and remains operational. Both instruments provide measurements of NO_2, ozone, oxygen and water absorption, as well as aerosol extinction.

The Improved Limb Atmospheric Spectrometer (ILAS) was a short-lived instrument (August 1996 - June 1997) on board the ADEOS (Advanced Earth Observing Satellite, latter renamed "MIDORI"). This instrument made spectral measurements of solar absorption at the Earth's limb. A new version of this instrument (ILAS II) has been launched on the ADEOS-II spacecraft. It consists of four separate spectrometers covering different wavelength ranges in the

infrared.

A major new mission for Earth observation is ENVISAT which was launched by the European Space Agency (ESA) in early 2002. This craft carries a variety of instruments capable of making aerosol measurements. The Global Ozone Monitoring by Occultation of Stars (GOMOS) experiment is made up of one UV and two infrared spectrometers and makes occultation measurements using approximately 25 bright stars. The Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) is a spectrometer which makes measurements in the UV, visible and near infrared. It can be used in three modes: nadir, limb viewing and solar/lunar occultation. Measurements include tropospheric and stratospheric trace gas concentrations, aerosol extinction, cloud top height, fractional cloud cover and albedo.

The Medium Resolution Imaging Spectrometer (MERIS) is another spectrometer on board ENVISAT. It's primary aim is to make spectrally resolved images of the oceans, but it also provides information on clouds, water vapour, aerosol and land vegetation. ENVISAT also carries the satellite version of the MIPAS sensor (of which MIPAS-B was a prototype). The satellite version is a limb viewing spectrometer measuring wavelengths between $4.15 - 14.6$ μm. The instrument produces vertical profiles of many atmospheric trace gases between 5 and 100 km in altitude.

Future instruments for atmospheric aerosol measurement from space include the High Resolution Dynamics Limb Sounder (HIRDLS) to be launched on the NASA Aura mission. This instrument is a limb-scanning radiometer which will provide trace gas and aerosol concentrations, cloud top heights and the location of polar stratospheric clouds. Details of the design and operation of these instruments can be found on the world wide web, at the sites given in Table (2.1). Information on many instruments can also be found at the German Aerospace Centre web site (DLR, 2002).

Instrument or satellite name	World wide web URL
ATMOS	http://remus.jpl.nasa.gov/atmos/
CLAES	http://badc.nerc.ac.uk/data/claes/
HALOE	http://haloedata.larc.nasa.gov/home.html
HIRDLS	http://www.eos.ucar.edu/hirdls/home.html
ILAS	http://www-ilas.nies.go.jp/
ILAS II	http://www-ilas2.nies.go.jp/
ISAMS	http://www.atm.ox.ac.uk/group/isams/
SAGE	http://www-sage1.larc.nasa.gov/
SAGE II	http://www-sage2.larc.nasa.gov/
SAGE III	http://www-sage3.larc.nasa.gov/
POAM II/III	http://wvms.nrl.navy.mil/POAM/
MIPAS	http://www-imk.fzk.de:8080/imk2/imk2-e.html
	http://envisat.esa.int/instruments/mipas/
GOMOS	http://envisat.esa.int/instruments/gomos/
MERIS	http://envisat.esa.int/instruments/meris/
SCIAMACHY	http://envisat.esa.int/instruments/sciamachy/
ENVISAT	http://envisat.esa.int/
UARS	http://umpgal.gsfc.nasa.gov/uars-science.html

Table 2.1: The world wide web addresses for satellite instruments making observations of stratospheric aerosol.

Chapter 3

Instrument design

The ultimate goal of this study has been to design and build a prototype instrument which would enable the measurement of the number density, size distribution and composition of aerosol particles of the type found in the stratosphere. The instrument was also required to be suitable for adaptation to use as part of a self-contained, *in situ*, balloon borne package capable of operating in the stratosphere.

Given the nature of atmospheric aerosol and the requirement of balloon borne operation one can list the specific requirements of the instrument:

- The instrument must be light and compact.
- The various systems on board must be simple and robust enough to withstand both the vast range of ambient conditions experienced in travelling from the ground to the stratosphere (i.e. an ambient pressure change from $\sim 10^3$ to ~ 10 hPa and a temperature change from ~ 300 to ~ 200 K), as well as the mechanical stresses likely to be experienced during launch and landing.
- The ability to measure predominantly liquid aerosols across a broad range of particle sizes and concentrations. In particular, the instrument should be able to measure the low concentration (< 10 cm^{-3}) of small particles (< 1 μm) found in the stratosphere.

These requirements determine the measurement technique selected and the detailed design of the instrument. This chapter describes why the SPARCLE instrument has been designed as an optical particle counter, as well as the

detailed design of the different systems which make up the instrument.

3.1 Aerosol measurement techniques

This section gives a description of the major types of instrument used for *in situ* measurement of aerosol particles and discusses their benefits and disadvantages, with emphasis on their use in our application. For a full and detailed description of all aspects of *in situ* aerosol measurement the reader is referred to Willeke & Baron (1993).

Probably the most common method employed for *in situ* aerosol measurement is to draw air through a filter, or over some form of impactor, which will trap the aerosol particles. The filter can then be retrieved and analysed. Examples of the use of such instruments in the measurement of stratospheric aerosol are the Casella impactors and the aircraft direct flow collector probe used by Junge *et al.* (1961) and Junge & Manson (1961). The most common use of impactor instruments for stratospheric aerosol today is as part of aircraft borne instrument suites. The advantage of this technique is that the analysis is carried out in a laboratory and can therefore make use of much more precise measurement techniques than would be possible or practical for a small automated measurement device. This is also a disadvantage however, since it means that there is a substantial delay between the collection of the sample and its analysis. There is therefore a great danger that the sample may become contaminated or undergo some form of chemical or physical change in transit. This problem is compounded by the fact that collecting the aerosol sample brings the particles in contact with some solid surface. It is also likely that the filters will experience conditions quite different from those found at the collection point while in transit. Collection in this way is therefore best suited to situations where the particles are solid and fairly inert, or where information on the size and number of particles is either not required or can be determined in some other way. Another consideration is that because the filter must be exposed to the aerosol laden air for a finite time this method has a very limited time resolution – it is difficult to tell when a particle hit the filter. These factors combine to make this approach completely unsuitable for the instrument required here – we are interested in measuring mostly liquid aerosol, are interested in both the size

and number density of the particles, and it would be very difficult to protect a sample collected in the stratosphere from adverse conditions on it's return to the surface. We are also interested in the vertical distribution of particles, determined by variation with time as the balloon ascends into the stratosphere.

Another widely-used type of aerosol instrument is the differential mobility analyser. Particles suspended in the air flow are deflected towards a small aperture by a force field, but due to different accelerations induced in different sized particles only those in a well defined mass interval will pass through it. The particles which do pass through the aperture are then counted by another method, usually optical. The field usually used is electric, though the principle of operation would be the same for any field[1]. One complicating factor in the use of an electric field is that the particles must have a known charge introduced to them. Such instruments enable very accurate and precise determination of the mass (and therefore size) distributions of the particles. They also have the advantage of not actually exposing the particles to any physical contact and therefore are suitable for measuring volatile or reactive particles. Unfortunately they require extremely stable and adjustable high voltage supplies as well as another type of instrument which can count the particles. This means that they are generally large, delicate and expensive pieces of equipment suitable only for laboratory use. They are also not very sensitive to particle composition, unless the device counting the size selected particles has this capability, as is the case with the DAWN-A instruments (see Chapter (5)).

The condensation nuclei (CN) counter is another instrument which pre-prepares particles before passing them on to a counting instrument. The work of Junge *et al.* (1961) made use of a CN counter and the Wyoming dust sonde record (see Section (2.3.1)) includes CN measurements. In this case the goal is to detect all particles larger than molecular size. This is done by passing the air through a region containing some supersaturated vapour. Because of this supersaturation, droplets will condense on any suitable surface and hence any particle within the sample airstream will grow, enabling even the smallest particles to be easily counted. Of course this sort of instrument only allows particles to be counted – practically all size and composition information is lost in the condensation process.

[1]Namely gravitational or magnetic.

One useful and common way to determine the properties of aerosol particles is to measure their effect on some form of radiation. Whether this radiation be a stream of particles like an electron beam or some form of electromagnetic radiation the principles are basically the same, and such devices will be referred to as *optical instruments* in this text. The four main types of optical instrument are particle imagers, extinction photometers, scattering photometers and optical particle counters. Particle imagers, as implied by the name, image the aerosol particles as they pass through the instrument. This type of instrument allows the determination of the particle size, phase and shape, as well as giving a good measure of particle concentration. The particles do need to be reasonably large to be imaged however, and the quantity of data one can expect from such a device makes it unsuitable for a remote instrument.

Both extinction and scattering photometers use the interaction of aerosol particles with a beam of light, as they pass through a well defined sample volume, to determine their properties. An extinction photometer measures the change in the apparent intensity of the light source itself due to particles passing between the light source and the detector. The extinction is due to both scattering and absorption by the particles and expressions exist relating the extinction to particle size distribution and mass concentration. The main disadvantage of this form of photometer is their poor sensitivity - the concentration of particles needs to be fairly high for them to have much effect on the intensity of the light. Scattering photometers get around this problem by detecting the light scattered out of the illuminating beam by the particles. The problem here is that the relationship between the signal and the particle properties becomes very complicated and such instruments must be very carefully calibrated with aerosols similar to those being measured. The Polar Nephelometer is a good example of a scattering photometer (see Section (2.3.2)).

An optical particle counter is a variation on the scattering photometer concept. Here the sampling volume (the volume where the sample air flow and light beam cross) and sample air flow rate are designed so that there should never be more than one particle present in the sample volume. This means that scattering occurs from only one aerosol particle at a time, which hugely simplifies the relationship between the light detected and the particle properties.

Obviously these instruments will provide erroneous results if more than one particle does enter the sample volume at once, and they will also give a smaller signal than a scattering photometer which detects the combined scattering of many particles. The Wyoming OPC (Section (2.3.1)) belongs to this category of instruments.

3.1.1 Selection of the instrument type

Given the above techniques, it is relatively easy to reduce the viable options for a instrument for use as part of a balloon payload to either a scattering photometer or an optical particle counter. Both these systems are relatively simple and robust in nature, and can be made small and light. They are also both suitable for measuring volatile particles, such as are likely to be encountered in the free atmosphere, as neither requires the particles to come into physical contact with the instrument. An extinction photometer might also appear appealing, but due to the generally low concentration of aerosol in the stratosphere such an instrument would be unlikely to have the required sensitivity.

To choose between these two options one must carefully consider exactly where the instrument will be operating and what information it is required to provide. The ultimate goal of the project is to produce an instrument suitable for stratospheric use, where the aerosol concentration is very low. A scattering photometer will be able to measure over a wide range of concentrations, but in the stratosphere a small instrument is likely to have a few particles in its sample volume at any one time. Hence the signal from a scattering photometer is unlikely to be much stronger than that given by a particle counter. Because we want to measure the refractive index of the aerosol particles, as well as the size distribution and concentration, the retrieval of these quantities will be a complex task. The number concentration of particles can be directly determined in the optical particle counter (it is simply the rate of particle counts divided by the sample air flow rate). The measurement retrieval process for an optical particle counter involves one less dependant variable than the scattering photometer case. We can thus expect the retrieval problem to be more straightforward for an OPC.

On the basis of these decisions it was decided that an optical counter would prove the most suitable type of instrument for our specific application.

3.2 Design details

The three main systems that are required to construct an optical particle counter are:

- A detection system. This includes illumination for the sample volume and detectors to sense the scattered light. The main considerations here are matching the detector and light source performances with the required measurement goals, as well as ensuring that the dominant signal given by the detectors is that due to aerosol scattering (as opposed to Rayleigh scattering from the air and scattering from instrument components).

- A sample air delivery system. This must deliver ambient air to the sample volume at the required rate and then remove it from the instrument. It is important that the sample air reaches the sample volume in a state as close to the ambient conditions as possible. The sample should also take the most direct route possible because the aerosol particles may be altered en route due to inlet and wall deposition as well as evaporation/condensation. It is also important that all of the aerosol particles that enter the sample volume leave the instrument since any free aerosol within the counting chamber could cause false counts.

- Data recording systems. For the instrument to be useful the signals received from the detectors must be stored, along with data pertaining to the instrument operation (light source intensity, air pump speed etc). The prime consideration here is balancing the amount of data required to accurately determine the aerosol properties with the required sampling rate and limits on the amount of information that can be stored or transmitted.

Each of these tasks will be described separately in the following three sections.

3.3 Detection system

In selecting which light detection systems would be used in this instrument the two main goals were – (i) measuring the smallest possible particle; (ii) measuring the shape of the scattering pattern to a high degree of precision. These two

goals are not particularly compatible; a system which can measure the very small change in light level over the very short illumination time of a very small particle is unlikely to be capable of measuring the spatial variation of the scattering pattern of a larger particle, and vice versa. It was thus decided that the instrument would use two separate detection systems, one devoted to each of the measurement goals.

3.3.1 Light source

To make a high resolution measurement of a scattering pattern, a primary condition is that the measurement be made using light of a very narrow wavelength band. As Mie scattering depends on the wavelength of the light being scattered, having a broad range of wavelengths will simply "wash out" the fine detail of the scattering pattern. Also it is highly advantageous for both the measurement of the scattering pattern and the detection of very small particles for the illumination to be very intense, because the greater the intensity the larger the observed signal. These considerations, as well as the requirements that the light source be light and compact and that the light be delivered to the scattering volume as a well defined beam, immediately suggest one light source as being appropriate: a laser diode.

Laser diodes, particularly high power ones, cannot guarantee a single lasing mode, nor do they offer a well-collimated beam. They also require a very stable power supply and temperature conditions for reliable and constant output in both wavelength and power. The compact size, high efficiency (compared to other types of laser), low supply voltage and current requirements, and very high intensity make them ideal for this sort of application, however.

If one wants to make measurements of scattering patterns from small particles, with a view to determining the size and refractive index of those particles it is advantageous to use the shortest wavelength of light that is practical. This is because the scattering pattern depends on the ratio of particle size to light wavelength, expressed in the quantity know as the size parameter:

$$x = \frac{2\pi r}{\lambda}, \tag{3.1}$$

where r is the particle radius and λ is the wavelength[2]. For very small particles

[2]See Appendix(A)

with size parameters of approximately 1 or less the scattering pattern is basically that described by Rayleigh scattering. The pattern is very smooth and there is very little dependence on the refractive index, which makes its determination much more difficult (see Chapter (4)). By reducing the wavelength of the light being used one can reduce the particle size at which the Rayleigh regime takes over and thereby determine the refractive index of smaller particles.

Because of the prohibitive cost of high power laser diodes and the availability of a suitable laboratory laser, the instrument developed here did not use a diode laser. The laser used was a Coherent Inc. argon gas laser, capable of producing 514.5 nm light at powers of up to a few Watts.

To enable the optical power reaching the scattering volume to be monitored the instrument design also included a simple meter for measuring the laser intensity. This power meter consisted of a large area photodiode placed behind a plain mirror (see Figure (3.4)). The mirror was positioned at an angle of 45 degrees and reflected the majority of the incident light into a beam dump. A small percentage of the light would pass through the mirror, giving a signal on the photodiode. This signal was then monitored via a small digital display.

3.3.2 Photomultiplier

For the measurement of very low light levels with a high temporal resolution the photomultiplier tube (PMT) is still the detector of choice. Photomultipliers offer extremely high photosensitivity, providing a signal which is representative of the total irradiance across the photocathode. The PMT chosen for SPARCLE was the Electron Tubes model 9828B. This PMT was selected because of its high sensitivity throughout the visable spectrum, large signal gain and relatively low cost. Details of the PMT electric and mechanical specifications as well as its data collection system can be found in Appendix (C). The PMT is of end window design and has a large 23 mm photocathode. The inner surface of the front window (on which the photocathode material has been directly applied) is moulded to produce an array of corner reflectors. This design greatly reduces the chance that incident light will be lost due to reflection off the photocathode.

Originally the PMT was to be used in conjunction with a light collecting

mirror, as shown in Figure (3.1). By placing the scattering volume at one focus of an elliptical mirror and the PMT at the other, light scattered across a broad solid angle can be focused onto the photocathode. However, in practice it was found that this system was extremely prone to stray light problems and was essentially unworkable. The instrument was therefore redesigned so that the PMT collected light directly. Because of the large size of the photocathode in the 9828B it is possible to collect scattered light over an angular range directly simply by placing the PMT close to the scattering volume, without the use of additional optics. This approach has the advantage of simplicity, however it is possible that it also more prone to stray light problems than more traditional designs. Most optical particle counters use a microscope arrangement to focus light from the scattering volume onto the detector. Because of the limited depth of field of the optics it is possible that stray light which would affect a directly exposed detector will be removed. Despite this the SPARCLE prototype was constructed using the direct method because of its simplicity. As a result of the need to minimize the stray light reaching the photocathode and design restrictions introduced by the PMT housing supplied by Electron Tubes, the closest the photocathode could practically be mounted to the sample volume in the SPARCLE prototype was (21 ± 0.5) mm from the scattering volume. At this distance the photocathode subtends a solid angle of 0.94 steradians. The photocathode is centred 90° in both polar and azimuth angles, i.e. its axis of symmetry is at right angles to both the sample air delivery pipe and the illuminating laser beam.

One of the major drawbacks of most optical particle counters is the response function of the instrument with respect to particle size is not single valued. This results from the complex nature of the scattering pattern for larger particle sizes. For example a certain scattering angle may lie on a peak in the scattering angle for one size, but in a trough for a larger size – in this case the signal for the larger particle would be the lower. This problem can be solved, at least for a specified range of particle sizes and refractive indices, by using white light and/or measuring over a range of angles. The use of white light, as has already been mentioned, 'smears out' the scattering pattern, making it much smoother. By careful selection of the scattering angles to be measured the multivalued nature of the particle size response curve can be greatly reduced. This is the technique, along with measuring over a wide angular range, used by the Wyoming dust

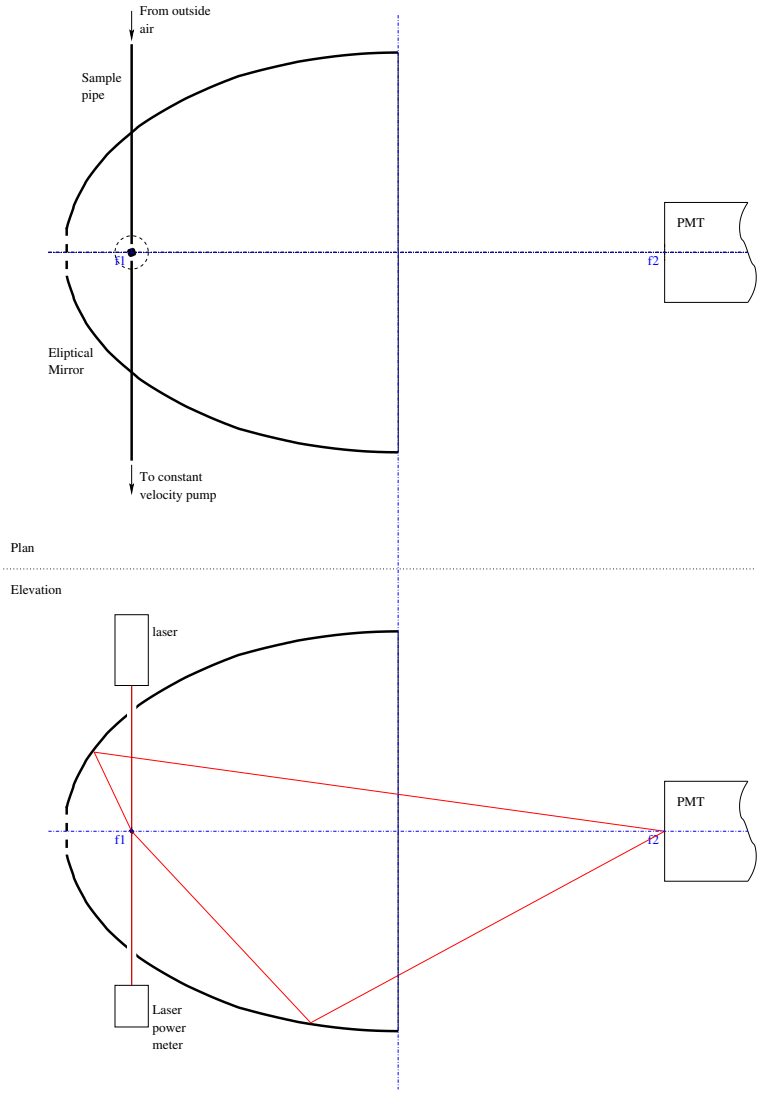


Figure 3.1: A schematic of the original PMT light collection design incorperating an elliptical mirror. The labels f1 and f2 refer to the two foci of the ellipse. The elevation view includes two example light paths from the scattering volume to the PMT photocathode.

sonde (see Section (2.3.1)). This technique can also greatly reduce the signal's dependence on the refractive index of the particle; however this dependence can never truly be accounted for by a single angle measurement.

As a single wavelength light source was required for this instrument we must rely on the smoothing effect of measuring a range of angles to reduce the multivalued nature of the response curve. As can be seen from Figures (3.2) and (3.3) this smoothing effect was found to be minimal for the solid angle subtended by the PMT and at larger particle sizes the response function is highly degenerate. However, at these sizes a signal is also measured by the LDA system and this extra information alleviates this degeneracy. For small particles, where LDA information is unavailable, the PMT response becomes single valued.

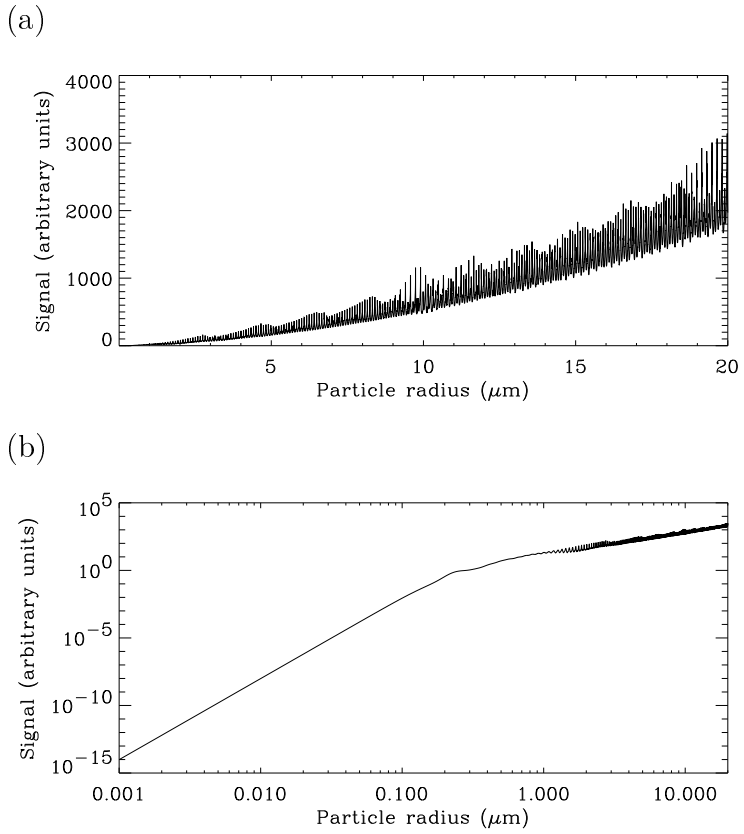


Figure 3.2: The model response curve for the photomultiplier signal with the size of a particle with a refractive index of $(1.4 - 0.001i)$. Both plots show the same data, plot (a) on a linear scale and plot (b) on the log-log scale.

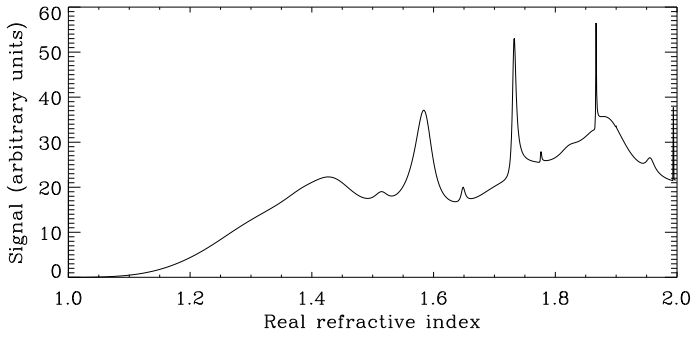


Figure 3.3: The model response curve for the photomultiplier signal with the real part of the refractive index of a 1 μm particle. As in Figure (3.2) the imaginary refractive index is set at $-0.001i$.

As the photomultiplier is to be used to measure a broad angular range of the scattering pattern it is of key importance that any variation of PMT signal with angle of incidence and location of incidence on the photocathode be taken into account. The manufacturer claims[3] that the PMT photocathode response is uniform for angles up to $\sim 60^\circ$ from normal incidence. However a $\sim 5\,\%$ variation in photocathode sensitivity can be expected across the sensitive area.

3.3.3 Photodiode array

The second light detection system employed in the SPARCLE instrument was designed to make measurements of the shape of the scattering pattern required for the retrieval of particle size and refractive index. The primary requirement of this system was that it provide a detailed profile of the scattering pattern across a broad range of scattering angles, with sensitivity and time resolution being secondary concerns.

The selection of an appropriate detector to use for this role presented some difficulty. The performance required is at the limits of sensor technology, due to the low light levels and the very short duration (see Section (3.4.4)) of the signals. If a high resolution ($\gtrsim 100$ pixels) image is to be made the only readily available sensors are CMOS linear sensors (here referred to as linear diode

[3]Garrad, C. (chrisg@electron-tubes.co.uk). *9828 PMT*. Email to Gareth Thomas (g.thomas@phys.canterbury.ac.nz). 28 September 1998.

arrays or LDAs) or linear charged coupled devices (CCDs). These devices are in principle very similar - they both consist of an array of photosensitive elements in which a charge proportional to the number of photons absorbed is collected. This charge is then read off each element in turn, usually using a MOS switch and operational amplifier arrangement. The difference in the two sensors lies in the design of the sensitive elements themselves – in a CCD the charge is collected by a pn junction within a layer of doped silicon, whereas an LDA uses a MOS capacitor to generate the required depletion layer within the silicon.

CCD detectors have lower noise characteristics than LDAs, however the manufacturing process used in making LDAs allows all of the control and amplification electronics to be included on one chip. This means that LDAs are cheaper and have much simpler power supply and driver circuit requirements. Modern linear CCD detectors universally have thousands of pixels in very small packages. This makes them unsuitable for this application due to their small size and the large amount of data produced. For reasonably large sensors with a resolution of a few hundred pixels one is limited to LDAs.

The sensor selected was an S5461-512Q C-MOS Linear Image Sensor manufactured by Hamamatsu. The primary reason for choosing this particular sensor was the fact that it has a large active area (25.6 mm long) while still having relatively few individual elements. This large size provides two main benefits: firstly it means the sensor can be mounted a reasonable distance from the scattering volume and still cover a broad range of scattering angles, secondly the relatively large size of each element means the sensitivity (per element) of the senor is higher than would be the case with smaller pixels. As with the photomultiplier, a detailed description of the LDA system's electronics can be found in Appendix (C).

The sensor is mounted so that the sensitive area was centred at a polar and azimuth angles of 90° and is aligned so that it sampled light scattered at a constant azimuth angle. The sensitive area is placed at a distance of 12.8 mm from the centre of the scattering volume so that it subtends scattering angles from 45° to 135°. This arrangement was chosen because it gave the broadest practical angular coverage of the sensor while keeping the angle of incidence of scattered light on the detector to within 45°. The positioning of the sensor at a

polar angle of 90° was also a matter of practicality – positioning the LDA such that it was not parallel to the laser beam would have presented a great increase in the difficulty of manufacturing the instrument body.

The LDA has 512 equally spaced elements, which gives a resolution of $\sim 0.22°$ at the centre of the detector and $\sim 0.11°$ at the ends. This arrangement allows the fine structure to be resolved in the scattering patterns of particles of up to $\sim$ 50 μm in diameter (assuming 0.66 μm light). The lower detection limit depends mainly on the thermal noise and on the random noise introduced by any background light, both of which in turn depend linearly on the sampling interval used. A major problem presented by the use of an LDA is this sampling interval, which with present technology is limited to a minimum of 5 ms. This is longer than the duration of the signal from a scattering event (see the following section). This not only means we have a noise level higher than would otherwise be the case, but that there is a chance that a single exposure might record the scattering patterns produced by multiple particles.

One possible solution to the problem of the high thermal noise associated with the LDA is to cool the chip. Miniature Peltier coolers specifically designed to perform this task are readily available, but the use of such a device introduces further complications. Firstly the 'hot' side of the cooler must be in good contact with an effective heat sink. Secondly and more critically there is a very high chance of considerable condensation on the LDA itself, possibly leading to electrical short circuits and obstruction of the LDA window. For these reasons it was decided that the LDA would be used uncooled in the prototype. There is no solution to the problem of the possibility of multiple scattering events being captured in a single exposure. The only option is to use the superior time resolution of the PMT to identify affected measurements.

Figure (3.4) shows the layout of the complete detection system. The laser beam enters the instrument through the laser baffle at the top of the instrument and intersects the sample air flow at right angles before passing into a beam dump via a simple optical power meter. The two light detectors are situated opposite each other and are mounted so that their sensitive surfaces are parallel to both the laser beam and the direction of the sample air flow through the scattering chamber.

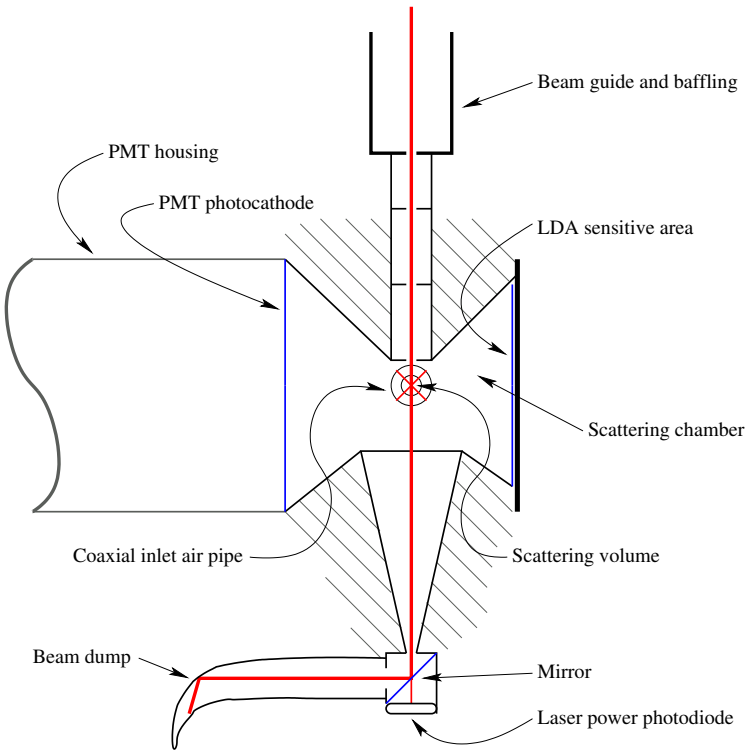


Figure 3.4: Schematic view of SPARCLE's scattering chamber. The view is a cross section taken at the scattering volume, looking down the axis of the inlet pipe.

3.4 Sample air delivery system

In this case the precedent of the Wyoming dustsonde instrument was followed reasonably closely, Figure (3.5) shows a schematic of the system. As can be seen the flow is controlled by two separate pumps, one of which pulls air out of the sample volume and releases it back into the surroundings (pump 2), the other which pushes ambient air through a filter into the sample volume (pump 1). The flow generated by pump 1 will be referred to as the sheath flow, as it enters the sample volume through the outer layer of a coaxial pipe and thus acts as a sheath of clean air around the sample air entering through the inner sample pipe.

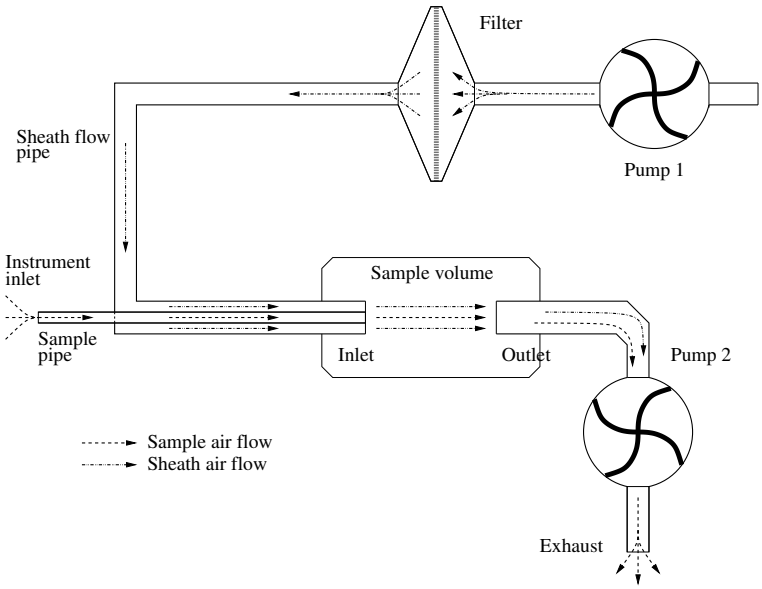


Figure 3.5: A schematic representation of the sample air delivery system used. Here pump 2 controls the sample flow rate and pump 1 the sheath flow rate.

The sample flow is generated by careful balancing of the two pump speeds – if pump 2 is pulling more air out of the sample volume than is being supplied by pump 1 the difference will be made up by air being drawn down the sample pipe. As this pipe is straight, leads directly to the exterior of the instrument, and is insulated by the sheath flow, aerosols contained in the sample air should not experience any drastic change in conditions before reaching the sample volume.

At the sample volume the two flows will form a laminar jet from the inlet pipes to the outlet, with the sample flow contained within the aerosol-free sheath flow. This arrangement ensures that all particles entering the scattering volume exit directly, there being little chance of particles becoming trapped in the sampling chamber and causing false counts.

The filter used to clean the sheath air is GF 10 glass fibre paper made by Schleicher & Schuell. The papers are mounted in a FP 050/1 50 mm diameter plastic housing for inline filtering, also made by Schleicher & Schuell. This filter is specifically designed for the removal of aerosols from air.

3.4.1 Flow measurement

All flows measured during the course of this work have been measured using a Gilian Gilibrator 2 flowmeter manufactured by Sensidyne Inc. The unit was used with the standard flow wet cell (model 800266-1). A schematic of the instrument can be seen in Figure (3.6).

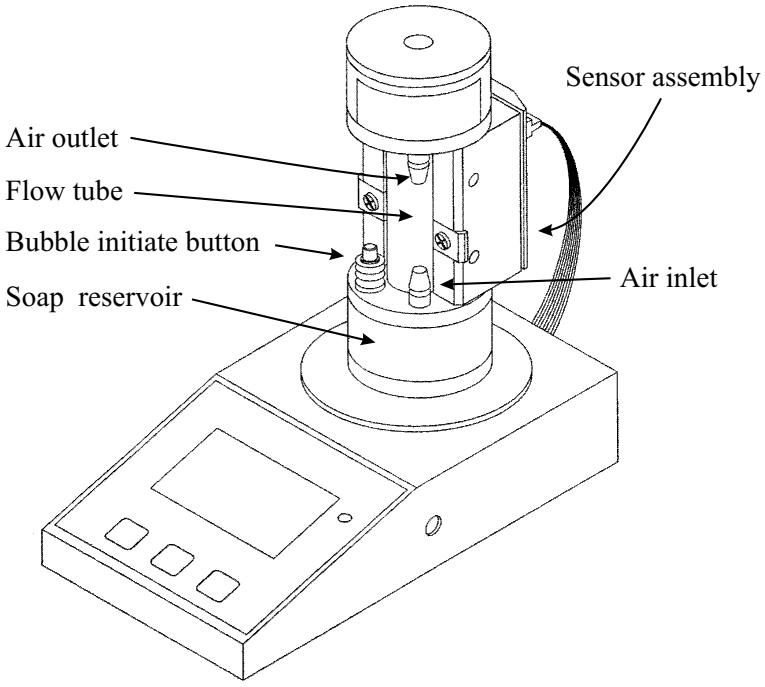


Figure 3.6: Schematic of the Gilibrator 2 flowmeter.

The instrument measures the flow by timing the transit of a soap bubble along the flow tube. The air flow enters through the soap reservoir at the base of the unit and exits at the top, and a bubble is generated by the operator pressing a button and as it is pushed up the flow tube by the air flow its passing is detected by two infrared sensors in the sensor assembly. The measured flow is then displayed on a screen. The unit is sold as a primary standard flowmeter, i.e. the flow measured is absolute and no user calibration is required. The standard flow cell used here has a measurement range of 20 ml/min to 6 l/min.

An instrument operating on a similar principle to the gilibrator would clearly be unsuitable for use in an *in situ* instrument. In this case the most convenient way to monitor the flow rate would be to measure the pump speeds via a tachometer. Once the pumps' performance over the expected range of operating conditions had been established this would provide an accurate and simple measure of the flow rate.

3.4.2 Pumps

The pump utilized must provide a very stable flow at a fixed volume (as opposed to a fixed mass) and be free of vibration (such as that produced by piston pumps for instance). Given the ultimate aim of using this instrument as a balloon payload it is also important that the pumps work over a wide range of ambient pressures and temperatures.

The Wyoming dustsonde makes use of gear pumps of a similar design to those used in car oil pumps. As can be seen in Figure (3.7) gear pumps consist of two intermeshing gears contained in an enclosed, close fitting body. One of the gears is spun via a drive shaft through the pump body and thus drives the other free-spinning gear. Air enters and exits the pump through two holes in either side of the body, aligned with the points where the gears mesh. Air is trapped between the gear teeth and the interior wall of the body and is transported to the outlet hole by the rotation of the gears. When the gears mesh the space between the teeth is filled and the air is forced out of the outlet hole, producing the movement of a fixed volume of air from one side of the pump to the other.

For this project gear pumps of a very similar design to those used in the

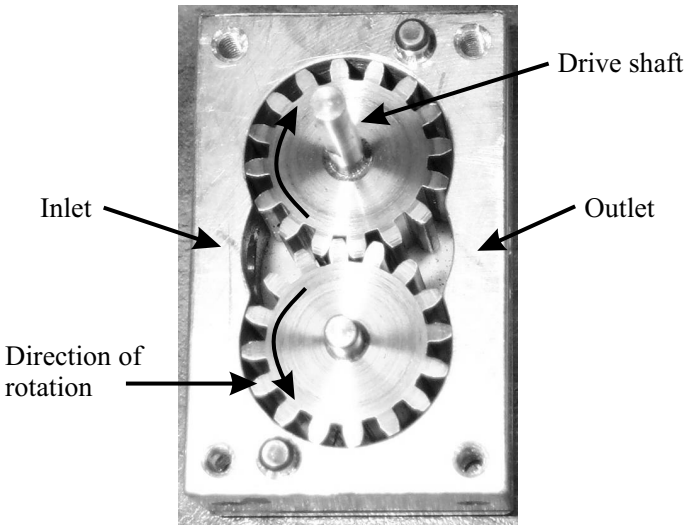


Figure 3.7: An example of the interior of a gear pump.

Wyoming instrument were produced. These pumps were found to have superb linearity over a wide range of flows, the relationship between flow rate and pump speed being linear over the entire range of speeds that the drive motor could supply. Because of their simple design pumps of this nature operate well over a large range of ambient pressures and temperatures, and experience with the Wyoming instrument has shown that they appear to maintain a constant volume flow rate from the ground to stratospheric altitudes.

Buying commercially available constant volume pumps, mostly gear pumps and peristaltic pumps, was considered: however because of the difficultly in finding a pump to match our needs and the high cost it was decided that tailor-made pumps were the best option. Figure (3.9) shows the variation of the flow produced by the two pumps when run at a constant voltage. As can be seen the small pump, after a $\sim$ 30 minute run in, maintains constant flow to within approximately 5 ml/min. The larger pump does not perform as well, with changes of up to 20 ml/min in flow speed. The power supply used for these measurements was the same as that used in the prototype instrument and was found to provide a constant voltage to within 0.01 V.

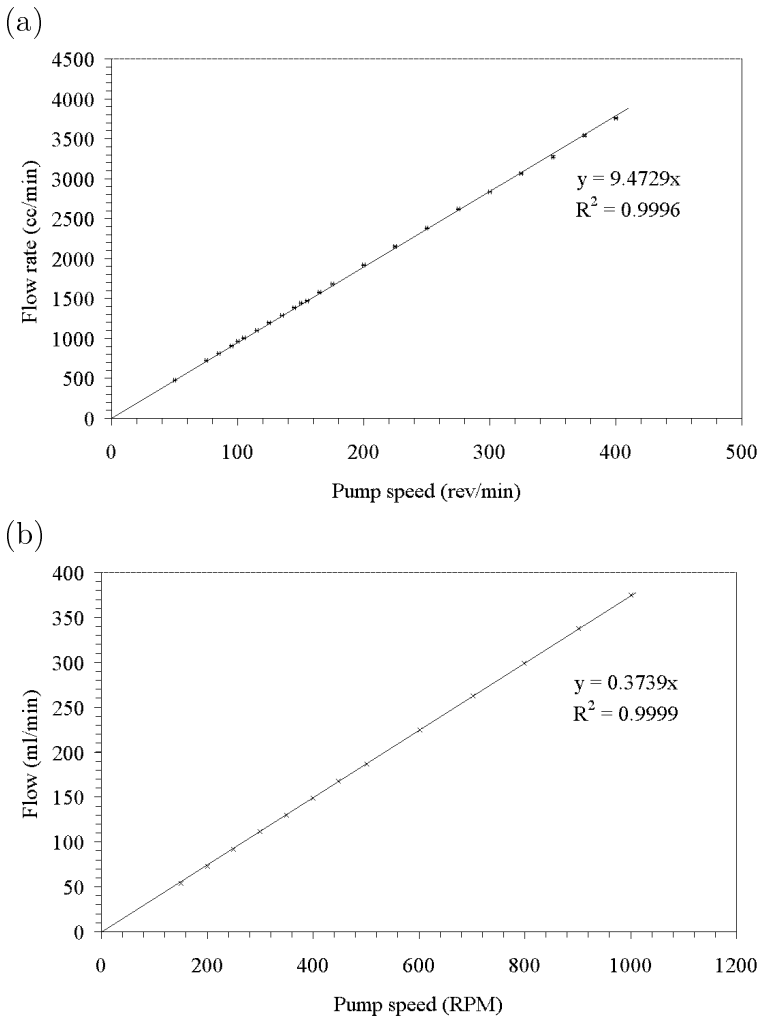


Figure 3.8: The air flow speed generated by the two gear pumps used in the SPARCLE prototype. Plot (a) is for the first pump made and used as pump 1 and plot (b) is for the smaller second pump used as pump 2.

3.4.3 Numerical modelling of air flow

To verify that this system would behave as predicted some simple numerical fluid dynamics modelling was done using the Fluent package. This software, along with the mesh generating software Gambit, allows complex fluid dynamics systems to be modelled in two or three dimensions and as a steady state or with temporal variations. There are several models of viscous turbulence and energy transfer available, and the package can model almost any mixture of fluid and suspended particles imaginable.

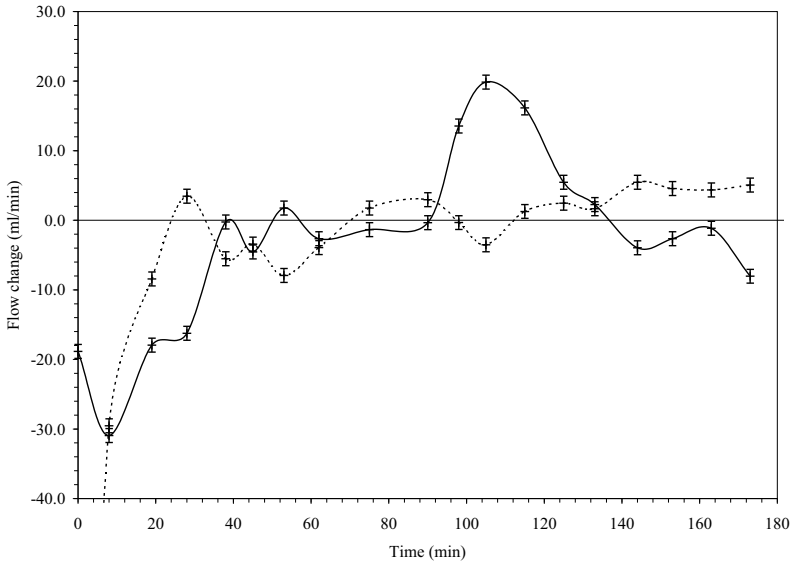


Figure 3.9: The change in the flow generated by both gear pumps over approximately 3 hours, plotted as a difference from the mean flow generated. Pump 1, with a mean flow of 421 ml/min through the sheath flow filter, is plotted as the solid line; pump 2, with a mean flow of 936 ml/min, is plotted as the dotted line.

Fluent is an example of a finite element analysis package – a type of numerical programme which is in widespread use in various fields of engineering for modelling everything from complex mechanical systems to fluid dynamics systems. Finite element analysis works, as the name suggests, by approximating continuous systems with a series of small 'boxes' (or elements) within which conditions are taken to be constant. The state of each element is then calculated, based on the states of the surrounding elements; a process which is repeated iteratively until the system converges to some steady solution.

Commercial packages like Fluent make the modelling of complex systems a reasonably straight forward operation, with intuitive graphical interfaces and extensive manuals and tutorials to guide the user. However the performance of the model is very dependant on how the grid is produced. This task is referred to as mesh generation and is the major task in performing a numerical analysis of this type.

For the purposes of this project a simple two dimensional model of the pipe system at the sampling chamber was generated using the mesh generating program Gambit. This program partially automates the generation of a mesh appropriate for use in Fluent, but a great deal of care must be taken in setting the mesh type and boundary conditions to get a workable result. The model did not include the pumps or the filter – the outlet and sheath flows were simply set to be at a constant velocity in the fluid dynamics package. The sampling chamber itself was represented as a circle of 75mm radius. This geometry was based on an earlier design of the instrument and hence doesn't represent the size or shape of the scattering chamber eventually used. However, the results produced should be applicable to the design used, as there was found to be very little interaction between the sample and sheath flows and the chamber air.

The model used to generate the mesh is shown in Figure (3.10) and indicates how the system was broken up into different zones for mesh generation. Each of the pipes, the area between the pipes, the area around the pipe gap and the remaining interior of the sample chamber were dealt with separately (although not independently as the boundary conditions for each mesh were taken from adjacent zones). Thus there were nine separate meshes generated for this model, the resulting overall mesh can be seen in Figure (3.11).

Fluent is actually a collection of different numerical solvers under one interface. For application to the SPARCLE model the segregated, two dimensional, steady state solver was used with the realizable $\kappa - \epsilon$ viscous turbulence model with the energy (heat) calculations enabled, but the species model (chemistry/mixing) disabled. In this mode the equations for momentum, continuity, energy and two relating to viscosity are solved sequentially (segregated) in a 2 dimensional space. As mentioned, the realizable $\kappa - \epsilon$ viscosity model uses two equations to model turbulence; these two equations calculate the turbulent kinetic energy (κ) and it's dissipation rate (ϵ) – the form of these equations used in the 'realizable' version of this model were derived by Shih *et al.* (1995). This particular turbulence model was used because it is considered highly successful in describing both boundary flows and free jets - both of which are central to the problem at hand. For further explanation of the different solvers and models available in Fluent refer to (Fluent Inc., 2000).

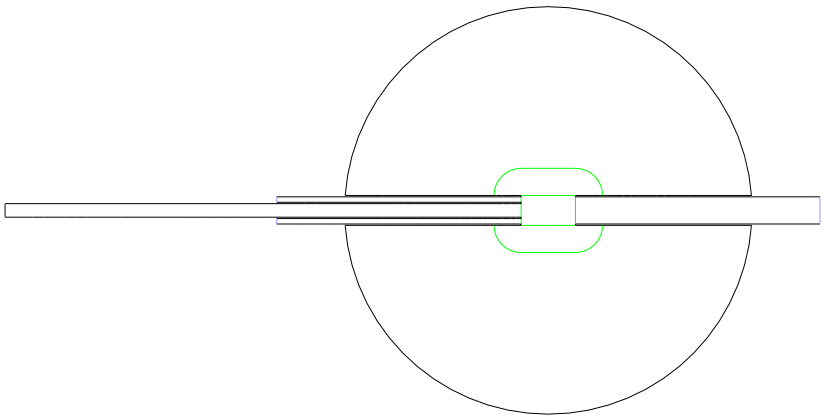


Figure 3.10: The model used for generating the mesh in Gambit.

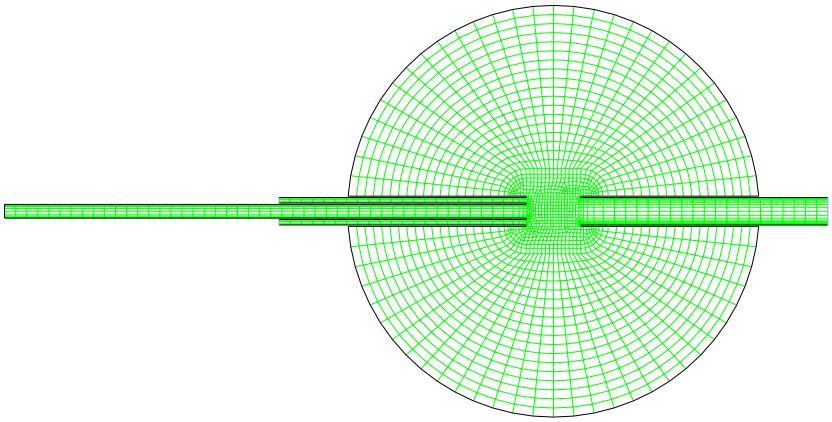
density ρ (kgm^{-3})	dynamic viscosity μ (Nsm^{-2})	specific heat C$_p$ (J/kgK)	thermal conductivity
1.225	1.7894×10^{-5}	1006.43	0.0242

Table 3.1: The physical properties of the standard air model used in the flow simulations.

This solver was then used on the mesh shown in Figure (3.11) with the following boundary conditions: the sheath flow inlet was set as a constant velocity inlet with a 1 ms^{-1} flow in the direction of the pipe, the outlet was set as a velocity outlet with a flow of 0.9 ms^{-1} in the direction of the pipe and the sample inlet was set as an inlet vent. The fluid model used was the standard air model included with Fluent, the properties of which are given in Table (3.1). The steady state solution given by these parameters was then used to refine the Gambit-produced mesh, based on the velocity gradients of the solution. This decreased the mesh spacing in areas where the flow was more complicated, thus allowing for more accurate modelling of potentially turbulent regions. The resulting refined mesh is displayed in Figure (3.12).

With the improved grid the solver was again allowed to converge to a steady

(a)



(b)

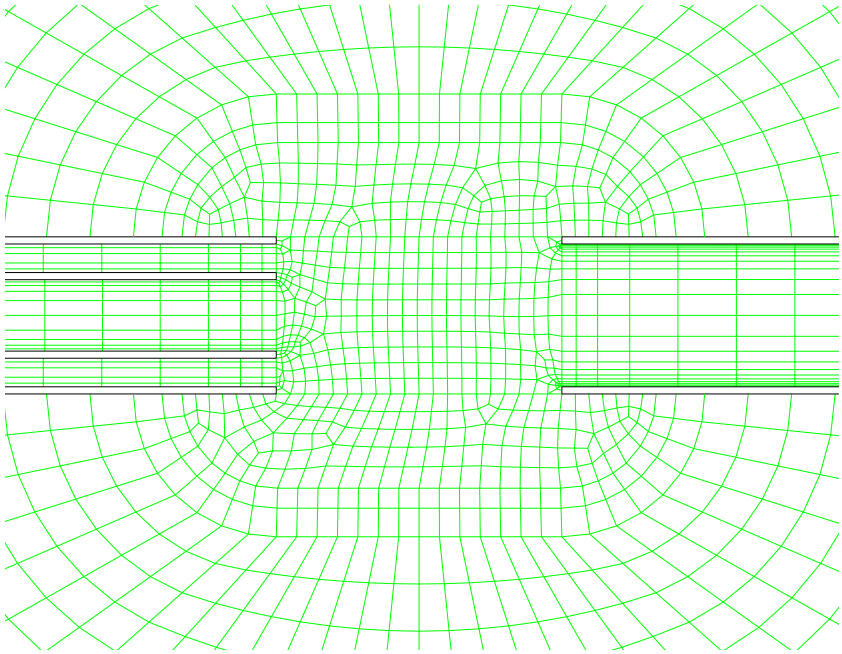
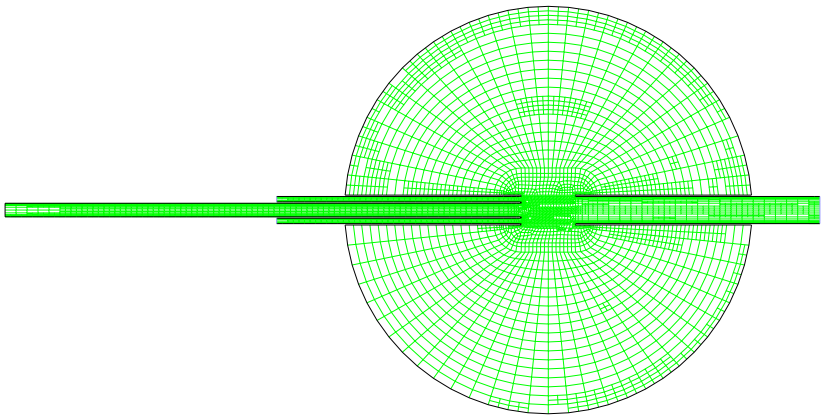


Figure 3.11: The initial grid used in the numerical modelling of the air flow through the instrument. (a) shows the entire mesh, while (b) shows the finer mesh used for the sampling volume area and within the pipes.

(a)



(b)

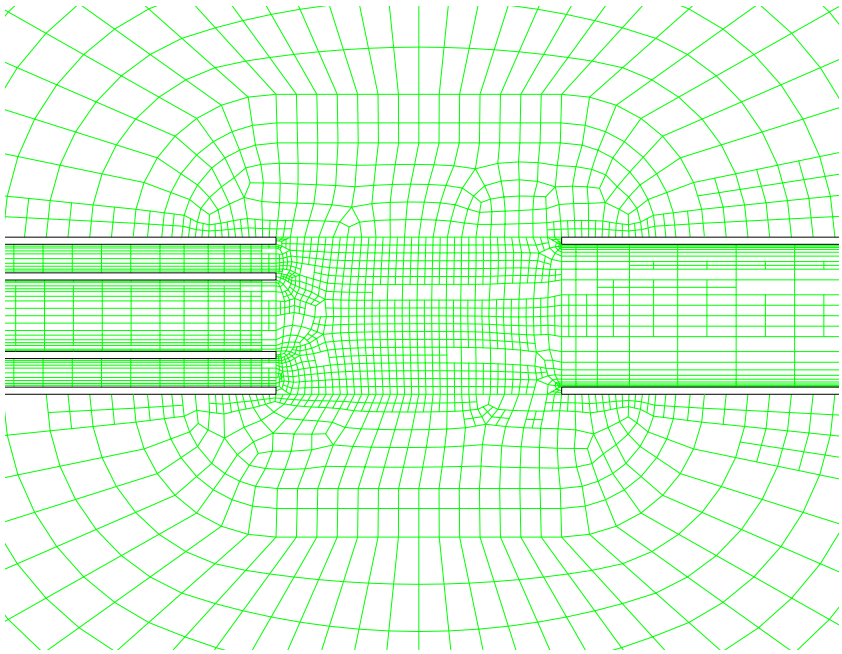


Figure 3.12: The refined grid used in the numerical modelling of the air flow through the instrument. Again (a) shows the entire mesh, while the scattering volume region is detailed in (b).

state solution. Figure (3.13) shows the resulting velocity vectors for each element and the path lines of air molecules entering both the sample and sheath flow inlets. It is clear that the flow does indeed form a laminar jet from the inlet pipes to the outlet pipes, as can be seen by the smooth path lines, and that there is very little interaction between the jet and the air within the sample chamber. The latter point is shown by two features; firstly almost all flow lines that enter through the inlet pipes exit through the outlet pipe, i.e. there is very little air exchanged between the sheath flow and the air within the chamber. Also the very low velocity of the circulation outside the jet indicates that there is very little momentum transfer between the jet and the surrounding air.

Also evident in Figure (3.13) is the effect of the pipe walls on the flow within them. Because friction slows the flow near the walls we can expect particles to undergo changes of velocity of up to ~ 0.4 ms^{-1} at the entry and exit of each pipe. Due to the small size of the particles this acceleration should have no noticeable direct effect on the nature of the aerosol. However such rapid changes in flow speed could result in rapid pressure changes, which could affect the particles due to evaporation or condensation. Figure (3.15) shows the pressure field relative to atmospheric pressure and shows a pressure gradient along the length of the sample tube of approximately 3 Pa. This change can be treated as an adiabatic expansion of the gas, provided that there is no significant heat exchange between the walls of the pipe and the gas. For pure particles (i.e. single component liquids) larger than the mean free path (λ) of the surrounding air the rate of change in particle size is given by the equation:

$$\frac{\mathrm{d}(d_p)}{\mathrm{d}t} = \frac{4D_v M(p - p_d)}{\rho_p d_p RT} \tag{3.2}$$

where d_p is the particle diameter, D_v is the vapour diffusion coefficient, M is the molecular weight of the aerosol material, R is the ideal gas constant, T is the temperature, p is the partial pressure of vapour surrounding the particle and p_d is the partial pressure of vapour at the particles surface. The change in temperature of the air is given by application of the ideal gas law and this can then be used to calculate the appropriate vapour pressure. Corrections can be made for the effects of droplet cooling due to the latent heat of evaporation and complications to diffusive mass transfer within one mean free path length of the droplet surface

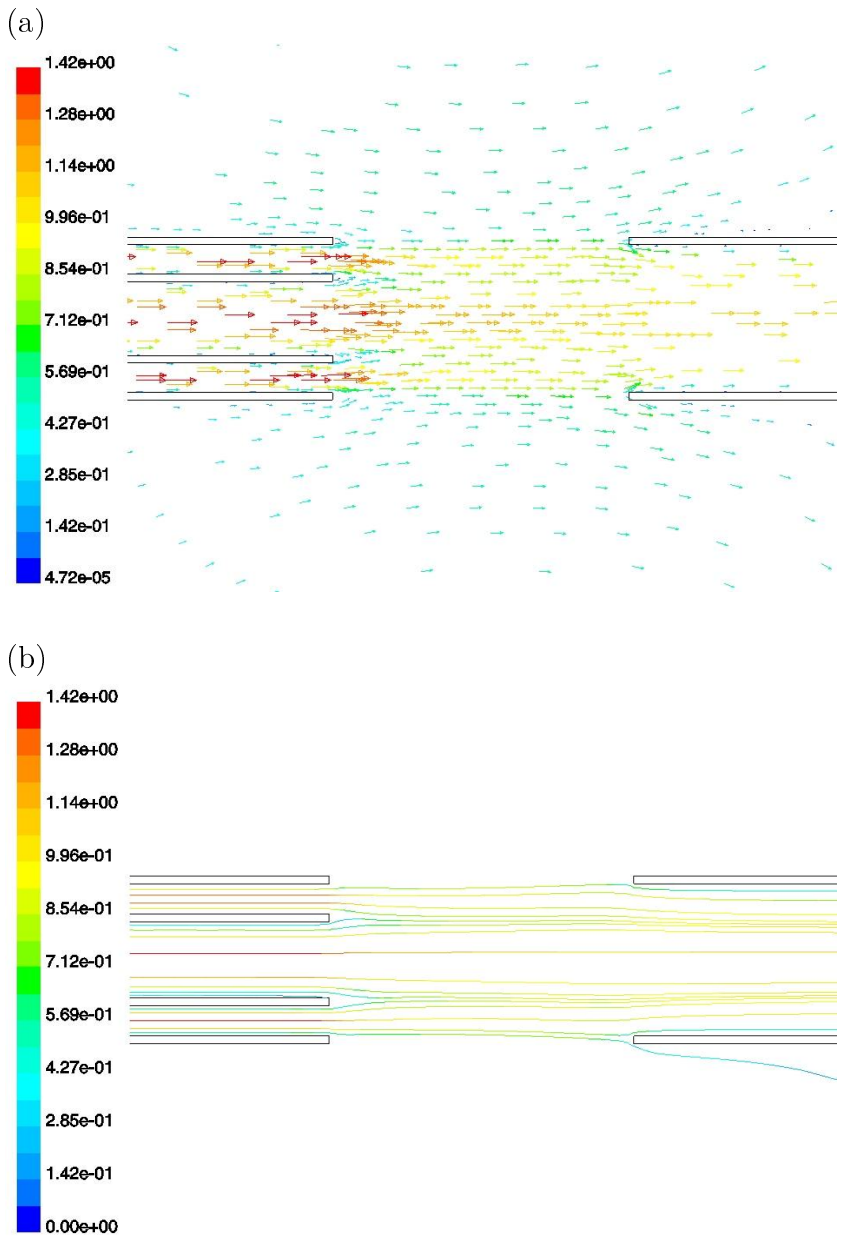


Figure 3.13: (a) Velocity vectors and (b) inlet flow paths for standard conditions. The colour scale indicates velocity magnitude in ms^{-1}.

(the Fuchs correction), resulting in the equation:

$$\frac{\mathrm{d}(d_p)}{\mathrm{d}t} = \frac{4D_v M}{\rho_p d_p R}\left[\frac{p}{T} - \frac{p_d}{T_d}\right]\left[\frac{2\lambda + d_p}{d_p + 5.33(\lambda^2/d_p) + 3.42\lambda}\right]. \tag{3.3}$$

Here T_d refers to the temperature at the particle surface and T is the temperature of the surrounding air. For further information on these equations, as well as expressions giving the vapour pressure of water as a function of temperature see Willeke & Baron (1993).

To determine the change in a particles size as it travels along the inlet pipe this expression needs to be integrated along the path. If we assume the pressure changes linearly along the length of the pipe and that the particle moves along the pipe at a constant velocity this integral can be simply calculated numerically. By breaking the path into a series of small sections, and taking the pressure and temperature to be constant within each interval, an evaporation rate can be calculated. The particle size can then be updated using the flow speed to give the time required to traverse each section.

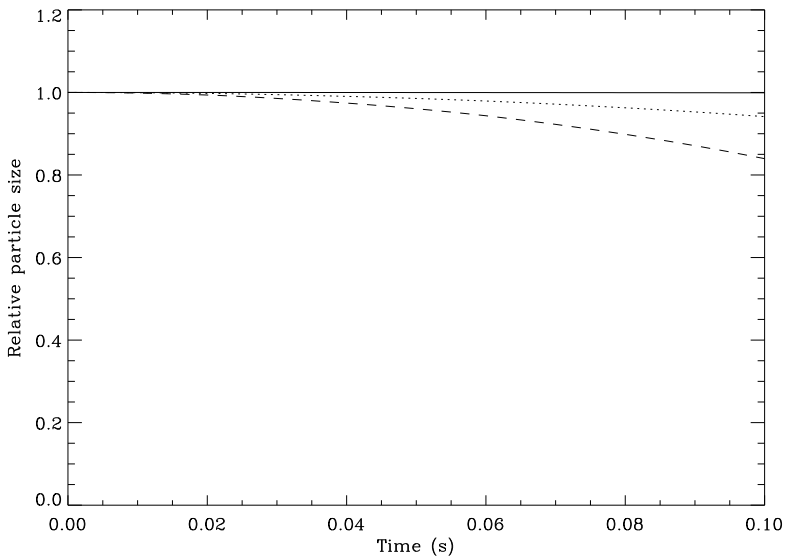


Figure 3.14: The relative change in particle size from entry to the sample tube to arrival at the sample volume for water droplets at standard conditions. The solid curve represents a 1 μm radius particle, the dotted curve is for 0.1 μm and the dashed curve is for 0.05 μm.

Figure (3.14) shows the resulting size change for water particles of three differing sizes entering the sample pipe at standard conditions (temperature = 293 K and pressure = 101 kPa). It is clear that larger particles (radius $\gtrsim 1$ μm) are essentially unchanged by this pressure drop, however for particles of the order 0.1 μm the effect has become significant. In the case of the measurements made in this study (see Chapter (6)) this effect will not be important, as the particles measured are all well above the 1 μm size and glycerol (the aerosol material used) is a less volatile liquid than water. However, more work is needed to characterise this effect across a wide range of ambient conditions and in the case of non-pure aerosol particles.

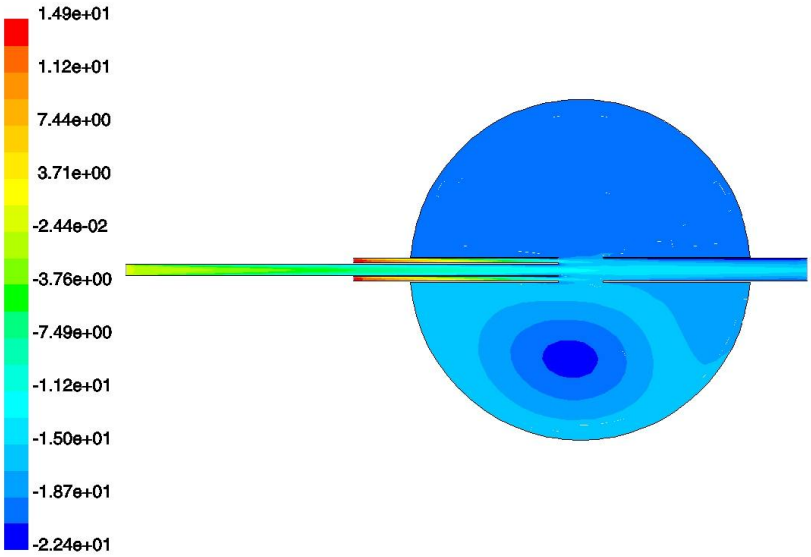


Figure 3.15: Pressure variation in sample flow (Pa).

density ρ (kgm^{-3})	dynamic viscosity μ (Nsm^{-2})	specific heat C$_p$ (J/kgK)	thermal conductivity
8.4634×10^{-3}	1.5287×10^{-5}	1006.43	2.12×10^{-5}

Table 3.2: The physical properties of the stratospheric air model used in the flow simulations. Note that the specific heat of air is relatively constant with altitude.

To ensure the air flow system would perform well under a wide variety of ambient conditions the flow was also modelled under stratospheric conditions. Here the fluid properties were modified to those given by the 1976 U.S. Standard

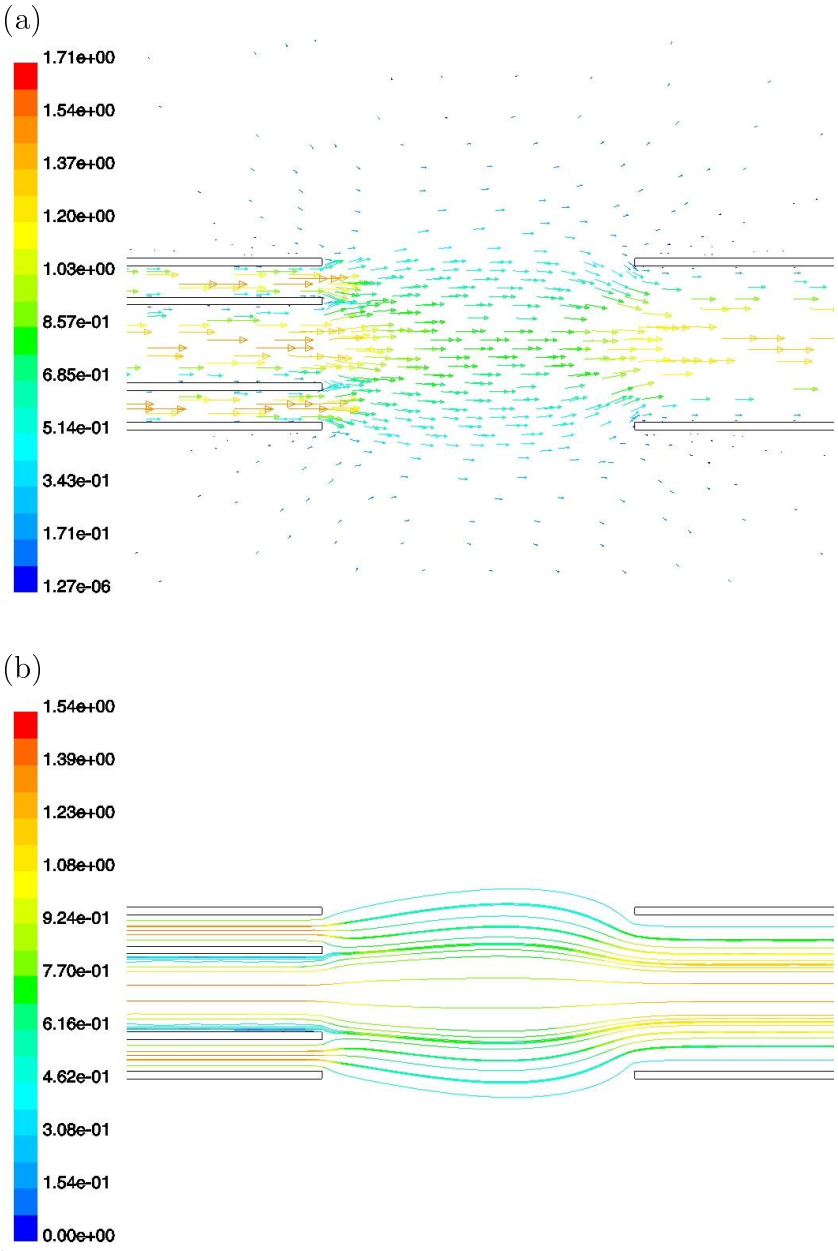


Figure 3.16: (a) Velocity vectors and (b) inlet flow paths for stratospheric conditions. The colour scale indicates velocity magnitude in ms^{-1}.

Atmosphere at 35 km altitude, as listed in Table (3.2). The velocity vectors and pipe flow path lines for this simulation can be seen in Figure (3.16). The obvious difference between these results and Figure (3.13) is a bulging of the air stream as it passes from the inlet to the outlet. This bulging does result in a slowing of the flow across the sample volume as well as increasing the cross sectional area of the jet of sample air. Although the change is minor, both of these effects could influence the measurement of aerosol by the instrument, thus this change in flow will have to be taken into account when later versions of the instrument are used at altitude. Pressure and temperature gradients for the stratospheric case were very similar to those at ground level.

3.4.4 Defining the sample volume

The size of sample volume, and the rate at which air moves through it, are important parameters of the instrument for two reasons. Firstly, to measure the number concentration of particles we need to know how much air has been sampled in a given period. Secondly, for an optical particle counter to be useful it must allow only one particle in the sample volume at once. This condition is met by ensuring that the sample volume is much smaller than the average volume of sample air per particle. Due to the statistical nature of this quantity there will always be instances of multiple particles cohabiting the sample volume: all that can be done is to ensure that these events are minimized.

The sample volume is defined by the intersection of the sample air flow with the illuminating laser beam. If we assume the sample air jet does not undergo significant expansion or contraction as it passes across the sampling volume, it can be described by a 2 mm diameter cylinder. Thus if we neglect the effects of diffraction, we can model the laser beam as a 1 mm diameter cylinder intersecting the airflow at right angles. The equation for the volume enclosed by two cylinders intersecting at right angles in this fashion can be shown to be

$$V = 4 \int_{-a}^{a} \sqrt{a^2 - x^2} \sqrt{b^2 - x^2} dx \tag{3.4}$$

where a is the radius of the smaller cylinder and b is the radius of the larger. Applying this expression to our problem and integrating numerically gives a sample volume of 1.5 μl. If we then take a factor of 100 between the size

of the sample volume and the average volume per particle we get a particle number concentration of 6.6 ml^{-1}. Measurements of aerosol populations with concentrations substantially higher than this would run the risk of being severely affected by multi-particle scattering. As was discussed in Chapter (2) particle number densities in the stratosphere are typically below 10 ml^{-1} (c.f. Figure (2.1)), thus this sampling rate is well suited to stratospheric conditions. However, particle concentrations of tropospheric aerosol and polar stratospheric clouds can be many times higher than this (c.f. Figure (2.5)).

Another important parameter defined by the sample volume is the sample time, or the time taken for a particle to transit the sample volume. This, of course, depends on which part of the laser beam the particle transits – the transit time will be longer towards the centre of the beam. The prototype has been used with a sample flow of approximately 1 m s^{-1}, corresponding to a maximum sample time of 1 ms. If a balloon ascent rate of 5 m s^{-1} was matched, the sample time would be 0.2 ms. This extremely short interval places strong requirements on temporal resolution of the measurement systems used in the instrument. As has already been discussed the LDA is simply unable to measure at this rate, hence we must rely on the PMT system.

The measurement of aerosol number concentration using an optical particle counter is a matter of counting how many particles pass through the sample volume, and is thus governed by Poisson statistics. The standard deviation in the number of particles present in a given volume of air is simply the square root of the number counted in a sample of the same volume. The number concentration (C_n) is given by,

$$C_n = NFt \tag{3.5}$$

where N is the number of particles counted over a time interval t and F is the sample flow rate. If we assume a measurement time interval of 20 s using the sample flow of the prototype instrument (150 ml/min) we find that a particle concentration of 100 ml^{-1} will give an average of 5000 counts, with a standard deviation of 71. The relative size of standard deviation rises as $\frac{1}{\sqrt{N}}$ as the concentration drops: at a particle concentration of 0.1 ml^{-1} we can expect 5 counts with a the standard deviation of 2.2.

3.5 Data recording systems

The final task in designing and constructing the SPARCLE prototype was the creation of systems to retrieve data from the two measurement systems described in Section (3.3). The requirements for these systems are:

- Signals from both the PMT and LDA must be recorded in a format which can easily be modified for use in the retrieval scheme.
- The data recording process must have a minimal impact in the actual measurement process (i.e. there must be as little 'dead time' after a measurement as possible).
- The amount of data recorded must be manageable (i.e. we do not want a large record of high resolution measurements of an empty scattering volume).

The acquisition system designed to fulfil these requirements is shown in Figure (3.17). The system is based around two microcontrollers interfaced to a PC. One of these microcontrollers, which will be referred to as the LDA controller, is devoted to controlling and transferring data from the LDA. The other, which will be referred to as the PMT controller, collects data from the PMT and performs the various measurement collection tasks when requested by the PC. Details of the measurement system are given in Appendix (C.1).

Prior to digitization, the PMT signal is amplified and passed through a low pass filter to remove the majority of the shot noise. To deal with the very large dynamic range of the PMT it is sampled through a 16 bit analog to digital converter (ADC), at a sample rate of 71.4 kHz. In normal data collection mode the PMT controller stores the data in a buffer of user defined size, and compares each sample to a threshold level. If 10 samples in a row are above the threshold a measurement event is said to have occurred. At this point the PMT controller sends a signal to the LDA controller via the pulse detected line and initiates data transfer to the PC. The PMT controller uses an RS232 serial link to the PC, which enables control signals to be transmitted between the computer and microcontroller using a simple command set created for the system (see Table (C.3)). The sampling rate of the PMT controller allows a temporal intensity profile of each scattering event to be captured, with each pulse lasting on the

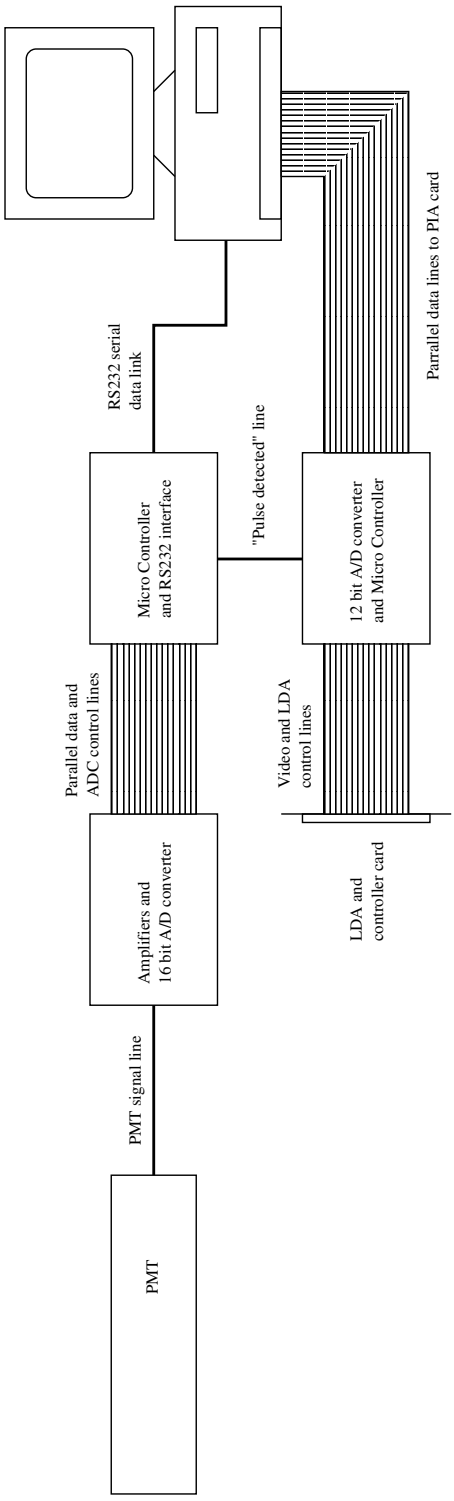


Figure 3.17: Schematic view of the data collection system.

order of 100 samples.

