Inversion of Nighttime Electron Densities Using Satellite Observations

by

Jeffrey David Stalvey

Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Science in Mathematics with Operations Research Option

New Mexico Institute of Mining and Technology
Socorro, New Mexico
December, 2001

ACKNOWLEDGEMENT

I would like to thank Dr. Ron Thomas and the Office of Naval Research for the financial support they provided and for Dr. Thomas's invaluable assistance in this research project.

I would also like to thank my wife for her support and patience during this process.

I would like to give special acknowledgment to my advisor Dr. Brian Borchers for all his help in this project and through out all of my graduate course work as well as the other member of my committee Dr. Anwar Hossain.

This thesis was typeset with LaTeX[1] by the author.

---

[1] LaTeX document preparation system was developed by Leslie Lamport as a special version of Donald Knuth's TeX program for computer typesetting. TeX is a trademark of the American Mathematical Society. The LaTeX macro package for the New Mexico Institute of Mining and Technology thesis format was adapted by Gerald Arnold from the LaTeX macro package for The University of Texas at Austin by Khe-Sing The.

ABSTRACT

This thesis addresses the problem of inverting nighttime electron densities using satellite observations. The LORASS spectrograph on board the ARGOS satellite counts photons that are emitted from radiative recombination and neutralization reactions that occur due to de-ionization in the F region. The forward function parameterizations that will be investigated are the three and four parameter Chapman functions as well as a free from model that interpolates using cubic splines. To obtain a two-dimensional mapping of the ionosphere, a grid of electron densities is calculated using the Chapman function and an array of optimal Chapman parameters. The optimal Chapman parameters come from inverting 911Å and 1356Å data simultaneously. This leads to physically unrealistic mappings of the ionosphere which are corrected using regularization.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis concerns itself with the problem of inferring nighttime electron densities in the F region (altitudes of 150 Km to 750 Km) of the atmosphere. The data has been obtained from the limb scans made by the Low Resolution Airglow and Aurora Spectrograph (LORAAS) instrument on board the Advanced Research and Global Observing Satellite (ARGOS). A fairly general description of the physics and instrumentation used will be given, while most of the material will focus on the research done in trying to find a best fit to the data.

Let us first give a brief history on the development of the theory of using ultraviolet and visible emissions in the upper atmosphere (airglow) to calculate the electron densities in the F region of the nighttime sky. During the day, a variety of molecules are ionized by solar radiation. At night, ions recombine, resulting in the decay of ionization from dusk till dawn. The recombination reactions emit photons with distinct wavelengths. The de-ionization emissions have wavelengths that range from the visible spectrum to the ultraviolet. The chemical reactions from nightglow (nighttime airglow) emissions have been measured in the visible spectrum from the ground [Tin 72, Tin 73] and in the ultraviolet spectrum from space [Hic 70] since the late 1960's and early 1970's. From these measurements it was deduced that two reactions could be responsible for the majority of the emissions. These two reactions are radiative recombination of O^+ ions [Han 69] with electrons and neutralization reactions between O^+ and O^- ions [Knu 70]. Further measurements at mid-latitudes [Bru 78] confirmed what was proposed by Hansen [Han 69], that radiative recombination reactions account for 75% of measured 1356Å intensities and that the majority of the remaining 25% of the 1356Å intensities are from O^+ and O^- neutralization reactions. This 25% number has been modified due to the discoveries by Feldman et al. [Fel 92] who determined in December, 1990 that the 25% contribution from the O^+ and O^- neutralization is really closer to 40%. Thus, the neutralization rate factor needed to be increased by a factor of 1.9.

The rate factor is determined from the electron temperature which Feldman [Fel 01] also inferred by spectrally resolving the shape of the 911Å continuum. This means that if the electron temperature is known and the 1356Å emissions can be measured then we can quantify the electron density in the measured region. We can do this since we know (from the radiative recombination rate coefficient) how many free electrons it would take to create the measured intensity of the 1356Å emissions. The 911Å emissions that have been measured are due entirely to radiative recombination since O^+ and O^- neutralization does not produce 911Å emissions [Tin 72].

The 911Å and 1356Å emissions are also present in the Earth's dayglow spectrum and the mechanisms for radiative recombination are still active. However the 1356Å emission measurement is dominated by photo electric impact excitation (from the sun) of atomic oxygen. The 911Å emission is unaffected by photoelectrons during the daytime [Tin 75] but Gentieu [Gen 79, Gen 81, Gen 84] showed using rocket observations that the 911Å dayglow emissions are contaminated by nearby emissions from nitrogen and oxygen, which are not present at night. Later, Feldman [Fel 01] showed that contamination of O^+ 911Å emissions to be negligible along with Dymond [Dym 01] who was able to closely estimate daytime electron densities using 911Å emissions as compared with the International Reference Ionosphere IRI-90 [Bil 90] at mid-latitudes. In light of these difficulties the 911Å and 1356Å intensities will be used to infer only nighttime, F region, electron densities.

There are other emissions in the far ultraviolet region that need to be considered such as 912Å emissions produced by radiative recombination of protons and electrons. Feldman [Fel 92] was not able to detect this emission however. Also, the H^+/O^+ density ratio can be fairly low in the F region except at low altitudes. For these reasons, the H^+ electron recombination emission is neglected.

Meier [Mei 91] saw the application of ultraviolet remote sensing from spaced-based platforms as a means of mapping the F region electron density globally. Meier's work and that of Tinsley and Bittencourt [Tin 75] and Chandra et al. [Chn 75] motivated the Naval Research Laboratory (NRL) to develop these space-based platforms. One of the first of these is the High Resolution airglow and Aurora Spectroscopy (HIRAAS) experiment. HIRAAS was launched on board the ARGOS satellite and put into sun-synchronous circular orbit at 840 Km altitude on February 23, 1999. LORAAS is one of three spectrographs that make up the HIRAAS experiment. LORAAS is an extreme and far ultraviolet spectrograph operating in the 800Å to 1700Å passband at 17Å spectral resolution. LORAAS simultaneously observes both the 911Å and 1356Å emissions. The observations cover the 100 km to 750 km altitude range with one limb scan occurring every 90 seconds. The latitude spacing of the limb scan is approximately 5.8 deg. The daytime sky also triggers mechanisms for the H^+ electron 1216Å (known as Lyman-α) and O^+ electron 1304Å emissions, so the instruments sensitivity was reduced by leaving the micro-channel plate on the SiC detector without a photo cathode in the region where these lines fall. This helped to limit the detector photon count rate for the spectra during daytime observations. Blocking the Lyman-α also helps the nighttime observations since this spectrum tends to bleed into the nighttime sky near the edges. For more information on the LORAAS instrument see references [McC 92, McC 94, McC 95, Dym 93, Tho 99].

At this point in time, all that was needed was an algorithm that could take the data from the photon counter (LORAAS) and back into a solution that is the electron density that caused the 911Å and 1356Å intensities measured by LORAAS. The algorithm will be discussed in detail in the following chapter but, in short, it is an inversion of the limb scan data using an iterative approach to solving the nonlinear least squares problem of estimating the maximum likelihood parameters to a given forward function that calculates the electron densities. If we equate O^+ density to electron density, the 911Å and 1356Å emissions that LORAAS should see can then be calculated. The maximum likelihood parameters to the forward function that generate data with least square errors (compared to the real data measured on LORAAS) are computed to minimize a χ^2 error value.

There has been quite a lot of research previously done on the data sets obtained from LORAAS. Both one-dimensional and two-dimensional fits have been performed using daytime and nighttime 911Å, 834Å and 1356Å intensities measured by LORAAS. Since this thesis is primarily concerned with the nightglow from electron densities, results will be presented from Dymond et al. [Dym 02, Dym 03] for the one and two dimensional inversions using the 1356Å spectra.

The one dimensional model assumes the ionosphere is spherically symmetric and does not vary across nearby latitudes and longitudes. This assumption is reasonable during the night at mid-latitudes but not at low altitudes near the equator. The two dimensional algorithm takes into account changes in the electron density by setting up a grid that represents a more realistic model of the ionosphere. This is the best model for areas where the gradient of the ion density is high. The results obtained are also compared with Ionosonde measurements. Ionosonde is a ground based radar that is widely accepted as an accurate measurement of the peak electron density and the altitude of the peak electron density directly over the instrument.

An arbitrarily chosen day of limb scan data acquired on November 24, 1999 was used for both the one and two dimensional models. The section of data used (out of the 800 profiles acquired that day by LORAAS) was selected to include nighttime measurements. Also, this data had two measurements that were close to two Ionosondes', the Millstone Hill, MA Ionosonde and the Rome, Italy Ionosonde. These Ionosonde's were the only two that day that had taken reliable measurements themselves, were close enough to ARGOS's (and thus the LORAAS instruments) flight path, and where LORAAS's data was uncontaminated by auroral precipitation. Aurorally charged particle precipitation, unlike the ionization sources already described, is characterized by its storm-like behavior: long-term unpredictability, highly variable strength and spatial inhomogeneity [Ree 89]. Since LORAAS's line of sight through the ionosphere covers 50 deg of latitude, auroral precipitation is sometimes hard to avoid. It shows up in the data as a large quantity of spikes far out of the normal range of intensities for nightglow. If LORAAS is pointing towards a star it can have a similar effect on the data. Spikes in the data from stars are usually very few in number and can easily be removed without a significant loss of information in the data. However, auroral precipitation can sometimes effect the majority of the measurements taken and thereby render the data useless.

There are two ways of spotting auroral precipitation. Examining a plot of the spectra for the presence of N_2 Lyman-Birge-Hopfield bands (1200-1700Å spectra), which are excited by auroral electrons, is one way. The other is to check the rate that LORAAS was rejecting photo events. The LORAAS instrument is capable of rejecting counts that it deems atypical. If the LORAAS rejection rate is greater than 25%, the data is considered contaminated.

