Understanding Uncertainties for Polar Mesospheric Cloud Retrievals and Initial Gravity Wave Observations in the Stratopause from the Cloud Imaging and Particle Size Instrument

Justin Neal Carstens

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Electrical Engineering

Scott M. Bailey
Gary S. Brown
Gregory D. Earle
Troy A. Henderson
Wayne A. Scales

October 25th, 2012
Blacksburg, Virginia

Keywords: polar mesospheric clouds, gravity waves, CIPS, AIM, retrieval uncertainties

Understanding Uncertainties for Polar Mesospheric Cloud Retrievals and Initial Gravity Wave Observations in the Stratopause from the Cloud Imaging and Particle Size Instrument

Justin N. Carstens

ABSTRACT

The Cloud Imaging and Particle Size (CIPS) instrument on the Aeronomy of Ice in the Mesosphere satellite images in the nadir at the UV wavelength of 265 nm. The camera array has an approximately 120° along track (2000 km) by 80° cross track (1000 km) field of view at a horizontal resolution of 1 by 2 km in the nadir. The satellite is in a sun synchronous orbit with an approximately noon local time equator crossing. The observed albedo is due to Rayleigh scattered sun light from an altitude of approximately 50 km and sunlight scattered from Polar Mesospheric Clouds (PMC) which occur in the summer mesosphere at 83 km.

The goal of the CIPS instrument is to retrieve high horizontal resolution maps of PMC albedo and the mode radius of the particle size distribution. The first manuscript analyzes the uncertainties involved in the retrieval. The ability to infer mode radius from the PMC signal is made significantly harder by the presence of the Rayleigh signal. Much of the difference between PMC signals of different mode radii is also consistent with possible changes in the Rayleigh signal. The signal is decomposed into components which isolate the portion of the PMC signal's dependence on radius which is not consistent with changes in the Rayleigh signal. This isolated component is compared with the measurement noise to estimate and understand the uncertainties in the CIPS retrieval.

The presence of the Rayleigh signal is a difficulty in the PMC retrieval, but it is also a valuable data product. The second manuscript highlights the initial findings of a new gravity wave data set developed by the author. The data set provides relative ozone variations at the stratopause with a horizontal resolution of 20 by 20 km. An abundance of gravity wave signatures can be seen in the data which appear to emanate from weather events like thunderstorms and hurricanes as well as orographic sources such as the Andes and the Antarctic Peninsula. The data set fills a gap that presently exists in our observational coverage of gravity waves, so the data set should help significantly in constraining Global Climate Models.

Acknowledgements

Thanks to my family for providing me with the love and support that has motivated me to work through any obstacles that have come in my way. Thanks to my wife Padma who put up with me during the stressful times and for loving me all of the time. Thanks to my mom who was always there to listen to support me, but who was hard on me when I was letting things slip. Thanks to my daughter Emma for whom I'm willing to go through anything to provide a better life and be a positive example. Thanks to my family for always being proud of my accomplishments.

Thanks to Scott Bailey my research advisor going back to my junior year as an undergraduate, for providing me with countless opportunities. As an undergraduate he allowed me the opportunity to prove myself and make meaningful contributions that I doubt many undergraduates received the opportunity to do. His approach of leaving me to develop my own ideas while at the same time remaining an encouragement and promoting my ideas to the team (when they were good ideas) was ideal for me. I consider him to be not only the best adviser I could have asked for but also one of my best friends. Thanks to the other students in my research group Cissi, Karthik, Justin and Padma for making work a fun place to be.

Thanks to the entire AIM team for developing an excellent satellite and finding ways to make it work even when it didn't want to listen to us. Thanks to the CIPS team for creating a unique and intriguing instrument that provided me with countless hours of pondering what information it provided. Thanks to the POLAR NOX and NSROC teams for participating in one of the most rewarding experiences of my life, and helping to really cement my confidence in my

choice of field. And of course, thanks to NASA for funding most of the projects I have worked on as both a Graduate and as an Undergraduate student.

Thanks to all the teachers whom have helped to further my education over my now 26 years as a student. Thanks to Virginia Tech for providing an excellent graduate student experience. Thanks to Cindy Hopkins for being so helpful in getting me through the process for obtaining my degree under rather unusual circumstances, and thanks to Karen DePauw for making allowances for my time constraints.

Table of Contents

List of Figures

1. Introduction

The Cloud Imaging and Particle Size (CIPS) instrument on board the Aeronomy of Ice in the Mesosphere (AIM) satellite is a UV imaging array that was designed to observe Polar Mesospheric Clouds (PMC) at high horizontal resolution allowing for observation of mesoscale structures. The observed signal is a superposition of Rayleigh scattered sunlight emanating from the stratopause (~50km) and sunlight scattered by PMCs at an altitude of 83km. The array is pointed in the nadir with a wide field of view, and the imaging cadence is such that an atmospheric parcel is observed approximately seven times as AIM passes over it. The set of observations for a given parcel called the "scattering profile" is the primary data product from which PMC properties are retrieved.

The upper mesosphere is the coldest region on earth and it is surprisingly coldest in the summer instead of the winter. This reversal is due to gravity wave drag which forces the mean flow, and ultimately leads to a summer to winter circulation consisting of upwelling in the summer causing the region to cool and sinking in the winter which causes the region to heat [*McIntire*; 1989]. In addition to the global impact of gravity waves, they also pose significant local effects as they carry energy and momentum from lower altitudes. Knowledge of the gravity wave spectrum and altitude dependence is crucial to understanding the mesosphere. Without this, Global Climate Models (GCM) cannot reproduce the behavior of the mesosphere. At present these waves are insufficiently constrained by observations [*Lindzen*; 1981, *Fritts and Alexander*; 2003].

Polar mesospheric clouds are the highest altitude clouds in Earth's atmosphere occurring in a thin layer peaking around 83 km just below the mesopause. The mesosphere is only cold

enough for the formation of PMCs, which consist of water ice particles, during the summer near the poles. The fact that they are so near the edge of being able to exist makes PMCs very sensitive to the conditions in the mesosphere, so changes in PMCs provide a strong indication of changes in the mesosphere. Gravity waves perturb the clouds significantly, so the horizontal structure of the clouds provides a way to observe the local impact of gravity. The CIPS instrument allows for observation of the horizontal cloud structure at the resolution necessary to observe gravity wave effects. No other instrument has this capability together with the global coverage.

The scattering profile observed for each observation parcel provides albedo as a function of scattering angle. The scattering phase function of ice particles in the range of sizes expected to be observed is a strong function particle size. Large particle sizes have a large asymmetry where forward scattering is much more efficient than backward scattering. Small particles approach the Rayleigh limit where forward verses backward scattering is symmetric. Prior to this work, it was assumed that the scattering profile should easily allow for retrieval of particle size by finding the best fit of the observed scattering profile to model scattering profiles corresponding to different particle sizes. This is due to the leverage provided by this asymmetry as a function of particle size; however, since the observed scattering profile is a superposition of a Rayleigh scattered signal controlled by ozone and a PMC signal controlled by particle size and density, fitting to the observed signal has the additional degrees of freedom allowed by the changes in the Rayleigh signal due to ozone. This additional freedom reduces the differences between net scattering profiles corresponding to PMCs of different particle size. The impact of this potential loss of leverage had not been adequately analyzed before this work. By decomposing the observed scattering profile in a way that isolates the portion of the scattering profile which is unique to

changes in particle size, the first manuscript (Chapter 2) analyzes in detail the leverage available to retrieve particle size from the scattering profile and the sensitivity to measurement errors. This decomposition reveals that most of the difference between clouds of different particle size is also consistent with a change in the ozone parameters, so much of the leverage that was assumed to be available in reality is not available. Fortunately, the analysis also reveals that there is enough leverage left over to achieve meaningful results, but the problem is much more complex than was previously realized. The analysis here also simplifies the problem greatly for CIPS retrievals such that it is reduced from having to be understood in terms of a four parameter fit to the approximately seven observed albedos in the scattering profile to understanding it in terms of a one parameter fit to two observed components. This simplification yields a much greater ability to predict and understand the sensitivity to errors and limitations and strengths of the measurement. This article is being prepared for publication as part of a Journal of Atmospheric and Solar-Terrestrial Physics special issue regarding layered phenomenon in the mesopause region.

While the Rayleigh scattered background provides a difficulty in the retrieval of PMCs, it can also be used as a useful data product on its own when PMCs are absent. The ozone at the stratopause is perturbed by gravity waves and results in the presence of many wave signatures in the Rayleigh scattering data. The structures seen in the Rayleigh signal are significantly weaker than those seen in the clouds which makes the signal to noise requirements different from the PMC retrieval. Also, since the Rayleigh signal comes from 50 km instead of 83 km like the clouds, the data must be geo-located to 50 km in order to resolve the structures. The default algorithm geo-locates to the 83 km layer in which any Rayleigh structures would be smeared out due to the parallax in the scattering profiles. For these two reasons, retrieval of the Rayleigh

structures required an independent algorithm. The stratopause at 50 km is not currently covered by any instrument which observes the horizontal structure of waves and which observes them at the global scale provided by a satellite, so the addition of this data set to the community promises to aid significantly in the observational constraints of the gravity wave spectrum. The second manuscript included in this dissertation (Chapter 3), outlines a new independent algorithm for retrieving stratopause gravity wave structures from CIPS data and some initial observations this data set has provided. This manuscript is being prepared as an article for Geophysical Review Letters (GRL) because it is expected that this data set will result in a significant advancement in the observational constraints of gravity waves in the middle atmosphere. This journal has a high impact factor, but it is limited to four pages and it should be written for an audience with a broader range of backgrounds. Further development of this data set and an analysis of the impact this data set has on knowledge of gravity wave constraints will be the subject of the author's postdoctoral research.

2. Understanding uncertainties in the retrieval of polar mesospheric clouds from the cloud imaging and particle size instrument

2.0 Abstract

This paper presents a framework for understanding and quantifying the leverage available for inverting the radiance signal in order to retrieve particle size distribution mode radius and albedo from Polar Mesospheric Clouds (PMC) observations using the Cloud Imaging and Particle Size (CIPS) instrument on board the Aeronomy of Ice in the Mesosphere (AIM) satellite. The observed signal is a superposition of the scattering angle dependence of the cloud albedo and the Rayleigh scattered albedo controlled by ozone at the stratopause. The leverage is a quantification of how much the net scattering angle dependence changes as a function of mode radius. The leverage is determined by decomposing the observed signal into orthogonal components which isolate the parts of the signal that are unique to changes in mode radius from those that could be due to changes in either the background or the cloud parameters. This leverage is considered along with instrument noise performance to determine retrieval uncertainties and to understand minimum thresholds in the cloud retrieval parameters.

2.1 Introduction

The Aeronomy of Ice in the Mesosphere (AIM) satellite is the first satellite dedicated to the study of Polar Mesospheric Clouds (PMC) [*Russell et al.*; 2008]. The satellite is in a sun synchronous orbit with an approximately noon local time at the equator. The Cloud Imaging and Particle Size (CIPS) instrument is a 4 camera UV imaging array pointed in the nadir. The array images at 265 nm with a field of view of about 2000 km along track by 1000 km cross track at a nadir resolution of 1 km by 2 km respectively [*McClintock et al.*; 2008].

The goal of CIPS is to provide global albedo and mode radius maps of PMC at high horizontal resolution [*Rusch et al.*; 2008]. Many instruments can measure albedo and mode radius, and some of those measurements have better leverage on that information than CIPS, but no instrument has this information at such high horizontal resolution with near global coverage. With the horizontal maps one can study gravity waves which play an important role in creating the environment for the clouds to exist through gravity wave drag, but they also perturb the clouds locally which allows for their observation [*Chandran et al.*; 2009, *Chandran et al.*; 2010, *Taylor et al.*; 2011]. The maps are also useful for adding information on the horizontal structure to the temporal information in coincident lidar observations [*Baumgarten et al.*; 2012].

Each air parcel in the map is observed several times with different scattering angles as AIM passes over it. This scattering angle dependence for each parcel is called the "scattering profile". The scattering angle dependence of the albedo observed is a result of Rayleigh scattering by the atmosphere and the ice scattering phase function of the PMC. The magnitude and angular dependence of the PMC phase function depends on the albedo and mode radius of the particle size distribution for the cloud observed, so analysis of this scattering angle dependence allows for retrieval of the albedo and mode radius maps. The magnitude of the change in the scattering profile angular dependence as a function of mode radius is quantified using its "leverage" which will be defined in section 4.3. The details of the algorithms used to produce these maps have been published for the operational CIPS retrieval, version 3, in *Bailey et al.* [2008] and for version 4 in *Lumpe et al.* [this issue]. This paper instead takes a step back and attempts to provide an understanding and quantification of the uncertainties in the CIPS measurement. How much leverage on the mode radius of the cloud particles does the measurement really have? Is the mode radius inversion unique? How bright must a cloud be in

order to be retrievable using CIPS measurements, and how does one define its brightness from the standpoint of the ability to retrieve its properties? How do the leverage and minimum brightness change as a function of solar zenith angle, sampling and mode radius? This paper attempts to answer the above questions and to provide a frame work for visualizing and diagnosing the impact of errors on the CIPS retrieval.

In section 2 we start by describing the CIPS scattering profile. In section 3 the scattering profile is transformed using orthonormal basis vectors which clearly separate the portion of the scattering profile that is exclusively due to the cloud and that which is consistent with both clouds and a Rayleigh background. In section 4 the retrieval problem is analyzed using the transformed signal. In this section many of the questions above are answered. Finally in section 5, simulated retrievals are done in the presence of noise, and the results of the simulation are checked for consistency with expectations based on the conclusions drawn in section 4.

2.2 The CIPS Scattering Profile

2.2.1 Models

The CIPS instrument observes air parcels at several different scattering angles as AIM passes overhead. These observations create a scattering profile which includes the scattering angle dependence of cloud ice albedo near 83 km and Rayleigh scattered sunlight from about 50 km. The Rayleigh scattering component of the scattering profile (referred to as the "background") obeys the nearly symmetric Rayleigh scattering phase function. The cloud particles are larger in comparison to the observed wavelength, so their scattering is not symmetric in scattering angle with much stronger efficiency for forward scattering. **Figure 2-1** is

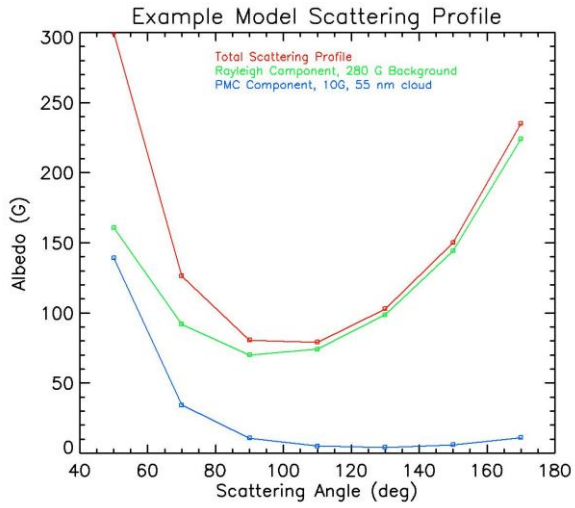

Figure 2-1 A simulated example of a scattering profile. The red signal is what would be observed and the green and blue curves are the Rayleigh and PMC components respectively.

an example modeled scattering profile illustrating the components of a CIPS signal. The albedo is expressed in the unit 10^{-6} sr^{-1} (G) which will be used throughout this paper.

For this analysis, an analytical model which describes the Rayleigh scattering background albedo, $R(\theta, SZA, VWA, C, \sigma)$, is used [*McPeters;* 1980, *Bailey et al.;* 2008]. We will refer to this model as the "*C*-σ model". The model is used in the following form,

$$R = C\ P(\theta)\frac{\{[Chap(SZA) + Chap(VWA)]/2\}^{-\sigma}}{\cos(VWA)} \qquad (1)\ ,$$

where $P(\theta)$ is the rayleigh scattering phase function as a function of the scattering angle θ, $Chap(SZA)$ is the chapman function [*Chapman;* 1931] as a function of the solar zenith angle *SZA*, *VWA* is the angle between the zenith at the observation point and the line of sight (referred to as the viewing angle hereafter) and σ is the ratio of the ozone scale height to the atmospheric scale height. The C parameter is the phase adjusted albedo one would observe if the parcel where to be viewed in the nadir at the sub solar point. It is approximately proportional to the

atmospheric pressure level at which the ozone optical depth reaches 1 [*Bhartia et al.;* 1996]. This form is convenient for fitting since the C parameter is simply a scale factor on the Rayleigh scattering profile. The Chapman function is calculated using the asymptotic expansion method of *Huestis* [2000].