The LDA chip is used in conjunction with a driver card supplied by Hamamatsu. This card provides a mounting for the chip and greatly simplifies the power supply and control signal requirements of the LDA. The driver card receives 7 control signals from the microcontroller and supplies an analog voltage signal as the data output. This signal is digitised using a 12 bit ADC. The microcontroller continuously samples the LDA at a rate of 200 readings per second. At any one time the LDA controller buffers the last two LDA profiles, as well as the one currently being read out. When the system receives the "Pulse detected" signal from the PMT system it finishes reading the current profile, reads one more and then sends the three buffered profiles to the PC. The data transfer is via an 8 bit parallel interface connected to the PC's ISA bus.

The use of a thresholding system in the PMT controller keeps the data rate at a reasonable level (3.5 kBytes per measurement) while still allowing very high resolution measurements of each event to be made. The use of external microcontrollers to provide the control signals for the LDA driver board and ADCs ensures accurate timing and allows the data collection system to be controlled by a single PC. The system is also capable of driving the LDA and ADCs at their optimum speeds, thus reducing the over-exposure of the LDA (as described in Section (3.3.3)).

The limiting factor with this system is the data transfer rate to PC. The PMT controller serial link runs at 38.4 kbit/s, which results in a transfer time of ~ 50 ms for each PMT profile. The LDA data transfer is limited by the speed of the ISA bus, and also takes a few tens of milliseconds to complete a transfer. This amounts to a not insignificant dead time, which results in a higher than ideal number of missed particles. This problem could be overcome by using a data collection system which could continue buffering data while transmitting a measurement to the PC, or by using substantially faster data transfer protocols.

Chapter 4

The retrieval scheme

The determination of particle characteristics from scattering measurements is not a trivial process. The relationship between the scattering pattern produced by a particle and the properties of interest (namely particle size and refractive index) is highly non-linear, leading to a difficult retrieval problem. Although it is possible to infer particle properties analytically from the Mie scattering pattern (Ludlow & Everitt, 2000), the full scattering function (from foward to back scatter) needs to be known to a high degree of precision for this method to work (see Appendix (A) for details). Such measurements are not practical for a small in situ instrument and thus we are limited to a numerical retrieval method. At the time of writing, a successful application of retrieval theory to this problem has yet to appear in the literature. Thus, development of a retrieval scheme to tackle this problem was the major challenge in this work.

Most retrieval problems can be considered as consisting of two parts: firstly a computational *forward model* of the instrument which accurately predicts the measurements given by the instrument for a given state; secondly a retrieval algorithm that accurately predicts the properties of the particle given a signal generated by the instrument. In the case of a highly complicated relationship between measured quantity and quantity of interest, such as the one we have here, it is not possible to analytically invert the forward model and the retrieval must be performed by some numerical method.

This chapter describes the forward model developed for the SPARCLE instrument, the algorithm developed to perform the retrieval and the results of applying this retrieval scheme to simulated SPARCLE data. Details of retrieval

theory are given in Appendix (B).

4.1 The forward model

Since SPARCLE uses the light scattering pattern generated by a particle to infer its size and refractive index, the prime requirement of the forward model is that it is able to predict the scattering pattern produced by a polarized, near monochromatic, light source (in our case a laser diode). Mie theory enables the intensity of light scattered by a given spherical particle at any given scattering angle to be easily computed. The development of a forward model for spaced measurements of intensity is therefore relatively straight-forward, provided procedures are available to do the Mie computations. In this study the Mie routine originates from the work of Grainger (1990). A brief description of Mie scattering theory and its application to this project is provided in Appendix (A).

The design of the proposed instrument, described in Chapter (3), calls for two forward models, one for the measurement of the scattering pattern by a linear diode array and another for the intensity measured by a photomultiplier. In the former case, the model must predict the optical power measured by each element of the linear diode array (LDA) used to make this measurement. In practice this is fairly simple; due to the very high resolution of the LDA (each pixel subtends an angle of between only 0.11° and 0.22°) they can be well modelled as point detectors. In other words the signal of a pixel centred at 55° is taken as being proportional to the intensity at 55° – no variation across the area of the pixel needs to be accounted for.

The forward model for the photomultiplier is more complicated as it must predict the total scattered power over a wide angular range. To do this the solid angle subtended by the PMT photocathode is divided into a grid over which the scattering pattern can be integrated. Determining the vertices of this grid is in itself a complicated procedure, due to the need to produce a computationally efficient forward model.

4.1.1 Quadrature of the scattering pattern

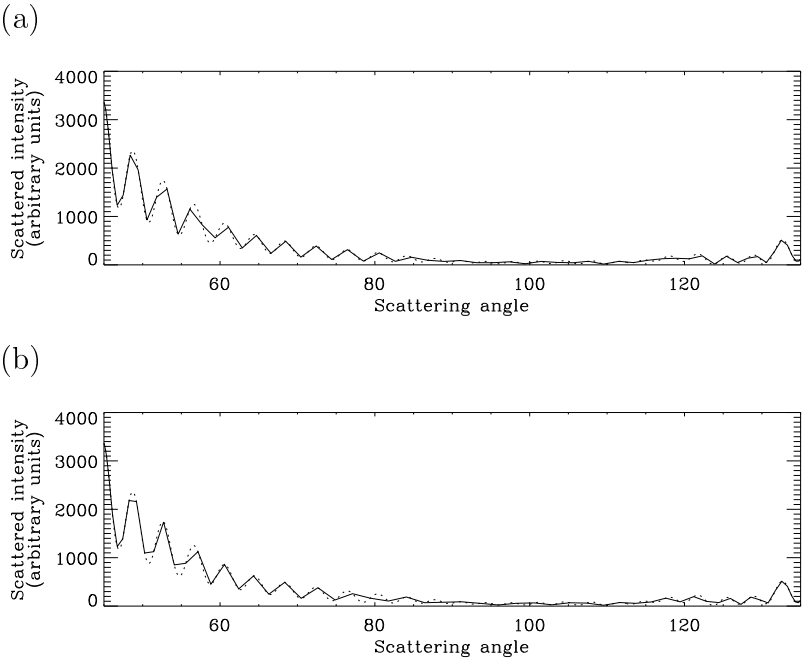
The calculation of Mie scattering intensities is computationally expensive. In order to keep the forward model from becoming too cumbersome, the number of individual Mie calculations required for integration of the scattering pattern across the extent of the photomultiplier photocathode must be kept to a minimum. The efficiency of a numerical integration method depends solely on the method used to determine at which points the function of interest is evaluated, a process called quadrature. Developing an efficient forward model is therefore a matter of finding a quadrature which works well over a wide range of Mie scattering patterns.

Work of this nature has been done before by those who are interested in calculating radiative transfer in an atmosphere containing aerosols and clouds (Hobbs, 1993, Chapter 5 and references therein). This task involves the summation of the scattering/absorption of a very large ensemble of aerosol particles over a wide angular range. Several schemes have been used to supply the quadrature points; early work used Gauss-Legendre quadrature over the interval [-1,1] (the cosine of the scattering angle from 0 – 180°) but later work by Wiscombe (1977) showed that Lobatto quadrature over the interval [0,180°] (i.e. the scattering angles themselves) provided far superior accuracy. Qualitatively this superiority is due to:

- The fact that Lobatto quadrature includes the end points, 0° and 180°, explicitly.
- Using the [0,180] interval produces many more quadrature points on the forward scattering peak, the description of which is vital to producing an accurate estimation of the scattering.

Figure (4.1) shows an example of these two quadrature schemes across scattering angles between 45° and 135°. Lobatto quadrature produces a better estimate of the integral of the intensity across this range. The reason for this superiority is the increased density of points given by Lobatto quadrature towards the ends of the interval. In general, scattering patterns for particles with a size well above the Rayleigh limit show a very strong forward peak, with their values decreasing several orders of magnitude to a minimum around 80° – 120°, which is then followed by about an order of magnitude increase towards

backscatter. The implication of this is that when integrating a scattering pattern over an interval centred on a scattering angle of $\sim 90°$ it is much more important to accurately characterize the regions near the ends of the interval, where the magnitude of the function is greatest.



Integral values:

True	Gauss-Legendre	Lobatto
565.77	517.07	565.78

Figure 4.1: Quadrature of the scattering pattern for a water particle (refractive index $m = 1.33 - 0i$) with a size parameter of $x = 50$: (a) using Gauss-Legendre quadrature on the interval $\left[\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]$ and (b) using Lobatto quadrature on the interval $[45°, 135°]$. In both cases the quadrature was of order 60. The true value of the integral was calculated using a five-point Newton-Cotes integration with 10,000 evenly spaced samples.

In the problem at hand the use of Lobatto quadrature provides an accurate integral of the phase function for a large particle with less than 100 quadrature points (compared to approximately 500 if evenly spaced scattering angles are used). Smaller particles require even fewer points than this due to their much smoother and more homogeneous phase function. However the calculation of the quadrature points is computationally extremely expensive and thus is

only performed once, using enough points to describe an arbitrarily large particle.

To characterize the signal received by the PMT it is necessary to integrate across the range of angle subtended by the photocathode in the scattering plane and in azimuth (the angle from the polarization plane). Thus the forward model involves multiple quadrature calculations, one for the azimuth plane and then one for each of the azimuth quadrature points in the scattering plane. The calculation of the limits for these integrals is not a trivial matter, as the angular extent of the photocathode in azimuth depends on the scattering angle. To define these limits we will move to a new coordinate system (denoted by a tilde) where the axis of symmetry of the photomultiplier lies along the $\tilde{z}$-axis and the laser propagates along the $\tilde{y}$-axis in the negative direction. Following the convention given in Figure (4.2) the edge of the photocathode will be at a constant $\tilde{\theta}$ angle in this frame. We then transform to a new frame by rotating about the $\tilde{x}$-axis so that

$$\begin{aligned} x &= \tilde{x} \\ z &= -\tilde{y} \\ y &= \tilde{z}. \end{aligned} \tag{4.1}$$

Noting that

$$\begin{aligned} \hat{x} &= \cos\phi \, \sin\theta \\ \hat{y} &= \sin\phi \, \sin\theta \\ \hat{z} &= \cos\theta \end{aligned} \tag{4.2}$$

(where $\hat{x}$, $\hat{y}$, $\hat{z}$ are the unit vector in the x, y, z directions, respectively) we find that, in the new frame, θ is the scattering angle and ϕ is the azimuth angle in our scattering problem. Simply comparing Equations (4.1) and (4.2) results in the expression

$$\cos\tilde{\theta} = \sin\phi \, \sin\theta.$$

This gives an equation which allows the limits on the azimuth angle to be calculated from the scattering angle and the angular extent of the photocathode

$$\phi = \sin^{-1}\left(\frac{\cos\tilde{\theta}}{\sin\theta}\right), \tag{4.3}$$

where $\tilde{\theta}$ is half the angular extent of the photocathode. This equation has two possible solutions over the area of the photocathode, one for the lower limit on ϕ and the other for the upper limit.

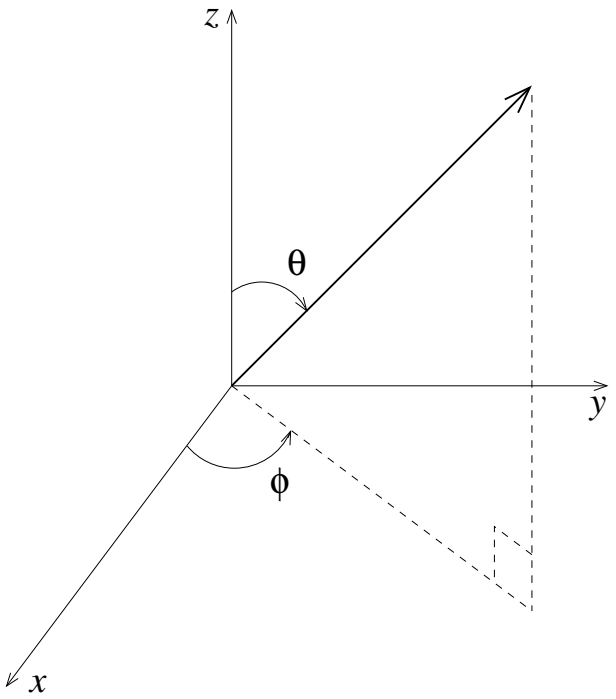


Figure 4.2: The standard spherical polar coordinate system relative to Cartesian axes.

The procedure for the calculation of the intensity detected by the PMT is as follows:

- Lobatto quadrature is used to subdivide the interval across the centre of the photomultiplier. These points are the angles at which the Mie scattering intensities will be calculated.
- At each angle determined in the first quadrature and in the direction perpendicular to it (i.e. in azimuth) Gaussian quadrature is performed across the extent of the PMT photocathode at that point. This creates an array of points covering the angular extent of the photomultiplier.
- The Mie scattering intensity distribution functions (equations (A.3) and (A.4)) are then calculated at each of the angles given by the first quadrature.
- Equation (A.9) is then used to calculate the intensity at each quadrature point across the PMT photocathode.

- Finally, the integrals are calculated, first in azimuth, then in the scattering angle to give a total intensity.

4.2 Constructing the measurement vector

Initial examination may suggest that the construction of a measurement vector to be used in the retrieval process is a trivial task. The obvious procedure to follow would be to form the measurement vector from the raw intensities measured by the LDA and that measured by the PMT. Unfortunately such a simple approach introduces a number of problems. Firstly, this results in a measurement vector with 513 elements, which leads to cumbersome and inefficient computation. Worse still, the extremely high sensitivity of the PMT means that its signal will usually be of a much higher magnitude than that given by the LDA – without some scaling of the two measurements we can expect numerical instability to become a problem in an iterative retrieval algorithm. Finally, the cost function given by such a measurement vector is, in general, extremely complex. The cost function for many particles contain so many deep, sharp local minima that they will usually defeat even the most robust retrieval algorithm.

It was therefore decided to find a way of presenting the information contained in a measurement to the retrieval algorithm in a more compact and concise manner. As discussed in Chapter (3) our measurement vector needs to include information on the structure of the scattering pattern at both large and small scales as well as the magnitude of scattered light. We therefore need a way of presenting both the general shape and the fine structure of a scattering pattern using a vector with relatively few elements.

Steiner *et al.* (1999) present a method of representing this fine structure in a very compact way. They note that the small scale structure of a Mie scattering pattern has, in general, a fairly uniform spacing. A fast Fourier transform (FFT) of a scattering pattern is an efficient way of characterizing this fine structure, as it will result in a single spatial frequency peak.

The measurement vector presented to the retrieval system is constructed by first removing the fine structure of the scattering pattern by smoothing the LDA

measurement by calculating a running mean. Trial and error has shown that, for the instrument layout used, an unweighted average with a window width of approximately 200 pixels provides suitable smoothing for a wide range of particle sizes and refractive indices. The difference between the smoothed and unsmoothed patterns is then taken, thus producing a pattern containing only the fine structure. This is done so that the application of the FFT produces a spectrum with one clear peak. A 32 element vector is then constructed. The first 30 elements of the vector are the smoothed scattering pattern, normalized by dividing by their mean value. The last two elements are the location of the highest value in the FFT spectrum and the intensity measured by the photomultiplier, respectively. To account for the different magnitudes of the various quantities in the measurement vector the logarithm of each element is then taken.

This formulation of the measurement vector has four major advantages over the raw measurements as presented by the linear diode array and photomultiplier; these are:

1. The cost function is somewhat smoother than that for the raw intensities.

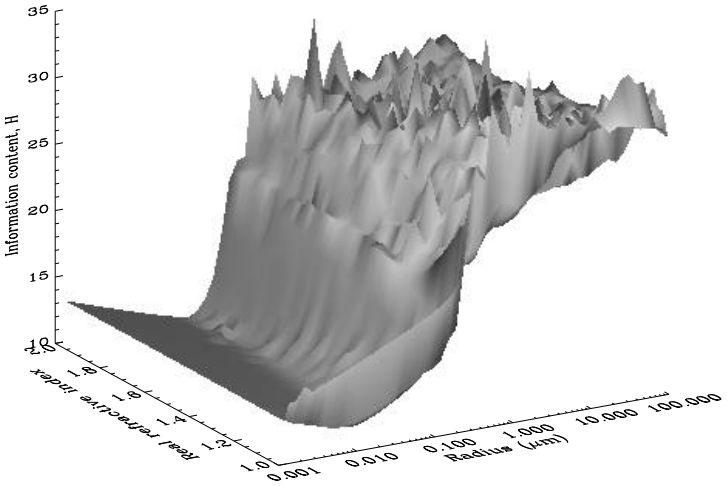
2. The measurement errors are much reduced compared to the raw intensity measurements.

3. The measurement vector is reduced from 513 to 32 elements and significant improvements in the computational effort required to perform the retrieval are made, although these gains must be offset by the cost of the increased complexity of the forward model.

4. The use of a normalized measurement from the diode array greatly simplifies the calibration of the instrument and thereby removes one potential source of error. The use of the intensity measured by the photomultiplier offsets the loss of information due to this normalization.

A disadvantage of constructing a measurement vector in this way is that its elements will no longer be independent. This will result in the covariance matrix for the measurement uncertainty having a non-diagonal form. This in turn increases the likelihood of numerical problems ocurring during the retrieval calculations, since they require this matrix to be inverted (see Appendix (B)).

(a)



(b)

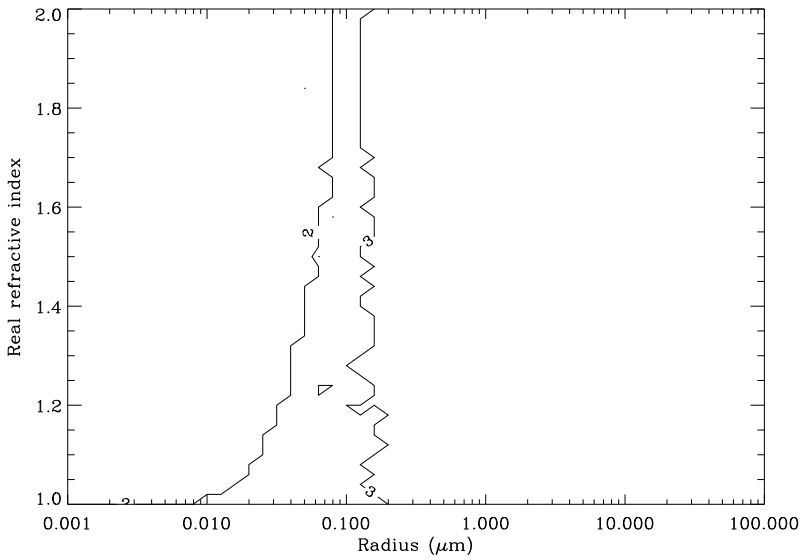
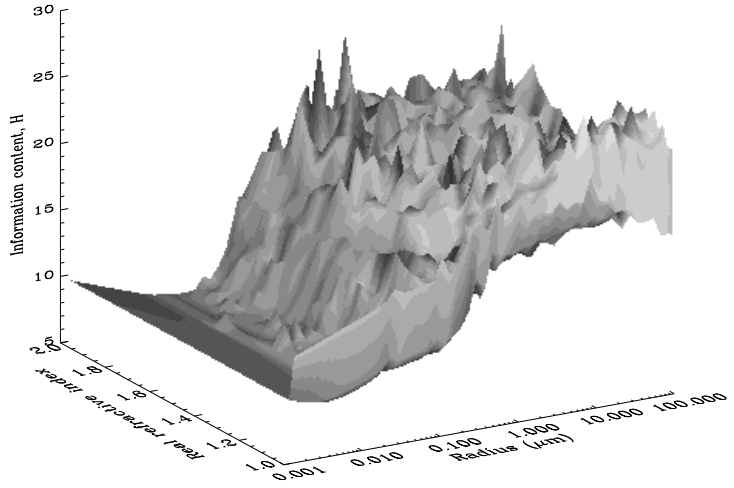


Figure 4.3: The value of information content, H, (plot (a)) and the number of independent measurements (plot (b)) calculated for the measurement vector consisting of 512 intensity measurements between the scattering angles of 55 and 135°, as well as the photomultiplier measurement. The imaginary part of the refractive index was set to -1×10^{-4} for these plots.

(a)



(b)

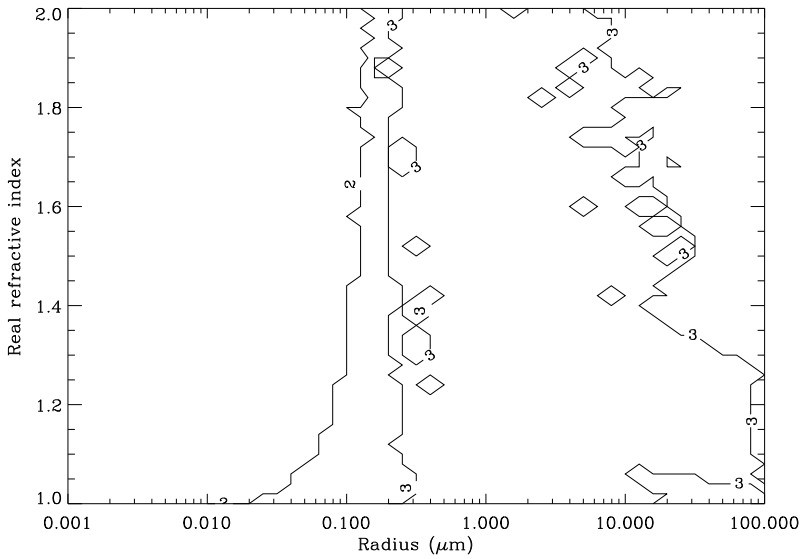


Figure 4.4: The value of information content, H, (plot (a)) and the number of independent measurements (plot (b)) calculated for the measurement vector used in the retrievals. The imaginary part of the refractive index was set to -1×10^{-4} for these plots.

To test the suitability of this new measurement vector the information content (as described in Section (B.4)) of both the raw intensity measurements and the new measurement vector were examined. Figures (4.3) and (4.4) show the results of this analysis for the raw and processed measurement vectors, using non-absorbing particles respectively. It should be noted that the formulation of the information uses a local, linear approximation to the forward model at the true solution. As the forward model is highly non-linear in our case, this approximation cannot be assumed to be accurate. The rapid fluctuations of the information content value, H, seen for larger particles (for which the forward model will be at its most non-linear, due to the complexity of the scattering pattern) in these figures are probably a result of the linear approximation breaking down.

With this caveat in mind, Figures (4.3) and (4.4) show that results for both measurement vectors are similar, with the raw measurement providing a slightly higher value of H on average, as would be expected. In both cases there is only one independent measurement for small particles rising to two at around 0.01 – 0.1 μm and then to three for particles slightly larger than 0.1 μm in radius. This too is as expected, as the smaller particles will act as Rayleigh scatterers. Rayleigh scattering is almost uniform for all scattering angles and the scattered intensity is proportional to r^4, regardless of the refractive index. See Figure (A.1) for examples of Rayleigh scattering patterns.

Figure (4.4b) shows that the number of independent measurements drops below three again for large particles - particularly those with high values of real refractive index. This is probably due to the scale of the fine structure on the scattering pattern becoming too fine for the LDA to resolve clearly, resulting in a loss of information in the FFT section of the measurement.

Figures (4.5) and (4.6) similarly show the information content for absorbing particles (with the imaginary part of the refractive index set to -0.66). Immediately obvious from these figures is the fact that the information content is considerably lower for absorbing particles than for non-absorbing ones. It should also be noted that the manipulation of the measurement vector again results in a lowering in information. The decrease in information content for absorbing particles is likely to be due to the reduction in fine structure for these particles

(see Appendix (A)). The further reduction in H which occurs once the raw measurement is processed is likely to be due to the difficulty in characterizing the fine structure as it becomes less apparent.

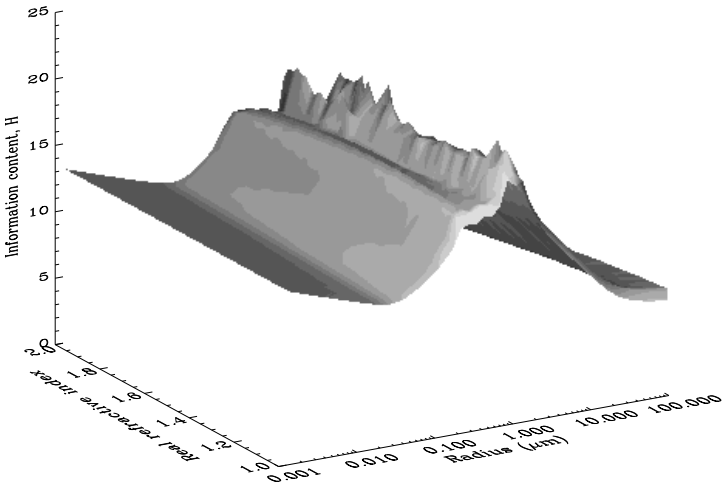
4.3 The retrieval method

As discussed in the previous section, the cost function which our retrieval method must minimize is in general extremely complicated. This means that the retrieval method used to locate the global minima must not be susceptible to becoming trapped in local minima, even if those local minima are extremely sharp and of comparable depth to the global minima. The algorithm chosen as meeting this requirement is a modified simulated annealing (Press *et al.*, 1992). Unfortunately simulated annealing does not provide an optimal solution – it does not give the absolute minima of a function but rather the end point of a constrained random walk. Simulated annealing does not provide any information on the nature of the minima found, so there is no indication of the precision of the result. In addition to these two shortcomings it was also found that even simulated annealing was not reliably able to identify the global minima of the surface. To address these three problems the annealing scheme was modified to incorporate the optimal Levenburg-Marquardt method. This method is prone to becoming trapped in local minima, but will give the exact location of the minima in a surface to within a known precision.

This hybrid retrieval scheme performs a simulated annealing run from which the 100 best states are returned. These best states are the set which have the lowest cost and lie at least some minimum distance in state space from all other members of the set. Each member of this set is then used as the starting point of a Levenburg-Marquardt run and the state which results in the lowest final cost is taken to be the estimate of the true state.

In this work the retrieval has been used with built-in limits to the allowed state, as shown in Table (4.1). These limits are a combination of *a priori* information and physical limits. An imaginary refractive index greater than zero implies a negative absorption (i.e. the particle would be scattering more light than was incident on it), which is clearly nonphysical. The other limits on

(a)



(b)

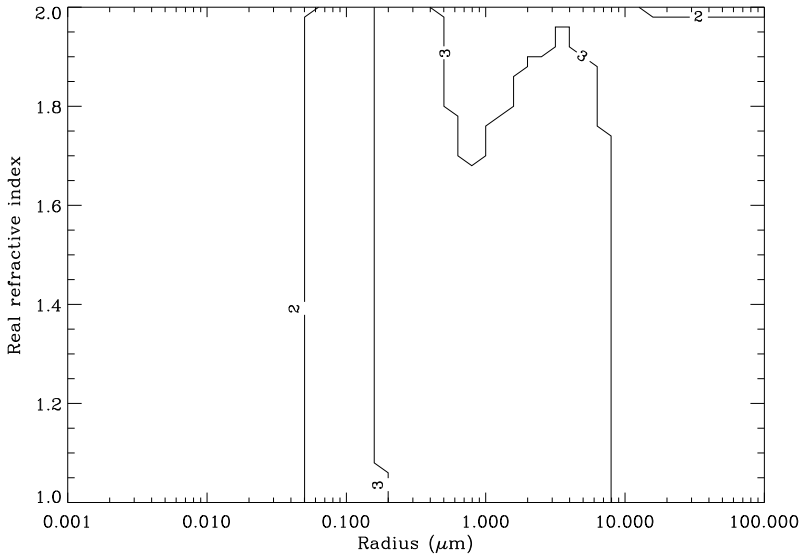
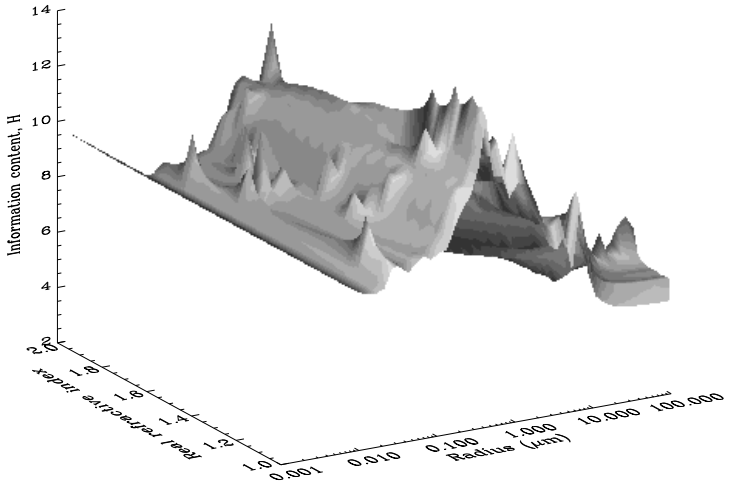


Figure 4.5: The value of information content, H, (plot (a)) and the number of independent measurements (plot (b)) calculated for the measurement vector consisting of 512 intensity measurements between the scattering angles of 55 and 135°, as well as the photomultiplier measurement. The imaginary part of the refractive index was set to -0.66 for these plots.

(a)



(b)

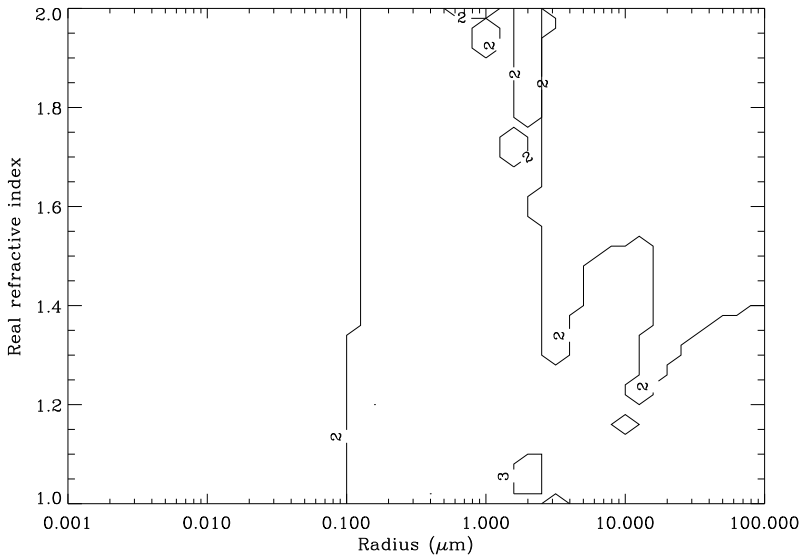


Figure 4.6: The value of information content, H, (plot (a)) and the number of independent measurements (plot (b)) calculated for the measurement vector used in the retrievals. The imaginary part of the refractive index was set to -0.66 for these plots.

the refractive index are essentially *a priori* information and do not represent a physical limit. For example metals have refractive indices outside the allowed range[1], but such particles are unlikely to be found in the free atmosphere. As well as explicitly preventing the retrieval from converging to unrealistic states these limits also avoided numerical problems associated with the discontinuity at a real refractive index of 1.

State element	Lower limit	Upper limit
Radius	1 nm	500 μm
Real refractive index	1	2
Imaginary refractive index	-1	0

Table 4.1: The limits on the allowed state included in the retrieval scheme.

For a more complete description of retrieval theory, simulated annealing, the Levenburg-Marquardt method and the retrieval scheme used here see Appendix (B).

4.4 Retrieval using simulated SPARCLE measurements

The development and initial testing of the retrieval method made extensive use of synthetic measurements generated using the forward model. The use of the forward model to simulate instrument measurements has obvious advantages – the true state is known exactly and so is the nature of any measurement error introduced. Therefore the performance of the retrieval software can be examined with full knowledge of the true state.

The limits of analysis of simulated measurements are determined by the accuracy of the forward model. If the forward model exactly describes the measurement process then the results of synthetic measurement analysis will be indistinguishable from real measurements. Of course, before physical measurements are available it is impossible to predict all peculiarities and problems

[1]For example chromium: $3.48 - 4.36i$ and gold: $0.13 - 3.16i$ at a wavelength of 633 nm (Liu *et al.*, 1995).

that will be present in instrument data. As the SPARCLE retrieval code was developed in parallel to the instrument itself, real measurements only became available after the retrieval code was at an advanced stage of development. Thus, the retrieval code was developed on the basis that the forward model provides a fairly accurate indication of the measurements produced by SPARCLE.

Presented here are results from simulations which retrieve a state vector consisting of particle radius (r) and the real part of the refractive index (n) and simulations which also attempt to retrieve the imaginary part of the refractive index (k). By fixing the value of the imaginary refractive index we reduce the number of independent variables available to the retrieval while the information content of the measurement remains unchanged. Thus if the assumed value of k is correct we would expect the 2-dimensional retrieval to perform better than the 3-dimensional. However, if our assumed k is incorrect the true state will lie outside the allowed state space and we can expect the retrieval to fail. Section (4.4.1) discusses in considerable depth the analysis of the 2-dimensional results, while Section (4.4.2) gives the results of a similar analysis for the 3-dimensional results.

The simulated measurements were generated from a bimodal, log-normal size distribution of particles. Table (4.2) gives the parameters of the two modes of the distribution. The refractive index ranges of the modes were chosen to emulate non-absorbing 'sulphate' or 'organic' particles and absorbing 'elemental carbon' particles which might be found in the free atmosphere (Dick, 1998). The distribution of refractive indices within these ranges was uniform. Figure (4.7) shows the two modes as well as the overall shape of the size distribution generated. Note that the non-absorbing particles were given non-zero imaginary refractive indices. This was done for numerical reasons as zero is the limit of allowed values for absorption.

From this distribution 500 particles were selected at random and a measurement was generated from them using the forward model. Measurement error was simulated by adding 2% Gaussian noise to the raw LDA and PMT intensity values. The noise on the current from both PMT and photodiode will be largely due to shot noise, which has a Poisson distribution, under conditions of constant illumination. However, the signals from both PMT and LDA are integrated with

Percentage of particles	mean radius (μm)	radius σ (in log space)	Real refractive index range	Imaginary refractive index
65%	5.0	2.0	1.43 – 1.50	-0.0001
35%	0.9	3.3	1.60 – 1.80	-0.66

Table 4.2: Parameters of the two components of the particle distribution used in generating the simulated SPARCLE measurements.

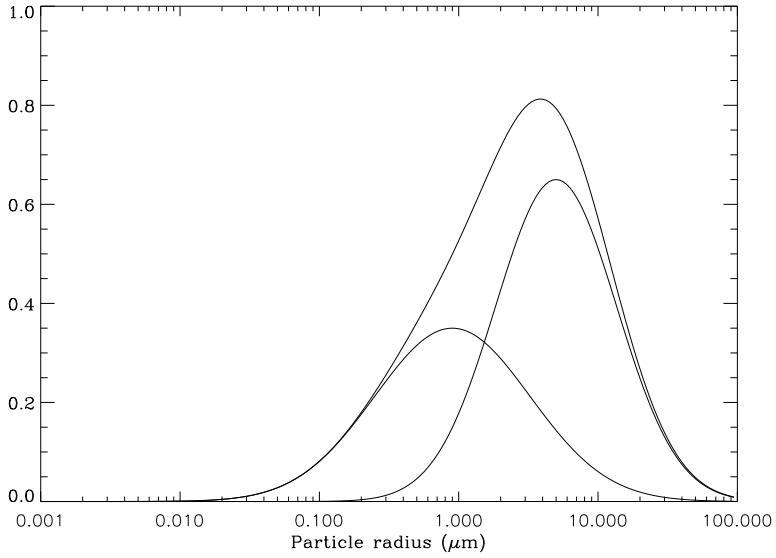


Figure 4.7: The relative particle size distribution, in terms of number density, from which the simulated SPARCLE measurements were generated. The bold line shows the overall distribution, the fine lines the two modes.

respect to time and the measurement system is subject to other sources of error (such as spatial variations in laser intensity and noise from the analog signal processing), it is unlikely that the measurement noise will match the Poisson model. No attempt was made to introduce a background signal (due to stray light, dark current etc) as any observable background signal in the LDA and PMT are minimized.

The *a priori* state used for both 2-dimensional and 3-dimensional retrievals was radius = 10 μm and refractive index = $1.5 - 0.001i$, with an associated

covariance matrix:

$$\mathbf{S_a} = \begin{bmatrix} 625.0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Only the first two elements of the *a priori* state and the top left 2×2 submatrix of $\mathbf{S_a}$ were used in the 2-dimensional case. This *a priori* does very little to constrain the state and was selected so that performance of the retrieval based on the measurement alone could be studied.

4.4.1 2-Dimensional retrieval

For the retrieval of particle radius and the real part of the refractive index only, the imaginary part was assumed to be zero by the retrieval scheme (i.e. the particles were assumed to be non-absorbing). With such an assumption we would expect the retrieval to fail for absorbing particles, but perform better on non-absorbing particles than the full 3-dimensional retrieval. Figure (4.8) shows the retrieved versus the true states and indicates this is the case for larger particles. As can be seen the retrieved radii for the larger absorbing particles are clustered around values of 0.3 to 0.4 μm. This behaviour can be understood in terms of the shape of the scattering pattern of a large absorbing particle. Two points must be considered: (i) because the particles are absorbing they will result in a lower scattered intensity than an equally sized non-absorbing particles; (ii) The scattering pattern for absorbing particles is generally much smoother than that for non-absorbing ones (see Appendix (A)). Taking these two points into consideration it can be predicted that the 2-dimensional retrieval code, using an assumption of non-absorbing particles, would under-predict the size of absorbing particles. What is not obvious from these considerations is the loss, not only of accuracy, but of sensitivity to the particle state in the retrieval. Figure (4.8) clearly shows that for absorbing particles above a radius of approximately 0.4 μm almost all measurements are best fitted by either a 0.3 or 0.4 μm non-absorbing particle[2]. The retrieved refractive indices for the absorbing particles also show this tendency to converge to a fixed value, regardless of true state, although to a lesser degree. A clustering of refractive indices of approximately 1.9 is apparent

[2]Initially these points have the appearance of an artifact caused by a problem with the retrieval code. However, since the retrieved radii are not identical and correspond to widely spread retrieved refractive indices, this is unlikely to be the case.

in Figure (4.8b). For smaller particles however, the retrieved particle radius is very close to the true value, exemplifying the low dependence of the scattering pattern on refractive index for small particles.

This lack of sensitivity to the refractive index for small particles is also evident in Figure (4.8b). There is a large spread in the size of the uncertainties for the absorbing particles. Those states which display the largest uncertainties in refractive index correspond to the smallest particles. Since their scattering patterns have little dependence on the refractive index, the retrieved state is poorly determined with respect to this variable and large uncertainties result.

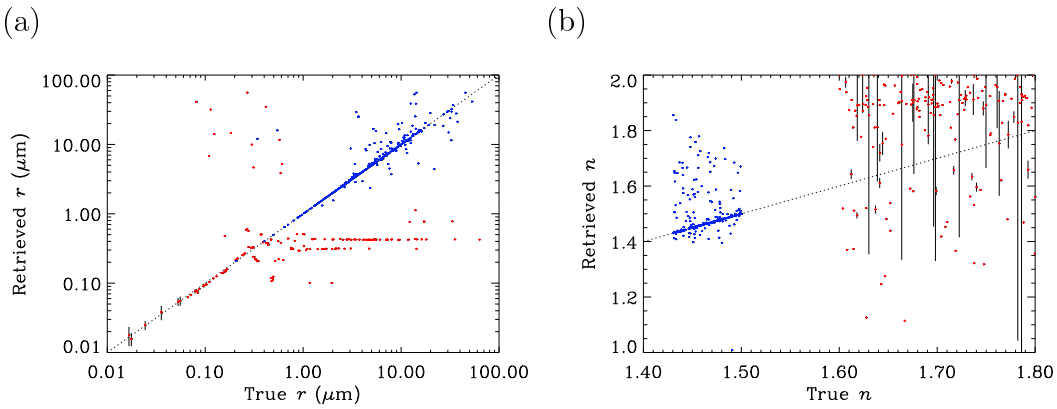


Figure 4.8: Retrieved state versus true state for the 2-dimensional retrieval. The blue points denote particles from the non-absorbing mode of the particle distribution and the red points denote particles from the absorbing mode.

Figure (4.9) shows example cost functions for various sizes of both absorbing and non-absorbing particles, as listed in Table (4.3). These figures clearly indicate the tendency of the cost function to become more complex for larger particle sizes. The cost functions for smaller particles (namely those shown in Figures (4.9a,b,d)) show one clearly defined deep minimum, as well as a much weaker dependence on the refractive index compared to surfaces for larger particles. Figure (4.9b) also clearly shows that, for such a small particle, the location of the minimum in r / n space is not much changed by a large offset in imaginary refractive index. In such a case the weak dependence of the measurement on the refractive index of small particles is actually an advantage.

Plot	True radius	refractive index	Retrieved radius	refractive index
(a)	0.2080 μm	$1.4771 - 0.0001i$	0.2081 μm	$1.4765 - 0.0001i$
(b)	0.0789 μm	$1.7468 - 0.66i$	0.0764 μm	$1.8921 - 0.0001i$
(c)	1.4366 μm	$1.4829 - 0.0001i$	1.4368 μm	$1.4826 - 0.0001i$
(d)	0.2048 μm	$1.7335 - 0.66i$	0.2174 μm	$1.5623 - 0.0001i$
(e)	14.6779 μm	$1.4431 - 0.0001i$	16.5815 μm	$1.4457 - 0.0001i$
(f)	3.3152 μm	$1.7256 - 0.66i$	0.4260 μm	$1.8998 - 0.0001i$

Table 4.3: The true and retrieved states corresponding to the cost functions shown in Figure (4.9).

Figures (4.9c,f) and especially (4.9e) clearly show the presence of multiple deep minima which result in a very difficult retrieval problem. As the particle size increases the general trend of the dependence on the refractive index (both n and k) increasing can also be observed. In fact in Figure (4.9e) the dependence on n is stronger than on radius.

Examination of the cost functions therefore indicates that the retrieval problem behaves as expected. Small particles have relatively simple, well behaved cost functions with little dependence on the refractive index. Larger particles have more complex cost functions with a strong dependence on the refractive index, both refractive and absorptive parts.

It should be noted that the results in Figure (4.8) show that there is a poor characterization of many of the states. All the points in this figure are plotted with one standard deviation error bars, as given by the state covariance calculated as part of the optimal estimation. However these error bars are too small to be seen in most cases and only 25% of the retrieved states agree with their corresponding true states to within one standard deviation. This is a consequence of the forward model not meeting the linear approximation in the region of the solution, as defined by the solution covariance. Since the linear approximation is implicit in the Levenburg-Marquardt method, the covariance of the solution can not be expected to provide a good estimate of the precision of the solution. The detail of the cost function shown in Figure (4.10) shows the situation in a graphical form. The cost function is shown to contain a series of very deep minima closely surrounding the true minima. The retrieved state is marked by the white cross, with the size of this cross denoting one standard

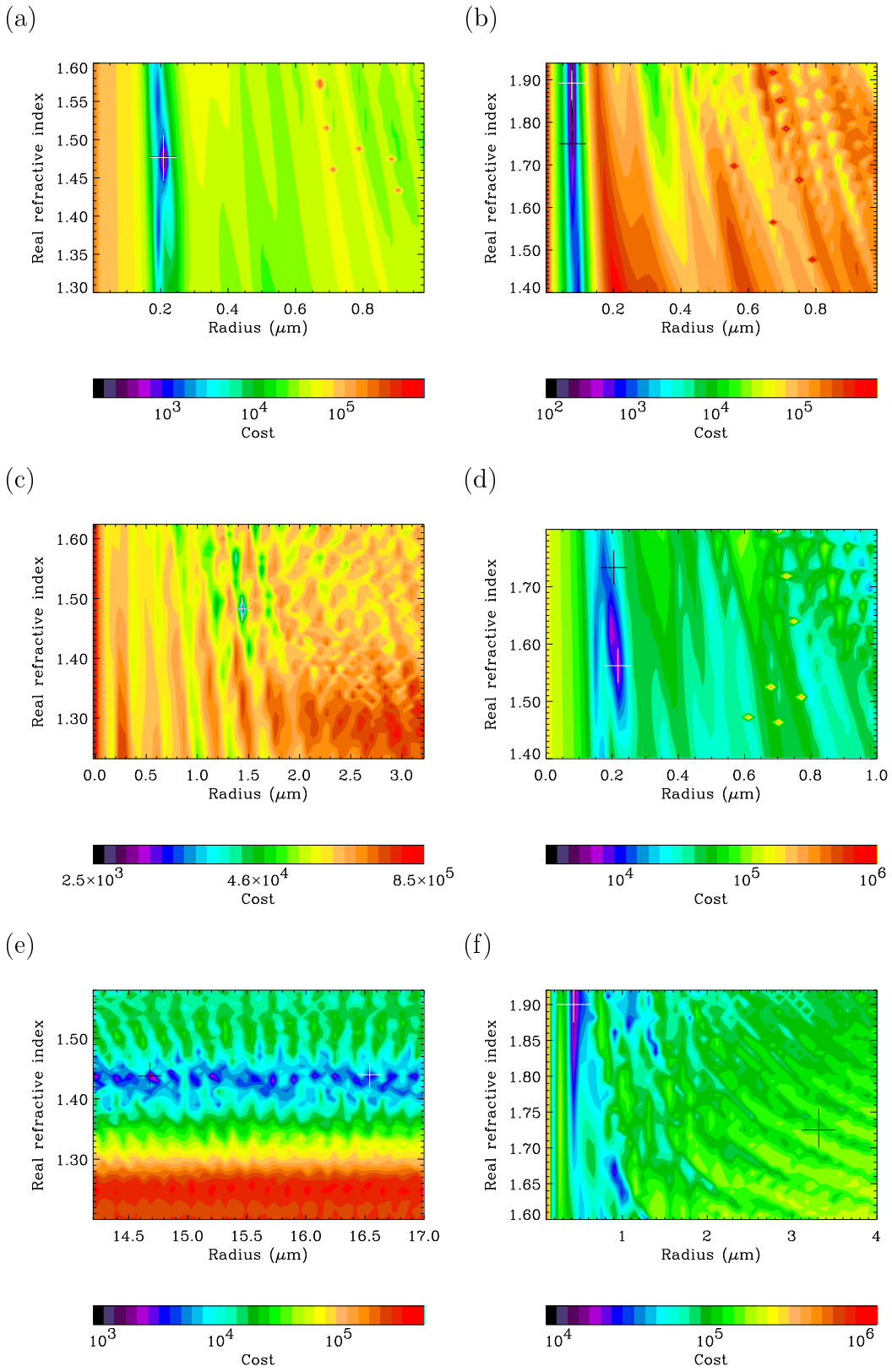
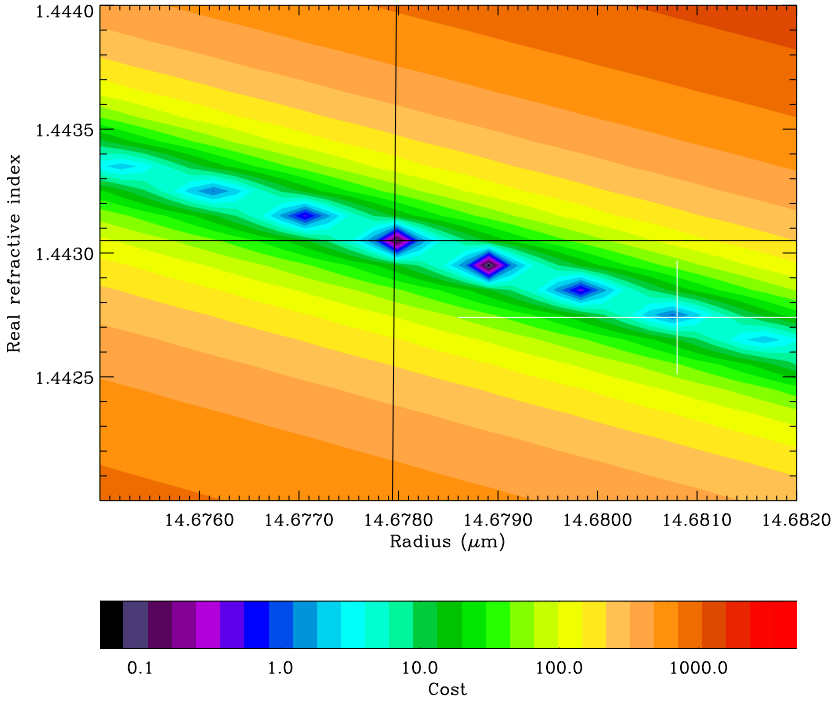


Figure 4.9: 2-Dimensional retrieval example cost functions, see Table (4.3) for the numerical values of the true (marked as black crosses) and retrieved (marked as white crosses) states of the corresponding particles. The plots are arranged with the cost functions of non-absorbing particles on the left and absorbing particles on the right.

deviation error bars; the true state is at the intersection of the two black lines. Figure (4.10) indicates that the retrieved state is a very good estimate of the true one in this case (the discrepancy being only 0.02% in both elements of the state vector) but the error given by the retrieval still underestimates the distance from the true minima. The error bars given by the optimal estimation are based on a single minimum of a Gaussian shape, and hence cannot provide an accurate characterisation in a situation like this.



	Radius (μm)	Refractive Index	
True state:	14.6779	1.44305	-0.0001
Retrieved state:	14.6808 ± 0.0022	1.44274 ± 0.00023	-0.0001

Figure 4.10: Detail of a cost function around the solution. The true state lies at the intersection of the black lines, the retrieved state with its associated error estimates is shown by the white cross.

This picture is further complicated by the addition of measurement noise. As was mentioned in the previous section the retrievals analysed incorporate a 2% random noise. As the generation of the cost functions displayed in this chapter was completed after the retrievals had been performed, this noise is not included in these plots, although the measurement covariance was of the

appropriate form. The addition of noise at this level does not seem to shift the location of the minima in the cost function, but it does change their depths. In the case shown in Figure (4.10) the retrieved minima had a final cost of 43.4 compared with that at the true state of 45.1 (i.e. the retrieved state had a lower cost than the true one, meaning that the location of the deepest minima does not necessarily indicate the location of the true state). Thus, in general, the error estimates presented by the Levenburg-Marquardt algorithm are not to be trusted due to the highly non-linear nature of the forward model. Although the performance of the retrieval scheme in characterising the retrieved states is poor, the states themselves are still very accurate estimates of the true states. The Levenburg-Marquardt part of the retrieval scheme is working successfully with the caveat that the linear approximation to the forward model is not accurate and thus the retrieval cannot provide information on the precision of the retrieved state.

A key element in a retrieval process is the ability to select which retrieved states are acceptable estimates of the true state. This is done using the difference between the measurement and the measurement vector described by the true state. Simply put, we check to see whether the estimate of the state agrees with the measurement to the expected precision. Rodgers (2000) gives the χ^2_s (where the subscript denotes the χ^2 value for a particular solution) comparison between the measurement and the solution of the retrieval as:

$$\chi^2_s = (\mathbf{y} - \mathbf{F}(\hat{\mathbf{x}}))^T \mathbf{S}_{\hat{y}}^{-1} (\mathbf{y} - \mathbf{F}(\hat{\mathbf{x}})) \backsimeq m, \tag{4.4}$$

where $\mathbf{S}_{\hat{y}} = \mathbf{S}_\epsilon(\hat{\mathbf{K}}\mathbf{S}_a\hat{\mathbf{K}}^T + \mathbf{S}_\epsilon)^{-1}\mathbf{S}_\epsilon$ is the covariance of $(\mathbf{y} - \mathbf{F}(\hat{\mathbf{x}}))$, $\hat{\mathbf{K}}$ is the weighting function and m is the number of elements in the measurement vector. (see Appendix (B)). In a problem where the assumption of a linear forward model holds, this value will be similar to the value of the cost function at the solution,

$$-2\ln P(\mathbf{x}|\mathbf{y}) = [\mathbf{y} - \mathbf{F}(\mathbf{x})]^T {\mathbf{S}_\epsilon}^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x})] + [\mathbf{x} - \mathbf{x}_a]^T \mathbf{S}_a^{-1} [\mathbf{x} - \mathbf{x}_a],$$

provided that the contribution of the *a priori* is not large. The highly non-linear nature of the forward model in our case results in large weighting function values at the solution in many cases. This results in a $\mathbf{S}_{\hat{y}}$ which is much smaller than would otherwise be the case and, thus, much higher values of χ^2_s than would be expected.

Figure (4.11a) shows the result of this problem: the χ_s^2 distribution is far from the bell shape expected and most of the retrievals have a χ^2 far greater than predicted. This is true even for those states which agree with the true state within the estimated uncertainty (displayed as dashed line). Figures (4.11b,c), respectively, show the RMS and absolute maximum relative differences between true and retrieved states with χ_s^2 less than a given value. Both show a large increase at a χ_s^2 value of approximately 1100. This value corresponds with the point at which the overall χ_s^2 and that for the states that agree within uncertainties diverge. Thus we can say that if a particular retrieval has a χ_s^2 lying within the first peak of the χ_s^2 distribution it probably agrees with the true state to within uncertainties.

In the case of this retrieval the cost function itself has been found to be a better measure of the accuracy of a retrieval. Figure (4.12) is the equivalent of Figure (4.11) but using the cost function (equation(B.8)):

$$-2\ln P(\mathbf{x}|\mathbf{y}) = [\mathbf{y} - \mathbf{F}(\mathbf{x})]^T \mathbf{S}_\epsilon{}^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x})] + [\mathbf{x} - \mathbf{x}_a]^T \mathbf{S}_a^{-1} [\mathbf{x} - \mathbf{x}_a].$$

This figure shows that the cost distribution has two main peaks. Examination of the relative differences between true and retrieved states associated with these two peaks reveals that all retrievals in the low cost peak are fairly accurate retrievals, while those in the high cost peak are generally inaccurate. Both plots of the relative differences between the true and retrieved states show a drastic increase for retrievals with a cost of greater than approximately 730. Examination of the states with a cost below this value shows that 86% of them are accurate to within 5% of the correct state and, conversely, that 91% of retrievals within 5% of the true state have a cost less than this threshold. This strongly suggests the value of the cost function at the solution is a powerful way of determining the accuracy of a retrieval, regardless of the problems caused by the non-linearity of the system.

Given the provisos that the absorption of the particles must be close to the assumed value and the unreliability of the uncertainties generated by the retrieval, the retrieval scheme does a remarkably good job in describing the aerosol population as a whole. Figures (4.13) and (4.14) show the actual and retrieved state distributions fed into the retrieval, as well as 3 subsets of the

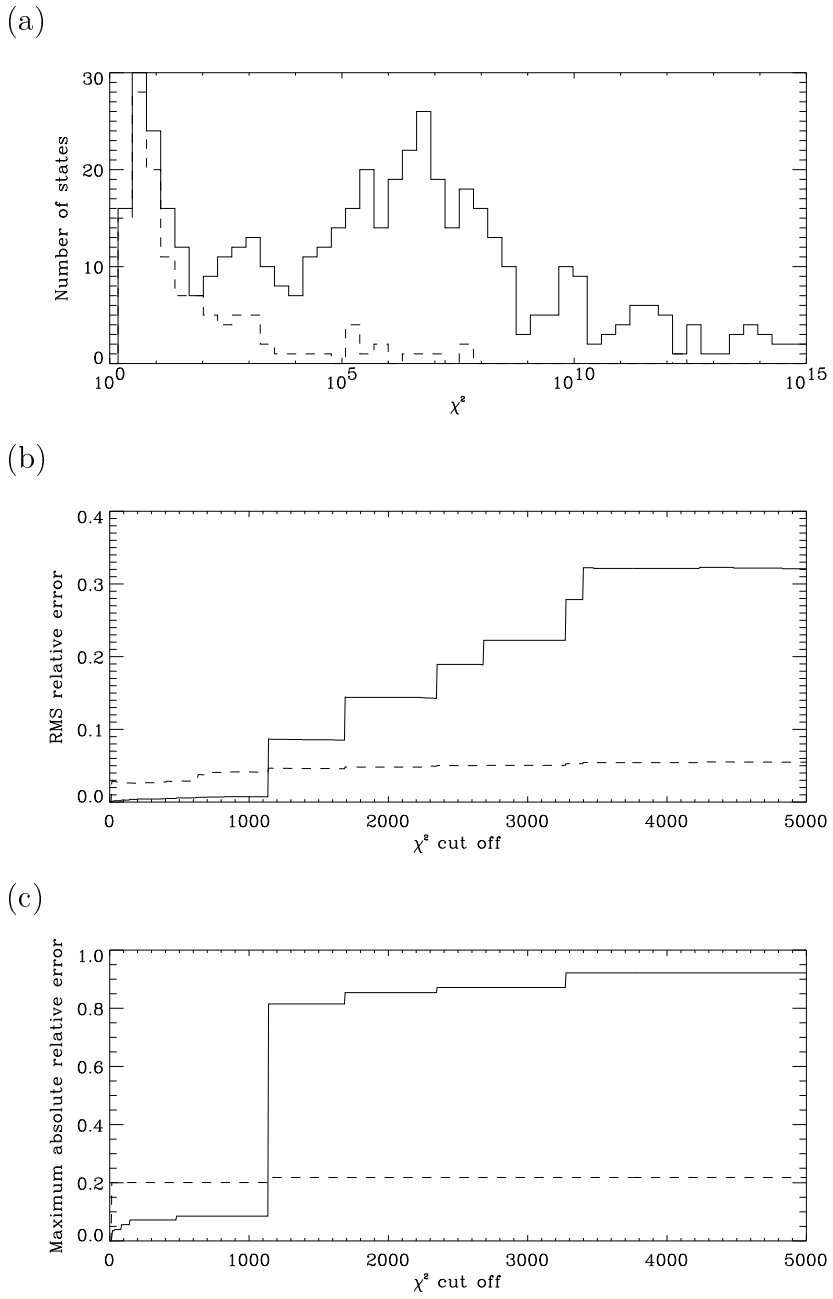
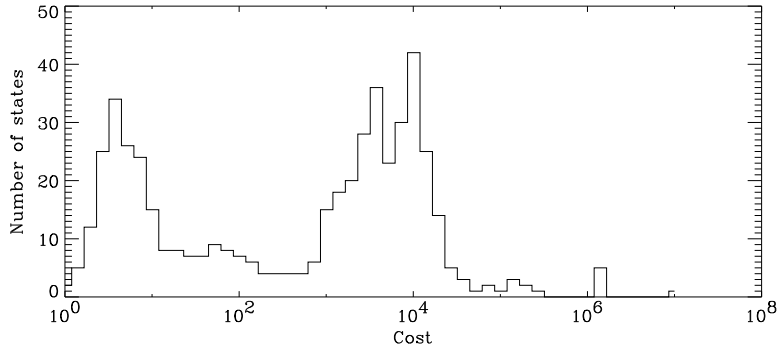
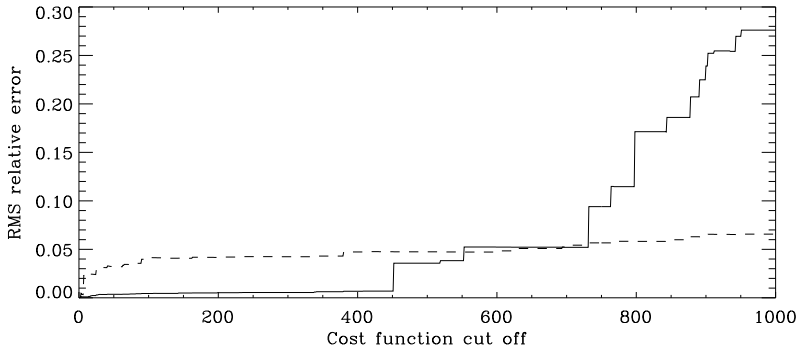


Figure 4.11: The χ_s^2 distribution for 2-dimensional retrievals as given by Equation (4.4) (plot (a)). The distribution resulting from all states is given by the solid line, and that resulting from only those states which agree within their error estimates given by the dashed line. Plots (b) and (c) show the RMS and maximum absolute difference between true and retrieved states for all retrievals with a χ_s^2 less than the given value. Here the solid line denotes the radius difference, the dashed line the refractive index difference.

(a)



(b)



(c)

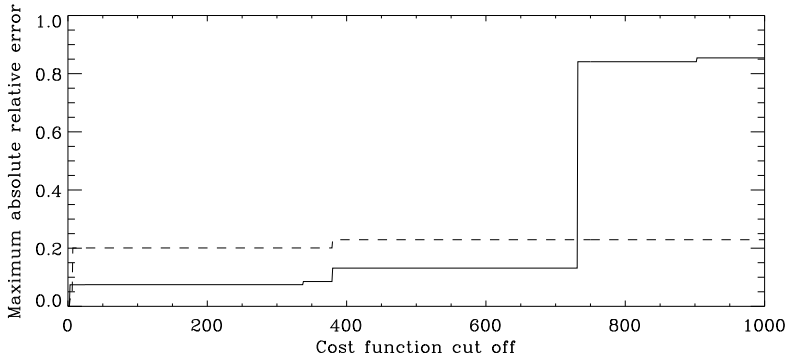


Figure 4.12: The same as Figure (4.11) but for the cost function.

states returned by the retrieval. There are several points to note about these figures. One of the most obvious features is the large peak in the retrieved size distribution at a radius of approximately 0.4 μm which does not correspond to any feature in the true distribution. This is an artifact generated by the larger absorbing particles for which the retrieval seems to have mostly converged to two very narrow size ranges centred on 0.3 and 0.4 μm (see Figure (4.8)). A similar effect is apparent in the refractive index plot for the absorbing particles (i.e. those with higher real refractive indices). Here the retrieved values are around $n = 1.8$ to 1.9 rather than the expected distribution around $n = 1.7$.

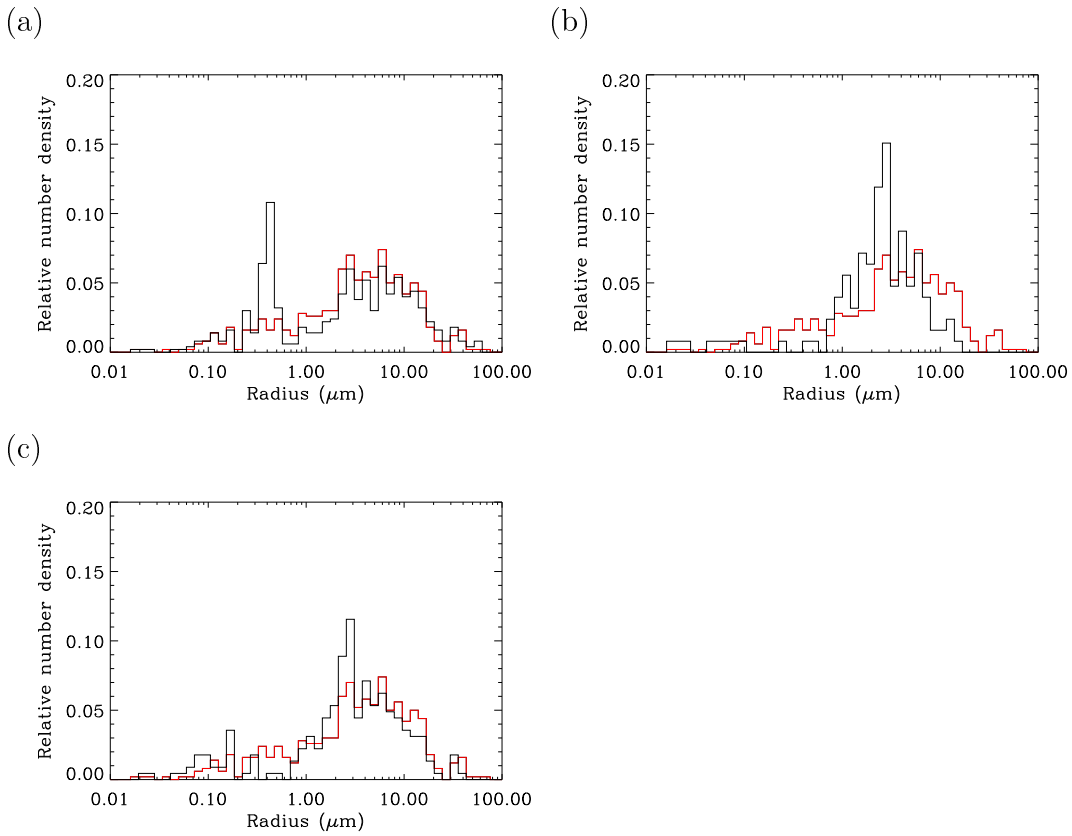


Figure 4.13: Size distributions. In all three plots the red line shows the distribution of the true states and the black the retrieved states. Plot (a) shows the distribution of all of the retrieved states, (b) those states which agreed with the true state within uncertainty and (c) those states with a final cost less than 730.

It should be noted that the distributions constructed from data using the cost function cut off discussed above lack the artifacts due to the inclusion of absorb-

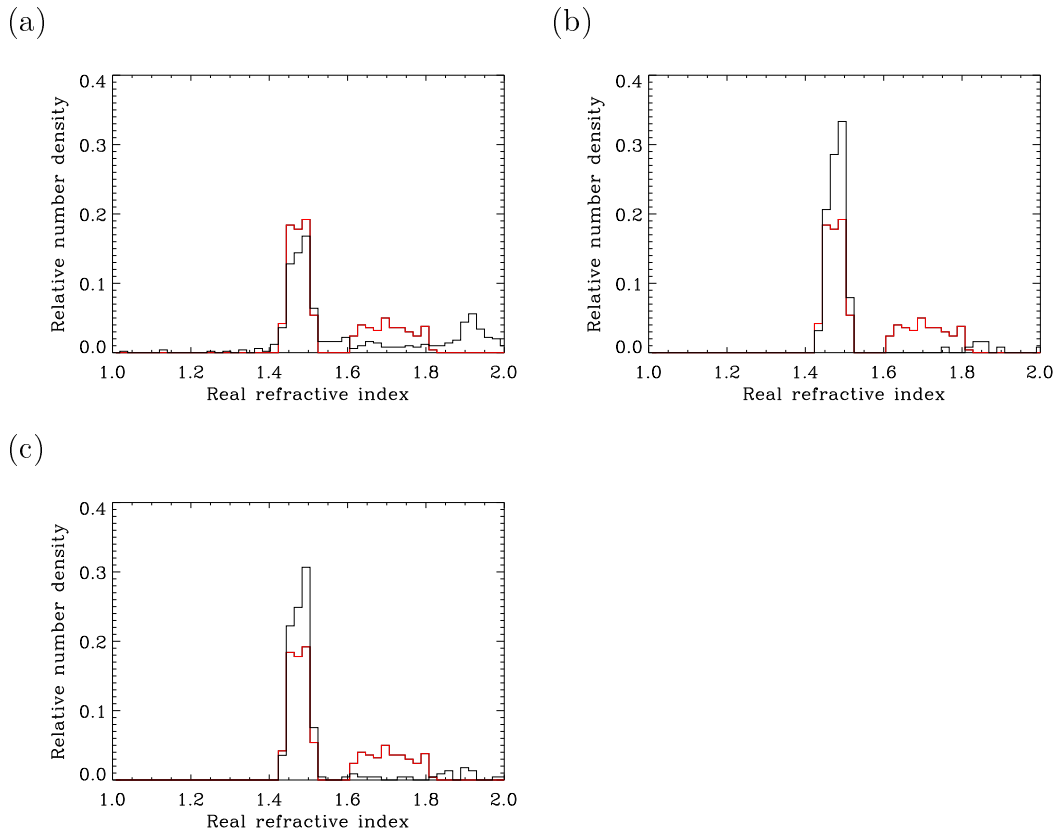


Figure 4.14: As for Figure (4.13), but showing the distributions of the real part of the refractive index.

ing particles. This is further evidence of the reliability of using the location of the minima in the cost function distribution as selection criteria for acceptable states.

Because the error estimates given by the retrieval process are, in general, a poor indicator of the actual discrepancy between the retrieved and true states for any given retrieval we require a new error estimate. Figure (4.15) displays the distributions of actual error for both the radius and refractive index, as well as two Gaussian fits to the distributions, for the retrievals which meet the cost function cutoff criteria. These distributions are far from Gaussian in nature because of their large tails. This means that an error estimate based on Gaussian statistics will be a poor estimator of the true error (as evidenced by the retrieval's poor error characterization). In such a case one can use the mean absolute deviation (MAD) as an error estimator. The mean absolute deviation is defined for a

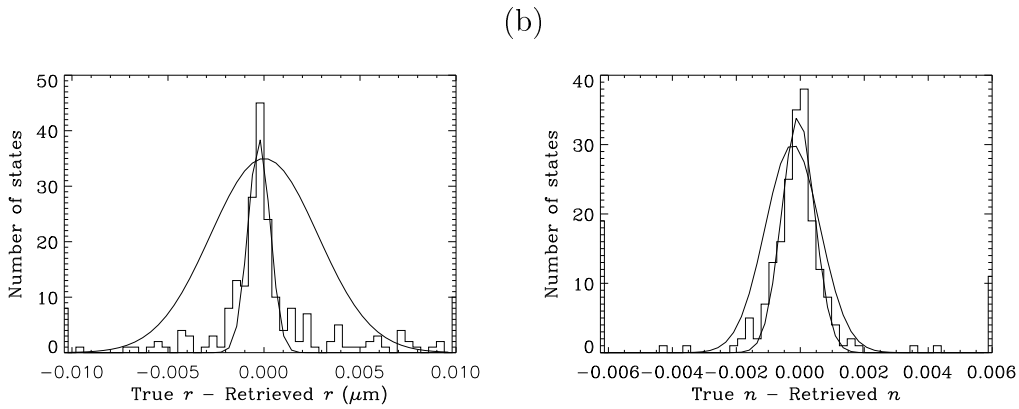


Figure 4.15: 2-Dimensional retrieval size and refractive index error distributions. The narrower of the two Gaussians is a fit to the plotted distribution in each case; the broader is the Gaussian given by the mean and standard deviation of the raw data used in the distributions.

sample of N values, x_i, as

$$MAD = \frac{1}{N}\sum_{i=1}^{N}|x_i - x_{av}|, \tag{4.5}$$

where x_{av} is either the mean, median or mode of the distribution. Because the MAD only depends on the first moment of the distribution, it is far more tolerant to skewed distributions than the standard deviation. This is especially true if the median is used as x_{av} because, in the case of a skewed distribution or one with long tails, the median is a much better estimator of the centre of the distribution. Because the distributions shown in Figure (4.15) are reasonably symmetrical (although they do have very long tails) we would expect the two variants of the MAD to be quite similar, and this is indeed the case:

	MAD (mean)	MAD (median)
Radius r	0.035	0.033
Refractive index n	0.028	0.023

Thus, if the uncertainty in a retrieved state which meets the cost function criteria is given as $x = r \pm 0.033$, $n \pm 0.023$, the true error can be expected to be within this range for approximately 50% of cases.

The ultimate goal of the retrieval process is to reproduce the true distribution of particles from which measurements are taken. Its performance in this regard

has already been discussed in a qualitative fashion earlier in this chapter. To provide a more qualitative measure a χ^2 test was also applied to the resutls[3]. The χ^2 test calculates the probability that a given sample distribution could be generated by randomly selecting samples from a given population distribution. The states fed into the retrieval scheme come from a sample taken from the population distribution shown in Figure (4.7). Thus, if the retrieval performs well we should expect the distribution of retrieved states to resemble a random sample of states from the true distribution.

The value of χ^2 for comparison between sample and population distributions is

$$\chi^2 = \sum_{i=1}^{N} \frac{(S_i - n_i)^2}{n_i} \tag{4.6}$$

where N is the number of bins in the distribution, S_i is the number of states in the ith bin of the sample distribution and n_i is the number expected according to the population distribution. In our case, however, it is more straightforward to compare the distribution of true states presented to the retrieval with the retrieved values. This can be performed using the equation for comparison of two binned data sets,

$$\chi^2 = \sum_{i=1}^{N} \frac{\left(\sqrt{S_t/R_t}R_i - \sqrt{R_t/S_t}S_i\right)^2}{R_i + S_i}, \tag{4.7}$$

where S_t and R_t are the total number of samples in each set, while S_i and R_i are the number in each bin. This χ^2 value can be used to determine how probable it is that the two distributions are sampled from the same population. This approach was taken because calculating the probability of a state lying in a given bin (the value n_i in Equation (4.6)) of the population distribution is not straightforward, whereas the sample distribution presented to the retrieval is immediately available.

The χ^2 value computed can be used to determine the probability that the two distributions are from the same population by comparison to a χ^2 distribution with the appropriate number of degrees of freedom, ν. In the case where the number of samples in each distribution are not constrained to be equal, the

[3]Note that this χ^2 is different from the χ^2_s discussed earlier.

number of degrees of freedom is $\nu = N$ (i.e. equal to the number of bins). The χ^2 probability function $P(\chi^2|\nu)$ gives the probability that a random sample from a χ^2 distribution should be less than the given value. In general we would expect the χ^2 for a given comparison to be approximately equal to the number of degrees of freedom. This follows because selecting a random sample from a population follows Poisson statistics where the standard deviation of the number of samples in a given bin is given by $\sigma = \sqrt{n_i}$ (where n_i is the expected number of samples in a given bin). Noting that the expectation value of the deviation of the sample from the population is approximately equal to the standard deviation, i.e. $\varepsilon\left[(S_i - n_i)^2\right] \backsimeq n_i$, Equation (4.6) becomes,

$$\chi^2 \backsim \sum_{i=1}^{N} \frac{n_i}{n_i} = N. \tag{4.8}$$

The probability function for a χ^2 equal to the number of degrees of freedom will therefore be approximately 0.5 – there is an equal probability of a random sample having a χ^2 higher or lower than this value.

The values for both the χ^2 and $P(\chi^2|\nu)$ in this comparison are given in Table (4.4). The fit between the radius distributions is reasonably good but that for the refractive indice is extremely poor, reflecting the loss of virtually all of the high refractive index, absorbing particles (see Figures (4.13) and (4.14)).

	χ^2	ν	$P(\chi^2\|\nu)$
Radius r	44.8	43	0.603
Refractive index n	116.5	25	1.000

Table 4.4: The χ^2 and their probability function values for the distributions of radius and refractive index given by the true states used in the retrieval and those retrievals which meet the cost function criteria.

4.4.2 3-Dimensional retrieval

The results of applying the full 3-dimensional retrieval scheme to the simulated data described previously are displayed in Figure (4.16). It can be immediately seen that the artifacts introduced by the non-absorbing assumption used in the

2-Dimensional retrieval (Figure (4.8)) are no longer observed. The absorbing particles have been identified as such and the retrieved values for all three elements of the state are distributed around their true values. It is also clear that the retrieved states for the non-absorbing particles show slightly more scatter, particularly in the real part of the refractive index (n). These results behave as predicted in Section (4.4).

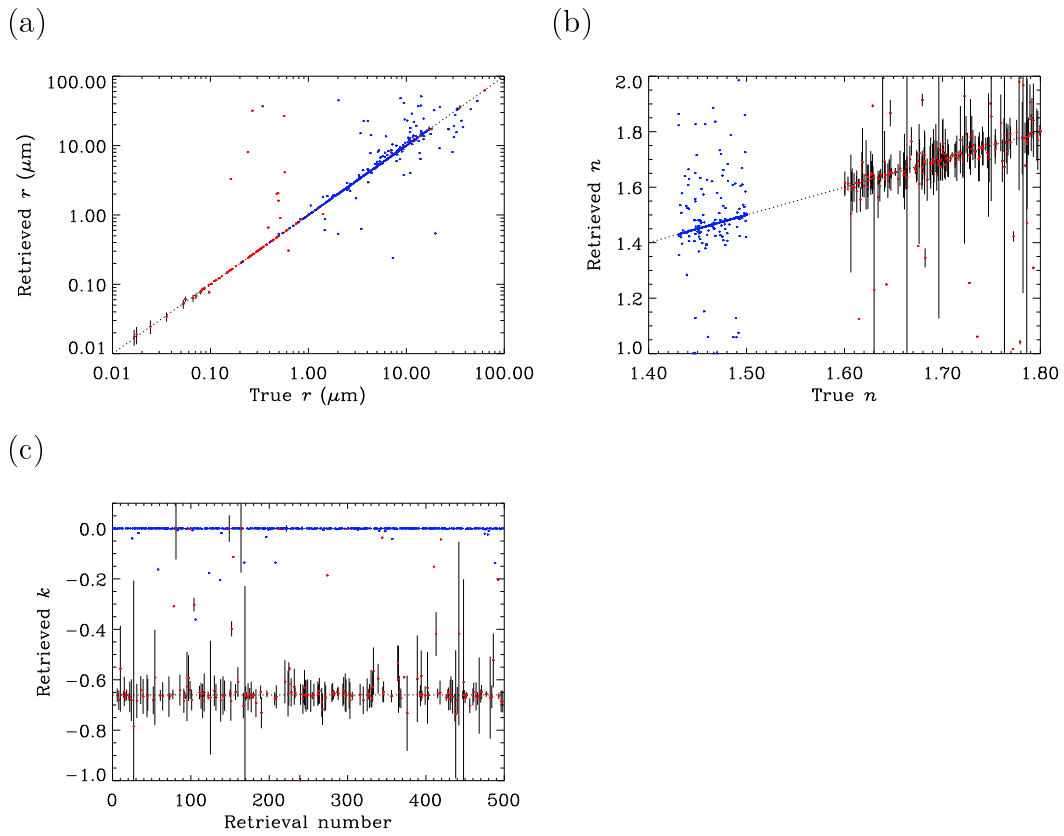


Figure 4.16: Retrieved state versus true state for the 3-dimensional retrieval. As with Figure (4.8), the red points denote absorbing particles and the blue indicate non-absorbing ones. For clarity (c) is plotted with the retrieval number as the independent variable, rather than the true state as is the case with (a) and (b), as there are only two possible values for the true k.