The results of the one dimensional model presented by Dymond et al. [Dym 02] follow. First they presented a comparison of their results to ground truth. The Millstone Hill Ionosonde is located at 41 deg N and -71 deg E, while the closest tangent altitude (in LORAAS's line of sight) to this point is 43 deg N and -75 deg E. The reduced χ^2 value obtained from this fit was 0.795 which is well within the 95% confidence level of about 1. The peak electron density from the model was found to be $5.3 \pm .65 \times 10^5\,\mathrm{cm}^{-3}$ at 7:28 UT. The Millstone Hill Ionosonde measured a peak electron density of $5.0 \times 10^5\,\mathrm{cm}^{-3}$ and $3.7 \times 10^5\,\mathrm{cm}^{-3}$ at 7:30 UT and 7:45 UT respectively. This fluctuation is not typical of most nights in November 1999 at Millstone Hill. LORAAS detected auroral precipitation 10 deg N of Millstone Hill at the time it flew over. The higher than average electron densities (average being about $3.0 \times 10^5\,\mathrm{cm}^{-3}$ for November 1999) are attributed to this auroral precipitation. This abnormally high density was detected by LORAAS and the one dimensional model and agrees with the Millstone Hill Ionosonde readings. The height of the peak electron density measured by the Ionosonde was 333 km and 276 km at 7:30 UT and 7:45 UT, respectively. The one dimensional model found the peak height to be 294 ± 16 km at 7:28 UT. Despite the fluctuations due to auroral precipitation, the Ionosonde and the model give very similar results.

The next result is from the Rome, Italy fly-over. The closest tangent point altitude to Rome was located at 44 deg N and 26 deg E. The Rome, Italy Ionosonde is located at 42 deg N and 13 deg E. The retrieved electron densities from the model are in very close agreement with the Ionosonde measurements. The Ionosonde showed the peak electron density to be $1.9 \times 10^5\,\mathrm{cm}^{-3}$ at an altitude of 258 km. The model retrieved values of $2.0 \pm .3 \times 10^5\,\mathrm{cm}^{-3}$ at a height of 278 ± 73 km. The large confidence intervals in the model estimates are due to small signal to noise ratios in the LORAAS measurement. The intensities measured by LORAAS were near it's detection limits. Still, the two forms of measurement are in very close agreement. The χ^2 value for this pass was 0.23, a very good fit.

The results of the two dimensional model, using the same data set as one dimensional model, yielded similar results. The reduced χ^2 value for the Millstone Hill and Rome fly-overs were 0.977 and 0.906 respectively. The peak electron density retrieved from over Millstone Hill, as found by the two dimensional model, was $5.2\times10^5\,\mathrm{cm}^{-3}$ at a height of 297 km. The peak electron density retrieved from over Rome, as found by the two dimensional model, was $1.6 \times 10^5\,\mathrm{cm}^{-3}$ at 300 km. Both of these fits are in agreement with their respective Ionosonde measurements.

These papers conclude that inversion of UV limb scans [Dym 02, Dym 03] produce accurate electron density profiles. The inversion method produces results that are in very close agreement with Ionosonde measurements. The big advantage to using the inversion of limb scan method is that 800 altitude profiles are gathered daily over land and water thus giving an accurate global picture (or weather map) of the ionosphere. The papers state that additional validation of the limb scan inversion method is needed over different times and geomagnetic and solar activity levels. The authors state that, "Clearly, the technique provides an accurate means for globally characterizing the ionospheric state."

Part of the research presented in this thesis will make improvements to the already existing software that will result in faster, more accurate convergence. This will allow us to tighten the stopping condition tolerances to achieve more accurate parameter estimations. Other improvements will be discussed in detail in the next chapter. We will then test the improved algorithm on both the one dimensional and two dimensional models. Tests will be run on the one dimensional model using 911Å emission data and 1356Å emission data simultaneously. We will fit three different parameterizations of the forward model in an attempt to find the best one. Next, we will test the two dimensional model using 911Å and 1356Å emission data simultaneously. Results from the one-dimensional model will be compared against Ionosonde results.

To determine which model is best we will examine the residuals for normality and randomness. Chi square tests will be performed on the results of all parameterizations to judge goodness of fit. F statistic tests will be performed to determine if one parameterization is more effective than another. The F test will be the deciding factor in most cases as to which model is best.

In the two dimensional model we will also look at the reality of the solution. Finding optimal parameters using a two dimensional grid and non-linear regression leads to rough and unreal solutions that have to be smoothed out (regularized). In the chapter on the two-dimensional model we will describe the way in which we smoothed our grid to obtain a more realistic solution. At the end of each chapter the results found will be discussed and conclusions drawn.

Chapter 2

The One Dimensional Model

2.1 The Data

As previously mentioned in chapter one, the data have been obtained from the limb scans made by the Low Resolution Airglow and Aurora Spectrograph (LORAAS) instrument on board the Advanced Research and Global Observing Satellite (ARGOS). An arbitrarily chosen day of limb scan data was acquired on November 24, 1999. This day was also a day of moderate geomagnetic activity.

The term "limb scan" comes from the fact that the LORAAS spectrograph is mounted on an arm that protrudes from the satellite. LORAAS then sweeps its field of view by using a scan mirror than reflects the emissions onto its lens. The scan mirror is mechanically swept from 10 deg to 27 deg below ARGOS's local horizon. This range of angles relates to tangent point altitudes of 100 km to 750 km. The scan rate varies from .28 deg per second above the altitude of 300 km to .14 deg per second at altitudes below 300 km. The reason for the change in the scan rate is that the change in the electron density is more gradual at altitudes above 300 km compared to below 300 km where the electron density gradient is high. At this scan rate LORASS collects ninety spectra per scan. The spectra are measured simultaneously in the 800Å to 1700Å range with 17Å spectral resolution.

Each second, LORAAS counts the photons that enter the instrument in this wavelength range. The higher the count the higher the intensity of the spectra measured. What is obtained then, is a count rate profile that reflects the quantities of electrons in LORAAS's line of sight (again using the safe assumption in the F-Region that O^+ density equals electron density). The count profiles were produced by using the multiple linear regression technique to fit each spectrum in a limb scan and to extract the count rate in the spectral line [Tho 99]. The line shapes used in the multiple linear regression analysis were measured during pre-flight calibration.

It is well established that counts follow a Poisson distribution. This is the distribution used in calculating the data standard deviation. The standard deviation of a Poisson random variable is the square root of the mean count. However, there is some contamination in the data that must be accounted for in the standard deviation. As mentioned in chapter one, the spectra that will be measured during nighttime in the 800Å to 1700Å range will predominately be 911Å (from radiative recombination), 1216Å (Lyman-α that bleeds over from the dayside H^+ recombination),1304Å and 1356Å emissions (from radiative recombination and O^+ to O^- neutralization). Since the Lyman-α signal can be very strong, its spectra bleeds over onto the 911Å and 1356Å spectra. The 1304Å emission also contaminates the 1356Å counts but not the 911Å since it is far enough away on the photo cathode plate. The 911Å and 1356Å counts are adjusted by subtracting the portion of Lyman-α and the 1304Å spectra that contaminate the counts.

If we let c911 = 911Å counts, r911_1216 = the ratio of 911Å to 1216Å in the 911Å counts, and lya = Lyman-α (1216Å spectrum) measured at that time, then the adjusted 911Å count will be ac911 = c911 - r911_1216 * lya (the ratios are based on Scott Budziens' lineshape model [Tho 99]). Likewise, the adjusted 1356Å counts will be ac1356 = c1356 - r1356_1216 * lya - r1356_1304* c1304. Since the counts are independent Poisson random variables, VAR[c911] = c911. The variance of the adjustment factor is VAR[r911_1216 * lya] = r911_1216^2 * VAR[lya]. This leads to the result:

$$\mathrm{STD[ac911]} = \sqrt{\mathrm{c911} + \mathrm{lya} * r911\_1216^2} \tag{2.1}$$

$$\mathrm{STD[ac1356]} = \sqrt{\mathrm{c1356} + \mathrm{lya} * r1356\_1216^2 + \mathrm{c1304} * r1356\_1304^2} \tag{2.2}$$

which is calculated for each data element gathered.

This was later changed so that the standard deviations of the counts is just the square root of the mean of the counts. This was made possible by leaving the background noise in the data set. This results in a purely Poisson distribution. The noise is later added into the intensities which are used to fit the data set.

Another step in processing the data is to convert the counts into radiance units called Rayleigh's. The counts are measurements of the photon flux at the spectrograph. This is usually less than the emitted flux due to attenuation of the photons by absorption and scattering. A Rayleigh is defined as a measure of the omni-directional emission rate in a column of unit cross section along the line of sight [Ree 89]. An assumption, that the field of view of the spectrograph is uniformly filled, must be made. This is usually true for nightglow emissions. A Rayleigh is, therefore, a measure of surface "brightness", not a unit of energy. As we will see, this conversion to Rayleigh's is crucial to our algorithm. The goal of our algorithm is to find electron density profiles. The parameters estimated for our model functions generate estimated electron density profiles. These parameters, using an iterative approach, can be optimized to generate electron density profiles that we can use to calculate units of Rayleigh's that LORAAS would see if the estimated profiles were real. We want to minimize the errors between the model estimated data values (in Rayleigh's) and the data values (in Rayleigh's) measured by LORAAS. The counts are converted to Rayleigh's by simply dividing the adjusted counts by the instruments' sensitivity (which is in units of counts/s/Rayleigh). The sensitivity for the 911Å data is 0.31 counts/s/Rayleigh and 0.8 counts/s/Rayleigh for the 1356Å data. These multiplicative factors take into account the fraction of the line shape used in the line shape model created by Budzien [Tho 99].

2.2 The Algorithm

The algorithm is based on Discrete Inverse Theory [Men 89] which is composed of a forward function (that models our data) and an iterative inversion process that seeks the maximum likelihood estimate of the parameters to the forward function (assuming our data and residuals are independent and normally distributed). As was pointed out in the subsection 2.1 of this chapter, counts follow a Poisson distribution. If the sample size is large enough, the Poisson distribution can be approximated by a normal distribution.

The inversion process uses a nonlinear least squares iterative algorithm (employing the Levenberg-Marquardt method [Nas 96, Pre 92]) to vary the parameters so as to minimize the χ^2 statistic. The χ^2 statistic that is minimized is $\sum_{i=1}^{n} \mathrm{r}_i^2/\sigma_i^2$ where n is the sample size (data points) and r_i, σ_i

are the i_{th} residual error and standard deviation. Minimizing the residuals is equivalent to maximizing the likelihood that the data observed were produced by the forward function model we proposed for our fit, *if* each data sample is drawn from a normal (or Gaussian) distribution.