The cloud scattering phase functions are simulated using the T-matrix codes [*Mishchenko*; 1998]. The particle size distributions are assumed to be Gaussian with widths equal to 0.355 times the mode radius, *r*, up to a maximum width of 16 nm. Since in this paper the radius of individual particles is rarely discussed, "radius" will often be used in place of mode radius for brevity. In instances where the mode radius is not what is being referred to, this will be made clear. The axial ratio of the cloud particles is assumed to be two. An additional $1/\cos(VWA)$ factor is added to account for the increased line of sight path length through the cloud with increasing *VWA*. The cloud albedo *A* is defined as the albedo one would observe if viewing the cloud in the nadir with a scattering angle of 90°. These are the same models used for the operational CIPS retrieval.

2.2.2 Radius/Background Ambiguity

When looked at in the absence of a Rayleigh background, the cloud scattering phase functions of different radii look very different. The most notable difference is that the asymmetry where forward scattering is stronger than the backward scattering is larger for larger radii. If the observed signal only consisted of the cloud signal, this would provide a great deal of leverage with which to retrieve radius, but the scattering profile is a superposition of an unknown cloud and background signal, so at least some component of the difference between cloud signals of different radius is likely also consistent with a change in background parameters. The retrieval is done by finding the least squares best fit between the observed and modeled scattering profiles,

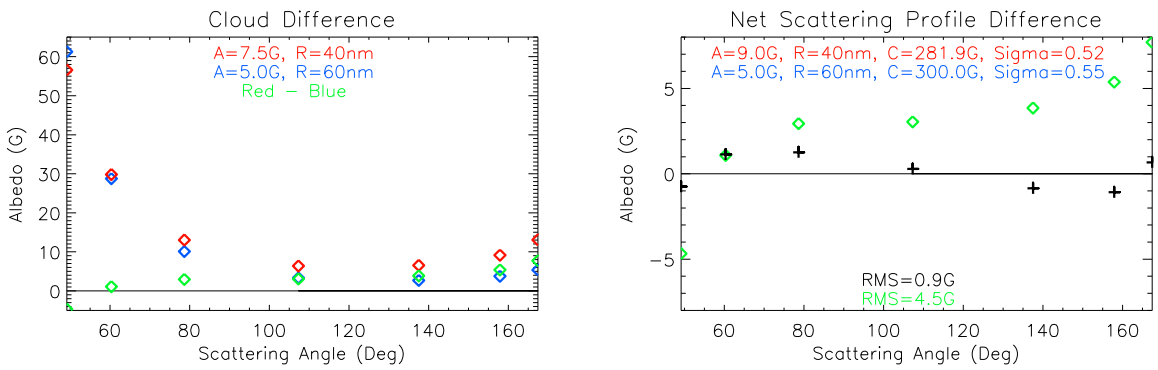

Figure 2-2 Example illustrating the ability to distinguish a 60 nm radius 5 G albedo cloud from a cloud with a 40 nm mode radius. For the cloud component of the scattering profile alone, the left plot shows the 60 nm cloud scattering profile in blue along with the best fit 40 nm cloud scattering profile in red. In green is the difference between those two profiles. The black points in the right plot shows the difference of the best fit 40nm cloud when the entire Rayleigh plus cloud scattering profile is used. The shown in the left plot is shown again in the right plot in green for comparison. The additional degrees of freedom allowed by the Rayleigh background make it much more difficult to distinguish the 60 nm cloud from a 40 nm cloud.

so this component reduces the available leverage with which radius can be retrieved. It will be shown that although this ambiguous component turns out to be quite significant, the residual component that is independent of the Rayleigh background is sufficient to obtain meaningful radius retrievals.

To understand this ambiguity and to motivate the approach that follows it's useful to look at a specific example. Shown in the left plot of **Figure 2-2** are two example cloud scattering profiles without the Rayleigh added. The 40 nm cloud in red has a larger albedo than the 60 nm cloud in blue and it is less asymmetric in a relative sense. The choice of albedo is such that the least squared error between the two clouds is minimized, so if for example a cloud with a 60 nm mode radius and albedo of 5 G is observed, the closest possible 40 nm cloud is the one plotted. The ability to resolve the difference between the two clouds in this case comes down to the ability to resolve the difference signal plotted in green; however, the scattering profile observations include the Rayleigh component as well, so the retrieval has the additional degrees of freedom allowed by the Rayleigh background in order to produce a best fit.

The black points in the right plot of **Figure 2-2** show the difference between the r=60 nm, A=5 G cloud with background parameters C=300 G and σ=0.55 and the least squares best fit scattering profile for a cloud of 40 nm mode radius. The green is the same difference shown in the left plot. By using a brighter cloud albedo relative to the cloud only case and a dimmer but still reasonable Rayleigh albedo relative to the 60 nm cloud, the best fit of the 40 nm cloud to the 60 nm cloud is much closer than in cloud only case. The RMS difference in the cloud only case was 4.5 G and in the case of fitting both a background and a cloud to the scattering profile it is only 0.9 G, so the measurement uncertainties need to be about 5 times smaller in order to distinguish the two clouds when the freedom in the background parameters is added to the fit. With the exception of the end points, the smaller radius cloud favors forward scattering more so than the larger radius cloud. This is the opposite of what would be expected when considering the cloud scattering profiles alone.

The example above has revealed that most of the difference between clouds of different radii is also consistent with a change in the background parameters, and what is left over has a very different scattering angle dependence than what is expected from the cloud signals alone. It will be shown below that the residual difference is sufficient to achieve meaningful radius retrievals, but the problem is more difficult than it initially appears. In order to understand the uncertainties in the retrieval, care must be taken to isolate the component of the signal which contains unambiguous information on the cloud radius from the component which is not unique in its inversion into cloud and Rayleigh parameters. It is not the entire signal which needs to be compared to the errors in order to understand the retrieval uncertainties. It is the component unique to the dependence on radius that needs to be compared with the errors in order to

understand the uncertainties. The goal of the scattering profile decomposition that follows is to isolate this component.

2.3 Scattering Profile Decomposition

2.3.1 Components

A useful tool to aid in the understanding of the retrieval is to decompose the scattering profile into orthogonal components. The components are constructed such that they clearly separate the portion of the scattering profile that is consistent with a C-σ background, from the portion that can only be due to the cloud. To accomplish this, basis vectors are constructed such that the first few are those needed to span most of the variance in the set of reasonable C-σ backgrounds (this will be referred to as the background subspace). Additional basis vectors are added which along with the background basis vectors span most of the variance in set of expected cloud signals (this will be referred to as the cloud subspace). Where "most of the variance" is determined by demanding that the residual variance is not a significant source of error in the retrieval. This criterion determines the number of basis vectors used. Since C and A are simply scale factors on the background and cloud components as a function of σ and r respectively, the sets correspond to expected ranges for σ and r. Since the background basis vectors span the C-σ background set, all of the background signal will project into the background subspace; however, some of the cloud signal will also project into the background subspace, so the magnitude of the scattering profile projection into this subspace will be a sum of the cloud and background projections. The projections of the signal into the cloud subspace will be exclusively determined by the cloud, so if the cloud's mode radius and albedo can be determined from these components alone, we can be sure that measurement is not biased by incorrect C-σ background subtraction.

The basis vectors are generated using the inner product definition below where $\vec{a}$ and $\vec{b}$ are vectors with each element as an albedo for an observation in a scattering profile, and *n* is the number of elements in the profile.

$$\vec{a} \cdot \vec{b} \equiv \frac{\sum_{i=1}^{n} a_i b_i}{n} \quad (2) .$$

This definition results in a length parameter or magnitude of,

$$|a| = \sqrt{\vec{a} \cdot \vec{a}} = \sqrt{\frac{\sum_{i=1}^{n} {a_i}^2}{n}} \quad (3) .$$

This definition approximately preserves length if a scattering profile is observed with a different number of elements. This is convenient since the complicated nature in which CIPS images overlap results in the number of elements in a scattering profile ranging from 1 to 12.

A principle component analysis done for a set of *C*-σ profiles covering the range of reasonable values for σ in a CIPS observation reveals most of the variance in the set is found in the primary component with a relatively small but not insignificant fraction of the variance in the second component. The fraction of the variance in the third component is many orders of magnitude smaller than the second component and insignificant for the purposes of a CIPS retrieval. This means that the dimension of the *C*-σ scattering profile is effectively two, so if any two *C*-σ profiles in the set are taken and used in a Gram Schmidt process to produce orthonormal basis vectors, they must span approximately the same space as the first two principal components. Once the basis vectors are constructed, one can calculate the residual between the two component approximation and **Equation 1** for the *C*-σ set. This residual is shown in the next section to be small enough to make this method for obtaining the basis vectors more than

adequate, but the residual would likely have been somewhat smaller if the principle component basis vectors had been used. Constructing the basis vectors in this way is less computationally expensive than principal component analysis. This is important because the CIPS data volume is so large and basis vectors need to be constructed for every parcel. The resulting basis vectors are also more intuitive since the first or "primary" basis vector is simply a normalized *C*-σ profile for a particular σ. The secondary component can be thought of as the shape of the difference between profiles of different σ. A similar situation exists in the cloud subspace. In this subspace three basis vectors are needed, and again, it is more computationally efficient and intuitive to construct the basis using three cloud scattering profiles of different mode radii than to use principle component analysis.

2.3.2 Background Subspace

This section describes how the background subspace basis vectors are constructed. They are created using values for σ within the range appropriate based on CIPS observations. The range is small enough that only two components are needed to reproduce enough of background signal to not influence the radius retrieval significantly. Background projections onto the second basis vector are very small in comparison to the primary component, but they are still significant compared to cloud signals, so the secondary component must be used.

For a given observation parcel looking at 100 orbits of cloud free CIPS data (approximately one week of out of PMC season data), the standard deviation (abbreviated as SD hereafter to avoid confusion with the scale height ratio σ) about the mean observed σ is approximately 0.01. The mean for each parcel is slightly different and changes slowly seasonally and as a function of *SZA*; however, the mean has little influence on this analysis so long as it is approximately known and the basis vectors are constructed with σ values near it. What is

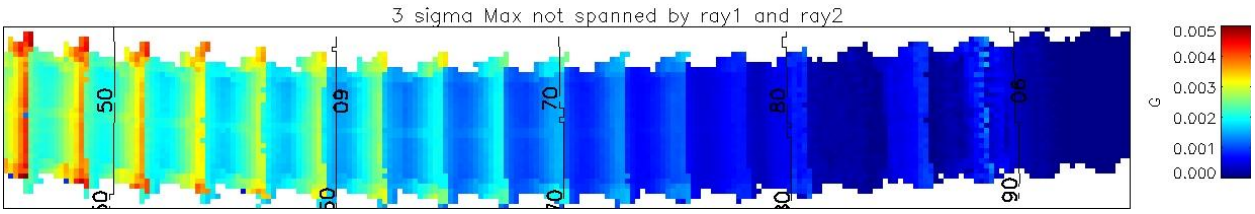

Figure 2-3 The magnitude of the *C-σ* which is not spanned by the two background components over a CIPS orbit. This is the available part of the background signal which could potentially project into the cloud subspace and bias the radius retrieval. The contour lines indicate solar zenith angle.

important is the width of the range in possible values of σ, and that is captured in the SD of σ. If the range is too large, the need for a third basis function to effectively span the *C-σ* set increases. Fortunately, the range needs to be at least an order of magnitude larger than the observed SD before this becomes important, and an approximate parameterization of the mean value for a given parcel can easily be done to the level of accuracy needed to choose basis vectors which span the observed range in σ well enough to not need a third component.

To simplify things this paper constructs the basis vectors assuming that the mean σ is 0.55 in all parcels (approximately the mean in the one week data set used above). In practice, the basis should be more optimally centered on a more appropriate estimate of the mean value for each observed parcel based on time of year and *SZA*. The background basis vectors are constructed using *C-σ* scattering profiles with a σ of 0.52 and 0.58 respectively. These are chosen since they are relatively far apart but close enough (±3 standard deviations) to the mean σ so that we expect to observe them. Placing them far apart makes the vectors as far from parallel as possible so the Gram Schmidt process is more numerically stable. The first basis vector, $\hat{r}_1$, is simply a σ = 0.52 scattering profile calculated using **Equation 1** and normalized using **Equation 3**. The second basis vector, $\hat{r}_2$, is constructed with σ = 0.58 and using the Gram Schmidt process to produce a vector orthonormal to $\hat{r}_1$.

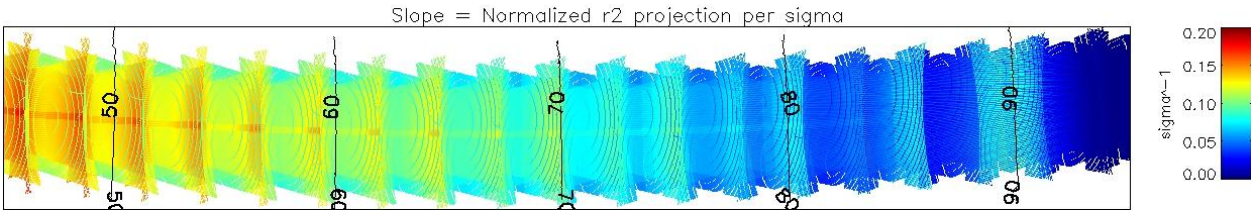

Figure 2-4 An orbit strip map of the slope relating σ to the $\hat{r}_2$ component of the normalized background signal. The slope here can be used with Equation 4 to give approximately the normalized background signal as a function of σ.

The maximum RMS error in the two component approximation to **Equation 1** for a range of σ encompassing three standard deviations about the mean (0.52 to 0.58) is shown in **Figure 2-3.** As can be seen from the figure, this error is less than approximately 0.005 G. A typical total signal magnitude is on the order of 300 G, so this is a small fraction of the background signal, but since some component of this error will project into the cloud subspace, the true test is whether it's small compared to the cloud signal's dependence on mode radius. It will be shown in section 4.3 that it is not a significant error, so it is safe to assume that the $\hat{r}_1$ and $\hat{r}_2$ basis vectors span enough of the *C*-σ background to leave the cloud subspace effectively free of *C*-σ contamination. Additionally, the absolute error is approximately quadratic in σ with zeros at 0.52 and 0.58, so the maximum error over this range actually occurs in the middle or around 0.55. This means that the range could be increased by a factor of $\sqrt{2}$ (0.508 to 0.592) without an increase in the maximum error.

Since the background signal is typically much brighter than the cloud it is necessary to use at least two basis vectors to remove enough of the background signal to render it insignificant to the cloud; however, for all *C*-σ background signals within the set, well over 99% of the signals magnitude projects onto $\hat{r}_1$. The projection onto $\hat{r}_2$ is zero at σ = 0.52 (by construction) and the normalized signal magnitude that projects onto it is approximately linear in

σ. The slope for this linear relationship is shown in **Figure 2-4**. In terms of this slope, m_{r2}, the normalized signal rayleigh signal, $\hat{R}_n$, can be approximated by,

$$\hat{R}_n = \sqrt{1 - ((\sigma - 0.52)m_{r2})^2}\hat{r}_1 + (\sigma - 0.52)m_{r2}\hat{r}_2 \qquad (4)\ .$$

Even for a 3 SD positive deviation (σ = 0.58) from the mean (σ = 0.55) the $\hat{r}_2$ component only has a relative magnitude on the order of 0.005 or approximately 0.0025% of the signal variance. This may seem small, but for solar zenith angles not near to or greater than 90º, it is on the order of 4G which is significant compared to a cloud signal.

2.3.3 Cloud Subspace

The cloud subspace is the space spanned by the set of expected cloud signals with the background components removed. Similar to the background case, a principle component analysis reveals the dimension of the cloud subspace is approximately three, so again we can get basis vectors for approximately the same subspace using the Gram Schmidt process on vectors within the set. As before, the check is to look at the residual between the three component approximation and the original cloud signal. The basis vectors are constructed using phase functions with mode radii of 50, 85 and 25 nm to obtain basis vectors $\hat{c}_1$, $\hat{c}_2$ and $\hat{c}_3$ respectively. The primary component is chosen to be near center of the range of radii expected for PMC observable by CIPS. The second component is chosen on the high end because CIPS is more sensitive to clouds of large mode radii. A look at the residual between the two component approximation and the original cloud signal shows that the error is largest at small particles, so a 25 nm mode radius cloud is chosen as the seed for the third component since anything much smaller than that cannot be reasonably retrieved from CIPS measurements. The third component was the only one for which the radius chosen was at all critical. Clouds of all mode radii have

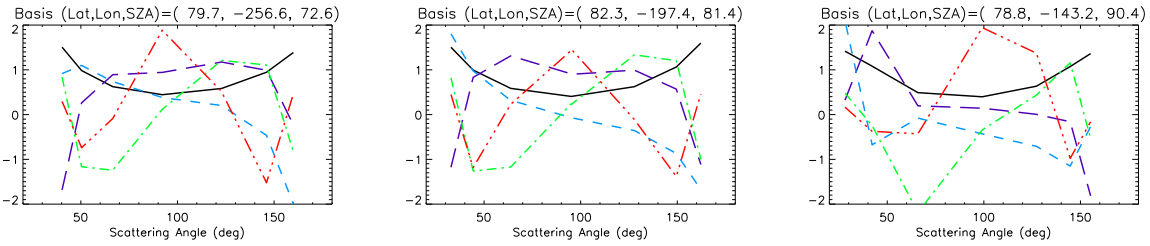

Figure 2-5 These are the derived basis functions for three scattering profiles in the orbit strip. The solid black line is $\hat{\boldsymbol{r}}_{\boldsymbol{1}}$, long dash purple line is $\hat{\boldsymbol{r}}_{\boldsymbol{2}}$, short dashed blue line is $\hat{\boldsymbol{c}}_{\boldsymbol{1}}$, and the green dash-dot and red dash-dot-dot-dot lines are $\hat{\boldsymbol{c}}_{\boldsymbol{2}}$ and $\hat{\boldsymbol{c}}_{\boldsymbol{3}}$ respectively. All these scattering profiles lie at approximately the center of the orbit strip and from left to right are at approximately 70°, 80° and 90° solar zenith angle.

significant projections onto the first two basis vectors, but projections onto the third basis vector are only significant for clouds with small radii, so at least one of the radii used should be small in the construction of the basis. The $\hat{c}_1$, $\hat{c}_2$ and $\hat{c}_3$ basis vectors are made orthonormal to each other as well as to $\hat{r}_1$ and $\hat{r}_2$ via the Gram Schmidt process. **Figure 2-5** shows examples of the derived basis functions for a selection of observation parcels. The solid black line is $\hat{r}_1$, long dash purple line is $\hat{r}_2$, short dashed blue line is $\hat{c}_1$, and the green dash-dot and red dash-dot-dot-dot lines are $\hat{c}_2$ and $\hat{c}_3$ respectively. This convention will be used throughout the paper to indicate the component or component projection plotted. The primary background basis vector $\hat{r}_1$is roughly symmetric as is expected for Rayleigh scattering while the primary cloud basis vector $\hat{c}_1$ accounts for the fact that the clouds scatter more efficiently in the forward direction. The remaining basis vectors represent the more detailed differences between scattering profiles of different *r* or σ.