Another interesting feature of the retrieval that these plots make apparent is the difference between the error estimates for the absorbing and non-absorbing particles. In the previous section it was shown that the covariance based error estimates are unreliable, due to the non-linearity of the forward model, and thus

these values cannot be used to indicate accuracy or precision of a given set of retrieved states. However, they do give an indication as to the 'width' of the retrieved minima in the cost function. The refractive index estimates of the absorbing particles have much greater error estimates than the non-absorbing particles, indicating that the cost functions for the absorbing particles have a much weaker dependence on refractive index. Indeed, as Table (4.5) indicates, the error estimates for the refractive indices generally do encompass the true state for absorbing particles, unlike those for non-absorbing partilces.

As the non-absorbing particles are generally larger than the absorbing ones it is tempting to attribute the difference in error estimate to the reduction in sensitivity to the refractive index as particle size is reduced. However, although this effect probably plays some role in the error discrepancy, it is unlikely to explain the differences observed. A comparison between the average error estimates on the refractive index values for particles with radii between 1.0 and 2.0 μm (an interval where the number of absorbing and non-absorbing particles are reasonably similar) is shown in Table (4.6). It should be noted that error estimates are considerably smaller for the non-absorbing particles. This suggests that increasing the absorption of a particle may have a similar effect to decreasing the size (i.e. the dependence of the scattering pattern on the refractive index of the particle is decreased). Indeed increasing the absorption of a particle does result in a lessening of the amplitude of the fine scale structure of the scattering pattern, and it is probably this effect which we detect.

The distribution of the χ_s^2 and cost function at the solution (c.f. Figures (4.11) and (4.12) in the previous section) are shown in Figures (4.17) and (4.18), respectively. As with the 2-dimensional retrievals the χ_s^2 value does not appear to be a good measure of the accuracy of the solution. There isn't a consistent increasing trend between the χ_s^2 value and the difference between the true and retrieved states. Comparison also shows the 3-dimensional retrieval

	Absorbing	Non-absorbing
Real part n	88.6	18.5
Imaginary part k	85.7	1.8

Table 4.5: The percentage of states from the 3-dimensional run for which the retrieved state is within 1σ of the true state.

	Real refractive index		Imaginary refractive index	
	absorbing	non-absorbing	absorbing	non-absorbing
mean	0.062	0.001	0.056	< 0.0001
median	0.063	0.001	0.058	< 0.0001

Table 4.6: A comparison between the average (both mean and median) error estimates on the refractive index for absorbing and non-absorbing particles.

cost function distribution also shows two distinct modes. In this case, however, there is not really a clear jump in the differences between true and retrieved states, although Figures (4.18b,c) do show a more marked increase in the average and maximum discrepancy between cost function values of approximately 700 and 1300. Nevertheless the minimum in the cost function distribution is still a reasonable criterion for determining the acceptability of a given solution.

Taking the location of the minima in the cost function distribution to be at approximately 100 we find that only 51% of these retrievals are within 5% of the correct state and almost all of these correspond to absorbing particles[4]. The reason for this poor agreement is the very small value of the imaginary refractive index for non-absorbing particles. As indicated in Section (4.4) the value of k for the non-absorbing particles was set to -10^{-4} (approaching this value to within 5% would require extraordinary accuracy from the retrieval). Thus we compare particle radii and real refractive indices only. We now find that all of the states which are accurate to within 5% lie below the threshold and that these states make up 96% of all the states below the threshold.

Figures (4.19), (4.20) and (4.21) show the distribution of the three state elements (c.f. Figures (4.13) and (4.14)). It is observed that the retrieval scheme is more successful with smaller particles than larger ones. The distribution for states which satisfy the cost function criteria show a noticeable drop-off for radii greater than ~ 10 μm. The fit between these states and the true distribution is very good in the refractive index distributions however, with only a very slight spreading of the sharp peaks. This plot also clearly shows the effect of the difference in error estimates between the absorbing and non-absorbing particles

[4]There are 344 states with a cost of less than 100. For comparison, using 1000 as the threshold results in 358 – an increase of only 4%.

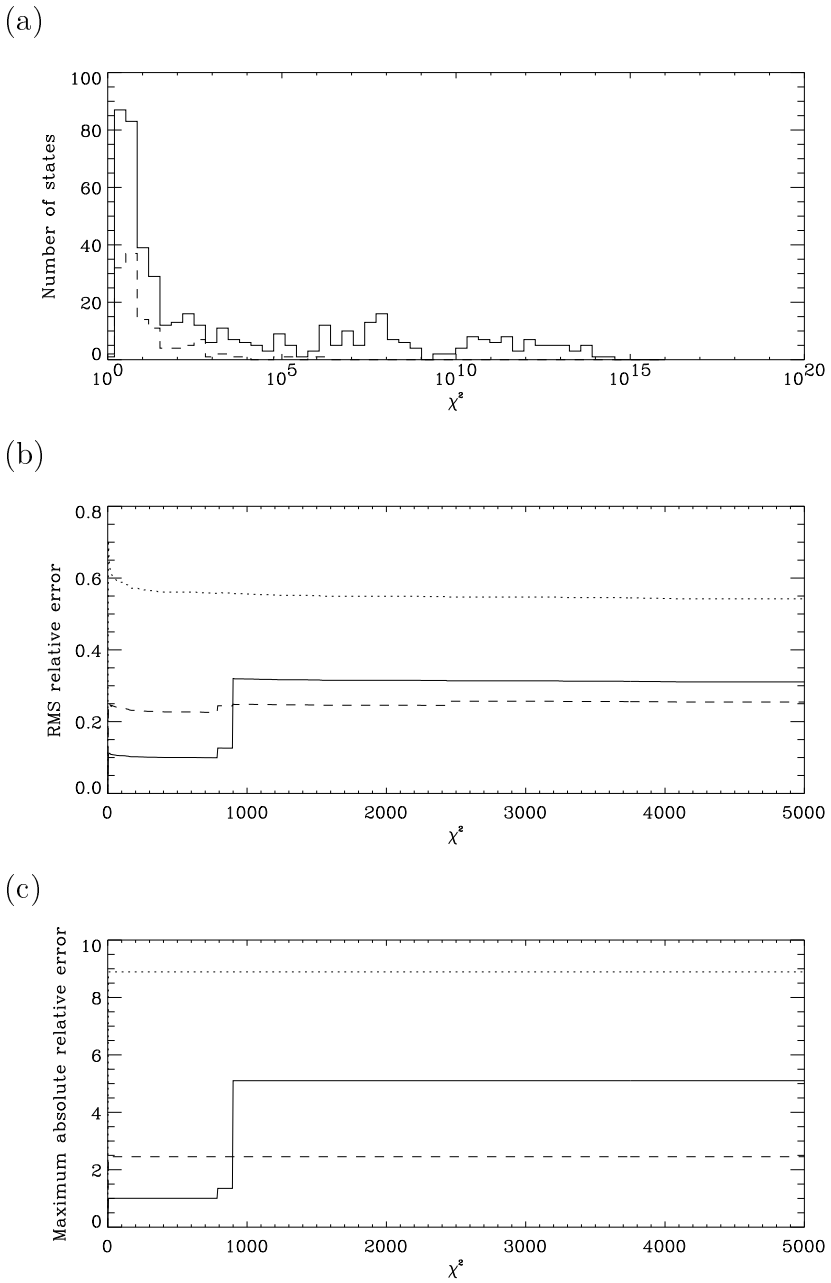


Figure 4.17: The χ_s^2 distribution for 3-dimensional retrievals as given by Equation (4.4) (plot (a)). The distribution resulting from all states is given by the solid line, and that resulting from only those states which agree within their error estimates given by the dashed line. Plots (b) and (c) show the RMS and maximum absolute difference between true and retrieved states for all retrievals with a χ_s^2 less than the given value. Here the solid line denotes the radius difference, the dashed line the difference of the real part of the refractive index and the dotted line the difference of the imaginary part.

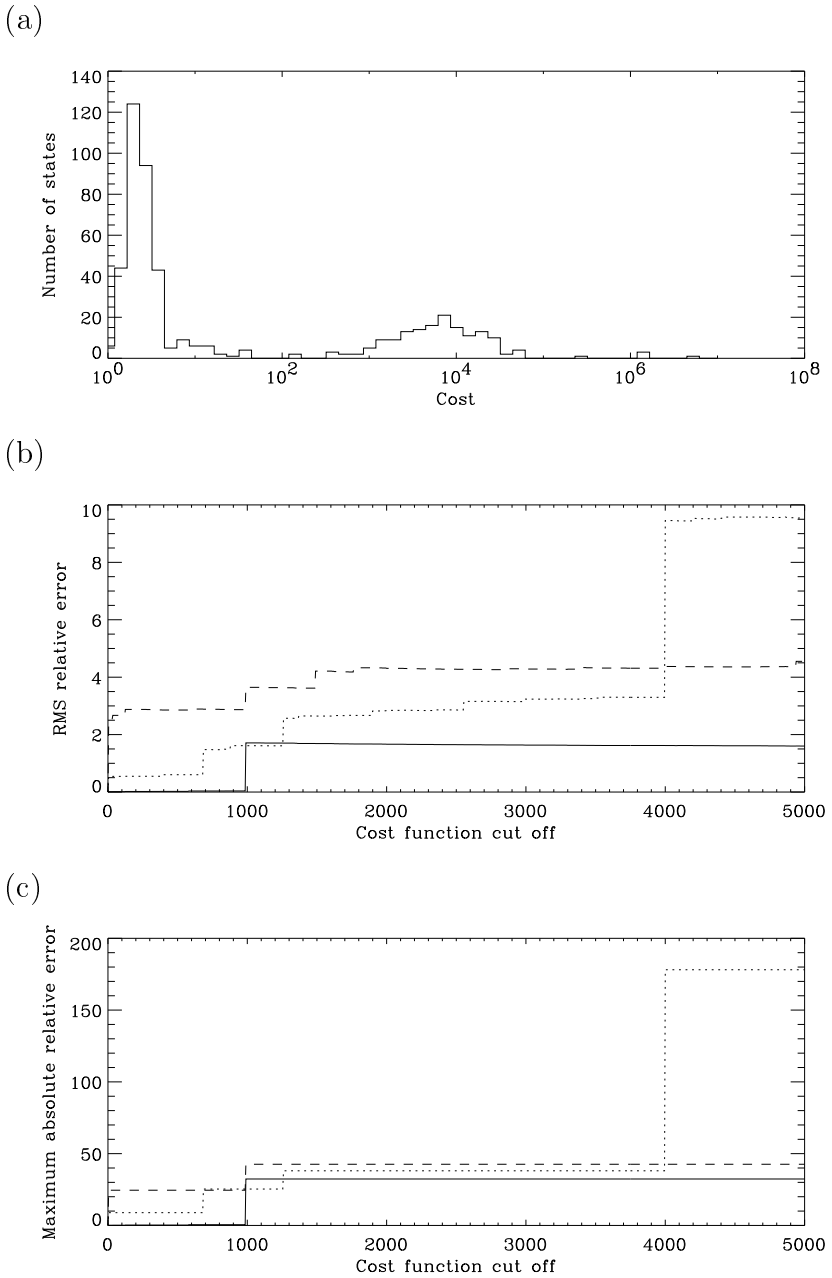


Figure 4.18: The cost function distribution for 3-dimensional retrievals as given by Equation (4.4) (plot (a)). The distribution resulting from all states is given by the solid line, and that resulting from only those states which agree within their error estimates given by the dashed line. Plots (b) and (c) show the RMS and maximum absolute difference between true and retrieved states for all retrievals with a χ^2 less than the given value. Here the solid line denotes the radius difference, the dashed line the difference of the real part of the refractive index and the dotted line the difference of the imaginary part.

(seen in Figure (4.16)). The distributions for those states which agree within their error estimates show a very pronounced bias towards the smaller, absorbing particles evident in all three state elements, the refractive index peaks for the non-absorbing particles having been almost completely removed. Interestingly, the best fit to the true radius distribution is given by the distribution for all retrieved states. This indicates that there is no bias in the radius errors and thus a retrieval is as likely to underestimate a particles size as overestimate it.

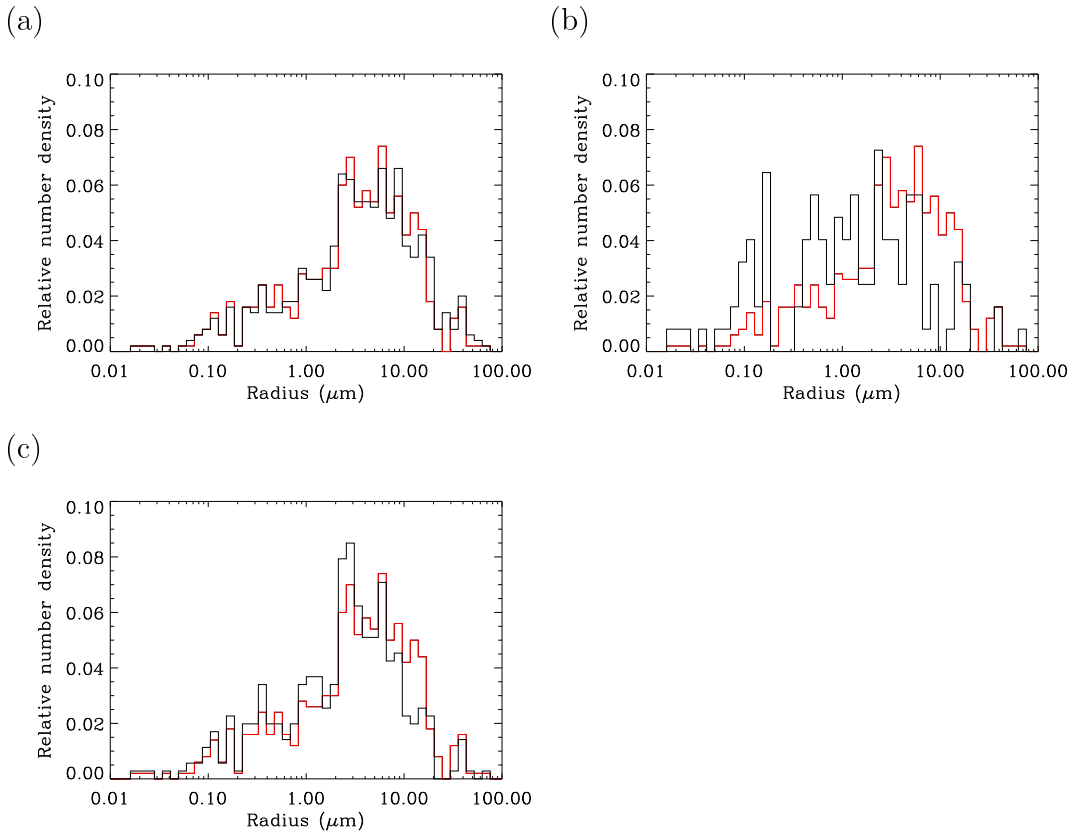


Figure 4.19: Size distributions. In all three plots the red line shows the distribution of the true states and the black the retrieved states. Plot (a) shows the distribution of all of the retrieved states, (b) those states which agreed with the true state within uncertainty and (c) those states with a final cost less than 100.

As with the 2-dimensional retrieval the distribution of true errors of the state are highly non-Gaussian, hence we again use the mean absolute deviation as our error estimator. For the states which comply to the cost function cutoff criteria of 100 we find:

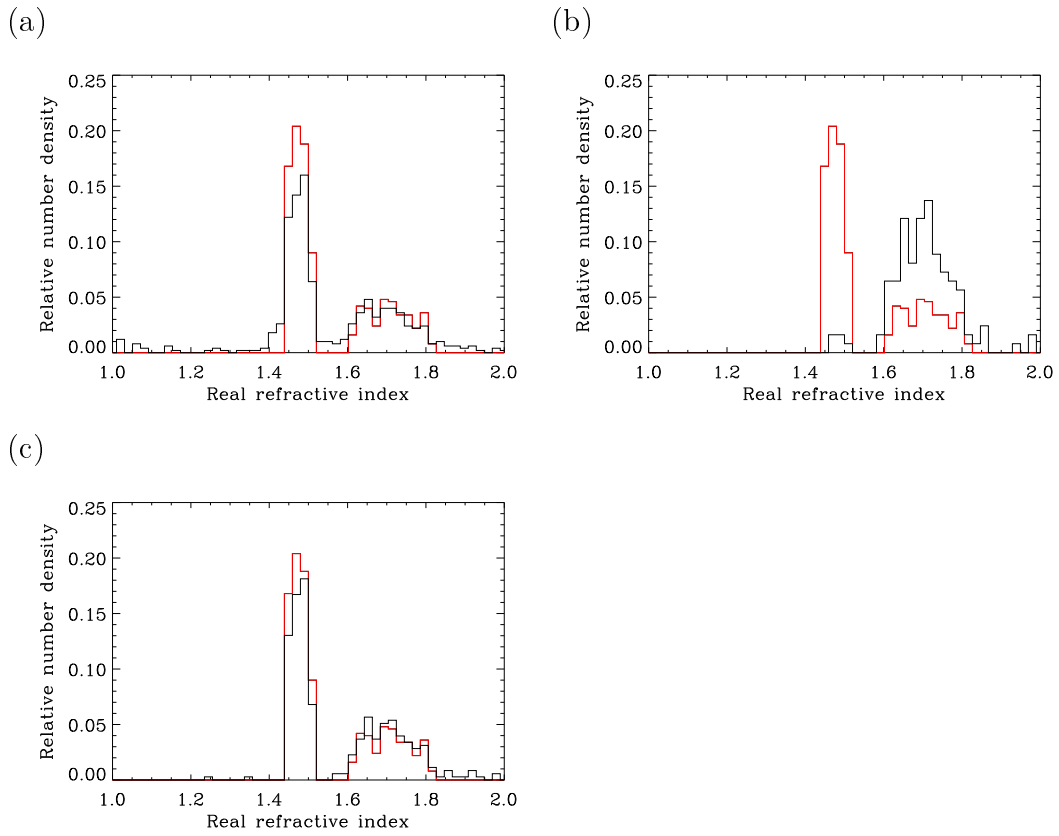


Figure 4.20: As for Figure (4.19), but showing the distributions of the real part of the refractive index.

	MAD (mean)	MAD (median)
Radius r	0.029	0.021
Real refractive index n	0.017	0.016
Imaginary refractive index k	0.016	0.013

These values compare well with the MAD values for the 2-dimensional retrieval, suggesting the cutoff criteria used in each case are equivalent.

Table (4.7) shows the results of applying the χ^2 test comparing the true distributions and distributions of those retrieved states meeting the cost function cutoff criteria. As can be seen the result for the radius distribution is very good, but that for the real refractive index is not, despite the encouraging appearance of Figure (4.20). This apparent discrepancy is due to the non-continuous nature of the distributions. Any spreading of the distributions caused by the retrieval will result in large increases in the χ^2 value, even if relatively few points are involved. If we restrict the calculation of the χ^2 to the limits of the true

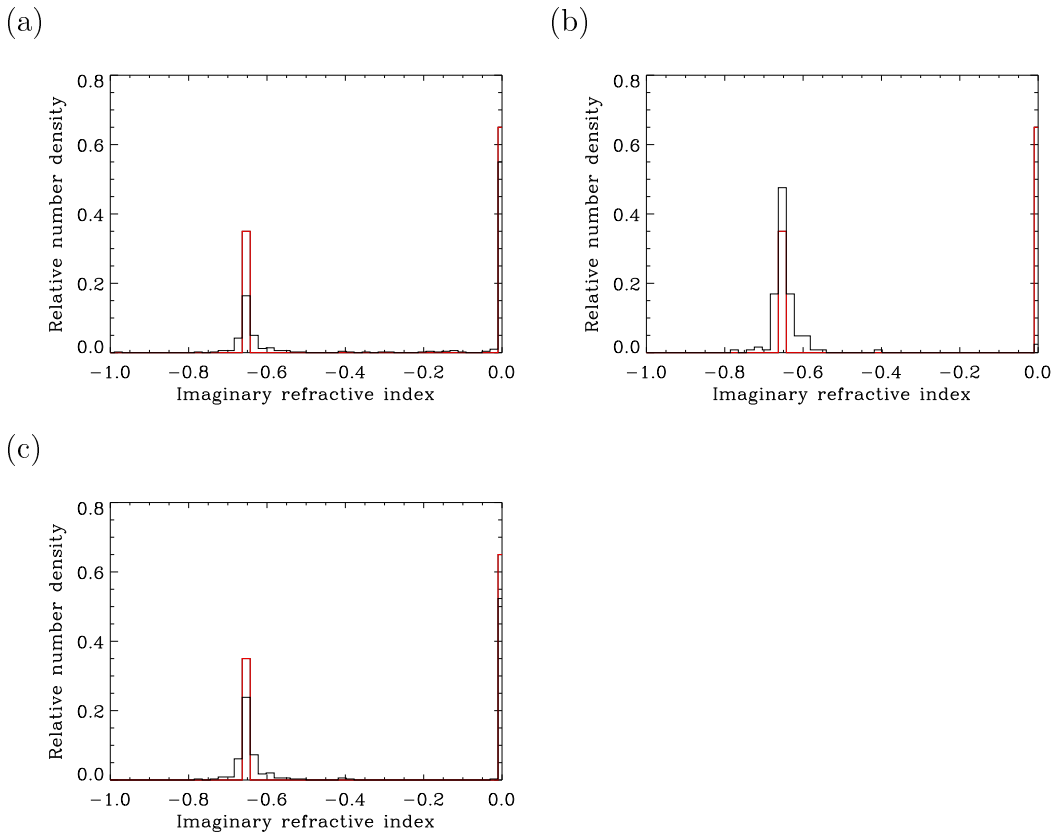


Figure 4.21: As for Figure (4.19) and (4.20), but showing the distributions of the imaginary part of the refractive index.

distribution (rather than including all bins which contain data from either true or retrieved distributions) the values become $\chi^2 = 8.46$ and $P(\chi^2|\nu) = 0.096$ with $\mu = 15$ degrees of freedom. This enormous change in the χ^2 is achieved by removing 18 of the 353 points (5%) included in the retrieved distribution. This problem would be even more apparent in the imaginary refractive index case, since there we do not have a continuous distribution of true states at all, but two discrete values. For this reason no χ^2 test was performed on k.

Overall comparison of the 2-dimensional and 3-dimensional retrievals suggests that there is no advantage in using the 2-dimensional method. There is very little difference in the success rate in retrieving the properties of non-absorbing particles between the 2 and 3-dimensional schemes. The summary of the results the analysis discussed here is that, given a good measurement of the scattering pattern, the retrieval code is able to produce the true properties of a wide range of particles to within a few percent accuracy with a success rate of approximately

	χ^2	μ	$P(\chi^2\|\nu)$
Radius r	24.5	42	0.014
Real refractive index n	34.4	26	0.875

Table 4.7: The χ^2 and χ^2 probability function values for the distributions of radius and refractive index given by the true states used in the retrieval and those retrievals which meet the cost function criteria.

70%.

Chapter 5

Analysis of DAWN-A measurements

This chapter presents the results of applying a modified version of the SPARCLE retrieval scheme to measurements made by the Dual Amplitude Weighted Nephelometer A (DAWN-A). This analysis provided the first test of the scheme on real data, since it was performed before a working prototype of the SPARCLE instrument was available.

The DAWN-A instrument consists of a differential mobility analyser coupled to a multi-angle scattering photometer. These were developed by Wyatt Technology Corporation (in Santa Barbara, California) in the late 1980s, with two individual instruments (models 1 and 2) being constructed. The original design and configuration of DAWN-A is described by Wyatt *et al.* (1988).

The system uses a differential mobility analyser (DMA) (as described by Knutson & Whitby, 1975; Liu & Pui, 1974) to select particles of a given size which are then passed to a light scattering chamber, where their individual scattering patterns are measured. The scattering chamber consists of a stainless-steel sphere which has, in addition to entry and exit points for the sample air flow and illuminating laser beam, 72 detector ports. Fibre-optic feeds can be connected to a maximum of 14 of the detector ports at any given time, with the rest of the ports being plugged. The arrangement of the ports allows for measurements at 9 different angles in the scattering plane, and 8 different angles of azimuth for each scattering angle. At the flow rates used in the work of interest here the DMA selected an approximately triangular shaped size distribution with a half-width

of ~ 10 % of the selected size. This distribution has been approximated in this work as having a standard deviation of 10 % of the mean.

The data presented in this section was collected during the SEAVS atmospheric visibility study carried out in 1995 (Dick, 1998). In this study the Model 1 DAWN-A was one of several instruments set up at Look Rock (near Townsend, Tennessee in the USA) over a 42-day period for the purpose of making measurements of ambient atmospheric aerosol. In the configuration used for this campaign DAWN-A used a $120 - 180$ mW circularly polarized argon ion laser, at a wavelength of 488 nm, as a light source. The detector configuration used 14 fibre-optically fed photomultipliers. Seven of these detectors made measurements at scattering angles of $\theta = 40,\ 55,\ 75,\ 90,\ 105,\ 125,\ 140°$. The remaining seven detectors were equally spaced at 45° intervals in azimuth at the $\theta = 55°$ scattering angle position and were used in the study of particle shape. The angle subtended by each detector in this configuration was 2.5°.

The DAWN-A data, as well as the the original results derived from them are presented in Dick (1998). The data consist of twelve separate scattering patterns: one of a laboratory calibration aerosol of 0.15 μm radius ammonium sulphate droplets, and eleven of size-selected ambient aerosol particles with radii between 0.1 and 0.4 μm. Each scattering pattern is an average of the patterns produced by many individual particles measured for each size. Amongst the instrumentation used in the SEAVS campaign were MOUDI cascade impactors, which were used to provide a measure of the amount of elemental carbon present in the ambient aerosol. By assuming that element carbon was the only constituent contribution to absorption, these measurements provided an estimate of the imaginary part of the refractive index. The accuracy of these estimates is limited by contributions from other absorbing aerosol types, and uncertainties in measurements of the total aerosol volume mass required in calculating the carbon fraction. Nevertheless, this left the real part of the refractive index as the only unknown quantity contributing to the scattering pattern. In this simple one-dimensional case, a weighted least squares fit to the scattering pattern proved adequate in retrieving this value. In the example of the ammonium sulphate calibration aerosol, the technique produced a value for the real part of the refractive index of 1.422 compared with the expected value of 1.413. In general, the error in retrieved refractive indexes for known calibration

aerosols was given to be between ± 0.01 and ± 0.03. Here it is assumed that the uncertainty on the real part of the refractive indices given by the original analysis is ± 0.02.

The analysis of data from DAWN-A was undertaken using the retrieval scheme designed for use with the SPARCLE instrument as a way of testing the scheme for robustness and performance with real measurements. As the original analysis of the DAWN-A measurements gives high precision estimates of the corresponding states, these data are ideal for this purpose. The DAWN-A instrument measures the scattering pattern with a much lower resolution than the SPARCLE instrument. Since the DAWN-A measurements are of small particles, it was hoped that DAWN-A measurements would prove adequate in characterizing the scattering patterns. It is important to note that this study does not attempt to infer the physical characteristics of the aerosol population measured during the SEAVS study. As the relationship between the physical properties of a particle and its scattering pattern is highly non-linear one would not expect, in general, the average of an ensemble of scattering patterns to correspond to the scattering pattern of the average of the corresponding ensemble of particles. By applying the SPARCLE retrieval scheme to the DAWN-A measurements we are simply comparing the performance of two independant analysis methods.

The data was analysed using two separate methods. In the first case the measured intensities were normalized by dividing them by their mean value. As discussed in Chapter (4) this is the method used in the analysis of SPARCLE measurements and it has the advantage of greatly simplifying the calibration of the instrument. The results of this analysis are discussed in Section (5.1).

The second method simulated the use of absolutely calibrated measurement vectors – i.e. measurements of the magnitude of the scattering pattern as well as its shape. To do this a simulated measurement at the original state (as given in Table (5.1)) was generated using the forward model. The simulated measurement was then scaled by a factor which minimized the difference between it and the corresponding actual measurement. This factor was then applied to all the forward model output, providing values which model the magnitude as well as the shape of the orginal measurements. This method should result

Reference number	Original Figure number	Number of measurements	Radius (μm)	DAWN-A refractive index
			± 10 %	± 0.02 (real part)
0	5.11	7	0.15	$1.413 - 0i$
1	5.13a	6	0.10	$1.398 - 0.0043i$
2	5.13b	6	0.10	$1.493 - 0.0083i$
3	5.13c	7	0.15	$1.390 - 0.0043i$
4	5.13d	7	0.15	$1.500 - 0.0083i$
5	5.13e	6	0.20	$1.370 - 0.0001i$
6	5.13f	6	0.20	$1.484 - 0.0002i$
7	5.13g	6	0.25	$1.407 - 0.0001i$
8	5.13h	7	0.25	$1.476 - 0.0002i$
9	5.13i	6	0.30	$1.410 - 0.0024i$
10	5.13j	6	0.35	$1.420 - 0.0024i$
11	5.13k	7	0.40	$1.448 - 0.0024i$

Table 5.1: Summary of the data extracted from Dick (1998). The reference numbers correspond to those shown in the following figures and text. The second column lists the original figure numbers. The number of measurements is the number of separate scattering angles used in the original derivation of the particle refractive index and, correspondingly, the number of points included in the re-analysis discussed here.

in a good approximation to an absolute calibration of the measurement vector based on the following two assumptions. The first is that the original estimate of the state is accurate – given the limited access to the data this is difficult to verify. The second assumption is that the response of the instrument is linear over a wide range of measured intensities. Dick (1998) extensively discusses the response of the DAWN-A detectors and they were found to be reasonably linear with intensity.

One would expect this method to be the more robust of the two because it takes into account the large changes in the total amount of light scattered by different particles, particularly with respect to changes in radius. This analysis is discussed further in Section (5.2).

5.1 Normalized measurement vectors

Figures (5.1) and (5.2) show the results of applying the SPARCLE retrieval system to the DAWN-A data (both the ammonium sulphate calibration aerosol and the ambient aerosol). Here the data has been normalized as described in the previous section. In the case of Figure (5.1) the retrieval was performed with a two dimensional state vector containing the particle radius and the real part of the refractive index, with the imaginary part of the refractive index set to 0. Figure (5.2) shows the results when all three quantities (particle radius, and real and imaginary parts of the refractive index) are retrieved. The χ_s^2 value for a comparison between the measurement and retrieval, as defined in Equation (4.4), and the cost function at the solution are shown for each retrieval.

In the previous chapter it was found that the χ_s^2 value was not a good measure of the accuracy of a given retrieval, due to the highly non-linear nature of the forward model. It was also found that error estimates for retrieved states based on the linear assumption were also unreliable. When applying the retrieval to the DAWN-A measurements the forward model is considerably simpler, and we are restricted to small particle sizes. Figures (5.1) and (5.2) indicate that, in this case, the linearisation of the forward model is more accurate than was found to be the case with simulated SPARCLE measurements. The χ_s^2 values for both retrieval runs are much closer to their expected value (approximately the number of elements in the measurement vector) and the error estimates given by the retrieval are much more realistic than was the case in the previous chapter. Furthermore, there is a much better correlation between the accuracy of a given state (as measured against the original DAWN-A estimates) and it's χ_s^2 value, than exists with the cost function (Figures (5.1d) and (5.2e)). However, it should still be noted that, particularly in the case of the 2-dimensional retrieval, the χ_s^2 values are somewhat smaller than would be expected in the ideal case.

Figure (5.2) shows that the retrieval scheme struggles to provide accurate estimates in the 3-dimensional case. In seven out of eleven cases the retrieved radius is not in agreement with that measured by DAWN-A. In this case the χ_s^2 shows some extreme outliers. The rest of the χ_s^2 values are of the expected size, but their value does not reflect the accuracy of the solution. As with the 2-dimensional retrieval, the cost function at the solution does not appear to be indicative of the accuracy of the retrieval. The retrieval of spurious states

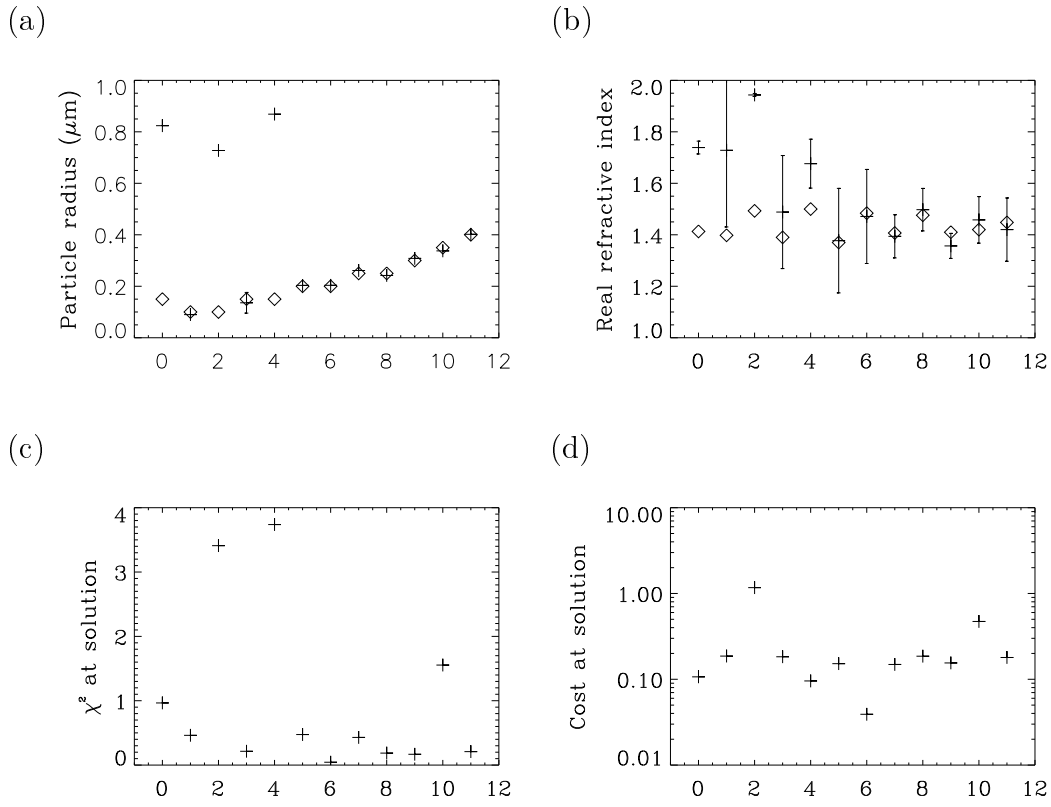


Figure 5.1: The results of applying the SPARCLE retrieval scheme to scattering patterns measured by the DAWN-A instrument in the 2-dimensional case (imaginary refractive index taken to be zero) using the normalized form of the measurement vector. The original estimates are given by the diamonds, with the new retrieved values displayed as crosses with 1σ error bars. Graph (a) shows the radius comparison, graph (b) the real part of the refractive index comparison. Graphs (c) and (d) the value of the χ^2 and the cost function at the retrieved state, respectively. The numbers on the x-axes correspond to the reference numbers given in Table (5.1).

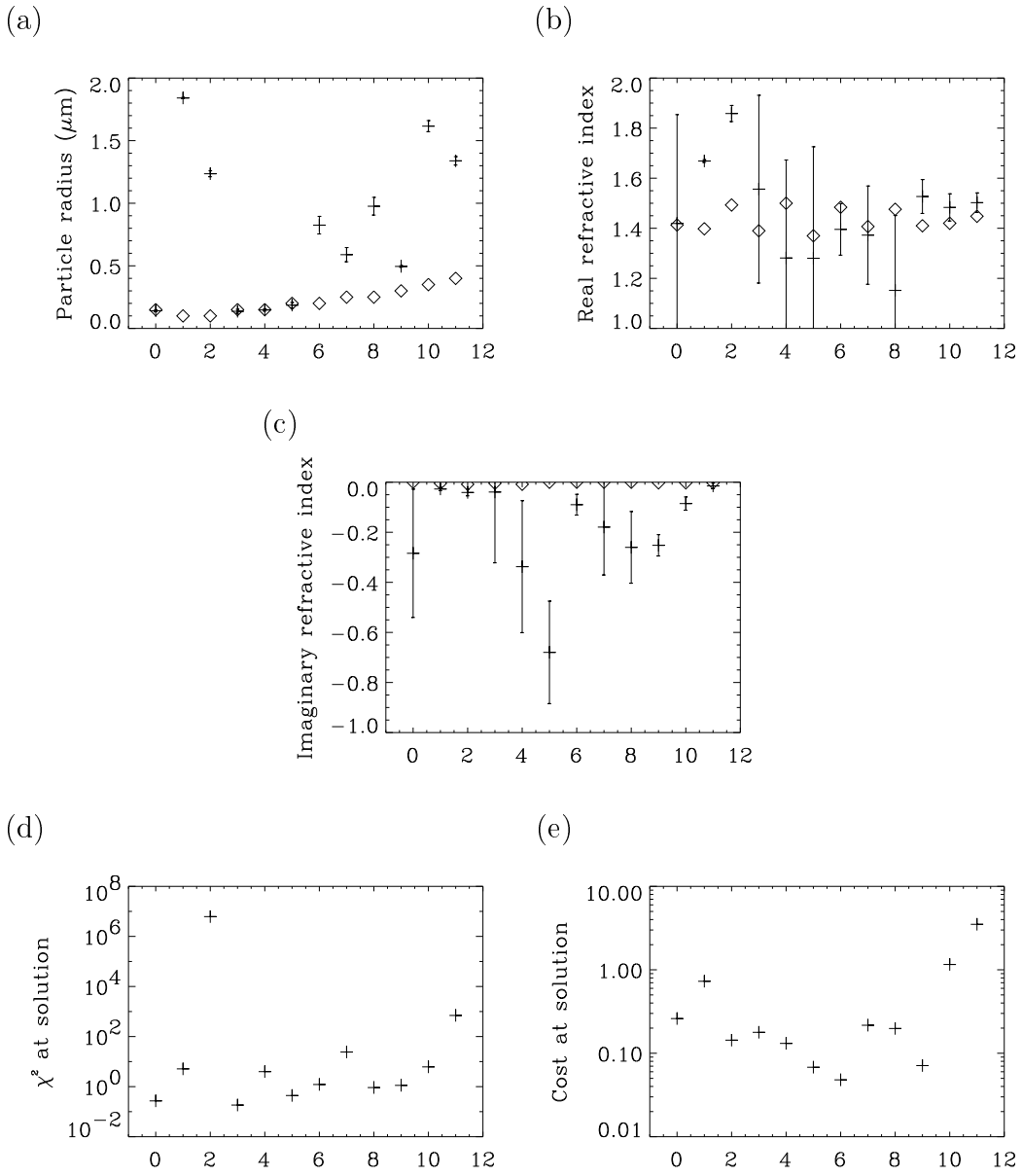


Figure 5.2: As for Figure (5.1) but in the 3-dimensional case. Graph (a) shows the radius comparison, graph (b) and (c) show the comparisons for the real and imaginary parts refractive index, respectively. Graphs (d) and (e) the value of the χ^2 and the cost function at the retrieved state, respectively.

with low χ_s^2 values implies that the cost function contains multiple minima of near equal depth. This is a result of the measurement not containing sufficient information to allow it to be associated with a unique state and thus making reliable retrieval of the correct state unlikely.

Figure (5.3) shows the raw data as well as the fitted scattering patterns generated by the original least squares fit and both the 2-dimensional and 3-dimensional retrievals for all particles. These scattering patterns are scaled by the average of their values at the measurement points, as was the case in the retrieval. They clearly show the scope for ambiguity in the correct solution. Because the magnitude of the pattern is scaled it is possible for a much larger particle (with a much more complicated scattering pattern) to match the measurements as well as a particle of the correct size. In the 2-dimensional case, all the values for particle radius and refractive index agree within uncertainties above a radius of 0.15 μm, and the particle size is very well defined.

The χ^2 surfaces for the 2-dimensional retrievals 1, 5 and 11 are shown in Figures (5.4a,b,c) respectively. Examination of these surfaces indicates that at small sizes the location of the minima in the refractive index direction becomes ill-defined. This effect can be easily understood by looking at the size parameter of these small particles. With the laser wavelength used to make these measurements (λ = 488 nm) an r = 1.0 μm particle has a size index ($x = 2\pi r/\lambda$) of only 1.3; in other words the scattering pattern is beginning to resemble that given by the Rayleigh approximation. In the Rayleigh regime the shape of the scattering pattern contains very little information: it is rather the magnitude of the scattering that is important, particularly as an indicator of size. Also apparent in the χ^2 surfaces are deep secondary minima corresponding to completely inaccurate solutions (at around $\sim$ 0.65 μm radius in Figure (5.4a) for example). It is minima like these which are responsible for the spurious solutions seen at small sizes in Figure (5.1a).

Also of interest in these plots is the apparently over-large error estimates in the retrieved refractive index values, especially when compared with the error estimates for the particle radius. These uncertainties result from the fact that the scattering patterns used in the retrievals are actually averages of many separate scattering patterns – this process resulted in large standard deviations on

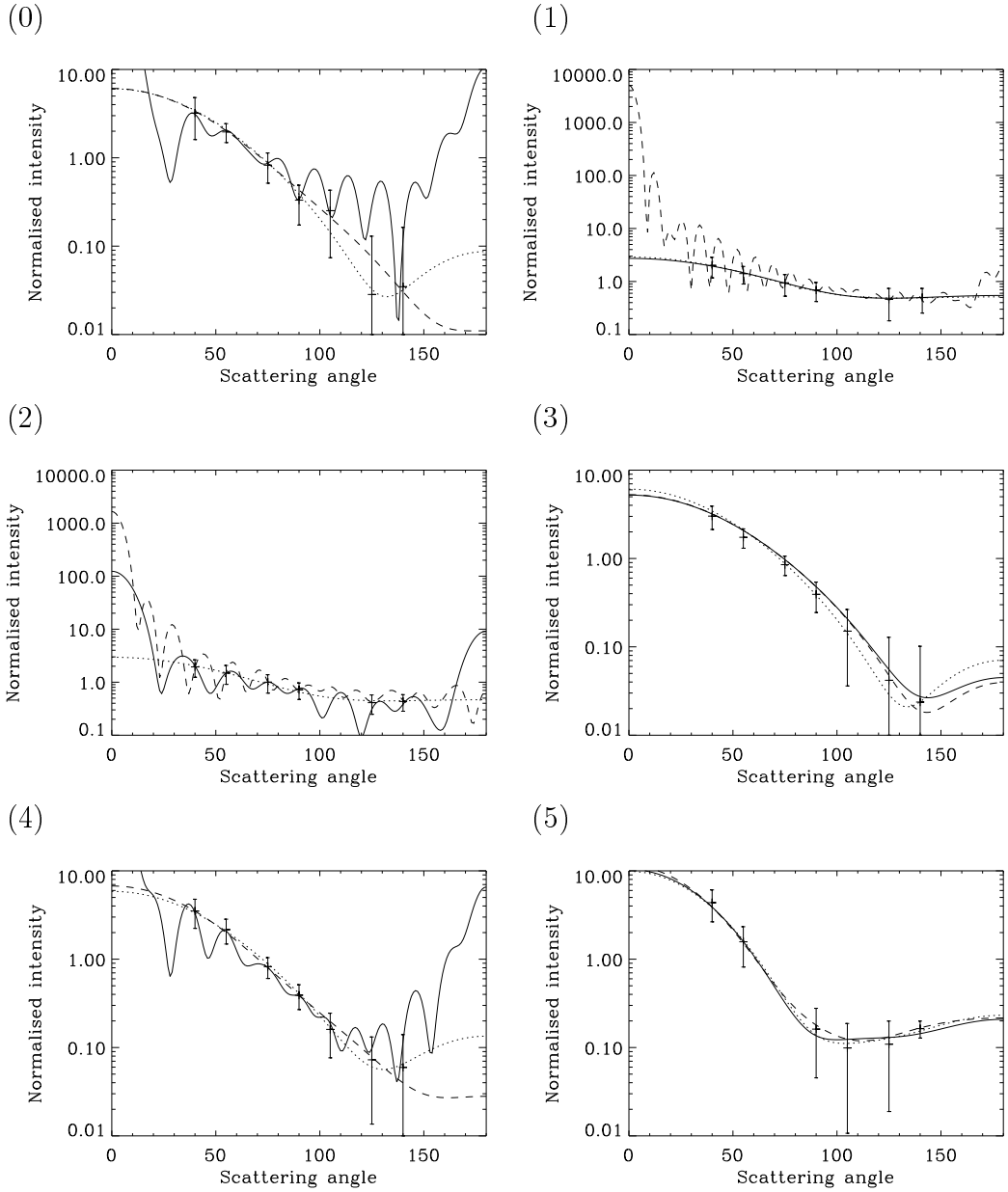
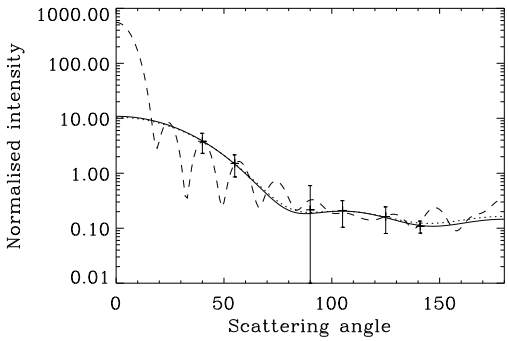
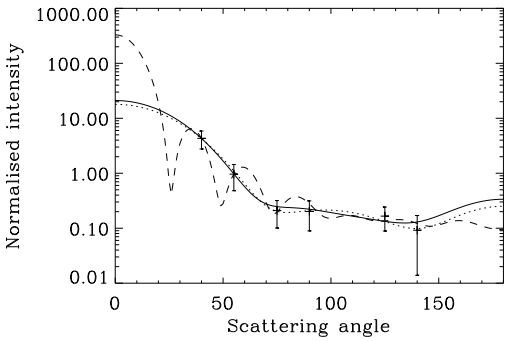


Figure 5.3: The original DAWN-A intensity measurements from Dick (1998) in their normalized form, as well the scattering patterns that result from the original estimates of the state (dotted line), the 2-dimensional retrieval (solid line) and the 3-dimensional retrieval (dashed line). The number assigned to each plot corresponds to the reference number given in Table (5.1).

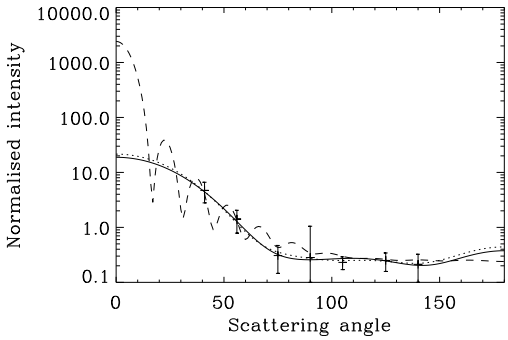
(6)


1000.00
100.00
10.00
1.00
0.10
0.01
Normalised intensity
0
50
100
150
Scattering angle

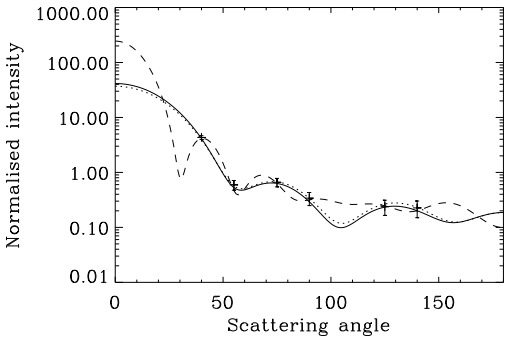
(7)


1000.00
100.00
10.00
1.00
0.10
0.01
Normalised intensity
0
50
100
150
Scattering angle

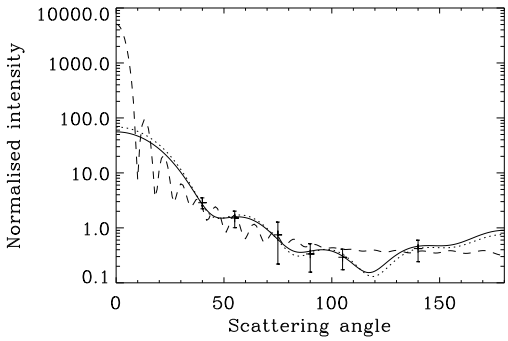
(8)


10000.0
1000.0
100.0
10.0
1.0
0.1
Normalised intensity
0
50
100
150
Scattering angle

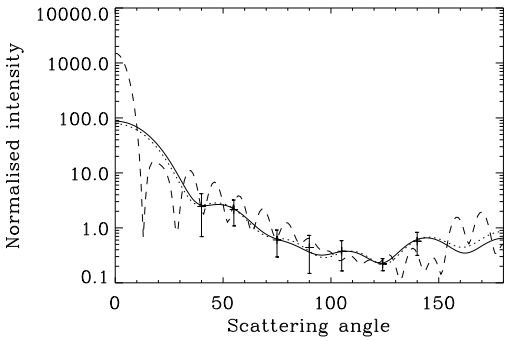
(9)


1000.00
100.00
10.00
1.00
0.10
0.01
Normalised intensity
0
50
100
150
Scattering angle

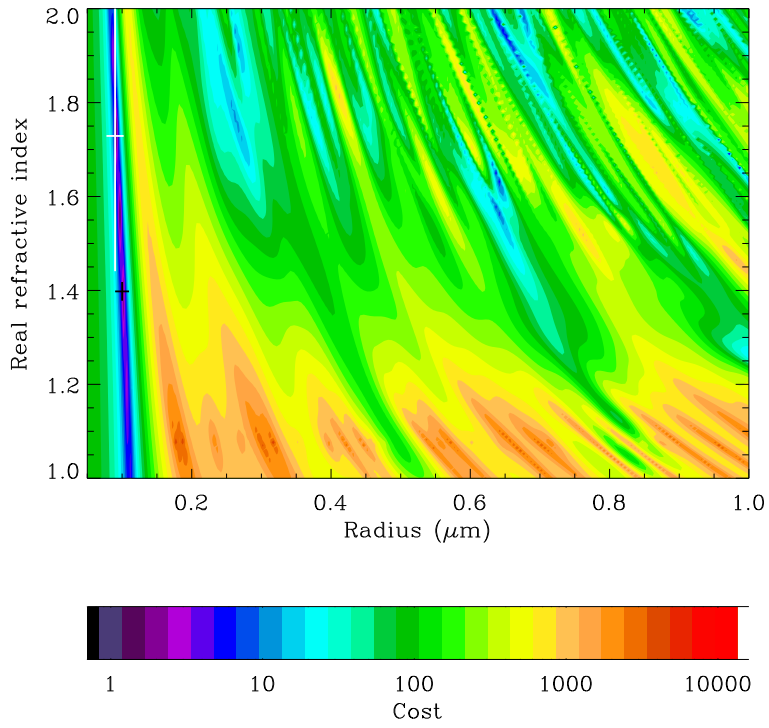
(10)


10000.0
1000.0
100.0
10.0
1.0
0.1
Normalised intensity
0
50
100
150
Scattering angle

(11)


10000.0
1000.0
100.0
10.0
1.0
0.1
Normalised intensity
0
50
100
150
Scattering angle

(a)



(b)

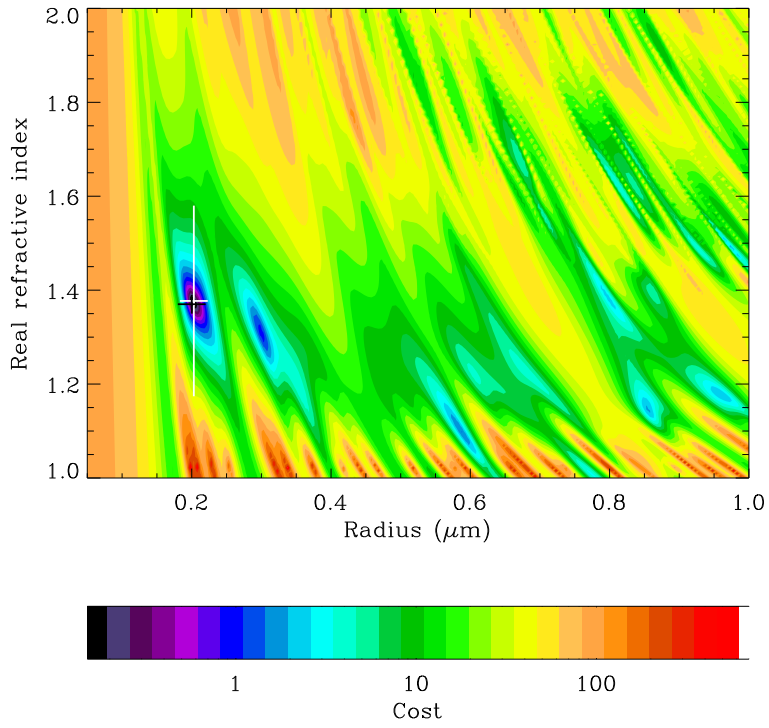
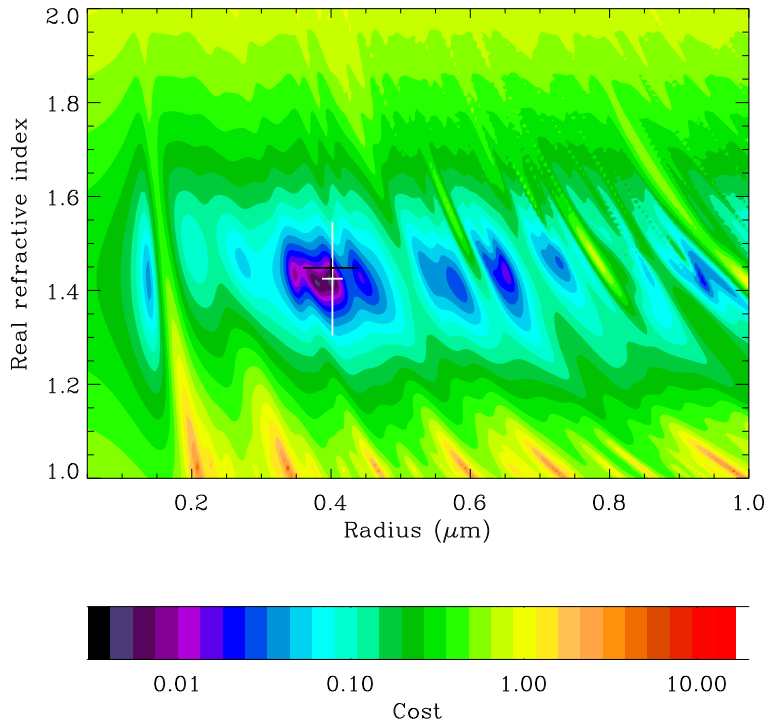


Figure 5.4: The cost function surfaces for retrievals (a): 1, (b): 5 and (c): 11 in the 2-dimensional, normalized case. Both the retrieved states (white) and the original estimates (black) are included with error bars.

(c)



mean measurements. The standard deviations given by Dick (1998) range from approximately 25 % to 75 % of the mean. As discussed in Appendix (B) the covariance of the solution is given by,

$$\hat{\mathrm{S}} = (\mathrm{S}_a^{-1} + \mathrm{K}^T \mathrm{S}_\epsilon^{-1} \mathrm{K}^T)^{-1}, \tag{5.1}$$

where S_a is the covariance of the *a priori*, S_ϵ is the covariance of the measurement and K is the weighting function matrix (basically the local gradient of the cost function). Hence large measurement error (resulting in a large measurement covariance) leads to large uncertainties in the solution. The difference in scale between the uncertainty in radius and refractive index can be attributed to the local gradient of the cost function near the solution – the minimum in the cost function is usually much better defined with respect to radius than refractive index when viewed at small scales.

5.2 Absolute measurement vectors

This section examines the results of applying the retrieval to measurements that have not been normalized. As noted previously, this method would be expected to be more robust than the normalized measurement vector method, simply because normalizing the measurement vector reduces the information it contains, particularly with respect to particle size. Figures (5.5) and (5.6) (c.f. Figures (5.1) and (5.2)) show that this is in fact the case. All but one of the 2-dimensional retrievals and three of the 3-dimensional show good agreement with the values given by Dick (1998) and, particularly in the 2-dimensional case, the uncertainties on the retrieved refractive index are much better than those found in Section (5.1). The χ_s^2 values for both 2-dimensional and 3-dimensional retrievals are of similar size apart from retrievals 4 and 8 in the 2-dimensional case, and retrievals 8 and 9 in the 3-dimensional case. In the latter case the larger χ_s^2 can be attributed to the inaccuracy of the retrieved state. There is no readily identifiable reason for the two large χ_s^2 values seen in the 2-dimensional case, however. Once again there is no apparent correlation between the values of the cost function at the solution and the accuracy of the retrieved state in either 2 or 3-dimensional retrievals.

Examining the scattering patterns that result from these retrievals shows that there is still scope for ambiguity in the 3-dimensional retrievals. As can be seen in Figure (5.6) retrievals 8, 9 and 11 all converged to much larger radii than was expected. Figure (5.7) shows that the total amount of light scattered by these large particles would indeed be much greater than for the expected particle. The intensity of the forward scattering peak is between one and two orders of magnitude higher for the retrieved particle in all three cases. Unfortunately, due the lack of measurements in the forward scatter direction, these profiles still fit the measurements well. Examining the patterns for 2-dimensional retrieval numbers 4 and 8 sheds no light on why their χ^2 values should be so high.

Figure (5.8) shows the 2-dimensional cost function surfaces for retrievals 1, 5 and 11 (cf. Figure (5.4)). Comparison with Figure (5.4) makes clear how the use of absolute measurement vectors improves the retrieval performance, particularly in the case of small particles. The minimum in the case of Figure (5.8a) is much more clearly defined than in (5.4a), and the resulting retrieval is consequently more satisfactory. The same is also true in the cases of Figures

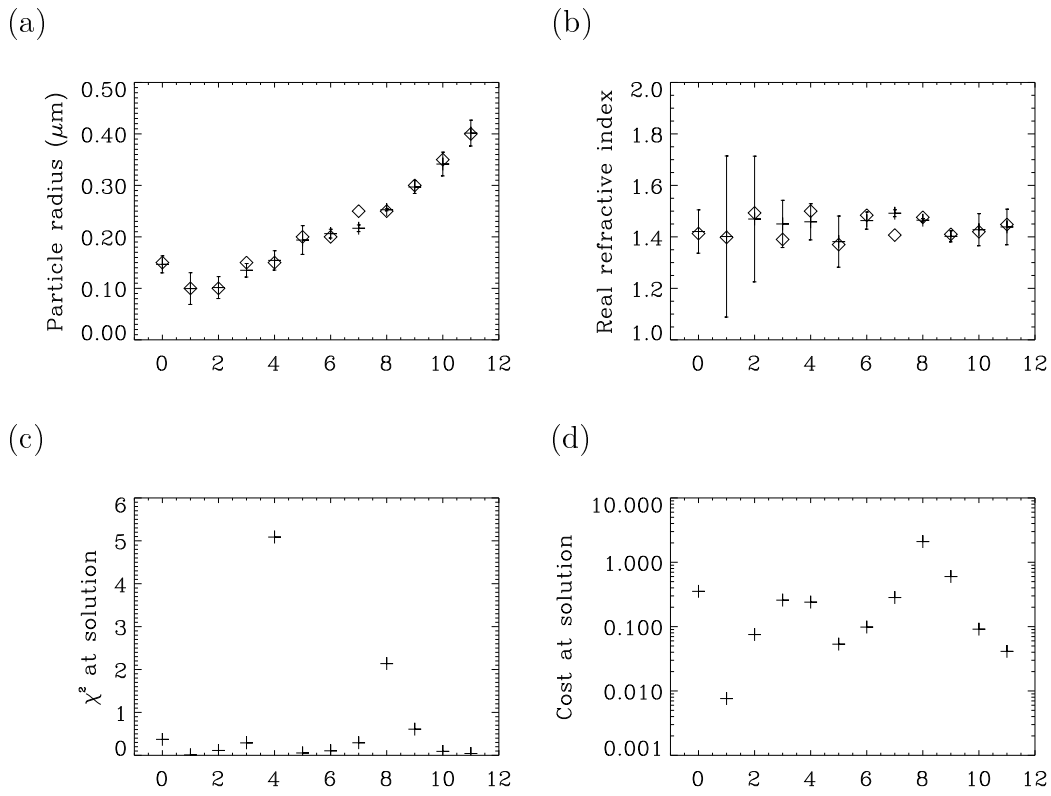


Figure 5.5: The results of applying the SPARCLE retrieval scheme to scattering patterns measured by the DAWN-A instrument in the 2-dimensional case (imaginary refractive index taken to be zero) using the absolute form of the measurement vector. The values quoted by Dick (1998) are given by the diamonds, with the new retrieved values displayed as crosses with 1σ error bars. Graph (a) shows the radius comparison, graph (b) the refractive index comparison. Graphs (c) and (d) show the χ_s^2 value and the cost function at the retrieved state, respectively. The numbers on the x-axes correspond to the reference numbers given in Table (5.1).

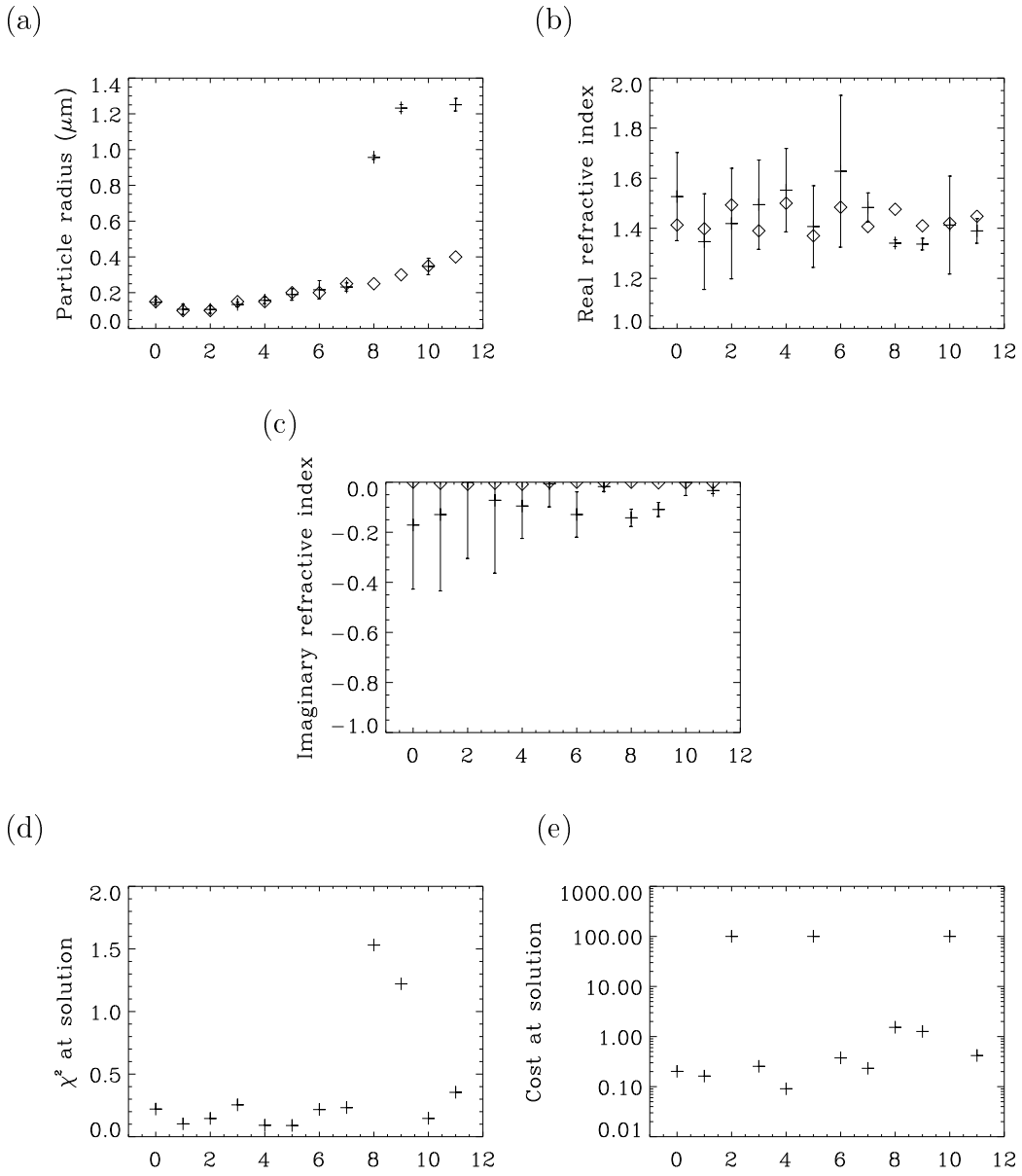


Figure 5.6: As for Figure (5.5) but in the 3-dimensional case. Graph (a) shows the radius comparison, graph (b) and (c) show the comparisons for the real and imaginary parts of the refractive index, respectively. Graphs (d) and (e) show the χ^2_s value and the cost function at the retrieved state, respectively.

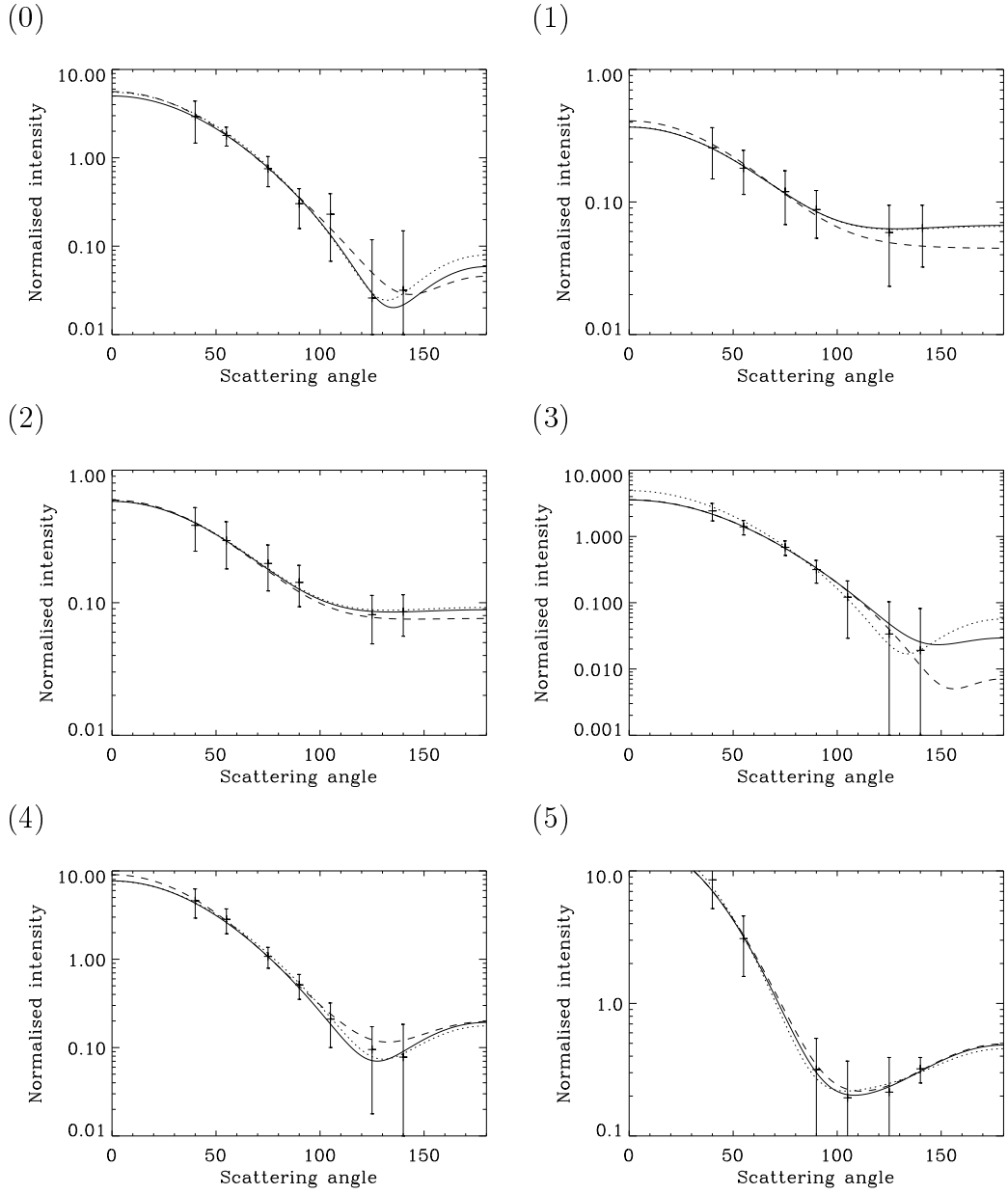
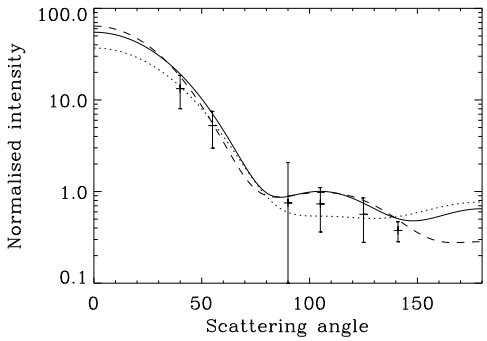
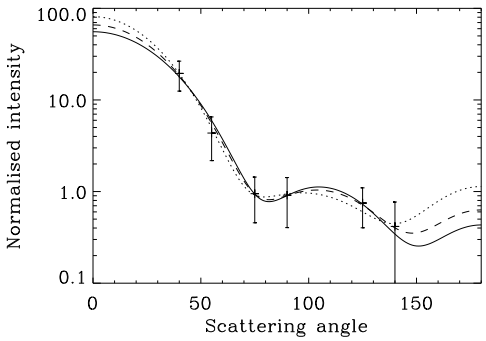


Figure 5.7: The original DAWN-A intensity measurements from Dick (1998) in their absolute form, as well as the scattering patterns that result from the original estimates of the state (dotted line), the 2-dimensional retrieval (solid line) and the 3-dimensional retrieval (dashed line). The number assigned to each plot corresponds to the reference number given in table (5).

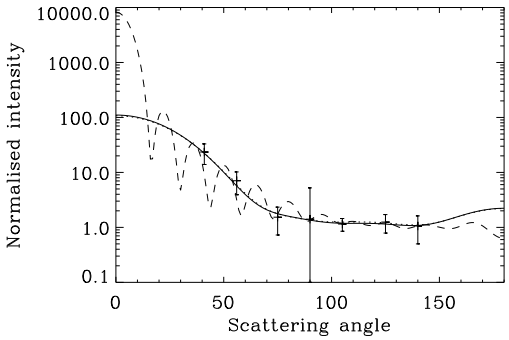
(6)


Normalised intensity
100.0
10.0
1.0
0.1
0
50
100
150
Scattering angle

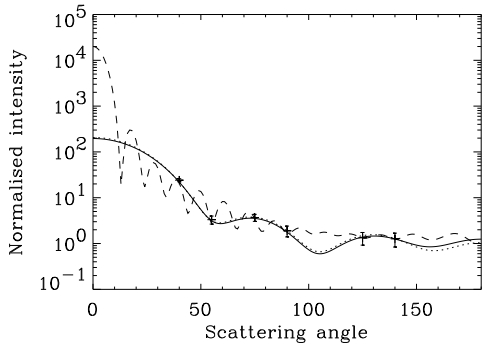
(7)


Normalised intensity
100.0
10.0
1.0
0.1
0
50
100
150
Scattering angle

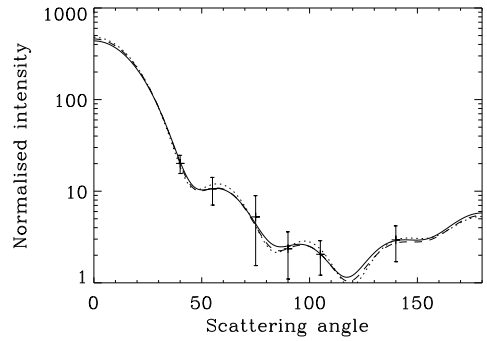
(8)


Normalised intensity
10000.0
1000.0
100.0
10.0
1.0
0.1
0
50
100
150
Scattering angle

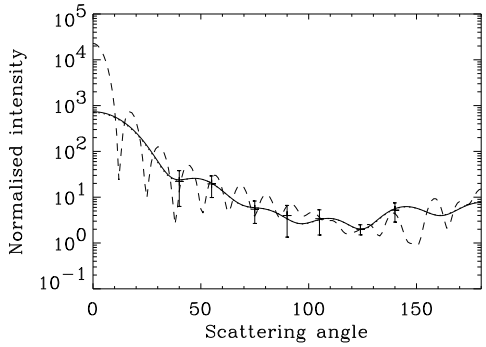
(9)


Normalised intensity
10^5
10^4
10^3
10^2
10^1
10^0
10^{-1}
0
50
100
150
Scattering angle

(10)


Normalised intensity
1000
100
10
1
0
50
100
150
Scattering angle

(11)


Normalised intensity
10^5
10^4
10^3
10^2
10^1
10^0
10^{-1}
0
50
100
150
Scattering angle

(5.8b,c), but to a lesser extent. In Figures (5.4b,c) there is at least one other minimum which is comparable in depth with the true minimum, whereas with absolute measurement vectors the secondary minima become much more shallow.

5.3 Summary of results from the DAWN-A retrievals

The analysis of the DAWN-A data has provided a useful check of the validity of the retrieval scheme developed for the SPARCLE instrument using independent data. Indeed, this analysis was the first application of the retrieval scheme to real measurements.

Four separate formulations of the retrieval problem were tested; the measurement vector was used in both normalized and absolute form and the state vector was retrieved in both 2-dimensional (particle radius and the real part of the refractive index) and 3-dimensional (particle radius, real and imaginary parts of the refractive index) forms.

The 2-dimensional case with absolute measurement vectors was found to be the most robust. In this case the output of the retrieval scheme agreed with the original DAWN-A results within uncertainties in all but one of the twelve cases. The 1σ uncertainties on the retrieved radius in this case were all ≤ 0.05 μm while those on the refractive index were ≤ 0.1 for particles with radius ≥ 0.15 μm. The worst-performing retrieval was the 3-dimensional case with normalized measurements. Here only two of the twelve retrievals agreed with the original results within uncertainties for all three retrieved quantities, and even where agreement was achieved the uncertainties on the refractive index values were extremely large.

These results suggest that the resolution at which the DAWN-A instrument samples the scattering pattern is not great enough to completely define the particle responsible for producing the pattern. It is, however, adequate for the retrieval of particle radius and the real part of the refractive index if the absorption of the particles (i.e. their imaginary refractive index) is known. Because DAWN-A instruments make a direct measurement of the particle size independently of the scattering pattern this retrieval is of limited use in

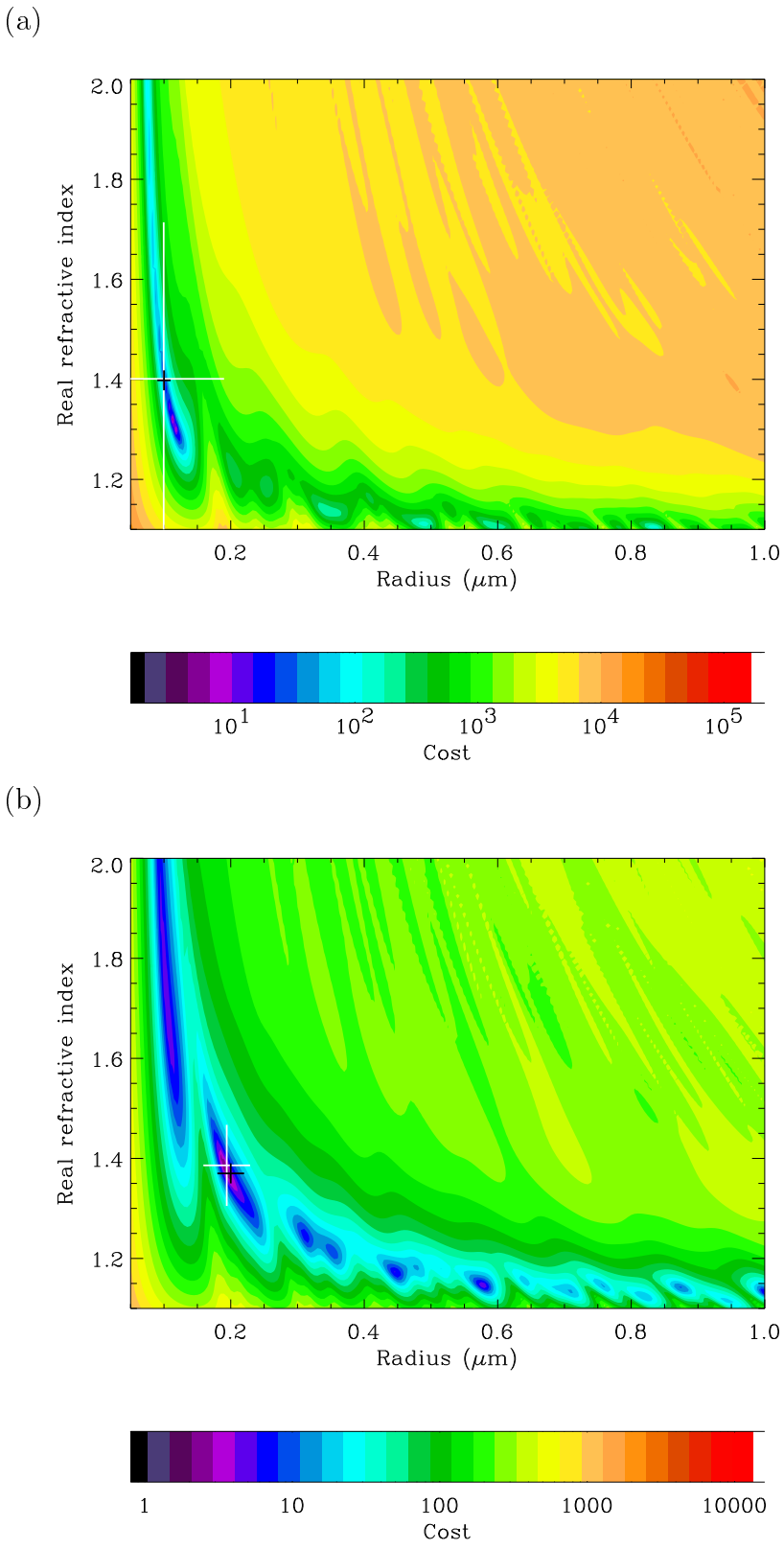
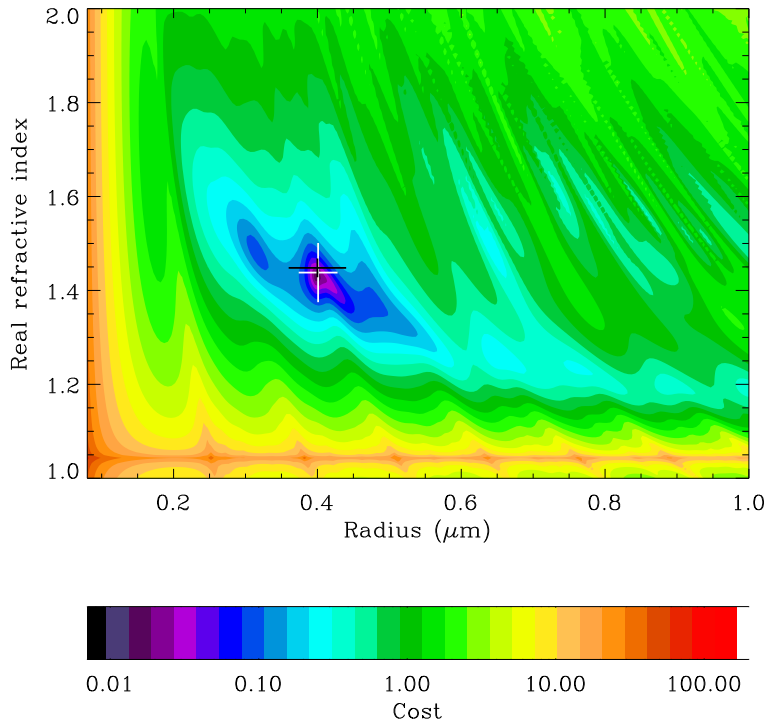


Figure 5.8: The cost function surfaces for retrievals (a): 1, (b): 5 and (c): 11 in the 2-dimensional, absolute measurement vector case. Both the retrieved states (white) and the original estimates (black) are included with error bars.