A nonlinear least-squares problem is an unconstrained minimization problem of the form:

$$\text{minimize over } x : f(x) = \frac{1}{2}\sum_{i=1}^{n}(f_i(x))^2 \tag{2.3}$$

where $f_i(x)$ equals the residual of the i^{th} data point. If we let $F(x) = (f_1(x)\ f_2(x)\ \ldots\ f_n(x))^T$ be a vector valued function then we can write the unconstrained minimization problem as

$$\text{minimize over } x : f(x) = \frac{1}{2}F(x)^T F(x) \tag{2.4}$$

It follows that the elements of $\nabla f(x)$ can be written (after differentiating using the chain rule with respect to x) as $\nabla f(x) = \nabla F(x)F(x)$. This is the convenient notation for the gradient of $f(x)$. To find the Hessian we need to differentiate one more time with respect to x. The formula for the Hessian is then

$$\nabla^2 f(x) = \nabla F(x)\nabla F(x)^T + \sum_{i=1}^{n} f_i(x)\nabla^2 f_i(x) \tag{2.5}$$

Once we are close to the optimal solution we would expect $f_i(x)$ to be small. This leads us to the conclusion that near the solution

$$\nabla^2 f(x) \approx \nabla F(x)\nabla F(x)^T \tag{2.6}$$

especially if our function is quadratic. The further we get away from a quadratic function the larger the error in equation 2.6 will be.

In the Gauss-Newton method, we apply Newton's method with Equation 2.6 as the approximate Hessian. The Levenberg-Marquardt (LM) method is a variation of the Gauss-Newton method [Nas 96] for finding a descent direction with the Hessian approximated by $\nabla F(x)\nabla F(x)^T + \lambda I$ rather than just $\nabla F(x)\nabla F(x)^T$. This additional term has several different functions. With $\lambda > 0$ properly chosen [Mad 99], the coefficient matrix is always positive definite, and this ensures that our step is a descent step. For large λ our step is much like a steepest descent step [Mad 99] and for very small λ our step is much like a Gauss-Newton step which converges quadratically. The main advantage here is that the Hessian does not have to be computed directly. This saves a great deal of processing time.

The model parameterization used is a four parameter Chapman Layer [Cha 87] that is generally considered to be a reasonable approximation for describing the ionospheric density given the assumption that the ionosphere is isothermal. The Chapman layer is a function of an altitude vector z and has the form:

$$\mathrm{n_{O^+}(z) = nmF2\ exp}\left\{\frac{1}{2}\left[1 - \frac{\mathrm{z - hmF2}}{\mathrm{H}} - \exp\left(-\frac{\mathrm{z - hmF2}}{\mathrm{H}}\right)\right]\right\} \tag{2.7}$$

where hmF2 is the altitude where the ionospheric density peaks, nmF2 is the density at the peak, and $\mathrm{H = H_{top} + (H_{con} - 1.0)\,(z - 400.0)}$. Here $\mathrm{H_{top}}$ = the topside scale height at 400 km and $\mathrm{H_{con}}$ = the slope of the scale height plus 1. The scale height at 400 km is the altitude interval that corresponds to the ionospheric density decreasing by a factor of e at 400 km (given the assumption that the ionosphere is isothermal).

In the experiments that follow, both a three parameter Chapman Layer and a four parameter Chapman Layer are fit. Fixing $\mathrm{H_{con}} = 1$ or some other initial value gives the three parameter Chapman Layer. As mentioned, it is assumed the electron density equals the O^+ density, a safe assumption below the H^+/O^+ transition height.

After we calculate the electron density using the Chapman function we must retrieve the model data values that will be used for comparison against the actual data. This amounts to evaluating two separate line integrals. One for the 911Å data and one for the 1356Å data. The 911Å intensity I (in Rayleigh) at look angle φ is given by

$$\mathrm{I}_{911}(\varphi) = 10^{-6}\alpha_{911}\int_0^{\infty} \mathrm{n_e}(\mathrm{z(s)})\mathrm{n_{O^+}}(\mathrm{z(s)})\,\mathrm{ds} \tag{2.8}$$

where z is a function that gives the tangent point altitude in the line of sight at the location of ds, n_e is the electron density at altitude z, N_{O^+} is the O^+ density at z, ds is the differential path length from the observer, and α_{911} is the radiative recombination rate coefficient. The radiative recombination rate coefficient is an empirical value that is used to describe the rate that O^+ and electrons recombine. The value used for α_{911} was 3.5×10^{-13} at 1160 Kelvin [Mel 99].

In calculating the 1356Å intensities, oxygen and nitrogen densities as well as temperature in the F-Region are needed. At the beginning of the algorithm a call to the procedure MSIS-86 is made. MSIS-86 stands for the Mass-Spectrometer-Incoherent-Scatter model 1986. The MSIS-86 model is based on the extensive data compilation and analysis work of A.E. Hedin and his colleagues. Data sources include measurements from several rockets, satellites, and incoherent scatter radars (Ionosonde's). The density of O is one of the outputs from the model. For a full explanation of MSIS-86 see [Hed 87]. The 1356Å intensities are somewhat more complicated to calculate than the 911Å intensities. A discussion on the intensity calculations can by found in [Dym 97].

At this point the square of the L_2 norm of the scaled residuals (our objective function we want to minimize) is calculated. For the one dimensional problem this is equivalent to calculating the χ^2 value. The algorithm continues to adjust λ, if necessary, until a descent direction is found for the parameters. This process of adjusting the parameters, calculating the Chapman Layer using the altitude vector z, calculation of the retrieved intensities from the Chapman Layer profiles, and calculation of the square of the L_2 norm is continued until certain termination rules are satisfied. The algorithm is terminated when three different stopping criteria are met.

The first is the norm of the gradient divided by one plus the objective function must be less than $\sqrt{\epsilon}$. This test will make sure that the gradient, relative to the objective function value, is close to zero. This is a necessary condition to insure we are at a local extremum. The fact that $\nabla F(x)\nabla F(x)^T + \lambda I$ is always positive definite guarantees we are at a local minimum and not at a maximum or saddle point.

The second test is to check if the magnitude of the change in the parameters divided by one plus the magnitude of the current parameter vector is less than $\sqrt{\epsilon}$. This test ensures that the parameter values are as accurate as our forward function and machine accuracy will allow.

The third test tests the change in the objective function value. If the absolute value of the change in the objective function relative to the objective function (i.e. divided by one plus the objective function value) is less than ϵ then the minimum objective value we have reached is as accurate as the machine and forward function will allow. The value of ϵ is chosen to be equal to the accuracy of the forward function or to the accuracy of the computer, which ever is greater. A discussion on termination rules can be found in [Nas 96].

The LM method of minimizing an objective function requires that starting values be given to the parameters of the forward function. As we will see, when testing different data sets for the one dimensional model, the convergence to the minimum objective value is not very sensitive to starting values. This is not true of the two dimensional model which we will see is very sensitive to the initial value given to the parameter for the altitude of the peak density, hmF2, and the parameter $\mathrm{H}_{\mathrm{con}}$ that is the slope of the scale height at 400 km. The initial values for the one dimensional model are set at average values that represent the F-Region of the ionosphere. The values are hmF2=350 km, $\mathrm{nmF2} = 8 \times 10^5/\mathrm{cm}^3$, $\mathrm{H}_{\mathrm{top}} = 70$ km, and $\mathrm{H}_{\mathrm{con}} = 1$ for the three parameter Chapman model or $\mathrm{H}_{\mathrm{con}} = 1.1$ for the four parameter Chapman model.

2.3 Improvements Made to the Already Existing Software

Before work began on this thesis, a large amount of the code had already been written. The job ahead was to improve the already existing code which would make it converge faster to more accurate parameter values and to achieve more suitable fits to the data. Any changes made to the one dimensional model carry over into the two dimensional model. There were also changes made to the two dimensional model only. These are discussed in the next chapter.

First, the accuracy of the forward function was checked. The step size for taking the finite difference derivative estimates was originally set at 1×10^{-3}. It is optimal to have the step size set at the square root of machine accuracy $\sqrt{\epsilon_{mach}}$ (see [Nas 96]). If the code was set up to be double precision, this would put $\sqrt{\epsilon_{mach}} = 1 \times 10^{-8}$. The only way (even at double precision) an epsilon that small could be used is if the forward function was accurate close to machine accuracy. All the variables in the procedures and functions were set to double precision[1].

A modification was made to the procedure that calculates the Jacobian ($\nabla F(x)$) to allow the option to use centered differencing derivative estimates [Nas 96] if needed. The advantage is that even with a step size of 1×10^{-4} (which would require far less machine accuracy) a derivative as accurate as the $\sqrt{\epsilon_{mach}} = 1 \times 10^{-8}$ step size using the finite difference derivative estimate can be obtained. The only problem with centered differencing is that twice as many forward function calculations are required. This is not a big issue for the one dimensional model that takes a few minutes to run. It is a big issue for the two dimensional model; the two hour run time on a fast computer would turn into three or four hours.

It turns out that setting all the variables to double precision increased the forward function accuracy so that finite differencing was adequate and the

---

[1] All computer code is available from the author or Dr. Borchers

step size used could be $\sqrt{\epsilon_{mach}} = 1 \times 10^{-8}$. Figure 2.1 shows how accurate the forward function was at single precision. The top plot shows an arbitrary parameter that was normalized to one and then perturbed down ten times and up ten times by a factor of 1×10^{-6}. The corresponding function values show that the function is sensitive to the changes in the parameter. The bottom plot of Figure 2.1 shows the function values breaking down when the parameter is perturbed by a factor of 1×10^{-7}. This indicates that the forward function is accurate to changes of up to 1×10^{-6} in the parameter values. Figure 2.2 shows the same parameter and corresponding function value after all variables were set to double precision. This time, the function values do not start losing integrity until the change in the parameter is a small as 1×10^{-15}. Clearly this is the desirable choice for more accurate derivatives.

In the process of checking the forward function accuracy, a bend in the curve of one of the retrieved sets of data values was observed. One of the basic assumptions in the Gauss-Newton or Levenberg-Marquardt methods is that the function be smooth (twice continuously differentiable). This was not the case for our forward function. After investigating the functions and procedures calls[2] of the forward function, it was discovered that hard coded limits on different function values had been set.