It is not possible to create more orthogonal basis vectors than the number of observations, *n*, so for $n < 5$, fewer cloud basis vectors can be created. If $n = 5$, the basis vectors will span all possible (bounded) signals. For $n > 5$, the basis vectors will span all possible *C*-σ and cloud signals (apart from the small residuals discussed above), but not all possible signals. While it is true that there is some residual component of the cloud set that is not spanned by the five

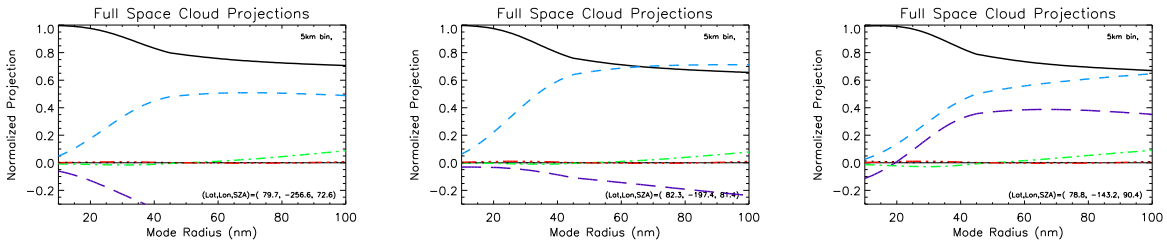

Figure 2-6 The projections of the normalized cloud signal onto the basis functions in Figure 5 as a function of mode radius. The color and line style indicates the component projection using the same convention as **Figure 2-5**.

components, it is very small as was the case with the background subspace. This is particularly true considering the presence of noise which completely washes out this small residual signal.

The result of projecting the normalized cloud phase functions onto the basis vectors in **Figure 2-5** can be seen in **Figure 2-6**. Although the details are different in every observation parcel, the following observations are generally true. The cloud signals are dominated by their $\hat{r}_1$ and $\hat{c}_1$ components with $\hat{r}_2$ being the third most dominant component. For small particles the cloud signal is primarily within the background subspace. This should be expected because the small particle limit should approach Rayleigh scattering. The bulk of the difference between clouds of different radii is in the relative amount of $\hat{r}_1$ and $\hat{c}_1$ in the signal; smaller particles have larger $\hat{r}_1$ contribution (more Rayleigh like), and larger particles have more $\hat{c}_1$ and thus scatter forward more efficiently. This is consistent with the example given in section 2.2. The $\hat{c}_2$ component is small but positive for large particles and negative for small particles, but the $\hat{c}_3$ component is significant only for small particles; however, this is hard to see in **Figure 2-6**. See **Figure 2-7** and **Figure 2-8** for a better view of the $\hat{c}_2$ and $\hat{c}_3$ components.

Similar to the case with the background, most of the variance within the cloud subspace for the cloud set is in the primary cloud component. This is unfortunate since this means most of the variance in the signals is in the three components $\hat{r}_1$, $\hat{r}_2$and $\hat{c}_1$which are a function of four

variables *C, σ, A* and *r*. Unique determination of all four parameters with only three measurements is not possible. At best, measurement of these components provides *r* as a function of the other three variables. This means that at least one more component is needed to obtain a solution. This is made difficult because only a small fraction of the signal projects onto the $\hat{c}_2$ and $\hat{c}_3$ components, so sufficient signal to noise is required to resolve them. Fortunately CIPS is able to acheive the necessary noise performance provided the cloud is sufficiently bright and/or the data is binned into large enough pixels. The cloud brightness and noise requirements are described in the following sections.

2.4 Mode radius Retrieval

2.4.1 Least Squares Fitting

The retrieved parameters are chosen such that they produce the least squares best fit of the model scattering profile to the observed profile. If the background is bright in comparison to the cloud signal, as it is over most of the orbit, the background components cannot be used for radius determination; however, at the highest *SZA* the contribution to the background components from the *C*-*σ* background drops off and the background components can provide some help in determining cloud parameters. The cloud albedo that minimizes the fitting error in the cloud subspace can be solved for as a function of *r*, so the least squares fitting problem when the background components are not usable reduces to finding the radius for which the normalized model cloud subspace vector is closest to the normalized observation's cloud subspace vector.

In the absence of knowledge of the background parameters, the background components are difficult to use to aid in the determination of radius. This can be seen by a look at the squared error in the transformed basis. Using the new basis we can represent a *C*-*σ* background $\vec{b}$ as,

$$\vec{b} = b_{r1}(\sigma)\hat{r}_1 + b_{r2}(\sigma)\hat{r}_2 \qquad (5)\ ,$$

where $b_i(\sigma)$ are the projections for a 1G background. The model cloud scattering profile $\vec{c}$ is,

$$\vec{c} = c_{r1}(r)\hat{r}_1 + c_{r2}(r)\hat{r}_2 + c_{c1}(r)\hat{c}_1 + c_{c2}(r)\hat{c}_2 + c_{c3}(r)\hat{c}_3 \qquad (6)\ ,$$

where $c_i(r)$ are the projections for a 1G cloud. The squared error E^2 for a given observed scattering profile $\vec{o}$ with a model background albedo, *C*, and cloud albedo, *A*, is,

$$\begin{aligned} E^2 = {} & (o_{r1} - Cb_{r1} - Ac_{r1})^2 + (o_{r2} - Cb_{r2} - Ac_{r2})^2 \\ & +(o_{c1} - Ac_{c1})^2 + (o_{c2} - Ac_{c2})^2 + (o_{c3} - Ac_{c3})^2 \end{aligned} \qquad (7)\ .$$

Over most of the orbit the background is bright enough such that over the range of reasonable C-σ pairs the range of terms Cb_{r1} and Cb_{r2} are larger than the range of Ac_{r1} and Ac_{r2} as a function of mode radius. This has the consequence that over a wide range of radii the background subspace error terms can be made to be identically zero with the appropriate choice of *C* and *σ*, so the radius which minimizes the cloud subspace error terms also minimizes the total squared error and is therefore the least squares best fit. This is what was done in the example given in section 2.2. This condition of being able to zero the background subspace error contribution will be referred to as the "bright background condition".

It can be seen in **Figure 2-4** that the relative contribution of the $\hat{r}_2$ component drops significantly at the highest zenith angles. Since the absolute magnitude of the background also

decreases, the significance of the background contribution to the $\hat{r}_2$ component decreases significantly at high zenith angles. At some point it decreases to the point that the bright background condition will not be satisfied, so the secondary background component can aid in the radius determination. At around 92° *SZA* the entire background drops to nearly zero, so for the very highest zenith angles, all 5 components are purely cloud and can be used for retrievals; however, for the analysis in this paper, only the cloud components are used, so it should be kept in mind that some of the issues that appear at high *SZA* may be remedied by the additional information in the background components.

For any *r*, the cloud albedo which minimizes the least squared error in the cloud subspace is,

$$A = \frac{o_{c1}c_{c1} + o_{c2}c_{c2} + o_{c3}c_{c3}}{c_{c1}^2 + c_{c2}^2 + c_{c3}^2} = \frac{\vec{o}_c \cdot \hat{c}_c}{|\vec{c}_c|} \qquad (8)\ .$$

where the *c* subscript on the right hand side indicates it's the component of the vector within the cloud subspace. The least squared error is then,

$$E^2 = \left(\vec{o}_c - \frac{\vec{o}_c \cdot \vec{c}_c}{c_c^2}\vec{c}_c\right)^2 = \mathrm{o}_c^2(1 - (\hat{o}_c \cdot \hat{c}_c)^2) \qquad (9)\ ,$$

so the least squared best fit is the radius for which the normalized model cloud vector is nearest the normalized observation vector within the cloud subspace. This implies that it is only the normal vectors, or equivalently the relative projection magnitudes, within the cloud subspace which are needed to determine *r*. The absolute magnitude has no influence on the retrieved *r*.

2.4.2 Radius Phase Space

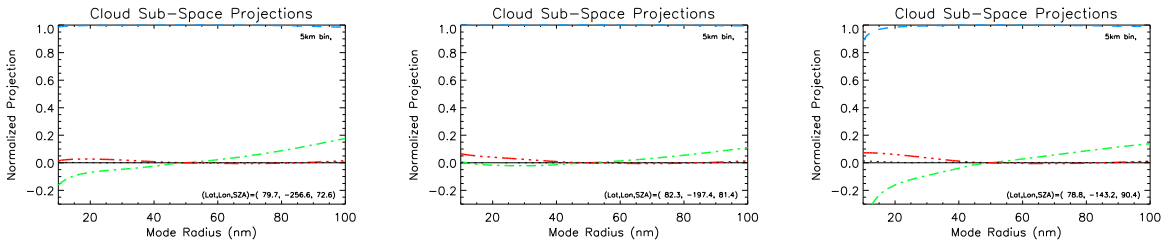

Figure 2-7 The projections of the normalized cloud signal within the cloud subspace projections. The color and line style indicates the component projection using the same convention as **Figure 2-5**.

Most of the cloud subspace projection for the cloud set residing in the primary component results in errors in the primary component having little effect on the normalized cloud subspace vector and thus radius retrieval. The retrieved value for r is determined primarily by the normalized values for the $\hat{c}_2$ and $\hat{c}_3$ projections and the errors in these projections, so looking at a plot in the $\hat{c}_3$ vs $\hat{c}_2$ phase space of the radius parameterized curve in addition to the noise distribution in this space is useful for understanding the effect of noise on the least squares best fit.

Using the same parcels as those in **Figure 2-5** and **Figure 2-6**, **Figure 2-7** shows the normalized projections within the cloud subspace vs. r. Most of the cloud subspace magnitude is in the primary cloud component, so errors in the measurement of this component have very little effect on the cloud subspace observation vector direction and thus the normalized vector, so the effect on the radius retrieval is minor; however, it is errors in the $\hat{c}_1$ component which dominate the total cloud subspace norm, $|\vec{o}_c|$, thus cloud albedo is affected. Errors in the radius retrieval are almost entirely due to measurement of the $\hat{c}_2$ and $\hat{c}_3$ components, and since the radius effects the magnitude of the background subspace projections, albedo is also indirectly affected.

Since r is determined from the $\hat{c}_2$ and $\hat{c}_3$ components, it's useful to visualize things from a phase space prospective. In **Figure 2-8** we see the r parameterized curve in the $\hat{c}_2$ vs. $\hat{c}_3$ phase

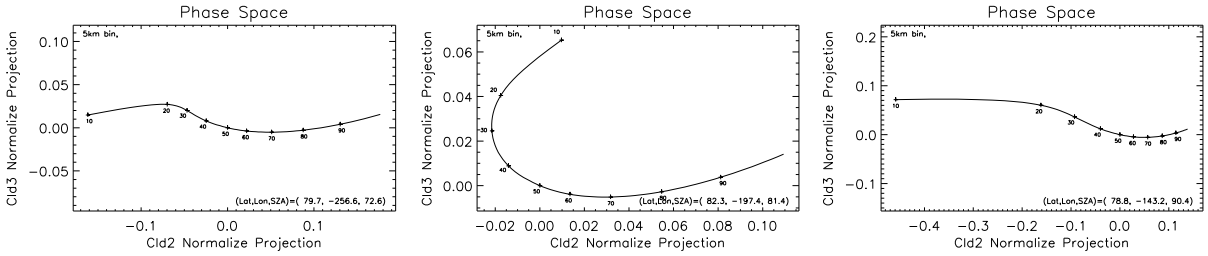

Figure 2-8 The $\hat{c}_3$ vs $\hat{c}_2$ radius parameterized curve for the normalized cloud subspace projections. The labels along the curve indicate radius.

space. A given observed scattering profile will have some position in this phase space. This position is not necessarily exactly on the curve due to measurement noise and imperfections in the assumptions in the fitting model. Since the primary cloud component's impact is negligible (it's always approximately 1), from **Equation 9** we can see that the least squared best fit *r* is the one for which this curve passes nearest the observed point in the phase space. In cases like the center plot, there are phase space points for which more than one point on the curve is equidistant. For these points, the radius inversion is multi-valued.

Measurement noise projects into the phase space in a very straight forward way. Using the component functions and the expected random error on each point in the scattering profile δo_x, we can propagate the errors into the new basis. The SD in the observed projection δc_i is,

$$\delta c_i = \frac{1}{n}\sqrt{\sum_{x=1}^{n}(\delta o_x \, c_{ix})^2} \qquad (10) \quad ,$$

where the *x* index indicates the element of the scattering profile and the *i* index indicates the component. The probability distribution associated with the observation of a given cloud in the presence of noise is a 2-D Gaussian in the phase space with widths given by **Equation 10**.

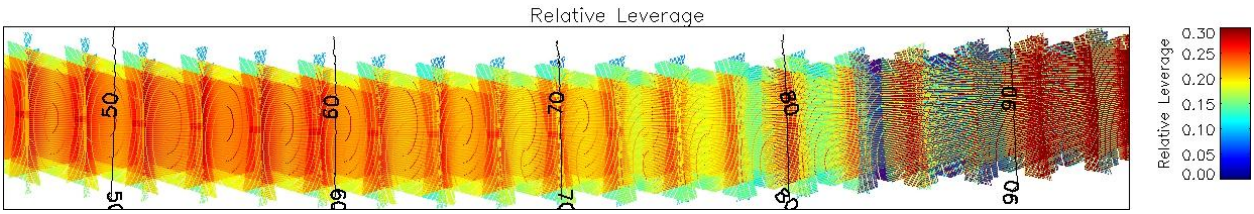

Figure 2-9 Orbit map of the radius leverage for a cloud of unit cloud subspace magnitude. This can be multiplied by the cloud visibility to obtain the absolute radius leverage.

2.4.3 Radius Leverage

The scale of the phase space is a useful metric for determining the effect of errors on radius retrieval. An error which is small compared to this scale size will have a small impact on radius retrieval. Assuming the bright background condition holds, errors which are large compared to this scale size will make retrieval difficult. For the remainder of the paper this scale size will be estimated using the diagonal size of the box containing the phase space curve from 25 to 100 nm. This measure will be referred to as "radius leverage" or sometimes just "leverage".

The relative radius leverage which is the leverage within the normalized phase space shown in **Figure 2-8**, is shown in **Figure 2-9**. It is relatively constant over most of the orbit with higher values around 0.25 near the center of the orbit strip and lower values on the edges. At high SZA it is more variable. One reason for this is that the shape of the Rayleigh background is fairly constant over much of the orbit, but as the SZA starts to approach 90°, the shape starts to change quickly due to the chapman function. This changes the intersection between the background subspace and total cloud space which can reduce or increase the leverage in the cloud subspace.