(c)



application to DAWN-A measurements. However, these results do suggest that a retrieval scheme similar to the one implemented here but designed to retrieve both parts of the refractive index for a known particle size might be used with DAWN-A measurements with some success. However, this work is outside the scope of the current study.

In terms of the SPARCLE experiment however, perhaps the most significant result of this analysis is the confirmation that the retrieval software produced can both accurately recreate a measurement of particle scattering and successfully retrieve particle properties from a real measurement.

Chapter 6

Instrument calibration and testing

This chapter describes the application of the SPARCLE prototype in measuring known, monodisperse aerosol populations. The experiments described here represent the first application of the SPARCLE instrument and provide valuable information on the strengths and shortcomings of the design.

Section (6.1) details the equipment and procedures used in the experiments. Section (6.2) details the analysis of independent particle size information measured with a Aerodynamic Particle Sizer. Section (6.3) describes the preparation of data collected by the SPARCLE instrument for presentation to the retrieval scheme. Details of the results of applying the retrieval to the measurements described here are given in Chapter (7).

6.1 Experimental setup

6.1.1 Aerosol generation and sizing equipment

To produce a known, mono-disperse aerosol for testing and calibration of SPARCLE two commercial instruments were used: a Model 3450 Vibrating Orifice Aerosol Generator (VOAG) and a Model 3320 Aerodynamic Particle Sizer (APS), both manufactured by TSI Incorporated. Detailed information, including complete specifications and operating manuals, on both products can be downloaded from the TSI web site: http://www.tsi.com

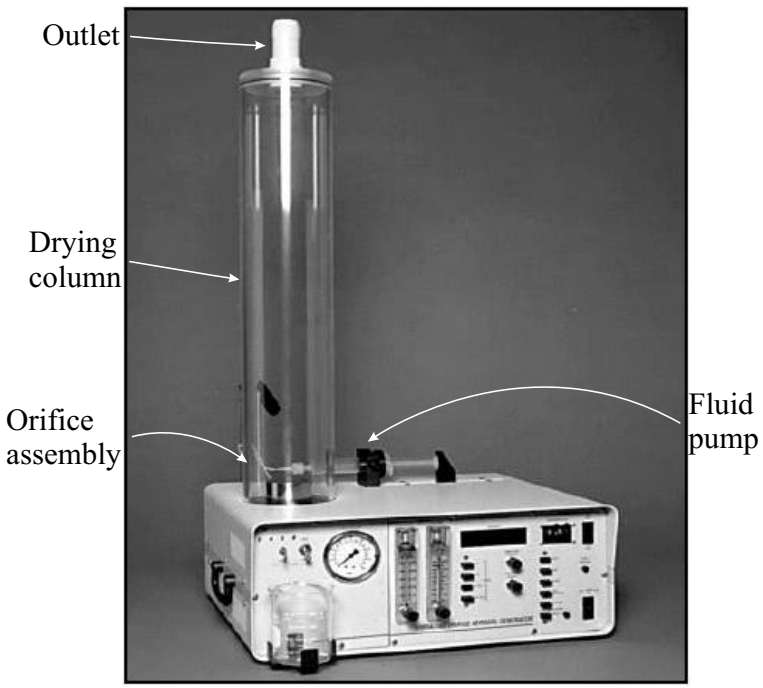


Figure 6.1: The TSI model 3450 VOAG.

The Model 3450 VOAG is shown in Figure (6.1). The instrument operates by forcing a solution from the fluid pump through a small circular orifice (20 μm in diameter in the case of this work). By vibrating this orifice at several kilohertz, the stream of fluid emerging through it can be broken into a series of uniform particles. Once produced, the particles travel up the drying column where they are diluted in a flow of clean air, before passing through the outlet. The particle size can be controlled by diluting the aerosol solution with a volatile liquid which evaporates from the particles in the drying column, resulting in a proportionate reduction in particle volume. The use of an evaporating solvent not only allows particles of different sizes to be produced but also makes it possible to produce solid particles from a solute.

The APS was used to provide an independent size measurement of the particles produced by the VOAG. Although the VOAG can produce a highly monodisperse aerosol, achieving this has been found to be highly dependent on the composition of the aerosol solution, the frequency of the orifice and the pressure exerted by the fluid pump. It was therefore impossible to predict exactly what size particles would result from a given solution, and difficult to create a

stable monodisperse aerosol, making the ability to independently measure the aerosol size vital.

The APS determines the aerodynamic size of particles by measuring their time of flight as they are accelerated by an air stream. As sample air approaches the measurement volume it is accelerated by an aerodynamic lens system. Any particles in the air stream are detected by an avalanche photodetector as they pass through a dual laser beam. The time of flight is measured between the two laser beams – the larger the particle, the lower the acceleration, and the longer the time of flight. The APS also provides a size estimate from the scattered light itself (i.e. using Mie theory), however this measurement is at a much lower size resolution and is only useful as an approximate check on the aerodynamic size.

The APS has a measurement size range between 0.5 and 20 μm which is divided into 52 logarithmically spaced size bins. The manufacturer claims that the instrument can measure a concentration of 1000 particles/cm^3 of 10 μm diameter particles with less than 6% coincidence. Table (6.1) lists the 52 size bins and their associated resolution. The APS can be entirely controlled by a PC via a serial interface using TSI proprietary software and most features are completely automated, making this instrument ideal for monitoring the aerosols generated by the VOAG.

The material used for aerosol generation in this work was a glycerol/propan-2-ol solution. This material has several key properties which make it ideal for this application:

- Glycerol has a refractive index similar to many aerosol particles found in the environment. The refractive index at 20 °C and 660 nm wavelength is $1.475 - 0.0i$ (Weast, 1988). The variation of this value across visible wavelengths is of the order of 0.01 in the real part[1].

- Propan-2-ol is an ideal solvent as glycerol is highly soluble in it, and it evaporates very rapidly once exposed to air at room temperature.

- Both chemicals are reasonably safe to handle. The only precautions that need to be taken are to avoid inhalation of the aerosol and to isolate aerosol and bulk forms of the chemicals from ignition sources.

[1]Reeves, R. J. Private communication. 2002.

Centre	Max	Centre	Max	Centre	Max	Centre	Max
< 0.523	0.523	1.286	1.334	3.278	3.398	8.354	8.660
0.542	0.562	1.382	1.433	3.523	3.652	8.977	9.306
0.583	0.604	1.486	1.540	3.786	3.924	9.647	10.000
0.626	0.649	1.596	1.655	4.068	4.217	10.37	10.746
0.673	0.698	1.715	1.778	4.371	4.532	11.14	11.548
0.723	0.750	1.843	1.911	4.698	4.870	11.97	12.409
0.777	0.806	1.981	2.054	5.048	5.233	12.86	13.334
0.835	0.866	2.129	2.207	5.425	5.623	13.82	14.330
0.898	0.931	2.288	2.371	5.829	6.043	14.86	15.399
0.965	1.000	2.458	2.548	6.264	6.494	15.96	16.548
1.037	1.075	2.642	2.738	6.732	6.978	17.15	17.783
1.114	1.155	2.839	2.943	7.234	7.499	18.43	19.110
1.197	1.241	3.051	3.162	7.774	8.058	19.81	20.535

Table 6.1: The aerodynamic particle diameter bins, in microns, used by the APS. The bin size increases logarithmically with increasing particle size and the 'centre' values (i.e. the particle diameter quoted for each bin) is the central value of each bin in log space. The 'max' columns list the maximum diameter for each bin, thus the bin with the centre value of 0.542 μm includes all particles between 0.523 and 0.562 μm.

- Both substances are also water soluble, which aids in clean-up.

6.1.2 Experiment configuration and measurement procedure

A schematic of the experimental setup can be seen in Figure (6.2). As can be seen the VOAG was connected to supply both SPARCLE and the APS with aerosol simultaneously. This was possible because the dilution flow introduced into the drying column was between 30 and 50 l/min, compared to the 1 l/min sample of the APS and the $\sim$ 100 ml/min flow used by SPARCLE. Therefore only $\sim$ 2% of the aerosol-laden air produced by the VOAG was actually required, the remainder was vented using an extractor fan system.

The sample delivery system was arranged so as to provide the most direct route for the aerosol to reach SPARCLE's measurement volume. This was done because of the small size of the inlet pipe and low flow rate made it likely that particle deposition would be of greatest concern in the SPARCLE air flow system.

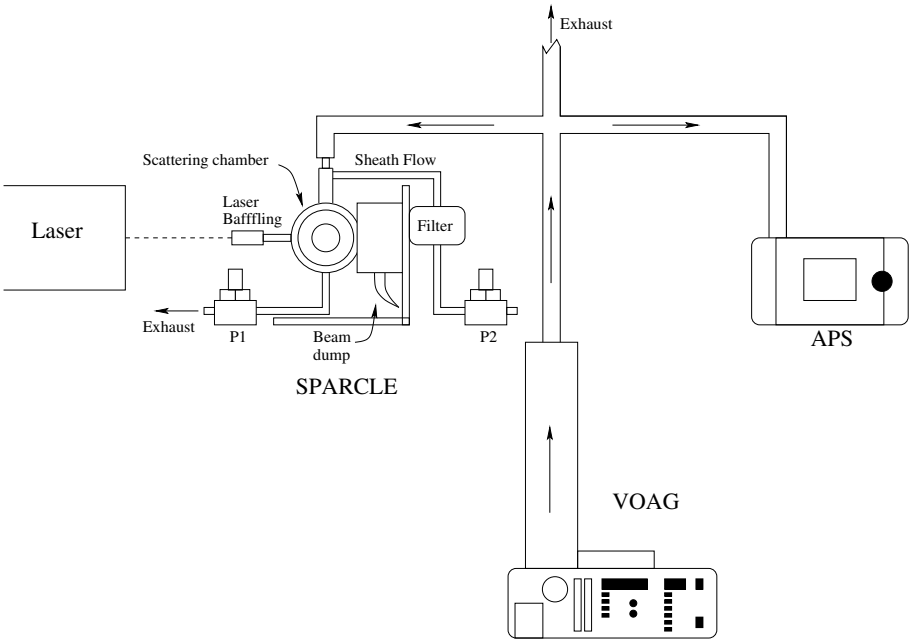


Figure 6.2: Schematic representation of the experimental setup. P1 and P2 refer to the SPARCLE air pumps.

The laser used as a light source for SPARCLE for these measurements was a continuous wave Coherent Ar laser, capable of providing light at a wavelength of 514.5 nm at powers of up to several Watts. In these experiments, the laser was run at approximately 70 mW and passed through the laser baffle shown in Appendix (C). With this arrangement the background stray light signal was just detectable across the entire LDA detector, thus any signal above the noise generated by this background was detectable. To accommodate this high background the PMT supply voltage was set at a relatively low -500 V. At this level the dynamic range of the PMT and it's ADC roughly matches that of the LDA system – i.e. any light pulse detected on the PMT will also be visible on the LDA.

The SPARCLE pump system was set to provide a sample flow of approximately 200 ml/min. This corresponds to the maximum outlet flow which can be provided by the outlet pump of approximately 550 ml/min. Figure (6.3) shows the flow rates measured for each of the measurements. Both the outlet and sheath inlet flows were measured before sampling started and after measurements were completed. The sample flow was not directly measurable during these experiments as the sample inlet was connected to the particle generation system.

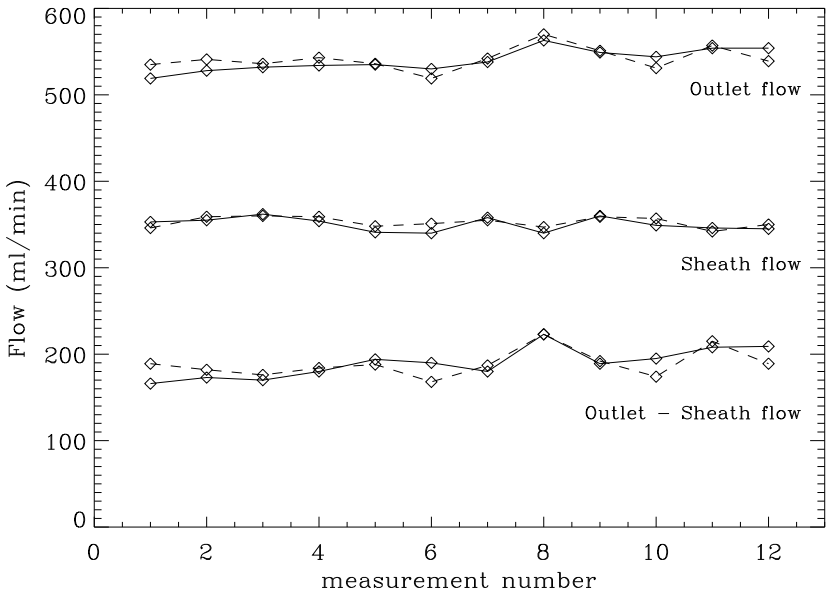


Figure 6.3: The measured flows for all experimental measurements. The solid line represents the flow at the start of each measurement run and the dashed line that at the end of run. Each point represents an average of four separate flow measurements taken consecutively. The variation of the flow across each four measurments was within the size of the plot symbols.

As can be seen in Figure (6.3), the flows remained reasonably stable throughout the approximately 3 weeks of measurements and were, for the most part, similar at the start and finish of each measurement run. The flow at the outlet minus the sheath flow can be viewed as a first approximation of the inlet flow, although it will be an overestimation due to air being drawn through the laser baffle and also to leakage from the pumps themselves (see Section (3.4)). Despite this however, it is likely that the true inlet flow would follow the outlet-sheath value, thus it is likely that the variation shown in the plot is representative of that in the true flow. Table (6.2) characterizes the variations of the flows across the 12 measurements as well as within the measurements. As can be seen the variation in sample flow can be expected to be of the order of 10 ml/min, or 5%, over the course of each sample.

Figure (6.4) shows the variation of laser power as seen by both the laser

Mean flow (ml/min)		Within measurement runs Mean change (ml/min)	Within measurement runs Max change (ml/min)	Across measurement runs Mean change (ml/min)	Across measurement runs Max change (ml/min)
Outlet	541	8.17 (1.5%)	16 (3.0%)	9.90 (1.8%)	29 (5.3%)
Sheath	352	5.33 (1.5%)	11 (3.2%)	6.08 (1.7%)	11 (3.3%)
Outlet-Sheath	189	10.7 (5.6%)	23 (12%)	12.1 (6.4%)	34 (18%)

Table 6.2: The mean and maximum flow variations seen during the measurement runs.

power meter included in SPARCLE and a commercial power meter. The laser power was monitored in the same way as the pump flows, i.e. readings were taken immediately before and after each measurement run. As with the pump flows there is noticeable variation, both between runs and within them.

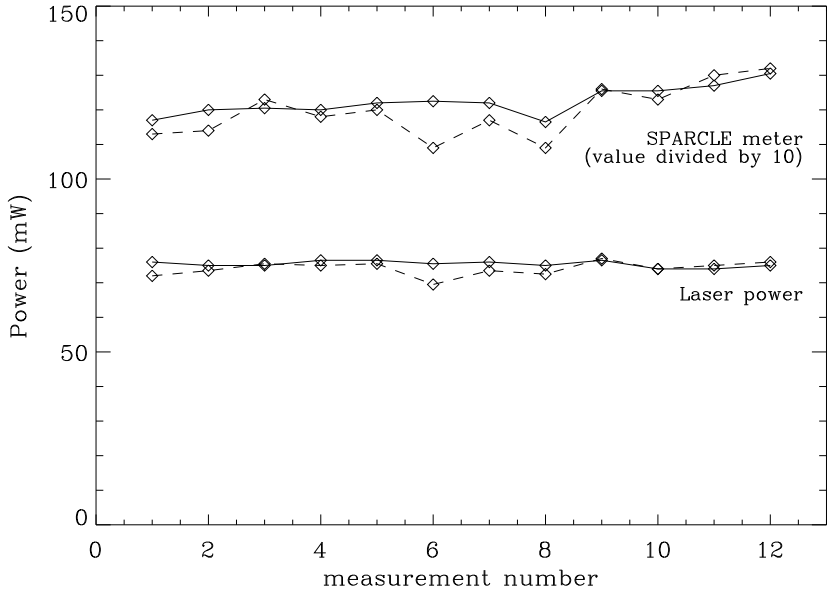


Figure 6.4: The laser power for all experimental measurements. The Solid line represents the power at the start of each measurement run and the dashed line that at the end of run. The units of the SPARCLE meters measurements are ADU/10.

During these experiments the laser entered SPARCLE horizontally as shown in Figure (6.2). In reality the laser was not simply aligned as shown in this figure: it was necessary to translate the beam vertically by $\sim$ 10 cm for it to

enter the baffle. This was achieved with two adjustable mirrors which were arranged to provide a horizontal beam at the correct height. Once the beam entered SPARCLE in the correct direction fine adjustment to the alignment was done to maximize the signal detected on the SPARCLE power meter. This fine adjustment was performed at the start of each measurement run, which offers a likely explanation for the gradual increase in laser power observed by the SPARCLE meter across the course of the measurements. It would appear that the repeated readjustment of the laser has resulted in general improvement of the alignment. It is important to realize here that a change in the the intensity measured by the SPARCLE meter doesn't necessarily imply this change will be seen at the sampling volume – due to the aperture in front of the detector it is possible to reduce the detected power without reducing the amount of light passing through the sample volume (see Chapter (3) and Appendix (C)).

The changes in both laser power readings during each measurement are due to variations in alignment and overall power output resulting from temperature effects. As the laser runs, heating causes changes in the laser cavity resulting in changes in the power output.

To alleviate the variations in both pump flow and laser power the system was given a run-in time of about 30 minutes, during which the worst of the temperature effects would occur. During this time the aerosol generation would be established and then the laser cavity tuning and beam alignment would be fine-tuned before measurements began. Because of the attention required in running the VOAG and collecting data using SPARCLE it was not possible to monitor either the pump flows or laser power directly. However a record of the background signal of the PMT was kept throughout the measurement run and this value should be fairly representative of the amount of light reaching the scattering volume.

Values of the PMT background taken at the start and end of each measurement run are displayed in Figure (6.5). These measurements resemble the variations in actual laser power, rather than the variation measured by the SPARCLE meter, suggesting that the slight change in alignment indicated by the latter has not affected the intensity at the scattering volume. Also, with the exception of measurement 8, all of the measurements agree with each other

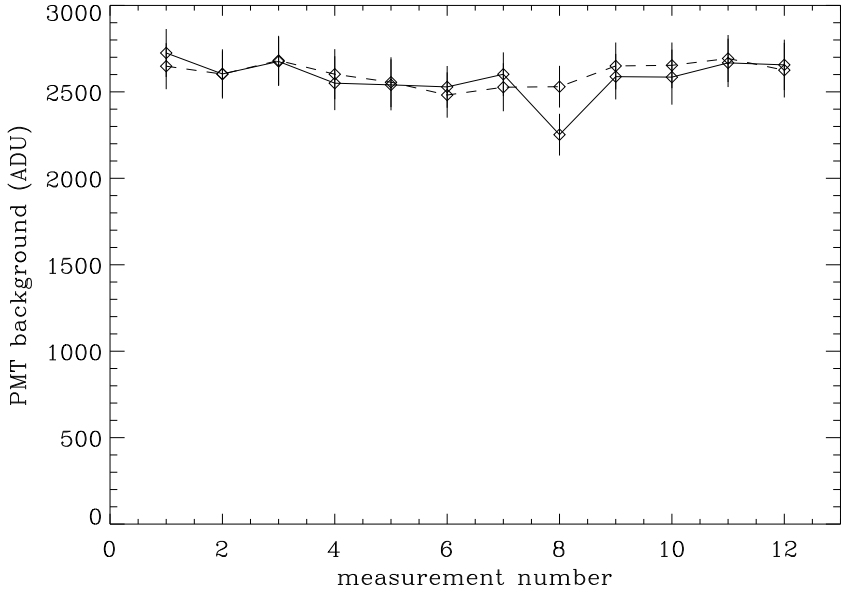


Figure 6.5: The background signal measured on the PMT at times close to the measurements shown in Figure (6.4). Each value corresponds to the mean of 1000 PMT samples (taking $\sim$ 14 ms) and the error bars show 1 standard deviation. Again, the solid line represents the signal at the start of each run, the dashed line is that at the end.

(both within measurement runs and between them) to within one standard deviation. Thus it is likely that the change in laser intensity has not affected the measurements any more than random noise.

For each of the measurement runs, data was collected by the APS throughout the experiment. The APS was controlled by a PC running TSI's proprietary Aerosol Instrument Manager Software under Microsoft Windows. The software allowed for continuous sampling with an integration time of 20 s. The cumulative distribution was also displayed in real time, thus enabling the aerosol being generated by the VOAG to be monitored during measurement. When a stable, monodisperse aerosol was available data collection from SPARCLE was begun. SPARCLE data was collected in samples of 100 particles and each measurement run consisted of between 10 and 15 successful samples. If the aerosol underwent a significant change in size and/or distribution during a sample, then that sample was discarded. Measurement would then be suspended while the aerosol generator was readjusted to produce the desired particles again. Figure (6.6)

shows an example of the APS data collected during a measurement run – in this case run 12 which was of nominally 6.9 μm particles. Points at which the particle generator ceased producing the required aerosol can be seen at 15 and 21 minutes into the run. While the desired distribution was being produced, the concentration of particles with radius near 6.9 μm varies between $\sim$ 1.25 and 1.75 cm^{-3}. The total particle count was between 2.0 and 2.5 cm^{-3}. The difference between these two values is mostly due to particles with radii between $\sim$ 0.25 and 3 μm, with an exponentially decreasing concentration with size.

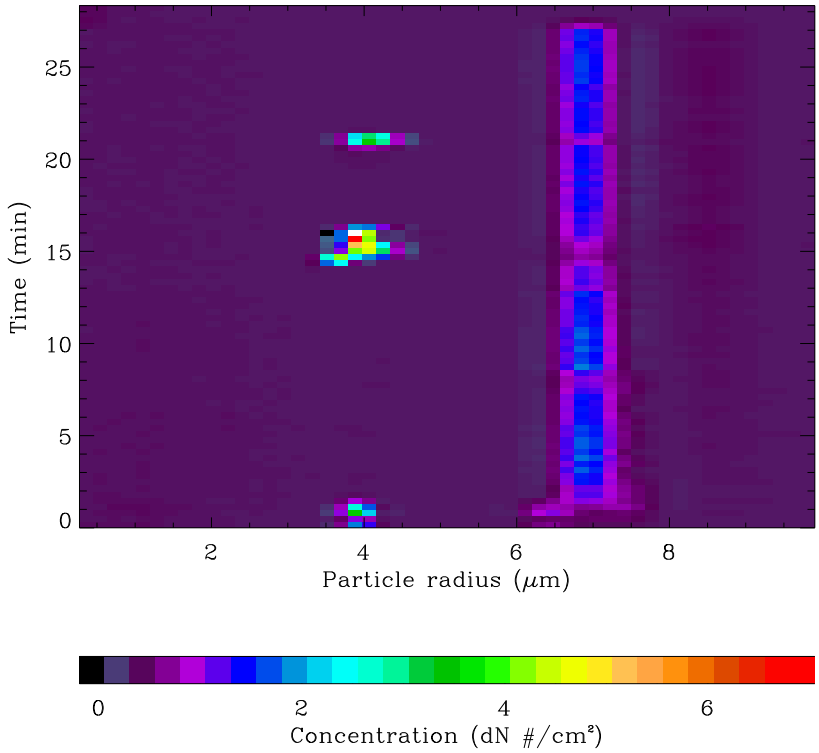


Figure 6.6: The data collected from the APS during the course of measurement number 12. The distribution has been interpolated onto a linear size scale (rather than the native logarithmic scale used by the APS).

When ordered to collect a sample, the SPARCLE control software first displayed a background PMT signal and its associated standard deviation and then requested a series of 100 measurements from the instruments dedicated controller circuitry. For these experiments the SPARCLE data acquisition was set up to provide 250 photomultiplier samples and 3 complete profiles from the

LDA (see Appendix (C) for a detailed description of this hardware). Once the data had been successfully collected it was stored in its raw form on disk. On average, a measurement run would span an hour and a half, approximately half of which would actually involve data collection from SPARCLE.

6.2 APS data

To provide a well-defined picture of the aerosol measured by SPARCLE, the APS measurements which were coincident with a SPARCLE sample were averaged over each run. This provided an average particle size distribution for each measurement run which could be compared with the distributions produced by the SPARCLE measurements. These average distributions are shown in Figure (6.7), and related statistics are given in Table (6.3).

The most important thing to note about these distributions is that the main limiting factor in characterizing them is the low size resolution of the APS. The nominal radii correspond to the centre value of the APS size bin with the most data (which in all cases corresponds to the median of the distribution). Since the width of each bin is a significant proportion of the width of all 12 distributions, the mean and mean absolute deviation (MAD) values are misleading. A better indicator of the spread of the distributions and the uncertainty in their central value is the width of the distribution, taken as the difference between the maximum radius value of the largest bin and the minimum radius value of the smallest bin. In the case of these measurements this width has been calculated at 20% of the peak value of the distribution.

The distributions also show the VOAG's tendency to produce particles of certain sizes. For example the distributions shown in Figures (6.7d,e,f) were all produced with differing concentrations of aerosol fluid and yet all resulted in very similar particles being produced. A few of the distributions also clearly show small secondary size modes, for example, Figure (6.7a) shows a peak at 4.2 μm as well as the dominant peak at 3.4 μm. These secondary peaks are actually present in all of the distributions (although often they are very small) and could not be entirely eliminated from the VOAG-produced aerosol.

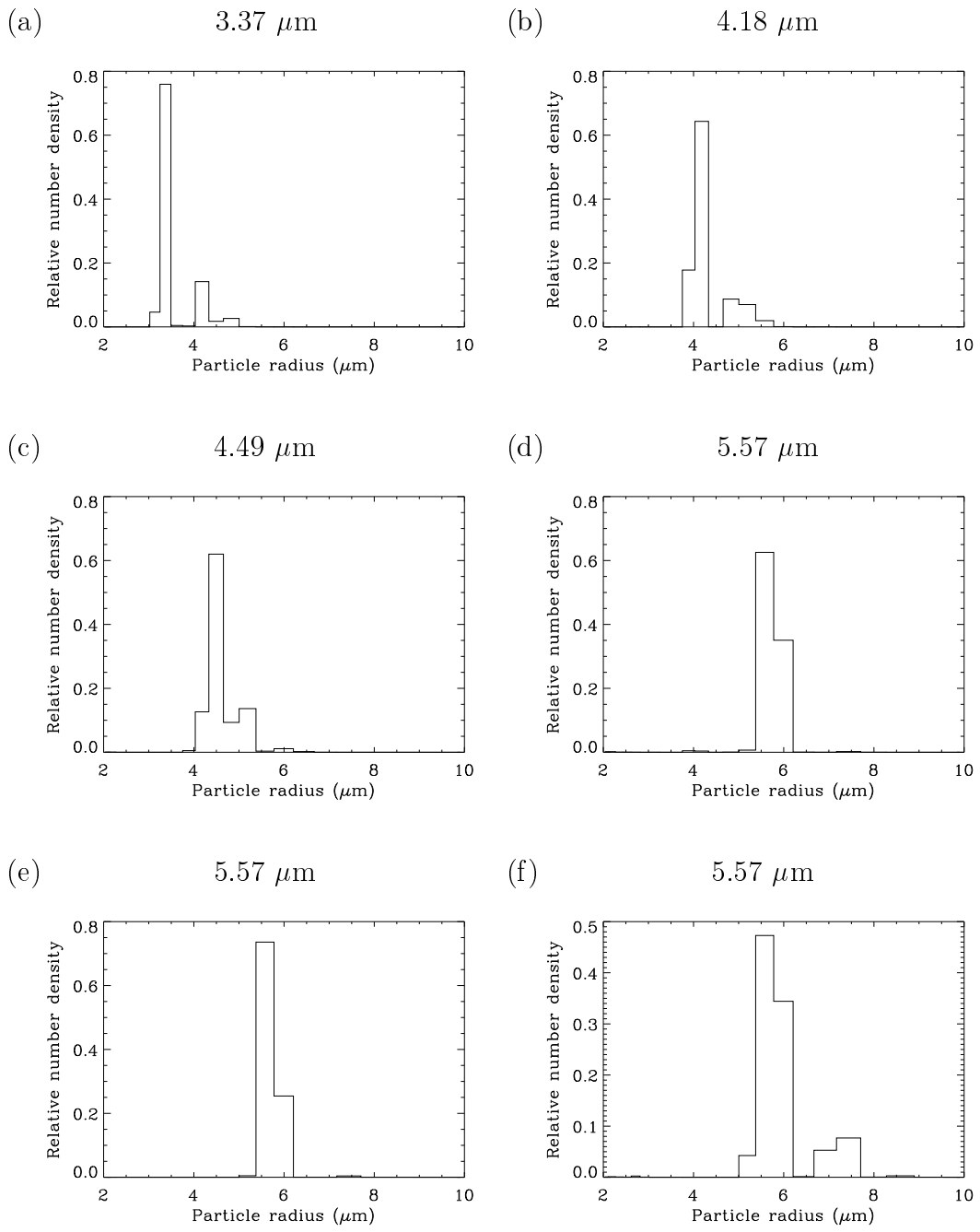
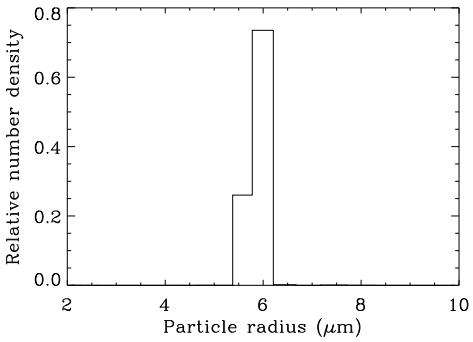
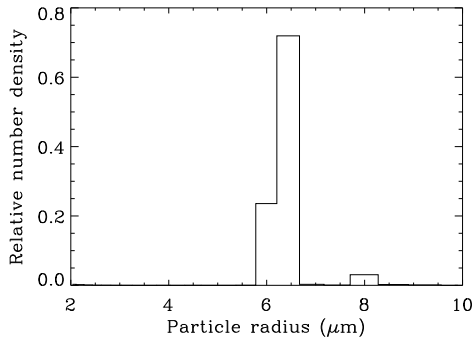


Figure 6.7: The averaged particle size distribution for each run as measured by the APS, in order of increasing particle size. The nominal radius is quoted for each distribution.

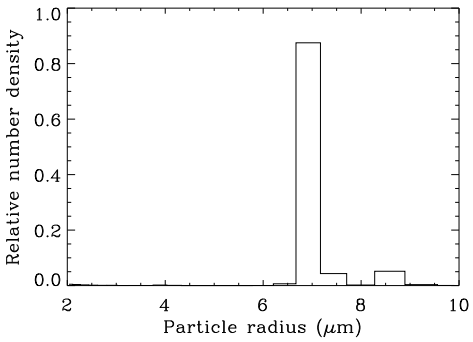
(g) 5.99 μm



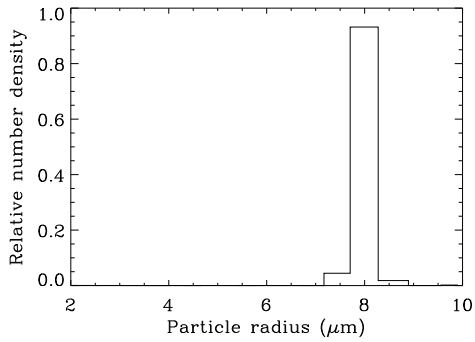
(h) 6.43 μm



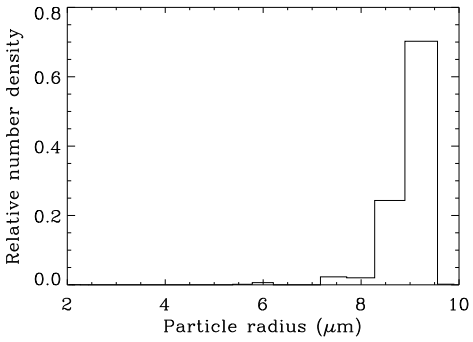
(i) 6.91 μm



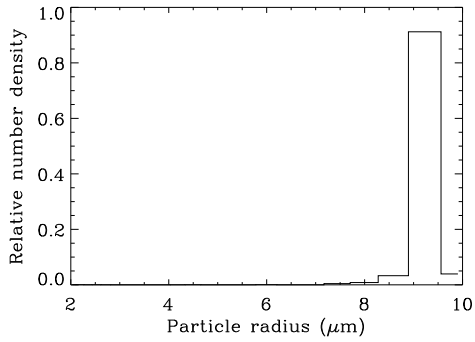
(j) 7.98 μm



(k) 9.22 μm



(l) 9.22 μm



	Radius (μm)					
	Nominal	Median	Mean	σ	MAD	$\frac{1}{2}$ width
(a)	3.37	3.37	3.53	0.39	0.29	0.24
(b)	4.18	4.18	4.28	0.40	0.28	0.58
(c)	4.49	4.49	4.60	0.36	0.25	1.34
(d)	5.57	5.57	5.69	0.37	0.22	0.83
(e)	5.57	5.57	5.68	0.23	0.17	0.83
(f)	5.57	5.57	5.90	0.64	0.42	0.83
(g)	5.99	5.99	5.88	0.23	0.17	0.83
(h)	6.43	6.43	6.36	0.51	0.23	0.89
(i)	6.91	6.91	6.95	0.71	0.23	0.50
(j)	7.98	7.98	7.95	0.34	0.08	0.58
(k)	9.22	9.22	8.96	0.54	0.36	1.28
(l)	9.22	9.22	9.19	0.39	0.11	0.66

Table 6.3: Statistical properties of the APS size distributions shown in Figure (6.7). The nominal radius corresponds to the mode of the distributions. The statistics exclude particles with a radius less than 2 μm, as these particles would be excluded from the SPARCLE measurements by the PMT threshold level.

6.3 SPARCLE measurements

The first task in preparing the data was to remove any spurious measurements. The main cause of a 'false' measurement in SPARCLE is a coincidence event, i.e. they result from two or more particles passing through the measurement volume either at the same time, or very shortly after one another. Such events can have one, or a combination, of three results, depending on the gap between the particles and the number of particles involved:

1. The light pulse detected can be distorted and lengthened.
2. Particle scattering can be recorded on all three LDA sweeps.
3. The acquisition system can be triggered by a particle before all of the LDA data from the previous measurement have been transferred to the PC, resulting in corrupted LDA data.

To remove such data from those to be passed onto the retrieval the following filtering scheme is applied to the data:

1. A data profile from the PMT must start below some threshold for background signal.

2. The PMT signal must have returned to background levels (i.e. similar to the level seen at the beginning of the profile) by the end of the 250-sample profile.

3. The data from the LDA must be below a reasonable level (a very high and constant reading from the LDA indicates corrupted data).

4. Once combined into one profile the LDA measurement must not start below zero (a situation which indicates scattering data on the 3rd raw LDA profile – see Section (6.3.2)).

Table (6.4) shows the nominal radius determined from APS measurements as well as the total number of SPARCLE measurements taken and the number which passed the above criteria. On average, approximately 1/3 of the raw measurements are rejected because of coincidence errors, and that figure is almost 1/2 for some measurement runs (those for which the particle concentration produced by the VOAG was particularly high). It is quite likely that not all measurements rejected by the above criteria are coincidence events; however, it was decided for the purpose of this work that it was best that some reasonable measurements be discarded rather than allowing spurious ones to contaminate the retrieval runs.

The number of measurements rejected due to coincidence errors is very large. The main reason for this is the long 'dead time' of the SPARCLE data acquisition hardware – the several milliseconds it takes for data transfer is unacceptably long for the purposes of counting the number of particles passing through the instrument. Thus the SPARCLE instrument, as used in this work, is not suitable for measuring aerosol concentration and no analysis of the counting statistics has been performed.

The two major problems faced in using the data accepted by the criteria discussed above for retrieval purposes were: the very low signal to noise ratio of the LDA data at larger scattering angles, and the degeneracy of the PMT signal for a given particle size. Methods for dealing with these problems are detailed in the following two sections.

Measurement number	letter	Nominal radius μm	Total number of measurements	Number of non-spurious measurements
1	(j)	8.0	1100	733 (66.6%)
2	(g)	5.9	1100	830 (75.4%)
3	(a)	3.4	1100	629 (57.2%)
4	(b)	4.2	1100	795 (72.3%)
5	(c)	4.5	1100	780 (70.9%)
6	(k)	9.1	1500	863 (57.5%)
7	(d)	5.7	1200	904 (75.3%)
8	(l)	9.2	1100	920 (83.6%)
9	(e)	5.7	1200	737 (61.4%)
10	(f)	5.7	1300	803 (61.8%)
11	(h)	6.4	1400	842 (60.1%)
12	(i)	6.9	1300	841 (64.7%)

Table 6.4: The nominal particle size (as measured by the APS) and number of SPARCLE measurements made for each measurement run. The reference letters used to denote the measurement run in order of ascending particle size are also included.

6.3.1 PMT data analysis

The problem with the PMT measurements arises because of the Gaussian profile of the laser beam and the fact that it is smaller than the beam of sample air. The Gaussian laser beam profile means that a cross-section through the edge of the beam will look identical to one through the centre of the beam, apart from being lower in overall intensity. Because the sample air stream is larger than the laser beam, particles can pass through any point across the beam and hence identical particles can produce an infinite number of different measurement intensities. Thus it is impossible to distinguish between a large particle passing through the edge of the beam and a smaller one passing closer to the centre.

The solution applied to this problem is to use the PMT measurements for each measurement run as an ensemble. If we assume all the particles being detected are the same, the amplitude of each PMT profile can be viewed as a random, uniform sample from a truncated Gaussian. The truncation is introduced by the detection threshold in the PMT data acquisition level. The amplitude distribution of such a sample can be approximated by the inverse of the gradient from the cut-off to the near maximum of the curve,

since the likelihood of picking a given amplitude is inversely proportional to how rapidly the amplitude is changing at that point. An example of this curve for a Gaussian which has a cut off at 5% of its maximum amplitude is given in Figure (6.8), along with a histogram of the amplitude distribution of 20,000 random samples from such a function. Obviously the inverse gradient approximation fails near the maximum amplitude, where the gradient approaches zero. The key feature of this curve is that there is a very sharp and well defined peak for values near the maximum and, naturally, an absolute cut off above the maximum. Increasing the height at which the Gaussian is truncated simply moves the left hand cut off of the curve shown in Figure (6.8) further to the right. Thus, a Gaussian cut off at an amplitude greater than approximately 30% of it's maximum height will only produce the right hand peak.

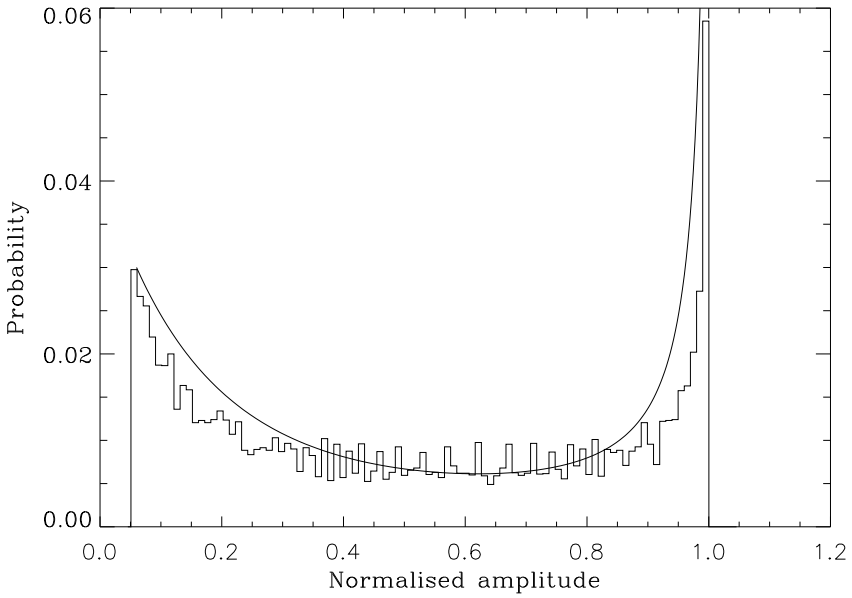


Figure 6.8: The inverse gradient approximation (smooth curve) to the amplitude distribution of uniform random samples taken from a Gaussian with a cut off at 5% of its maximum amplitude and the actual distribution of 20,000 such samples (histogram).

The distribution of values determined from PMT measurements can not be expected to follow the profile shown in Figure (6.8) exactly. This is because of noise on the PMT measurements, the fact that we have a distribution of

particle sizes, deviations of the laser beam profile from strictly Gaussian, and bias introduced by the circular cross section of the sample air flow. Figure (6.9) shows that the PMT signal distributions do follow the general pattern of Figure (6.8) however. Many of these distributions show evidence of secondary peaks (for instance Figure (6.9d,g)) which are probably due to the presence of a secondary particle size peak, or possibly some distortion of the laser beam profile from the expected Gaussian shape. The other major difference from the theoretical prediction is the tail found on the high signal side of the distributions – the width of this tail gives a good indication of the spread of the PMT signals due to spread in the particle size and noise from the PMT.

The peaks of these distributions are used to define the PMT signal for each measurement run, with the width of the high signal tail defining the associated uncertainty. This uncertainty is defined by fitting 1/2 of a Gaussian to the tail above the peak intensity – the error being the standard deviation defined by the fitted curve. Figure (6.9) shows the location of the maximum in each distribution and gives the value taken as the PMT signal. These values are plotted in Figure (6.10) with a scaled curve showing the output of the PMT forward model. The error bars show the standard deviation associated with each distribution and the half-width of the particle size distribution measured by the APS. Here we can see that it is possible to attribute the uncertainty in the averaged PMT measurements almost entirely to the rapid fluctuations expected in the PMT signal with increasing particle size. This agreement suggests that the method described above is a valid way of presenting a PMT signal which is suitable for use in the retrieval problem. It is important to note, however, that this method is only useful for measurements of a monodisperse aerosol.

6.3.2 LDA data analysis

As mentioned in Section (6.1), each LDA measurement consisted of three successive LDA profiles. All being well, the particle scattering pattern is found on the first two profiles (it can be on the first, second or both depending on the exact timing of the particle's flight through the laser beam) and the final profile shows the background signal. An example of a raw LDA measurement is shown in Figure (6.11). This is an example where the particle scattering pattern is present on both of the first two LDA profiles. Here the particle has entered

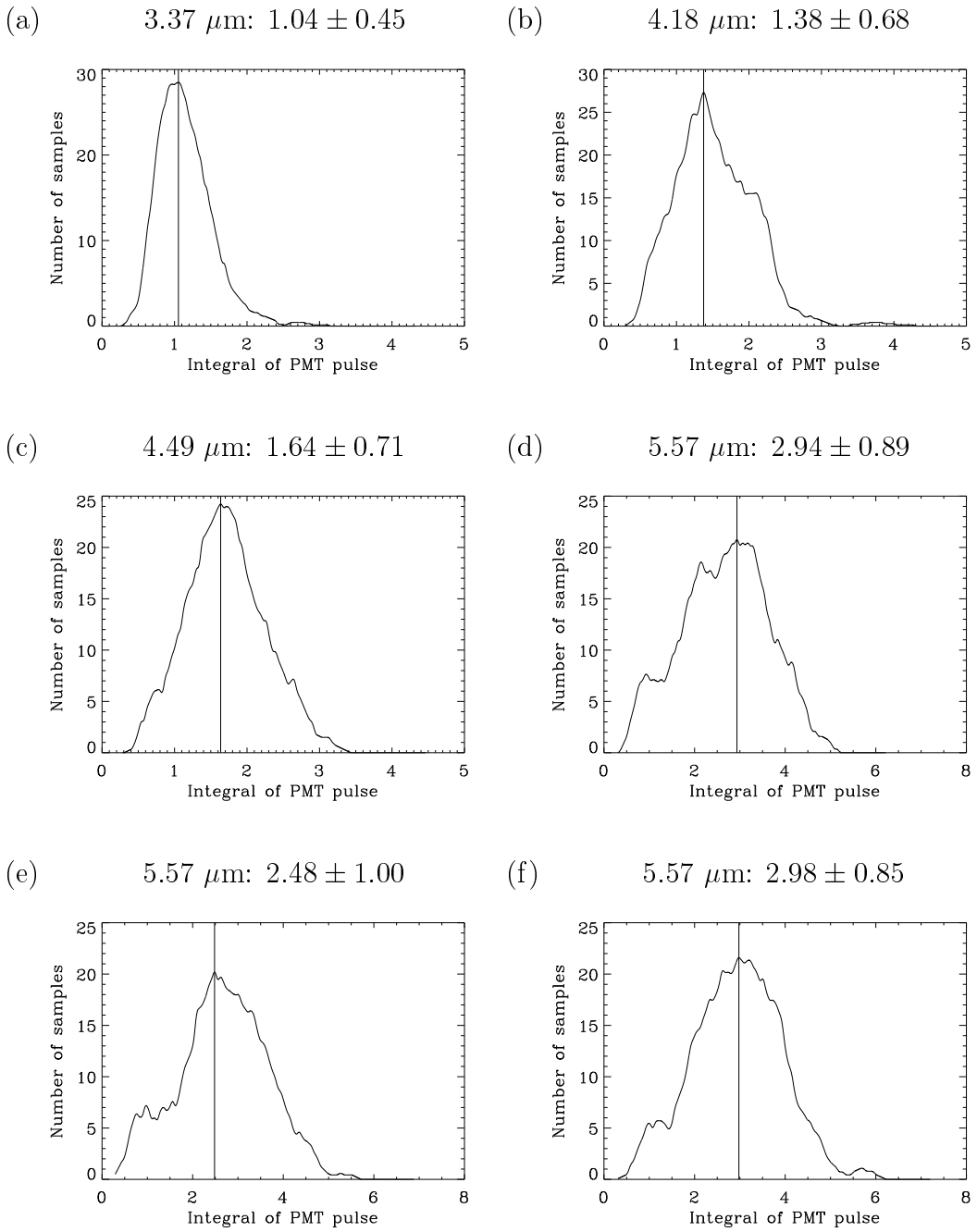
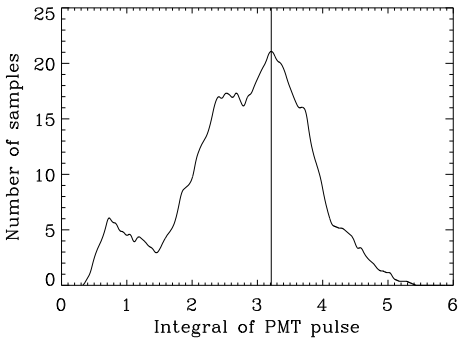
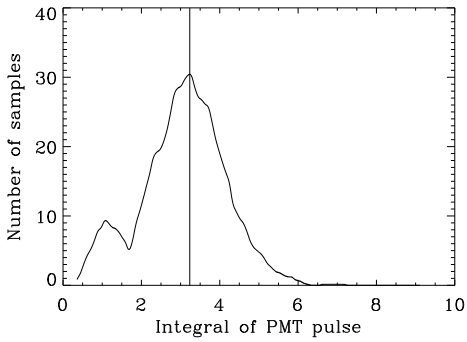


Figure 6.9: The distribution of the PMT signals for each measurement run. The distributions have been smoothed with a ~ 0.5 wide (in PMT integral units) boxcar averaging filter. The peak value of each distribution is given along with its associated standard deviation.

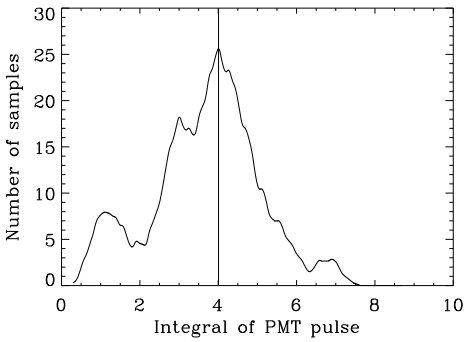
(g) 5.99 μm: 3.21 ± 0.72



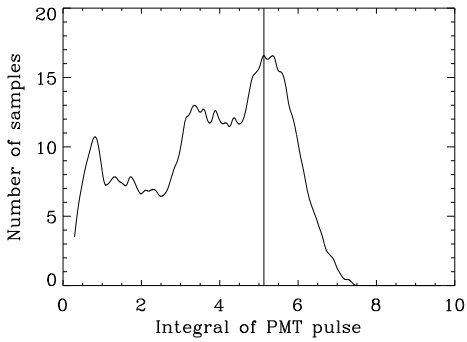
(h) 6.43 μm: 3.24 ± 0.99



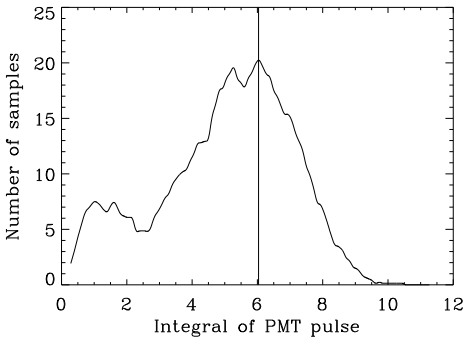
(i) 6.91 μm: 4.01 ± 1.33



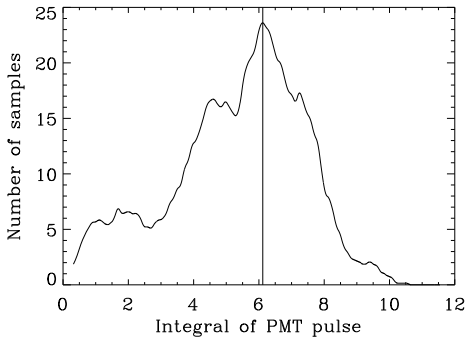
(j) 7.98 μm: 5.31 ± 0.76



(k) 9.22 μm: 6.04 ± 1.35



(l) 9.22 μm: 6.12 ± 1.41



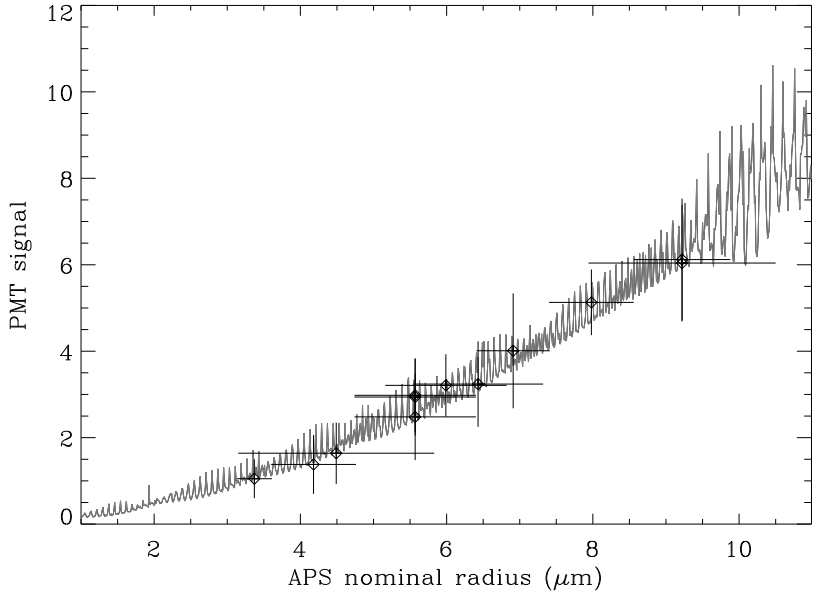


Figure 6.10: The PMT signal for each measurement run compared to the output of the forward model. The PMT measurements include the error estimates for the PMT signal and APS measured radii. The model output has been uniformly scaled to fit the data.

the beam while the LDA was transferring data off the sensitive pixels, resulting in part of the profile being visible on two profiles. The first profile shows a scattering signal on the last pixels to be transferred to the readout buffer (i.e. those in the forward scatter direction), while the second profile shows signal on those pixels which had already been transferred at the time at which the particle entered the laser beam. The pixels which were transferred onto the readout buffer at the time of the particle transit show a scattering signal on both the first and second profiles. A single LDA measurement can be constructed from a measurement like this by adding the first two profiles together and then subtracting twice the background signal observed on the third profile. The scattering profile resulting from applying this procedure to the measurement shown in Figure (6.11) is presented in Figure (6.12).

Figure (6.13) shows the same measurement again, this time in the form used by the retrieval (as described in Chapter (4)), along with the corresponding forward model output based on the APS measured nominal size for this particle. As can be seen in Figure (6.13b), the agreement between the theory

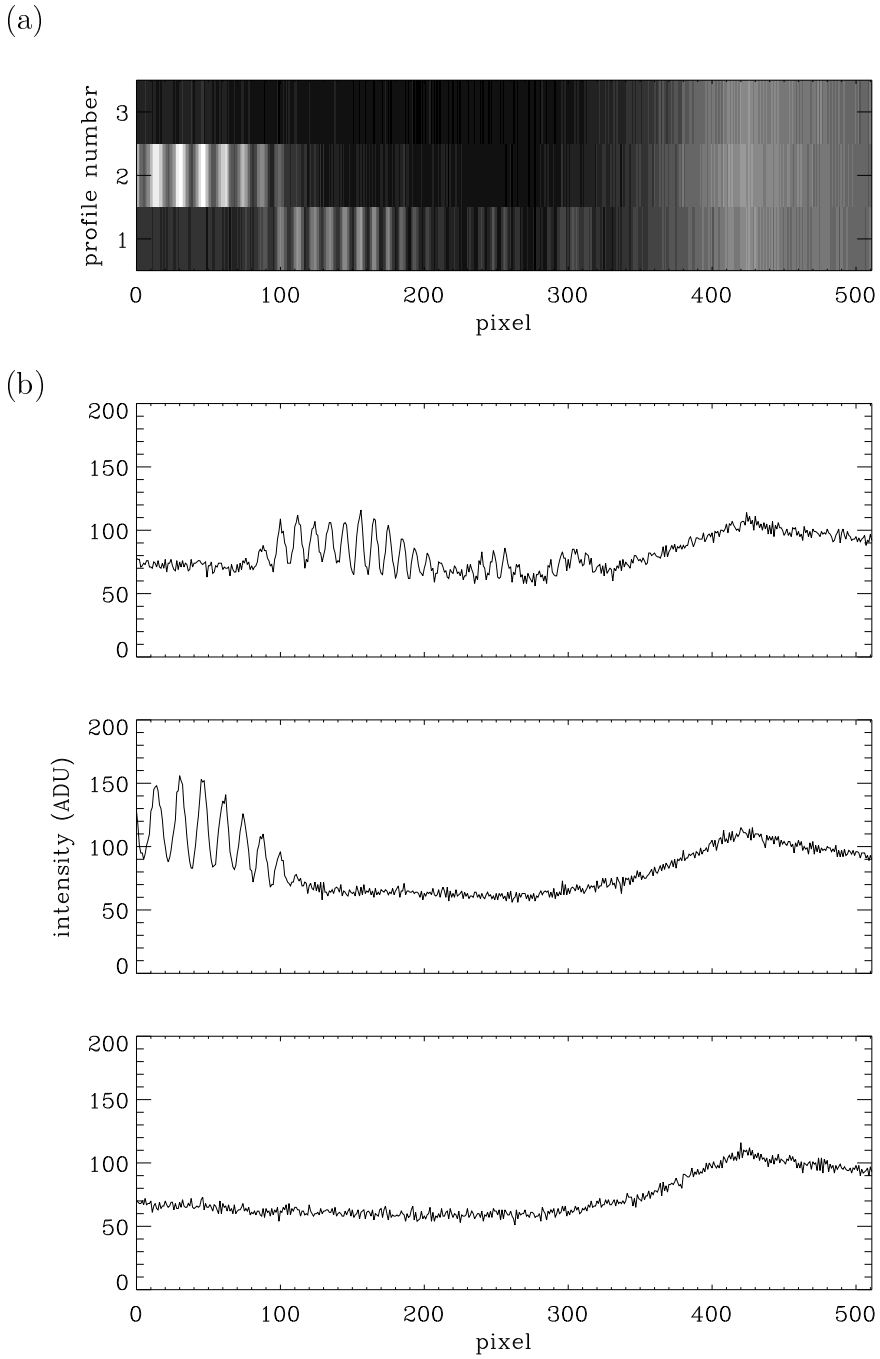


Figure 6.11: An example of an LDA reading in its raw form. (a) An image plot of the data, with the first profile at the bottom. (b) Plots of the three profiles, with the first at the top.

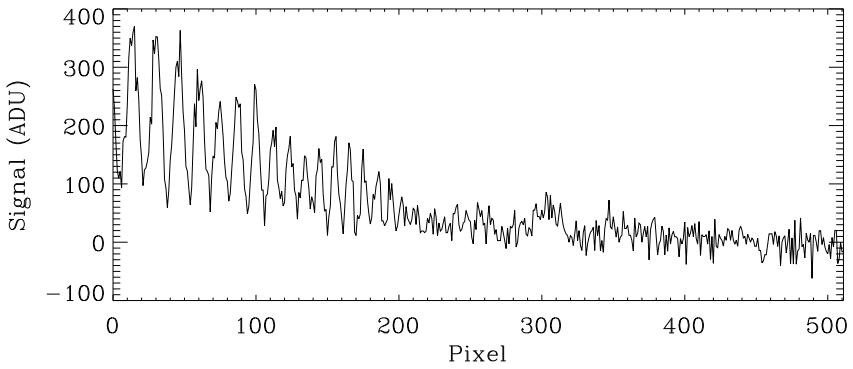


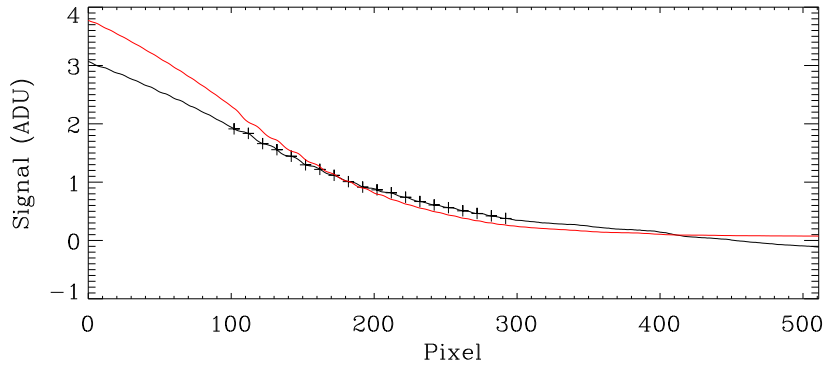
Figure 6.12: The same LDA measurement as shown in Figure (6.11) after the profiles have been combined and the background removed.

and measurement is excellent with respect to the small scale structure of the scattering pattern. However, Figure (6.13a) shows the overall shape to be not as satisfactory. The pattern measured by SPARCLE is too "flat" when compared to that given by Mie theory. An obvious cause of this discrepancy is the the loss of particle scattering signal into the background signal towards higher scattering angles. This effect is clearly apparent in Figure (6.11) and is due not only to the rapid decrease in scattering signal towards the backscatter direction, but also to the corresponding increase in the background signal.

Another problem encountered was that, because of the finite width of the sample air beam, any given LDA measurement could cover a range of scattering angles. A particle located at one edge of the sample air stream will produce a scattering pattern measured from 43.0° to 141.4°, whereas one located at the other edge will produce a pattern from 38.5° to 137°. A shift of this size has little effect on FFT measurement of the fine structure, but can have a strong effect on the shape of the smoothed profile near its edges where the smoothing is truncated.

In order to alleviate the problems mentioned above, the construction of the measurement vector was changed slightly from that used on simulated data (see Chapter (4)). Firstly the number of samples taken from the smoothed scattering profile was reduced from 30 to 20, thus limiting the points to the region of the profile where signal is available and is free from edge effects. Secondly, rather than normalizing the profile by the mean of the entire profile, the mean of the

(a)



(b)

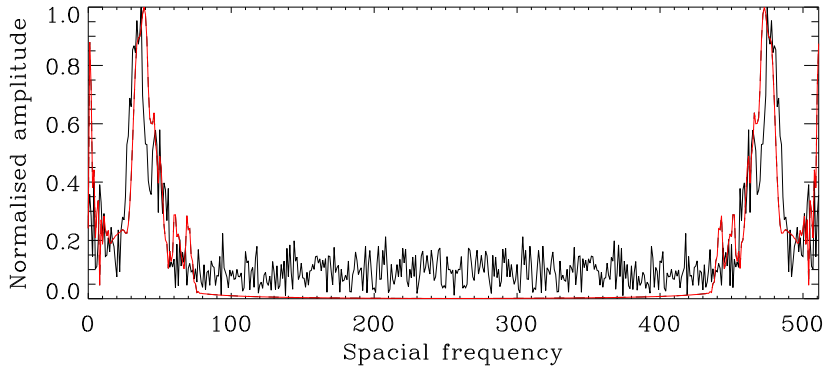


Figure 6.13: The LDA measurement shown in Figure(6.11) in the form used by the retrieval. Plot (a) shows the smoothed, normalized profile with the points used in the retrieval and the curve predicted by Mie theory in red. Plot (b) shows the Fourier transform of the pattern's small-scale structure, again with the theoretical prediction plotted in red.

20 samples was used. As shown in Figure (6.13) this resulted in only a small section of the LDA profile being available for use in the retrieval, however trial retrievals using a greater range of the LDA measurement were found to result in poorer performance.

The presentation of LDA measurement errors to the retrieval scheme also proved difficult. The ideal method of producing error estimates is to make multiple measurements of the same signal. The measurement vectors, $\mathbf{y}_i$, produced by

this process can then be used to construct the measurement covariance matrix,

$$S_\epsilon = E\{[\mathbf{y}_i - \mathbf{y}_j][\mathbf{y}_i - \mathbf{y}_j]^T\}, \quad i \neq j \tag{6.1}$$

(see Appendix (B)). The LDA data from SPARCLE included a background profile for each measurement which can be used to produce an estimate of S_ϵ using the above method. Unfortunately there are several practical problems with implementing this approach:

- The background profiles lack the fine structure produced by particle scattering, hence the covariance matrix will not reflect the influence of measurement noise on the fine structure.
- The signal strength of the background profile is lower than found in a measurement profile and the measurement noise on the LDA is proportional to the signal strength.
- Since the measurement profiles are actually a sum of three separate LDA profiles, an error estimate on the background profiles alone will underestimate errors in the measurement profiles.
- The covariance matrices produced using this method have been found to be singular, or very nearly so, and thus cannot be inverted for use in the retrieval code.

To overcome these difficulties, the forward model was used to calculate measurement covariance matrices on a measurement by measurement basis, using a method similar to that used with the simulated measurements (see Chapter (4)). This method involves randomly perturbing the raw LDA profile (as shown in Figure (6.12)) based on its uncertainty and passing it through the forward model. This produces a sample of perturbed measurement vectors, $\acute{\mathbf{y}}$, which can then be used to estimate the value of the covariance matrix:

$$S_\epsilon = E\{[\mathbf{y} - \acute{\mathbf{y}}][\mathbf{y} - \acute{\mathbf{y}}]^T\}. \tag{6.2}$$

It is possible to estimate an uncertainty for each value in each LDA profile based on the variances in a series of background profiles, since such profiles provide separate measurements of the same signal. However, this method still suffers from most of the problems listed above. In addition the background signal is effected by changes in the laser intensity during each measurement run. The

resulting systematic difference will result in overly large error estimates, since this signal will be removed from the measurements when they are normalized. Computation of error estimates in this way indicates that the uncertainty in the background profiles is relatively constant across the LDA, however. Based on this it was decided to estimate the uncertainty in each LDA profile using the standard deviation of the last 200 pixels of that profile. Because these pixels do not show evidence of signal above the background level this standard deviation should provide a reasonable estimate of the level of noise in each profile. The estimate of this expectation value was determined by taking the mean of a series of 1000 separate difference matrices. However, in practice this method was found to produce matrices which were near singular and thus were impossible to accurately invert.

The occurance of off-diagonal elements of significant size is not an unexpected phenomenon in this case. Off diagonal elements indicate that the elements of the measurement vector are correlated. This is a consequence of the smoothing and FFT applied to the measurement before it is presented to the retrieval.

To ensure the measurement covariance matrices produced were invertible it was decided to use only variance estimates from Equation (6.2) and set all off diagonal elements in the covariance matrix to zero. As well as producing a numerically stable diagonal matrix this approach also has the advantage that the more robust median can be used to estimate the expectation value of each variance, since each measurement is considered independent. Examples of the measurement vector errors for each measurement run are given in Figure (6.14). All of the error vectors show the same general form. The first 20 values in each profile correspond to the uncertainty of each element in the smoothed LDA profile. The minima found at approximately element 7 or 8 is due to the normalization of the smoothed profile about its mean – the mean value being located at the point of this minimum. The profiles also demonstrate the stability of the location of the FFT with respect to random noise in the LDA profile. In most cases the standard deviation for this value was below 0.1. The final element of the error profiles corresponds to the PMT signal error estimate discussed in Section (6.3.1).

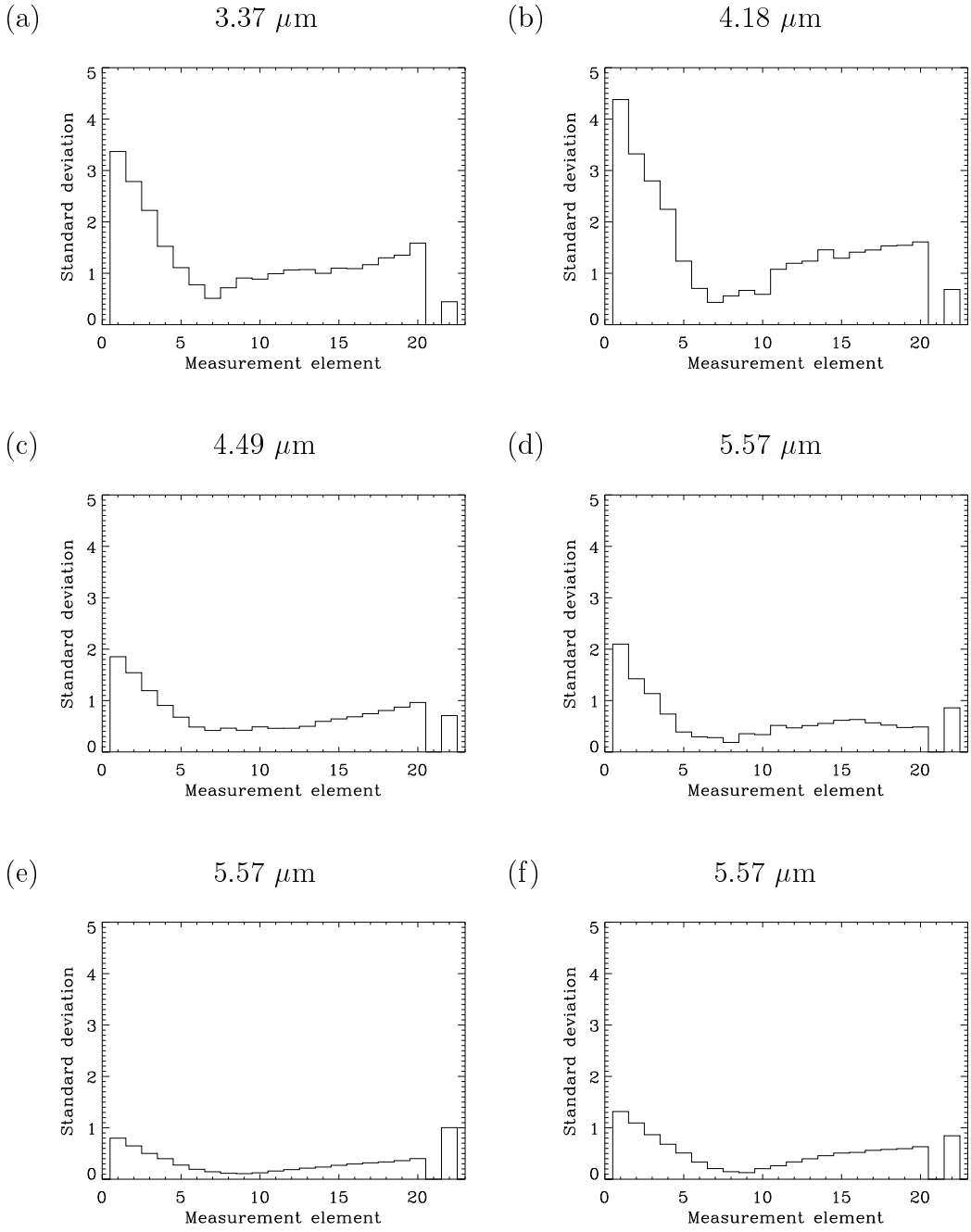
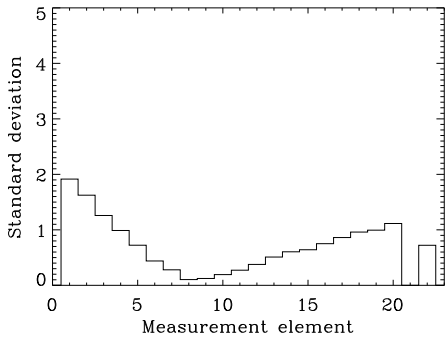
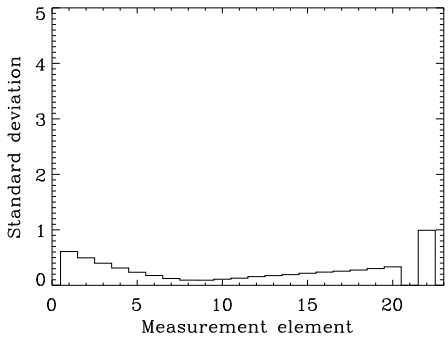


Figure 6.14: Example variances from each measurement run.

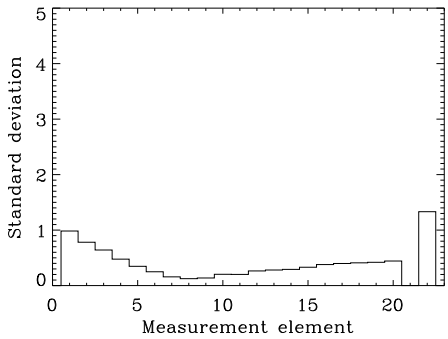
(g) 5.99 μm



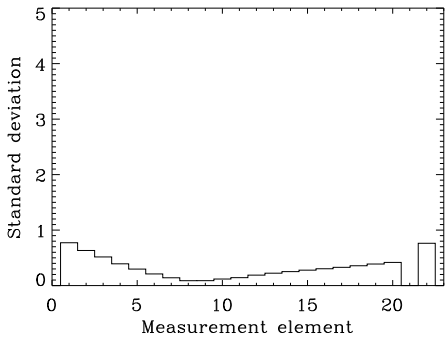
(h) 6.43 μm



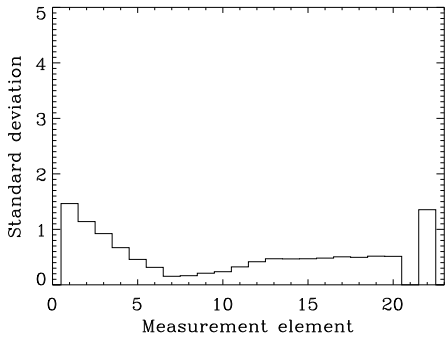
(i) 6.91 μm



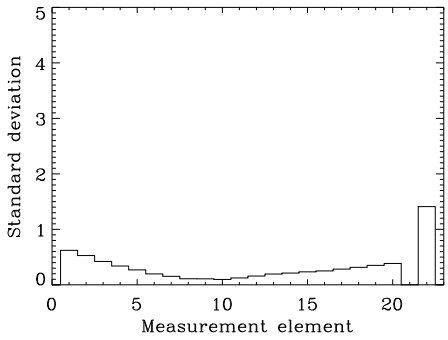
(j) 7.98 μm



(k) 9.22 μm



(l) 9.22 μm



6.4 Summary

The overall results of the experiments described in this chapter have been satisfactory. The major systems of the SPARCLE instrument (i.e. the pumping system, laser and detectors) have been found to be consistent in their behaviour. The pumping system was found to provide constant flow to within $\sim 10\%$ within each measurement run. The laser was found to be less consistent, due to the continual repeated adjustments made necessary by alignment changes due to thermal effects. However, this should have little affect on the measurements due to the normalization of the data before they are presented to the retrieval scheme.

The two instruments used to produce the aerosol and provide independent particle size information (the VOAG and APS, respectively) also performed well. The VOAG was used to produce 12 different aerosol distributions, all of which were close to monodisperse. The nominal radii, as measured by the APS, of these distributions ranged between 3.37 and 9.22 μm. The average half-width of the distributions was found to be approximately 0.8 μm. It was found that the size bins used by the APS in the size range of interest was too coarse to define the distributions produced by the VOAG. In many cases it is the resolution of the APS which defines the width of the aerosol distribution.

Initial analysis of the data collected by the SPARCLE instrument has made clear that the signal to noise ratio is the primary factor limiting its performance. The necessity of baffling the laser beam (thus lowering the background signal to a level where LDA measurements can be made) results in ambiguity in the PMT data. Even with the laser baffling in place, the particle scattering signal is lost in the background noise of the LDA at high scattering angles. Despite these difficulties data from the SPARCLE prototype can be presented in a form suitable for use in the retrieval scheme.

Because of the degeneracy in the PMT measurements all data from each measurement run were combined into a single measurement, based on the expected distribution of random samples from a Gaussian beam shape. The uncertainty of these measurements was based on the deviation of the actual distribution of PMT signals from the expected one.

The presentation of the LDA data was also changed from that used in the

retrieval of simulated data (Chapter (4). Because the particle scattering signal is lost in the background noise at high scattering angles, the number of samples taken from the smoothed LDA profile was reduced from 30 to 20. Also, rather than normalizing by the mean of the entire profile, the mean of the 20 samples was used. Uncertainty values on the LDA measurements were estimated from the area of each profile where background signal dominated. It was found that the covariance matrices produced from this method were found to be near-singular. Thus, only their diagonal elements (i.e. the variances) were used in forming the measurement covariance matrices.

Chapter 7

Analysis of SPARCLE measurements

In this chapter the results of applying the retrieval scheme to real measurements of known, monodisperse aerosol distributions are given. Details of the experimental procedure used to obtain these measurements and description of the data itself are given in Chapter (6). Similarly to the retrievals of simulated data, as well as data from the DAWN-A instrument, the SPARCLE data was retrieved using both 2-dimensional (particle radius and the real part of the refractive index) and 3-dimensional (particle radius and complex refractive index) retrieval algorithms. The 2 and 3-dimensional results are discussed in Sections (7.1) and (7.2), respectively.