The limits were removed when possible and in cases where it was not possible (i.e. dividing by close to zero or infinity) a penalty function was introduced. The penalty function data value is stored as another data point in the data vector but is not included in the χ^2 calculation. It should always

---

[2] All computer code is available from the author or Dr. Borchers

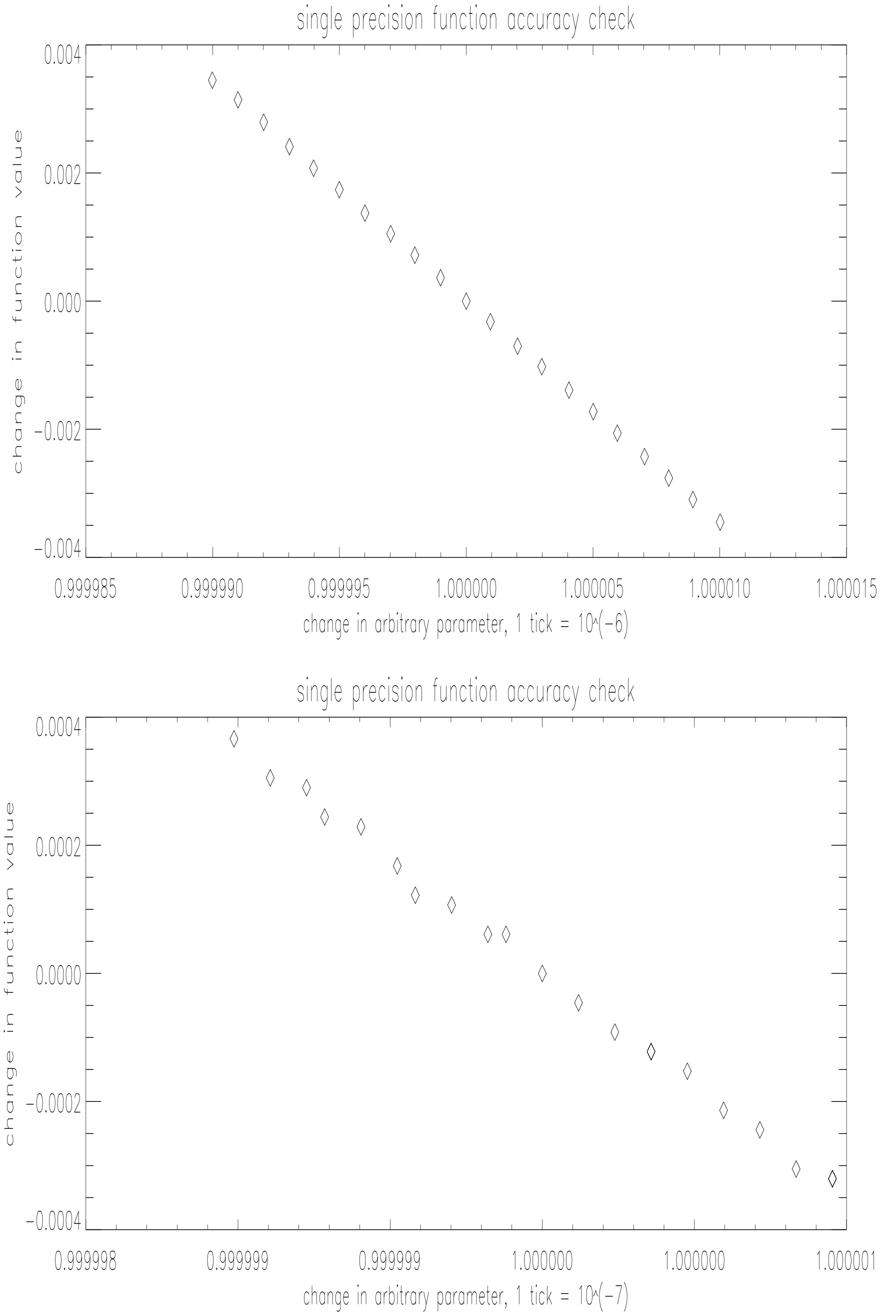
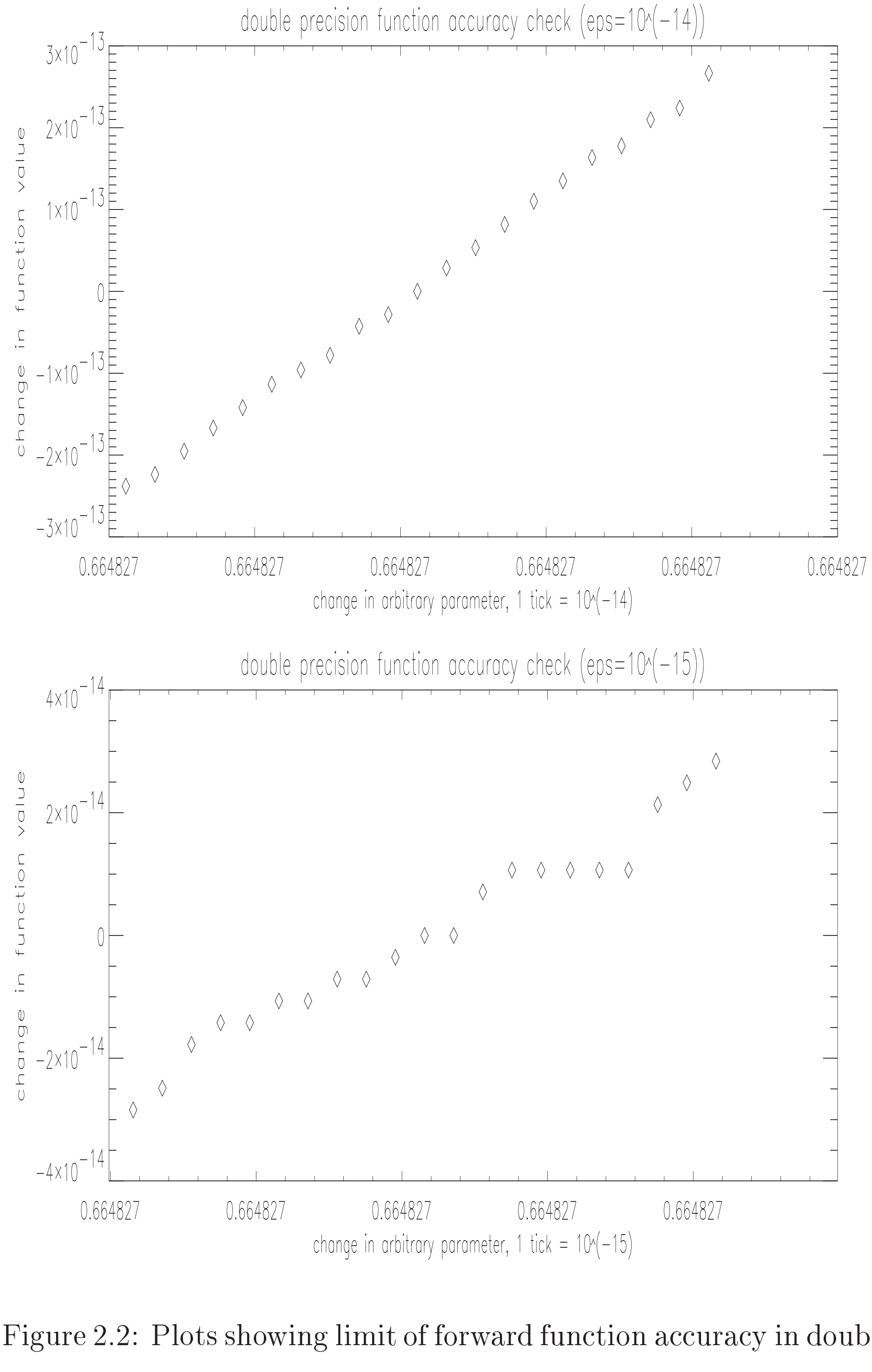


Figure 2.1: Plots showing limit of forward function accuracy in single precision.


double precision function accuracy check (eps=10^(−14))

double precision function accuracy check (eps=10^(−15))

Figure 2.2: Plots showing limit of forward function accuracy in double precision.

have a value of zero unless one of the parameters or variables in the forward function goes outside its upper or lower bound. If this happens then the penalty function takes on the value penalty $= 10^{15} * (\text{variable} - \text{ulbound})^2$. When the square of the L_2 norm is calculated, having this huge value in one of the data points makes the square of the L_2 norm larger than it was on the previous step no matter what was gained by the descent direction found by stepping out of bounds. Therefore, the step is not taken. Lambda (the damping factor in the LM method) is increased until a descent direction is found without having to step outside the upper and lower bounds imposed on the variables.

Figure 2.3 shows the discontinuity found in the forward function. The value of the parameter (0.8725 of its starting value) at the discontinuity is definitely in the range of values the parameter could take on at optimality which means the discontinuity would probably be involved in a derivative calculation at some point in the iterative steps. The bottom plot of Figure 2.3 shows the same parameters' effect on the same retrieved data value after the changes described above and the bend is gone. After an exhaustive search the forward function is found to be smooth everywhere for the values encountered.