Using the radius leverage, the claim made in section 3.2 that the two background basis vectors are sufficient to effectively remove the *C-σ* background from the signal can be justified. Shown in **Figure 2-10** is the maximum cloud subspace norm for *C-σ* backgrounds with σ

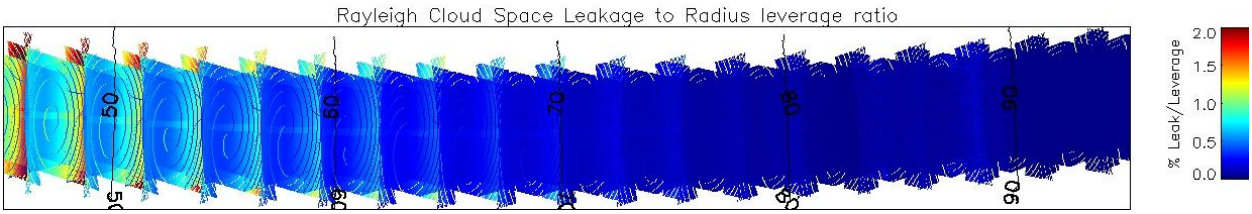

Figure 2-10 Orbit map of the ratio of the maximum *C-σ* projection into the cloud subspace to the radius leverage for a dim cloud (1 G, 50 nm). Since this ratio is very small the impact on radius retrieval is also very small. For brighter clouds the impact is even smaller.

between 0.52 and 0.58 divided by the radius leverage for a cloud with a 1 G albedo and a 50 nm radius. This is a dim cloud at the edge of detectability with a typical mode radius; however, the maximum error is under 1% for most of the orbit. This is much less than the random errors for a cloud this dim. Brighter clouds will have even smaller relative error, so for clouds bright enough to be retrieved, the residual error introduced by assuming the cloud components are free of *C-σ* contamination is insignificant; however, although very small, this is an additional error introduced by this analysis. Unlike most of the conclusions in this paper, it does not apply to a general curve fitting algorithm such as the one used in the operational CIPS retrieval.

The "noise level' in the phase space δc_r is taken as the total noise added in quadrature, so $\delta c_r = \sqrt{\delta c_2^2 + \delta c_3^2}$. This noise level is plotted in **Figure 2-11** for a dim relatively symmetric cloud (2 G, 30 nm) and bright asymmetric one (20 G, 70 nm). The noise level is very similar in both the bright and dim cloud cases except at the highest SZA. Typically, the background's contribution to the albedo is much larger than the clouds, so the clouds impact on the noise level is small. This has the convenient consequence that the noise level is usually independent of the cloud, and since the background is roughly constant on scales relevant to the noise it produces, we can think of the noise level at a given location in the orbit strip as a constant; however, as observations approach the terminator, the background drops off sharply, so at the highest SZA, the total signal, and therefore the noise level, is determined by the cloud.

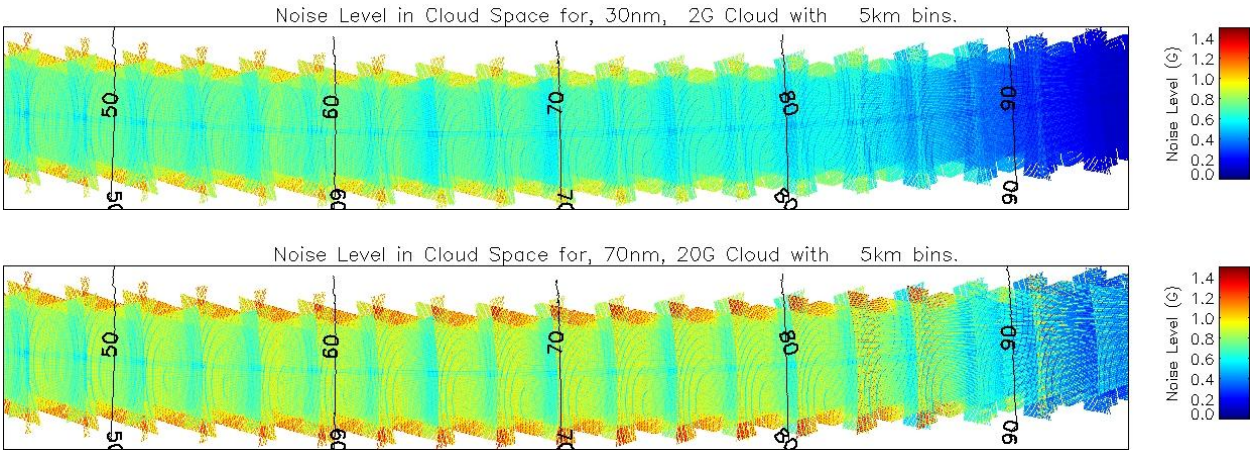

Figure 2-11 Orbit maps of the noise level projecting into the $\hat{c}_3$ and $\hat{c}_2$ components for dim (top) and bright (bottom) cloud cases.

To estimate the impact of noise on the radius retrieval a useful metric is the leverage to noise ratio (LNR) with the noise being the total noise level calculated above. The LNR should be significantly greater than one to allow radius retrieval. When the LNR is significantly larger than one, 1/LNR is a fairly good predictor of the SD in the retrieved radius as one might expect; however, in some pathological cases, the phase space curve can be very non-linear. In these cases, using LNR to predict the SD in *r* is too simple. The SD vs. LNR relationship is covered in more detail in section 5.5.

2.4.4 Cloud Visibility

The cloud subspace magnitude of the best fit model for any *r* is,

$$\left|\frac{\vec{o}_c \cdot \vec{c}_c}{c_c^2}\vec{c}_c\right| = |\vec{o}_c \cdot \hat{c}_c| = |\vec{o}_c|\cos[\angle(\vec{o}_c, \hat{c}_c)] \qquad (11)$$

From **Figure 2-7** we can see that the angles between vectors in the cloud subspace are small since they are all nearly parallel to $\hat{c}_1$, thus $|\vec{o}_c|\cos[\angle(\vec{o}_c, \hat{c}_c)] \cong |\vec{o}_c|$. So the cloud subspace magnitude of the best fit is approximately independent of *r* and equal to the cloud subspace magnitude of the observed scattering profile. The cloud subspace magnitude, which will be

referred to as “visibility”, is directly observable from the scattering profile without the need to retrieve radius. Much can be inferred from the LNR and visibility alone, so the fact that they are independent of the radius retrieved is very convenient.

When divided by the total noise power in the cloud subspace, the visibility to noise ratio (VNR) makes a natural cloud presence indicator. The noise in this case would be δc_1, since uncertainties in visibility are almost exclusively due to this component. Since the noise transformation into the cloud subspace preserves the Gaussian distribution, the VNR can actually put a firm number on the confidence of cloud presence in the parcel. For example, a VNR of greater than 3 has less than a 1% chance of being produced by noise, since it is greater than a 3 SD outlier; however, at this VNR the LNR is not likely to be good enough for radius retrieval since the relative leverage is typically less than 0.3 and $\delta c_1 \cong \delta c_r$, therefore a VNR of 3 implies a LNR of around one. This implies that for most of the orbit the visibility required to determine a cloud is present is less than the visibility required to make a meaningful radius retrieval.

Setting a threshold for detection in VNR naturally accounts for the fact that CIPS sensitivity to a cloud is a function of *r*, *A*, observing geometry and the *C-σ* background present. The relation between cloud albedo, mode radius and visibility can be found using **Equation 8**, so the VNR can be used to determine if a given cloud should be visible above the noise. Alternatively, the visibility can be used to determine the noise level required to view a given cloud. Since binning the data at lower resolution reduces the level of noise, the required resolution can be determined.

Using the visibility, the normalized radius leverage shown in **Figure 2-9** can be scaled to the absolute radius leverage for the observed parcel. The phase space noise power can be

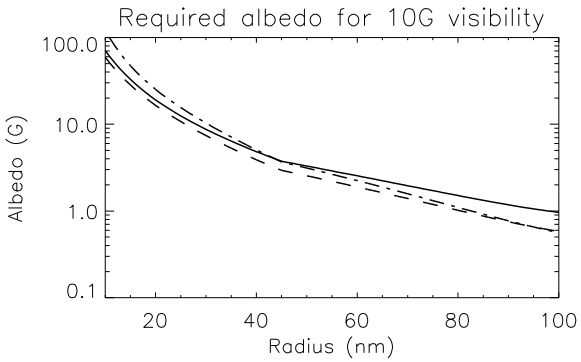

Figure 2-12 Required cloud albedo for a fixed visibility of 10 G for parcels at: Solid 70°, dash 80°, dash-dot 90° *SZA*. These are again the same example parcels used in previous figures.

calculated and used to calculate the LNR, so the uncertainty in radius retrieval can be estimated without retrieving radius. The cloud albedo as a function of *r* for a visibility of 10G is shown in **Figure 2-12**. For a given visibility, the retrieved albedo is a strong function of mode radius, so obtaining a small error in mode radius is important for getting a small error in albedo, particularly if the particle size is on the small end.

Putting a threshold on the retrieval based on a fixed cloud albedo may lead to problems. Since radii of the same visibility have approximately the same retrieval uncertainty and smaller radii fits to a given observation will have larger albedos, a fixed threshold in albedo will tend to filter out large radii clouds with similar retrieval uncertainties to the small radii clouds which would not be filtered. If a cloud with a LNR too low to precisely retrieve radius happens to be retrieved with a small mode radius, its albedo would be high, so it may pass a fixed albedo threshold even though the retrieval is highly uncertain.

2.5 Simulated Retrievals

2.5.1 Noise Simulation

In order to better understand the uncertainties, concrete examples are useful. In this section simulated retrievals are done in the presence of noise. This provides a better understanding of the tools developed above, and provides a check on some of those conclusions. Three example cloud layers are used. In the first case a cloud is chosen which is too dim based on its LNR to provide meaningful retrievals. In the second case a cloud is chosen that is borderline in the sense that in some regions in the orbit the LNR is sufficient for radius retrieval and in other regions it is insufficient. The third case uses a cloud for which its LNR is more than large enough everywhere. While the LNR in these examples tends to be a reasonably good metric for determining how well the radius retrieval will work, there is a small subset of cases where it fails due to pathological phase space curves which make unique radius determination difficult.

The simulated retrievals are done with noise levels appropriate for CIPS images. Using Fig. 16 of *McClintock et al.* [2008], the one SD noise uncertainty for a level 1A data pixel is approximated to be $\sqrt{A/50}$, where A is the total pixel albedo in units of G. For retrievals the 1A images are transformed onto an equal area Lambert grid at the 83km cloud deck. The data transformed in this way is called 1B data. In this analysis, the 1B value is set to the mean of all the 1A pixels which fall into a given 1B bin for a given image. The contribution from each image in the bin is one element of the scattering profile for that bin.

Transforming the data in this way provides noise reduction by a factor of $1/\sqrt{n}$ where n is the number of 1A pixels falling within the 1B bin. In this way the random errors can be

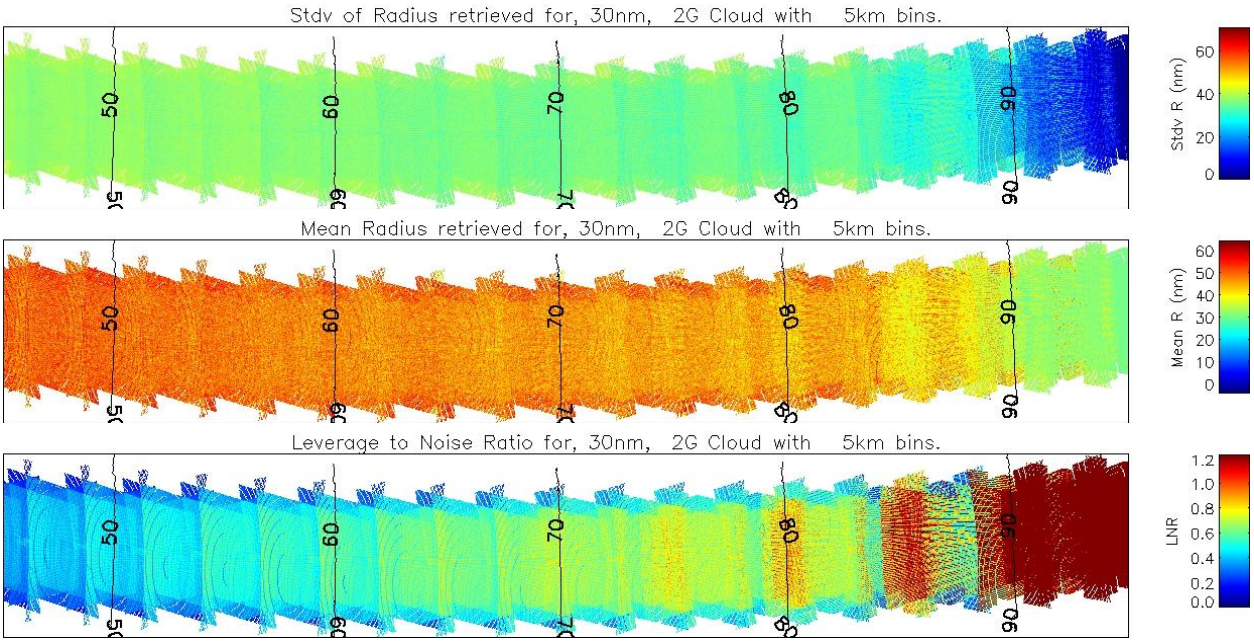

Figure 2-13 Simulated retrievals for a dim cloud (2 G, 30 nm, 5 km bins). Top is an orbit map of the standard deviation in the retrieval, middle is the mean radius retrieved and bottom is the LNR.

reduced at the expense of resolution. It should be noted that in the operational retrieval the transformation from 1A to 1B is an interpolation, so this potential noise reduction is not currently utilized; however, some noise reduction is still achieved by using multiple 5km 1B pixels for retrieval.

For each 1B bin and model scattering profile, simulated Gaussian noise is added to the signal appropriate for the level of 1B binning and scattering profile albedo. The least squares best fit is then retrieved using **Equation 9**. This process is repeated 200 times for every model scattering profile. The distribution of the retrievals can be used to estimate the uncertainties in retrieving a given cloud and checked for consistency with the expectations derived in section 4.

2.5.2 Dim Cloud retrieval

Shown in **Figure 2-13** are the results of the simulated retrieval for a cloud layer with a 2 G albedo and a 30 nm radius with 5km bins. The LNR, except for at the highest SZA, is well

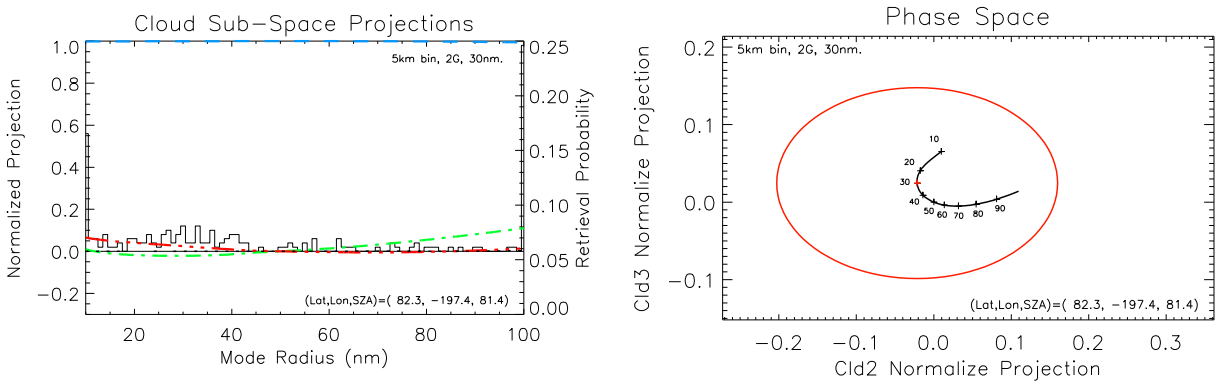

Figure 2-14 Retrieval details for dim cloud case (2 G, 30 nm, 5 km bins) using the 80° example parcel in figures above. Left is a plot of the normalized cloud subspace projections as in Figure 7, but with the simulated retrieval histogram over plotted in black. The histogram has a Mean = 49.9 and a SD = 34.4. Right is a phase space plot for the parcel as in **Figure 2-8** but with the one SD error ellipse for the 2-D Gaussian representing the noise spread of the observation plotted in red. The red point on the radius curve represents the noise free cloud vector. This parcel has a LNR = 0.6.

below 1, so as expected, the retrievals are not good. The mean retrieved radius is near the center of our retrieval range at about 50 nm instead of the 30 nm radius input, and the SD of the retrievals is on the order of half the retrieval range. For a cloud with a LNR this low the radius inversion is more a result of the noise than the input radius, and as a result the retrieved radius is essentially a random number.

At high SZA, the LNR is significantly higher saturating the color scale of the image. The primary reason for this increase in LNR is the sharp decrease in noise shown in **Figure 2-11**. The LNR is also more variable above 80°. This is a result of variations in leverage as can be seen in **Figure 2-9**. Although for most of the orbit the LNR is too low for meaningful retrievals, spikes in the LNR at high SZAs are clearly correlated with improvements in the retrievals.