In both 2 and 3-dimensional cases the *a priori* was set to be $r = 5.0$ μm and $m = 1.4 - 0.0001i$ with an associated covariance of

$$\mathbf{S_a} = \begin{bmatrix} 625.0 & 0 & 0 \\ 0 & 0.04 & 0 \\ 0 & 0 & 0.25 \end{bmatrix}$$

In the 2-dimensional case the appropriate submatrix of this was used. Clearly this puts very little constraint on the particle size while the refractive index is much more tightly constrained. The range of the real part of the refractive index for atmospheric particles is between approximately 1.33 and 1.65, with a value near 1.4 being the most common. Also, most particles found in the atmosphere are essentially non-absorbing at visible wavelengths, although particles with absorption values around 0.6 have been observed (Dick, 1998). The *a priori* and its covariance have been chosen to represent these features. However, it was

found that, to ensure reasonably stable retrieval performance, the real part of the refractive index needed to be more tightly constrained than was the case in the retrieval of simulated measurements (Chapter (4)).

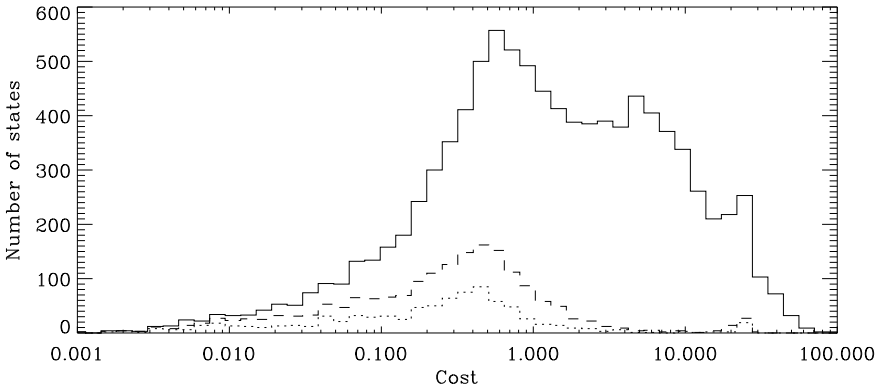
7.1 2-Dimensional retrieval

The cost function distribution and the associated cumulative error distributions from the 2-dimensional retrievals are shown in Figure (7.1). The limitations to the retrieval introduced by the highly non-linear forward model are again clear to see. Out of the 9677 retrievals performed 841 agree with the nominal state within one standard deviation as defined by the covariance of the retrieved state, $\hat{\mathbf{S}}$, and 1854 agree within two standard deviations. The distribution of all the states shows little indication of a clear minimum when compared to the cost function distribution produced by 2-dimensional retrieval of simulated measurements (c.f. Figure (4.12)). However there is a small minimum around a cost of two and very few of the states which agree with the expected value to within either one or two standard deviations have a cost higher than this value. On average, the deviation of the retrieved state from the expected state also shows a strong correlation with its cost (Figure (7.1b,c)). However there is no point at which there is a clear rapid increase in the cost of the retrieved state which could be associated with a threshold in the cost function corresponding to 'good' and 'bad' retrievals – the rapid increase in the maximum absolute error of the retrieved radius at a cost of approximately 10 (Figure (7.1c)) is due to a single outlying point.

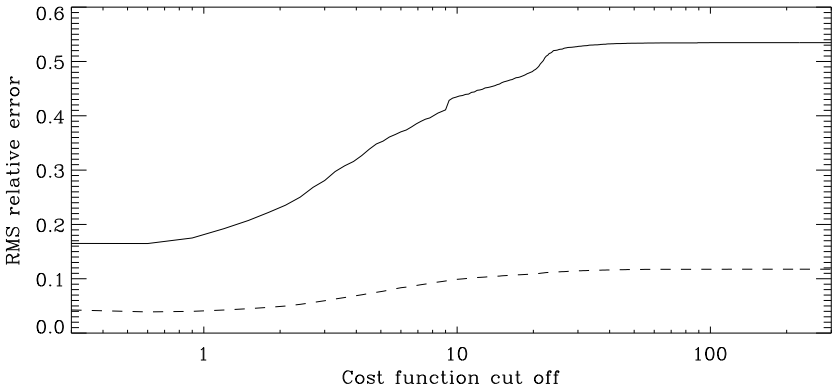
The χ^2_s distribution relating the retrieved state to the measurement (see Chapter (4) and Appendix (B)) is given in Figure (7.2). As is expected, this distribution is far from the form expected if the linear approximation was accurate. As with the retrieval of simulated measurements the cost function (which does make use of a linear approximation to the forward model) was used as an estimator of the accuracy of a solution, with those states with a cost of less than two being deemed acceptable.

The retrieval size distributions for each of the twelve measurement runs, using the same size bins as the APS, are shown in Figure (7.3). It is clear that the retrieval has resulted in a significant broadening of the distributions, but

(a)



(b)



(c)

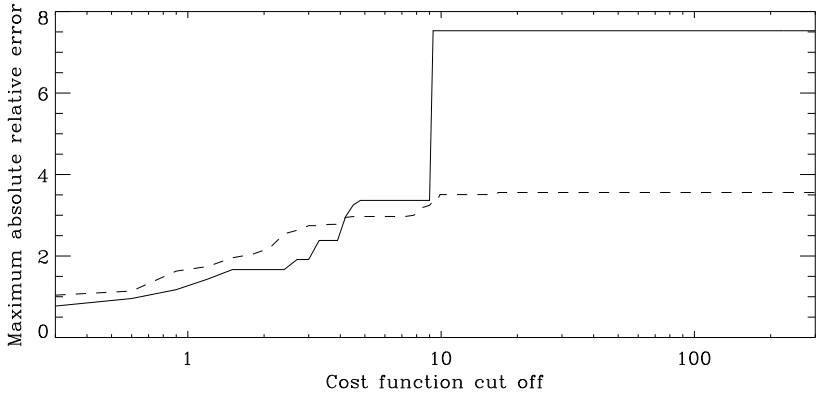
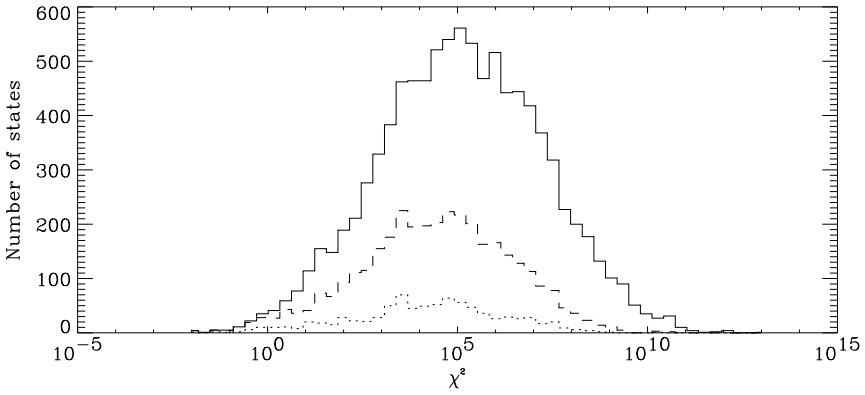
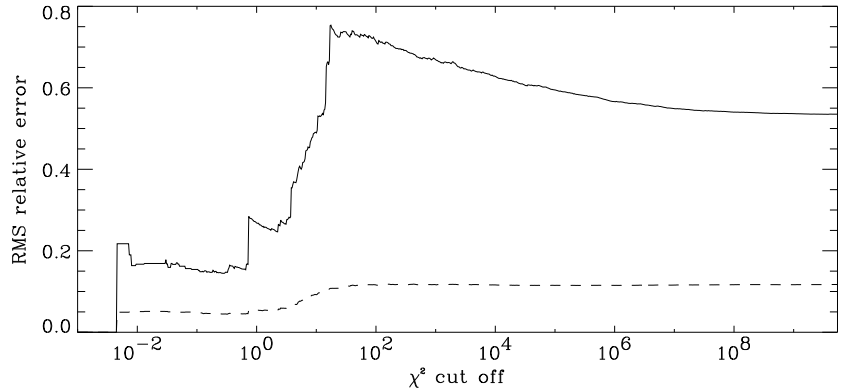


Figure 7.1: Retrieved state cost function distribution for 2-dimensional run. The three distributions shown in (a) are: all retrieved states (solid line), states which agree with the nominal state within the retrieved error estimate (dotted line) and states which agree to within twice the error estimate (dashed line). In plots (b) and (c) the solid line shows the particle size deviation and the dashed lines shows the refractive index deviation. The relative error for the refractive index has been scaled by a factor of 10 in plot (c).

(a)



(b)



(c)

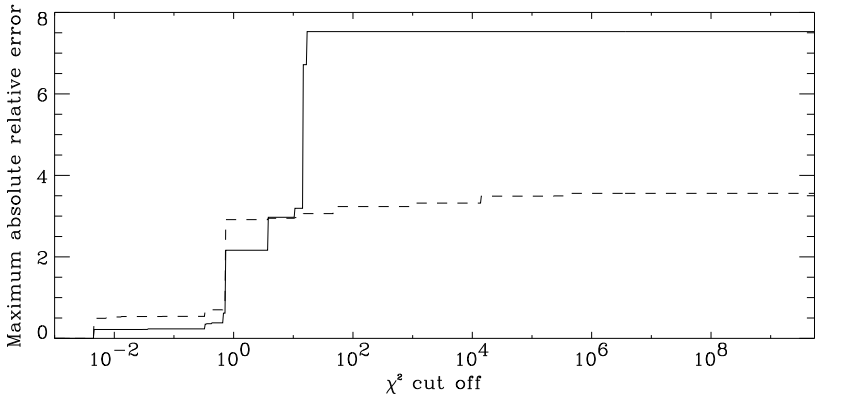


Figure 7.2: The same as Figure (7.1) but showing the χ^2 distribution. Again the relative error for the refractive index has been scaled by a factor of 10 in plot (c).

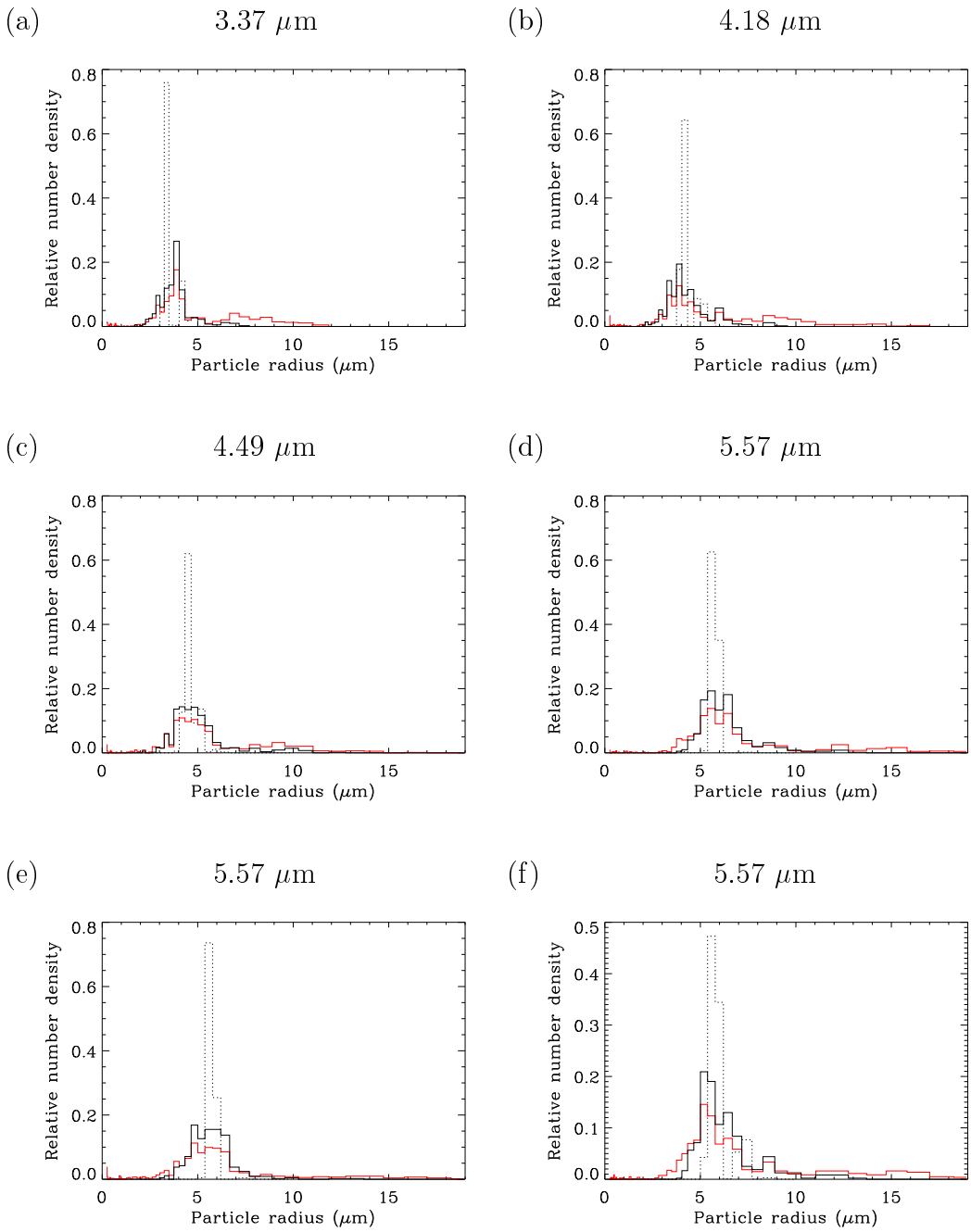
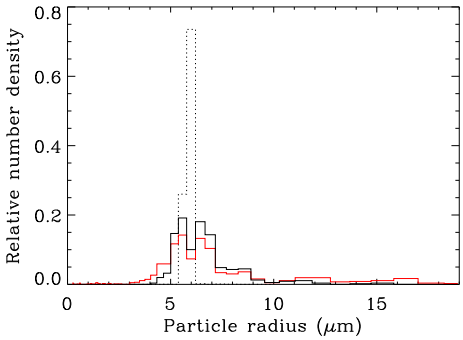
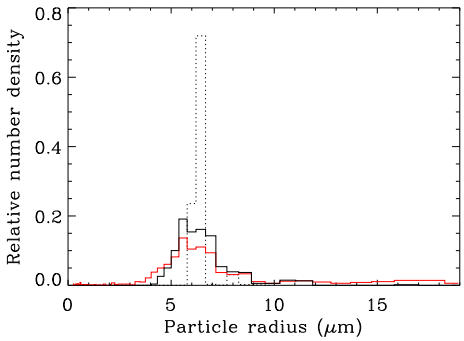


Figure 7.3: Retrieved radius distributions from 2-dimensional retrieval run, plotted using the APS size bins. The distribution of all retrieved states is shown in red, those states which meet the cost function criteria are plotted in black and the APS measured distribution is shown with the dotted line. The distributions are normalized by the number of states they contain.

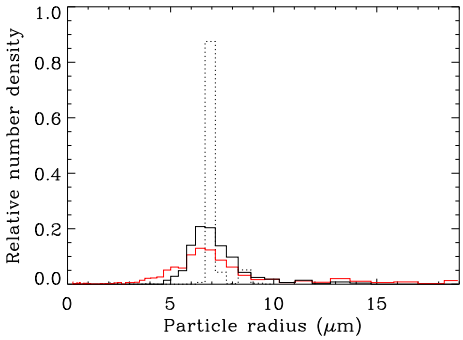
(g) 5.99 μm



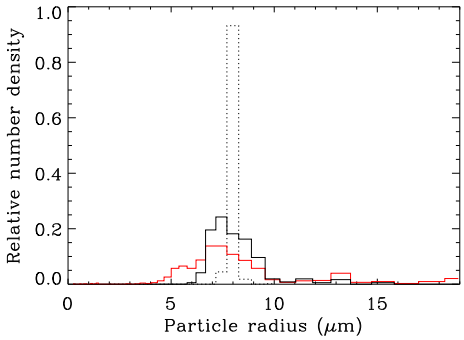
(h) 6.43 μm



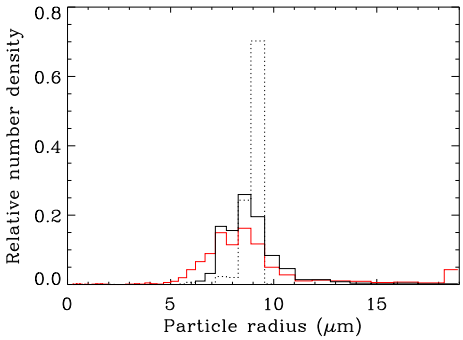
(i) 6.91 μm



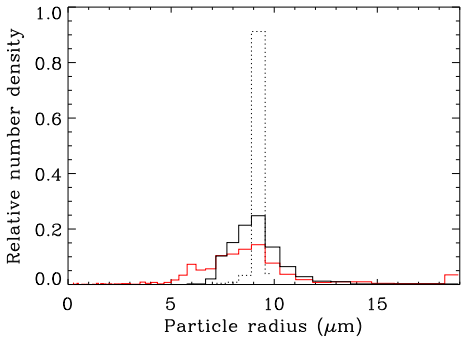
(j) 7.98 μm



(k) 9.22 μm



(l) 9.22 μm



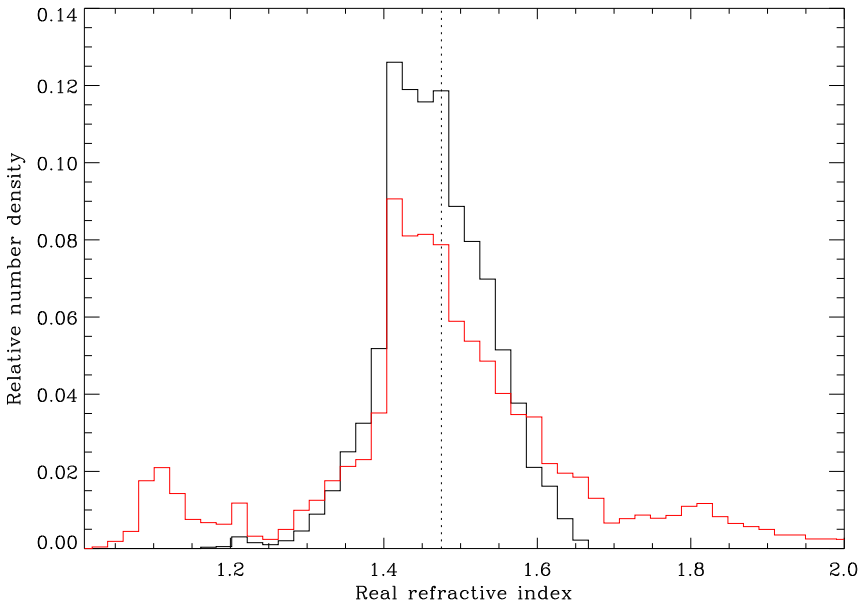


Figure 7.4: The distribution of retrieved values of the real part of the refractive index for the 2-dimensional retrieval runs. As with the size distributions (figure (7.3)) all the retrieved states are plotted in red and those meeting the cost function criteria are shown in black. The distributions are normalized by the number of states they contain and the expected value is denoted by the dotted line.

the peak size shows good agreement with that measured by the APS. It is also clear that the use of the cost function criteria removes most of the outlying data points. Figure (7.4) shows the distribution of retrieved refractive indices for all data. Once again the cost function criteria has resulted in the removal of most outlying values. However, the peak of the distribution occurs at a value of 1.4 (equal to the *a priori* value).

Figures (7.5) and (7.6) show the retrieved radii and refractive indices for each measurement run, respectively. The data in these figures have been sorted by the maximum amplitude found in their raw LDA profiles, i.e. those measurements with the greatest signal strength appear first. The behaviour of the retrieval is strongly dependent on the size of the particles being measured. The retrievals done on the larger particle sizes show evidence of a decrease in performance with decreasing signal strength, the effect being most clearly observed in the results from runs (i), (j) and (l), where the scatter in both radius and refractive

index increases with retrieval number. In general all retrievals for particles with nominal radii $r \geq 5.57$ μm have given results clustered about their expected values, but with considerable spread.

For the particles with radii less than 5.57 μm however the retrieved states are less satisfactory. The retrieved values of radius for these runs remains clustered about the nominal value, but there are clear oscillations in the preferred state as the signal strength decreases. Hints of these oscillations can also be seen in the $r \geq 5.57$ μm retrievals, particularly run (d). The retrieved radii also appear to improve with decreasing signal strength, particularly in the case of run (a). Examination of the retrieved refractive indices for these runs shows similar oscillatory behaviour as seen in the radius values and the values seem to be more clustered about the *a priori* ($n = 1.4$) than the expected one. Here we are seeing the effects of a low signal to noise ratio throughout the runs. For the measurements with the lowest signal strength at these sizes the information content of the LDA measurements is very small and the retrieval is dominated by the PMT signal and the *a priori*. This situation is similar to a simple one-measurement, one-dependent variable retrieval: the *a priori* fixes the refractive index value and the radius which best fits the PMT signal will be the solution. Thus the retrievals at the end of each of the smaller particle size runs become more consistent in their radius estimate and tend to converge to the *a priori* refractive index. It is this effect which causes the peak of the refractive index distribution to occur at the *a priori* value.

The oscillatory nature of the retrieved values as the signal strength is varied is more difficult to explain. Extensive examination of the data used in the retrievals showing this effect has led to the conclusion that it probably results from a combination of variations in the range of scattering angles sampled by the LDA sensor (see Section (6.3.2)) and the method used to describe the fine structure of the scattering pattern. At small sizes, not only is the signal strength much reduced but the fine structure used to generate the FFT component of the measurement vector becomes much broader, i.e. the location of the FFT peak is based on fewer complete cycles of the fine structure. Since the periodicity of the fine structure is not entirely constant, such measurements will display a variation in the peak spatial frequency determined by the FFT as the position of the particle in a sample flow (and thus the range of scattering angles intercepted by the LDA)

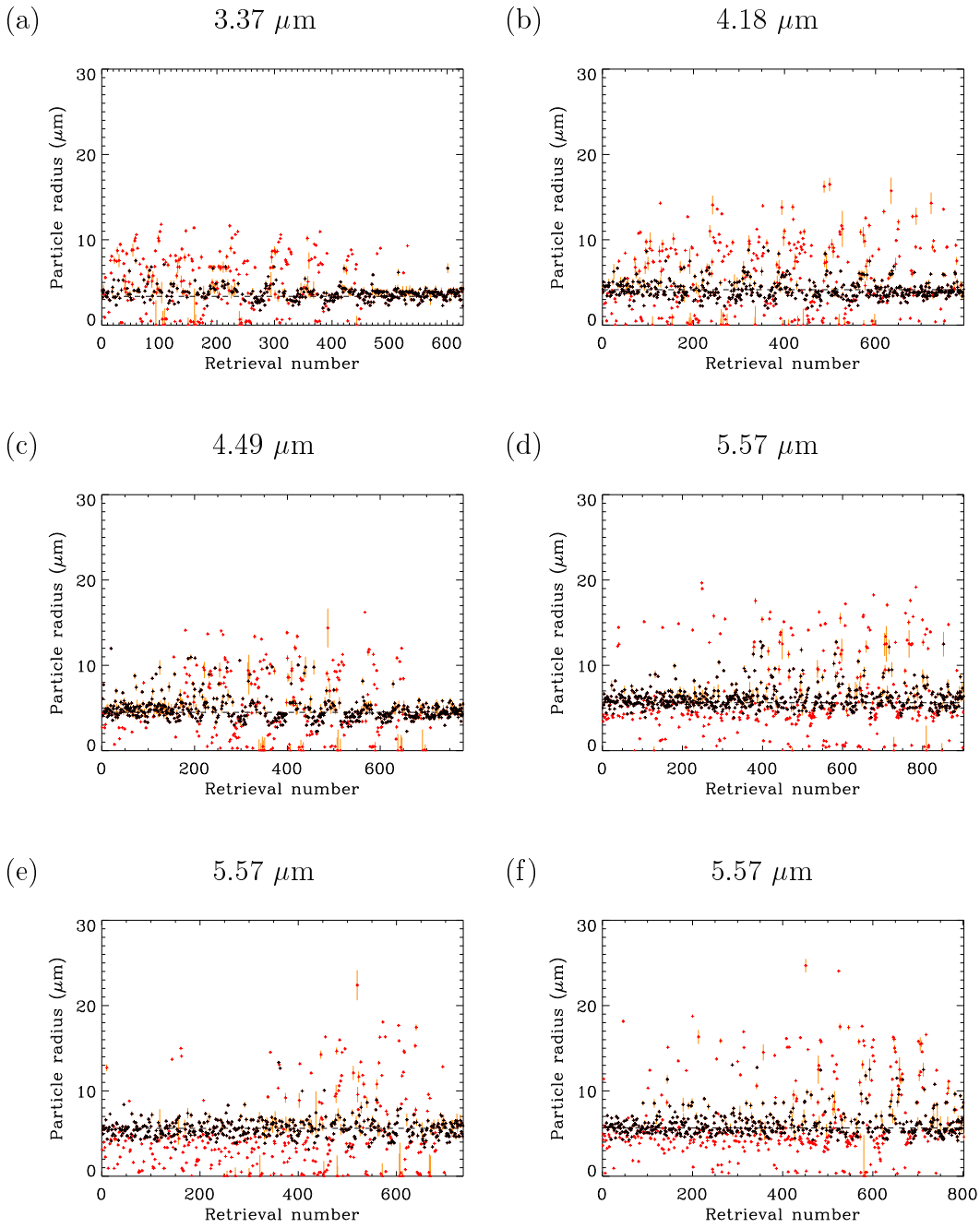
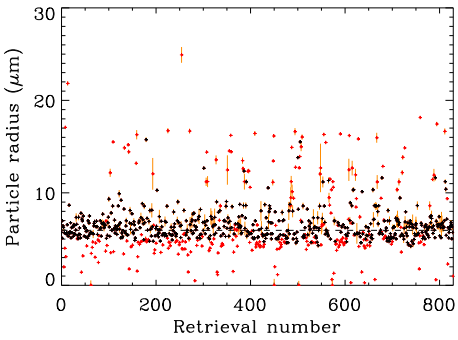
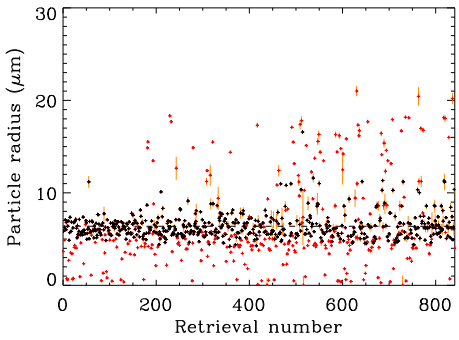


Figure 7.5: Retrieved radii from 2-dimensional runs. States which met the cost function criteria are plotted in black, while those failing the criteria are red. The error estimates given by optimal estimation are shown in orange and the nominal state is denoted by the dashed line.

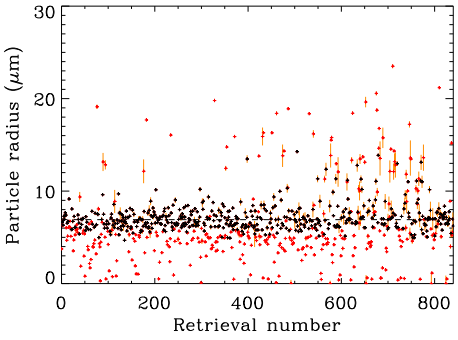
(g) 5.99 μm



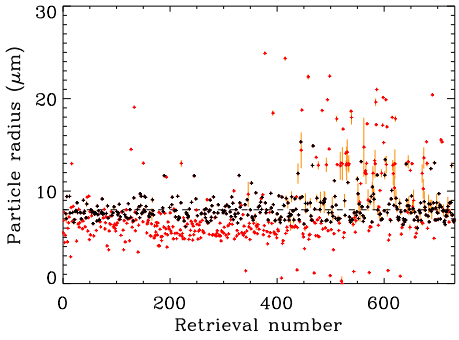
(h) 6.43 μm



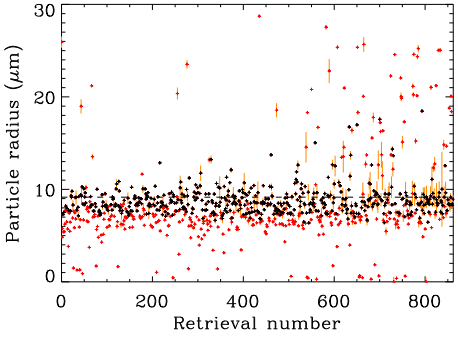
(i) 6.91 μm



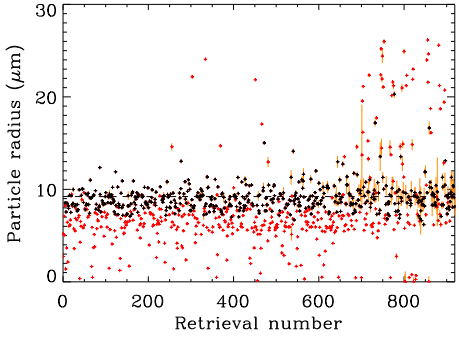
(j) 7.98 μm



(k) 9.22 μm



(l) 9.22 μm



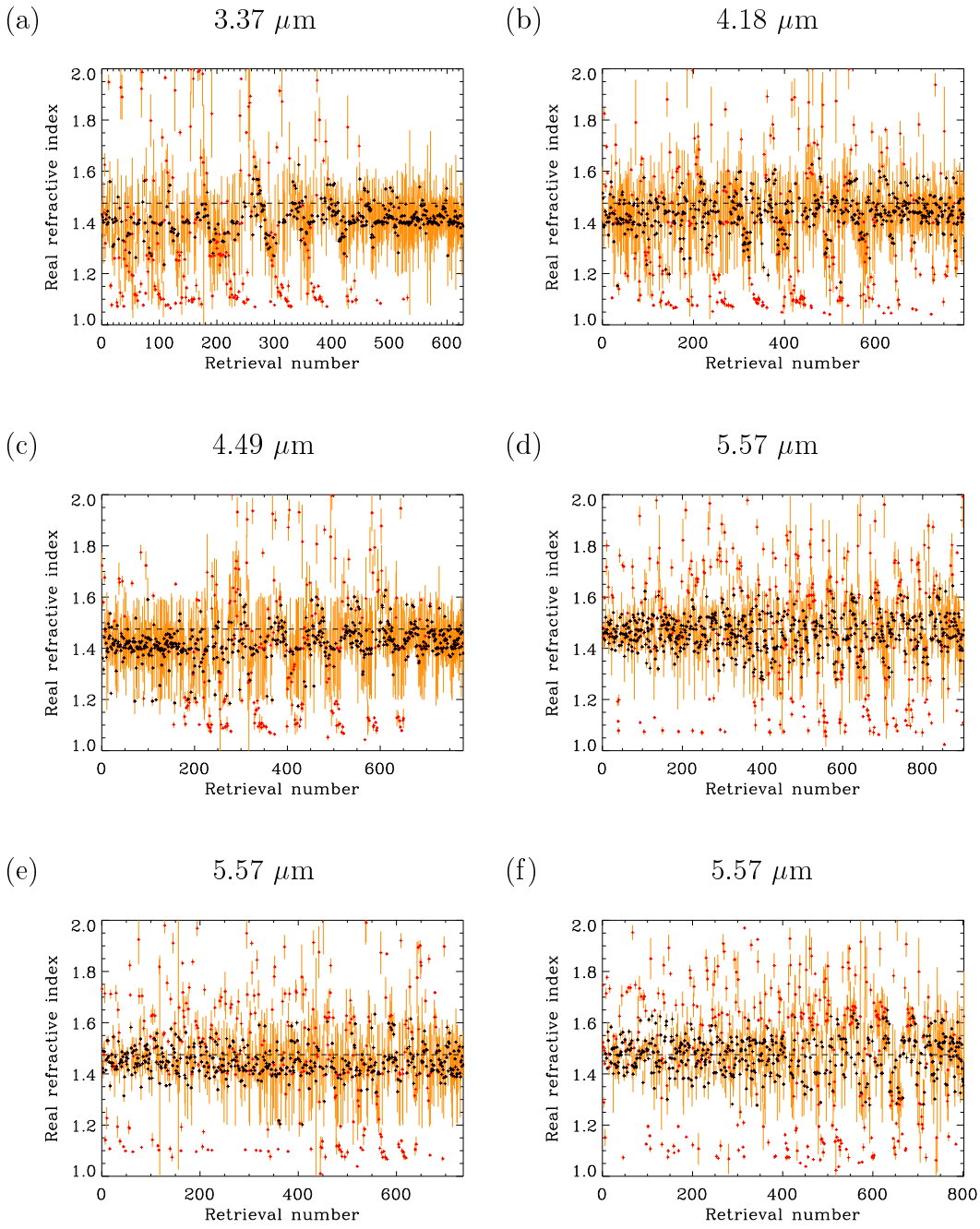
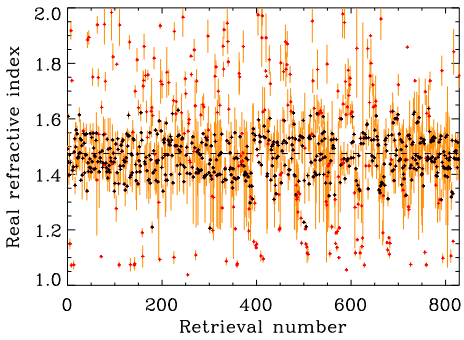
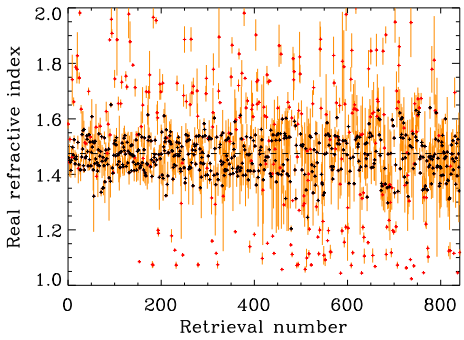


Figure 7.6: The same as Figure (7.6) but showing the retrieved values of the real part of the refractive index for the 2-dimensional runs.

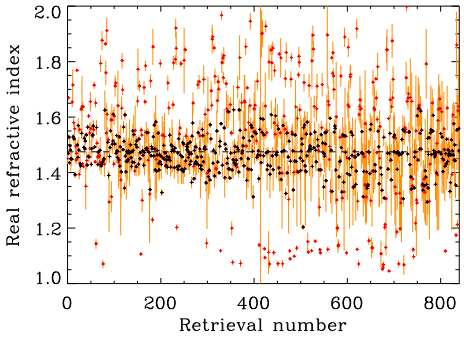
(g) 5.99 μm



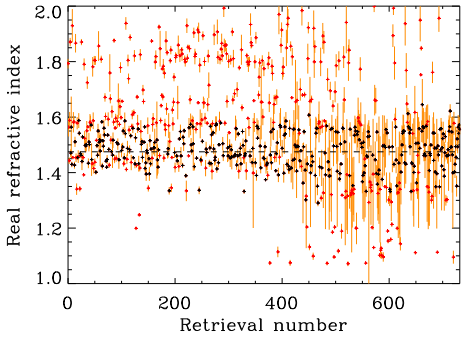
(h) 6.43 μm



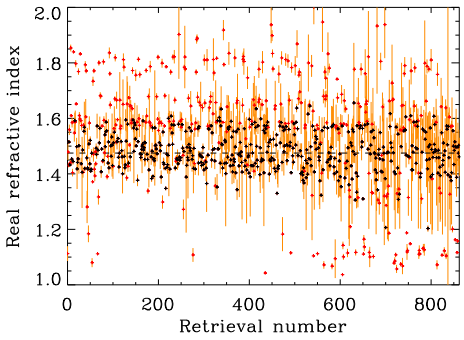
(i) 6.91 μm



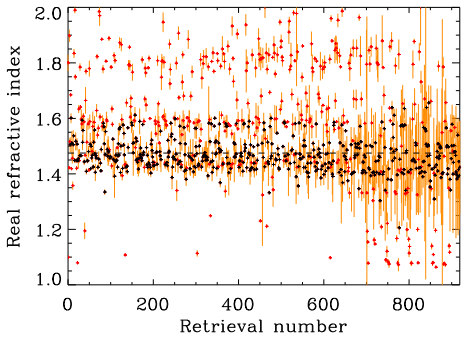
(j) 7.98 μm



(k) 9.22 μm



(l) 9.22 μm



is changed. Because the intensity of scattered light is rapidly decreasing with increasing scattering angle in this region, variations in the angular range of the LDA measurement will also result in changes in the peak amplitude measured by the sensor. This complicates the relationship between the location of a particle within the sample air flow and the maximum amplitude of the resulting LDA profile and should result in similarly offset particles appearing near each other in the sorted data set. The interaction of these two effects seems the most plausible explanation of the observed variation. Other possibilities should be considered, but can generally be discounted:

- The effect could be explained by some fluctuation in the smoothed LDA profile with changing signal strength, however no periodic pattern could be discerned in these data.

- The possibility of the oscillation being caused by a numerical problem within the retrieval code is extremely unlikely since the retrievals were not done in the order displayed. There is no periodic behaviour present when the results are displayed in the order they were obtained.

Data in Figures (7.5) and (7.6) are plotted with error bars which are simply the appropriate standard deviation inferred from the state covariance matrix, $\hat{\mathbf{S}}$, calculated as during the optimal estimation. Although these values are generally poor estimates of the actual errors in the retrieved states, they are included because of the light they shed on the retrieval process. The covariance of the state vector is given by

$$\hat{\mathbf{S}} = (\mathbf{S_a^{-1}} + \mathbf{K^T S_\epsilon^{-1} K^T})^{-1} \tag{7.1}$$

where $\mathbf{S_a}$ and $\mathbf{S}_\epsilon$ are the *a priori* and measurement covariance matrices, and $\mathbf{K}$ is the weighting function at the solution (see Appendix (B)). A value of $\hat{\mathbf{S}}$ close to that of $\mathbf{S_a}$ will occur if $\mathbf{K}$ is small, which in turn, implies that the cost function has little dependence on the state. Recalling that the *a priori* variance on the refractive index is 0.04, corresponding to a standard deviation of 0.2, we can see that many of the retrieved refractive indices have error estimates which are approximately equal to the *a priori* value. This is particularly true at the smaller particle sizes, but even at the largest sizes the runs show large error bars for those profiles with the lowest LDA amplitude.

The scatter plots of the retrieved refractive indices also suggest the important role the *a priori* plays in the selection of acceptable states. In the hypothetical case where the refractive index component of the *a priori* cost was the only contributing factor, the cost would be given by:

$$\text{Cost} = \frac{1}{0.04}[\hat{\mathbf{n}} - \mathbf{n_a}]^2, \tag{7.2}$$

where $\hat{\mathbf{n}}$ is the estimated refractive index and $\mathbf{n_a}$ the *a priori* refractive index. In this case a difference of 0.28 between $\hat{\mathbf{n}}$ and $\mathbf{n_a}$ would result in a cost of 2. Examination of Figure (7.4) shows that virtually no states with a retrieved refractive index outside the 1.2 – 1.6 interval have been accepted by the cost function criteria and this can be attributed to the *a priori* cost of the refractive index alone. To base the selection of acceptable states on their ability to reproduce the measurement alone one could use the first term of the cost function alone:

$$\text{Fit cost} = [\mathbf{y} - \mathbf{F}(\hat{\mathbf{x}})]^T \mathbf{S}_\epsilon{}^{-1} [\mathbf{y} - \mathbf{F}(\hat{\mathbf{x}})]. \tag{7.3}$$

However, the *a priori* does represent valid information on the true state and thus this knowledge should be used in selecting acceptable states.

Three key questions which must be addressed in quantifying the usefulness of the SPARCLE measurements are:

- Do the size distributions and refractive indices agree with the expected (APS measured) values within error appropriate estimates?
- Are the size distributions, as measured by SPARCLE, significantly different from one another?
- What is the resolving power of the SPARCLE instrument with respect to particle size and refractive index?

It is unreasonable to expect statistical tests such as the χ^2 test performed in Chapter (4) to produce good results because of the discontinuous nature of the distributions measured by the APS. A χ^2 comparison between the APS and SPARCLE determined size distributions was performed however as it does provide a quantitative ranking of the different runs (Table (7.1)). As expected the χ^2 values are such that the theory predicts that none of the SPARCLE distributions could have been sampled from the same population as the APS

distributions. This can be attributed to the broadening of the distributions by SPARCLE, as observed in the refractive index tests given in Section (4.4). Disregarding the probability values and comparing the χ^2 and ν values directly we can see that there is no apparent trend in the χ^2 comparisons with increasing particle size. The best fit between the two data sets is found in run (f) (nominal radius 5.57 μm) and the worst is run (g) (nominal radius 5.99 μm).

	χ^2	ν	$P(\chi^2\|\nu)$		χ^2	ν	$P(\chi^2\|\nu)$
(a)	196	21	1	(g)	242	24	1
(b)	190	22	1	(h)	165	26	1
(c)	133	25	1	(i)	163	27	1
(d)	147	25	1	(j)	192	28	1
(e)	175	23	1	(k)	126	25	1
(f)	113	26	1	(l)	145	24	1

Table 7.1: The results of applying a χ^2 comparison between the size distributions given by the 2-dimensional retrieval and the APS. ν is the number of degrees of freedom (number of non-zero size bins). $P(\chi^2|\nu)$ is the probability that a random value taken from a χ^2 distribution with ν degrees of freedom will have a value larger than that given – all were equal to one to six significant figures.

Table (7.2) shows the mean, median and peak APS size bin of the size distributions given by the 2-dimensional retrieval, along with the standard deviation, mean absolute deviation and half width of the distributions on the APS size bins scale. Also shown are the nominal radii (i.e. the peak size bin of the APS measured distribution) along with the half width of the APS measured distribution (see Section (6.2)). As discussed previously, the peak size bin and half width of the distribution provide the best description of the APS measured distributions because of the coarse binning. As the values provided by the SPARCLE retrieval are essentially continuous they are better described by more usual statistical values. It cannot be judged which average provides the best measure of the centre of a given size distribution from Table (7.2) as both mean and median values fit the data equally well. It is also difficult to determine whether the standard deviation or mean absolute deviation provides the better measure of the spread of the distributions. To clarify the situation a third measure of the distribution is needed.

	APS radii (μm)		SPARCLE radii (μm)					
	Nominal	$\frac{1}{2}$ width	Peak	Mean	Median	σ	MAD	$\frac{1}{2}$ width
(a)	3.37	0.24	3.89	3.70	3.73	0.83	0.57	0.76
(b)	4.18	0.58	3.89	4.16 ←	3.90	1.09	0.80	1.48
(c)	4.49	1.34	4.17	4.80	4.56 ←	1.34	0.89	1.48
(d)	5.57	0.83	5.57	5.91	5.61 ←	2.53	1.48	1.71
(e)	5.57	0.83	4.82	5.54 ←	5.44 ←	1.10	0.79	1.71
(f)	5.57	0.83	5.19	6.07	5.67 ←	1.43	1.01	1.42
(g)	5.99	0.83	5.57	6.43	6.22	1.44	0.99	1.95
(h)	6.43	0.89	5.57	6.37 ←	6.15	1.34	0.93	1.81
(i)	6.91	0.50	6.43	7.07 ←	6.87 ←	1.35	0.93	1.77
(j)	7.98	0.58	7.43	8.06 ←	7.75	1.33	0.93	1.44
(k)	9.22	1.28	8.58	8.77	8.50	1.45	0.98	1.55
(l)	9.22	0.66	9.22	9.05 ←	8.92 ←	1.39	0.96	1.94

Table 7.2: Comparison between the APS measured size distributions and those given by the 2-dimensional retrieval of SPARCLE measurements. Note that both the nominal APS and peak SPARCLE radii are the central radius of the APS size bin where the peak of the measured distribution occurs. States whose median or mean values lie within the APS size bin of the nominal radius are indicated with an arrow.

In Figure (7.3) the SPARCLE retrieval measurements have been displayed using the same size bins as the APS uses to allow direct comparison between the two data sets. However, the APS bins are not ideal as they are logarithmically spaced and are quite coarse in the size range of interest. Figure (7.7) shows the distribution of particle sizes retrieved in the 2-dimensional runs for those states which meet the cost function criteria using 0.1 μm size bins, along with Gaussian fits. Although the distributions are not Gaussian in nature, the fitted curves do characterize the width of the distributions quite well. The parameters of the fitted Gaussians are given in Table (7.3). Comparison with Table (7.2) shows that the mean and standard deviation of the fitted curves is much more closely related to the median and mean absolute deviation of the states than their mean and standard deviation. The discrepancy between the two means can be attributed to the outliers found on the right of the SPARCLE distributions, and once again points to the median and mean absolute deviation being more useful in characterizing the distributions.

Given estimates of the average particle size and the spread of the data, the

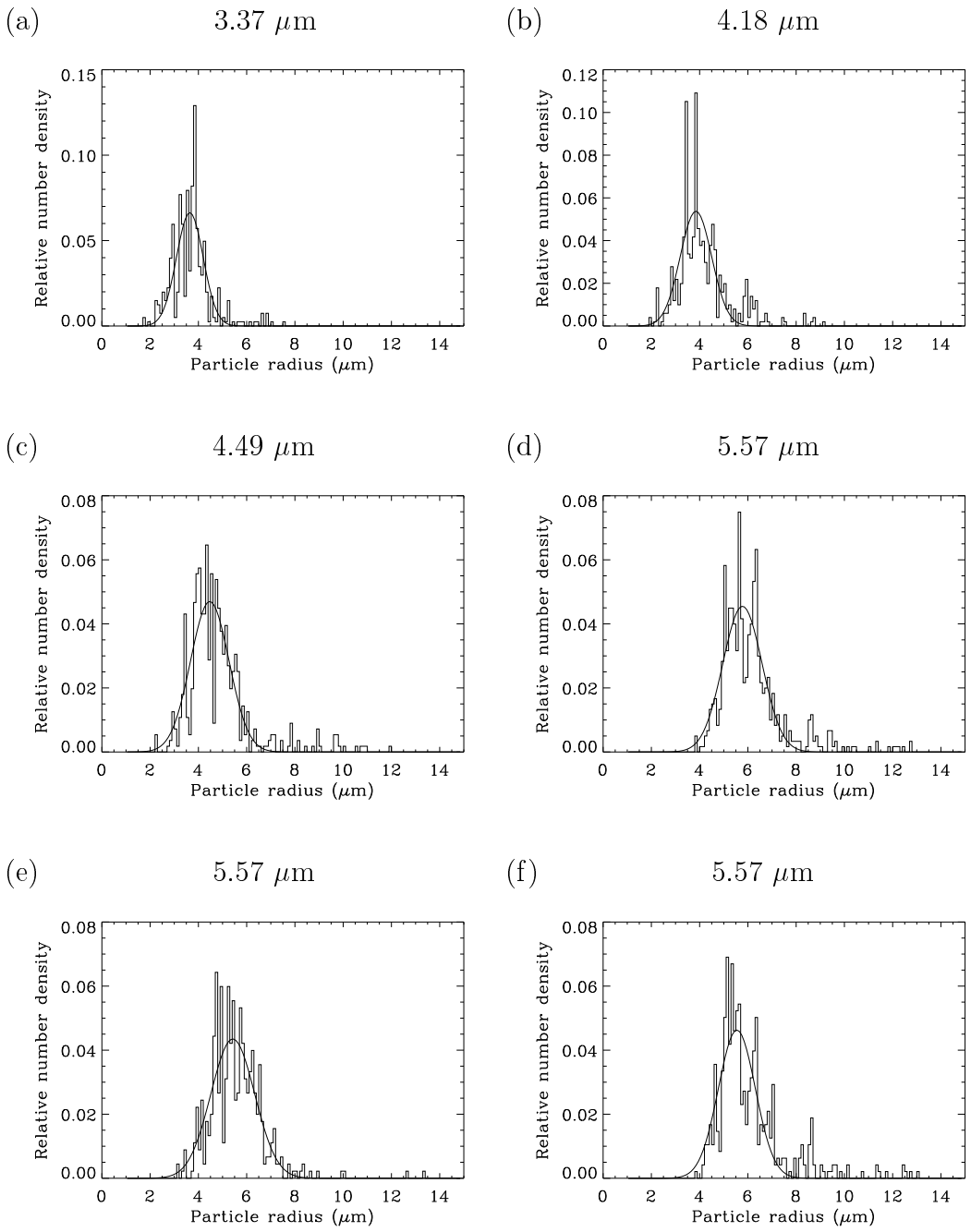
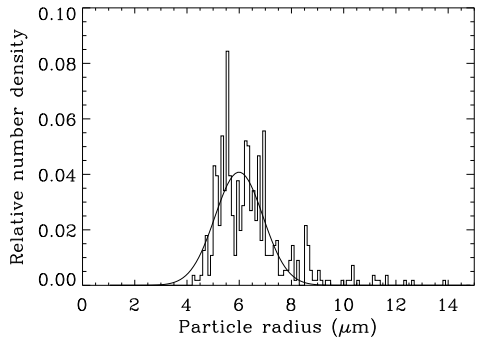
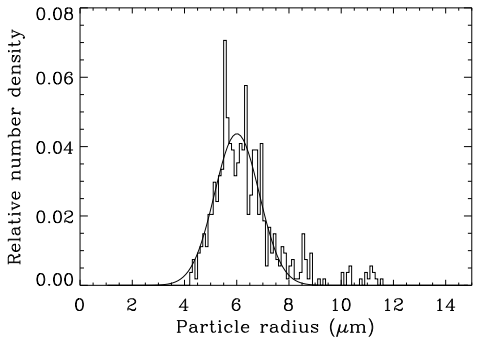


Figure 7.7: Retrieved radius distributions using uniform 0.1 μm bins from 2-dimensional retrieval run along with Gaussian fits.

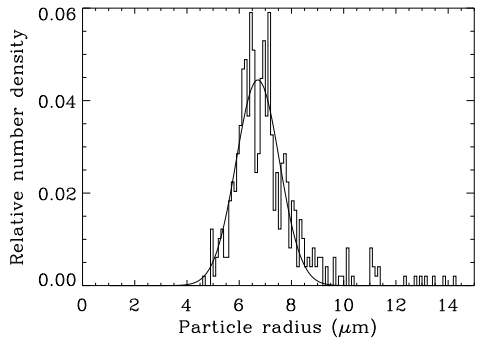
(g) 5.99 μm



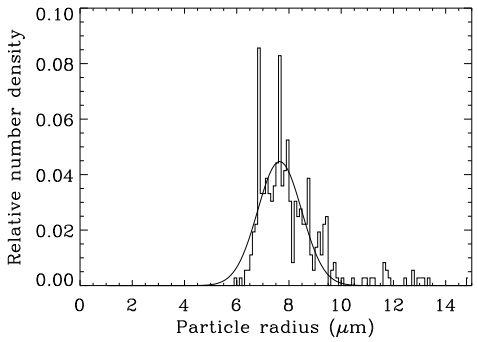
(h) 6.43 μm



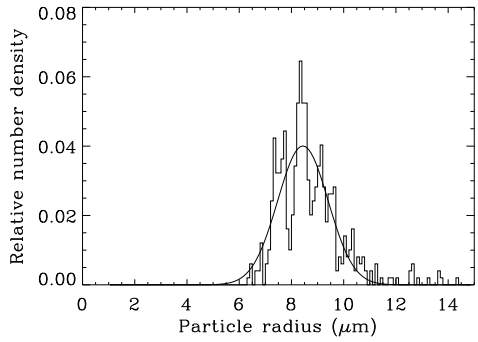
(i) 6.91 μm



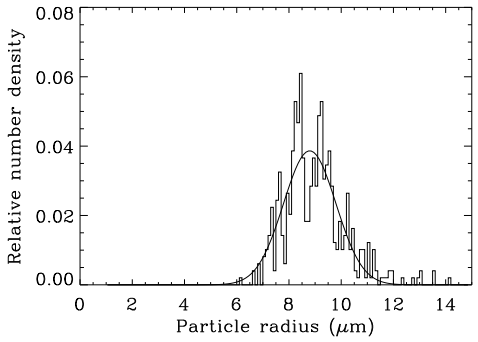
(j) 7.98 μm



(k) 9.22 μm



(l) 9.22 μm



	Nominal	Mean	σ		Nominal	Mean	σ
(a)	3.37	3.64	0.55	(g)	5.99	5.99	0.92
(b)	4.18	3.85	0.65	(h)	6.43	6.00	0.85
(c)	4.49	4.47	0.80	(i)	6.91	6.71	0.83
(d)	5.57	5.78	0.81	(j)	7.98	7.64	0.84
(e)	5.57	5.42	0.91	(k)	9.22	8.44	0.96
(f)	5.57	5.54	0.77	(l)	9.22	8.80	1.00

Table 7.3: The parameters of the Gaussians fitted to the rebinned size distributions produced by the 2-dimensional retrieval after the cost function acceptance criteria has been applied. For reference the nominal radii, as measured by the APS are also included.

questions of agreement with the APS and resolution can now be addressed. Figure (7.8), as well as Figure (7.7) and Table (7.3), show that the size distributions measured by both instruments are centred on very similar values in most cases. However the SPARCLE distributions are, on average, broader than the corresponding APS distribution. It should be noted that if poly-disperse aerosols were measured, rather than a series of pseudo mono-disperse distributions, this broadening would probably not greatly affect the overall shape of the measured distribution (as was found in Chapter (4)).

To give a statistical measure of how different the retrieved distributions are from each other we can use the Kolmogorov-Smirnov (K-S) test. The K-S statistic for two data sets is simply the maximum separation of their respective cumulative probability functions, S_{N1} and S_{N2}. Thus the K-S statistic is simply:

$$D = \max_{-\infty<x<\infty} |S_{N1}(x) - S_{N2}(x)|, \tag{7.4}$$

or, in the case of retrieved radius,

$$D = \max_{0<x\leq r_{max}} \left| \frac{M_{r\leq x}}{M_{r\leq r_{max}}} - \frac{N_{r\leq x}}{N_{r\leq r_{max}}} \right| \tag{7.5}$$

where $M_{r\leq x}$ and $N_{r\leq x}$ are the number of measurements in each distribution with radius $\leq r$. The K-S statistic can then be used to determine the probability the two samples were taken from the same distribution, as described in Press *et al.* (1992). Table (7.4) shows both the K-S statistic and the associated probability for each combination of the twelve SPARCLE size distributions as

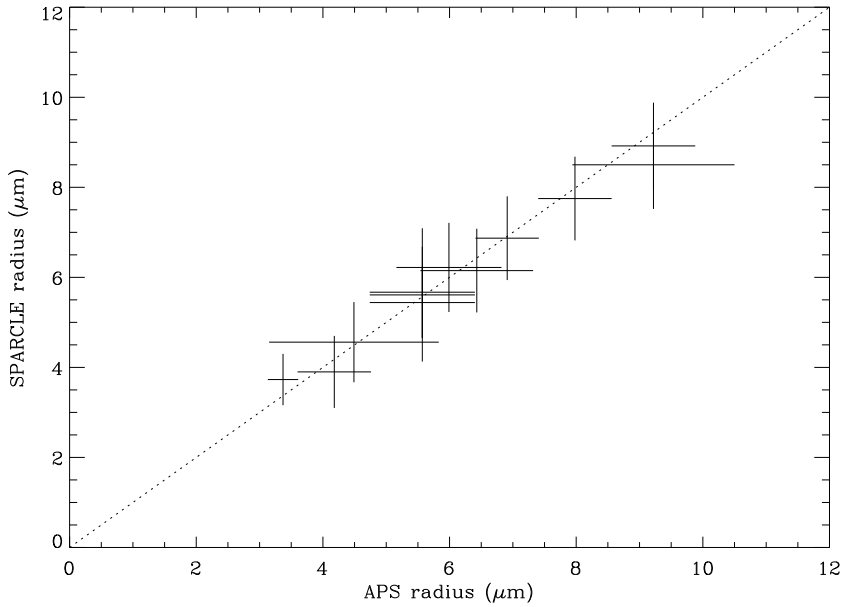


Figure 7.8: The median values of the size distributions given by 2-dimensional retrieval of SPARCLE measurements as a function of the peak radius given by the APS. The error bars on the APS measurements are the half width of the distributions, while those on the SPARCLE are the mean absolute deviation of the distributions.

generated by the 2-dimensional retrieval. Only runs (g) and (h) can be said to result in similar size distributions. This is in agreement with the results shown in Table (7.3), where it can be seen that the Gaussian fits to the SPARCLE size distributions for these two samples are by far the most similar in mean and standard deviation. Examination of the directly inferred parameters given in Figure (7.3) also shows many similarities between these two samples.

It is of some concern that the K-S test rejects the hypothesis of samples (d), (e) and (f) being from the same distribution, when both the nominal radii and half width as measured by the APS are equal for all three of these distributions. Similarly samples (k) and (l) also have identical nominal radii, but result in poorly correlated radius distributions. This problem is also apparent in Table (7.3) where the mean radii of the Gaussian fits for distributions (d), (e) and (f) are in no better agreement with each other than with those adjacent (i.e. samples (c) and (g)). This effect is, once again, due to the size resolution of the APS. As shown in Table (7.2), all three of the median radii of distributions

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)
(a)	0 1	0.23 0	0.54 0	0.88 0	0.80 0	0.87 0	0.90 0	0.89 0	0.94 0	0.97 0	0.98 0	0.99 0
(b)		0 1	0.32 0	0.74 0	0.62 0	0.72 0	0.79 0	0.77 0	0.86 0	0.96 0	0.97 0	0.97 0
(c)			0 1	0.57 0	0.40 0	0.55 0	0.63 0	0.63 0	0.81 0	0.92 0	0.92 0	0.93 0
(d)				0 1	0.22 0	0.11 0.002	0.14 0	0.12 0.001	0.41 0	0.75 0	0.83 0	0.84 0
(e)					0 1	0.20 0	0.29 0	0.58 0	0.88 0	0.88 0	0.91 0	0.93 0
(f)						0 1	0.18 0	0.21 0	0.47 0	0.75 0	0.83 0	0.84 0
(g)							0 1	0.05 0.54	0.31 0	0.66 0	0.78 0	0.79 0
(h)								0 1	0.32 0	0.68 0	0.79 0	0.81 0
(i)									0 1	0.42 0	0.64 0	0.69 0
(j)										0 1	0.35 0	0.45 0
(k)											0 1	0.14 0
(l)												0 1

Table 7.4: The results of applying the Kolmogorov-Smirnov test to all combinations of 2-dimensional retrieved radius results, subject to the cost function criteria. The top number is the K-S statistic itself, with the probability of the two distributions being sampled from the same population below. The zero entries include all values less than 0.001.

(d), (e) and (f) lie within the nominal APS size bin. Examination of the APS distributions for these samples (Figure (6.7)) reveals that they are quite different in shape, and the distributions for samples (c) and (g) peak at adjacent bins. Therefore there is little reason to expect comparison of measurements based on these samples to appear as samples from the same distribution.

The matter of the size resolution of the SPARCLE measurements as presented by the 2-dimensional retrieval is a subjective problem. If we assume that the size distributions produced are Gaussian (remembering that Figure (7.7) showed that a Gaussian is a good descriptor of the overall width of these distributions) then we can say that a separation greater than 2σ is required for two identical distributions to be resolved (i.e. for the total distribution to have two distinct peaks). A 4σ separation, with the distributions crossing at their respective 2σ points, provides very clear separation. Given the mean absolute deviations of the radius distributions, as well as the σ values of the Gaussian fits presented in Table (7.3), a conservative estimate for the minimum radius separation of two monodisperse aerosol distributions for them to be clearly resolved would be $\sim 3-4\ \mu$m.

7.2 3-Dimensional retrieval

Due to computation time constraints, full 3-dimensional retrievals were only carried out on 500 measurements from each measurement run. The states had been pre-sorted based on how well their smoothed LDA profiles agreed with one computed from the nominal state as measured by the APS. Thus these retrievals can be considered to be of the 500 best measurements from each run, given that the same PMT signal was used for all measurements in each run and the relative stability of the FFT parameterization of the LDA profile's fine structure. The absence of the 'worst' measurements from this analysis must be taken into account when comparing these results with those from the 2-dimensional retrieval.

Figure (7.9) shows the cost function distribution for retrieved states from the 3-dimensional retrieval runs, along with the cumulative error plots (cf. Figure (7.1)). Unlike the cost function distributions produced by previous retrieval runs, there is no clear bimodal structure apparent in Figure (7.9). However, most states which are within two standard deviations (based on the state covariance

calculated during the retrieval) have a cost below ten and almost all states which are within one standard deviation fall below a cost of 2. There is also some similarity between the shape of the cumulative error plots between the 2 and 3-dimensional cases, with most of the increase in both the RMS and maximum absolute errors occurring between costs of 1 and 10. Given these facts, and to allow a more direct comparison between the data sets, the same cost function cut off value (two) has been applied to the 3-dimensionally retrieved states as was used in the 2-dimensional case. However, it must be emphasized that this selection cannot be justified from the final cost function distribution of the 3-dimensional retrievals alone.

The retrieved size distributions for each measurement run are given in Figure (7.10), along with the distributions measured by the APS and those resulting from applying the cost function selection criteria (cf. Figure (7.3)). Clearly the retrieval of particle size in the 3-dimensional case is not as successful as in the 2-dimensional case. The retrieved size distributions are broad and have long tails in the direction of increasing particle size. It is also clear that selecting states based on the cost function cut off does not greatly improve the distribution. For the smaller particle sizes there is no clear peak in the distribution near the expected value, although it must be remembered that the APS size bins at the smaller radii are much smaller than those at larger particle sizes (see Table (6.1)). There is a general improvement of retrieval performance with increasing particle size however, with a maxima in the distributions near the expected radii apparent for particles with $r \geq 5.57$ μm. The large peak seen at the largest particle size bin in most of the plots is an artefact of the binning scheme as the last bin contains all particles with radii $\gtrsim 18.3$.

Figure (7.11) shows the distribution of the real part of the refractive index. Both of these distributions are much more encouraging than the size distributions shown in Figure (7.10). There is a strong peak at the expected value and the application of the cost function criterion removes many of the outlying values. Unlike the 2-dimensional retrievals, the peak of the real refractive index distribution is centred at the expected value rather than the *a priori* one. Figure (7.12) shows the real refractive index distributions produced by the 2-dimensional retrieval using the same data as utilized in the 3-dimensional case. This distribution shows that it is the absence of the poor measurements

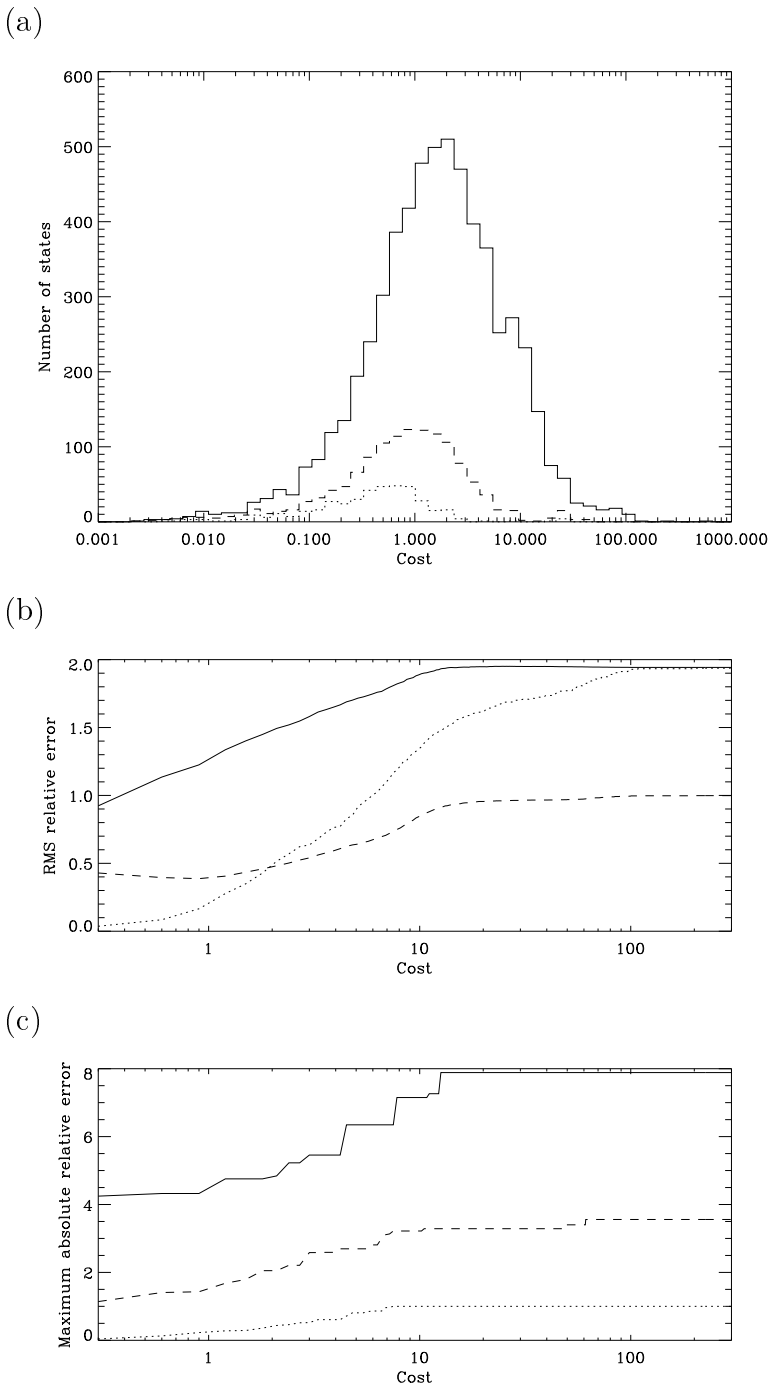


Figure 7.9: Retrieved state cost function distribution for 3-dimensional run. The three distributions shown in (a) are: all retrieved states (solid line), states which agree with the nominal state within the retrieved error estimate (dotted line) and states which agree to within twice the error estimate (dashed line). In plots (b) and (c) the solid line shows the particle size deviation and the dashed the refractive index deviation. The relative error for the refractive index values has been scaled by a factor of 10 in plot (c).

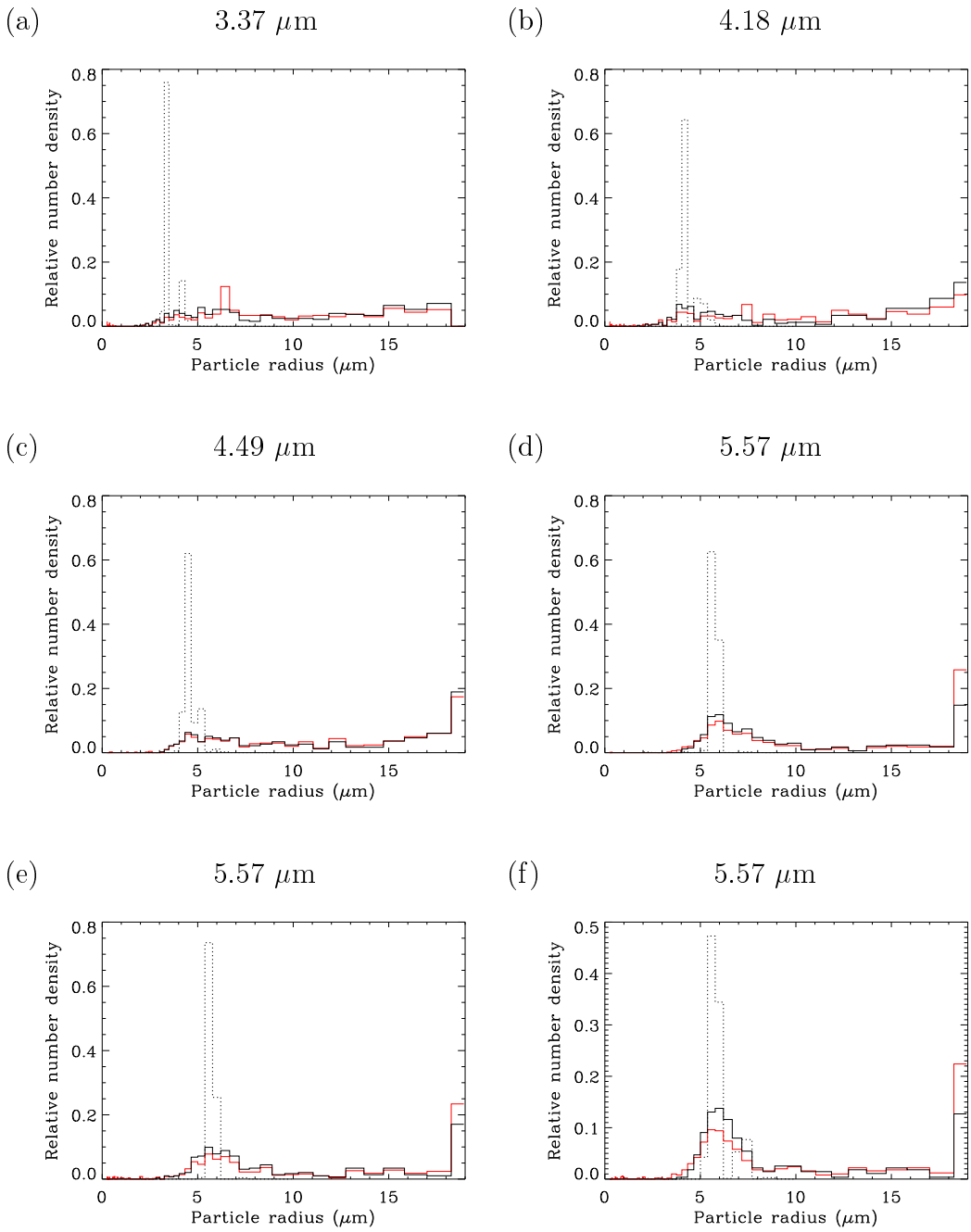
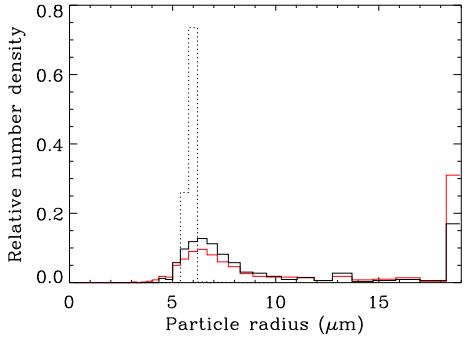
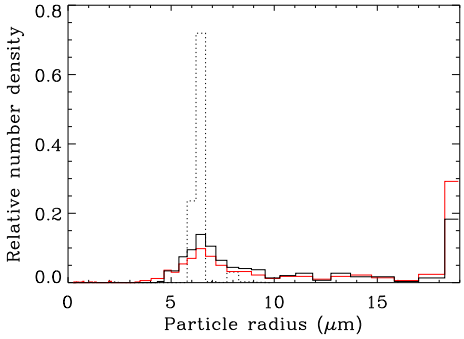


Figure 7.10: Retrieved radius distributions from the 3-dimensional retrieval run, plotted using the APS size bins. The distribution of all retrieved states is shown in red, those states which meet the cost function criteria are plotted in black and the APS measured distribution is shown with the dotted line. The distributions are normalized by the number of states they contain.

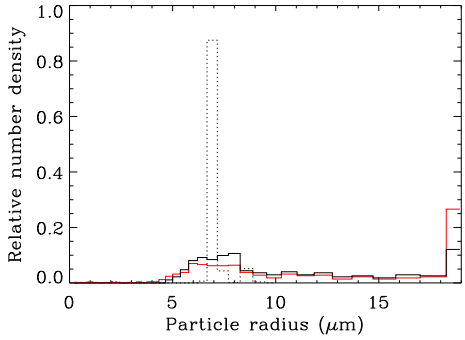
(g) 5.99 μm



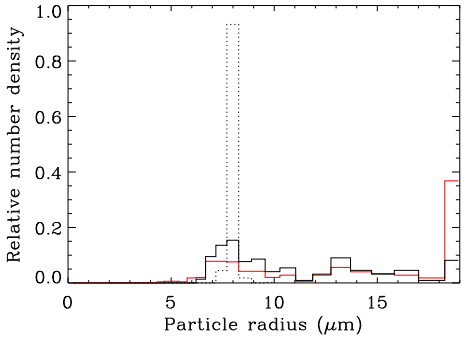
(h) 6.43 μm



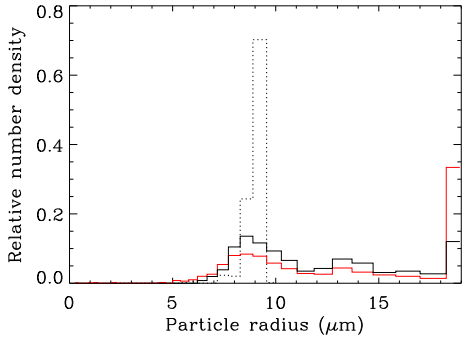
(i) 6.91 μm



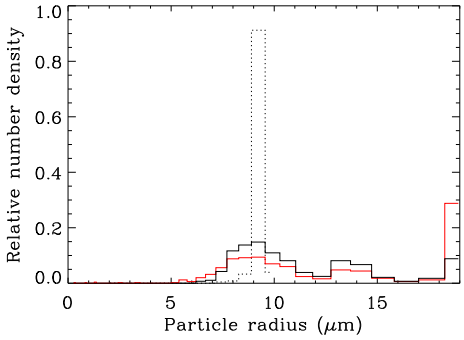
(j) 7.98 μm



(k) 9.22 μm



(l) 9.22 μm



from the 3-dimensional retrieval runs which has resulted in the better agreement between retrieved and expected refractive indices. The truncated 2-dimensional data set is very similar in form to the 3-dimensional one, particularly once the cost function criterion has been applied.

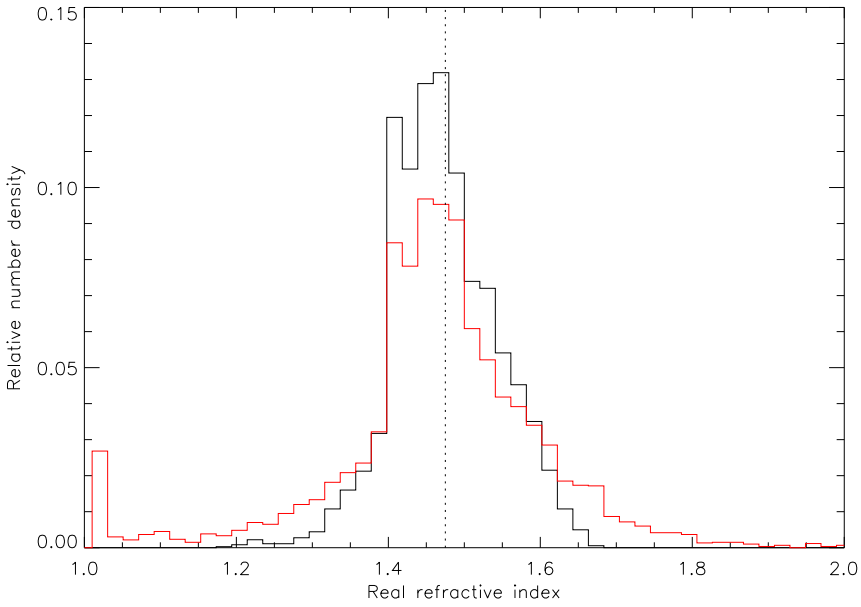


Figure 7.11: The distribution of retrieved values of the real part of the refractive index for the 3-dimensional retrieval runs. As with the size distributions (Figure (7.10)) all the retrieved states are plotted in red and those meeting the cost function criteria are shown in black. The distributions are normalized by the number of states they contain and the expected value is denoted by the dotted line.

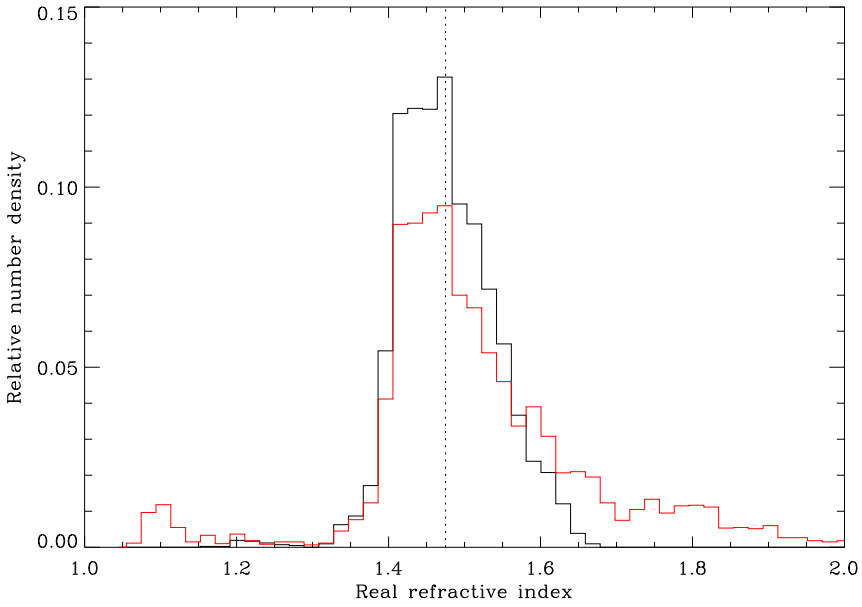


Figure 7.12: The distribution of real refractive index produced by the 2-dimensional retrieval using the same measurement used in the 3-dimensional retrievals.

Figure (7.13) shows the retrieved distribution for the imaginary part of the the refractive index – for clarity a logarithmic x-axis has been used. It is of interest that the imaginary refractive index shows generally good agreement with the expected value and that the real refractive index is similarly distributed to the 2-dimensional results, whereas the retrieved radii are much more spread in the 3-dimensional case. This suggests that the large spread in the radius distributions must be largely due to small fluctuations in the retrieved imaginary refractive index. Clearly, if a particle absorbs a portion of the incident light the amount scattered will be reduced. Thus, to produce a given PMT signal, an absorbing particle would have to be larger than a non-absorbing one. Examination of Figure (7.13) shows that most states with an absorption $\gtrsim 10^{-3}$ have poor radius estimates and virtually all of the the states with an absorption $\gtrsim 10^{-2}$ also have very inaccurate radius estimates.