As previously mentioned, if the scan mirror for LORAAS points toward a star, the next few data elements will have much larger counts then they should. Since these data are not measures of 911Å or 1356Å emissions, they must be omitted. To do this, a data masking vector is used. In the initial stages of the program, data structures in IDL are built from the LORAAS data set. One of the elements in the data structure is the data mask vector. It is initially set to all ones. Then it is up to the user to plot the data set

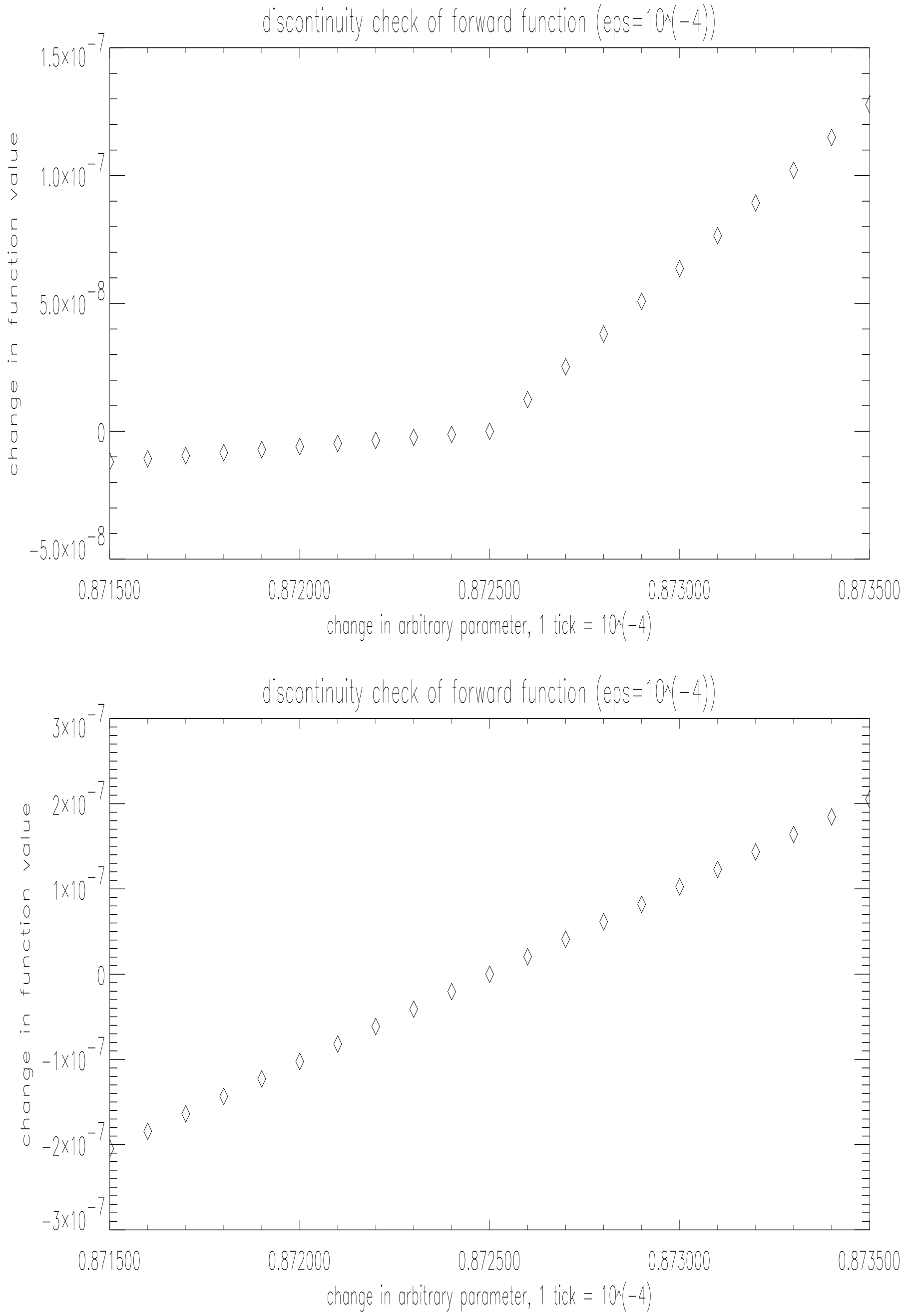


Figure 2.3: Example of effect of removing upper and lower bounds .

and locate any of the contaminated data points. The corresponding elements in the data mask are then set to zero. The algorithm proceeds with all the data points intact but when it comes to calculating the objective function, the function we are trying to minimize (in this case its $\chi^2 = \sum_{i=1}^{n} \text{residuals}_i^2/\sigma_i^2$), only the data points corresponding to the elements in the data mask with ones are used. The goodness-of-fit measure used is the reduced χ^2 which is simply the ratio of χ^2 divided by the degrees of freedom. In this case, the degrees of freedom are $n - m$ where n is the number of indices in the data mask vector that are set to one and m is the number of parameters being used to fit the model.

The data mask is also used to mask out data points that correspond to very low counts. Some data points come from very low photon counts (one to three counts per data point). This is not a large enough number for the Poisson distribution to be approximately normal. Also these data values are small in comparison to the noise level of the data so they do not contribute much information toward the model. Omitting these data points does not effect the outcome of the optimal parameter values but does effect the normality of the residuals which can become distorted for these data elements.

The method used employs a moving average of eight consecutive data points calculated while reading through a data set. When this average is below a cutoff value (usually set at 3 Rayleigh's) the data mask is zeroed at the corresponding data indices. The routine then jumps to the other side of the profile peak and starts checking the moving averages again. This leaves the defining part of the Chapman curve intact but removes data points that are essentially zero. The data mask also helped the convergence rate. Without the extra noise both the one and two dimensional models converged in less than ten iterations.

Another method tried in the implementation of the 1-D and 2-D models leaves the background noise in. Since the background noise is not subtracted off but instead added to the retrieved intensities and then fit to the data, there are no negative data points to worry about. This method is best since the counts are purely Poisson and are high enough to allow the distribution to be approximately normal. Another benefit is in the fact that no data points are removed (except ones that have been contaminated by starlight) and thus no information is lost. This method was checked against the results for the 1-D model thus far and the optimal parameter values were in agreement for both methods.

A simulated test on low count data was run to see how far the worst case scenario would deviate from normality. Ten thousand random samples from a Poisson distribution were generated with a mean value of 14 to simulate the 911Å and 1356Å counts. This generates random numbers as small as six which corresponds to the smallest counts in our real data set. Another ten thousand samples from a Poisson distribution where generated with a mean of 150 to simulate the Lyman-α counts. A small amount of the simulated Lyman-α was subtracted from the simulated low count data set corresponding to what would be subtracted in the real data set.

Anderson-Darling [Law 00] normality tests were run on the differences between these two cooked up Poisson distributed data sets. The Anderson-

Darling (A-D) test for normality is a weighted average of the squared difference between the cumulative distribution of the tested data set against the cumulative distribution of a true normal distribution with corresponding mean and standard deviation. The main difference between the A-D test and other tests for normality is that more weight is given to the difference in the tails of the normal distribution where distributions tend to differ the most. The A-D test is therefore very sensitive (especially with 10,000 data points) and the test is expected to fail when applied to our cooked up data set.

A Q-Q plot for our simulated data set is shown in figure 2.4. The simulated data set differs from a true normal distribution mainly in the tails which is reflected in the very poor A-D test statistic. For a data set with thousands of observations, the A-D statistic (known as the A^2 statistic) should be less than one. For information on the A-D statistic and the corresponding p-value see [And 54, Ste 74].

Figure 2.5 shows the results of the difference when 100 is used for the mean instead of 14. It would be expected that with $\mu = 100$, the simulated data would approximate a normal distribution more closely and in fact Figure 2.5 shows a much better A^2 value but the simulated data still fails the A-D test with a p-value of 0.0000. The point is that the Poisson distribution is approximately normal when μ is large enough. The histogram plot in Figure 2.6 shows that the simulated data with mean 14 does approximate a normal distribution. How does this approximation effect the residuals of the real data? Later, the results will be compared to runs of real data with and without the low count data elements included.

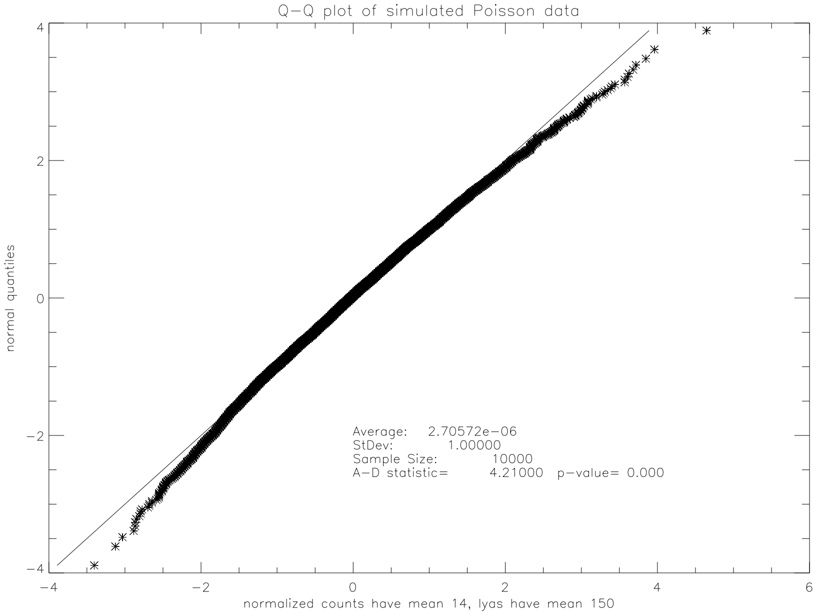


Figure 2.4: Q-Q plot showing how normal the Poisson simulated data is when the mean is 14.

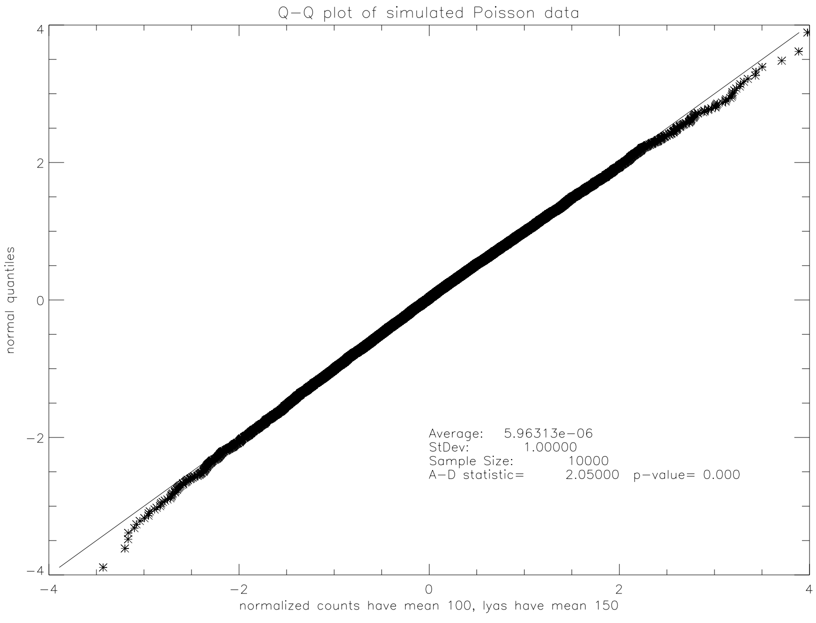


Figure 2.5: Q-Q plot showing how normal the Poisson simulated data is when the mean is 100 .