Shown in **Figure 2-14** is a more detailed look at how a low LNR effects the retrievals. This is the same 81.4° SZA parcel used in previous figures. Added to the cloud subspace projection plot is the retrieval histogram. Essentially all radii are retrieved with a slight skew towards small particles and many instances of either a 10nm or 100nm retrieval. This can be

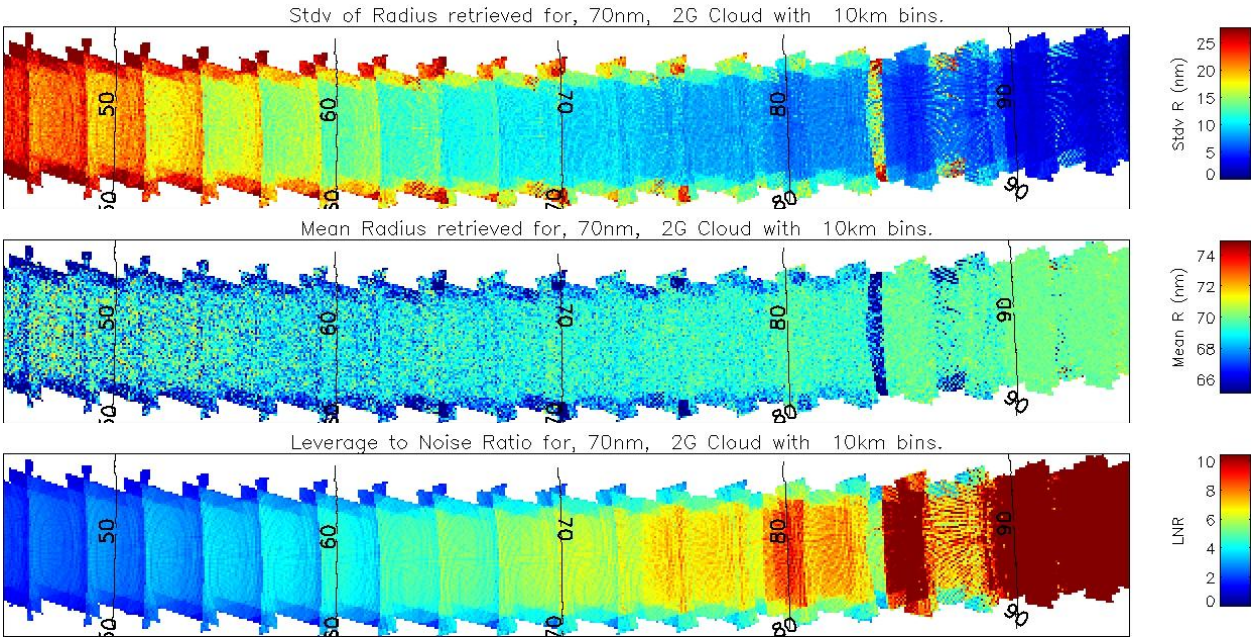

Figure 2-15 Simulated retrievals for a borderline cloud (2 G, 70 nm, 10 km bins). Top is an orbit map of the standard deviation in the retrieval, middle is the mean radius retrieved and bottom is the LNR.

understood by looking at the phase space plot where the one SD error ellipse has been added in red. The error distribution is much wider than the range allowed by the clouds, and since 10 nm and 100 nm are at the end points of the curve, a large section of the noise probability distribution will be nearest to those points. Similarly, the section of the curve from 20 to 50 nm is going to be closer to a larger fraction of the distribution than the large radii are. In cases of low LNR, the retrieval histogram is more a result of overall shape of the phase space curve than it is of the true cloud mode radius.

2.5.3 Borderline Cloud Retrieval

For the borderline case in **Figure 2-15**, the cloud still has a 2 G albedo, but it has a 70 nm radius instead of 30 nm as in the dim cloud case. Even though the albedo is the same, it can be seen from **Figure 2-12** that it has approximately a factor of five larger visibility. Also, the binning has increased to 10km. Since four 5km pixels fit in a 10 km pixel, the noise will be reduced by approximately a factor of two. This is reflected in the LNR which for this case is

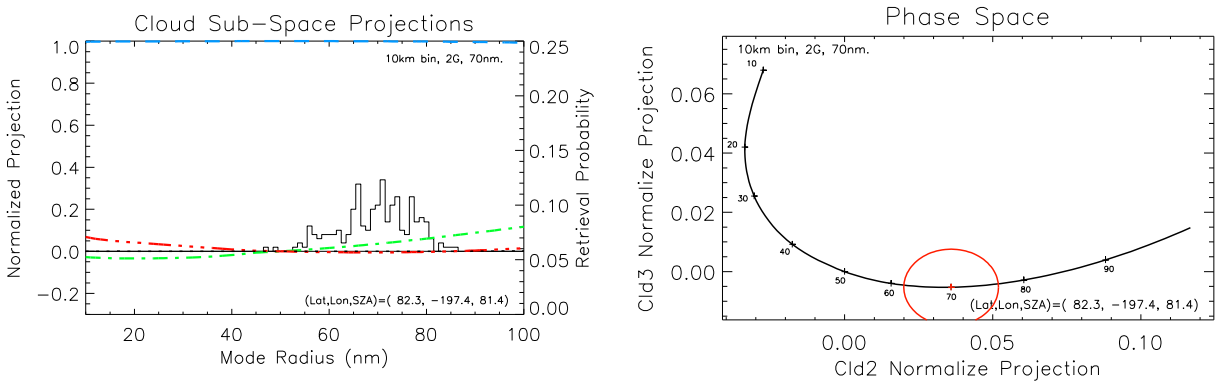

Figure 2-16 Retrieval details for borderline cloud case (2 G, 70 nm, 10 km bins) using the 80° example parcel in figures above. See **Figure 2-14** for a legend to the plots. The retrieval simulation in this case had Mean = 69.6, SD = 7.4. The LNR = 7.5.

about an order of magnitude larger than the dim cloud case. This case is borderline in the sense that at the lower SZA end of the orbit the LNR is on the order of 1, so it is right on the edge of producing a meaningful radius retrieval. As the SZA increases, the LNR tends to improve with the exception of a few spots between 80° and 90°. The standard deviation and mean error mirror this general increase in LNR with SZA. The edges of the orbit strip tend to have a larger SD and also a bias toward slightly lower mean radius retrieval.

This small radius bias is due to the scale of the non-linearity in the phase space being significant compared to the width of the noise distribution. The radii below 70 nm tend to be more closely spaced than the radii above 70 nm, so noise which pushes the observation vector towards smaller radii has a larger effect than noise which pushes the observation vector towards larger radii. An example of this can be seen in the phase space plot of **Figure 2-18.**

Shown in **Figure 2-16** is the 81.4° example parcel used in previous figures. The noise ellipse is now small in comparison to the phase space curve, so radius retrievals are much better than in the dim cloud case, but the non-linearity of the phase space appears as though it could introduce a small bias. With a 7.4 nm SD and 200 simulated retrievals, the uncertainty in the

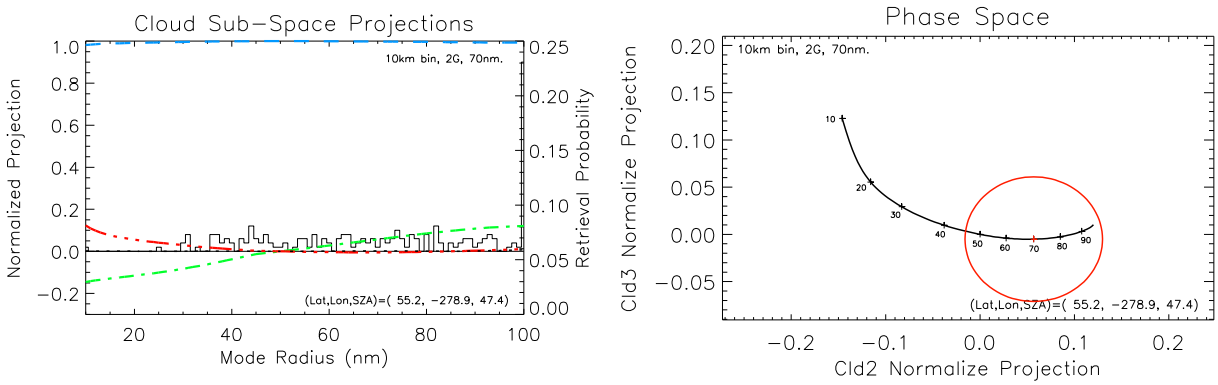

Figure 2-17 Retrieval details for borderline cloud case (2 G, 70 nm, 10 km bins) using a low SZA parcel at 47° near the middle of the orbit strip. See **Figure 2-14** for a legend to the plots. The retrieval simulation in this case had Mean = 71.8, SD = 23.1. The LNR = 2.3.

mean is 0.5 nm, so the 0.4 nm mean error is on the edge of statistical significance; however, looking at neighboring pixels there is a similar negative bias, so together with the small asymmetry in the phase space curve, it's likely some of this bias is not simply due to the random error in the mean. This bias is a small source of error compared to the SD, and unlike typical bias errors it will improve with improved signal to noise.

In **Figure 2-17** is an example of a low SZA retrieval. This parcel has a much smaller LNR. The error ellipse is still smaller than the radius curve, but only by about a factor of two. The SD is very high at 23 nm, and the mean radius is off by 1.8 nm; however, a closer look at the histogram reveals that many retrievals are being retrieved as 100 nm. This is happening because the noise distribution has significant probability off the right edge of the phase space plot all of which would be retrieved as 100 nm. If the retrieval had allowed higher radii, the mean retrieval would likely have been biased high by a more significant amount. For this parcel a cloud of this LNR will not produce very reliable retrievals, but it is good enough to say whether the particle size is on the large end or small end.

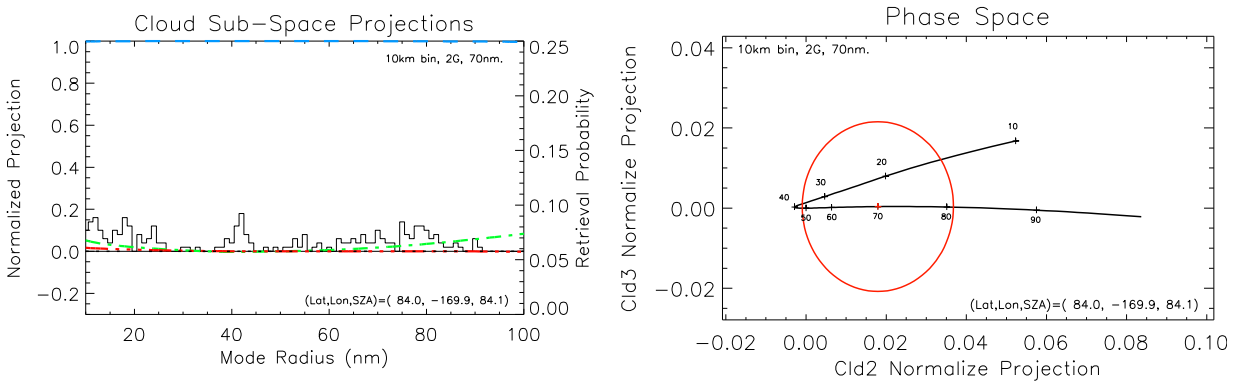

Figure 2-18 Retrieval details for borderline cloud case (2 G, 70 nm, 10 km bins) using a 84° SZA example near the top of the orbit strip. See **Figure 2-14** for a legend to the plots. The retrieval simulation in this case had Mean = 47.1, SD = 26.3. The LNR was 3.1.

An obvious feature in **Figure 2-15** is the regions for which there is a sudden drop in retrieval performance between 80° and 90°. Shown in **Figure 2-18** is a parcel contained within one of these features near the top of the orbit strip at 84.1°. A close look at this strip when comparing it to regions of similar LNR at lower SZAs shows that the retrievals are atypically bad even for the lower LNR. The reason for this is that the radius curve folds back over on itself within the phase space. Although there is a subset of the retrieval histogram that is roughly Gaussian and centered near the correct radius, another subset of the distribution is closer to the top branch of the curve and so many low radius retrievals are found as well. There is an additional subset of retrievals at approximately 40nm. Although it is slightly outside the one SD error ellipse, it's positioned such that all observation vectors which are on the left hand side and a little farther than one SD away will be retrieved as 40 nm. It is clear from the shape of the phase space curve that any retrieval in this parcel with a true radius from 30 to 60 nm is going to be difficult to retrieve even for a very high LNR, so LNR will not be a good metric in these cases. Fortunately, pathological phase space curves like this make up a small subset of the orbit strip, and for those that are like this, a more detailed look at the phase space curve can diagnose the problem when LNR is insufficient.

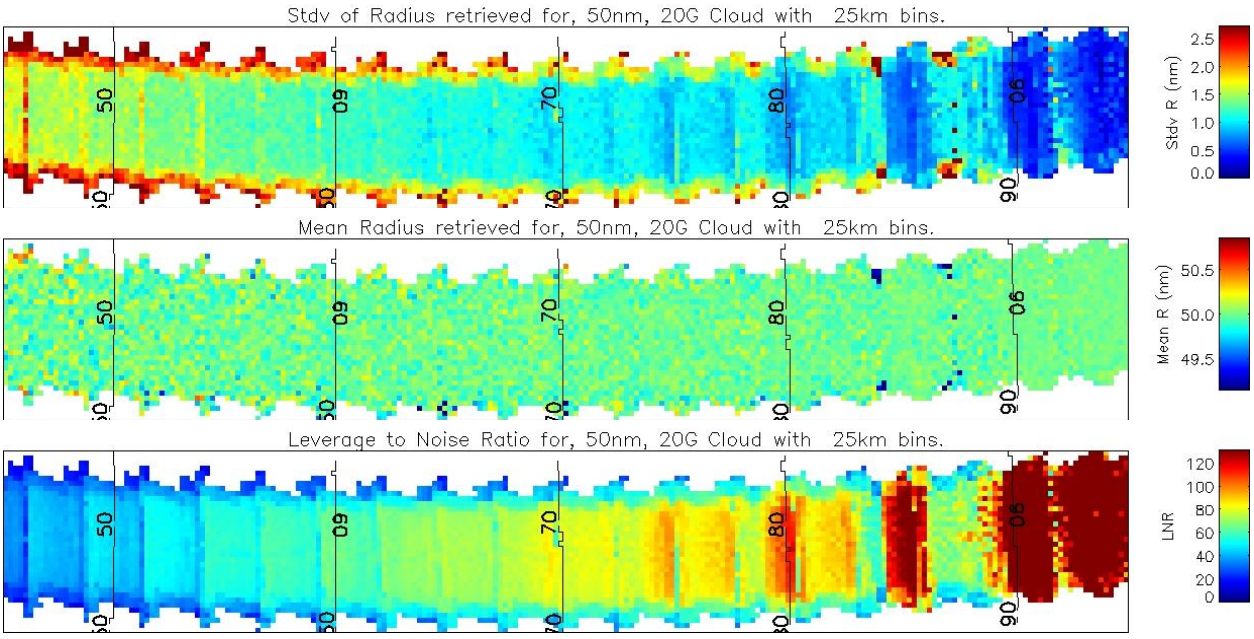

Figure 2-19 Simulated retrievals for a bright cloud (20 G, 50 nm, 25 km bins). Top is an orbit map of the standard deviation in the retrieval, middle is the mean radius retrieved and bottom is the LNR.

2.5.4 Bright Cloud Retrieval

For this case shown in **Figure 2-19,** the LNR is another order of magnitude larger as a result of the cloud having approximately 20 times the visibility and the noise level being about a factor of 5 lower than the dim cloud case. The LNR is large enough for good retrievals at any *SZA* in the orbit strip. The standard deviations are on the order of 1 nm for the vast majority of the parcels in the orbit and the mean errors in most cases are much less than 1 nm. The problem areas between 80° and 90° are still there, but the size of these areas is reduced significantly when looking at the mean radius errors. This indicates that fewer of the bins are suffering from multi valued radius inversions.

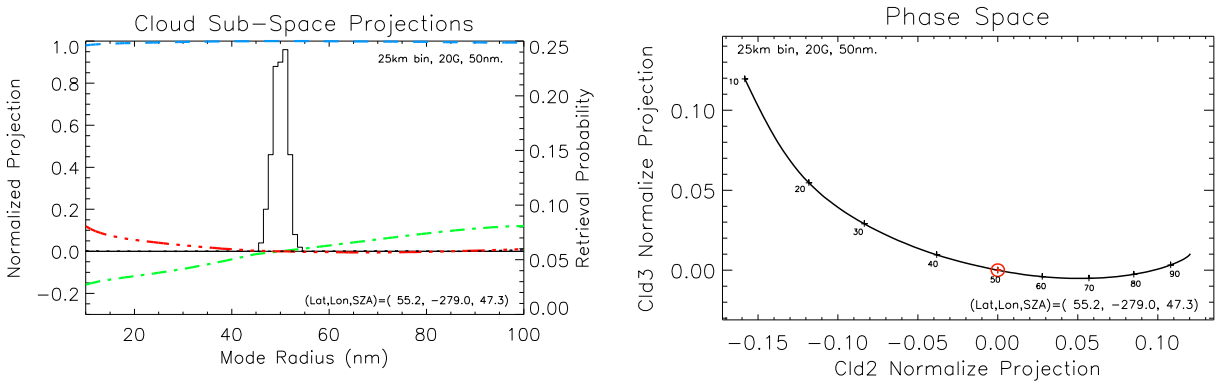

Figure 2-20 Retrieval details for bright cloud case (20 G, 50 nm, 25 km bins) using a low SZA parcel at 47° near the middle of the orbit strip. See **Figure 2-14** for a legend to the plots. The retrieval simulation in this case had Mean = 49.9, SD =1.5. The LNR was 40.2.