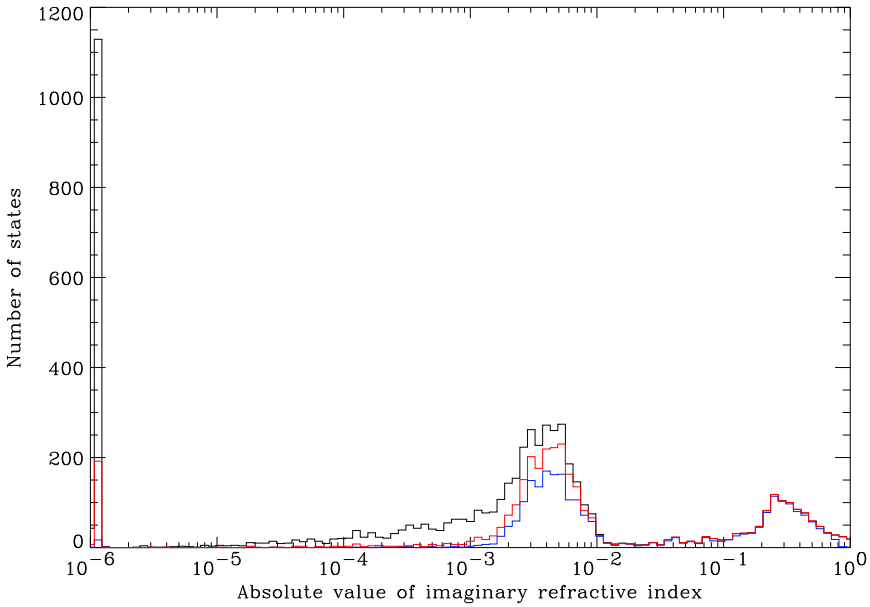


Figure 7.13: The distribution of retrieved absolute values of the imaginary part of the refractive index on a logarithmic scale. The black line shows the distribution for all states, the red and blue lines show the distributions for states where the retrieved and expected radii disagree by more than 2 and 5 μm, respectively.

The retrieved states are plotted in Figures (7.14) (radius), (7.15) (real part of the refractive index) and (7.16) (imaginary part of the refractive index). As with

Figures (7.5) and (7.6) in the previous section, the uncertainties are the standard deviations derived from the state covariance matrix calculated as part of the optimal estimation. As previously done in the 2-dimensional case, the data shown in these figures is sorted according to the maximum amplitude found in each LDA profile so that those measurements with the strongest signal are plotted first.

Brief examination of these plots shows that overall the retrieval is behaving as would be expected. The most accurate and well defined estimates of the state tend to be associated with measurements which have the strongest signal strengths. As the size of the particles increases both the particle radius and imaginary refractive index estimates become less scattered and more concentrated about their nominal values. The real refractive index for the smallest sizes tends to be clustered about the *a priori* value (1.4), with generally large error estimates, suggesting that the *a priori* is dominating the retrieval. At larger particle sizes the error estimates tend to be smaller and the retrieved values are more scattered about the expected value, with a similar spread to that observed in the 2-dimensional retrievals.

A notable difference between the scatter plots produced by the 2 and 3-dimensional retrievals is the absence in the 3-dimensional case of the unusual periodic pattern seen at smaller sizes in the 2-dimensional case (Figures (7.5a,b,c) and (7.6a,b,c)). There is some suggestion of a periodic pattern in Figure (7.14a), but it is very slight and not present in any of the other plots. This is not surprising as there is now an extra variable with which the interaction discussed in Section (7.1) can be described. The scatter plots also indicate that the inclusion of the data excluded by only using the 500 'best' measurements would not provide any significant insight into the retrieval performance. Even in the case of the largest particles the retrieved states, particularly the radii, become clustered around values far from the expected one. This suggests that these measurements are not adequate for a sucessful retrieval – the inclusion of measurements of even lower quality would be unlikely to yield any new information.

Table (7.5) shows the statistical properties of the size distributions given by the 3-dimensional retrieval runs. The best agreement between the APS and SPARCLE distributions is given by the location of the peak size in the

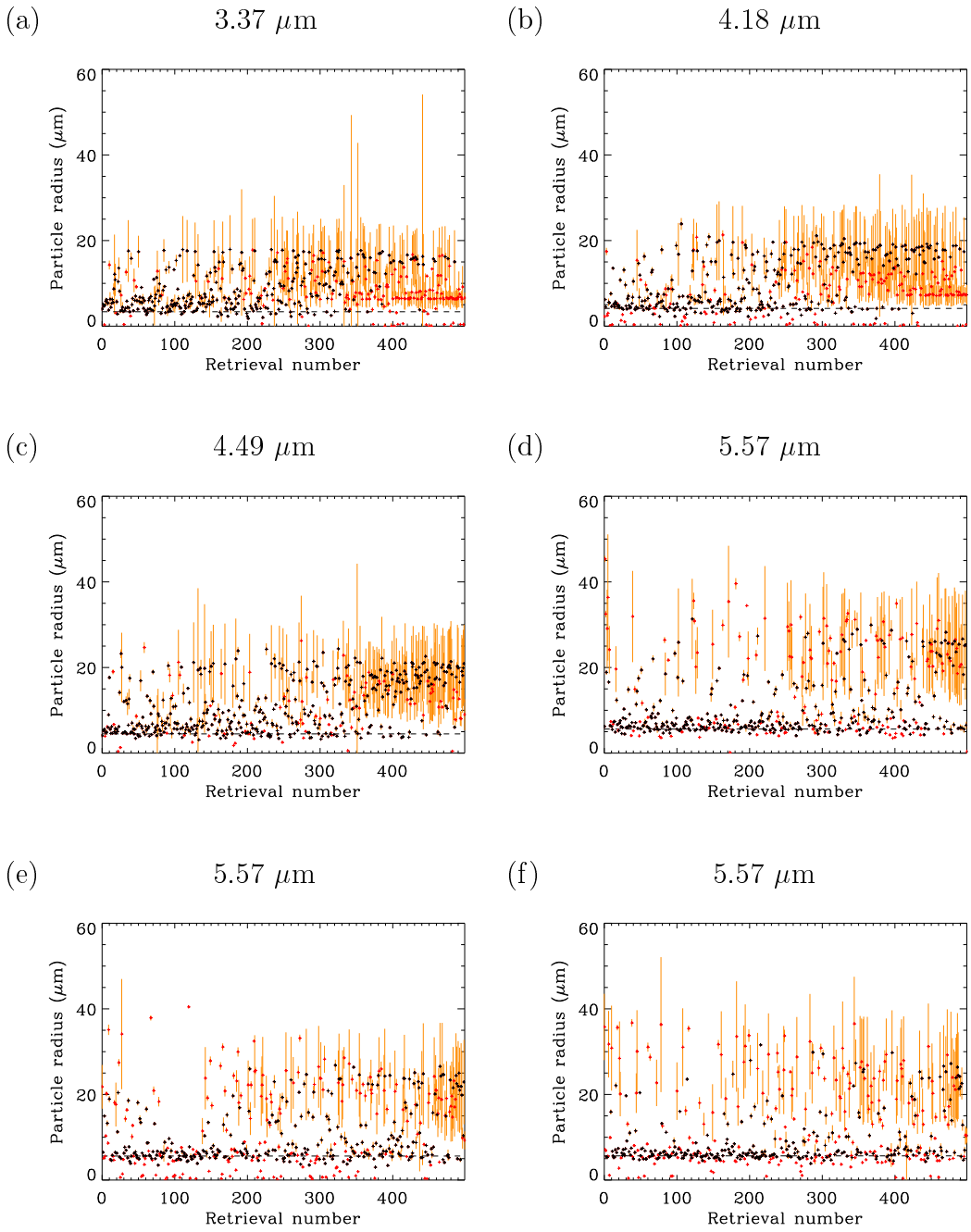
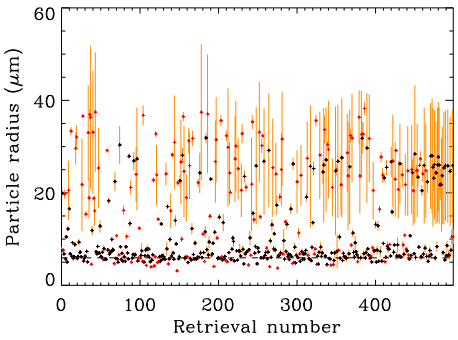
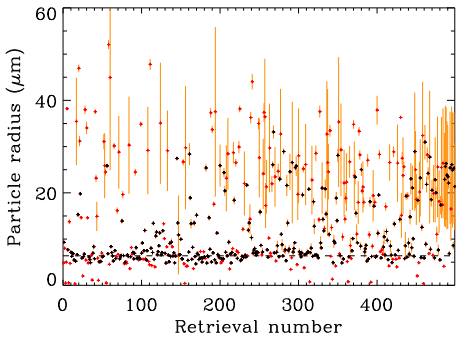


Figure 7.14: Retrieved radii from 3-dimensional runs. States which met the cost function criteria are plotted in black, while those failing the criteria are red. The error estimates given by optimal estimation are shown in orange and the nominal state is denoted by the dashed line.

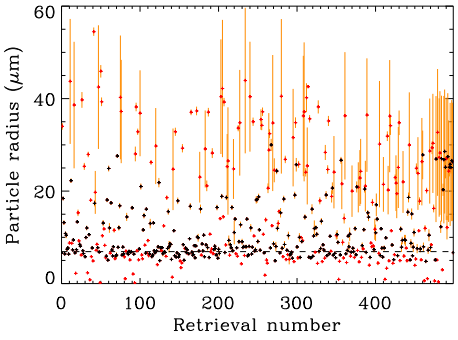
(g) 5.99 μm



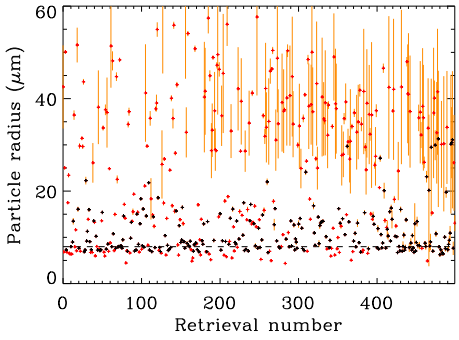
(h) 6.43 μm



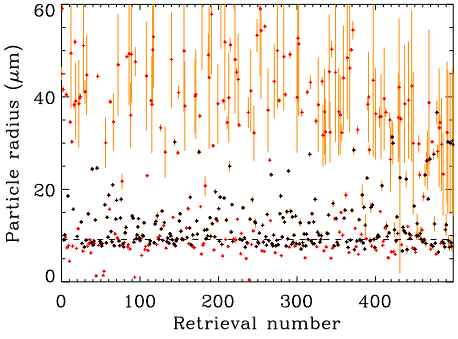
(i) 6.91 μm



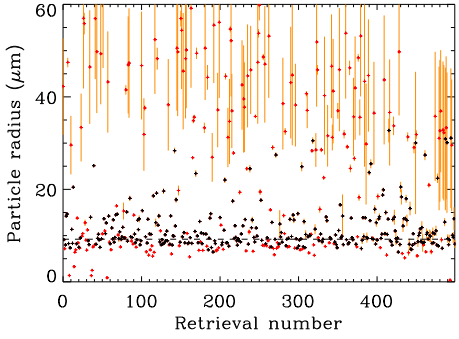
(j) 7.98 μm



(k) 9.22 μm



(l) 9.22 μm



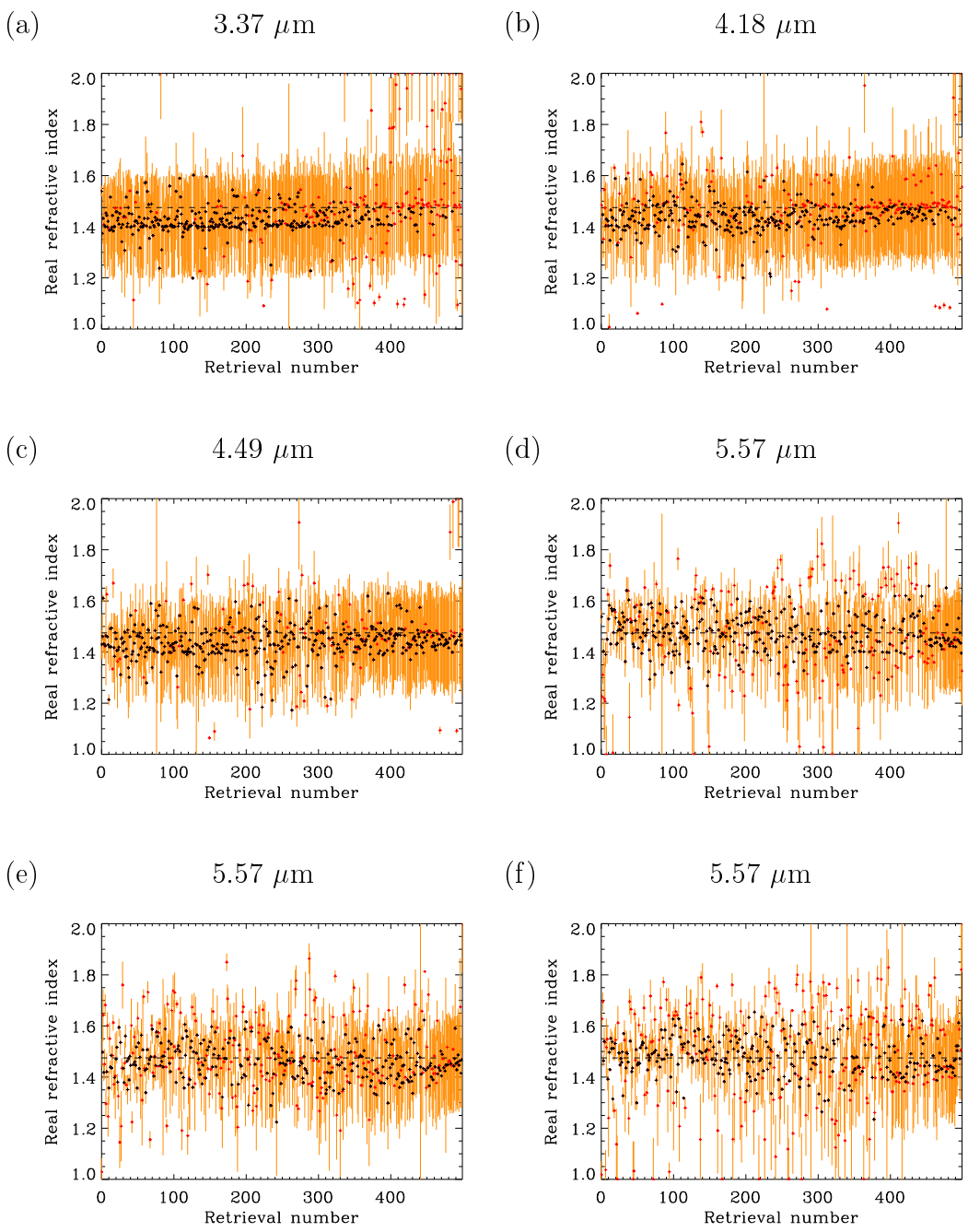
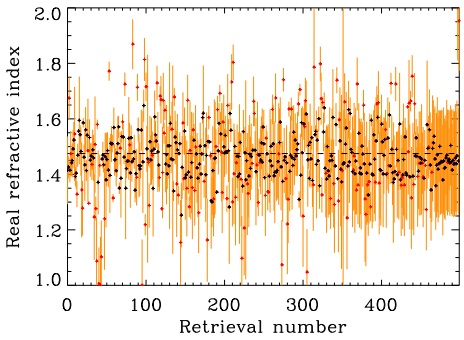
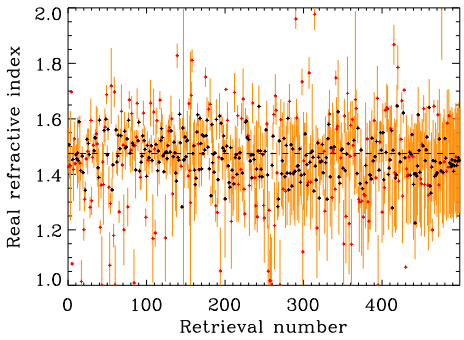


Figure 7.15: The same as Figure (7.14) but showing the retrieved values of the real part of the refractive index for the 3-dimensional runs.

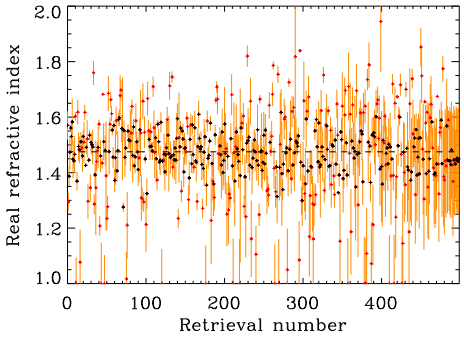
(g) 5.99 μm



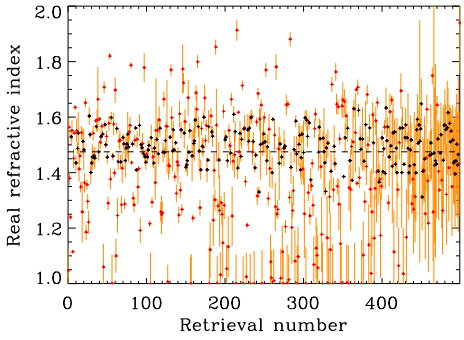
(h) 6.43 μm



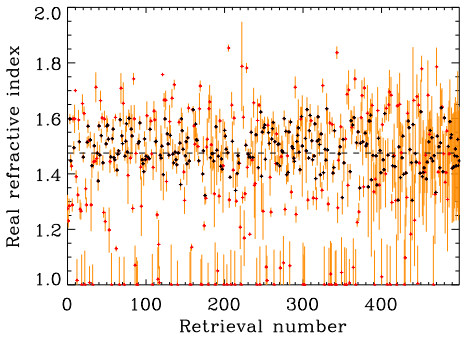
(i) 6.91 μm



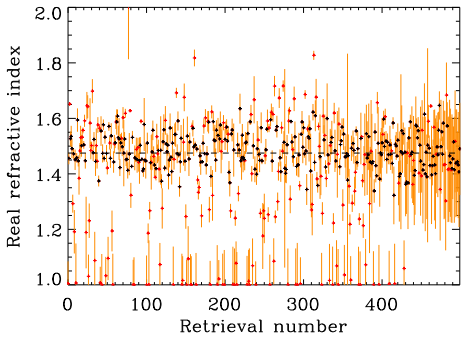
(j) 7.98 μm



(k) 9.22 μm



(l) 9.22 μm



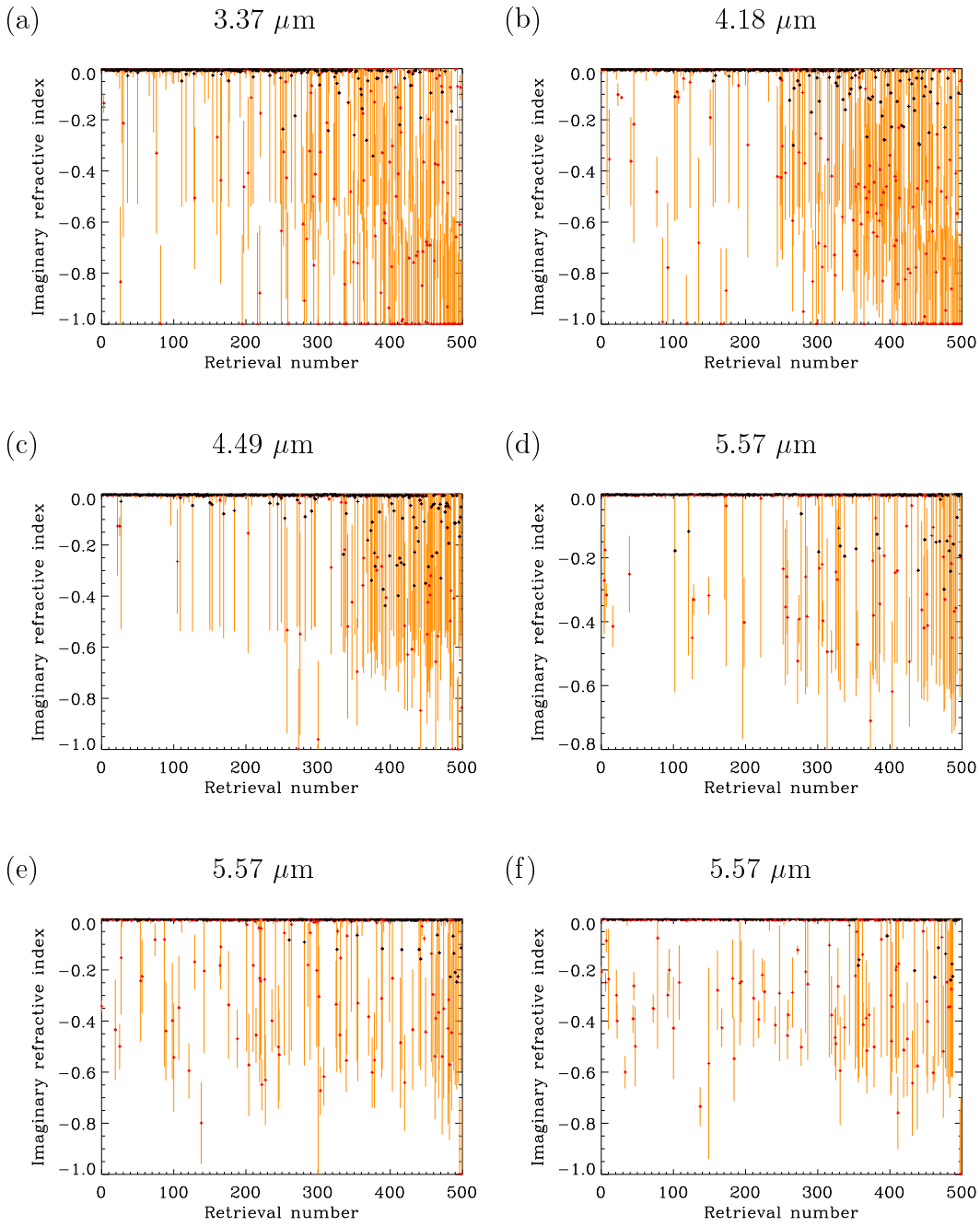
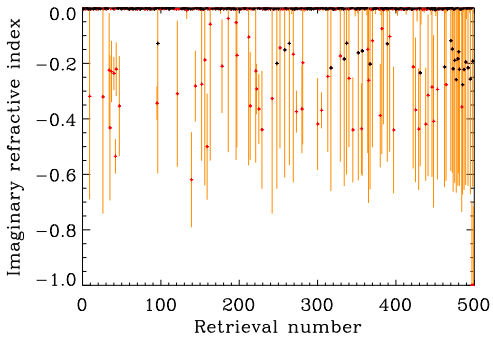
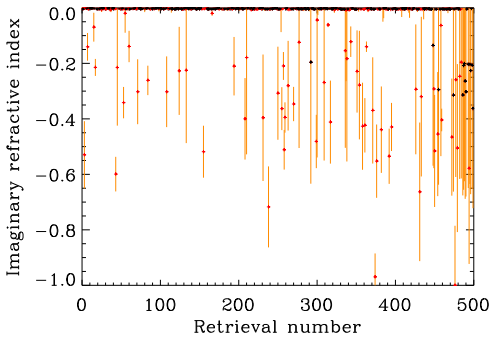


Figure 7.16: The same as Figure (7.14) but showing the retrieved values of the imaginary part of the refractive index for the 3-dimensional runs.

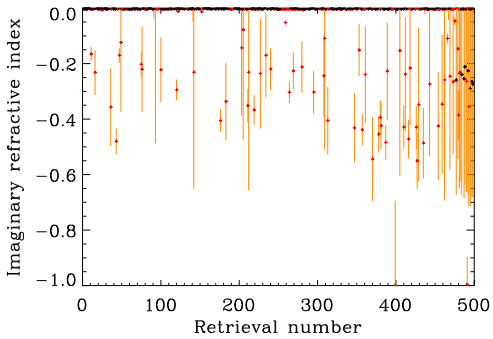
(g) 5.99 μm



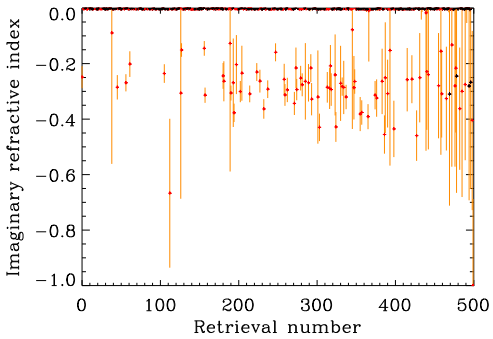
(h) 6.43 μm



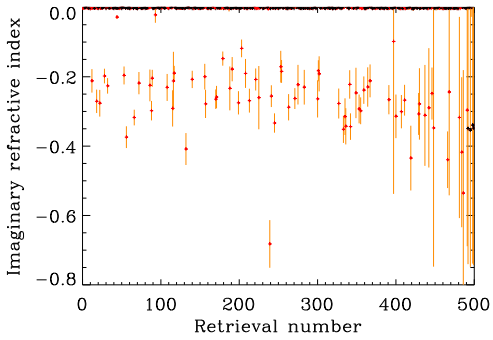
(i) 6.91 μm



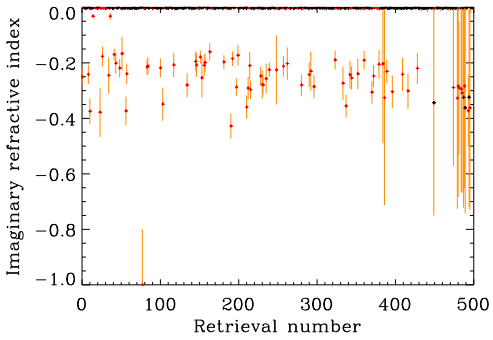
(j) 7.98 μm



(k) 9.22 μm



(l) 9.22 μm



	APS radii (μm)		SPARCLE radii (μm)					
	Nominal	$\frac{1}{2}$ width	Peak	Mean	Median	σ	MAD	$\frac{1}{2}$ width
(a)	3.37	0.24	15.3	8.67	6.79	4.86	4.27	15.6
(b)	4.18	0.58	3.89	10.2	7.50	6.04	5.60	3.92
(c)	4.49	1.34	4.49	10.9	8.83	6.17	5.53	2.84
(d)	5.57	0.83	5.99	10.2	7.13	6.73	5.28	4.24
(e)	5.57	0.83	5.57	10.1	6.74	6.54	5.40	4.24
(f)	5.57	0.83	5.99	9.32	6.50	6.28	4.68	3.05
(g)	5.99	0.83	6.43	10.4	7.07	7.23	5.58	3.27
(h)	6.43	0.89	6.43	10.7	7.22	6.79	5.43	3.78
(i)	6.91	0.50	7.98	10.6	7.96	5.81	4.47	12.9
(j)	7.98	0.58	7.98	11.2	9.10	5.17	3.81	11.6
(k)	9.22	1.28	8.58	12.4	10.1 $\leftarrow$	5.64	4.15	11.1
(l)	9.22	0.66	9.22	11.6	9.75 $\leftarrow$	4.87	3.40	7.55

Table 7.5: Comparison between the APS measured size distributions and those given by the 3-dimensional retrieval of SPARCLE measurements. Both the nominal APS and peak SPARCLE radii are the central radius of the APS size bin where the peak of the measured distribution occurs. States whose median value lies within the APS size bin of the nominal radius are indicated with an arrow.

binned distributions. In general the mean and median of the SPARCLE states show poor agreement with the APS estimates, pointing to the broad nature of the distributions and the bias towards larger particles seen in the size distributions in Figure (7.10). The large fluctuations in the half-width values for the SPARCLE distributions are due to the long tail being consistently above the peak height threshold on some distributions and below on others.

Figure (7.17) shows the the 3-dimensionally retrieved size distributions again, this time using equally spaced 0.1 μm bins (cf. Figure (7.7)). Because of the uniform binning the main peaks of the distributions are much clearer, although the long tails on the distributions are still apparent. The Gaussian fits to each distribution, the properties of which are given in Table (7.6), show interesting features. As might be expected the Gaussians in the 3-dimensional case are somewhat broader than in the 2-dimensional case (see Table (7.3)), particularly for the smaller particles. At small particle sizes the 2-dimensional radii estimates are considerably closer to the nominal values than in the 3-dimensional case, however for radii greater than 6 μm (sample (h) onwards) the situation is

reversed. For the samples (h) – (l) all of the mean values of the fitted Gaussians lie within the size range covered by the appropriate TSI size bin. There is no reason to expect the 3-dimensional retrieval to perform better at larger particle sizes, since the imaginary refractive index assumed in the 2-dimensional retrievals is known to be accurate. The exclusion of poor measurements from the 3-dimensional retrievals cannot explain this result, as applying the same binning and Gaussian fitting to the appropriate subsets does not significantly change the mean values given by the Gaussian fits (although the spread is somewhat reduced). The most likely explanation is that the spurious large radii given by the 3-dimensional retrieval are biasing the Gaussian fits towards larger particle sizes and thus offsetting a bias towards smaller radii in the main peak of the distribution.

	Nominal	Mean	σ		Nominal	Mean	σ
(a)	3.37	5.01	2.03	(g)	5.99	6.38	1.03
(b)	4.18	4.61	1.30	(h)	6.43	6.44	1.01
(c)	4.49	5.40	1.59	(i)	6.91	7.01	1.25
(d)	5.57	6.18	1.13	(j)	7.98	7.89	1.02
(e)	5.57	5.82	1.11	(k)	9.22	8.88	1.23
(f)	5.57	5.93	0.86	(l)	9.22	9.04	1.19

Table 7.6: The parameters of the Gaussians fitted to the rebinned size distributions produced by the 3-dimensional retrieval after the cost function acceptance criteria has been applied. For reference the nominal radii, as measured by the APS are also included.

Table (7.7) gives the results of applying the Kolmogorov-Smirnov test to the radii resulting from the 3-dimensional retrieval runs and subject to the cost function criteria (cf. Table (7.4)). Overall the results are similar to those found in the 2-dimensional case, with very few of the distributions being statistically similar. In this case, however, the two most similar distributions are those for samples (k) and (l), both of which have nominal radii of 9.22 μm. The reduction in the K-S statistics in the 3-dimensional case can be attributed to the increased spread of these data, and in particular, the long tail which all of the distributions have in common.

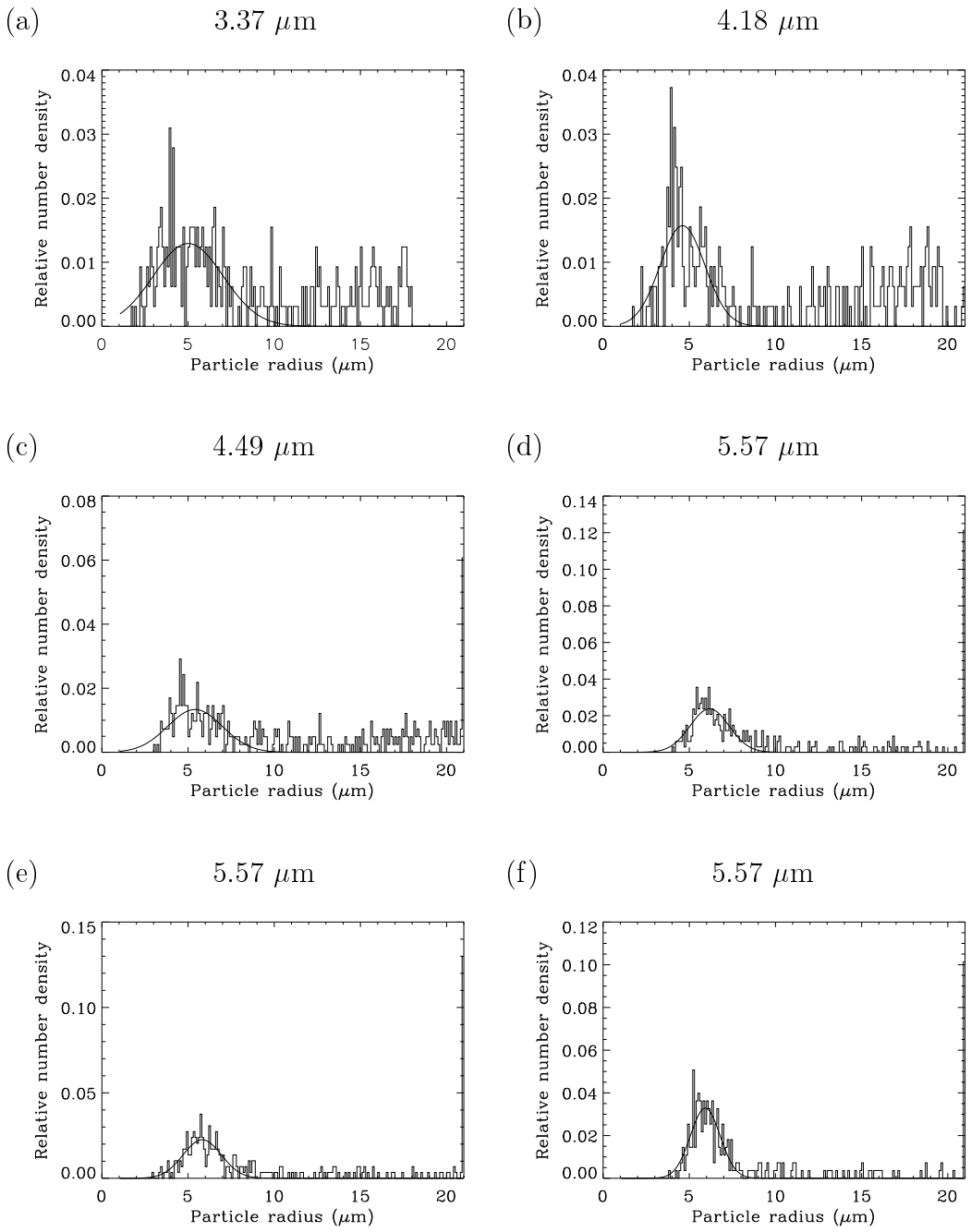
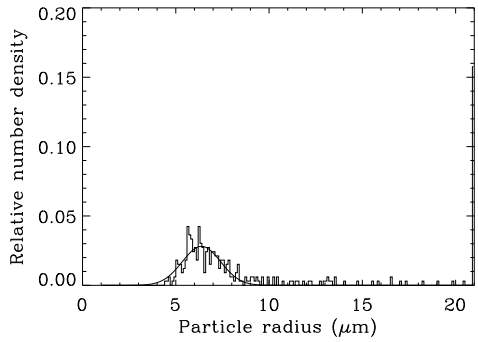
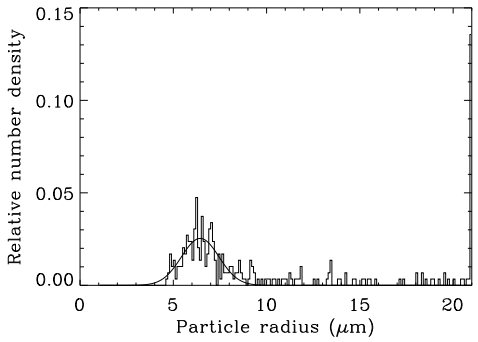


Figure 7.17: Retrieved radius distributions using uniform 0.1 μm bins from 3-dimensional run, along with Gaussian fits.

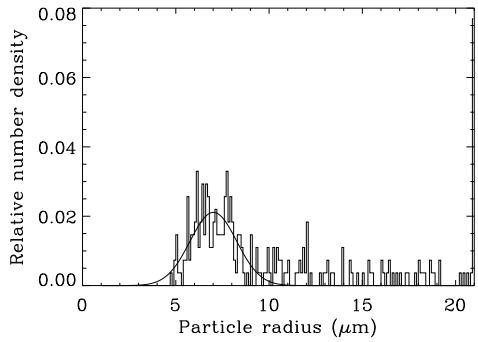
(g) 5.99 μm



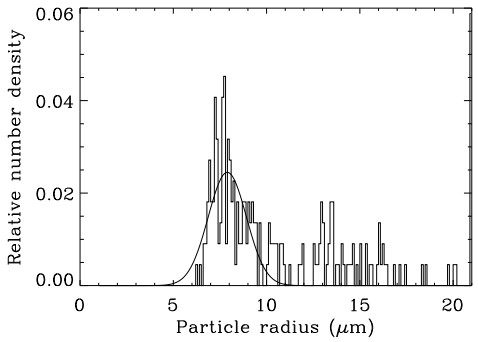
(h) 6.43 μm



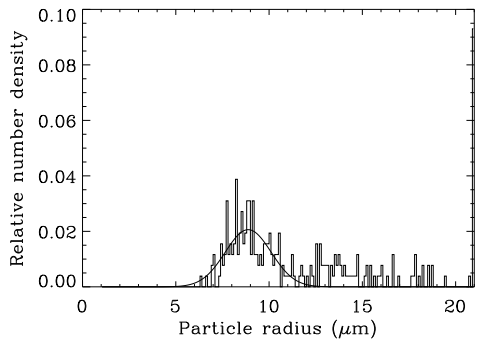
(i) 6.91 μm



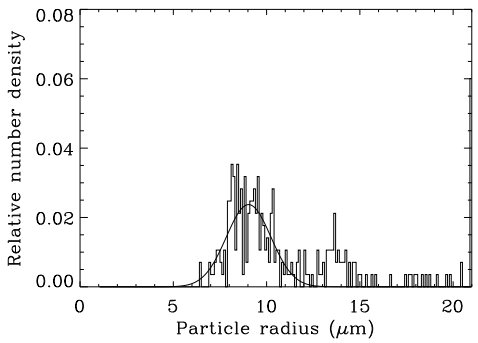
(j) 7.98 μm



(k) 9.22 μm



(l) 9.22 μm



	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)
(a)	0 1	0.17 0	0.21 0	0.25 0	0.20 0	0.25 0	0.28 0	0.28 0	0.33 0	0.48 0	0.52 0	0.52 0
(b)		0 1	0.13 0 4	0.26 0	0.21 0	0.26 0	0.28 0	0.28 0	0.32 0	0.45 0	0.47 0	0.48 0
(c)			0 1	0.17 0	0.15 0	0.24 0	0.21 0	0.18 0	0.23 0	0.38 0	0.42 0	0.41 0
(d)				0 1	0.10 0.09	0.13 0.008	0.08 0.22	0.12 0.02	0.21 0	0.43 0	0.52 0	0.53 0
(e)					0 1	0.11 0.05	0.16 0	0.18 0	0.26 0	0.49 0	0.53 0	0.53 0
(f)						0 1	0.15 0.003	0.20 0	0.31 0	0.55 0	0.63 0	0.63 0
(g)							0 1	0.09 0.14	0.20 0	0.43 0	0.56 0	0.58 0
(h)								0 1	0.16 0.002	0.39 0	0.49 0	0.49 0
(i)									0 1	0.26 0	0.40 0	0.42 0
(j)										0 1	0.24 0	0.25
(k)											0 1	0.09 0.25
(l)												0 1

Table 7.7: The results of applying the Kolmogorov-Smirnov test to all combinations of 3-dimensional retrieval results (cf. Table (7.4)). The top number is the K-S statistic itself, with the probability of the two distributions being sampled from the same population below. The zero entries include all values less than 0.001.

7.3 Summary of retrieval results

The application of the retrieval code to SPARCLE measurements has shown that measurements of the type made by the instrument do have the potential to provide reliable estimates of particle radius and refractive index. The overall behaviour of the retrieval scheme has been quite similar to that shown when applied to simulated data in Chapter (4), although the agreement with expected results has not been as satisfactory.

As would be expected, 2-dimensional retrieval (of radius and the real part refractive index, with the imaginary refractive index constrained at the expected value) has proved to be more reliable than the full 3-dimensional retrieval. In the case of the 2-dimensional retrieval, accepting only those states with a cost function value less than 2 was found to eliminate many of the outlying states and can be related to a minimum in the distribution of the costs of all retrieved states. However, it should be noted that this criterion was found to be heavily influenced by the *a priori* variance in the real refractive index. In the case of full 3-dimensional retrievals the cost function distribution was found to have no significant minima. Applying the same criteria as was used in the 2-dimensional case did result in an improvement in the distribution of retrieved states, but many states which were in poor agreement with expected results were still accepted and some states which showed good agreement were rejected.

As was found previously with simulated measurements (Chapter 4), the χ^2 value relating the retrieved state to the measurement was found to be a poor estimator of the accuracy of the retrieval. It was also found that the error estimates provided by the optimal estimation part of the retrieval code were, in general, poor estimators of the true accuracy of a given state. As with the simulated results these factors can be attributed to the high degree of non-linearity in the forward model.

Both 2 and 3-dimensional retrievals provided particle size distributions with strong peaks near the nominal radii (as measured by the APS), although the SPARCLE distributions were considerably broadened. In the 2-dimensional case the median radius for all states meeting the cost function criteria lay within the size range covered by the peak size bin measured by the APS in 50% of the retrieval runs. There was, however, a tendency for the SPARCLE

estimated radii to be lower than the APS values for larger particle sizes. Table (7.8) shows state estimates from the 2-dimensional retrieval. Each estimate is the median of the retrieved states from the appropriate retrieval run, subject to the cost function criteria. For the smaller particle sizes the refractive index retrieval appears to be biased towards the *a priori* value, suggesting that the *a priori* information is dominating the measurements for this element of these retrievals. This is confirmed by examination of the covariance of the retrieved state (as computed in the optimal estimation) which gives variances of the refractive index for these retrievals which are very similar to the *a priori* variance. Larger particles do not seem to be affected in this way however.

	APS radii (μm)		SPARCLE	
	Nominal	$\frac{1}{2}$ width	Radii, r (μm)	Refractive index, n
(a)	3.37	0.24	3.7±0.6	1.41±0.04
(b)	4.18	0.58	3.9±0.8	1.44±0.05
(c)	4.49	1.34	4.6±0.9	1.42±0.05
(d)	5.57	0.83	5.6±1.5	1.46±0.05
(e)	5.57	0.83	5.4±0.8	1.44±0.04
(f)	5.57	0.83	5.7±1.0	1.48±0.06
(g)	5.99	0.83	6.2±1.0	1.46±0.06
(h)	6.43	0.89	6.2±0.9	1.46±0.06
(i)	6.91	0.50	6.9±0.9	1.47±0.05
(j)	7.98	0.58	7.8±0.9	1.49±0.05
(k)	9.22	1.28	8.5±1.0	1.48±0.05
(l)	9.22	0.66	8.9±1.0	1.46±0.04

Table 7.8: A summary of the results given by the 2-dimensional retrieval. The SPARCLE state estimate is the median of the retrieved states, the uncertainty is the mean absolute deviation. The APS nominal radii are the central values of the peak APS size bin, the half widths are taken at 20% of the peak height. The nominal refractive index is 1.47 ± 0.01.

The 3-dimensional retrievals, in contrast, do not provide radius distributions which yield accurate estimates of the state via direct statistical methods. Although these distributions still show a peak at the expected radius, they also show large tails in the direction of increasing radius. The outliers are not reliably removed by application of the cost function criteria and are numerous enough to

significantly offset both the mean and median of the distributions away from the peak in the distribution. These spurious values appear to be the result of the retrieval converging to states with a slightly larger (more negative) imaginary part of the refractive index. In the 2-dimensional case it was found that the mean and standard deviation of a Gaussian fit to the SPARCLE size distributions are reasonable estimators of the median and spread of the raw data. Such a technique can also be used to isolate only the main peak of the 3-dimensionally retrieved radius distributions. Table (7.8) shows Gaussian fit parameters, as well as the mean absolute deviation of the data as a whole to define the 3-retrieval radii, while the refractive index values are the median and MAD alone.

	APS radii (μm)		SPARCLE			
	Nominal	$\frac{1}{2}$ width	Radii, r (μm)	Radius MAD	Refractive Index: real, n	imaginary, k
(a)	3.37	0.24	5.0±2.0	4.3	1.41±0.04	-0.01±0.02
(b)	4.18	0.58	4.6±1.3	5.6	1.43±0.04	-0.01±0.03
(c)	4.49	1.34	5.4±1.6	5.5	1.44±0.05	-0.00±0.04
(d)	5.57	0.83	6.2±1.1	5.3	1.47±0.06	-0.00±0.02
(e)	5.57	0.83	5.8±1.1	5.4	1.45±0.06	-0.00±0.01
(f)	5.57	0.83	5.9±0.9	4.7	1.48±0.06	-0.00±0.01
(g)	5.99	0.83	6.4±1.0	5.6	1.45±0.05	-0.00±0.03
(h)	6.43	0.89	6.4±1.0	5.4	1.46±0.06	-0.00±0.02
(i)	6.91	0.50	7.0±1.3	4.5	1.47±0.05	-0.00±0.01
(j)	7.98	0.58	7.9±1.0	3.8	1.50±0.05	-0.00±0.01
(k)	9.22	1.28	8.9±1.2	4.2	1.49±0.06	-0.00±0.01
(l)	9.22	0.66	9.0±1.2	3.4	1.48±0.05	-0.00±0.01

Table 7.9: A summary of the results given by the 3-dimensional retrieval. The radius estimate and uncertainty are the mean and standard deviation of the Gaussian fitted to the binned data distribution (see Figure (7.17)). The values for the refractive index are the median of the retrieved states, the uncertainty is the mean absolute deviation. The APS nominal radii are the central values of the peak APS size bin, the $\frac{1}{2}$ widths are taken at 20% of the peak height. The nominal refractive index is $(1.47 \pm 0.01) - 0i$.

The values of the real part of the refractive index given by the 3-dimensional retrieval show very close agreement with those calculated in the 2-dimensional case, both in their median and mean absolute deviation. The values of absorption are all very close to zero and are quite precise. Clearly the retrieval scheme

is also capable of giving radius estimates which are comparable to those given by the 2-dimensional retrieval, but this requires a reliable method of removing the spurious states with large radii. No reliable method based on information provided by the retrieval itself (without knowledge of the true state) has been found, however. In the case of multi-disperse aerosol measurements the Gaussian fitting method used here would be of no use.

It seems likely that most of the problems and limitations found in applying the retrieval to the SPARCLE data presented here arise from a lack of signal strength. As described in Section (6.3), much of each LDA profile was unusable due to the particle scattering pattern becoming lost in the background signal. In many cases, particularly at small sizes, there was very little large scale variation in the measurements made by the LDA, which resulted in nearly flat smoothed LDA profiles. It seems likely that an altered LDA measurement system, based around a more evenly illuminated sample volume, with less background signal and measuring more towards the forward scatter direction would result in much improved retrieval results.

On the basis of the results presented here it can be said that the prototype SPARCLE instrument, in conjunction with the 2-dimensional version of its retrieval scheme can provide accurate estimates of the radius and the real part of the refractive index mono-disperse particles of radius ~ 5 μm or larger. The precision of these estimates is approximately ± 1 μm in radius and ± 0.05 in refractive index. Below 5 μm the retrieval of refractive index becomes unreliable, but, given reasonable *a priori* knowledge, the retrieval of radius is still viable. Inclusion of the imaginary part of the refractive index in the retrieval state vector drastically reduces the accuracy of the retrieved radius estimates, but the refractive index is accurately determined. If the particle distribution is known to be mono-disperse, accurate radius estimates can be determined indirectly from the 3-dimensional retrieval results for particles larger than ~ 5 μm.

Chapter 8

Discussion, conclusions and future work

The aims of this project were to investigate the feasibility of using angularly resolved measurements of scattered light to determine the size and refractive index of air-borne particles, and to produce a laboratory instrument to make such measurements using methods which would be suitable for use in a small, balloon-borne *in situ* instrument. This chapter will a discuss the progress towards these two goals, and provide suggestions for furthering the techniques and hardware developed here.

Although the development of the retrieval scheme used to infer particle properties from light scattering and the design of the instrument are strongly linked, they are in many ways separate tasks. Thus they will be treated separately here, before the results of bringing them together in applying the retrieval to measurements made by the SPARCLE prototype are discussed.

8.1 Instrument design

The instrument designed and constructed during the course of this work is an optical particle counter, i.e. it is an instrument which detects single particles by optical means as they pass through an illuminated sample volume. The three key components of the instrument are a light source, detectors, and a sample delivery system. Each of these systems will be described separately before the strengths and weaknesses of the instrument as a whole are discussed.

The light source selected for use in the instrument was a laser. Because angularly resolved scattering measurements are required from the instrument the illumination used must be monochromatic. The illumination also needs to be of a high intensity to provide an observable scattering signal from single aerosol particles. Finally, the light needs to intersect the sample flow in a well defined beam to provide a well defined scattering volume and minimize background illumination. A laser is well suited to meeting all of these requirements and, if a laser diode based system is used, it can be implemented in a small, autonomous instrument. For the measurements made in this work a large laboratory Argon laser was used, which provides light at a wavelength of 514.5 nm at powers of up to several Watts. The output power used here was approximately 70 mW and the laser beam was passed through a baffling system to provide an approximately 1 mm beam at the sample volume. The instrument contained a simple photodiode based laser power meter which monitored the laser intensity after it had passed through the sample volume.

The design of the SPARCLE instrument is based around two separate measurement systems. The first of these systems uses a PMT to collect scattered light across a solid angle of 0.94 steradians, centered at 90° in both polar and azimuth angles. The second uses a solid state linear array sensor (referred to in this thesis as a Linear Diode Array or LDA) to measure a 512 element profile of the scattered light across the scattering angles $45 - 135°$. The LDA provides the core information of the measurement for the retrieval of particle properties, namely a profile of the scattering pattern, but it has limitations. These shortcomings are primarily related to the sensitivity of the LDA and its time resolution. It has been found that an incident optical power on the order of 50 mW is required at the sample volume before the LDA can detect scattering from particles with radii greater than 1 μm. Also the time required to read a profile from the LDA is of the order of a few milli-seconds, which is considerably longer than the period a particle is illuminated in the sample volume. Finally, because of the large amount of information delivered by each LDA profile, it is not practical to monitor this data in real time to determine when a particle has been detected or to store all LDA profiles for post-analysis. The inclusion of the PMT in the measurement system addresses these problems in the following ways:

- The PMT provides temporal resolution that the LDA lacks and thus enables

a profile of the particle scattering to be measured as it passes through the laser beam. The PMT also provides a convenient way of determining when a particle has entered the sample volume, and thus when data from the LDA needs to be collected.

- The PMT has far greater sensitivity than the LDA. Bearing in mind the long term aim is to measure very small particles with a SPARCLE type instrument this sensitivity will be of vital importance in future instruments.
- By making a single, integrated intensity measurement with the PMT the LDA signal can be used in a normalized form without significant loss of information, which greatly simplifies the calibration of the instrument.

Both measurement systems are controlled from a PC, via dedicated microcontrollers. The measurements used in this study consist of 250 PMT intensity samples covering 3.5 ms and three LDA profiles (each one taking approximately 5 ms). The collection of three LDA profiles ensures that at least one is a measurement of the background signal, allowing it to be subtracted from the measurement to leave only the particle scattering component.

The sample air delivery system used in the SPARCLE prototype is based on the design used in the Wyoming optical particle counter (see Section (2.3.1)) and is fully described in Section (3.4). The system makes use of two pumps, one of which is located at the outlet, while the other draws in ambient air and passes it through a filter, generating a flow of clean air. The clean air is then fed into the outer layer of a coaxial pipe (and is thus called the sheath flow), which then enters the measurement volume opposite the outlet pipe, so that the sheath flow travels across the sample volume from one pipe to the other. The inner layer of the coaxial pipe leads directly to the ambient air and forms the sample inlet. If the outlet flow is higher than the sheath flow, air will be drawn into the inlet and will pass through the sample volume, surrounded by the clean air of the sheath flow. By balancing the flows, the speed of sample and sheath flows can be balanced, resulting in a stable laminar jet across the sample volume.

Numerical modelling of this system using the Fluent finite element analysis package has shown that the flows are stable across a wide range of ambient conditions. It also demonstrated that the sheath flow ensures none of the particle-laden air of the sample flow can contaminate the measurement volume

and that there is very little change in the temperature and pressure of the sample air before it reaches the sample volume.

The pumps used in the SPARCLE instrument are small gear pumps driven by small 12 V DC electric motors. In both cases the pumps have been found to have excellent linearity between motor speed and flow rate and provide reasonably stable flows after an approximately 30 minute warm up.

8.2 The retrieval scheme

Mie scattering theory allows the scattering pattern produced by a spherical particle to be accurately predicted, given the particle's size and refractive index. This relationship is highly non-linear, making its inversion a difficult problem. By using Mie theory to model the measurement a given instrument would produce (i.e. producing a forward model $\mathbf{F}(\mathbf{x})$), it is possible to reduce this problem to that of finding the global minima of the function:

$$-2 \ln P(\mathbf{x}|\mathbf{y}) = [\mathbf{y} - \mathbf{F}(\mathbf{x})]^T {\mathbf{S}_\epsilon}^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x})] + [\mathbf{x} - \mathbf{x}_a]^T \mathbf{S}_a^{-1} [\mathbf{x} - \mathbf{x}_a]$$

(Equation B.8).

Traditional numerical approaches to solving this type of problem have been found to be inadequate. A new retrieval scheme was developed which combines the features of simulated annealing and the Levenburg-Marquardt optimal estimation method. The new method first applies a simulated annealing algorithm to the problem. Of all the states tested during the annealing, the one hundred that have the lowest costs and are separated by at least some minimum distance in state space, are stored and used as starting points for optimal estimation runs. From these one hundred Levenburg-Marquardt runs the state estimate with the lowest cost is chosen as the solution. The retrieval scheme was tested using both simulated SPARCLE measurements and data from the DAWN-A multiangle aerosol photometer.

8.2.1 Retrieval of simulated SPARCLE measurements

The design used in the SPARCLE instrument provides two measurements of particle scattering, the intensity of scattered light integrated across a broad solid angle (the PMT measurement) and a 512 element profile of the scattering pattern covering a 90° range of scattering angles (the LDA measurement). It was found that the number of local minima in the cost function could be reduced by applying the method of Steiner *et al.* (1999) to the LDA measurement. This method uses the Fourier transform of the scattering pattern to characterize its fine structure, which produces a single peak on the Fourier spectra. In the case of the SPARCLE measurement the fine structure was characterized by a single value, the location of the Fourier peak. The measurement vector was completed with a 30 element smoothed profile of the LDA measurement (i.e. without the fine structure), normalized to its average amplitude, and the intensity measured by the PMT.

To provide a quantitative measure of the information contained within the measurement vector the Shannon information content was computed for both the raw intensity measurements and the measurement vector described above. It was found that both forms of the measurements contained either two or three independent pieces of information, depending on the radius and refractive index of the particle in question. Particle radius was found to be the most important factor in the information content of a given measurement. In general the information content was found to fall rapidly for particles with radii less than ~ 0.2 μm and there was also a drop for particles with radii in the order of 10 μm or more, particularly for particles with high values of the real part of the refractive index. The decrease in information content for small particles can be attributed to the scattering beginning to resemble Rayleigh scattering, with a subsequent loss of angular variation in the scattering pattern. At large particle sizes the fine structure of the scattering pattern becomes small enough for aliasing to become a problem, due to the finite sampling width of the LDA. There is also a marked reduction in the information content of absorbing particles, which can be attributed to the reduction in the amplitude of the fine structure of the scattering patterns as the imaginary part of the refractive index becomes larger. Comparison of the information content of the raw intensity measurements and the measurement vector used in the retrieval code shows that the latter has a slightly lower information content, with the effect being most

notable for absorbing particles.

To test the performance of the retrieval scheme it was applied to 500 simulated measurements generated by the SPARCLE forward model. The state vectors of the simulated particles were randomly generated from a bimodal, log-normal size distribution, with one mode representing non-absorbing atmospheric particles (refractive index $= 1.43 - 0.0001i$ to $1.5 - 0.0001i$) and the other representing absorbing atmospheric particles (refractive index $= 1.6 - 0.66i$ to $1.8 - 0.66i$). Measurement error was simulated by adding 2% Gaussian noise to the raw LDA and PMT signals. Two versions of the retrieval scheme were applied to these simulated measurements. The first used a 2-dimensional state vector made up of the particle radius and the real part of the refractive index, and assumed all particles were non-absorbing. The second used a 3-dimensional state vector which included the imaginary part of the refractive index.

Although the retrieval provided quite accurate estimates of the states for many of the simulated measurements used, the highly non-linear nature of the forward model resulted in poor characterization of the state under the linear approximation used in the Levenburg-Marquart method. Both estimates of the uncertainty of the retrieved state and the χ^2 (denoted as χ^2_s) relating the forward model at the solution to the measurement were found to bear little relationship to the true accuracy of the retrieval. Examination of cost functions in the 2-dimensional retrievals revealed that the global minimum was, in general, very closely surrounded by several deep minima. With the addition of measurement noise one or more of these secondary minima could become deeper than the true solution. Even without this occurring it is possible that the minimization procedure would not converge to the global minimum, but rather one of the deep local minima – the possibility that the true state may not correspond to the global minimum of the surface makes this even more likely. Since the optimal estimation uses a local linear approximation of the cost function it is unable to account for this possibility.

Without prior knowledge of the true state it is not possible to distinguish between convergence to the true solution and convergence to a deep, nearby local minimum, as both will result in forward model output which is very close to the original measurement. It was found that the depth of these minima generally

decreased as they became more distant from the true solution. Thus the value of the cost function (the computation of which does not involve any assumptions about the form of the forward model) at the final solution was found to be a good estimator of the accuracy of a given retrieval. In both the 2 and 3-dimensional cases the solution cost was found to have a clear bimodal distribution with one mode comprised of generally accurate state estimates and the other of inaccurate ones. Acceptable states were thus selected by applying a cost function criterion. In the case of the 3-dimensional retrieval it was found that approximately 70% of the retrieved states meet the cost function criteria, and of these 96% lie within a 5% margin of error of the corresponding true states.

The 2-dimensional retrieval was applied to reveal whether removing one independent variable from the state vector would improve the performance of the retrieval for states where the assumed absorption was close to the true value. As was expected the retrieval performed poorly for those particles with a large imaginary part of the refractive index. The particles in the absorbing mode were all but absent from the states which met the cost function criteria, with the exception of particles with radii less than ~ 0.3 μm. The cost functions of these small particles show very little sensitivity to the refractive index, thus an accurate radius is retrieved while the measurement adds little to the *a priori* knowledge of the refractive index. It was also found that the 2-dimensional retrieval did not provide a better estimate of the states for which the assumed absorption was accurate. Thus, in the case of these simulations, the measurement contained sufficient information to allow the retrieval of the full 3-dimensional state.

8.2.2 Retrieval of DAWN-A measurements

The retrieval code was further tested on data from the DAWN-A multiangle aerosol photometer. The DAWN-A instrument uses a differential mobility analyser to select particles of a given size which are then fed into a light-scattering chamber. This chamber allows measurement of scattered intensity at 9 different scattering angles along 8 separate azimuth planes. Measurements are made by up to 14 photomultiplier tubes, via optical fibre feeds.

The data analysed in this study were measured during the SEAVS atmo-

spheric visibility study carried out at Look Rock, Tennesse, USA, in 1995 and is presented in Dick (1998). For the SEAVS campaign the DAWN-A made intensity measurements of ambient atmospheric aerosol at 7 scattering angles spaced between 40 and 140°. In the original data analysis the DMA gave the particle size directly and the imaginary part of the refractive index was inferred from independent measurements. The averaged scattering pattern measured for many particles of the same size was then used to provide an estimate of the real part of the refractive index by a least squares fit of a Mie scattering pattern. Thus, the data presented by Dick (1998) included complete estimates of the particle state.

In this study the averaged scattering patterns (as used in the linear least squares fitting during the original DAWN-A data analysis) were used as measurements for a modified version of the SPARCLE retrieval scheme. In this version the raw intensities formed the measurement vector, since the measurements were too sparse to allow generation of the SPARCLE style measurement vector described in the previous section. The data consisted of 12 separate scattering patterns for particles with radii between 0.15 and 0.40 μm. There were four separate runs of the retrieval performed:

- 2-dimensional retrieval (radius and the real part of the refractive index only) using measurement vectors normalized by their mean value.
- 3-dimensional retrieval using the normalized measurements.
- 2-dimensional retrieval using absolute measurement vectors and a calibrated forward model.
- 3-dimensional retrieval using the absolute measurements.

The retrieval of normalized measurements was performed because such measurements would greatly simplify the calibration of an instrument. In the case of 2-dimensional retrieval with the normalized measurements, eight of the twelve retrieved states agreed with the original estimates of radius and the real part of the refractive index within the uncertainty estimates given by the retrieval. The performance of the 3-dimensional retrieval with normalized measurements was poor, with only one retrieved state agreeing in all three of the state elements within uncertainties. The performance with absolute measurements was considerably better. The 2-dimensional retrieval returned the expected state

for all but one of the measurements and the 3-dimensional results agreed in all but 3 cases.

The relatively poor performance of the retrieval in the case of normalized measurements can be attributed to the low angular resolution of the DAWN-A measured scattering pattern. With only 6 or 7 point measurements defining the shape of the scattering profile it is possible to fit many different scattering patterns, especially when there are three degrees of freedom in the state. The use of absolute measurement vectors meant that the amount of scattered light was constrained as well as its shape, which removed much of the degeneracy present in the normalized measurement vectors, resulting in an improvement in retrieval performance.

An interesting feature of the DAWN-A retrievals was that both the error estimates and the χ_s^2 comparing the retrieved state to the measurements were found to behave much more as would be expected from retrieval theory. In all cases, apart from the 3-dimensional retrieval of normalized measurements, the χ_s^2 value is smaller than the expected value (being approximately the number of elements in the measurement vector) and there is some correlation between the value of the χ_s^2 and the discrepancy between SPARCLE retrieval and DAWN-A results. The relative sucess of the χ_s^2 value as an estimator of the retrievals accuracy in comparison with the results from simulated SPARCLE measurements can be attributed to the simpler forward model required for the DAWN-A instrument and the small particle sizes involved. Both of these factors will reduce the non-linearity of the forward model and the linear approximation can be expected to perform better. The fact that the χ^2 is generally smaller than might be expected could be explained by the measurement vector containing less pieces of independent information than the number of elements would suggest – remembering that the Shannon information content of a 512 element measurement across a similar angular range gave two or three independent measurements.

Restricting ourselves to the more successful retrieval using absolute measurements, it can be said that the SPARCLE retrieval has been shown to provide reliable estimates of the state based on DAWN-A measurements of particles in the size range 0.15 to 0.40 μm. 2-dimensional retrieval was found to be the most robust. The output of the 2-dimensional retrieval scheme agreed with

the original DAWN-A results within uncertainties for all but one of the twelve measurements. The one standard deviation uncertainties on the retrieved radius in this case were all ≤ 0.05 μm while those on the refractive index were ≤ 0.1 for particles with radius ≥ 0.15 μm. The 3-dimensional retrieval provided results that were in agreement with the DAWN-A estimates in 9 out of 12 cases, however the uncertainties, particularly on the refractive index values, were large.

DAWN-A measurements analysed here do not define the scattering pattern well enough to allow the properties of the measured particles to be completely defined, but with an accurate estimate of the imaginary part of the refractive index the SPARCLE retrieval scheme has been able to provide accurate and precise estimates of particle radius and the real part of the refractive index. It would be of interest to examine whether the scheme would be able to provide the full refractive index if the particle radius were fixed, since the DMA used in DAWN-A measures the radius explicitly, however this work lies beyond the scope of this study.

8.3 SPARCLE measurements

The SPARCLE prototype has been used to make measurements of twelve separate pseudo mono-disperse glycerol aerosol distributions. These distributions were generated by a Model 3450 Vibrating Orifice Aerosol Generator (VOAG) manufactured by TSI Incorporated and had peak radii between 3.4 and 9.2 μm. Size measurements of the aerosol, made concurrently with the SPARCLE measurements, with a TSI Model 3320 Aerodynamic Particle Sizer (APS) allowed the aerosol state to be well-characterized independently of the SPARCLE measurements. The APS measurements revealed the VOAG generated aerosol distributions to be between 0.5 and 2.6 μm wide (full width at 20% of maximum) with a value of approximately 1.6 μm being most common.

Analysis of the SPARCLE measurements from these runs has revealed three main problems with the SPARCLE design:

- The measurements show a large number of coincidence errors. The long dead time between measurements dictated by both the slow read out rate of the LDA and the time taken to transfer the data from the microcon-

trollers to the data collection PC resulted in a significant proportion of measurements (up to 42%) being discarded due to two or more particles passing through the sample volume within one measurement cycle. This problem effectively makes SPARCLE unsuitable for determining particle concentrations for the concentration range produced by the VOAG (approximately 1 cm^{-1}). It should be noted, however, that the concentration of stratospheric aerosol particles, with radii above ~ 0.1 μm is well below this level.

- The background signal was found to be too high. The background light level due to Rayleigh scattering from the air and stray reflected laser light within the instrument is of a level which swamps the particle scattering pattern signal for small particles. The use of laser baffling was found to reduce the background signal to a level where LDA measurements can be made, although approximately one third of the scattering pattern is still below the noise threshold for even the largest particles.

- The PMT signals were found to be ambiguous. The need for the laser baffling used to lower the background light signal results in an approximately Gaussian laser beam profile, $\sim$ 1 mm in width at the scattering volume. Since the sample air flow has a diameter of 2 mm it is possible for a particle to cross the beam anywhere across its width. As a given particle passes through the laser beam it will produce a Gaussian peak in the PMT signal, the height of which will be determined by which part of the laser beam it passed through. However, since the profile is Gaussian, all of the signals measured by the PMT will have the same width, thus it is impossible to distinguish between a small particle passing through the centre of the laser beam and a larger one passing closer to the edge.

To overcome these problems and allow the SPARCLE measurements to be used by the retrieval scheme, the following alterations were made. Because each particle distribution was close to mono-disperse, the PMT measurements could be considered to be a set of random samples from a truncated Gaussian (the truncation being due to the detection threshold set by the measurement system). By comparing the measured distribution of signals to the expected distribution for such a sample a single PMT measurement, corresponding to a particle passing through the centre of the laser beam, and the associated uncertainty could be derived for each measurement run. Because of the low signal to noise ratio of

the LDA measurements it is necessary to limit the portion of the LDA signal used to form the smoothed LDA profile to approximately the first two thirds of the total profile. This results in the overall shape of the scattering pattern being poorly defined in the retrieval scheme, but was found to be preferable to the inaccuracies introduced by the full profile.

It was found that the covariance matrix for the measurement vector ($\mathbf{S}_\epsilon$) was best generated by using the measured noise on each LDA profile to randomly perturb that profile and then to use this to produce a perturbed measurement vector. By repeating this many times and calculating the expectation value of the squared difference between true and perturbed measurement vectors a covariance matrix can be formed. It was also found, however, that to ensure numerical stability, the covariance matrix must be assumed to be diagonal (i.e. all of the off-diagonal covariances set to zero, leaving only the variances).

Due to the fact that the LDA measurements are used in a normalized form, only the PMT measurements needed to be calibrated for use by the retrieval scheme. This was done by performing a linear, weighted least squares fit between the photomultiplier forward model output and PMT signals from each run, as defined above. This fit yielded a scaling factor which matched the model output to the measurements. It was found that the error estimates on the PMT signals can be fully accounted for by the variation in the expected PMT signal for the size range of the measured aerosol distributions alone.

8.4 SPARCLE measurement retrieval

As with all other applications the retrieval of the SPARCLE measurements was done in both 2 and 3-dimensional forms. The 2-dimensional retrieval was run on all non-spurious (not affected by coincidence errors) measurements, whereas, due to its less satisfactory performance and time constraints, the 3-dimensional retrieval was only performed on the 500 best states from each measurement run. For the retrieval to produce reasonable results it was found to be necessary to provide much tighter *a priori* constraints on the real part of the refractive index than had been necessary in previous cases. The *a priori* standard deviation on the real part of the refractive index used in these retrievals was 0.2, compared

with 1.0 for the previous cases. Such restrictions were not placed on the particle radius or absorption (in the 3-dimensional case), where the *a priori* constraints were very weak.

The statistical characterization of the retrieved states given by 2 and 3-dimensional retrievals was found to be poor (in agreement with the findings of the retrieval of simulated SPARCLE measurements). Once again, both the χ^2 comparing retrieved state to measurement and the state covariance estimates are poor measures of the true accuracy of a solution. As with the simulated measurement retrievals, there is a strong correlation between the cost of the retrieved state and the discrepancy with the expected state. Unlike the simulated case however, the distribution of the cost function does not show clear bimodal structure. There is some evidence of the distribution being bimodal in the 2-dimensional case, although it is apparent that this is mostly due to states disagreeing with the *a priori* rather than the measurement. However, using the cost corresponding to the minimum in the cost function as an acceptance criterion was found to remove most of the outliers from the set of retrieved states. In the case of the 3-dimensional retrieval, there is no clear minimum in the cost distribution for retrieved states. However, to enable a more direct comparison between 2 and 3-dimensional results the same cost function criterion was applied as was done in the 2-dimensional case.

The 2-dimensional retrieval (using an accurate estimate of the imaginary part of the refractive index) was found to give the best results, producing particle size distributions with strong peaks near the nominal, APS measured, radii for all particle sizes. Of the twelve retrieval runs performed, six produced median radii which are within the peak size bin measured by the APS. The SPARCLE radius distributions are considerably broadened compared to the APS measured distributions however, with the majority having mean absolute deviations of the order of 1 μm. For particles with a nominal radius greater than approximately 5.5 μm the 2-dimensional retrieval also provides a refractive index distribution centred close to the expected value. Below this size, however, the retrieved refractive index is strongly influenced by the *a priori* estimate, with most states having error estimates of the same size as the *a priori* standard deviation.

Two statistical tests were applied to the retrieved SPARCLE measurements,

although the information they have yielded is limited. Firstly a χ^2 test comparing the APS and SPARCLE size distributions was applied. Due to the small number of size bins occupied by each distribution, as measured by the APS, and the broadening of the distributions by the SPARCLE system, the χ^2 value is extremely large for all 12 comparisons. It is suggested that these values are an overly harsh measure of the level of agreement between SPARCLE and APS estimates. To provide a quantitative measure of whether the size distributions given by the SPARCLE retrieval scheme can be considered independent, a Kolmogorov-Smirnov comparison was done between all possible combinations of measurement runs. Here it is found that only one of the 78 possible pairs of distributions were at all similar at a statistically significant level.

The 3-dimensional retrieval was less successful. The radius distributions, although still showing peaks near the expected radius for most runs, are much broader, in general, and all showed long tails in the direction of increasing radius. Normal statistical averages (i.e. the mean or median) do not provide satisfactory estimates of particle size – they are strongly influenced by the high radius tails found on the distributions. Normal statistical measures of spread produce similarly unsatisfactory results for the same reason. Using the median and mean absolute deviation of the distributions results in uncertainties on the order of 5 μm and state estimates which only lie within the peak APS bin for the two largest sizes. However, it has been found that Gaussian fits to the retrieved radius distributions (using evenly spaced bins) model the main peak quite well, in both 2 and 3-dimensional cases. Using the mean and standard deviation of the fitted Gaussians as estimates of the radius provide results that lie within the peak APS bin in 5 of the 12 cases with uncertainties ranging between 0.9 and 2.0 μm.

The distribution of the real part of the refractive index produced by the 3-dimensional retrieval is very similar to that determined by the 2-dimensional version. Once again there is evidence than many retrievals are converging to the *a priori* for particles with nominal radii under 5.5 μm. For sizes above this, however, the retrieved distributions are centred very close to the nominal value and have mean absolute deviations of 0.05 or 0.06.

In all twelve cases the distribution of the imaginary part of refractive index

produced by the 3-dimensional retrieval has a peak value very close to zero. The mean absolute deviations of these distributions are 0.01 in six out of twelve cases, with a maximum value of 0.04. It has been found that the long tails observed in the radius distributions result from retrievals where the particle's absorption has been slightly over estimated – it is interesting to note that a small discrepancy in one state element can lead to very large errors in another.

The χ^2 test between the APS and SPARCLE retrieved radius distributions was not performed in the 3-dimensional case, since the agreement between the distributions is so clearly worse than in the 2-dimensional case. The Kolmogorov-Smirnov tests were performed however. In this case, three of the 78 comparisons show a significant probability ($> 10\%$) that the two distributions could be modelled as random samples from the same distribution. Thus, once again it can be said that the retrieved radius distributions are significantly different from one another.

8.5 Summary

On the basis of the results presented here it can be concluded that the prototype SPARCLE instrument, in conjunction with the 2-dimensional version of its retrieval scheme can provide accurate estimates of the radius and the real part of the refractive index of mono-disperse particles of radius ~ 5 μm or larger. The precision of these estimates is approximately ± 1 μm in radius and ± 0.05 in refractive index. Below 5 μm the retrieval of refractive index becomes unreliable, but, given reasonable *a priori* knowledge of the refractive index, the retrieval of radius is still viable. Inclusion of the imaginary part of the refractive index in the retrieval state vector drastically reduces the accuracy of the retrieved radius estimates, but the refractive index is accurately determined. If the particle distribution is known to be mono-disperse, an accurate radius estimate can be determined indirectly from the 3-dimensional retrieval results for particles larger than ~ 5 μm. At present the SPARCLE instrument is unable to make useful measurements of poly-disperse aerosols due to ambiguity in the PMT measurement.

Despite the shortcomings of the instrument, the results presented here provide

proof of concept for the measurement system employed by the SPARCLE system and show that the instrument warrants further development. Application of the retrieval code to simulated measurements, DAWN-A data and SPARCLE measurements have clearly demonstrated the feasibility of using single particle Mie scattering to characterize not only the number density and size distribution of an aerosol population, but also to provide refractive index estimates as well. The benefit of such measurements is twofold. Firstly, the explicit retrieval of particle refractive index removes one major source of error which affects the size estimates derived from more traditional particle counters. Secondly, knowledge of particle refractive index is useful in determining composition and thus fully characterizing the aerosol.

8.6 Future work

The work presented here is a preliminary study into the feasibility of a new type of instrument. The current SPARCLE prototype is not at the stage where it could be used productively in aerosol research, as it suffers from several limiting factors which need to be addressed. Most of the problems arise from two key areas: the optical design of the instrument and the data collection hardware.

The current optical design leads to two main problems. First of all the background light levels severely limit the sensitivity of the instrument and result in approximately one third of each LDA profile being unusable. Secondly, the baffling of the laser beam required to limit the stray light problem results in ambiguous PMT measurements. Any further development of this instrument will require a redesign of the optical path to minimize scattering and reflections, as well as to provide an even illumination across the sample flow. The use of optical design software (such as Zemax or CODE V) in aiding the design process would undoubtedly help avoid many pitfalls.

Other steps to alleviate the optical design system problems should be investigated, these include mounting the LDA so that it samples scattering angles closer to forward scatter. Scattering from large particles is much stronger in the forward direction, and the angular variation in the scattered intensity is also at its strongest. Both of these factors could aid in improving the signal

to noise ratio of the LDA measurements. Ensuring an even illumination of the sample flow could be aided by applying some sort of aerodynamic focusing to the sample flow as it approached the measurement volume. The effects of focusing on volatile aerosol particles should be investigated, as should the effects of changing ambient conditions (most notably pressure) on the shape of the focused flow, before such a step is taken.

The system which limits data collection and storage in SPARCLE is the LDA and its control hardware. Any decrease in the minimum exposure time would improve the signal to noise ratio of the LDA profiles, since the sensor would spend less time looking at the background illumination. The transmission of a complete LDA measurement takes too long in the current system. One possible solution to this problem would be a system which is able to transmit a measurement to the PC while still collecting data from the LDA. Such a system would greatly reduce the number of coincidence errors found in the SPARCLE data.

The other major product of this work has been a retrieval scheme capable of determining particle size and refractive index from measurements of the scattering pattern. As with all algorithms, this scheme can certainly be improved upon. The growth in computing power over the past few decades has seen numerical optimization methods flourish and it is quite possible that an algorithm exists which can provide better performance than the one developed here. In addition, it is likely that measurement vector, and the way it is utilized, can be improved. For example, the use of an FFT to describe the fine structure of the scattering pattern can probably be improved upon. The periodicity of the high frequency structure is determined by the particle size alone (being the diffraction pattern produced by the particle, rather than the result of internal refraction). Thus it may be possible to define the particle size with a great deal of precision from this measurement alone, especially considering the low variance this value was found to have. Two possible improvements could be:

- Performing the FFT with respect to scattering angle, rather than LDA pixel number. Because the LDA pixels are not evenly spaced with respect to scattering angle, using them in the transform results in a broadening of the FFT peak. It must be noted that performing an FFT on unevenly spaced data is not possible; which means that a considerable penalty in

computational efficiency is incurred.

- Using the spacing of the fine structure to constrain the particle radius directly, rather than relying on the retrieval to do this, may also be useful. For example, given that we know this quantity to be related to particle size alone, we can introduce a strong *a priori* constraint on the radius before applying the retrieval algorithm. However, the retrieval does contain information on this fine structure already, so it is possible that this process may not produce a large benefit.

These alterations have the potential to significantly improve the performance of the retrieval and should be investigated as part of any further analysis of SPARCLE measurements.

As well as simply improving the retrieval algorithm there is also great scope for applying the scheme to other measurements of single particle scattering. For instance, as mentioned in Sections (5.3) and (8.2.2), it would be interesting to attempt the retrieval of the complex refractive index alone for use with instruments that provide an independent measurement of particle size, such as DAWN-A.