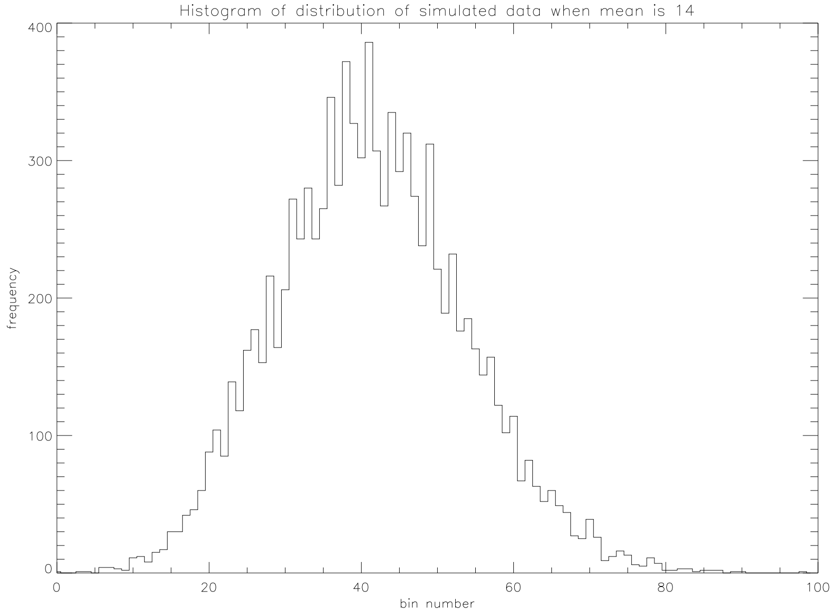


Figure 2.6: Histogram plot of simulated data with mean set at 14.

One more technique for data cleaning has to be considered other than removing stars and entries where the signal to noise ratio is very low. This is the background noise adjustment. In this case, the background noise comes from the 1216Å frequency (or Lyman-α) with a small contribution from 1304Å contamination in the 1356Å count case. The given ratios for this background noise where derived from the line shape model used in calibrating the LORAAS instrument. It was observed that the given ratios of 1356Å and 911Å to 1216Å and 1304Å contamination subtracts off too much of the 1216Å and 1304Å counts.

The way to correct this is to adjust these ratios so that when the data set is plotted, the section of data where it is expected to be nearly zero should have just as many data elements randomly distributed above and below zero. Before the algorithm is run, the data set under consideration is examined to make sure that all contamination by stars are excluded and that the subtracted background noise is correct.

Another way is to not use the ratios just discussed as the scaling factor for the Lyman-α but to use counts in nearby bins as the estimate of the background noise. Due to a “hot spot” found on the detector in the region near Lyman-α, the ratio used to scale the Lyman-α often calculates to much background. The total counts in bins nearby the 911Å and 1356Å regions work well as the total background noise.

2.4 The Three Parameter Chapman Model

The Chapman function was described in Section 2.2 and is given by equation 2.7. In this section we will explore results from the optimization of three of the four parameters in equation 2.7 using the algorithm described in Section 2.2. The fourth parameter is the slope of the scale height and it will remain fixed at $H_{\mathrm{con}} = 1$. To help find a more precise parameterization of the data, 911Å and 1356Å emissions will be fit simultaneously. Since the line of sight of LORAAS cuts across many degrees of latitude and longitude the readings can only be parameterized by a Chapman function (which gives a vertical profile) if the ionosphere is assumed to have only vertical variation. This is only true in the mid-latitude regions ($\pm 25\,\mathrm{deg}$ to $65\,\mathrm{deg}$) and so it only makes sense to run tests on limb scan profiles taken in the mid-latitude region.

The objective function we are minimizing for the one dimensional model is

$$\parallel \text{weighted residuals} \parallel_2^2 = \sum_{i=1}^{n} \text{residuals}_i^2/\sigma_i^2 = \chi^2_{n-m-1} \tag{2.9}$$

where the residuals are the differences between actual data and the retrieved data, σ is the standard deviation of each data point as calculated in equations 2.1 & 2.2 and n is the number of data points. The topics discussed will be the global convergence properties of the algorithm (an attempt will be made to find multiple local minima), the goodness-of-fit using the reduced χ^2 statistic, the normality of the residuals (which is essential to the χ^2 distribution), and comparison of the findings to corresponding Ionosonde measurements.

An mid-latitude limb scan profile was chosen arbitrarily from a data set obtained from LORAAS on November 24, 1999 at 17:67 Universal time. The tangent point latitudes and longitudes for this profile are 51 deg latitude and 135 deg longitude. Two of the data points were excluded from the fit because they were found to be contaminated by star light. All other data points where used, including those with small counts. Later in this section we will run a fit with small count data points removed and compare the results. The initial starting values are taken as the general average values for mid-latitude region Chapman parameters as discussed in Section 2.2. Table 2.1 shows the results of the objective function value and the optimized parameter values for different initial values (the usual initial values adjusted randomly up and down by the shown percentage). Testing convergence by adjusting the initial values of the parameters by more than 50% starts to get close to initial parameter values that are physically unrealistic. Other data subsets were tested by randomly adjusting the initial values up and down by up to

% param adj	initial param values	final param value	obj func
0.0%	hmF2 = 350 km	248 km	159
	nmF2 = 800000/cm^3	598400/cm^3	
	H$_{\text{top}}$ = 70 km	66 km	
±25.0%	hmF2 = 437 km	248 km	159
	nmF2 = 600000/cm^3	598400/cm^3	
	H$_{\text{top}}$ = 87.5 km	66 km	
±50.0%	hmF2 = 175 km	248 km	159
	nmF2 = 1.2×10^6/cm^3	598400/cm^3	
	H$_{\text{top}}$ = 35 km	66 km	

Table 2.1: Three parameter results from different initial values converging to same solution.

50% with the algorithm converging to the same objective value every time. This result shows that the one dimensional model using the three parameter Chapman function has good convergence properties within the range of realistic starting values. Confidence intervals for the parameters are hmF2=248 ± 13 km, nmF2=$5.98 \pm .33 \times 10^5$/cm^3, and $\text{H}_{\text{top}} = 66 \pm 8.33$ km. The reduced χ^2 value for this fit is 0.9109. The corresponding p-value is 0.7951 at 175 degrees of freedom, which is significant at the $\alpha = .05$ level. This indicates that the three parameter Chapman function is fitting the data well.

Another crucial step in checking the validity of the model is to verify that the residuals are uncorrelated and normally distributed. Figure 2.7 shows that the residuals are definitely normally distributed with the Anderson-Darling test for normality [Law 00] yielding a p-value of 0.893. One way to check for patterns or correlations is to plot the residual vector against itself offset by one index. If the residuals are correlated patterns will be visible in the

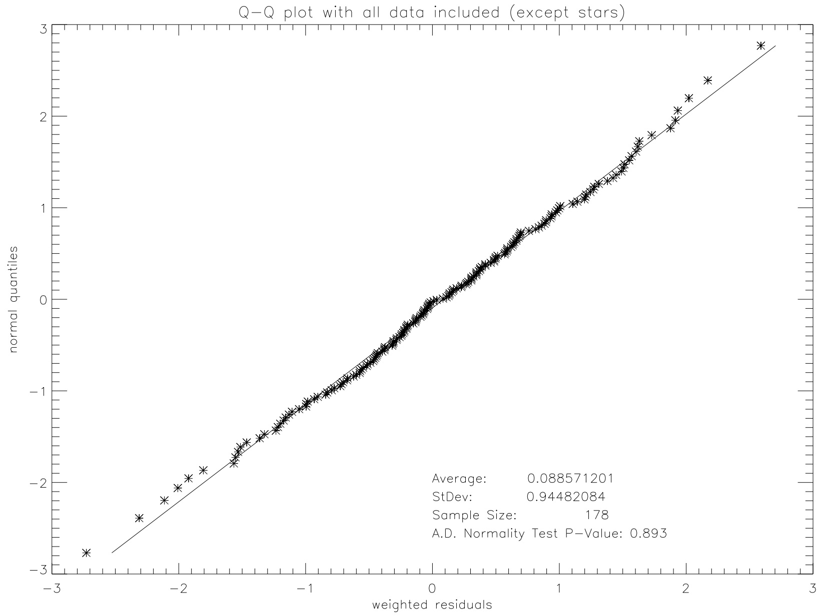


Figure 2.7: Q-Q plot showing residuals are normally distributed.

resulting scatter plot. In the scatter plot of Figure 2.8, there appears to be no significant correlation between the residual values, no patterns are obvious, and the residuals appear randomly distributed. Even though the errors are well within the acceptable region of normality, we will run the same test with small data elements (averaging a value of less than three Rayleigh's) removed. The optimal parameter values were the same as when all data points were used. The reduced χ^2 value for this run is 0.9723 with corresponding p-value of 0.5834 at 152 degrees of freedom. A little was lost in the fit with the low count data elements removed but we shall now look at the residuals to see if there was an improvement made there. The Q-Q plot of Figure 2.9 shows the A^2 statistic with a p-value of 0.915 thus indicating that the normality of the residuals was slightly improved over the run with all data points included.

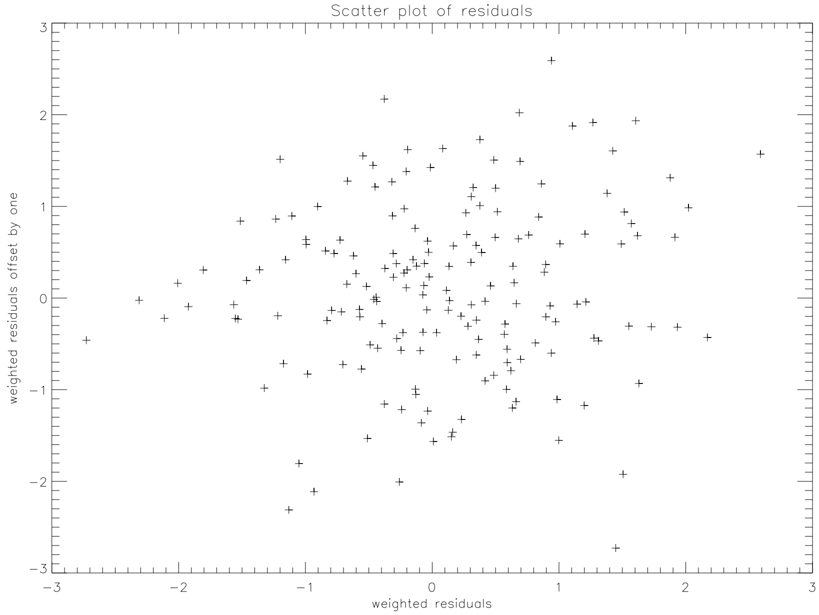


Figure 2.8: Scatter plot to show randomness of residuals.