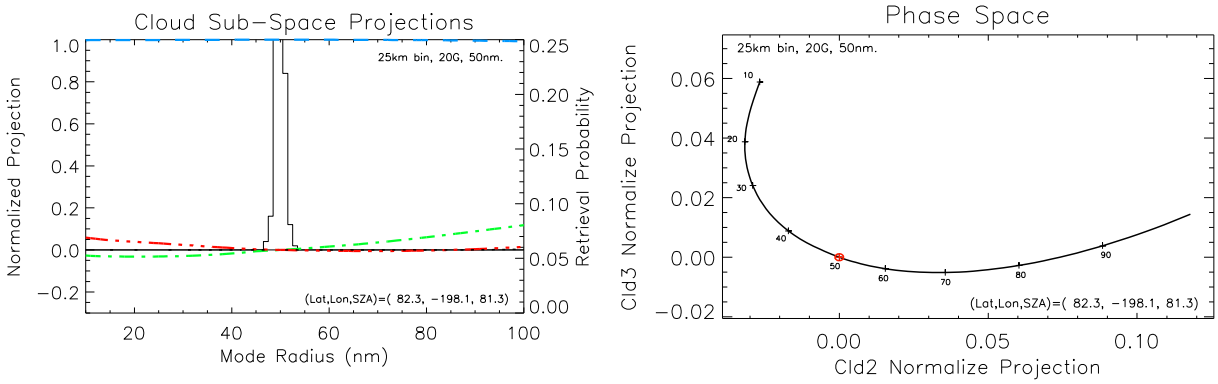

Figure 2-21 Retrieval details for bright cloud case (20 G, 50 nm, 25 km bins) using a high SZA parcel at 80° near the middle of the orbit strip. See **Figure 2-14** for a legend to the plots. The retrieval simulation in this case had Mean = 49.9, SD = 0.9. The LNR is 84.2.

Shown in **Figure 2-20** and **Figure 2-21** are parcels in approximately the same location as the parcels used in the borderline cloud case for high and low *SZA* respectively. As before, the high *SZA* retrieval is very good, but with an order of magnitude larger LNR, the SD has dropped to only 0.9 nm from 7.4 nm in the borderline case. The low *SZA* parcel which was unable to provide a good retrieval in the borderline case is now able to easily have its radius determined with a SD of only 1.5 nm.

As expected, the problem area at around 84.1° near the top of the strip shown in **Figure 2-22** still has problems even with a much brighter cloud. The retrieval histogram looks like 2 Gaussians, one centered on the correct answer and one centered a little below 30 nm. This

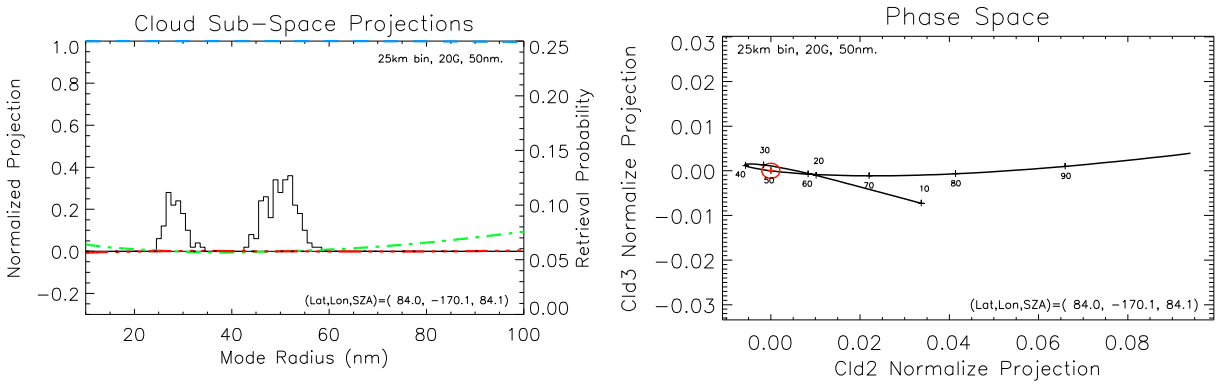

Figure 2-22 Retrieval details for bright cloud case (20 G, 50 nm, 25 km bins) using a 84° SZA example near the top of the orbit strip. See **Figure 2-14** for a legend to the plots. The retrieval simulation in this case had Mean = 43.2, SD = 10.4. The LNR is 38.0.

behavior is exactly what one would expect when looking at the phase space curve. The two solutions are almost on top of each other in the phase space, so it would take a very high LNR to resolve them. Even worse would be trying to retrieve at around 60 or 20 nm where the two branches actually cross one another. At these radii, no amount of LNR can remove the degeneracy in the retrieval.

Remember that restricting the inversion to the cloud subspace was based on the bright background condition defined in section 4.1. As the observations approach the terminator, the background drops off sharply, so at some point the condition fails to be satisfied. Since these bad parcels occur at high *SZA*, it may be possible to resolve the degeneracy by looking to see if in one of the two cases the background parameters retrieved are outside of some reasonable range.

2.5.5 SD vs. LNR

As stated earlier, LNR can be used to give a very rough idea of the uncertainty in the retrieved *r*. In cases where there are not significant nonlinear features in the phase space curve, LNR can be thought of as approximately the number of one SD divisions within the retrieval range. So one should expect a rough proportionality of SD ~ 1/LNR, but since the phase space

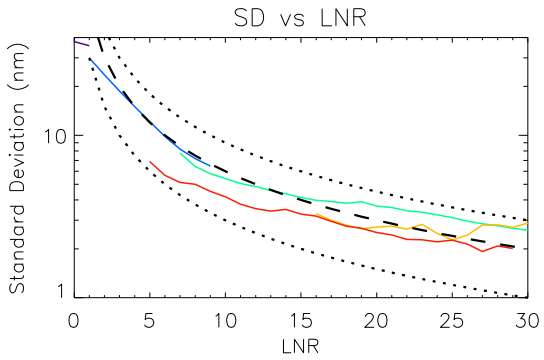

Figure 2-23 The mean SD vs LNR for 2 G 30 nm 5 km bin in purple, 2 G 70 nm 10 km in blue, 20 G 50 nm 10 km in blue-green, 20 G 30 nm 25 km in orange, 20 G 40 nm 10 km in red. The black dashed line is 60/LNR and the dotted lines are ± 50% from that curve.

curves always have some amount of nonlinearity and the measure for leverage is far from perfect, this relation is very rough; however, when considering errors, approximate values are often good enough.

Shown in **Figure 2-23** are the mean standard deviations vs. LNR for a select set of simulated cloud layer retrievals. There is a significant amount of spread in the curves for different cloud parameters, and there is an equally significant spread in the points making up each mean standard deviation point on those curves, but there is clearly a correlation between the standard deviation and LNR. The black dashed line is a plot of 60/LNR. This curve was simply chosen by eye to roughly fit the data, but it's good enough to predict the SD to about the 50% level. Looking at the cases above in **Figure 2-16, Figure 2-17, Figure 2-20** and **Figure 2-21** where the phase space curve doesn't fold over on itself, 60/LNR is well within 50% of the standard deviation from the simulation; however, in pathological cases such as **Figure 2-22**, the SD is well outside the 50% range, and a cloud of the same LNR at 60 nm for this parcel would presumably have an even higher SD.

2.6 Discussion and Conclusions

The general idea behind the CIPS retrieval is to use the difference in shape between a cloud signal and the Rayleigh background measured over a range of scattering angles to separate the cloud signal from the total signal. The cloud signal is then compared with model scattering phase functions of different mode radii to determine the best fit particle size of the cloud. Since a cloud signal is very asymmetric for typical particle sizes in PMC, and the magnitude of this asymmetry is a strong function of mode radius, it would seem that this retrieval would have significant leverage on the radius; however, the background signal and cloud signal are not orthogonal, and their overlap is such that the bulk of the difference between clouds of different radii is indistinguishable from a difference in background parameters. This leaves a relatively small component of the phase function differences left over from which to determine radius, but even with this relatively small leverage, CIPS noise performance allows for precise radius determination in most cases given a bright enough cloud or enough pixel binning.

Decomposing the CIPS scattering profile into orthogonal components based on the desired retrieval parameters reduces the dimension and number of free parameters of the fitting problem significantly. Looking at the full scattering profile, each of the n, typically around seven, observations is a function of C, σ, A and r. Transforming into the new basis reduces the dimension of the observation to five from n where two of them depend on all four parameters and three depend only on A and r. Under bright background conditions the background components provide no information on the cloud, so the problem further reduces to three observed components depending on A and r. From **Equation 9**, the radius can be determined from unit magnitude cloud subspace vector. This reduces the relevant dimensions of the observation to only two as a function of r. So the fitting problem has been reduced from looking

at approximately seven dimensions with four free parameters to looking at only two dimensions with one free parameter.

This reduction makes analysis of the retrieval much more intuitive and predictable. The effect of errors is simple to visualize in terms of the radius curve in the two dimensional phase space. For random errors based on measurement noise, as were analyzed in this paper, the fact that the transformation is linear preserves the Gaussian distribution of the noise in the phase space. This allows us to visualize the errors with a familiar Gaussian distribution centered at the true answer in the phase space. Systematic errors show up as displacements in the phase space and so the effect will also be straight forward to interpret.

It has been shown that an improved understanding of the retrieval uncertainties can be obtained by a more thorough analysis of the phase space curve and noise distribution. It is also possible in principle to calculate the retrieval distribution given the phase space curve and noise distribution in the phase space, but this may be numerically intensive. For general retrievals this may be too cumbersome and/or time consuming in comparison to the LNR approach since the CIPS data volume is very large, but it may be useful when analyzing case studies such as coincidences with other instruments. These include common volume observations taken routinely with SOFIE [*Bailey et al.*; this issue] and LIDAR coincidence measurements [*Baumgarten et al.*; 2012]. Since these data sets are relatively small, it may be practical to look at the errors in more detail.

There are important sources of error in addition to noise in the CIPS retrieval. The most important of these is the assumption that the C-σ model is an accurate representation of the Rayleigh background. Small calibration uncertainties can also systematically distort the signal

such that the Rayleigh background looks different than what is physically expected. These two effects combined result in the *C-σ* model being systematically different from the observed background by about 2%. Although 2% may seem small, this 2% projects into the cloud subspace which is significantly dimmer than the background space and can produce a significant bias on the cloud retrievals if not accounted for. This error is parameterized and removed from the CIPS data using the "error maps" in version 4 of the operational retrieval [*Lumpe et al.*; this issue]. This change should provide a significant improvement over previous versions of the retrieval algorithms. Work has been done to analyze the effect of this error using the component framework developed in this paper, and it is intended for this work to be published as part of a future paper.

As is the case in all PMC measurements, CIPS retrievals make assumptions about the particle size distribution. The strength of CIPS is its horizontal resolution and global coverage, but the measurement doesn't contain the leverage necessary to determine all of the details of the particle size distribution. Parameters such as the Gaussian shape, axial ratio and distribution width need to be assumed based on other measurements [*Baumgarten et al.*; 2002, *Lumpe et al.*; this issue] and theoretical work [*Rapp and Thomas*; 2006]. Changes in these assumptions will therefore change CIPS retrievals. Some of these issues are explored by *Bailey et al.* [this issue].

2.7 References

Bailey, Scott M., Gary E. Thomas, David W. Rusch, Aimee W. Merkel, Christian D. Jeppesen, Justin N. Carstens, Cora E. Randall, William E. McClintock, James M. Russell III (2009), Phase functions of polar mesospheric cloud ice as observed by the CIPS instrument on the AIM satellite, *Journal of Atmospheric and Solar-Terrestrial Physics*, Volume 71, Issues 3–4, Pages 373-380, ISSN 1364-6826, 10.1016/j.jastp.2008.09.039.

Bhartia, P. K., R. D. McPeters, C. L. Mateer, L. E. Flynn, C. Wellemeyer, Algorithm for the estimation for vertical ozone profiles from the backscattered ultraviolet technique, *Journal of Geophysical Research*, Vol. 101, No. D13, Pages 18,793-18,806, August 20, 1996.

Baumgarten, G., K. H. Fricke, and G. von Cossart (2002), Investigation of the shape of noctilucent cloud particles by polarization lidar technique, *Geophys. Res. Lett.*, 29(13), 1630, doi:10.1029/2001GL013877.

Baumgarten, G., A. Chandran, J. Fiedler, P. Hoffmann, N. Kaifler, J. Lumpe, A. Merkel, C. E. Randall, D. Rusch, and G. Thomas (2012), On the horizontal and temporal structure of noctilucent clouds as observed by satellite and lidar at ALOMAR (69N), *Geophys. Res. Lett.*, 39, L01803, doi:10.1029/2011GL049935.

Chandran, A., D.W. Rusch, S.E. Palo, G.E. Thomas, M.J. Taylor, Gravity wave observations in the summertime polar mesosphere from the Cloud Imaging and Particle Size (CIPS) experiment on the AIM spacecraft, Journal of Atmospheric and Solar-Terrestrial Physics, Volume 71, Issues 3–4, March 2009, Pages 392-400, ISSN 1364-6826, 10.1016/j.jastp.2008.09.041.

Chandran, A., D. W. Rusch, A. W. Merkel, S. E. Palo, G. E. Thomas, M. J. Taylor, S. M. Bailey, and J. M. Russell III (2010), Polar mesospheric cloud structures observed from the cloud imaging and particle size experiment on the Aeronomy of Ice in the Mesosphere spacecraft: Atmospheric gravity waves as drivers for longitudinal variability in polar mesospheric cloud occurrence, *J. Geophys. Res.*, 115, D13102, doi:10.1029/2009JD013185.

Chapman S., The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating Earth, Part II. Grazing Incidence, *Proc. Phys. Soc.*, 43, pp. 483-501, 1931.

Huestis, David L., Accurate evaluation of the Chapman function for atmospheric attenuation, *Journal of Quantitative Spectroscopy and Radiative Transfer*, Volume 69, Issue 6, 15 June 2001, Pages 709-721, ISSN 0022-4073, 10.1016/S0022-4073(00)00107-2.

McClintock, W.E., D.W. Rusch, G.E. Thomas, A.W. Merkel, M.R. Lankton, V.A. Drake, S.M. Bailey, J.M. Russell III (2009), The cloud imaging and particle size experiment on the Aeronomy of Ice in the mesosphere mission: Instrument concept, design, calibration, and on-orbit performance, *Journal of Atmospheric and Solar-Terrestrial Physics*, Volume 71, Issues 3–4, Pages 340-355, ISSN 1364-6826, 10.1016/j.jastp.2008.10.011.

Mcpeters, R. D. (1980), The Behavior of Ozone Near the Stratopause From Two Years of BUV Observations, *J. Geophys. Res.*, 85(C8), 4545–4550, doi:10.1029/JC085iC08p04545.

Mishchenko, M. I. and L. D. Travis, Capabilities and limitations of a current Fortran implementation of the T-matrix method for randomly oriented, rotationally symmetric scatterers, *J. Quant. Spectrosc. Radiat. Transfer*, Vol. 60, No. 3, pp. 309-324, 1998.

Rapp, M., G. E. Thomas, Modeling the microphysics of mesospheric ice particles: Assessment of current capabilities and basic sensitivities, *Journal of Atmospheric and Solar-Terrestrial Physics*, Volume 68, Pages 715-744, 2006.

Rusch, D.W., G.E. Thomas, W. McClintock, A.W. Merkel, S.M. Bailey, J.M. Russell III, C.E. Randall, C. Jeppesen, M. Callan (2009), The cloud imaging and particle size experiment on the aeronomy of ice in the mesosphere mission: Cloud morphology for the northern

2007 season, *Journal of Atmospheric and Solar-Terrestrial Physics*, Volume 71, Issues 3–4, March 2009, Pages 356-364, ISSN 1364-6826, 10.1016/j.jastp.2008.11.005.

Russell, James M. III, Scott M. Bailey, Larry L. Gordley, David W. Rusch, Mihály Horányi, Mark E. Hervig, Gary E. Thomas, Cora E. Randall, David E. Siskind, Michael H. Stevens, Michael E. Summers, Michael J. Taylor, Christoph R. Englert, Patrick J. Espy, William E. McClintock, Aimee W. Merkel (2009), The Aeronomy of Ice in the Mesosphere (AIM) mission: Overview and early science results, *Journal of Atmospheric and Solar-Terrestrial Physics*, Volume 71, Issues 3–4, March 2009, Pages 289-299, ISSN 1364-6826, 10.1016/j.jastp.2008.08.011.

Taylor, M.J., Pautet, P.-D., Zhao, Y., Randall, C.E., Lumpe, J., Bailey, S.M., Carstens, J., Nielsen, K., Russell III, J.M., Stegman, J., 2011. High-latitude gravity wave measurements in noctilucent clouds and polar mesospheric clouds. Aeronomy of the Earth's Atmosphere and Ionosphere, IAGA Special Sopron Book Series, vol. 2, part 1, 93-105, DOI 10.1007/978-94-007-0326-1_7.