8.6.1 An issue raised in the review of this thesis

During the review process of this thesis an issue was raised regarding the accuracy of the LDA forward model used in the retrieval algorithm. Because the LDA pixels are arranged in a straight line the intensity measured by each pixel will depend on it's location. There are two effects which need to be taken into account:

- Pixels at the ends of the sensor are a factor of $\frac{1}{\sqrt{2}}$ $(\cos 45^\circ)$ further away from the scattering volume that the central pixels. As intensity is inversely proportional to radius squared, this results in a $\cos^2$ drop in scattered intensity at the pixels as one moves away from the centre of the sensor.

- Pixels at the centre of the sensors are orthogonal to the scattered radiation, while those at the ends are at a 45° angle to the incoming light, reducing their apparent area when viewed from the scattering volume. This results in an additional cos dependence in the total irradiance received by each pixel as one moves away from the centre of the sensor.

Thus it is necessary to scale the intensities predicted by the LDA forward model by a factor $\cos^3(\theta - \frac{\pi}{2})$, where θ is the scattering angle. In the calculations done in this work only the second of these two corrections has been applied, meaning that all calculated LDA intensities are incorrect by a factor of $\cos^2(\theta - \frac{\pi}{2})$.

This omission obviously has a significant effect on the intensities calculated in the LDA forward model and will need to be taken into account in any further work and analysis of the SPARCLE instrument. However it's effect on the results presented here is relatively minor, for the following reasons:

- Only the analysis of the SPARCLE measurements is effected. This problem is not an issue with the DAWN-A analysis, and the analysis of simulated SPARCLE measurements is not effected because the same forward model was used to generate the measurements and perform the retrieval.

- The formulation of the measurement vector reduces the impact of the problem. Only the smoothed LDA profile will be effected, since the location of the FFT peak is essentially independent of scaling of the pattern. In addition, only the central portion of the smoothed scattering profile (where the correction is smallest) is used in the retrieval.

In order to test the impact of this omission a short, 2-dimensional retrieval run has been performed on ten states from the SPARCLE measurement run with a nominal radius of 7.98 μm. A direct comparison between the retrieved states from this run and those from the original run is displayed in Figure (8.1). These results indicate that the effect of applying the correction is relatively minor. The new retrieved states are close to both the nominal state and the previous retrieved values, and the spread of results appears approximately unchanged. The most notable difference between the new and old state estimates is a slight offset towards larger radii and smaller refractive indices. Given the tendency for the retrieval to slightly underestimate the particle size found in Section (7.1) (see Figure (7.3) and Table (7.2)), this observation suggests that the application of the correction might improve the retrieval results by reducing this offset. However, due to the limited number of results available here, it is not possible to quantify the effects of the correction in a statistically significant way.

To fully understand the influence of the omission of the sensor distance correction on the performance of the retrieval a substantial amount of re-analysis

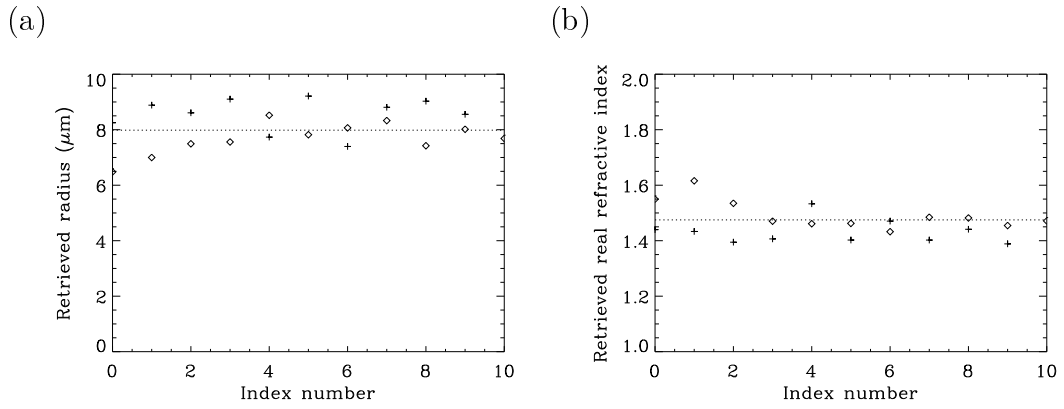


Figure 8.1: Comparison of the results of a 2-dimensional retrieval run using the corrected LDA forward model with results using the original forward model. The radius comparison is given in (a), while (b) shows the comparison of the real part of the refractive index. The original estimates are denoted by diamonds, while the new estimates are denoted by crosses. The nominal state is indicated by the dotted line.

needs to be done. However, at this late stage of the project, this is considered outside the scope of the current work.

Appendix A

Mie scattering theory

The interaction between an electromagnetic wave and physical matter can be thought of as the interaction between an oscillating electric field and one or more electric dipoles. If the electromagnetic wave encounters a particle much smaller than its wavelength the particle will act as a single dipole, the resultant scattering being known as Rayleigh scattering. As the particle size increases with respect to the wavelength the behaviour will deviate from that predicted for a single dipole since the particle begins to appear as a large array of multipoles. Mie (1908) first solved Maxwell's equations for this problem for the case of spherical particles. The scattered radiation at any given point is taken as the superposition of the secondary waves generated by the multipoles which constitute the particle as they oscillate in the incident electric field. The scattering predicted by Mie's equations becomes indistinguishable from that predicted by Rayleigh for small particles.

In the following discussion we will assume that the outside medium has a refractive index very close to unity, which simplifies the algebraic expressions slightly. Also used is the concept of the *plane of observation* for a measurement of such a scattering pattern. This plane is defined as being parallel to vectors from the particle in the direction of forward scattering and in the direction of some point in space where the measurement is being made. Similarly, we define the plane of polarization for a linearly polarized electromagnetic wave, this is simply the plane parallel to the electric field vector of the wave and it's direction of propagation.

The two parameters which determine the scattering for a given particle are

the size parameter,

$$x = \frac{2\pi}{\lambda} r, \tag{A.1}$$

and

$$y = mx, \tag{A.2}$$

where m is the refractive index of the particle, r is the radius of the particle and λ is the wavelength of the incident radiation.

The solution takes the form of two infinite series,

$$i_1(x, m, \theta) = |s_1|^2 = |\sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)}(a_n \pi_n + b_n \tau_n)|^2 \tag{A.3}$$

and

$$i_2(x, m, \theta) = |s_2|^2 = |\sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)}(a_n \tau_n + b_n \pi_n)|^2. \tag{A.4}$$

These two quantities are the intensity distribution functions in the directions perpendicular (i_1) and parallel (i_2) to the observation plane. The terms π_n and τ_n are defined as

$$\pi_n = \frac{1}{sin\ \theta} P_n^1, \tag{A.5}$$

$$\tau_n = \frac{d}{d\theta} P_n^1, \tag{A.6}$$

where P_n^1 is the nth order Legendre polynomial of the first type. Also,

$$a_n = \frac{\psi_n'(y)\psi_n(x) - m\psi_n(y)\psi_n'(x)}{\psi_n'(y)\zeta_n(x) - m\psi_n(y)\zeta_n'(x)} \tag{A.7}$$

and

$$b_n = \frac{m\psi_n'(y)\psi_n(x) - \psi_n(y)\psi_n'(x)}{m\psi_n'(y)\zeta_n(x) - \psi_n(y)\zeta_n'(x)}. \tag{A.8}$$

Here ψ_n and ζ_n are Ricatti-Bessel functions and $\psi_n'(x) = \frac{d\psi_n}{dx}$.

If the incident radiation is linearly polarized the scattering pattern will be dependent on the angle between the plane of polarization and the observation plane

$$I(\theta, \phi) = E_\phi \frac{\lambda^2}{4\pi^2}(i_1 sin^2\ \phi + i_2 cos^2\ \phi). \tag{A.9}$$

The sine and cosine terms are squared because we are dealing with intensity

(i.e. the square of the amplitude of the wave). From this equation it can be seen that when the observation plane and the polarization plane are the same the scattered radiation depends on the parallel intensity distribution function, i_2, and likewise when the two planes are at right angles I depends only on i_1 (as would be expected).

In the case of unpolarized incident radiation the two axis are equally illuminated but are not coherently related and the relationship becomes

$$I(\theta) = E\frac{\lambda^2}{4\pi^2}\left(\frac{i_1 + i_2}{2}\right). \tag{A.10}$$

The factor 2 in this equation is required because i_1 and i_2 are defined on the basis of polarized incident light of unit irradiance.

In computation successive terms of the infinite series (A.3) and (A.4) are calculated until convergence is achieved to the desired precision. This converged solution gives the distribution functions used to calculate the effects of the particle on the incident light (be it scattering or absorption). For a complete description of the subject of the scattering of light see Bohren & Huffman (1983).

Figures (A.1) and (A.2) show the phase function (a normalized version of Equation (A.10)) for a variety of particle sizes and refractive indices. For a particle with a radius smaller than, or of the same order as, the wavelength of the incident light there is very little information in the angular distribution of light. For such particles Rayleigh scattering theory can be applied, and it is unlikely that refractive index will be retrievable from the scattering pattern. Also evident from Figure (A.1) is that the larger the value of the size parameter, x, the more complex the scattering pattern. In other words the shorter the wavelength of the light used to detect the particles, the smaller the lower limit on detectable particle size and the larger the range of particles from which refractive index can be retrieved. Comparison of Figures (A.1) and (A.2) shows that scattering from absorbing particles results in a much smoother scattering pattern, particularly for large scattering angles. Note that structure in the scattering patterns of absorbing particles still becomes more fine with increasing size, but its amplitude is much reduced. There is also a reduced dependence on the real part of the refractive index apparent in the case of absorbing particles, with the curves

corresponding to differing values being more similar in Figure (A.2) than (A.1).

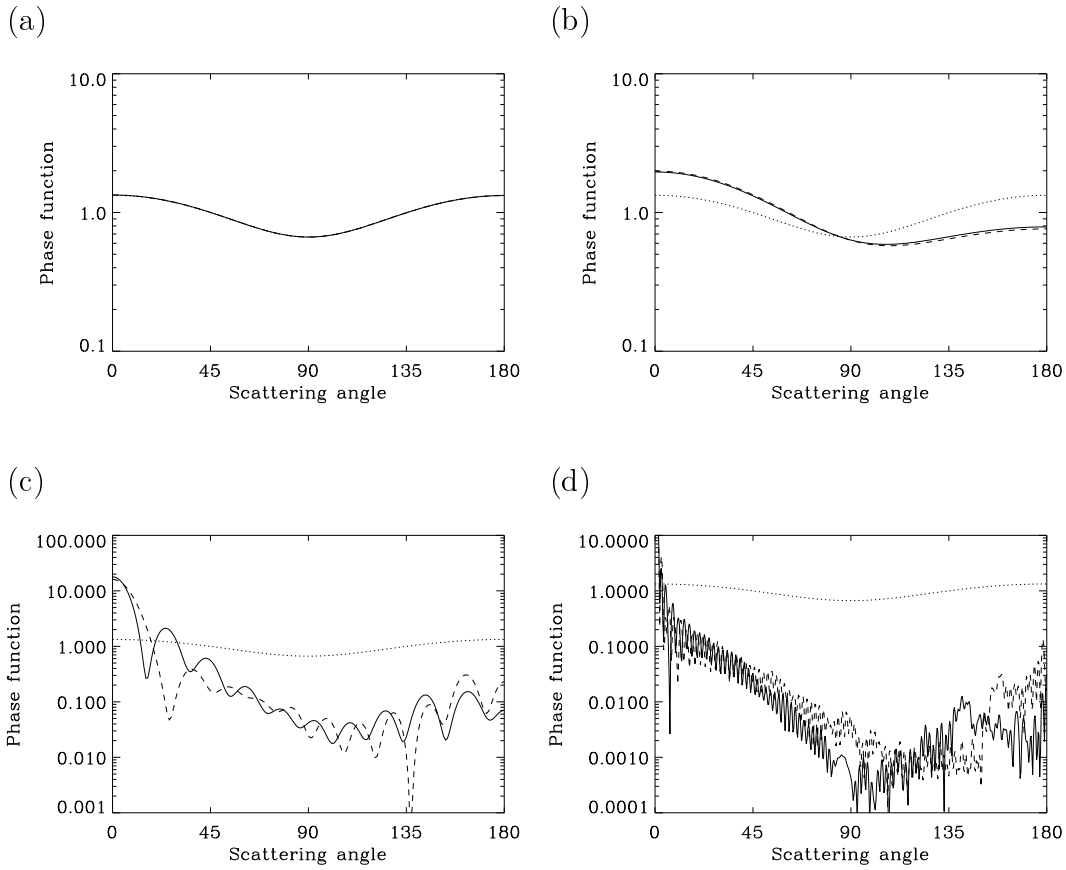


Figure A.1: Phase functions for droplets with a single wavelength and size parameters of: (a) $x = 0.1$, (b) $x = 1$, (c) $x = 10$ and (d) $x = 100$. Particles with a refractive index of $1.33 - 0.0i$ (solid line), $1.48 - 0.0i$ (dashed line) are plotted in each case, as well as the Rayleigh approximation (dotted line). The value of the phase function is a normalized intensity.

Figure (A.3) shows the effect on the measured phase function of increasing the wavelength band and/or the range of scattering angles used to make the measurement. It can be seen that both of these smooth the scattering pattern, removing the fine detail. Therefore, if a light source with a wide wavelength band (such as an incandescent lamp) or a detector that averages the signal over a broad range of scattering angles is used the chances of being able to retrieve refractive index are greatly reduced.

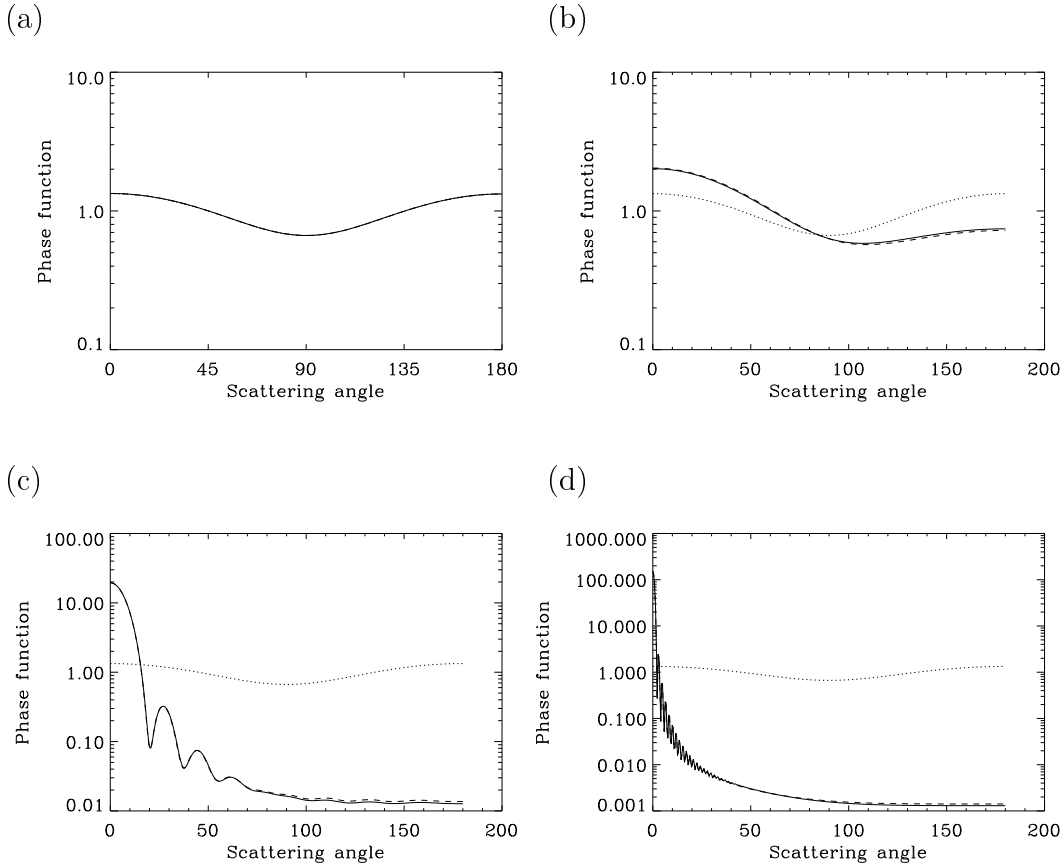


Figure A.2: Phase functions for droplets with a single wavelength and size parameters of: (a) $x = 0.1$, (b) $x = 1$, (c) $x = 10$ and (d) $x = 100$. Particles with a refractive index of $1.33 - 0.66i$ (solid line), $1.48 - 0.66i$ (dashed line) are plotted in each case, as well as the Rayleigh approximation (dotted line). The value of the phase function is a normalized intensity.

A.1 Analytical inversion of Mie scattering

Ludlow & Everitt (2000) present a method for analytically inverting a Mie scattering pattern to give the physical characteristics of the associated particle (i.e. its size and complex refractive index). This method is summarized here.

If the Mie intensity distribution functions (Equations (A.3) and (A.4)) are known, it can be shown that the scattering coefficients (a_n and b_n, Equations (A.7) and (A.8)) can be calculated by the integrals

$$a_n = \frac{1}{2n(n+1)} \int_0^{\pi} [s_{\parallel}(\theta)\tau_n(\theta) + s_{\perp}(\theta)\pi_n(\theta)] \sin\theta \mathrm{d}\theta \tag{A.11}$$

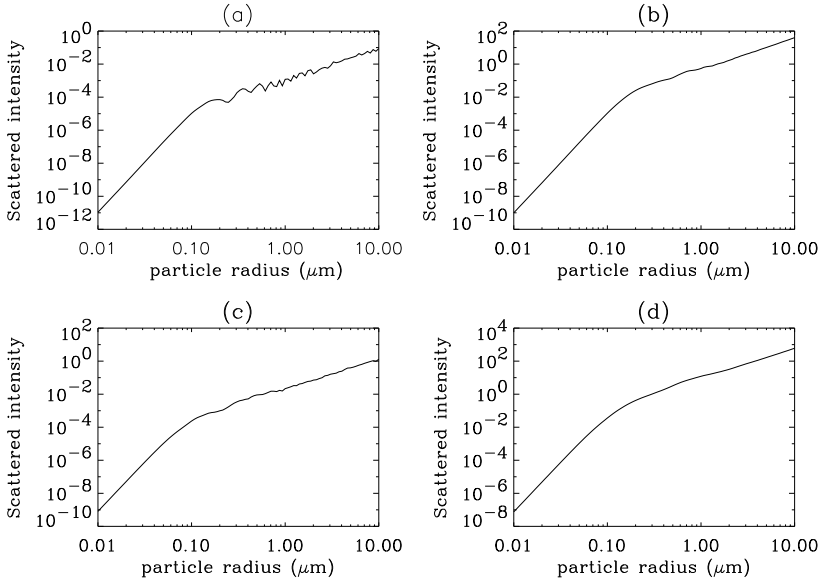


Figure A.3: The scattered intensity (in arbitrary units) from water droplets measured as a function of particle size for: (a) A scattering angle of 40° and a wavelength of 300 nm. (b) A scattering angle of 40° and a wavelength range of 300–700 nm. (c) All scattering angles from 40–140° and a wavelength of 300 nm. (d) All scattering angles from 40–140° and a wavelength range of 300–700 nm.

and

$$b_n = \frac{1}{2n(n+1)} \int_0^{\pi} [s_{\parallel}(\theta)\pi_n(\theta) + s_{\perp}(\theta)\tau_n(\theta)] \sin\theta \mathrm{d}\theta, \tag{A.12}$$

where $\pi_n(\theta)$ and $\tau_n(\theta)$ are as defined in Equations (A.5) and (A.6).

In the case of a homogeneous sphere we can relate the complex refractive index and the distribution functions by:

$$m^2 = \left[\frac{a_n\xi_n(x) - \phi_n(x)}{a_n\xi'_n(x) - \phi'_n(x)}\right] \left[\frac{b_n\xi_n(x) - \phi_n(x)}{b_n\xi'_n(x) - \phi'_n(x)}\right] \tag{A.13}$$

where $\xi_n(x)$, $\phi_n(x)$, $\xi'_n(x)$ and $\phi'_n(x)$ are as defined in the previous section. This equation can then be used to fully characterize the scattering sphere. If the function

$$f(z) = \left[\frac{a_n\xi_n(z) - \phi_n(z)}{a_n\xi'_n(z) - \phi'_n(z)}\right] \left[\frac{b_n\xi_n(z) - \phi_n(z)}{b_n\xi'_n(z) - \phi'_n(z)}\right] \tag{A.14}$$

is plotted against the order n for varying values of z, a straight line will result when $f(z) = m^2$ and $z = x$. Thus, the inversion of a Mie scattering pattern is

reduced to a trivial 1-dimensional minimization.

This method can also be extended to non-homogeneous but spherically symmetric particles. In this case the solution is given by a pair of radial differential equations:

$$\begin{aligned} \frac{\mathrm{d}^2}{\mathrm{d}r^2}\phi_n(kr) &+ \left[k^2 - \frac{n(n+1)}{r^2}\right]\phi_n(kr) \\ &= \begin{cases} V_1(r)\phi_n(kr) \\ V_1(r)\phi_n(kr) + V_2(r)\frac{\mathrm{d}}{\mathrm{d}r}\phi_n(kr), \end{cases} \end{aligned} \tag{A.15}$$

where

$$V_1(r) = -\left[m^2(r) - 1\right]k^2, \tag{A.16}$$

$$V_2(r) = -\frac{\mathrm{d}}{\mathrm{d}r}\left\{\ln\left[\frac{1}{m^2(r)}\right]\right\}, \tag{A.17}$$

n is again an integer order and $m_n(r)$ is the complex refractive index as a function of radius. Note that, in general, $m_n(r)$ will have a discontinuity at the particle surface.

Although this method significantly reduces the complexity of the retrieval process (particularly in the case of a homogeneous particle), it does require complete knowledge of the intensity distribution functions. Thus, obtaining a measurement of a scattering pattern suitable for use with this method is far from trivial and is not practical for an *in situ* instrument.

Appendix B

Retrieval theory

B.1 Introduction to the retrieval problem

Any situation where a measurement of a system gives you a quantity which is a function of the property of interest is a retrieval problem. Almost all measurements involve a retrieval problem, but in most cases these are so simple they are solved implicitly. An example of this is using the height of a column of mercury to measure temperature. However, when the functional relationship between the desired and measured quantities becomes more complex so does the retrieval problem; this is particularly true where the measurement of several different quantities is used to infer several other quantities which describe the state of some system. A good example of this situation is using spectral measurements of the Earth's (or some other planet's) atmosphere to infer everything from vertical temperature structure to composition. Such a measurement is related to the state by very complex relationships which are often not analytically invertable. For example, it is possible to write an expression relating the composition and temperature to a specific spectral signature, but it is not possible to write an expression relating a given spectrum to a specific state of the atmosphere.

The development of the computer has made it practical to solve such problems using iterative numerical methods. Central to the aims of the SPARCLE project is a retrieval problem of this type, namely using the Mie scattering pattern produced by a particle to determine its size and refractive index. This appendix gives a brief introduction to retrieval theory and the methods used in this work. Rodgers (2000) rigorously develops retrieval theory as applied to atmospheric remote sensing and much of the material and notation presented here is drawn

from that work.

A retrieval problem can be formulated in the following way. The quantities measured are arranged in a vector called the *measurement vector*, $\mathbf{y}$, of length m. The quantities we are interested in are arranged in a *state vector*, $\mathbf{x}$, of length n. These two vectors can be thought of as existing in two vector manifolds, *measurement space* (dimension m) and *state space* (dimension n) – the measurement space contains all of the possible measurements, the state space all of the possible states. A point in state space is mapped to a point in measurement space by the *forward function*, $\mathbf{f}(\mathbf{x})$, which depends on all the physics involved. We can only approximate this process with our *forward model*, $\mathbf{F}(\mathbf{x})$. The retrieval problem is then one of finding the inverse mapping, $\mathbf{R}(\mathbf{y})$, which maps a point in measurement space to a point in state space.

In reality we will not obtain a single exact solution to any retrieval problem, due to experimental uncertainty and forward model error. We will instead have a distribution of probable states from which we must select the 'best'. We achieve this by optimizing some quantity associated with the solution: for instance we might try to maximize the likelihood of the solution, or we might minimize the uncertainty in the solution. Not only that, but we would also like to include knowledge we already have about the solution in our retrieval, to ensure that the solution is a reasonable one.

Because we can no longer talk of unique solutions we will generalize to the notation of probability density functions (*pdf*s) and use Bayes theorem to formulate the solution. Firstly some definitions:

$P(\mathbf{x})$ The prior *pdf* of the state $\mathbf{x}$. This quantity embodies our knowledge about the state before we have measured it. $P(\mathbf{x})\mathrm{d}\mathbf{x}$ is the probability that the true value of $\mathbf{x}$ lies between $\mathbf{x}$ and $\mathbf{x} + \mathrm{d}\mathbf{x}$.

$P(\mathbf{y})$ The prior *pdf* of the measurement $\mathbf{y}$. As with $P(\mathbf{x})$, this is our knowledge of the measurement before it is made.

$P(\mathbf{x}, \mathbf{y})$ The joint prior *pdf*. $P(\mathbf{x}, \mathbf{y})\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}$ is the probability that the true value of $\mathbf{x}$ lies between $\mathbf{x}$ and $\mathbf{x} + \mathrm{d}\mathbf{x}$ and the true value of $\mathbf{y}$ lies between $\mathbf{y}$ and $\mathbf{y} + \mathrm{d}\mathbf{y}$. Note that $P(\mathbf{x}, \mathbf{y}) \equiv P(\mathbf{y}, \mathbf{x})$.

$P(\mathbf{y}|\mathbf{x})$ The conditional *pdf* of $\mathbf{y}$ given $\mathbf{x}$. $P(\mathbf{y}|\mathbf{x})\mathrm{d}\mathbf{y}$ is the probability that the true value of $\mathbf{y}$ lies between $\mathbf{y}$ and $\mathbf{y} + \mathrm{d}\mathbf{y}$ for given $\mathbf{x}$.

$P(\mathbf{x}|\mathbf{y})$ The conditional *pdf* of $\mathbf{x}$ given $\mathbf{y}$. This is the quantity of interest.

Bayes theorem links these *pdf*s together with the following relationships

$$\begin{aligned} P(\mathbf{x},\mathbf{y}) &= P(\mathbf{x}|\mathbf{y})P(\mathbf{y}) \\ P(\mathbf{y},\mathbf{x}) &= P(\mathbf{y}|\mathbf{x})P(\mathbf{x}), \end{aligned} \tag{B.1}$$

so that we can say

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y})}. \tag{B.2}$$

If we include Gaussian experimental error the relationship between measurement and state becomes

$$\mathbf{y} = \mathbf{F}(\mathbf{x}) + \epsilon \tag{B.3}$$

and we can generate a *measurement error covariance matrix*

$$\mathbf{S}_\epsilon = E\{\epsilon\epsilon^T\} = E\{[\mathbf{y} - \mathbf{F}(\mathbf{x})][\mathbf{y} - \mathbf{F}(\mathbf{x})]^T\}, \tag{B.4}$$

where $E\{\}$ denotes the expectation value of a quantity. We can then write down an equation for the conditional *pdf* of $\mathbf{y}$ given $\mathbf{x}$

$$-2\ln P(\mathbf{y}|\mathbf{x}) = [\mathbf{y} - \mathbf{F}(\mathbf{x})]^T \mathbf{S}_\epsilon{}^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x})] + c_1, \tag{B.5}$$

where c_1 is a constant.

If we also assume that the *pdf* of our prior knowledge of the state has a Gaussian pdf we can follow a similar procedure. Defining $\mathbf{x}_a$ as the prior value of the state vector we can generate a covariance matrix for the *a priori* state –

$$\mathbf{S}_a = E\{[\mathbf{x} - \mathbf{x}_a][\mathbf{x} - \mathbf{x}_a]^T\}, \tag{B.6}$$

giving rise to the expression

$$-2\ln P(\mathbf{x}) = [\mathbf{x} - \mathbf{x}_a]^T \mathbf{S}_a^{-1} [\mathbf{x} - \mathbf{x}_a] + c_2. \tag{B.7}$$

Substituting into Equation (B.2) gives

$$-2\ln P(\mathbf{x}|\mathbf{y}) = [\mathbf{y}-\mathbf{F}(\mathbf{x})]^T {\mathbf{S}_\epsilon}^{-1}[\mathbf{y}-\mathbf{F}(\mathbf{x})] + [\mathbf{x}-\mathbf{x}_a]^T \mathbf{S}_a^{-1}[\mathbf{x}-\mathbf{x}_a] + c_3, \quad \text{(B.8)}$$

giving a general form of the solution for the ensemble of possible solutions described by $P(\mathbf{x}|\mathbf{y})$. If we make a linear approximation to the function $\mathbf{F}(\mathbf{x})$ (thereby making Equation (B.8) quadratic in $\mathbf{x}$) then we can re-write the posterior *pdf* in the form

$$-2\ln P(\mathbf{x}|\mathbf{y}) = [\mathbf{x}-\hat{\mathbf{x}}]^{\mathbf{T}} \hat{\mathbf{S}}^{-\mathbf{1}} [\mathbf{x}-\hat{\mathbf{x}}] + c_4 \quad \text{(B.9)}$$

because it is also a Gaussian, with expected value $\hat{\mathbf{x}}$ and covariance $\hat{\mathbf{S}}$. Equating terms that are quadratic in $\mathbf{x}$ in Equations (B.8) and (B.9) gives an expression for the posterior covariance (in the linear approximation with Gaussian *pdf*s)

$$\hat{\mathbf{S}}^{-\mathbf{1}} = \mathbf{K}^{\mathbf{T}} \mathbf{S}_\epsilon^{-\mathbf{1}} \mathbf{K} + \mathbf{S}_\mathbf{a}^{-\mathbf{1}}. \quad \text{(B.10)}$$

Here $\mathbf{K} = \nabla_\mathbf{x}\mathbf{F}(\mathbf{x})$, the derivative of the forward model with respect to the state vector. $\mathbf{K}$ is an $n \times m$ matrix and is known as the *weighting function matrix*.

For our measurement to be useful, however, we need to be able to select one single solution. To achieve this aim it is necessary to optimize some characteristic of the solution. For the retrieval problem at hand it was decided to find the solution which maximized the posterior *pdf*, $P(\mathbf{x}|\mathbf{y})$, i.e. finding the mean state of this *pdf*

$$\hat{\mathbf{x}} = \int \mathbf{x} P(\mathbf{x}|\mathbf{y}) \mathrm{d}\mathbf{x}. \quad \text{(B.11)}$$

This is know as the *Maximum a Posteriori Solution* and provides the most probable solution given all the information about the state (measurement and *a priori*). Other possible approaches are to minimize some error quantity (such as the modelling error or the measurement error), or to seek the most precise solution. It was decided to maximize $P(\mathbf{x}|\mathbf{y})$ because it would seem the most physically reasonable solution and some of the other possible approaches are subject to effects unrelated to the quality of the resulting solution. For example, if the measurement error is minimized, the mathematically optimum solution will usually be that which uses the minimum number of measurements that allow a solution, as the addition of any more measurements and their associated uncertainty is likely to increase the error in the solution. This is obviously not a good approach.

Maximizing $P(\mathbf{x}|\mathbf{y})$ is equivalent to finding the minimum of the right hand side of Equation (B.8). We therefore refer to this expression as the *cost function* for our problem. If we take a two dimensional example (e.g. the retrieval of particle size r and the real part of the refractive index n of particles with a fixed imaginary part of the refractive index) the variation of the cost function can be pictured as a surface over these two dimensions. The solution is the r and n that correspond to the lowest point on the surface. The inversion problem can thus be considered to be that of finding the global minimum of a surface.

Many methods for finding the minimum of a surface exist and no one is ideal for all problems. The two methods used in the SPARCLE retrieval system are Simulated Annealing and Levenberg-Marquardt, both of which are discussed in detail in the following two sections. Simulated annealing is a so-called non-optimal estimate because it does not guarantee that the solution provided will lie at the minimum found to a given precision. However, it does cope well with cost functions that contain local minima. Levenberg-Marquardt is an optimal method, that is if it converges the solution will always lie at a minimum to a given precision. However, if the cost function has many local minima there is no guarantee that the minimum found will be the global one.

B.2 Simulated annealing

In the retrieval scheme developed for the SPARCLE project the initial method used to try to isolate the area of the cost function containing the global minimum is a modified version of simulated annealing. The physical process of annealing relates to the softening of 'work hardened' metal. When a metal is deformed below its re-crystallization temperature the crystal structure becomes distorted as atoms are moved away from their equilibrium positions. This process results in an increase in strength and hardness, and a decrease in ductility of the bulk metal[1]. The annealing process involves heating the metal above the re-crystallization temperature, allowing time for the crystal lattice to return to it's equilibrium state and then allowing the metal to slowly cool. This restores

[1]A similar effect can be achieved by rapidly cooling (quenching) metal from a temperature above its re-crystallization temperature.

the bulk metal to a more ductile (lower energy) state, the degree of which is time dependent. It is by emulating this time dependent re-crystallization process that simulated annealing works.

The standard implementation of simulated annealing operates on a discrete space of possible solutions, such as the famous travelling salesman problem[2]. At each iteration of the algorithm one element of the system configuration is changed (i.e. in the travelling salesman problem the order in which two adjacent cities are visited is swapped). The 'energy state', E (in our case the cost function, and the total distance travelled in the salesman problem) is then evaluated; if the step has lowered the energy state it is always accepted but, if the energy has increased there is still a chance that the step will be accepted. The probability of such an 'up hill' step is determined by a 'temperature' parameter, t, via the Boltzmann probability distribution:

$$P \propto \exp\left(\frac{-E}{kT}\right), \tag{B.12}$$

where k is a constant (not necessarily Boltzmann's constant as found in the thermodynamic problem) and P is the probability that the new state will be accepted. Each step which results in increased energy is compared to a random number, if the random number is less than P the state is accepted, otherwise it is rejected. A large number of these random steps (equivalent to random thermal changes of state in our hypothetical piece of steel) are made, the temperature being lowered slightly periodically throughout the procedure. Each reduction in the temperature reduces the likelihood of uphill steps being accepted, so the state gradually becomes confined to a minimum. The scheme for reducing the temperature parameter varies between implementations of simulated annealing. Common approaches are to reduce the temperature by either a fixed value or a fixed ratio after a certain number of random steps have been made.

There is a large amount of experimentation involved in getting a working and optimized simulated annealing scheme as the concepts of temperature and thermal energy are purely artificial constructs for most problems. Even a well-optimized annealing scheme will be an inefficient way of solving a well-behaved minimization problem. However where the value being minimized displays many local minima, simulated annealing shows a remarkable ability to converge to the

[2]The travelling salesman problem involves choosing the shortest route for a salesman who must visit a set of N cities, passing through each one only once.

global minimum, and therefore has the advantage over 'down hill' type methods for such problems.

The annealing scheme used in the SPARCLE retrieval is a modified version of the 'SimAnneal' C++ class written by Dr. Everett Carter (the original version of the code is free for use and modification and can be downloaded from the World Wide Web at http://www.taygeta.com/annealing/simanneal.html). The class is designed to interface with a separate function which provides the value of the mathematical function to be minimized (i.e. the cost function) and provides a method for determining the optimal initial value of the temperature parameter. The cooling scheme used reduces the temperature by a fixed ratio (approximately 10% by default) for every 20 random steps made in state space. The number of temperature steps required to reach a stable solution is approximately 400.

The code is designed to work with a continuous state space rather than a finite set of possible states. This is an extension to the original simulated annealing algorithm, but the alteration is extremely minor. In the continuous case a new state vector is generated by simply adding a random vector to the old state, the magnitude of which is proportional to the temperature parameter. The random number generator used by the class is the R250 generator, which is described in Maier (1991).

The code also allows the storage of the n best states (where n is some user defined number) encountered during the annealing of the system. This enables multiple minima to be tested further, thus enhancing the likelihood of finding the correct minimum. Although the solution given by simulated annealing will always be very close to a deep minimum there is no guarantee that it will be the global minimum. In this work the 100 best states found by the simulated annealing were passed on to optimal estimation code.

B.3 Optimal estimation methods

The term optimal estimation method is used here to describe a numerical method that makes an estimate of where the minimum state should lie based on the nature of the function at the current state. The new estimate of the state is then used to make another estimate and, all being well, over several iterations

these estimates converge to the true location of the minimum. The problem is said to have been solved when successive states are within the required precision of each other. This can be compared to simulated annealing, which does not claim to give you the exact minimum at some precision level, it simply gives you the last state of a constrained random walk through state space.

There are an enormous number of different optimal estimation methods available. These range from the simple and obvious steepest descent method which calculates the local gradient of the function and simply steps down hill at each iteration, to schemes which approximate the function being minimized with some simpler function and step to the location of the approximation's minimum value. All these methods suffer to varying degrees from the same weakness, namely if the function being minimized has many local minima and/or is poorly modelled by the assumed functional form the chances are high that the method will either converge to the incorrect minimum, or not converge at all.

This makes an optimal estimation method alone completely inappropriate for the problem presented by the SPARCLE instrument because the cost function is extremely complex and contains many minima. However, if the optimal estimation can be given a starting point which is sufficiently close to the global minimum it will be able to define the minima better than a non-optimal method alone.

Although Levenberg-Marquardt is the only optimal estimation scheme used in the SPARCLE retrieval code we will start by describing the Gauss-Newton method as the former method can be thought of as an extension of this scheme.

Both of the methods discussed in this section search for the maximum in in $P(\mathbf{x}|\mathbf{y})$ by locating the root in its first derivative. Taking the derivative of Equation (B.8) gives

$$\nabla_{\mathbf{x}}[-2\ln P(\mathbf{x}|\mathbf{y})] = 0 = -\left[\nabla_{\mathbf{x}}\mathbf{F}(\mathbf{x})\right]^T \mathbf{S}_\epsilon{}^{-1}[\mathbf{y} - \mathbf{F}(\mathbf{x})] + \mathbf{S}_a^{-1}[\mathbf{x} - \mathbf{x}_a]. \qquad \text{(B.13)}$$

This cannot be solved directly, because the term $\mathbf{F}(\mathbf{x})$ will be non-linear in gen-

eral. We therefore linearize $\mathbf{F}(\mathbf{x})$ about some point $\mathbf{x}_0$

$$\begin{aligned}\mathbf{F}(\mathbf{x}) &= \mathbf{F}(\mathbf{x_0}) + \nabla_{\mathbf{x}}\mathbf{F}(\mathbf{x})[\mathbf{x} - \mathbf{x_0}] \\ &= \mathbf{y_0} + \mathbf{K}[\mathbf{x} - \mathbf{x_0}],\end{aligned} \tag{B.14}$$

where $\mathbf{y_0} = \mathbf{F}(\mathbf{x_0})$ and $\mathbf{K} = \nabla_{\mathbf{x}}\mathbf{F}(\mathbf{x})$ is the weighting function matrix mentioned. This then gives an implicit equation for $\hat{\mathbf{x}}$ of the form

$$-\hat{\mathbf{K}}^T(\hat{\mathbf{x}}){\mathbf{S}_\epsilon}^{-1}[\mathbf{y} - \mathbf{F}(\hat{\mathbf{x}})] + \mathbf{S}_a^{-1}[\hat{\mathbf{x}} - \mathbf{x}_a] = 0. \tag{B.15}$$

In the linear case this can be solved explicitly, but not in the non-linear case since the linearization step will only be valid over a small region. Therefore some numerical solution must be found. It is important to note that this equation was derived using Gaussian *pdfs* which only have one maximum and, hence, one root.

B.3.1 Gauss-Newton iteration

Newton's method is a simple numerical method for determining the roots of a general one dimensional function, $f(x)$. Starting from some initial guess of the location of the root, the function is approximated by a straight line at each iteration, i, with the instantaneous gradient of f at x_i, $f'(x_i)$,

$$y = f'(x_i)(x - x_i) + f(x_i). \tag{B.16}$$

The root of this linear approximation is then used as the next iteration point, x_{i+1}. The equation relating x_{i+1} to x_i can be expressed as –

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}. \tag{B.17}$$

Newton iteration will not always provide a quick convergence and is not guaranteed to converge at all, but provided the starting point is close enough to the true root it will give the fastest convergence of any method.

Newton's method can be generalized to multiple dimensions by replacing x with a state vector, $\mathbf{x}$, and the derivative with the vector gradient, $\nabla_{\mathbf{x}}$. The

$i+1$th guess at the state vector is given by –

$$\mathbf{x_{i+1}} = \mathbf{x}_i - (\nabla_{\mathbf{x}}\mathbf{g}(\mathbf{x}_i))^{-\mathbf{1}}\mathbf{g}(\mathbf{x}_i) \tag{B.18}$$

where $\mathbf{g}(\mathbf{x})$ is some vector valued function. Since finding the minimum (or maximum) of a function is equivalent to locating the roots of that functions derivative we can think of $\mathbf{g}(\mathbf{x})$ as the derivative of the cost function, given in equation (B.15) as

$$\mathbf{g}(\mathbf{x}) = -\hat{\mathbf{K}}^T(\hat{\mathbf{x}})\mathbf{S}_\epsilon{}^{-1}[\mathbf{y} - \mathbf{F}(\hat{\mathbf{x}})] + \mathbf{S}_a^{-1}[\hat{\mathbf{x}} - \mathbf{x}_a], \tag{B.19}$$

thus

$$\nabla_{\mathbf{x}}\mathbf{g}(\mathbf{x}_i) = \mathbf{S_a^{-1}} + \mathbf{K^T}\mathbf{S}_\epsilon - \mathbf{1K} - [\nabla_{\mathbf{x}}\mathbf{K^T}]\mathbf{S}_\epsilon^{-\mathbf{1}}[\mathbf{y} - \mathbf{F}(\mathbf{x})]. \tag{B.20}$$

Remembering that $\mathbf{K}$ is the derivative of $\mathbf{F}$, then $\nabla_{\mathbf{x}}\mathbf{K^T}$ is the second derivative of the forward model. This is a very unwieldy object, being a vector whose elements are matrices, and is usually very costly to compute. Fortunately as long as the problem is only mildly non-linear the term involving it is small and it can be neglected[3]. Ignoring this term and including *a priori* information gives the Gauss-Newton iteration step,

$$\mathbf{x_{i+1}} = \mathbf{x}_i + (\mathbf{S_a^{-1}} + \mathbf{K}_i^{\mathbf{T}}\mathbf{S}_\epsilon^{-\mathbf{1}}\mathbf{K}_i)^{-\mathbf{1}}[\mathbf{K}_i^{\mathbf{T}}\mathbf{S}_\epsilon^{-\mathbf{1}}(\mathbf{y} - \mathbf{F}(\mathbf{x}_i)) - \mathbf{S_a^{-1}}(\mathbf{x}_i - \mathbf{x_a})]. \tag{B.21}$$

B.3.2 The Levenberg-Marquardt method

Because Gauss-Newton will provide very poor performance if the form of the problem is far from quadratic (i.e. the value of $\nabla_{\mathbf{x}}\mathbf{K}$ is not close to zero) Levenberg (1944) proposed the following variant on the Gauss-Newton iteration

$$\mathbf{x}_{i+1} = \mathbf{x}_i + (\mathbf{K}\mathbf{K}^T + \gamma_i\mathbf{I})^{-1}\mathbf{K}^T[\mathbf{y} - \mathbf{F}(\mathbf{x}_i)] \tag{B.22}$$

where γ_i is a integer factor which is updated during each iteration. Setting γ_i to 0 we obtain

$$\mathbf{x}_{i+1} = \mathbf{x}_i + (\mathbf{K}\mathbf{K}^T)^{-1}\mathbf{K}^T[\mathbf{y} - \mathbf{F}(\mathbf{x}_i)] \tag{B.23}$$

which is a normal Gauss-Newton iteration, whereas for $\gamma_i \rightarrow \infty$ the step direction becomes that of steepest descent. The scheme is to alter the value of γ_i based on the change in the cost function in the previous step. If the

[3]However, it is worth noting that the term's value is determined not only by the non-linearity, but also by the measurement noise.

previous step was poor γ_i is increased to make the algorithm behave like the very robust steepest decent method, whereas if performance is good γ_i is decreased to make the next iteration closer to that given by the Gauss-Newton step.

Applying this alteration to the Gauss-Newton iteration equation (B.21) gives

$$\mathbf{x_{i=1}} = \mathbf{x}_i + (\mathbf{S_a^{-1}} + \mathbf{K}_i^{\mathbf{T}}\mathbf{S}_\epsilon^{\mathbf{-1}}\mathbf{K}_i + \gamma\mathbf{D}_n)^{\mathbf{-1}} \left(\mathbf{K}_i^{\mathbf{T}}\mathbf{S}_\epsilon^{\mathbf{-1}}[\mathbf{y} - \mathbf{F}(\mathbf{x}_i)] - \mathbf{S_a^{-1}}[\mathbf{x}_i - \mathbf{x_a}]\right), \tag{B.24}$$

where $\mathbf{D}_n$ is some square, positive definite matrix of size n.

The scheme used to update the factor γ was developed by Fletcher (1971) and is based on the ratio of the change in the cost function computed with the full forward model and that calculated with the linear approximation. This ratio will be 1 if the linear approximation is exact and will become negative if the cost function increases in size. The scheme is as follows:

- If the ratio is greater than 0.75, reduce γ.
- If the ratio is less than 0.25, increase γ.
- Otherwise make no change.
- If γ is below some threshold, use zero.

This more complicated scheme (than simply using the change in value of the cost function to update γ) is used in order to keep the new estimate of $\mathbf{x}$ within linear range of the old estimate (the so called *trust region*).

For the SPARCLE retrieval scheme a Levenberg-Marquardt C++ class was developed. The class is formulated to work most efficiently in cases where the state vector has less elements than the measurement vector (the n-form discussed in (Rodgers, 2000)). It uses the same cost function class as the simulated annealing class as well as the 'NewMat' linear algebra class developed by Robert Davies[4]. The Levenberg-Marquardt class allows all of the important parameters to be set, including the covariances of both the measurement and *a priori*, the maximum number of iterations to allow and the absolute minimum and maximum values for the state vector elements.

[4]As with the simulated annealing class this is open source code and can be downloaded from http://webnz.com/robert/

B.4 Information content of the measurement

To be able to retrieve useful quantities from a measurement we must be sure that our measurement actually contains information on those quantities. This basically amounts to determining how many independent elements we have in the measurement vector: measuring 500 separate intensities in a scattering pattern by no means suggests we have 500 independent measurements.

Information theory was originally developed by Shannon & Weaver (1949) for determining the capacity of communication channels. The theory is analogous to that of entropy in thermodynamics. The thermodynamic entropy of a system is the logarithm of the number of internal states of the system that give rise to the same measured macro-state (given by macro-physical quantities such as temperature, pressure etc). The information content equivalent of this is essentially the number of states of the system that could result in the same measurement, given uncertainties. The information content, H, is defined simply as the change in entropy caused by taking the measurement.

Rodgers (2000) shows that the entropy, S, of a Gaussian *pdf* is related to its covariance matrix, $\mathbf{S}$, by $S[P] = \frac{1}{2}\ln|\mathbf{S}|$. It is therefore possible to calculate the information content of a measurement in state space,

$$\begin{aligned} H_s &= S[P(\mathbf{x})] - S[P(\mathbf{x}|\mathbf{y})] && \text{(B.25)} \\ &= \frac{1}{2}\ln|\mathbf{S_a}| - \frac{1}{2}\ln|\hat{\mathbf{S}}| && \text{(B.26)} \\ &= \frac{1}{2}\ln|\hat{\mathbf{S}}^{-1}\mathbf{S_a}| && \text{(B.27)} \\ & && \text{(B.28)} \end{aligned}$$

where $\hat{\mathbf{S}}$ is the covariance matrix of the retrieved state, which is given in Equation (B.10). Thus

$$H_s = \frac{1}{2}\ln|(\mathbf{K^T}\mathbf{S}_\epsilon^{-1}\mathbf{K} + \mathbf{S_a^{-1}})\mathbf{S_a}|. \qquad \text{(B.29)}$$

It is also possible to calculate the information content in measurement space,

$$H_m = S[P(\mathbf{y})] - S[P(\mathbf{y}|\mathbf{x})]. \qquad \text{(B.30)}$$

The information content in this case is the difference between the entropy of the

prior estimate of $\mathbf{y}$ and the posterior estimate. The covariance of$P(\mathbf{y})$ (the *pdf* of the measurement before it is made) is

$$\begin{aligned}\mathbf{S}_{\mathbf{y_a}} &= E\left\{[\mathbf{y}-\mathbf{y_a}][\mathbf{y}-\mathbf{y_a}]^{\mathbf{T}}\right\} \\ &= E\left\{[\mathbf{F}(\mathbf{x})-\mathbf{F}(\mathbf{x_a})][\mathbf{F}(\mathbf{x})-\mathbf{F}(\mathbf{x_a})]^{\mathbf{T}}+\epsilon\epsilon^{\mathbf{T}}\right\},\end{aligned} \tag{B.31}$$

applying the linear approximation gives –

$$\begin{aligned}\mathbf{S}_{\mathbf{y_a}} &= E\left\{\mathbf{K}(\mathbf{x}-\mathbf{x_a})(\mathbf{x}-\mathbf{x_a})^{\mathbf{T}}\mathbf{K}^{\mathbf{T}}+\epsilon\epsilon^{\mathbf{T}}\right\} \qquad \text{(B.32)} \\ &= \mathbf{K}\mathbf{S_a}\mathbf{K}^{\mathbf{T}}+\mathbf{S}_\epsilon. \qquad \text{(B.33)}\end{aligned}$$

Thus –

$$H_m = \frac{1}{2}\ln|\mathbf{S}_\epsilon^{-\mathbf{1}}(\mathbf{K}\mathbf{S_a}\mathbf{K}^{\mathbf{T}}+\mathbf{S}_\epsilon)|. \tag{B.34}$$

Perhaps a more useful way of determining the information content of a measurement is to calculate the *effective row space* of the problem. In the linear case, without noise there will be p pieces of independent information, where p is the rank of the weighting function matrix $\mathbf{K}$. The space spanned by the rows of $\mathbf{K}$ is termed the row space and has degree p. If, however, we take the various sources of uncertainty into account we may well find that some components of the row space become worthless – the signal is lost in the noise. We therefore define the *effective rank* as the number of *useful* measurements and the subspace they lie in as the effective row space.

To determine the effective row space we compare the measurement error covariance, $\mathbf{S}_\epsilon$, with the natural variability of the measurements, as expressed in the prior covariance, $\mathbf{S_a}$. If the measurement error on a component is greater than the natural variability then that component can provide virtually no useful information and is not part of the effective row space. Of course covariance matrices in general have off-diagonal elements (the covariances between separate measurements) and to compare them they must be diagonalized. Following the

method of Rodgers (2000) we transform to the basis where

$$\tilde{\mathbf{x}} = \mathbf{S_a}^{-\frac{1}{2}}\mathbf{x} \text{ and } \tilde{\mathbf{y}} = \mathbf{S_\epsilon}^{-\frac{1}{2}}\mathbf{y} \tag{B.35}$$

giving the forward model

$$\begin{aligned} \tilde{\mathbf{y}} &= \mathbf{S_\epsilon}^{-\frac{1}{2}}\mathbf{K}\mathbf{S_a}^{-\frac{1}{2}}\tilde{\mathbf{x}} + \mathbf{S_\epsilon}^{-\frac{1}{2}}\epsilon \\ &= \tilde{\mathbf{K}}\tilde{\mathbf{x}} + \tilde{\epsilon}. \end{aligned} \tag{B.36}$$

This has the useful property that both the covariances of $\tilde{\mathbf{x}}$ and $\tilde{\epsilon}$ are unit matrices, e.g. $\mathbf{S}_{\tilde{\epsilon}} = E\{\epsilon\epsilon^{\mathbf{T}}\} = \mathbf{I}$.

Since $\mathbf{S}_{\tilde{\epsilon}}$ and $\mathbf{S}_{\tilde{\mathbf{a}}}$ lie in different spaces they are still not directly comparable. We must therefore compare the prior covariance of $\tilde{\mathbf{y}}$ with $\mathbf{S}_{\tilde{\epsilon}}$:

$$\begin{aligned} \mathbf{S}_{\tilde{\mathbf{y}}} &= E\{\tilde{\mathbf{y}}\tilde{\mathbf{y}}^{\mathbf{T}}\} \\ &= E\{(\tilde{\mathbf{K}}\tilde{\mathbf{x}} + \tilde{\epsilon})(\tilde{\mathbf{K}}\tilde{\mathbf{x}} + \tilde{\epsilon})^{\mathbf{T}}\} \\ &= \tilde{\mathbf{K}}\tilde{\mathbf{K}}^{\mathbf{T}} + \mathbf{I}. \end{aligned} \tag{B.37}$$

Unfortunately this process does not produce a diagonal matrix either, so we make another transformation on the model by applying the singular value decomposition to $\tilde{\mathbf{K}} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\mathbf{T}}$ (where $\mathbf{\Lambda}$ is a diagonal matrix of the eigenvalues, λ_i, of $\tilde{\mathbf{K}}$):

$$\begin{aligned} \mathbf{y}' &= \mathbf{U}^{\mathbf{T}}\tilde{\mathbf{y}} \\ &= \mathbf{U}^{\mathbf{T}}\tilde{\mathbf{K}}\tilde{\mathbf{x}} + \mathbf{U}^{\mathbf{T}}\tilde{\epsilon} \\ &= \mathbf{\Lambda}\mathbf{V}^{\mathbf{T}}\tilde{\mathbf{x}} + \epsilon' \\ &= \mathbf{\Lambda}\mathbf{x}' + \epsilon'. \end{aligned} \tag{B.38}$$

This transformation leaves the covariance of ϵ' and $\mathbf{x}'$ as unit matrices and the prior covariance of $\mathbf{y}'$ is $\mathbf{\Lambda}^{\mathbf{2}} + \mathbf{I}$, which is a diagonal matrix. Thus the elements of $\mathbf{y}'$ which vary more than the noise are those which have eigenvalues greater than or about 1, $\lambda_i \gtrsim 1$.

From this we can therefore say that the number of independent measurements made to better than the level of measurement error is equal to the number of eigenvalues of $\mathbf{S_\epsilon}^{-\frac{1}{2}}\mathbf{K}\mathbf{S_a}^{-\frac{1}{2}}$ which are greater than unity.

If we apply the definition of $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ to Equations (B.29) and (B.34) we can re-write them as

$$\begin{aligned} H_s &= \frac{1}{2}\ln|\mathbf{S}_{\mathbf{a}}^{\frac{1}{2}}\mathbf{K}^{\mathbf{T}}\mathbf{S}_{\epsilon}^{-1}\mathbf{K}\mathbf{S}_{\mathbf{a}}^{\frac{1}{2}} + \mathbf{I}_{\mathbf{n}}| \\ &= \frac{1}{2}\ln|\tilde{\mathbf{K}}^{\mathbf{T}}\tilde{\mathbf{K}} + \mathbf{I}_{\mathbf{n}}|, \qquad \text{(B.39)} \\ & \qquad \text{(B.40)} \end{aligned}$$

and

$$\begin{aligned} H_m &= \frac{1}{2}\ln|\mathbf{S}_{\epsilon}^{-\frac{1}{2}}\mathbf{K}^{\mathbf{T}}\mathbf{S}_{\mathbf{a}}\mathbf{K}\mathbf{S}_{\epsilon}^{-\frac{1}{2}} + \mathbf{I}_{\mathbf{m}}| \\ &= \frac{1}{2}\ln|\tilde{\mathbf{K}}\tilde{\mathbf{K}}^{\mathbf{T}} + \mathbf{I}_{\mathbf{m}}|. \qquad \text{(B.41)} \\ & \qquad \text{(B.42)} \end{aligned}$$

Since $\tilde{\mathbf{K}}^{\mathbf{T}}\tilde{\mathbf{K}}$ and $\tilde{\mathbf{K}}\tilde{\mathbf{K}}^{\mathbf{T}}$ have the same non-zero eigenvalues, this shows H_s and H_m are equivalent.

Appendix C

Instrument hardware

This appendix is intended to provide detailed information on the mechanical and electronic components used in the SPARCLE prototype produced in the course of this study.

C.1 Electronic components

C.1.1 LDA circuitry

The linear diode array selected for use in this project is the Hamamatsu S5462-512Q CMOS Linear Image Sensor. This device consists of a linear array of 512 photodiodes with a built in CMOS shift register, integration amplifier and clamp circuit. It is coupled with the C5863 miniature driver/amplifier circuit, which is manufactured by Hamamatsu specifically for use with its range of CMOS linear sensors. This circuit includes a regulated power supply for the sensor chip, buffers for all of the sensor control lines and a gain amplifier for the data signal. Inclusion of this circuit allows the chip to be driven from a single 5V supply and relaxes the transition time requirements on the control signals. The layout and dimensions of both the sensor and driver circuit board are given in Figure (C.1).

Data is clocked off the S5462-512Q using six control lines, as shown in Figure (C.2). The LDA itself provides two output signals termed monitor video and data video – the latter having been passed through the internal clamp circuit, the former directly from the integration amplifier. The C5863 driver circuit only uses the data video output however and provides a positive analog voltage

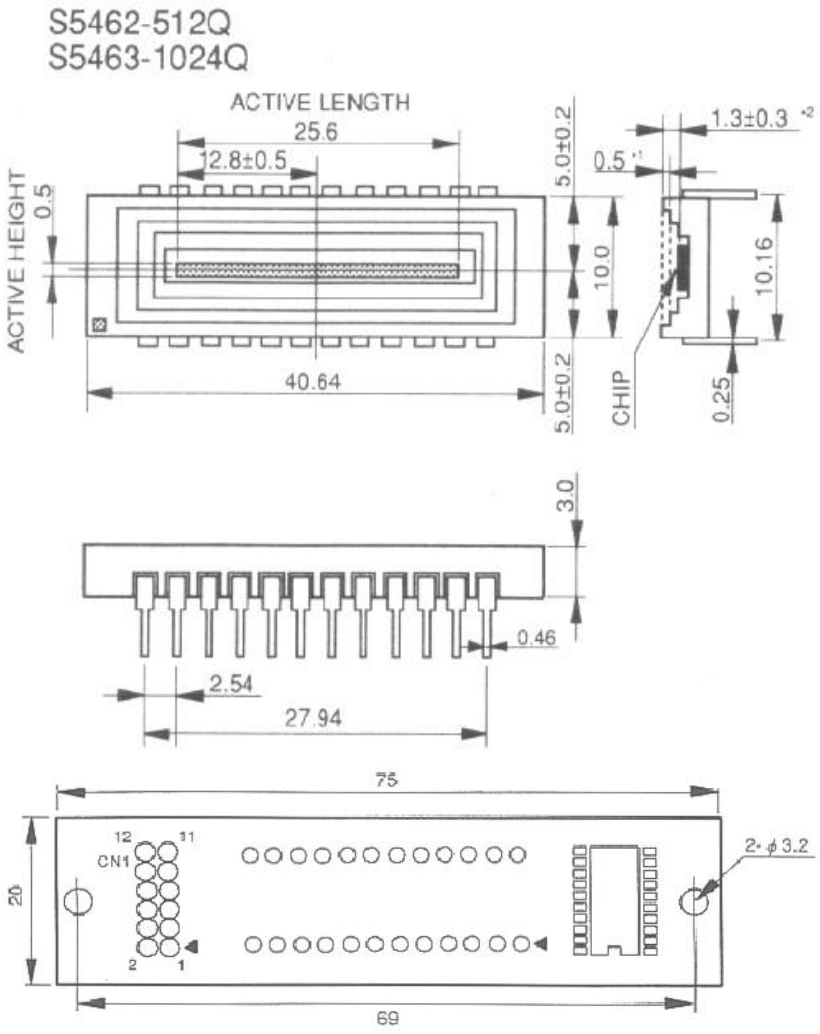


Figure C.1: The layout and dimensions of the linear sensor and the driver/amplifier circuit used. As can be seen the active area is 25.6 mm long and 0.5 mm wide, giving a diode pitch of 50 μm. The active area is also centred both length and width-wise in the package. (Source: Hamamatsu data sheet.)

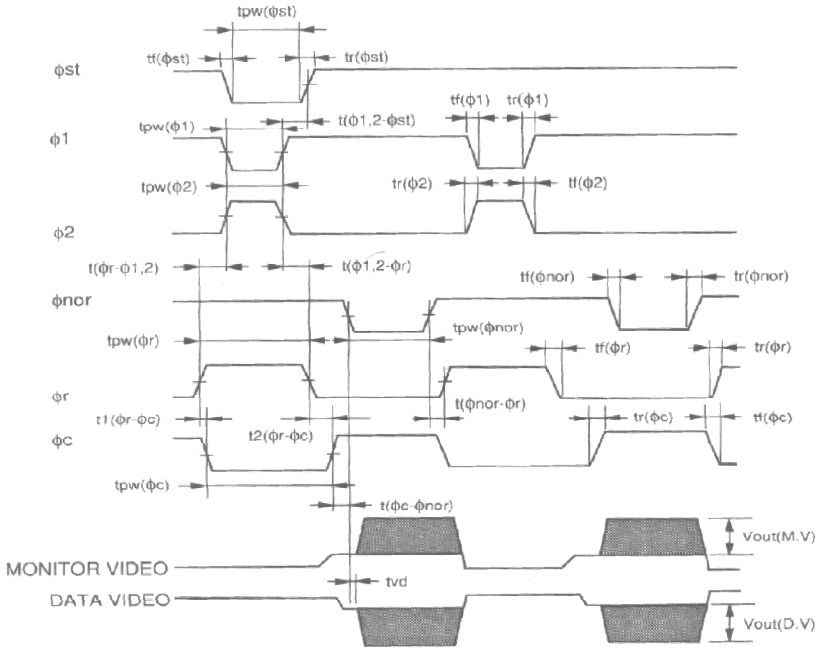
representing the charge accumulated on each pixel as the sensor is read out.

In addition to the six control lines mentioned there is also a Gain selection line which allows the gain of the integrating amplifier to be increased by a factor of approximately 3 (low logical level selects high gain). Finally, the chip also provides an 'End of Scan' output, which goes low to indicate all of the pixels have been read out.

Prior to digitization the data video output from the LDA driver board is passed through the simple analog processing circuit shown in Figure (C.3). As can be seen there are 3 manual adjustments on this circuit, labeled POT A, POT B and GAIN. POT A is the zero level control for the PGA207 OP-AMP. This adjustment is required because the LDA has an output voltage offset of approximately 100 mV, as well as allowing for compensation of a background signal. POT B is used to set the voltage range seen at the input of the ADC to be ± 5 V (rather than 0 to +10 V). This is done after the PGA207 so that the gain and zero level can be adjusted independently. Finally, the GAIN selection on the PGA207 is performed using jumpers to control the digital gain selection pins on the chip.

Setting the zero level of the PGA207 is done while the LDA is being continuously read out and being kept completely dark; this provides the zero light level signal. The signal is then monitored by an oscilliscope placed on the output of the PGA207 (pin 11 or 12) and POT A is adjusted until the lowest output (that is the lowest values of thermal noise) is zero volts.

The LDA control system is based around the Atmel AT90S8515 microcontroller and the Burr-Brown ADS7800 12-bit analog to digital converter (ADC). A schematic of the circuit is shown in Figure (C.4). The AT90S8515 is an AVR 8-bit RISC based processor with inbuilt programmable flash memory and 4 8-bit I/0 ports. The microcontroller provides the signals required by the LDA controller board, governs the digitization and buffering of the LDA data video signal and, when required, transmits data to a PC via an 8-bit parallel interface card. The use of this microcontroller allows data to be recovered from the LDA at an extremely high rate (approximately 5 ms per sample) and provides a very stable time base for the sampling.



Parameter	Symbol	Minimum	Maximum
Start pulse width	tpw(Øst)	200	-
Clock pulse width	tpw(Ø1,2)	200	-
Clock NOR width	tpw(Ønor)	300	-
Reset pulse width	tpw(Ør)	400	-
Clamp pulse width	tpw(Øc)	500	-
Clamp pulse rise/fall times	tr/tf(Øc)	0	1000
All other rise/fall times	tr/tf(ØX)	0	100
Start pulse - clock pulse	t(Ø1,2-Øst)	0	-
Reset pulse - clock pulse	t(Ør-Ø1,2)	100	-
Clock pulse - reset pulse	t(Ø1,2-Ør)	100	-
Reset pulse - clamp pulse 1	t1(Ør-Øc)	0	-
Reset pulse - clamp pulse 2	t2(Ør-Øc)	100	-
Clamp pulse - clock NOR	t(Øc-Ønor)	100	-
Clock NOR - reset pulse	t(Ønor-Ør)	100	-

Figure C.2: The timing requirements for reading out the S5462-512Q sensor. All times are in nano seconds. Note that the C5863 driver circuit allows for much longer rise/fall times. (Source: Hamamatsu data sheet.)

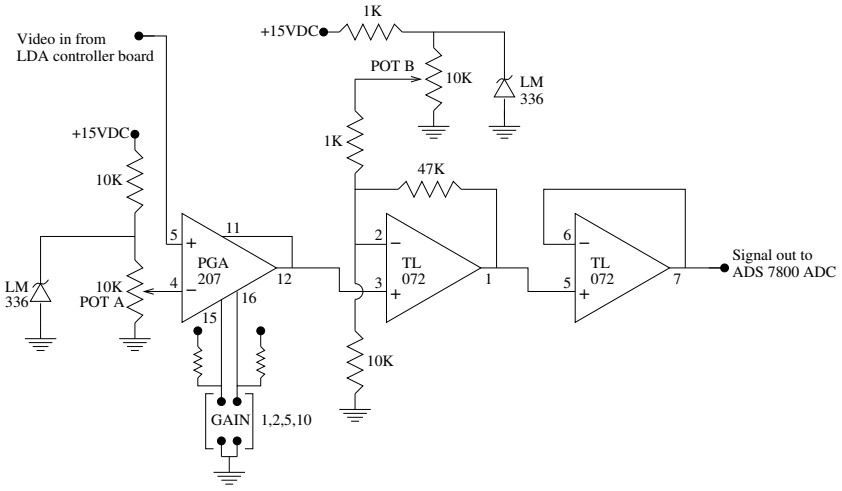


Figure C.3: Circuit diagram of the LDA signal amplifying circuit.

The ADS7800 provides a 3 μs conversion time and supplies the data in the form of two bytes. The ADC can be controlled with two lines alone: a start conversion (R/$\bar{\text{C}}$) line and a high/low byte selection (HBE) line. These are linked to the first two I/O lines of port B on the microcontroller, with the ADC Busy line taking the third I/O line. The analog to digital conversion is initiated by taking the R/$\bar{\text{C}}$ line low while holding HBE low; this sends the BUSY line low indicating conversion is in progress and after approximately 3 μs BUSY returns high and the low-byte (8 least significant bits) becomes available on the data lines. If HBE is then taken high the high-byte becomes available on the data lines, the 4 bits occupying the 4 least significant lines and '0s' occupying the other 4 lines.

As mentioned above the microcontroller provides the necessary control signals for the LDA control board. These signals are passed through the remaining five I/O lines on port B (PB3 – 7), plus the first two lines on port D (PD0 and 1), as shown in Table (C.1) and Figure (C.4). The direct connection of control lines of both the ADC and LDA control board allows the data collection process to be controlled directly from software. Figure (C.5) schematically shows the timing diagram for the signals for the reading of the first two pixels of an LDA profile.

As well as controlling the read-out of the LDA, the microcontroller must

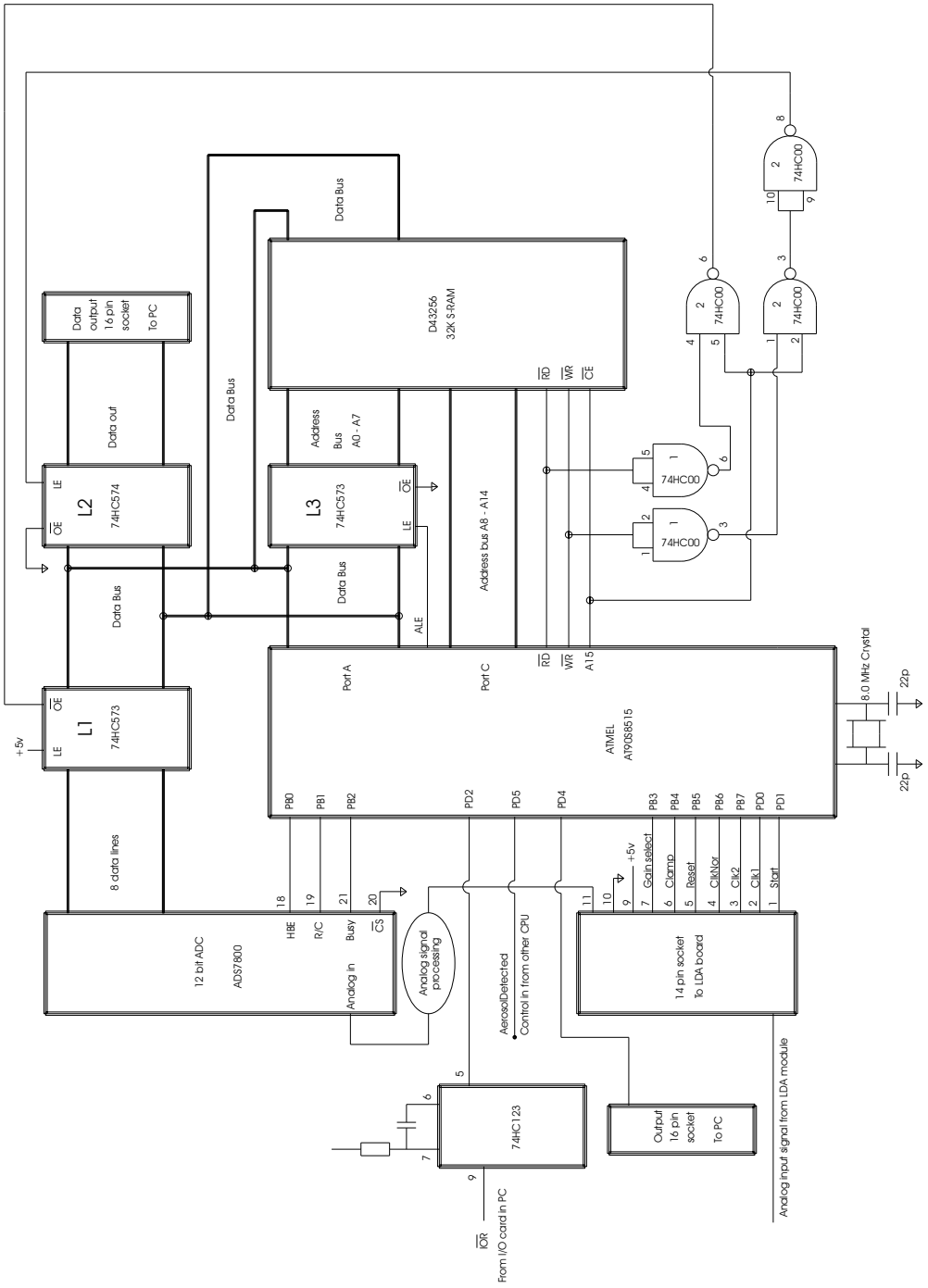


Figure C.4: Circuit diagram of the LDA microcontroller circuit. The "AerosolDetected" line is connected to the PMT microcontroller circuit (see Figure (C.9)) and provides a trigger signal for data to be recorded. The circuit shown in Figure (C.3) is denoted by the "Analog signal processing" ellipse in this diagram.

Data line	C5863 pin	Symbol	Name	Description
PD1	1	Øst	Start	Begins the shift register operation, i.e. initiates a new read out cycle. Negative logic.
PD0	2	Ø1	Clock 1	Operates the shift register. Negative logic.
PB7	3	Ø2	Clock 2	Second clock for timing purposes. Positive logic.
PB6	4	Ønor	Clock NOR	Video output is obtained and syncronized with the rise of Ønor.
PB5	5	Ør	Reset	High level resets the integration amplifier.
PB4	6	Øc	Clamp	High level enables the clamp circuit on the output.
PB3	7	Vg	Gain Select	High level selects "Low gain", low level selects "High gain".
n/a	8	EOS	End of Scan	Output signal to indicate all pixels have been read out. Negative logic.
n/a	9	$+V_D$		+5V power supply.
n/a	10	D.GND		Digital ground.
n/a	11	D.V.	Data video	Analog video output. Positive polarity.
n/a	12	A.GND		Analog ground.

Table C.1: Descriptions of the connector pins for the C5863 driver circuit, including their corresponding data bit on the PC interface card (where applicable). All digital inputs and outputs are H-CMOS compatable.

also deal with the data being measured by the ADC. This involves two separate tasks. First, the data from each LDA reading must be buffered using onboard RAM. Second, on recieving a trigger signal, the microcontroller must present the buffered LDA data to a data logging PC via an 8-bit ISA parallel interface card.

As mentioned above AT90S8515 has four 8-bit I/O ports. Two of these are used for controlling the LDA driver and the ADC, as well as for recieving the measurement trigger pulse and handshaking with the data logging computer. This leaves 2 ports (A and C) for control of the RAM chip and transfering data. The AT90S8515 provides port A specifically as a multiplexed address/data port and port C as a dedicated address port when using external RAM. In this case

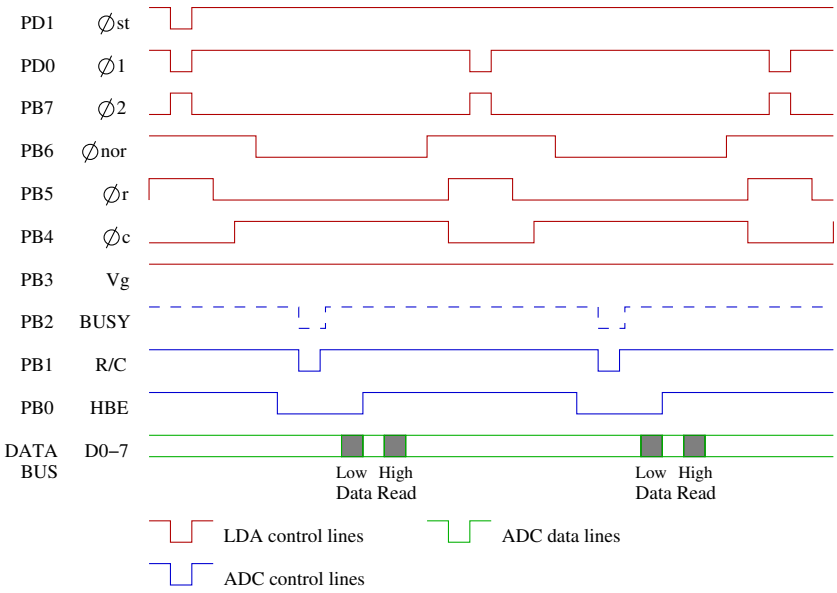


Figure C.5: The I/O timing used to control the LDA and ADC circuitry.

there are 32 kBytes of external RAM, requiring a 15 bit address bus and leaving one address line free.

As shown in Figure (C.4) the address/data bus on port A uses 3 latches to control data flow. L1 is used as a switch, either connecting the output of the ADC to bus, or appearing as an open circuit. L2 is a rising edge transition flip-flop used to latch data onto the parallel interface to the PC. L3 is used to latch address data from port A onto the address bus. Both L1 and L2 are controlled with the remaining address line from port C (A15) in conjunction with the data read ($\overline{\text{RD}}$) amd write ($\overline{\text{WR}}$) lines on the microcontroller. These three lines are also used to select and control the RAM chip, using its data read, data write and Chip Enable ($\overline{\text{CE}}$) lines.

This mulitplexing results in the following procedure for reading data from the ADC and buffering it in memory:

1. The microcontroller begins an ADC conversion.

2. When the ADC busy line returns to the high state at the end of the conversion the NAND gates are enabled by taking A15 high. This also has the effect of disabling the RAM chip by taking its $\overline{\text{CE}}$ line high (i.e. it will

ignore signals on its $\overline{\text{RD}}$ and $\overline{\text{WR}}$ lines as well as on the data bus).

3. The microcontroller then does a read on port A. This briefly puts the $\overline{\text{RD}}$ low, which in turn takes the output enable line $\overline{\text{OE}}$ of L1 low. With $\overline{\text{OE}}$ low L1 is a transparent connection to the data lines of the ADC, otherwise it is in a high impedence state.

4. With the data byte read into internal memory the microcontroller then deselects the data latches and enables the RAM chip by taking A15 low, while at the same time putting the memory address for this byte onto port A and the first seven lines of port C.

5. This address is then latched onto L3 by taking the Address Latch Enable line ALE on the microcontroller high.

6. The data byte is then written back to port A, which sends the $\overline{\text{WR}}$ signal to the RAM chip and the byte is stored.

7. The microcontroller then requests the high-byte from the ADC and steps 3 to 5 are repeated.

8. Repeat.

The number of profiles stored in memory is determined in software and for the measurements made in this work it was set to three (i.e. the profile currently being read out, plus the two previous profiles). This entire process is repeated continuously until the trigger signal is recieved. This trigger signal consists of a short pulse on the 6th line of port D (PD5) and in normal operation[1] is sent by the photomultiplier control circuit (see Section (C.1.2)).

On recieving the trigger pulse the LDA microcontroller performs the following sequence:

1. The profile currently being read off the LDA chip and the one following it are read and stored in memory according to the algorithm given above.

2. Next, the RAM chip is enabled (and the data latches disabled) by taking A15 low, as well as placing the memory address of the first byte onto the RAM chip's address bus.

[1]For testing purposes a suitable pulse from a signal generator can also be used, allowing LDA data collection independant of the PMT system.

3. The address is then latched onto L3 as described previously.

4. The microcontroller then reads port A. Again, this briefly takes the $\overline{\text{RD}}$ line low, causing the RAM chip to place the data byte onto the bus.