Figure 2.10 shows the simultaneous 911Å and 1356Å fit of the intensities derived from the Chapman function (smooth curve) plotted over the actual real data. The curve on the left is the 911Å data fit and the curve on the right is the 1356Å data fit. The two data points far removed from the rest are the stars that were excluded. In the second run with low count data removed, about 25 data points near data indices 0 and 100 were excluded from the fit. At this point, the model fits the data very well. It will be interesting to see how the model compares to corresponding Ionosonde results.

The Millstone Hill Ionosonde is located at 41 deg N and −71 deg E. The closest tangent point altitudes of the line of sight of LORAAS that could be obtained from our data set coincident with Millstone Hill are 40 deg N

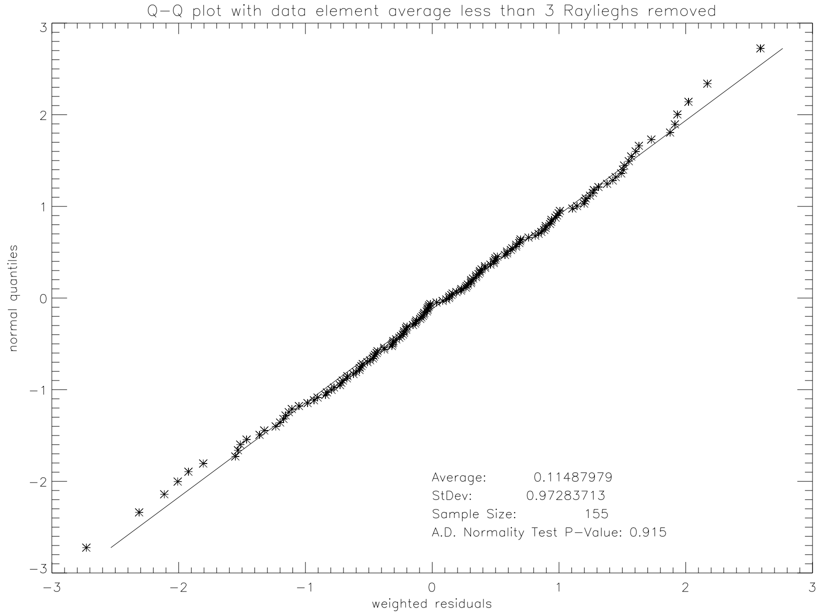


Figure 2.9: Q-Q plot of residuals with low count data removed.

and −76 deg E taken at 7:33 UT. There was auroral precipitation 10 deg N of Millstone Hill at that time which caused fluctuations and abnormally high Ionosonde measurements. The Millstone Hill Ionosonde measured a peak electron density of $5.0 \times 10^5 \text{cm}^{-3}$ and $3.7 \times 10^5 \text{cm}^{-3}$ at 7:30 UT and 7:45 UT respectively. The height of the peak electron density measured by the Ionosonde was 333 km and 276 km at 7:30 UT and 7:45 UT, respectively. Normally at this time of year the Ionosonde measurements at Millstone Hill do not fluctuate and average electron density is about $3.0 \times 10^5 \text{cm}^{-3}$. The parameters obtained from our algorithm show the peak electron density to be $4.14 \pm .38 \times 10^5\,\text{cm}^{-3}$ at an altitude of 319 ± 13 km. The model predicted the scale height parameter to be 62 ± 12 km. Figure 2.11 shows the retrieved Chapman Layer profile over Millstone Hill at 7:35 UT. The asterisk on the plot is where the Ionosonde

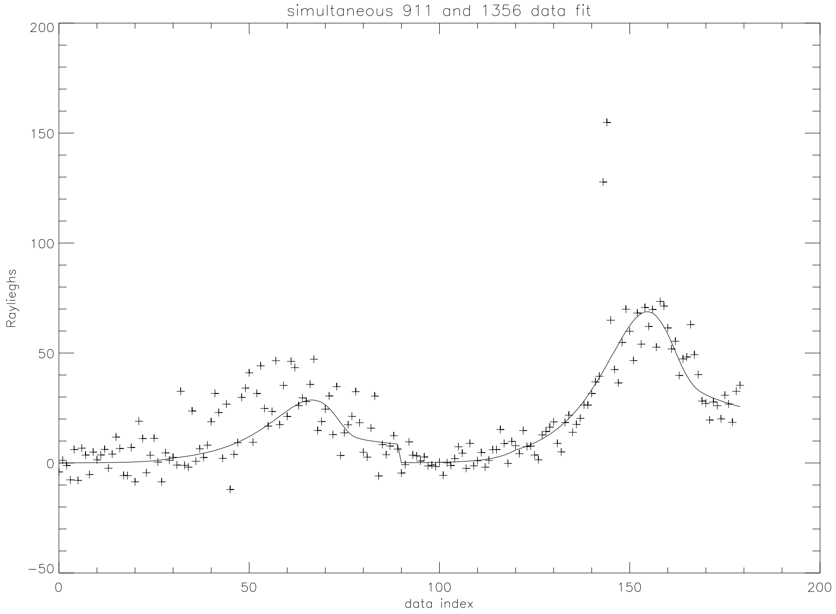


Figure 2.10: Simultaneous 911Å and 1356Å fit.

measured the peak electron density at 7:30 UT and the diamond is where the Ionosonde measured the peak electron density at 7:45 UT. The model predicts values that lie in between these two readings which makes sense if the electron density and peak altitude were decreasing over this fifteen minute period. The reduced χ^2 value for this fit is 1.03 with corresponding p-value of 0.3818 at 154 degrees of freedom.

The variance-covariance matrix used to infer confidence intervals on the parameters is based on the asymptotic variance-covariance matrix of the nonlinear regression coefficients. There is a nonlinear analog to the linear variance-covariance matrix $\sigma^2(\mathbf{X^TX})^{-1}$ where $\mathbf{X}$ is the model matrix described

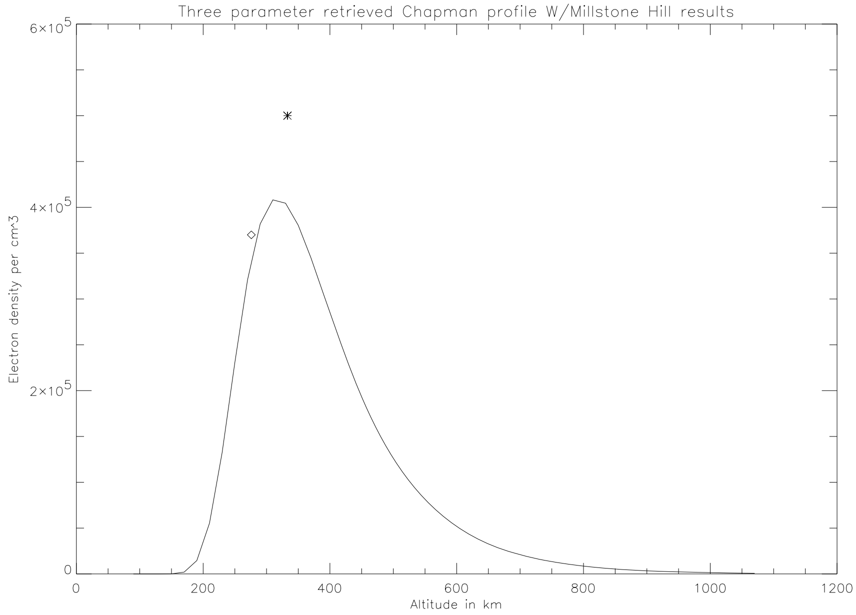


Figure 2.11: Retrieved three parameter Chapman profile with corresponding Ionosonde measurements.

in [Mye 90]. This estimate of the asymptotic variance-covariance is given by

$$\widehat{\mathrm{cov}}(m) = \sigma^2(\mathbf{J}^{\mathbf{T}}\mathbf{J})^{-1} \tag{2.10}$$

where $\mathbf{J} = \nabla f(x)$ is the Jacobian of the objective function given in equation 2.4 and σ^2 is the variance of the errors. The 95% confidence intervals were then calculated using the square roots of the diagonals of $\widehat{\mathrm{cov}}(m)$ multiplied by 1.96. This is the less conservative form of the confidence interval that does not take into account the off diagonal covariance terms.

2.5 The Four Parameter Chapman Model

We next move on to the results of the four parameter version of the Chapman Layer function. As discussed and displayed by Equation 2.7, the fourth parameter $\mathrm{H_{con}}$ is equal to the slope of the top side scale height at 400 km plus 1. We use exactly the same Chapman function for both the three and four parameterization versions, only in the three parameter version we set $\mathrm{H_{con}} = 1$ and don't allow it to change, which effectively cancels the fourth parameter. In the four parameter version the initial value of $\mathrm{H_{con}} = 1.1$ and it is allowed to change to a value corresponding to an optimal fit. Everything else remains the same in the algorithm.

The same two data sets will be fit as was done using three parameters and comparing our results to see if the fourth parameter improves the fit enough to warrant its use. Success of the fourth parameter will be gauged using an F-test.

The first result presented will be the convergence properties of the four parameter model. The same three tests were run as were run for the three parameter model. The four parameter results are summarized here in Table 2.2. As can be seen, the algorithm converged to the same objective function value each time. The parameter values changed a little bit each time but this is because changing the slope of the scale height warps the top side of the curve in ways where an equally good fit can be obtained even with slightly different parameter values. Confidence intervals for the parameters are hmF2=252 ± 18 km, nmF2=$5.98 \pm .45 \times 10^5/\text{cm}^3$, $\text{H}_{\text{top}} = 66 \pm 10.68$ km, and $\text{H}_{\text{con}} = .95 \pm .15$ km. The χ^2 value for this fit was 0.9217 with corresponding p-value of 0.76. The ratio $R = \frac{\chi_1^2/\nu_1}{\chi_2^2/\nu_2}$ follows an F distribution [Mye 90] as long as weighted data errors are distributed as $N(0,1)$, which, we have shown they are. If we take the corresponding reduced χ^2 values from our three and four parameter runs then we get $R = .9109/.9217 = .9883$. The upper 5% critical value F statistic with $\nu_1 = 175$ and $\nu_2 = 174$ is $F_{.95}(175, 174) = 1.2837$. Since R is less than this value, the improvement in fit from the three to the four parameter model is not significant at the 5% level.