3. Space based gravity wave imaging near the stratopause from the cloud imaging and particle size instrument

3.0 Abstract

The Cloud Imaging and Particle Size (CIPS) instrument has been analyzed for the presence of Gravity Waves (GW) in the ozone signal near the stratopause. The CIPS instrument is onboard the Aeronomy of Ice in the Mesosphere (AIM) satellite. Observations are made in the spring and summer hemisphere at latitudes poleward of approximately 35 degrees. This is a layer which is not presently imaged for gravity waves with the coverage provided by CIPS.

For the spring through summer period observed, the activity is the strongest in the early spring. Waves are observed which correlate with thunderstorm activity as well as known orographic sources such as the Andes and Antarctic Peninsula. Future work on this data set promises to provide vertical wavelength by using overlapping orbits at polar latitudes or coincident observations with other instruments such as AIRS.

3.1 Introduction

Gravity waves are a mesoscale phenomenon in which the buoyancy of air parcels in a stably stratified atmosphere is the restoring force for a disturbance which propagates information from one region of the atmosphere to another. Of particular importance is the transfer of energy and momentum from relatively energetic disturbances in the troposphere to the relatively rarefied middle and upper atmosphere. Gravity wave drag significantly alters the circulation of the middle atmosphere [*Lindzen*; 1981], thus a complete picture of the gravity wave spectrum, amplitude, geographical/seasonal dependence and intermittency is necessary for Global Circulation Models (GCM) to accurately reproduce and predict the behavior of the middle

atmosphere. Since GCM typically do not have the resolution necessary to resolve gravity waves, they are handled by parameterizations which are at present insufficiently constrained by observational data. These parameterizations are large source of uncertainty [*Fritts and Alexander*; 2003, *Alexander and Barnet*; 2007].

The Cloud imaging and Particle Size (CIPS) experiment on the Aeronomy of Ice in the Mesosphere (AIM) satellite has recently been used to develop a lower mesospheric Rayleigh scattering albedo data set. The albedo is controlled by the column density of ozone near 50 km. An abundance of clear gravity wave signatures can be seen in the data. The CIPS instrument was designed for observation of Polar Mesospheric Clouds (PMC), and it has been used in the observation of gravity waves in these clouds which occur at 83 km [*Chandran et al.*; 2009, *Taylor et al.*; 2011]. The coverage provided is in the spring and summer hemisphere for latitudes poleward of around 30 degrees since this is the time and region needed for PMC observation. The data set currently gives relative ozone column density variations near an altitude of 50 km with a horizontal resolution of 20 km x 20 km. This extends the observational coverage of gravity waves [*Alexander and Barnet*; 2007] up to the stratopause and lower mesosphere which should allow the CIPS data set to aid significantly in the constraint of gravity wave parameterizations needed by the GCMs.

3.2 Instrument and Retrieval

3.2.1 The CIPS instrument

The CIPS instrument is a 4-camera UV imager which observes at 265 nm. Images are taken in the nadir with a field of view of approximately 2000 km along track by 1000 km across track [*McClintock et al.*; 2008]. The AIM satellite is in a sun synchronous orbit with a noon local

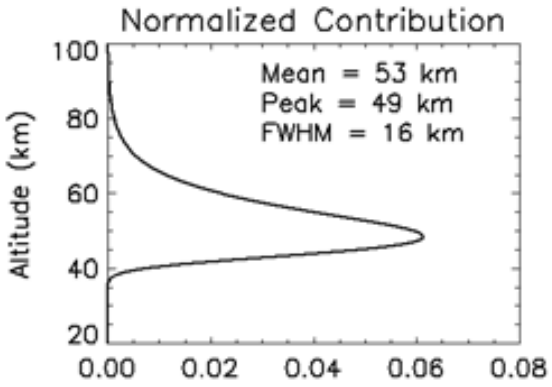

Figure 3-1 The normalized contribution to a CIPS observation with altitude. The contribution peaks at 49 km with a full width half maximum of 16 km.

time equator crossing. The cadence in which the images are taken is such that an air parcel is observed in 7 consecutive images as CIPS passes over head. The nadir pixel resolution is 2x1 km increasing to approximately 10x5 km to the forward and aft; however, for signal to noise reasons ozone retrievals are currently binned to 20x20 km.

In the absence of PMC, which the instrument was designed to observe [*Rusch et al.*; 2009], the observed signal is Rayleigh scattered sunlight. Through absorption, ozone abundance near the stratopause controls the optical depth and therefore the albedo of the CIPS signal. **Figure 3-1** shows an example contribution function for CIPS. The contribution to the signal has a mean altitude of 53 km with a full width half maximum of 16 km. This will create an observational filter [*Alexander et al.*; 1998] for gravity waves with vertical wavelengths less than 16 km.

3.2.2 The Ozone Retrieval

Since each air parcel is observed by CIPS multiple times, combining them provides signal to noise improvement; however, since each observation of an air parcel is done under

different observation geometries and therefore yield different albedos, they cannot simply be averaged. Instead, a simple analytical model is used,

$$A = \frac{P_{Ray}(\Theta)\Gamma(\sigma+1)\beta N(z_0)}{\mu_V\left(\frac{1}{\mu_V}+\frac{1}{\mu_Z}\right)^{\sigma}\alpha^{\sigma}C(z_0)^{\sigma}} \qquad (1)$$

This "*C*-σ" model (Equation 1) is an analytical Rayleigh scattering model where *A* is the Rayleigh albedo, $C(z_0)$ is the ozone column density above a reference altitude, σ is the ratio of the ozone scale height to the atmospheric scale height, $P_{Ray}(\Theta)$ is the Rayleigh scattering phase function, Γ is the Gamma function, β is the atmospheric scattering cross section, α is the ozone absorption cross section, $N(z_0)$ is the atmospheric column density above a reference altitude and μ_v and μ_z are the slant path geometry factors for sun to the scattering point and the scattering point to the satellite respectively. The model was modified from *McPeters* [1980] where it was used for nadir viewing geometry in *Bailey et al.* [2008] for use in CIPS PMC retrievals.

The CIPS retrieved *C* and σ parameters are fit to the observations taken in a given air parcel. These two parameters are combined to produce an inferred nadir albedo. This is the Rayleigh phase adjusted albedo one would observe viewing the parcel in the nadir. Under these conditions $P_{Ray}(\Theta) = 1$ and $\mu_v = 1$. The "ozone" product is then,

$$A_{nadir} = \frac{\Gamma(\sigma+1)\beta N(z_0)}{\left(1+\frac{1}{\mu_Z}\right)^{\sigma}\alpha^{\sigma}C(z_0)^{\sigma}} \qquad (2)$$

Since the σ value is typically very close to 0.5, a positive (negative) variation of 1% in A_{nadir} can be interpreted as an approximately negative (positive) variation of 2% in ozone column density in the 40 to 60 km region. This relation remains approximately linear over the range of values observed.

While the A_{nadir} value produces a smooth low noise product for structure analysis, the presence of the solar zenith angle dependence of μ_z produces gradients in the data along the orbit

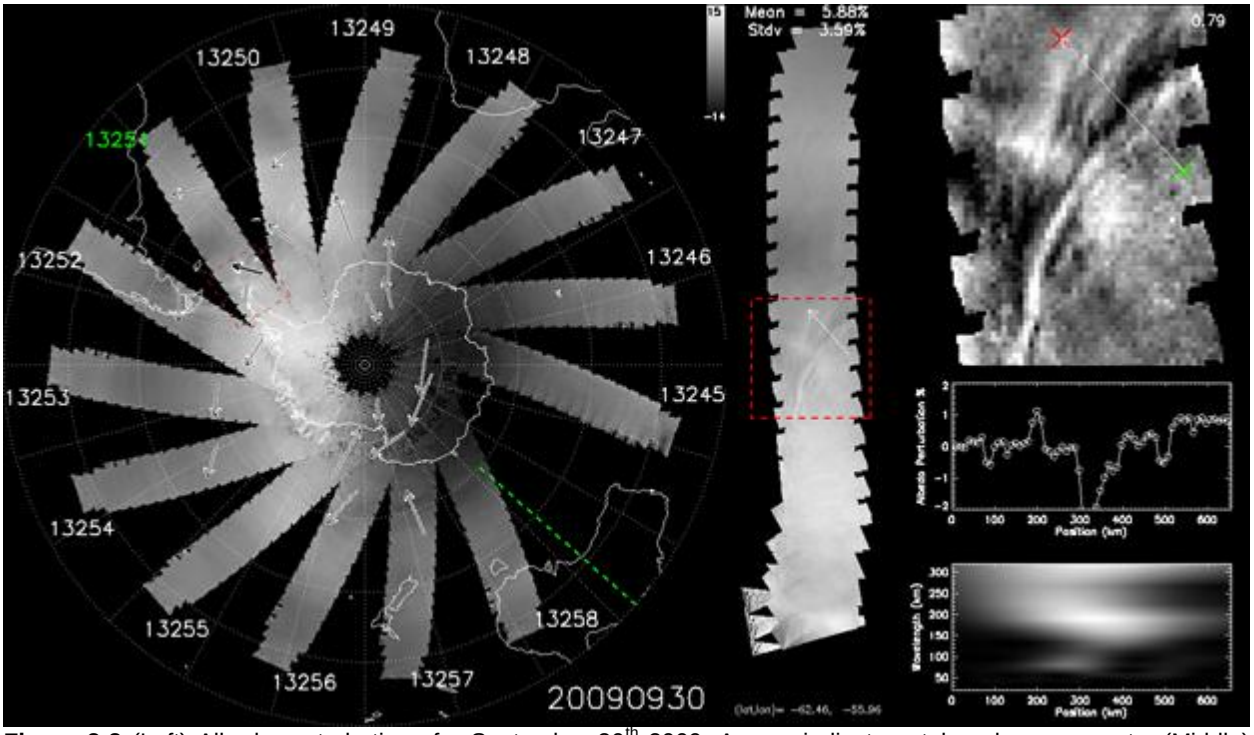

Figure 3-2 (Left) Albedo perturbations for September 30th 2009. Arrows indicate cataloged wave events. (Middle) Selected orbit strip indicated by green orbit number label on left. (Right top) Zoom in on event marked by red boxes with best fit quadratic surface removed. The bolded arrow within the red box marks the selected event on left plot. (Right bottom) Line sample of event and wavelet transform of the sample.

track which are much larger than the geophysical structures. To account for this, a 1-day running mean at each solar zenith angle is calculated. The final albedo perturbation product is the percentage difference between the observed albedo and the running mean albedo. An example of this data product is shown in **Figure 3-2**. The middle image is the selected orbit strip indicated by the orbit number with the green text on the left image. All the orbits within half a day of the selected orbits are shown in the left image, so these are the orbits included in the running mean for the selected orbit.

3.3 Observations

CIPS observes a given hemisphere from approximately the spring equinox to the fall equinox. Wave events can be observed throughout this period, but the occurance frequency and

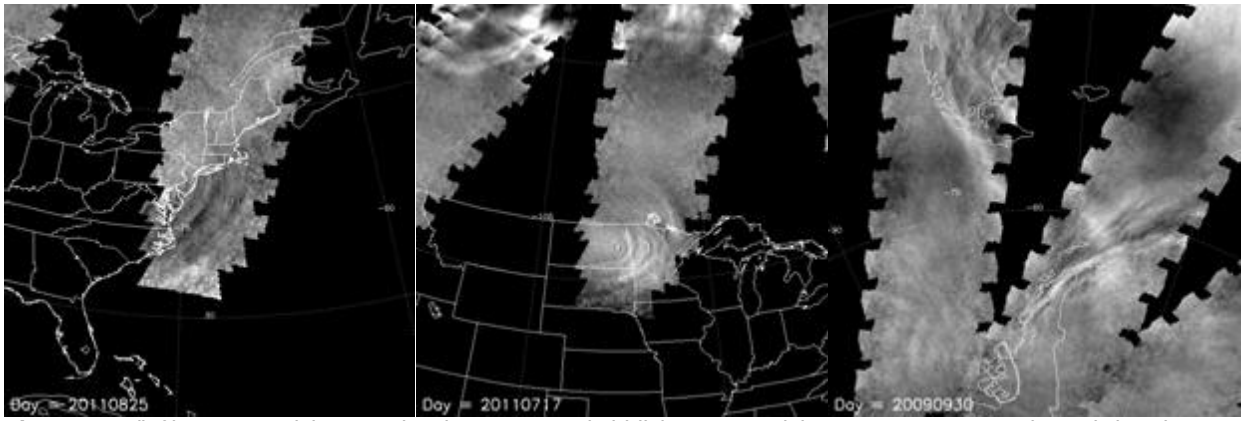

Figure 3-3 (left) Wave activity over hurricane Irene. (middle) Wave activity over strong tornado and thunderstorm activity in North Dakota. The features to the north are PMCs. (right) Mountain wave activity over the Andes and the Antarctic Peninsula.

the observed magnitude are much higher in the few weeks following spring equinox. The example given in **Figure 3-2** is typical of this period. During this period multiple wave signatures are seen per day and often multiple per orbit are observed. For the months following this period the wave activity drops off significantly. On many days there are no obvious wave signatures; however, the northern hemisphere seems to have more occurances of waves outside this early spring wave season. This is likely due to thunderstorm activity. An example of a likely thunderstorm related wave can be seen in **Figure 3-3**. The synoptic scale variability follows a similar seasonal pattern. With large scale albedo variability on the order of that seen in **Figure 3-2** (~10%) being typical for early spring. By mid-spring the synoptic scale variability in the albedo perturbation drops to approximately the 2% level. It remains at this level through the fall equinox when CIPS observations switch to the opposite hemisphere.

Since CIPS observes the whole polar cap, wave observations can be put into global scale prospective. Each arrow in the left image of **Figure 3-2** is manually flagged wave event for that day. Where possible the arrow was directed away from any apparent center of curvature. This can be seen in the middle and upper right images; however, the waves are not always noticably curved, so in some cases the direction is arbitrary. Near by wave events definitely seem to be

linked. For example on orbits 13251-13251, all the events are co-aligned. These could be all part of an extended wave field created by orographic sources in the Andes and Antartic peninsula, or perhaps the pattern occurs due to a common wind field. In any case, there is certainly some degree of correlation between gravity wave events on synoptic scales. The selected wave shown in the orbit strip in the middle and the zoomed image at the top right is the largest amplitude event seen on the day shown. Below the zoomed image is a line sample of the wave along the arrow shown in the images. It has a minimum to maximum amplitude of around 3% which implies a change in ozone column density of about 6%. The scale size of this feature is around 190 km as can be seen from the wavelet on the bottom right. This wave is on the high end of what is typically observed in amplitude and fairly typical in terms of wavelength.

Waves observed by CIPS appear to eminate from both weather based events such as thunderstroms and orographic sources such as mountains. **Figure 3-3** shows some examples of each. The waves observed in middle image of **Figure 3-3** coincided with strong thunderstrom and tornado activity in North Dakota on July 17th 2011. This event has a particularly interesting structure. The horizontal wavelength is much shorter and the radius of the concentric circular structures is much tighter than is typically observed. Waves have also been observed coinciding with other storm events such as hurricane Irene shown in the left image in **Figure 3-3**. The right image in **Figure 3-3** depicts waves which appear to be related to the orographic features of the Andes and the Antartic peninsula. The qualitative features of these mountain waves appear similar to those observed by the Atmosphere Infrared Sounder (AIRS) instrument in the same region [*Alexander and Barnet*; 2007] but at lower altitude, and the structure is what is expected based on 3D propagation effects in gravity wave resolving models [*Alexander and Teitelbaum*; 2011, *Sato et al.*; 2011]. In particular when looking at the wave features over the Antartic

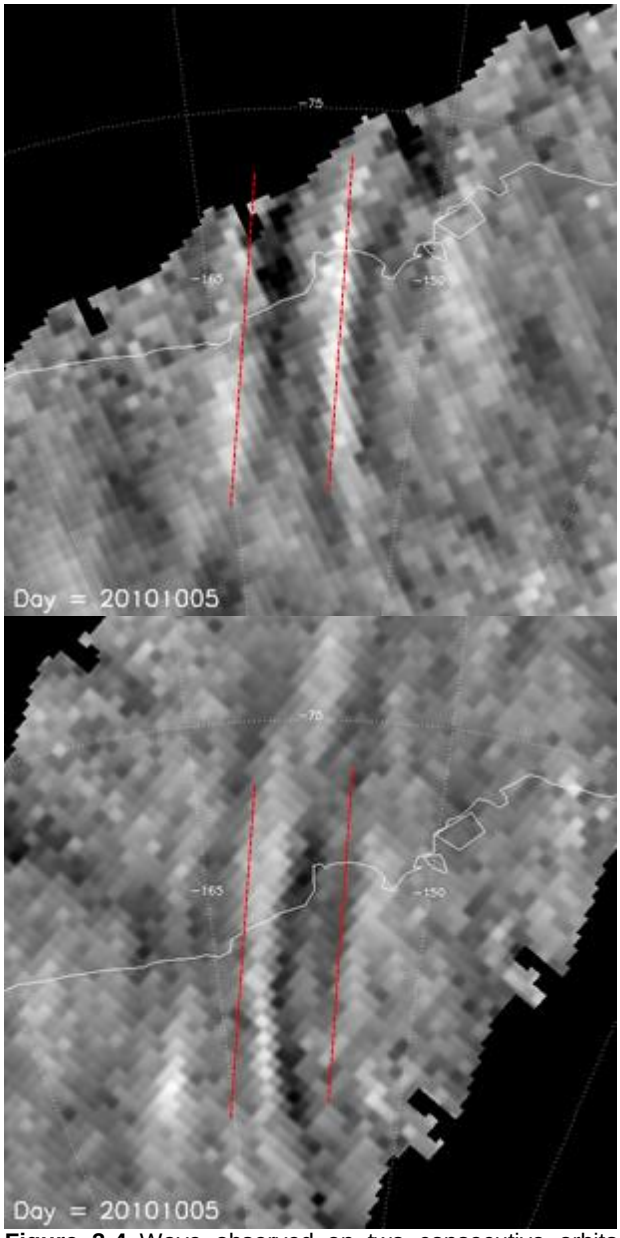

Figure 3-4 Wave observed on two consecutive orbits east of McMurdo station. The red lines are in the same position in both cases to aid the eye in the observation of the phase progression.

peninsula, the wave features follow the coastline, but also extend out beyond the end of the paninsula. This behavior is expected due to leeward propogation induced by meridional gradients in the zonal wind field.