5. The RAM is then disabled and the data latches enabled by taking A15 high and the data is written to port A. This causes a brief pulse on the LE line of L2, which latches this byte onto the data output.

6. If this is the first byte, line 5 of port D (PD4) (which is connected to the computers ISA interface card) is set high. This signals the computer that data is available on the parallel data lines.

7. The controller then waits for the $\overline{\text{IOR}}$ signal to be recieved from the PC (on line PD2). When the signal arrives steps 1 to 5 are repeated until all data has been transmitted.

8. Finally PD4 is taken low again (indicating to the computer that there are no more data available) and the microcontroller goes back to data reading as described above.

The interface to the PC used by this system is a direct connection to two ports on the PC 8-bit ISA bus. The interface card simply links two 16-pin connectors to two addresses on the ISA bus. One of these connectors is directly linked to the data output port on the controller card. The other carries the "data avaliable" line from the controller on its first bit, with all other bits floating. The wiring expected by the controller for the 16-pin connectors is shown in Table (C.2). The $\overline{\text{IOR}}$ is that generated by the ISA bus during a read on that particular port.

C.1.2 PMT circuitry

The photomultiplier tube (PMT) selected for this investigation was the Electron Tubes Inc. 9828. This tube was selected on the basis of the sensitivity of the tube at the original laser wavelength (near 650 nm), the overall gain of the tube and cost. This PMT has a S20 Trialkali (made up of a mixture of Sodium, Potasium, Antimony and Cesium) photocathode with a prismatic window. This combination gives the highest quantum efficiency for wavelengths in the approximate interval 600 nm to 900 nm of any photocathode, with the exception of GaAs

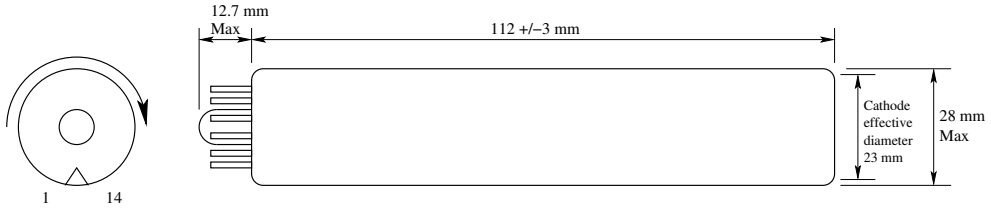
Pin	Connection	Pin	Connection
0	Data bit 1	8	NC
1	Data bit 2	9	Earth
2	Data bit 3	10	Earth
3	Data bit 4	11	Earth
4	Data bit 5	12	Earth
5	Data bit 6	13	Earth
6	Data bit 7	14	Earth
7	Data bit 8	15	$\overline{\text{IOR}}$ signal from computer

Table C.2: Pin connections for the LDA controller's interface to the PC ISA bus. Note that the pin 15 connection is only needed for the data port.

(Gallium-Arsenide). The quantum efficiency is quoted as being approximately 5% at 660 nm. A GaAs photomultiplier was not used because these are prohibitively expensive. The tube has an eleven element dynode chain, providing a gain of 10^6.

The tube's dimension and the arrangement of the connector pins are displayed in Figure (C.6). The tube was bought with a voltage divider and socket unit as well as a magnetically shielded stainless steel housing, as shown in Figure (C.7). This housing includes an o-ring on the mounting flange (to allow for a light tight mount). The voltage divider used is configured for a negative supply, meaning that the photocathode is placed at a high negative voltage and the dynode voltages become progressively less negative, with the anode being at earth. This arrangement produces a continuous DC current output from the anode, with the size of the current being approximately linearly related to the intensity of light falling on the photocathode.

As with the LDA control circuit the PMT data aquisition can be thought of as being made up of two parts, namely analog signal processing followed by digital signal processing. The analog part of the circuit is displayed in Figure (C.8) and consists of amplification and low pass filtering of the signal. The DC current supplied by the PMT is first converted into a voltage in the range 0 to 10 V by an OP-AMP. The signal then passes into a four pole low-pass filter with a cut-off frequency of approximately 20 kHz, which removes the higher frequencies of shot noise produced by the PMT. The signal is passed through a second amplification stage to provide the -10 to 10 V voltage range expected by the



Pin Number	Pin Name	Description	Pin Number	Pin Name	Description
1	ic	Internal connection.	8	d10	10th dynode.
2	d3	3rd dynode.	9	d8	8th dynode.
3	d5	5th dynode.	10	d6	6th dynode.
4	d7	7th dynode.	11	d4	4th dynode.
5	d9	9th dynode.	12	d2	2nd dynode.
6	d11	11th dynode.	13	k	Photocathode.
7	a	Anode (output).	14	d1	1st dynode.

Figure C.6: The physical dimensions and pin configuration of the 9828 photomultiplier tube. The pin number increases clockwise when the tube is viewed from below. The triangle indicates the position of the short locating pin.

ADC. This final OP-AMP includes 10 turn adjustment pots for both zero-offset and gain to allow for changes in background light levels and photomultiplier gain. The circuit uses the 16-bit ADS7805P analog to digital converter to deal with the huge dynamic range of the PMT signal (up to 5 orders of magnitude for the 9828 tube). This ADC is very similar in configuration to the ADS7800 used in the LDA control circuit, with two control lines (Start and H/L Byte) and 8 data lines presenting the result of each conversion as two 8 bit bytes.

A schematic of the digital portion of the circuit can be seen in Figure (C.9) and as with the LDA controller is based around the Atmel AT90S8515 microcontroller. The circuit uses an RS232 interface to transmit data to the data logging PC, and like the LDA controller the ADC and data transmission are controlled directly by the microcontroller. The use of serial data transmission removes the need for a multiplexed data bus and makes more complicated communication between the computer and microcontroller possible. This two way communication capability is of importance because, as well as performing data collection, the PMT microcontroller is used to determine when an aerosol particle has passed through the sample volume and triggers data transmission on the LDA control circuit.

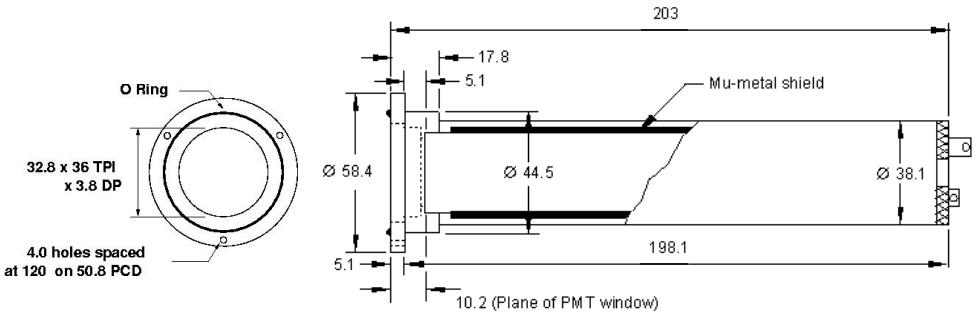


Figure C.7: The physical dimensions of the QL30 stainless steel housing supplied by Electron tubes for mounting their PMTs. (source Electron tubes catalog)

The serial comunication is controlled directly by the microcontroller, which expects and generates TTL logical signals (i.e. low logic level = 0 V, high = 5 V) at a baud rate of 38.4 kbit/s, using 8 bit bytes with 1 stop bit and no parity. The conversion between the TTL levels and RS232 levels (low = 12 V, high = -12 V) is performed by the MAX232 converter. The microcontroller is programmed with a simple instruction set to enable the setting of various parameters as well as receiving requests for measurements to be made; this instruction set is summarized in Table (C.3).

The algorithm followed by the microcontroller for collecting data from the PMT is similar to that followed by the LDA controller:

1. The microcontroller requests an A-D conversion by taking R/C (PD3) low momentarily.

2. Once the busy line (PD5) returns to the high state (indicating the conversion is completed) the low-byte (8 most significant bits) becomes available on the microcontroller's Port B data lines. At this point the RAM chip is enabled by taking C15 ($\overline{\text{CE}}$) low and the memory address for this data byte is placed onto ports A and C. The first 8 bits of the address are latched onto the address bus by taking ALE high, thus freeing Port A to act as the memory data bus.

3. The data byte is then placed on Port A and the RAM chip sent the read signal ($\overline{\text{RD}}$).

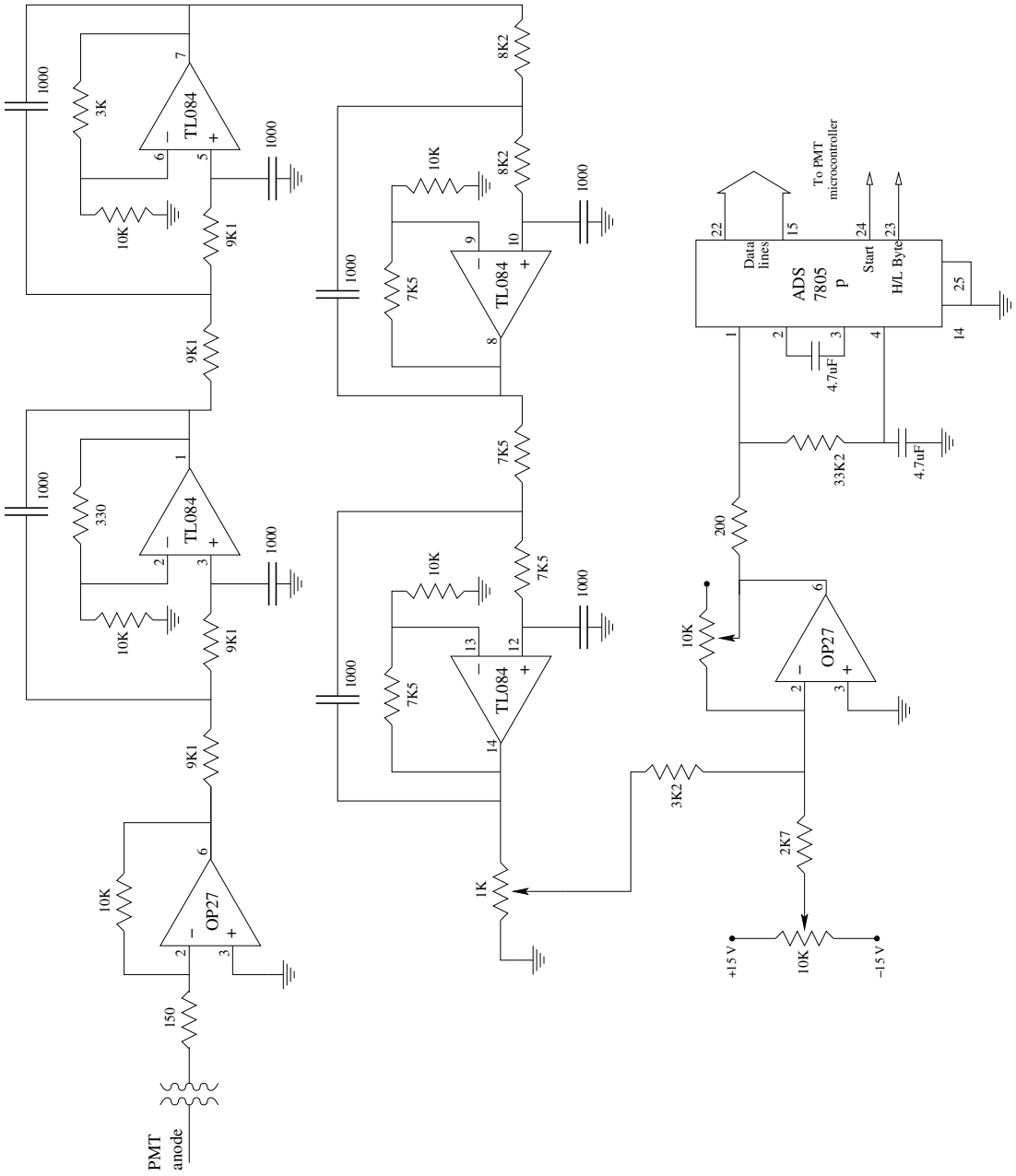


Figure C.8: The circuit which amplifies, low-pass filters and digitizes the PMT signal into a series of 16-bit words.

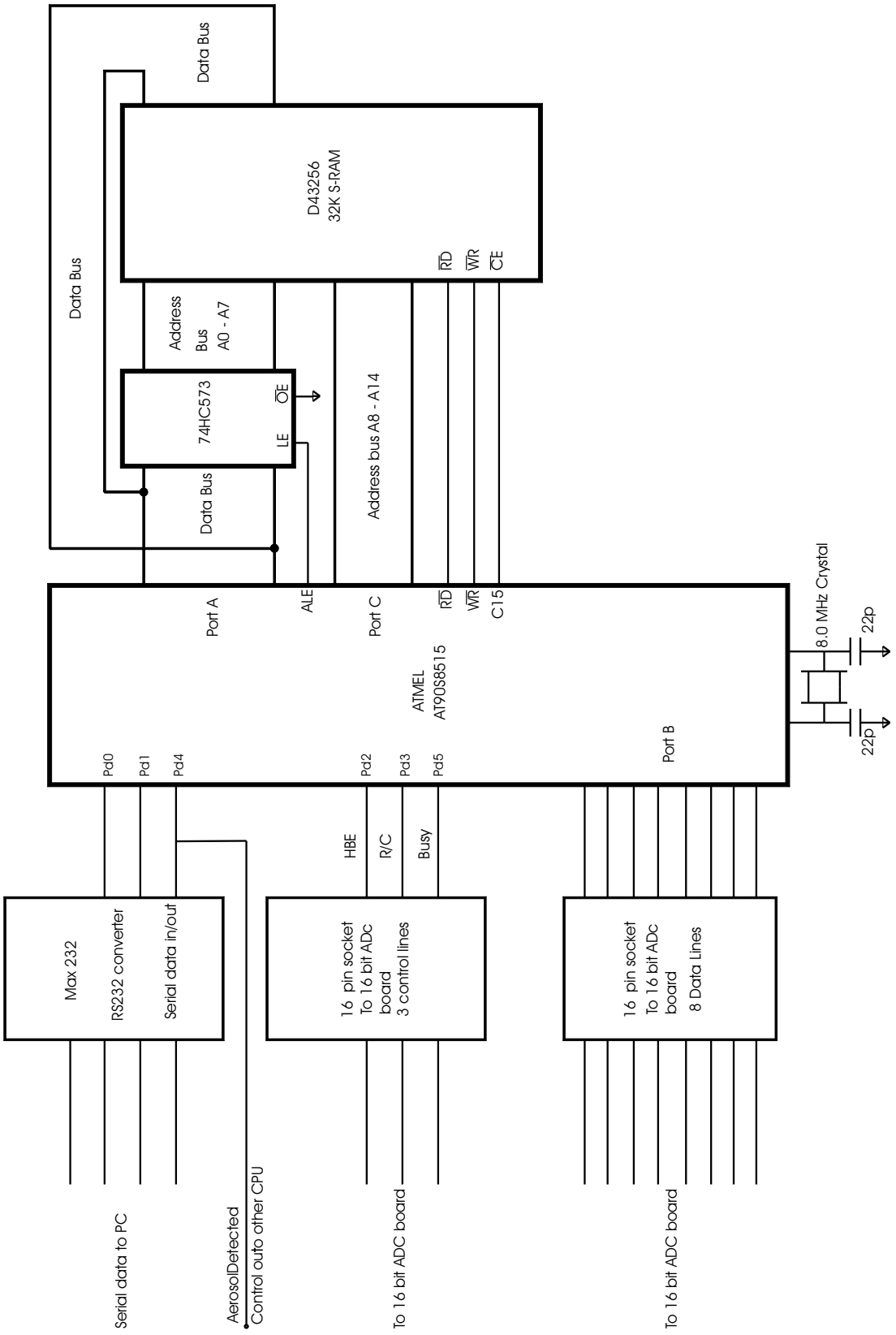


Figure C.9: The microcontroller for PMT data collection. The "AerosolDectected" line is used to trigger data collection on the LDA microcontroller (see Figure(C.4)).

Instruction	Description
'E' hi-byte low-byte	Collect 256×hi-byte+low-byte samples from the PMT and then transmit them to the PC.
'D' hi-byte low-byte	Set the length of the measurement buffer to 256×hi-byte+low-byte (i.e. the number of PMT samples sent to the PC when an aerosol is detected).
'C' hi-byte low-byte sign-byte	Set the aerosol detection threshold level to 256×hi-byte+low-byte. If the sign-byte = 0 then look for a signal greater than the threshold, if sign-byte = 1 look for one lower than the threshold.
'A'	Begin searching for an aerosol particle. When the threshold level is broken the controller continues collecting data until the detection point is half way down the buffer and stops. A 'Z' character is then sent to the computer and the microcontroller waits for any character to be returned from the PC – once this happens the data is returned as a series of bytes ordered as low-byte followed by high-byte.

Table C.3: The instruction set used by the PMT microcontroller. 'A' refers to the byte representing the letter A in the ASCII character set. The controller has a 2 second time-out for the multibyte commands, i.e. if all bytes haven't been received within 2 seconds of the first byte the controller returns to its previous state.

4. The high-byte (8 least significant bits) are then requested from the ADC by taking PD2 (HBE) high and the steps 2 and 3 are repeated to place this data into memory.

When performing a 'background measurement' (see Table (C.3)) this process is repeated until the requested number of measurements have been made, and then the data is re-read from RAM and transmitted to the PC via the RS232 serial interface. In usual measurement mode however the controller also compares each sample to the user set threshold: once 10 consecutive samples break the threshold a particle is said to have been detected. At this point the controller briefly takes line PD4 high, which provides the 'aerosol detected' pulse for triggering the LDA controller[2] (as described in Section (C.1.1)). Defining the length of the data buffer to be N samples, the controller then collects an additional $N/2$ samples (as described in Table (C.3)) before ceasing measurement and sending the data to the computer. The acquisition of a PMT sample and comparison with the threshold requires 14 μs, resulting in a sampling frequency of 71.4 kHz.

C.2 Instrument body

Figures (C.10) to (C.12) show scale drawings of the body of the instrument, as well as the mounts and air flow pipes.

1. Instrument body

Figure (C.10) shows the body of the SPARCLE prototype used in this study. The body is machined from a single piece of 100 mm free machining aluminium bar and is anodised black to reduce reflectivity. The scattering volume is located at the centre of the body and it provides mountings which accurately locate the rest of the instrument components (sensors, laser baffling and air pipes). As shown in the elevation views in Figure (C.10), the laser beam enters the instrument from the top and, after passing through the sample volume, exits the instrument through the tapered tube at the bottom. The air pipes can be mounted so the flow is from either left or right through the tubes entering the body from either side.

[2]This signal is also sent on one of the RS232 control lines, but is not used in the current setup.

The sample volume is open on both sides of the body and there are mounting points for the PMT on one side and the LDA on the other. The PMT housing provided by Electron Tubes (Figure (C.7)) attaches directly to the face of the body, while the LDA is mounted behind a circular plastic mask (not shown).

2. Laser Baffles

Figure (C.11) shows the laser baffle (top) and the mount used to attach it to the instrument body (bottom). The laser baffle is constructed from aluminium and a 75 mm section of thin wall, 7 mm OD, steel tube. This tube contains two circular baffles constructed from a black plastic, which are held in place by two sections of plastic tube which fit tightly within the steel tube. The aluminum section of the baffle was originally designed to hold a R6705GB-AP laser diode module from Intelite Inc., which was used in the development of a previous prototype.

The mounting which holds the baffle in place is machined from free machining aluminum. It fits into the opening at the top of the instrument body (Figure (C.10) and provides for the use of an o-ring seal to prevent air and light leaks. The steel tube on the laser baffle fits snugly within the central hole of the mount and is held in place by a combination of a grub screw and another o-ring. The mount is attached to the instrument body by three screws, which also serve to secure the o-ring holding the baffle in place (via the washer which is shown in Figure (C.11)). As with the instrument body the mounting is anodized black, while the laser baffle is painted black.

3. Air Pipes

The specifications of the air pipes and their mounts are shown in Figure (C.12). The outlet pipe is simply a length of 4.75 mm OD steel pipe with a section of brass tube to provide reinforcing partway along its length.

The inlet pipe is co-axial, as described in Section (3.4). The inner, sample tube passes straight through the component and is made from steel tubing. The outer, sheath flow is introduced by an elbow tube constructed of brass, before passing into another steel tube (which contains the sample tube). The brass elbow also provides a firm mounting for the sample tube and reinforcing for a

section of the sheath flow tube.

The pipe mounts are identical for inlet and outlet pipes. The brass renforced section of the pipes fit snugly in the semi-circular groves machined into the face of the mount and are then secured by four screws. The pipe mounts also secure o-ring seals between the pipes and the instrument body.

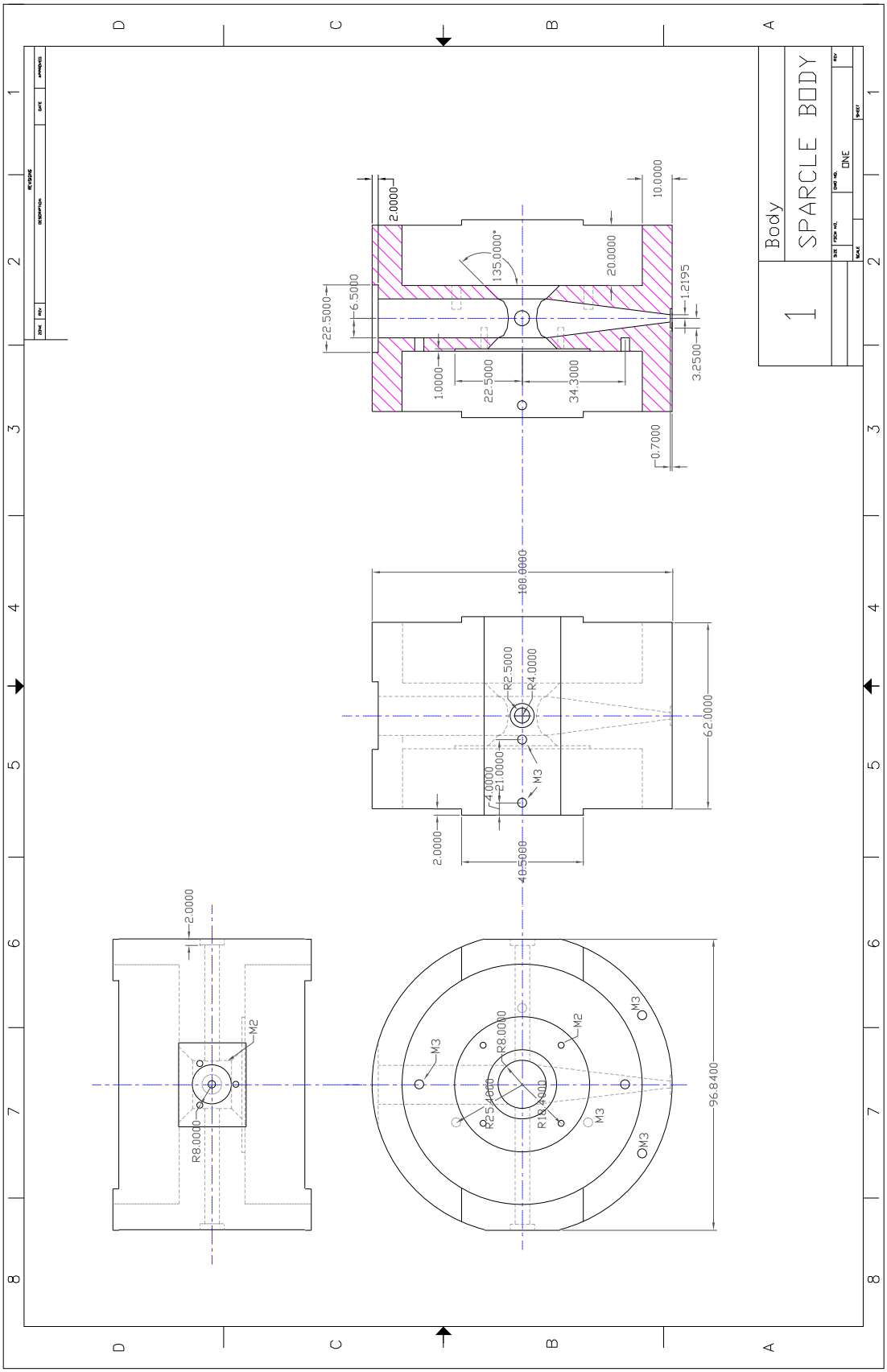


Figure C.10: Body of the SPARCLE prototype.

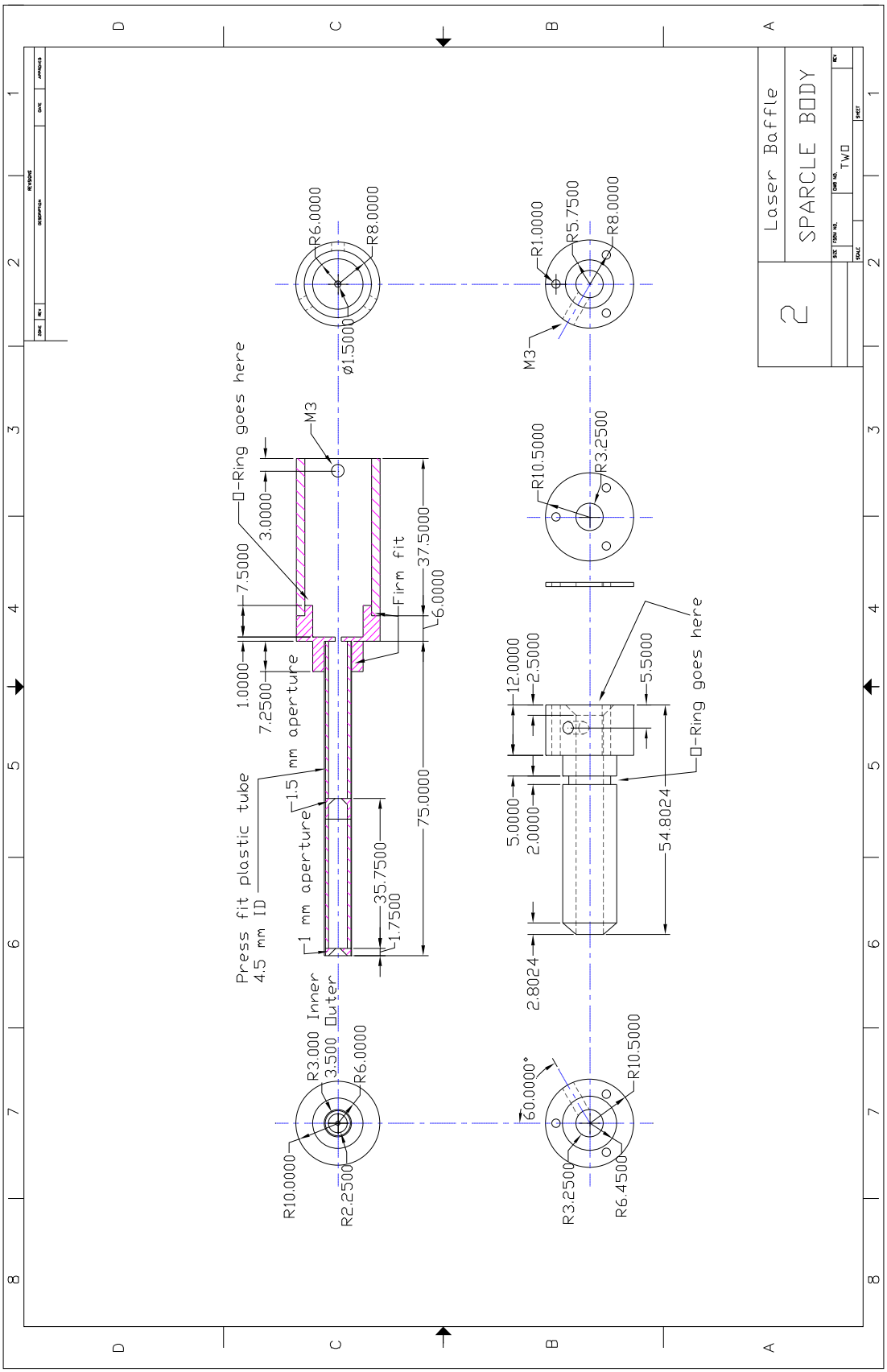


Figure C.11: Laser baffle used in the SPARCLE prototype.

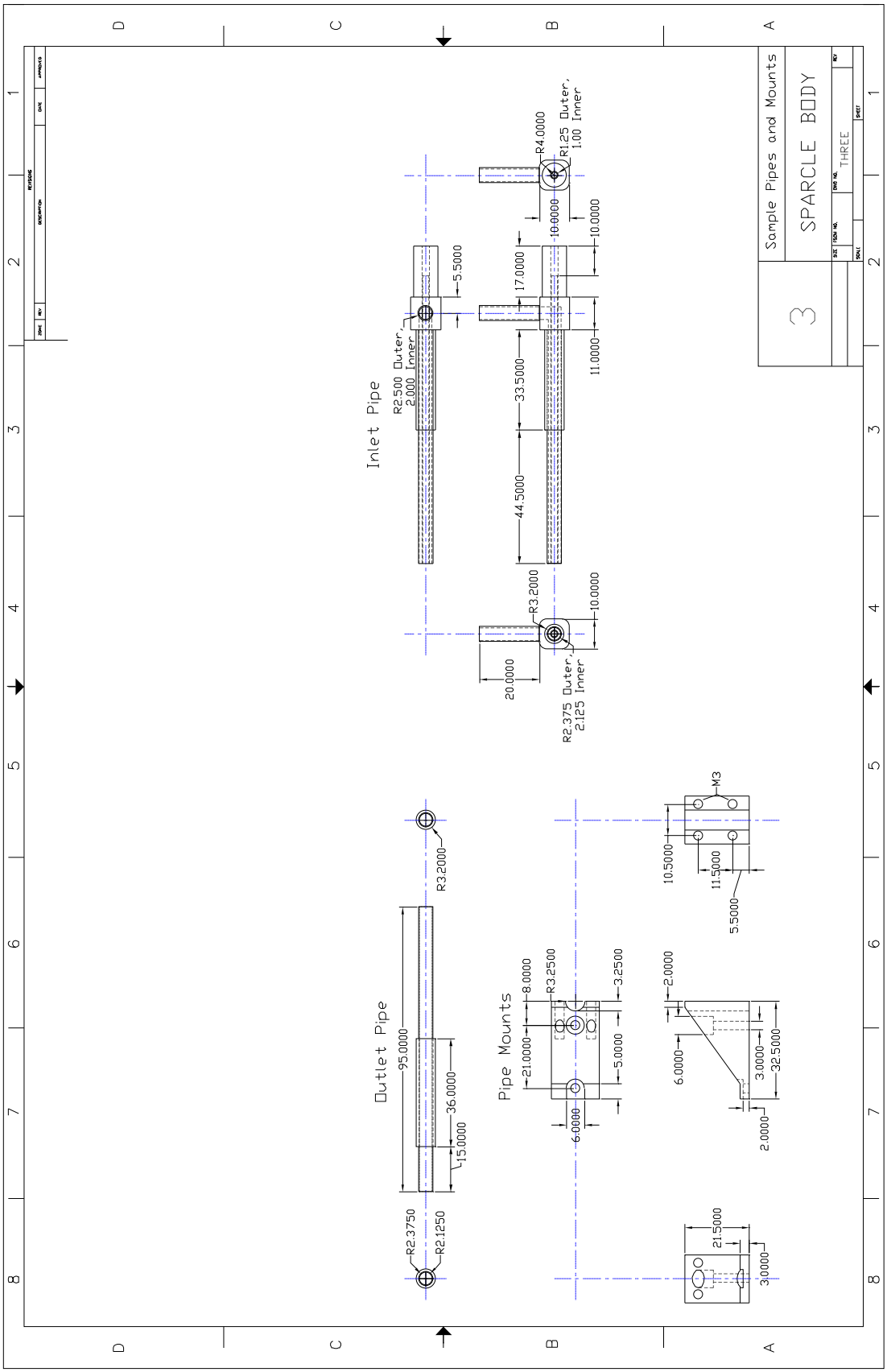


Figure C.12: Air inlet and outlet pipes from the SPARCLE prototype.

C.3 Pumps

Figure (C.13) shows exploded and assembled views of the large gear pump used to provide the sheath flow. The pump body is constructed from free machining aluminum in three sections – the dimensions of the central section being given in Figure (C.14). The gears used are from the oil pump of a 1600 cm^3 Volkswagen Beatle engine, milled down to the appropiate thickness. The axle of the driven gear runs on two sealed ball bearing races, one mounted in each of the pump cover plates, while the free gear runs on a roller bearing mounted between the gear and its axle. In both cases these axles are made of hardened steel. Not shown in these figures are the Teflon gaskets which separate the cover plates from the central body section and gears. These gaskets seal the body and provide lubrication for the gears.

The motor used on this pump was a Barber-Colman (now Colman Motor Products) FYQF 63310-9 12 V DC servo motor, supplied by Servo Systems. This motor includes a reducing gear head which enables it to provide high torque at low speeds.

The smaller gear pump used to generate the sample flow is shown in figures (C.15) and (C.16). As with the larger pump the body is of free machining aluminum construction, however it is constructed from five parts rather than three. Between the central section and the two cover plates in which the gear bearings are mounted, there are two thin plates which provide a firm backing for the teflon gasket. This is required because the bearings are approximately the same diameter as the gears, which would result in a poor seal as the gear teeth worked on the teflon.

The gears used here are Boston Gear 24 pitch, 12 tooth spur gears (part number 09630). Both driven and free gears are mounted on axels which are supported by sealed ball bearings mounted top and bottom. This pump used a Premotec CL29-3 Watt Coreless 12 V DC servo motor, which was directly coupled to the pump.

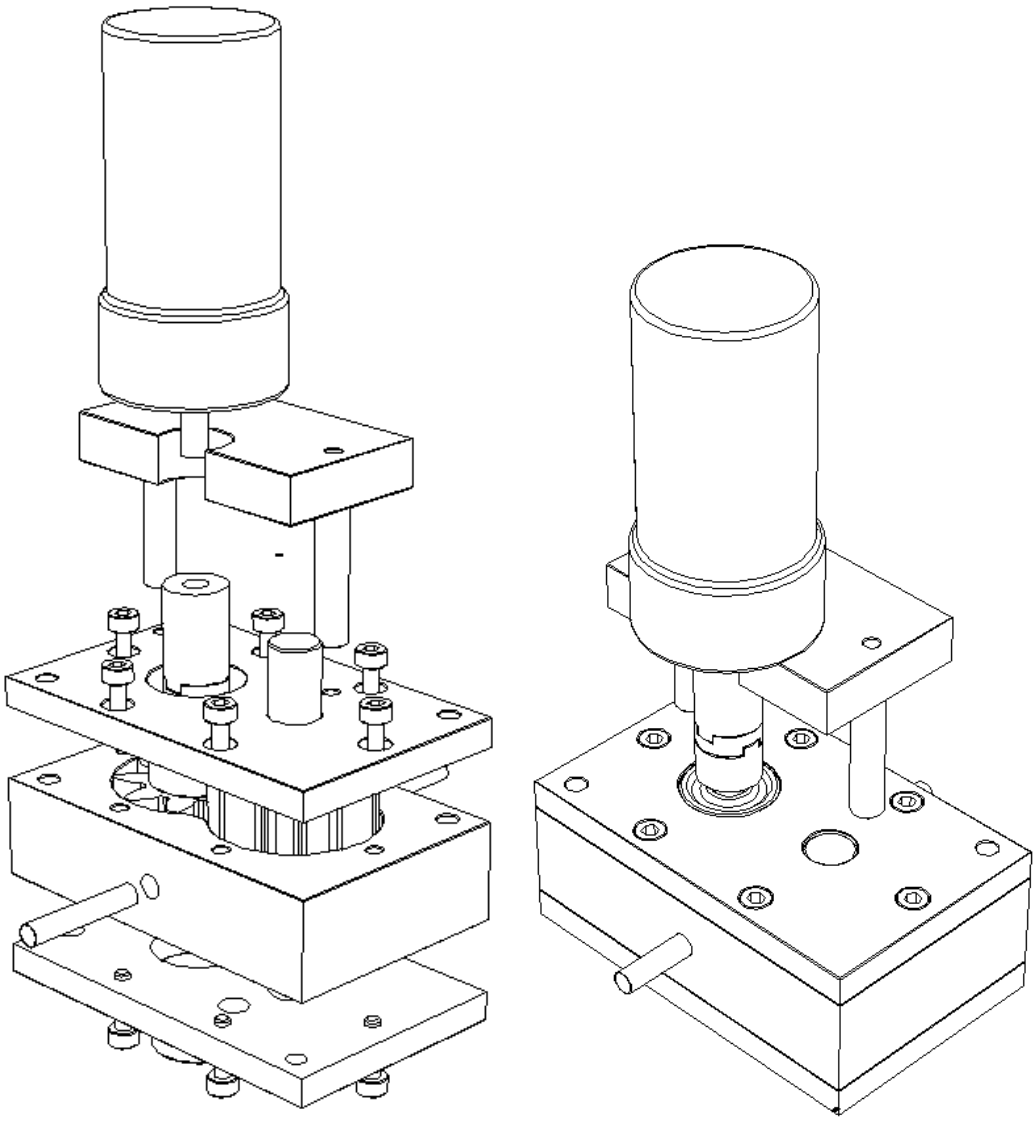


Figure C.13: Isometric views of the large gear pump used in the SPARCLE instrument.

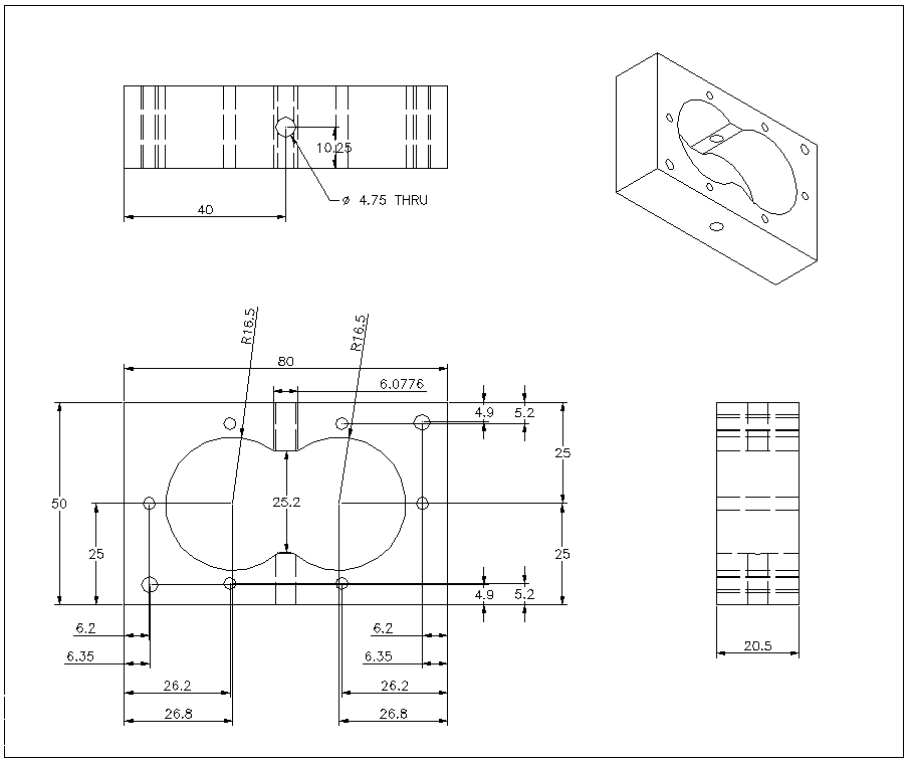


Figure C.14: Dimensions of the large gear pump body.

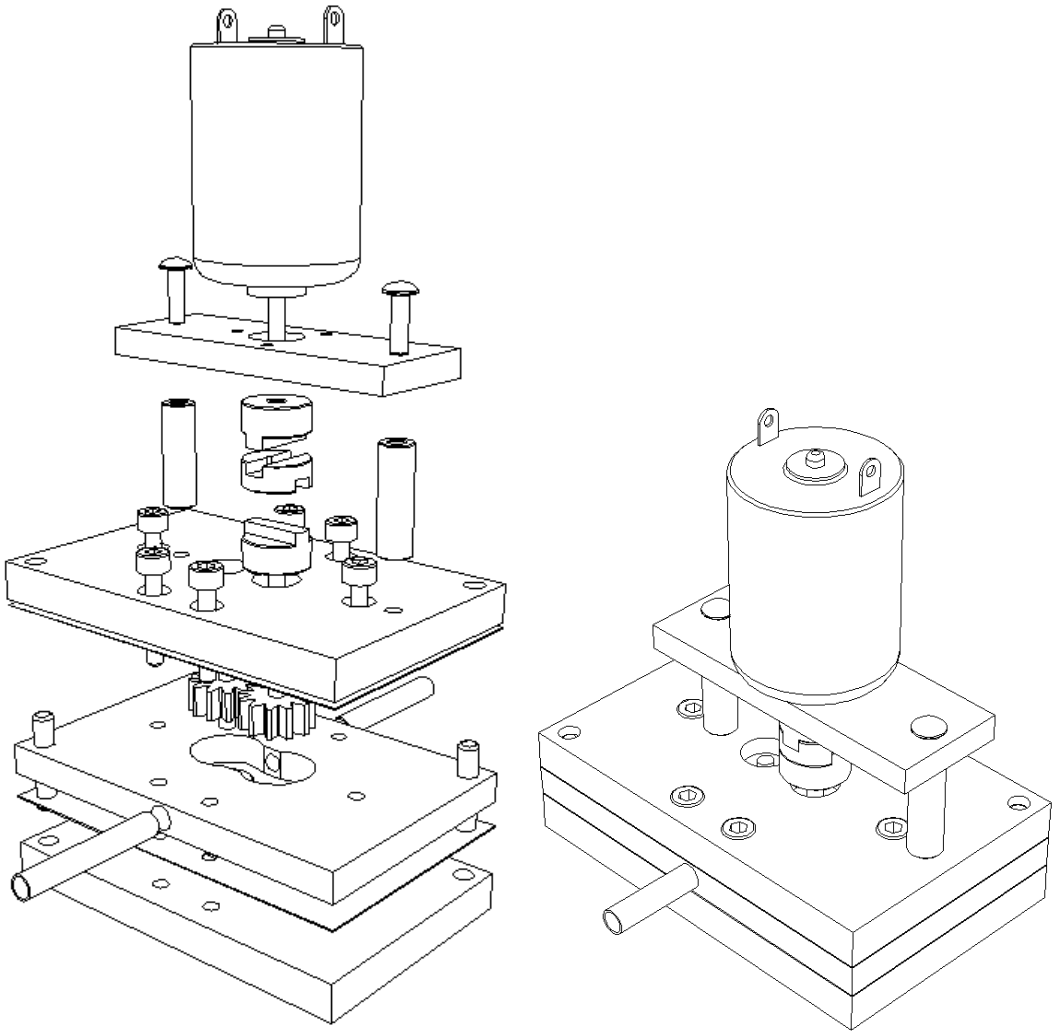


Figure C.15: Isometric views of the small gear pump used in the SPARCLE instrument.

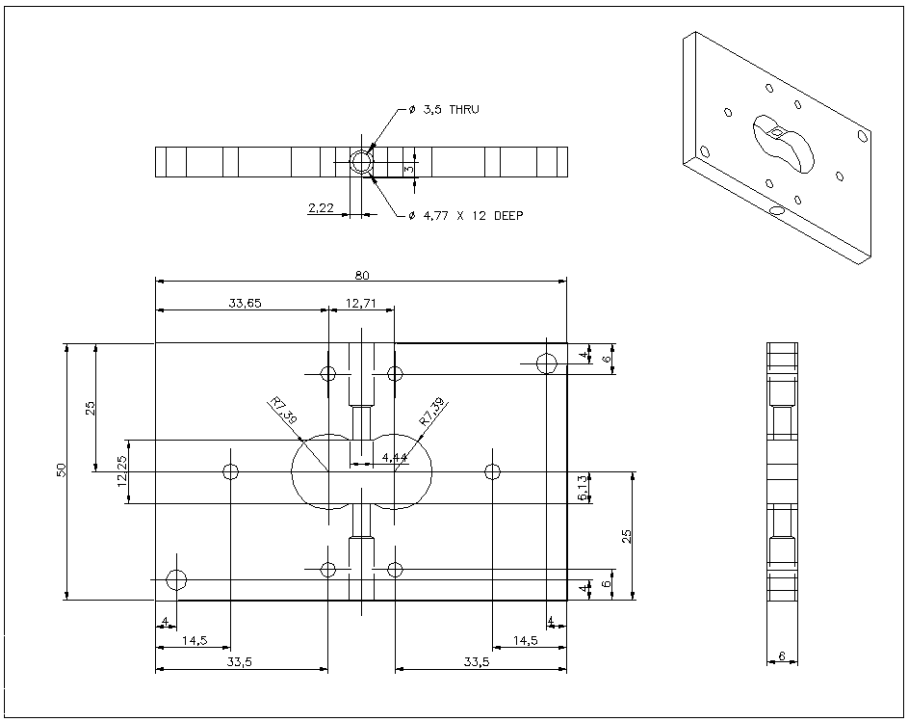


Figure C.16: Dimensions of the small gear pump body.

References

Adriani, A., Deshler, T., Gobbi, G. P., Johnson, B. J., & Di Donfrancesco, G. 1992. Polar stratospheric clouds over McMurdo, Antarctica, during the 1991 spring: lidar and particle counter measurements. *Geophys. Res. Lett.*, **19**(17), 1755–1758.

Adriani, A., Cairo, F., Pulvirenti, L., Cardillo, F., Viterbini, M., Di Donfrancesco, G., & Pommereau, J. P. 1999. Stratospheric background aerosol and polar cloud observations by laser backscattersonde within the framework of the European project "Stratospheric Regular Sounding". *Ann. Geophysicae*, **17**(10), 1352–1360.

Anderson, B. E., Cofer, W. R., Bsrrick, J. D., Bagwell, D. R., & Hudgins, C. H. 1998a. Airborne observations of aircraft aerosol emissions I: Total nonvolatile particle emission indices. *Geophys. Res. Lett.*, **25**(10), 1689–1692.

Anderson, B. E., Cofer, W. R., Barrick, J. D., Bagwell, D. R., & Hudgins, C. H. 1998b. Airborne observations of aircraft aerosol emissions II: Factors controlling volatile particle production. *Geophys. Res. Lett.*, **25**(10), 1693–1696.

Armistead, K. R., & Cannon, K. A. 2003. *CALIPSO.* World Wide Web. http://www-calipso.larc.nasa.gov/.

Arnold, F., Curtius, J., Spreng, S., & Deshler, T. 1998. Stratospheric aerosol sulfuric acid: First direct in situ measurements using a novel balloon-based mass spectrometer apparatus. *J. Atmos. Chem.*, **30**(1), 3–10.

Artaxo, P., & Maenhaut, W. 1988. Composition and sources of aerosols from the Amazon Basin. *J. Geophys. Res.*, **93**(D2), 1605–1615.

Barnes, J. E., & Hofmann, D. J. 2001. Variability in the stratospheric background aerosol over Mauna Loa Observatory. *Geophys. Res. Lett.*, **28**(15), 2895–2898.

Barthel, H., Sachweh, B., & Ebert, F. 1998. Measurement of airborne mineral fibres using a new differential light scattering device. *Meas. Sci. Technol.*, **9**(2), 210–220.

Baumgardner, D., Strapp, W., & Dye, J. E. 1985. Evaluation of the forward scattering spectrometer probe II : Corrections for coincidence and dead time losses. *J. Atmos. Ocean. Technol.*, **2**(4), 626–632.

Baumgardner, D., Dye, J. E., Gandrud, B., Rogers, D., & Weaver, K. 1995. The multiangle aerosol spectrometer probe. A new instrument for airborne particle research. *In: Proc. 9th Symp. on Met. Observ. & Instr.*

Baumgardner, D., Jonsson, H., Dawson, W., O'Connor, D., & Newton, R. 2001. The cloud, aerosol and precipitation spectrometer: a new instrument for cloud investigations. *Atmos. Res.*, **59**, 251–264.

Bertram, A. K., Dickens, D. B., & Sloan, J. J. 2000. Supercooling of type I polar stratospheric clouds: The freezing of submicron nitric acid aerosols having HNO_3 mole fractions less than 0.5. *J. Geophys. Res.*, **105**(D7), 9283–9290.

Blau, H. H. Jr., McCleese, D. J., & Watson, D. 1970. Scattering by individual transparent spheres. *Appl. Opt.*, **9**(11), 2522–2528.

Bohren, C. F., & Huffman, D. R. 1983. *Absorption and Scattering of Light by Small Particles.* New York: John Wiley & Sons.

Brewer, A. (Lead scientist). 2002. *NOAA Environmental Technology Laboratory lidar homepage.* World Wide Web. http://www.etl.noaa.gov/et2/instruments/.

Carslaw, K. S., Luo, B. P., Clegg, S. L., Peter, Th., Brimblecombe, P., & Crutzen, P. J. 1994. Stratospheric aerosol growth and HNO_3 gas phase depletion from coupled HNO_3 and water uptake by liquid particles. *Geophys. Res. Lett.*, **21**(23), 2479–2482.

Clark, R. J. H., & Hester, R. E. (eds). 1995. *Spectroscopy in environmental science.* Advances in spectroscopy, vol. 24. New York: John Wiley & Son. Chap. Lidar methods and applications (Thomas, L.).

Crutzen, P. J. 1976. The possible importance of CSO for the sulphate layer of the stratosphere. *Geophys. Res. Lett.*, **3**(2), 73–76.

Crutzen, P. J., & 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.

d'Almeida, G. A. 1986. A model for Saharan dust transport. *J. Clim. Appl. Metoer.*, **25**(7), 903–916.

d'Almeida, G. A., & Schütz, L. 1983. Number, mass and volume distributions of mineral aerosol and soils of the Sahare. *J. Clim. Appl. Metoer.*, **22**(2), 233–243.

Deshler, T., Hofmann, D. J., Johnson, B. J., & Rozier, W. R. 1992a. Balloonborne measurements of the Pinatubo aerosol size distribution and volatility at Laramie, Wyoming during the summer of 1991. *Geophys. Res. Lett.*, **19**(2), 199–202.

Deshler, T., Adriani, A., Gobbi, G. P., Hofmann, D. J., Di Donfrancesco, G., & Johnson, B. J. 1992b. Volcanic aerosol and ozone depletion within the Antarctic Polar Vortex during the austral spring of 1991. *Geophys. Res. Lett.*, **19**(18), 1819–1822.

Deshler, T., Johnson, B. J., & Rozier, W. R. 1993. Balloonbourne measurements of Pinatubo aerosol during 1991 and 1992 at 41°N: Vertical profiles, size distribution and volatility. *Geophys. Res. Lett.*, **20**(14), 1435–1438.

Deshler, T., Nardi, B., Adriani, A., Cairo, F., Hansen, G., Fierli, F., Hauchecorne, A., & Pulvirenti, L. 2000. Determining the index of refraction of polar stratospheric clouds above Andoya (69°N) by combining size-resolved concentration and optical scattering measurements. *J. Geophys. Res.*, **105**(D3), 3943–3953.

Deshler, T., Hervig, M. E., Kröger, C., Hofmann, D. J., Rosen, J. M., & Liley, J. B. 2003. Thirty years of in situ stratospheric aerosol size distribution measurements from Laramie, Wyoming (41°N), using balloonborne instruments. *J. Geophys. Res.*, **108**(D5), 4167. doi: 10.1029/2002JD002514.

Dick, W. D. 1998. *Multiangle light scattering techniques for measuring shape and refractive index of submicron atmospheric particles.* Ph.D. thesis, The University of Minnesota.

DLR, (German Aerospace Center). 2002. *The German Remote Sensing Data Center, ATMOS User Center.* World Wide Web. http://auc.dfd.dlr.de/.

Dye, J. E., & Baumgardner, D. 1984. Evaluation of the forward scattering spectrometer probe I: Electronic and optical studies. *J. Atmos. Ocean. Technol.*, **1**(4), 329–344.

Fairall, C. W., Davidson, K. L., & Schacher, G. E. 1983. An analysis of the surface production of sea-salt aerosols. *Tellus*, **35B**(1), 31–39.

Farman, J. C., Gardiner, B. G., & Shanklin, J. D. 1985. Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction. *Nature*, **315**, 207–210.

Fiocco, G., & Grams, G. 1964. Observations of the aerosol layer at 20 km by optical radar. *J. Atmos. Sci.*, **21**(3), 323–324.

Fischer, H., & Oelhaf, H. 1996. Remote sensing of vertical profiles of atmospheric trace constituents with MIPAS limb-emission spectrometers. *Appl. Opt.*, **35**(16), 2787–2796.

Fletcher, R. 1971. *A modified Marquardt subroutine for nonlinear least squares fitting.* Tech. rept. R6799. A.E.R.E., Harwell, England.

Fluent Inc. 2000. *Fluent Manual.* Fluent Inc.

Gayet, J. F., Crépel, O., Fournol, J. F., & Oshchepkov, S. 1997a. A new airborne polar Nephelometer for the measurements of optical and microphysical cloud properties. Part I: Theoretical design. *Ann. Geophysicae*, **15**(4), 451–459.

Gayet, J. F., Crépel, O., Fournol, J. F., & Oshchepkov, S. 1997b. A new airborne polar Nephelometer for the measurements of optical and microphysical cloud properties. Part II: Performance in natural cloud conditions. *Ann. Geophysicae*, **15**(4), 460–470.

Gayet, J. F., Auriol, F., Oshchepkov, S., Schröder, F., Duroure, C., Febvre, G., Fournol, J. F., Crépel, O., Personne, P., & Daugereon, D. 1998. In situ measurements of the scattering phase functions of stratocumulus, contrails and cirrus. *Geophys. Res. Lett.*, **25**(7), 971–974.

Gleason, J. F., Bhartia, P. K., Herman, J. R., McPeters, R., Newman, P., Stolarski, R. S., Flynn, L., Labow, G., Larko, D., Seftor, C., Wellemeyer, C., Komhyr, W. D., Miller, A. J., & Planet, W. 1993. Record low global ozone in 1992. *Science*, **260**, 523–526.

Goodman, J., Verma, S., Pueschel, R. F., Hamill, P., Ferry, G. V., & Webster, D. 1997. New evidence of size and composition of polar stratospheric cloud particles. *Geophys. Res. Lett.*, **24**(5), 615–618.

Grainger, R. G. 1990. *The calculation of cloud parameters from AVHRR data.* Ph.D. thesis, University of Auckland.

Grainger, R. G., Lambert, A., Rodgers, C. D., & Taylor, F. W. 1998. Towards a reference stratospheric aerosol loading. *Adv. Space Res.*, **21**(10), 1421–1424.

Gunson, M. R., Abbas, M. M., Abrams, M. C., Allen, M., Brown, L. R., Brown, T. L., Chang, A. Y., Goldman, A., Irion, F. W., Lowes, L. L., Mahieu, E., Manney, G. L., Michelsen, H. A., Newchurch, M. J., Rinsland, C. P., Salawitch, R. J., Stiller, G. P., Toon, G. C., Yung, Y. L., & Zander, R. 1996. The Atmospheric Trace Molecule Spectroscopy (ATMOS) experiment: Deployment on the ATLAS Space Shuttle missions. *Geophys. Res. Lett.*, **23**(17), 2333–2336.

Hayashida, S., & Horikawa, M. 2001. Anti-correlation between stratospheric aerosol extinction and the Ångström parameter from multiple wavelength measurements with SAGE II – a characteristic of the decay period following major volcanic eruptions. *Geophys. Res. Lett.*, **28**(21), 4063–4066.

Herman, J. R., & Larko, D. 1994. Low ozone amounts during 1992–1993 from Nimbus 7 and Meteor 3 total ozone mapping spectrometers. *J. Geophys. Res.*, **99**(D2), 3483–3496.

Heymsfield, A. J., Lawson, R. P., & Sachse, G. W. 1998a. Growth of ice crystals in a precipitating contrail. *Geophys. Res. Lett.*, **25**(9), 1335–1338.

Heymsfield, A. J., Aulenbach, S. M., & Jensen, T. L. 1998b. Shapes, sizes and light scattering of ice crystals in cirrus and a persistent contrail during SUCCESS. *Geophys. Res. Lett.*, **25**(9), 1331–1334.

Hobbs, P. V. (ed). 1993. *Aerosol-cloud-climate interactions.* International Geophysics Series, Volume 54. San Diego: Academic Press.

Hofmann, D. J. 1990. Increase in the stratospheric background sulphuric acid aerosol mass in the past 10 years. *Science*, **248**, 996–1000.

Hofmann, D. J. 1993. Twenty years of balloon-borne tropospheric aerosol measurements at Laramie, Wyoming. *J. Geophys. Res.*, **98**(D7), 12753–12766.

Hofmann, D. J., & Deshler, T. 1991. Stratospheric cloud observation during formation of the Antarctic ozone hole in 1989. *J. Geophys. Res.*, **96**(D2), 2897–2912.

Hofmann, D. J., & Rosen, J. M. 1977. Balloon observations of the time development of the stratospheric aerosol event of 1974-1975. *J. Geophys. Res.*, **82**(9), 1435–1440.

Hofmann, D. J., & Rosen, J. M. 1981. On the background stratospheric aerosol layer. *J. Atmos. Sci.*, **38**(1), 168–181.

Hofmann, D. J., & Solomon, S. 1989. Ozone destruction through heterogeneous chemistry following the eruption of El Chichón. *J. Geophys. Res.*, **94**(D4), 5029–5041.

Hofmann, D. J., Rosen, J. M., Pepin, T. J., & Pinnick, R. G. 1975. Stratospheric aerosol measurements I: Time variations at northern latitudes. *J. Atmos. Sci.*, **32**(7), 1446–1456.

Hofmann, D. J., Harder, J. W., Rolf, S. R., & Rosen, J. M. 1987. Balloon-borne observations of the development and vertical structure of the Antarctic ozone hole in 1986. *Nature*, **326**(59–62).

Hofmann, D. J., Rosen, J. M., Harder, J. W., & Hereford, J. V. 1989. Balloon-borne measurements of aerosol, condensation nuclei, and cloud particles in the stratosphere at McMurdo Station, Antarctica during the spring of 1987. *J. Geophys. Res.*, **94**(D9), 11253–11269.

Holz, R. E. 2002. *Measurements of cirrus backscatter phase functions using a high spectral resolution lidar.* M.Phil. thesis, University of Wisconsin Madison. Available on the world wide web: http://lidar.ssec.wisc.edu/papers/papers.htm.

Höpfner, M., Oelhaf, H., Wetzel, G., Friedl-Vallon, F., Kleinert, A., Lengel, A., Maucher, G., Nordmeyer, H., Glatthor, N., Stiller, G., Clarmann, T. v., Fischer, H., Kröger, C., & Deshler, T. 2002. Evidence of scattering of tropospheric radiation by PSCs in mid-IR limb emission spectra: MIPAS-B observations and KOPRA simulations. *Geophys. Res. Lett.*, **29**(8), 119 1–4.

Hornak, J. P. (ed). 2002. *The Encyclopedia of Imaging Science and Technology.* New York: John Wiley & Son. Chap. LIDAR (Argall, P. S. and Sica, R. J.).

Jäger, H., Deshler, T., & Hofmann, D. J. 1995. Midlatitude lidar backscatter conversions based on balloonborne aerosol measurements. *Geophys. Res. Lett.*, **22**(13), 1729–1732.

Jarisch, M., Offermann, D., Riese, M., & Wuebbels, D. J. 1997. Measurements of stratospheric trace gases by a balloon borne infrared spectrometer in France. *J. Atmos. Solar-Terr. Phys.*, **59**, 1747–1755.

Johnston, P. V., McKenzie, R. L., Keys, J. G., & Matthews, W. A. 1992. Observation of depleted stratospheric NO_2 following the Pinatubo volcanic eruption. *Geophys. Res. Lett.*, **19**(2), 211–213.

Jonsson, H. H., Wilson, J. C., Brock, C. A., Knollenberg, R. G., Newton, R., Dye, J. E., Baumgardner, D., Borrmann, S., Ferry, G. V., Pueschel, R., Woods, D. C., & Pitts, M. C. 1995. Performance of a focused cavity aerosol spectrometer for measurements in the stratosphere of particle size in the $0.06 - 2.0 - \mu$m-diameter range. *J. Atmos. Ocean Tech.*, **12**(1), 115–129.

Junge, C. E., & Manson, J. E. 1961. Stratospheric aerosol studies. *J. Geophys. Res.*, **66**(7), 2163–2182.

Junge, C. E., Chagon, C. W., & Manson, J. E. 1961. Stratospheric aerosols. *J. Meterol.*, **18**, 81–108.

Kawa, S. R., Fahey, D. W., Wilson, J. C., Schoeberl, M. R., Douglass, A. R., Stolarski, R. S., Woodbridge, E. L., Jonsson, H., Lait, L. R., Newman, P. A., Proffitt, M. H., Anderson, D. E., Loewenstein, M., Chan, K. R., Webster, C. R., May, R. D., & Kelly, K. K. 1993. Interpretation of NO_x/NO_y observations from AASE-II using a model of chemistry along trajectories. *Geophys. Res. Lett.*, **20**(22), 2507–2510.

Kaye, P., Hirst, E., & Whang-Thomas, Z. 1997. Neural-network-based spatial light-scattering instrument for hazardous airborne fiber detection. *Appl. Opt.*, **36**(24), 6149–6156.

Kerker, M. 1997. Light scattering instrumentation for aerosol studies: an historical overview. *Aerosol Sci. Technol.*, **27**(4), 522–540.

Knutson, E. O., & Whitby, K. T. 1975. Aerosol classification by electric mobility: apparatus, theory, and applications. *J. Aerosol Sci.*, **6**, 443–451.

Koutsemogii, P. 1992. *Measurements of remote continental aerosol in Siberia.* Ph.D. thesis, University of Mainz.

Lacis, A., Hansen, J., & Sato, M. 1992. Climate forcing by stratospheric aerosols. *Geophys. Res. Lett.*, **19**(15), 1607–1610.

Larsen, N., Mikkelsen, I. S., Knudsen, B. M., Schreiner, J., Voigt, C., Mauersberger, K., Rosen, J. M., & Kjome, N. T. 2000. Comparison of chemical and optical in situ measurements of polar stratospheric cloud particles. *J. Geophys. Res.*, **105**(D1), 1491–1502.

Lawson, R. P., Heymsfield, A. J., Aulenbach, S. M., & Jensen, T. L. 1998. Shapes, sizes and light scattering properties of ice crystals in cirrus and a persistent contrail during SUCCESS. *Geophys. Res. Lett.*, **25**(9), 1331–1334.

Levenberg, K. 1944. A method for the solution of certain nonlinear problems in least squares. *Quart. Appl. Math.*, **2**(2), 164–168.

Liley, B. J., Rosen, J. M., Kjome, N. T., Jones, N. B., & Rinsland, C. P. 2001. Springtime enhancement of upper tropospheric aerosol at 45° S. *Geophys. Res. Lett.*, **28**(8), 1495–1498.

Liu, B. Y. H., & Pui, D. Y. H. 1974. A submicron aerosol standard and the primary, absolute calibration of the condensation nuclei counter. *J. Colloid Interface Sci.*, **47**(1), 155–171.

Liu, B. Y. H., S.-H., Yo, & Chae, S.-K. 1995. Lower detection limit of aerosol particle counters. *J. IES*, **38**(4), 31–37.

Ludlow, I. K., & Everitt, J. 2000. Inverse Mie problem. *J. Opt. Soc. Am. A*, **17**(12), 2229–2235.

Maier, W. L. 1991. A fast pseudo random number generator: r250 for "better" random numbers. *Dr. Dobb's J.*, **16**(5), 152–157.

Massie, S. T., Deshler, T., Thomas, G. E., Mergenthaler, J. L., & Russell, J. M. III. 1996. Evolution of the infrared properties of the Mount Pinatubo aerosol cloud over Laramie, Wyoming. *J. Geophys. Res.*, **101**(D17), 23,007–23,019.

McCormick, M. P., & Veiga, R. E. 1992. SAGE II measurements of early Pinatubo aerosols. *Geophys. Res. Lett.*, 1295–1298.

McCormick, M. P., Swissler, T. J., Chu, W. P., & Fuller, W. H. Jr. 1979. Satellite studies of the stratospheric aerosol. *Bull. Am. Meteorol. Soc.*, **60**, 1038–1046.

McCormick, M. P., Chu, W. P., Grams, G. W., Hamill, P., Herman, B. M., McMaster, L. R., Pepin, T. J., Russell, P. B., Steele, H. M., & Swissler, T. J. 1981a. High-latitude stratospheric aerosols measured by the SAM II satellite system in 1978 and 1979. *Science*, **214**, 328–331.

McCormick, M. P., Chu, W. P., McMaster, L. R., Grams, G. W., Herman, B. M., Pepin, T. J., Russell, P. B., & Swissler, T. J. 1981b. SAM II aerosol profile measurements, Poker Flat, Alaska; July 16-19 1979. *Geophys. Res. Lett.*, **8**(1), 3–4.

McCormick, M. P., Kent, G. S., Yue, G. K., & Cunnold, D. M. 1982. Stratospheric aerosol effects from the Soufriè volcano as measured by the SAGE satellite system. *Science*, **215**, 3–4.

McCormick, M. P., Thomason, L. W., & Trepte, C. R. 1995. Atmospheric effects of the Mt Pinatubo eruption. *Nature*, **373**, 399–404.

McElroy, M. B., Salawitch, R. J., & Wofsy, S. C. 1986a. Antarctic O_3: Chemical mechanisms for the spring decrease. *Geophys. Res. Lett.*, **13**(12), 1296–1299.

McElroy, M. B., Salawitch, R. J., Wofsy, S. C., & Logan, J. A. 1986b. Reductions of Antarctic ozone due to synergistic interactions of chlorine and bromine. *Nature*, **321**(6072), 759–762.

Mie, G. 1908. Beiträge zur optik trüber medien speziell kolloidaler metallösungen. *Ann. Phys.*, **25**(3), 377–442.

Molina, L. T., & Molina, M. J. 1987. Production of Cl_2O_2 from the self-reaction of the ClO radical. *J. Phys. Chem.*, **91**(2), 433–436.

Molina, M. J. 1991. Heterogeneous chemistry on polar stratospheric clouds. *Atmos. Environ.*, **25A**(11), 2535–2537.

Molina, M. J., Zhang, R., Wooldridge, P. J., McMahon, J. R., Kim, J. E., Chang, H. Y., & Beyer, K. D. 1993. Physical chemistry of the $H_2SO_4/HNO_3/H_2O$ system: Implications for polar stratospheric clouds. *Science*, **261**, 1418–1423.

Mullen, J. 2002. *Aerosol Research Group, Department of Engineering, University of Denver*. World Wide Web. http://www.engr.du.edu/aerosol/.

NASA. 2000. *SOLVE home page.* World Wide Web. http://cloud1.arc.nasa.gov/solve/index.html.

NASA. 2002. *SOLVE II home.* World Wide Web. http://cloud1.arc.nasa.gov/solveII/index.html.

Naudet, J. P., Robert, C., & Huguenin, D. 1994. Balloon measurements of stratospheric trace species using a multichannel UV-visible spectrometer. *In: Proc. 14th ESA symposium.*

Offermann, D., Grossmann, K.-U., Barthol, P., Knieling, P., Riese, M., & Trant, R. 1999. Cryogenic Infrared Spectrometers and Telescopes for the Atmosphere (CRISTA) experiment and middle atmosphere variability. *J. Geophys. Res.*, **104**(D13), 16,311–16,325.

Oshchepkov, S., Isaka, H., Gayet, J. F., Sinyuk, A., Auriol, F., & Havemann, S. 2000. Microphysical properties of mixed-phase & ice clouds retrieved from *in situ* airborne "Polar Nephelometer" measurements. *Geophys. Res. Lett.*, **27**(2), 209–212.

Philbrick, C. R., & Mulik, K. R. 2000. Application of Raman lidar to aor quality measurements. *Pages 22–23 of: Proceedings of the SPIE conference on laser radar technology and applications V*, vol. 4035.

Phillips, D. T., & Wyatt, P. J. 1972. Single-particle light-scattering measurement: Photochemical aerosols and atmospheric particulates. *Appl. Opt.*, **11**(9), 2082–2087.

Pinnick, R. G., & Hofmann, D. J. 1973. Efficiency of light-scattering aerosol particles counters. *Appl. Opt.*, **12**, 2593–2597.

Pinnick, R. G., Rosen, J. M., & Hofmann, D. J. 1973. Measured light-scattering properties of individual aerosol particles compared to Mie scattering theory. *Appl. Opt.*, **12**, 37–41.

Pluchino, A. 1987. Scattering photometer for measuring single ice crystals and evaporation and condensation rates of liquid droplets. *J. Opt. Soc. Am.*, **4**(3), 614–620.

Pluchino, A. B., Goldberg, S. S.and Dowling, J. M., & Randall, C. M. 1980. Refractive-index measurements of single micron-sized carbon particles. *Appl. Opt.*, **19**(19), 3370–3372.

Poole, L. R., & McCormick, M. P. 1988. Airborne lidar observations of arctic polar stratospheric clouds: indications of two distinct growth stages. *Geophys. Res. Lett.*, **15**(1), 21–23.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992. *Numerical recipes in fortran 77, 2nd edition*. Cambridge University Press.

Rodgers, C. D. 2000. *Inverse methods for atmospheric sounding: Theory and Practice*. Singapore: World Scientific.

Rosen, J. M. 1964. The vertical distribution of dust to 30 km. *J. Geophys. Res.*, **69**(21), 4673–4676.

Rosen, J. M. 1967. Stratospheric dust. *Univ. Minnesota Atmospheric Phys. Annual Progr. Rept.*

Rosen, J. M. 1968. Simultaneous dust and ozone soundings over North and Central America. *J. Geophys. Res.*, **73**(2), 479–486.

Rosen, J. M. 1971. The boiling point of stratospheric aerosols. *J. Appl. Meteorol.*, **10**(5), 1044–1046.

Rosen, J. M., & Hofmann, D. J. 1977. Balloonborne measurements of condensation nuclei. *J. Appl. Meteor.*, **16**, 56–62.

Rosen, J. M., & Hofmann, D. J. 1980. A stratospheric aerosol increase. *Geophys. Res. Lett.*, **7**(9), 669–672.

Rosen, J. M, & Kjome, N. T. 1991. Backscattersonde: A new instrument for atmospheric aerosol research. *Appl. Optics*, **30**(12), 1552–1561.

Sassen, K. 1992. Evidence for liquid-phase cirrus cloud formation from volcanic aerosols: climate implications. *Science*, **257**, 516–519.

Schnell, R. C., & Tan-Schnell, S. N. 1982. Kenyan tea litter: A source of ice nuclei. *Tellus*, **34**(1), 92–95.

Schreiner, J., Voigt, Chr., Kohlmann, A, Arnold, F., & Mauersberger, K. 1999. Chemical analysis of polar stratospheric cloud particles. *Science*, **283**, 968–970.

Secker, D. R., Kaye, P. H., Greenaway, R. S., Hirst, E., Bartley, D. L., & Videen, G. 2000. Light scattering from deformed droplets and droplets with inclusions. *Appl. Opt.*, **39**(27), 5023–5039.

Seiler, W., & Crutzen, P. J. 1980. Estimates of gross and net fluxes of carbon between the biosphere and the atmosphere from biomass burning. *Climate Change*, **2**(3), 207–247.

Seinfeld, J. H., & Pandis, S. N. 1998. *Atmospheric chemistry and physics: From air pollution to climate change*. New York: John Wiley & Sons.

Sen, B., Toon, G. C., Osterman, G. B., Blavier, J. F., Margitan, J. J., Salawitch, R. J., & Yue, G. K. 1998. Measurements of reactive nitrogen in the stratosphere. *J. Geophys. Res.*, **103**(D3), 3571–3585.

Shannon, C. E., & Weaver, W. 1949. *The mathematical theory of communication*. Urbana: University of Illinois Press.

Shih, T. H., Liou, A., Shabbir, A., & Zhu, J. 1995. A new $\kappa - \epsilon$ eddy-viscosity model for high Reynolds number turbulent flows – Model development and validation. *Computers Fluids*, **24**(3), 227–238.

Solomon, S., Garcia, R. R., Rowland, F. S., & Wuebbles, D. J. 1986. On the depletion of Antarctic ozone. *Nature*, **321**(6072), 755–759.

Spurny, K. R. (ed). 2000. *Aerosol chemical processes in the environment*. Boca Raton, Florida: CRC Press LLC.

Steiner, B., Berge, B., Gausmann, R., Rohmann, J., & Ruehl, E. 1999. Fast *in situ* sizing technique for single levitated liquid aerosols. *Appl. Opt.*, **38**(9), 1523–1529.

Stowasser, M., Oelhaf, H., Wetzel, G., Friedl-Vallon, F., Maucher, G., Seefeldner, M., Trieschmann, O., Clarmann, T. v., & Fischer, H. 1999. Simultaneous measurements of HDO, H_20 and CH_4 with MIPAS-B: Hydrogen budget and indication of dehydration inside the polar vortex. *J. Geophys. Res.*, **104**(D16), 19,213–19,225.

Ström, J., & Ohlsson, S. 1998. In situ measurements of enhanced crystal number densities in cirrus clouds caused by aircraft exhaust. *J. Geophys. Res.*, **103**(D10), 11355–11361.

Sugita, T., Kondo, Y., Koike, M., Kanada, M., Toriyama, N., Nakajima, H., Deshler, T., & Imasu, R. 1999. Balloon-borne optical counter for in situ aerosol measurement. *J. Atmos. Chem.*, **32**(1), 183–204.

Tabazadeh, A., Turco, R. P., Drdla, K., & Jacobson, M. Z. 1994. A study of Type I polar stratospheric cloud formation. *Geophys. Res. Lett.*, **21**(15), 1619–1622.

Thomason, L. W., Poole, L. R., & Deshler, T. 1997. A global climatology of stratospheric aerosol surface area density deduced from Stratospheric Aerosol and Gas Experiment II measurements: 1984–1994. *J. Geophys. Res.*, **102**(D7), 8967–8976.

Toon, G. C. 1991. The JPL MkIV interferometer. *Opt. Photonics News*, **2**, 19–21.

Toon, G. C., Blavier, J.-F., Sen, B., Margitan, J. J., Webster, C. R., May, R. D., Fahey, D., Gao, R., Del Negro, L., Proffitt, M., Elkins, J., Romashkin, P. A., Hurst, D. F., Oltmans, S., Atlas, E., Schauffer, S., Flocke, F., Bui, T. P., Stimpfle, R. M., Boone, G. P., & Voss, P. B. 1999. Comparison of MkIV ballonn and ER-2 aircraft measurements of atmospheric trace gases. *J. Geophys. Res.*, **104**(D21), 26,779–26,790.

Toon, O. B., Hamill, P., Turco, R. P., & Pinto, J. 1986. Condensation of HNO_3 and HCl in the winter polar stratosphere. *Geophys. Res. Lett.*, **13**(12), 1284–1287.

Turco, R. P., Whitten, R. C., & Toon, O. B. 1982. Stratospheric aerosols: Observation and theory. *Rev. of Geophys. Space Phys.*, **20**, 233–279.

Twomey, S. A., Piepgrass, M., & Wolfe, T. L. 1984. An assessment of the impact of pollution on global cloud albedo. *Tellus*, **36B**(5), 356–366.

Weast, R. C. (ed). 1988. *CRC handbook of chemistry and physics.* Vol. 1st Student ed. CRC Press.

Webster, C. R., May, R. D., Toohey, D. W., Avallone, L. M., Anderson, J. G., Newman, P., Lait, L., Schoeberl, M. R., Elkins, J. W., & Chan, K. R. 1993. Chlorine Chemistry on Polar Stratospheric Cloud Particles in the Artic Winter. *Science*, **261**(5125), 1130–1134.

Webster, C. R., May, R. D., Allen, M., Jaeglé, L., & McCormick, M. P. 1994. Balloon profiles of stratospheric NO_2 and HNO_3 for testing the heterogeneous of N_2O_5 on sulphate aerosols. *Geophys. Res. Lett.*, **21**(1), 53–56.

Willeke, K., & Baron, P. A. 1993. *Aerosol Measurement: Principles, Techniques, and Applications.* New York: Van Nostrand Reinhold.

Winker, D. M., & Powell, K. A. 1998. *LITE Home page*. World Wide Web. http://www-lite.larc.nasa.gov/.

Wiscombe, W. J. 1977. The delta-M method: Rapid yet accurate radiative flux calculations for strongly asymmetric phase functions. *J. Atmos. Sci.*, **34**(9), 1408–1422.

Wulfmeyer, V., Brewer, W.A., Senff, C., Mayor, S., Marchbanks, R., Howell, J., Weickmann, A., Richter, R., Grund, C., & Hardesty, R. M. 1998. Performance and applications of the NOAA 2um High Resolution Doppler Lidar. *Pages 573–576 of: 19th International Laser Radar Conference*.

Wyatt, P. J., Schehrer, K. L., Phillips, S. D., Jackson, C., Chang, Y., Parker, R. G., Phillips, D. T., & Bottiger, J. R. 1988. Aerosol particle analyzer. *Appl. Opt.*, **27**(2), 217–221.

Yue, G. K., & Deepak, A. 1981. Modeling of growth and evaporation effects on the extinction of 1.0 μm solar radiation traversing stratospheric sulfuric acid aerosols. *Appl. Opt.*, **20**(20), 3669–3675.

Yue, G. K., Poole, L. R., Wang, P. H., & Chiou, E. W. 1994. Stratospheric aerosol acidity, density, and refractive index deduced from SAGE II and NMC temperature data. *J. Geophys. Res.*, **99**(D2), 3727–3738.