For latitudes poleward of approximately 70°, the field of view for CIPS overlaps from one orbit to the next. This allows for the observation of the same wave field on multiple orbits.

Shown in **Figure 3-4** is an example of such an observation east of McMurdo Station in Antartica. The red lines are aligned with the wave fronts in the first observation. Over the approximately 90 minutes between the orbits, the eulerian phase progression is approximately ¾ of the 150 km wavelength. With additional data on the winds from another source such as Met Office Stratospheric Assimilated Data or European Centre for Medium-Range Weather Forecasts (ECMWF), one can derive the intrinsic frequency, and from that, the vertical wavelength can be estimated. This would give all three components of the wavevector for these overlaping observations.

3.4 Discussion

The CIPS Rayleigh scattering data set is rich with gravity wave signatures that offer the potential for a wealth of results regarding sources, latitudinal redistribution, intermittency, and momentum flux in the stratopause region. A full realization of this potential requires a quantitative understanding of the physical mechanisms by which the gravity waves perturb the ozone and of the observational effects a wave induced structure in the ozone has on the CIPS Rayleigh scattering observation. Gravity waves can perturb the ozone in two ways. The ozone can be perturbed due to dynamical density perturbations and/or photochemical perturbations [*Eckermann et al.*; 1998]. The photochemical perturbations are due in large part to the change in temperature induced by the wave. Initial work suggests that the dynamical response should dominate at these altitudes [*Eckermann*; private communication], but more work needs to be done to confirm these conclusions. Additionally, observational filtering effects need to be further analyzed. As mentioned earlier, the altitude contribution function will filter out waves with short vertical wavelength. Three dimensional viewing effects such as the line of sight's orientation

with respect to the wave fronts will also play a role. All of these effects, physical and observational, need to be combined to produce a transfer function which maps CIPS observations to gravity wave parameters, and the observational filter this transfer function produces needs to be considered in any conclusions made from CIPS data.

Once the relationship between CIPS observations and gravity waves is determined, CIPS stratopause data can be used with GW imaging data taken at lower altitudes such as AIRS [*Hoffmann and Alexander*; 2010] to estimate vertical wavelength. At latitudes poleward of approximately 70 degrees, the neighboring orbits have overlapping fields of view (see left image in **Figure 3-2**). This overlap allows for observation of the same wave event on two or more orbits approximately 90 minutes apart. When combined with wind data, this provides another method for estimation of vertical wavelength. Having the amplitude and all three components of the wavelength will allow for estimation of momentum flux. Knowledge of the momentum flux is important because it is what forces the general circulation.

The CIPS data set opens up a new layer of the atmosphere to planetary scale GW imaging which increases the observational coverage of the gravity wave spectrum. Knowledge of the altitude dependence of the spectrum is important to our understanding of the middle atmosphere's coupling to the troposphere. One reason for this is gravity wave refraction which causes a latitudinal redistribution of the momentum flux as the waves propagate to higher altitudes [*Dunkerton;* 1984, *Sato et al.*; 2009]. Observational evidence of this would be helpful because the GW parameterizations presently in GCM do not account for it. The leeward propagation from mountain waves is one indication of this wave refraction that can already be seen in the data, but the real strength is the fact that CIPS adds a new altitude to the coverage.

Comparisons with measurements at lower altitudes can give a more direct picture of this latitudinal redistribution with altitude.

3.5 References

Alexander, M. J. (1998), Interpretations of observed climatological patterns in stratospheric gravity wave variance, *J. Geophys. Res.*, 103(D8), 8627–8640, doi:10.1029/97JD03325.

Alexander, M. Joan, Christopher Barnet, (2007), Using Satellite Observations to Constrain Parameterizations of Gravity Wave Effects for Global Models. *J. Atmos. Sci.*, 64, 1652–1665.

Alexander, M. Joan, Hector Teitelbaum (2011), Three-dimensional properties of the Andes mountain waves observed by satellite: A case study, *J. Geophys. Res.*, 116, D23110.

Bailey, Scott M., et al. (2009), Phase functions of polar mesospheric cloud ice as observed by the CIPS instrument on the AIM satellite, *J. Atmos. Terr. Phys.*, 71, 373-380.

Chandran, A., D.W. Rusch, S.E. Palo, G.E. Thomas, M.J. Taylor (2009), Gravity wave observations in the summertime polar mesosphere from the Cloud Imaging and Particle Size (CIPS) experiment on the AIM spacecraft, Journal of Atmospheric and Solar-Terrestrial Physics, Volume 71, 392-400.

Dunkerton, Timothy J., 1984: Inertia–Gravity Waves in the Stratosphere. *J. Atmos. Sci.*, 41, 3396–3404.

Eckermann, Stephen D., Dorthy E. Gibson-Wilde, Julio T. Backmeister (1998), Gravity Wave Perturbations of Minor Constituents: A Parcel Advection Methodology, *J. Atmos. Sci.*, 55, 3521-3539.

Fritts, D. C., and M. J. Alexander (2003), Gravity wave dynamics and effects in the middle atmosphere, *Rev. Geophys*., 41(1), 1003.

Hoffmann, L., and M. J. Alexander (2010), Occurrence frequency of convective gravity waves during the North American thunderstorm season, *J. Geophys. Res.*, 115.

Lindzen, R. S. (1981), Turbulence and Stress Owing to Gravity Wave and Tidal Breakdown, J. Geophys. Res., 86(C10), 9707–9714.

McClintock, et al. (2009), The cloud imaging and particle size experiment on the Aeronomy of Ice in the mesosphere mission: Instrument concept, design, calibration, and on-orbit performance, *J. Atmos. Terr. Phys.*, Vol. 71, Is. 3–4, Pages 340-355.

Mcpeters, R. D. (1980), The Behavior of Ozone Near the Stratopause From Two Years of BUV Observations, *J. Geophys. Res.*, 85(C8), 4545–4550

Rusch, D.W., et al. (2009), The cloud imaging and particle size experiment on the aeronomy of ice in the mesosphere mission: Cloud morphology for the northern 2007 season, *J. Atmos. Terr. Phys.*, Volume 71, Issues 3–4, Pages 356-364.

Sato, K., et al. (2009), On the origins of mesospheric gravity waves, *Geophys. Res. Lett.*, 36, L19801.

Sato, K., et al. (2011), Gravity Wave Characteristics in the Southern Hemisphere Revealed by a High-Resolution Middle-Atmosphere General Circulation Model, *J. Atmos. Sci.*, 69, 1378-1396.

Taylor, M.J., Pautet, P.-D., Zhao, Y., Randall, C.E., Lumpe, J., Bailey, S.M., Carstens, J., Nielsen, K., Russell III, J.M., Stegman, J., (2011). High-latitude gravity wave measurements in noctilucent clouds and polar mesospheric clouds. Aeronomy of the

Earth's Atmosphere and Ionosphere, IAGA Special Sopron Book Series, vol. 2, part 1, 93-105.

4. Conclusions

Gravity waves play a crucial role in the middle atmosphere. Without an understanding of gravity wave drag, general circulation and temperatures cannot be reproduced [*Lindzen*; 1981]. The CIPS instrument is unique in that it allows for the retrieval of PMC particle size and albedo at high horizontal resolution. The resolution allows for the analysis of gravity wave structures in the clouds, but it also allows for observation of wave structures in the Rayleigh scattering signal. This makes CIPS observations very useful since observations of gravity waves in both of these regions are sparse.

The general idea behind the CIPS retrieval is to use the difference in shape between a cloud signal and the Rayleigh background measured over a range of scattering angles to separate the cloud signal from the total signal. The cloud signal is then compared with model scattering phase functions of different mode radii to determine the best fit particle size of the cloud. A cloud signal is very asymmetric for typical particle sizes in PMC, and the magnitude of this asymmetry is a strong function of mode radius. It was assumed that this would provide the retrieval significant leverage on the radius. Unfortunately, the background signal and cloud signal are not orthogonal as is shown in 2.3.3, and their overlap is such that the bulk of the difference between clouds of different radii is indistinguishable from a difference in background parameters. Further, the difference is not even in the form of an asymmetry that favors forward scattering for large particles as had been thought before. This leaves a relatively small component of the phase function differences left over from which to determine radius. Fortunately even with this relatively small leverage, CIPS noise performance allows for precise radius determination in most cases given a bright enough cloud or enough pixel binning. The most recent CIPS retrieval algorithms include such binning.

Decomposing the CIPS scattering profile into orthogonal components based on the desired retrieval parameters reduces the dimension and number of free parameters of the fitting problem significantly. This simplifies the problem greatly. Looking at the full scattering profile, each of the n, typically around seven, observations is a function of C, σ, A and r. That is seven observations being fit to as a function of four variables. The results of the component decomposition analysis reduce the relevant dimension of the observation to two components as a function of r. Thus the fitting problem has been reduced from looking at approximately seven dimensions with four free parameters to looking at only two dimensions with one free parameter. This reduction makes analysis of the retrieval much more intuitive and predictable. The effect of errors is simple to visualize in terms of the radius curve in the two dimensional phase space. For random errors based on measurement noise, as were analyzed in Chapter 2, the fact that the transformation is linear preserves the Gaussian distribution of the noise in the phase space. This allows us to visualize the errors with a familiar Gaussian distribution centered at the true answer in the phase space. Diagnosing the behavior of the errors in a case like **Figure 2-22** where the phase space curve folds back over on itself would be very difficult without this simplification. The cloud phase functions for the two solutions look very different from one another, so it would be hard to predict that the retrieval would have a difficult time distinguishing the two solutions. The operational algorithms have always had problems in this region of the orbit strip, but prior to these results, it never made any sense.

The manuscript in Chapter 2 covers the component decomposition approach and investigates random errors using this framework; however, systematic errors have also been a significant source of error in CIPS retrieval algorithms. The author has done a great deal of work investigating these errors, and the findings have helped to motivate many of the changes that

have been implemented in the development of the operational retrieval algorithm. The main problem that was faced with respect to systematic errors is that under cloud free conditions the best fit C-σ model systematically differs from the observed scattering profile. In particular the forward scattering points are observed to be brighter than the best fit and the backward scattering points are observed to be dimmer. The scale of this systematic difference is on the order of 2% of the background signal. This may seem small, but 2% of a 300 G background is 6 G, and this is quite significant in comparison to the cloud component of the signal. To make things worse, almost all of this error signal projects into the cloud subspace, so within the radius phase space this error results in a significant artificial displacement of the observation vector and therefor biases the radius retrievals.

One of the reasons for this systematic error is calibration related. The "flat field" which accounts for relative sensitivity between pixels in the camera array had biases in it. Since each observation in a scattering profile comes from different regions of the camera, errors in the flat field artificially warp the background signal. Part of the reason for this bias was due to the way in which the flat fields were calculated. The flat fields were calculated by assuming constant ozone parameters across the field of view. If there was in reality a systematic gradient across the CIPS field of view, it would bias the flat field measurement. To remedy this, the author developed a method for measure the flat fields in a way which didn't require assumptions about the ozone parameters. This corrected a significant fraction of the systematic error. Another part of the systematic error is that the C-σ model is only an approximation to the Rayleigh scattering signal. To allow for analytical evaluation of the radiative transfer integral, several approximations where made. These approximations are fairly accurate most of the time, but at the high solar zenith angles the error can become significant. There is also evidence which

indicates some sort of memory effect and/or an uncorrected temperature dependence in the cameras. All of these errors act to produce a difference between the observed background and the C-σ model; however, these differences are very systematic and can be reliably parameterized. In the latest version of the operational retrieval algorithm, this error is corrected using the "error maps" as described in *Lumpe et al.* [2012].

It is the work on errors in the PMC retrievals that led to the development of the Rayleigh scattering gravity wave data set. The systematic errors presented an even bigger problem for observation of the gravity waves than they did for the clouds because the systematic artifacts that they produced in the images where much larger than the geophysical structures produced by the gravity waves. To analyze the impact of noise on the PMC retrievals, the change described in 2.5.1 where the CIPS pixels are binned into the Lambert projection grid instead of being interpolated was also necessary for the observation of waves. The noise level in the CIPS images is too high to observe the wave structures. So the noise reduction afforded by binning instead of interpolating was crucial. The only major change that was needed to observe the Rayleigh structures that was not part of the error investigation of the PMC retrieval was to geo-locate to 50 km instead of 83 km; however, future work towards creating a more quantitative gravity wave retrieval will result in a larger divergence from the PMC algorithm.

The CIPS Rayleigh scattering data set offers many interesting results in its current state; however, without knowledge of the amplitude and vertical wavelength the most important physical quantity of the waves, momentum flux, cannot be calculated. The vertical wavelength can be estimated at high latitudes when a wave is observed on multiple orbits. Derivation of the vertical wavelength via comparisons with AIRS observations and measurement of the amplitude generally requires a quantitative understanding of the physical mechanisms by which the gravity

waves perturb the ozone and of the observational effects a wave induced structure in the ozone has on the CIPS Rayleigh scattering observation. The ozone can be perturbed due to dynamical density perturbations and/or photochemical perturbations [*Eckermann et al.*; 1998]. The photochemical perturbations are due in large part to the change in temperature induced by the wave. Initial work suggests that the dynamical response should dominate at these altitudes [*Eckermann*; private communication], but more work needs to be done to confirm these conclusions. Additionally, observational filtering effects need to be further analyzed. As mentioned in Chapter 3, the altitude contribution function will filter out waves with short vertical wavelength. Three dimensional viewing effects such as the line of sight's orientation with respect to the wave fronts will also play a role. All of these effects, physical and observational, need to be combined to produce a transfer function which maps CIPS observations to gravity wave parameters.

The CIPS data set opens up a new layer of the atmosphere to planetary scale GW imaging which increases the observational coverage of the gravity wave spectrum. Knowledge of the altitude dependence of the spectrum is important to our understanding of the middle atmosphere's coupling to the troposphere. One reason for this is gravity wave refraction which causes a latitudinal redistribution of the momentum flux as the waves propagate to higher altitudes [*Dunkerton;* 1984, *Sato et al.*; 2009]. Observational evidence of this would be helpful because the GW parameterizations presently in GCM do not account for it. The leeward propagation from mountain waves is one indication of this wave refraction that can already be seen in the data, but the real strength is the fact that CIPS adds a new altitude to the coverage. Comparisons with measurements at lower altitudes can give a more direct picture of this latitudinal redistribution with altitude.

5. References for Introduction and Conclusions

Dunkerton, Timothy J., 1984: Inertia–Gravity Waves in the Stratosphere. *J. Atmos. Sci.*, 41, 3396–3404.

Eckermann, Stephen D., Dorthy E. Gibson-Wilde, Julio T. Backmeister (1998), Gravity Wave Perturbations of Minor Constituents: A Parcel Advection Methodology, *J. Atmos. Sci.*, 55, 3521-3539.

Fritts, D. C., and M. J. Alexander (2003), Gravity wave dynamics and effects in the middle atmosphere, *Rev. Geophys*., 41(1), 1003.

Lindzen, R. S. (1981), Turbulence and Stress Owing to Gravity Wave and Tidal Breakdown, *J. Geophys. Res.*, 86(C10), 9707–9714.

McIntyre, Michael E., On Dynamics and Transport Near the Polar Mesopause in Summer, *J. Geophys. Res.*, 94 (D12), 14617-14628.

Sato, K., et al. (2009), On the origins of mesospheric gravity waves, *Geophys. Res. Lett.*, 36, L19801.