A plot of the four parameter Millstone Hill fit is shown in Figure 2.12 with the same corresponding Ionosonde results that were shown in Figure 2.11. The asterisk on the plot is where the Ionosonde measured the peak electron density at 7:30 UT and the diamond is where the Ionosonde measured the peak electron density at 7:45 UT. The Millstone Hill Ionosonde measured a peak electron density of $5.0 \times 10^5 \text{cm}^{-3}$ and $3.7 \times 10^5 \text{cm}^{-3}$ at 7:30 UT and 7:45 UT respectively. The height of the peak electron density measured by the Ionosonde was 333 km and 276 km at 7:30 UT and 7:45 UT, respectively. The parameters obtained from our algorithm show the peak electron density to be $4.65 \pm .7 \times 10^5\,\text{cm}^{-3}$ at an altitude of 299 ± 13 km. The model predicted the

% param adj	initial param values	final param value	obj func
0.0%	$\mathrm{hmF}_2 = 350\,\mathrm{km}$	252 km	160
	$\mathrm{nmF2} = 800000/\mathrm{cm}^3$	$586224/\mathrm{cm}^3$	
	$\mathrm{H}_{\mathrm{top}} = 70\,\mathrm{km}$	67 km	
	$\mathrm{H}_{\mathrm{con}} = 1.1$	0.95	
±25.0%	$\mathrm{hmF}_2 = 437\,\mathrm{km}$	253 km	160
	$\mathrm{nmF2} = 600000/\mathrm{cm}^3$	$583854/\mathrm{cm}^3$	
	$\mathrm{H}_{\mathrm{top}} = 87.5\,\mathrm{km}$	65 km	
	$\mathrm{H}_{\mathrm{con}} = .825$	0.96	
±50.0%	$\mathrm{hmF}_2 = 175\,\mathrm{km}$	252 km	160
	$\mathrm{nmF}_2 = 1.2\mathrm{e}6/\mathrm{cm}^3$	$583845/\mathrm{cm}^3$	
	$\mathrm{H}_{\mathrm{top}} = 35\,\mathrm{km}$	67 km	
	$\mathrm{H}_{\mathrm{con}} = 1.65$	0.95	

Table 2.2: Four parameter results from different initial values converging to same solution.

scale height parameter to be 52 ± 12 km with the slope of the scale height being predicted at $1.13 \pm .07$. The χ^2 value for this fit is 0.9951. The ratio R needed for the F test against the three parameter model is $R = 1.0290/.9951 = 1.0301$. The upper 5% critical value $F_{.95}(154, 153) = 1.3050$. Since R is less than the critical value, the improvement in the fit of the model is not significant at the 5% level.

It is important to check the residuals of the four parameter Chapman fit to make sure the they are normally distributed. Figure 2.13 shows a Q-Q plot of the residuals with p-value from the A.D. test of 0.274 indicating the residuals are $N(0, 1)$. Figure 2.14 shows a scatter plot of the residuals indicating they are also random and uncorrelated. We can conclude that the fit of the model passed the χ^2 test.

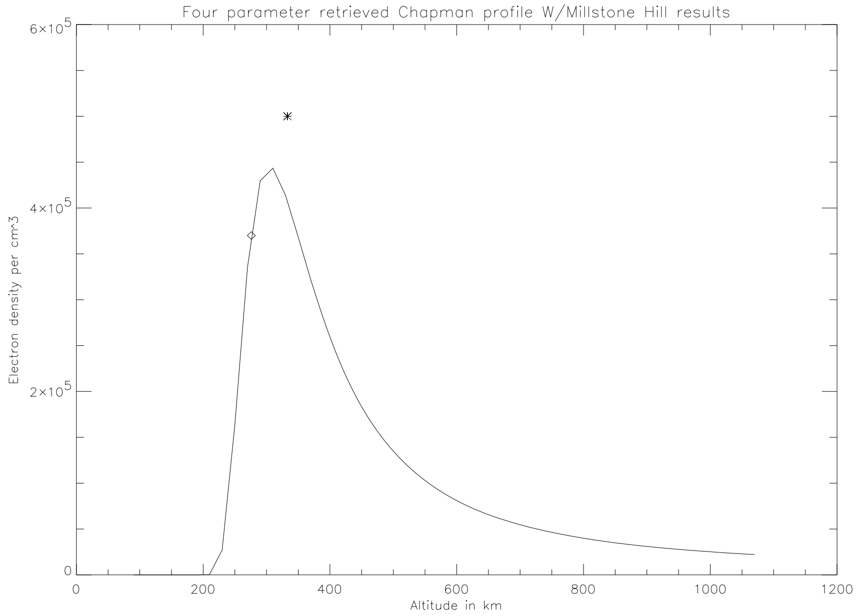


Figure 2.12: Four parameter Chapman profile with corresponding Ionosonde measurements.

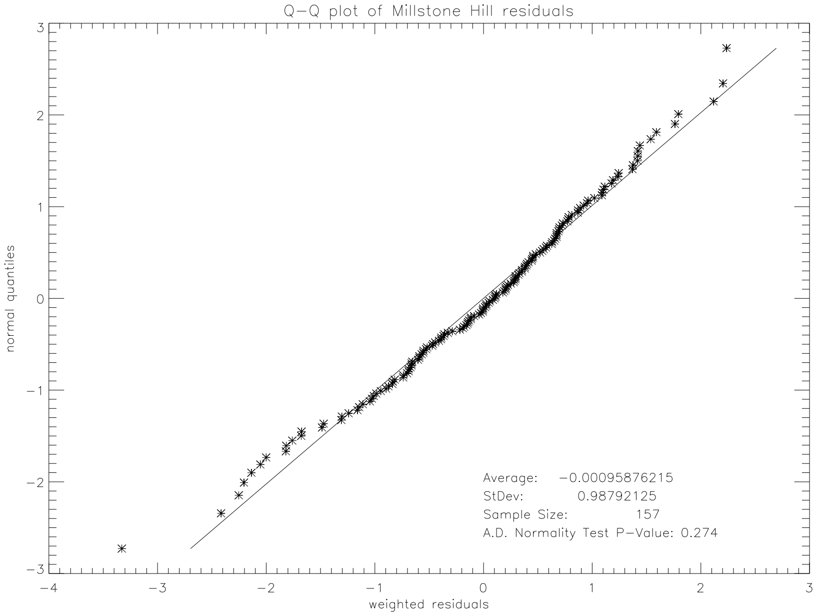


Figure 2.13: Q-Q plot showing four parameter Millstone Hill residuals are N(0,1).

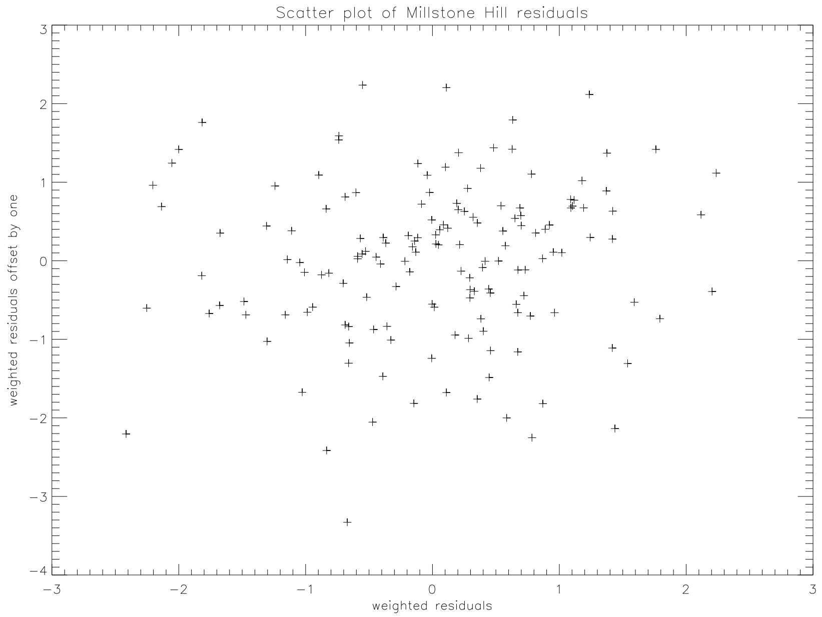


Figure 2.14: Scatter plot of Millstone Hill residuals showing no correlation.

2.6 The Free Form Forward Model

The free form forward model is an alternative to using the Chapman profile to obtain electron densities. Instead, the parameters that are adjusted are electron densities at chosen altitudes. A cubic spline function is used to find the electron densities in between the parameters at all altitudes needed for the intensity calculations. The cubic spline function used is an IDL function based on the spline function described in [IDL 95, Pre 92]. Essentially it computes a cubic polynomial curve in between the points in such a way as to keep the function differentiable at the points of connection.

At the beginning of the algorithm, the same initial values are used that where used for the Chapman models from which the electron densities at the different altitudes (chosen as the parameters) are calculated. The calculated electron densities at the sparsely chosen altitudes are saved in a vector and the log of the values are taken. The electron densities at these altitudes are the model parameters that will be iteratively adjusted in the fit. The log is used to help the stability of the spline function in fitting a curve. The altitudes chosen are up to the user. We used altitudes closer together at the peak of the curve and more spread out at very low and very high altitudes. Results are highly variable based on the altitudes chosen. The trick is not to use too many altitudes (i.e. parameters) since the χ^2 values will be effected by corresponding lower degrees of freedom.

The spline function calculates the log of the densities for all altitudes. The vector of electron densities will be used to calculate the intensities used to fit the data. The algorithm proceeds as before by adjusting the parameter values so as to minimize the squared difference between these intensities and the original data set intensities.

The principle advantage to using this method is that fluctuations in the electron density become evident. The Chapman function cannot, by the nature of the equation, show ripples in the electron density profile. The free form model is sensitive to these ripples, depending upon the region in which the parameter altitudes where chosen close together. We chose the parameter altitudes closer together at around the peak. This can be changed, for instance, to the top side of the profile if the user is more interested in fluctuations there.

Comparing the results obtained from the first data set with the results from the free form model, Figure 2.15 shows the obtained intensities from the

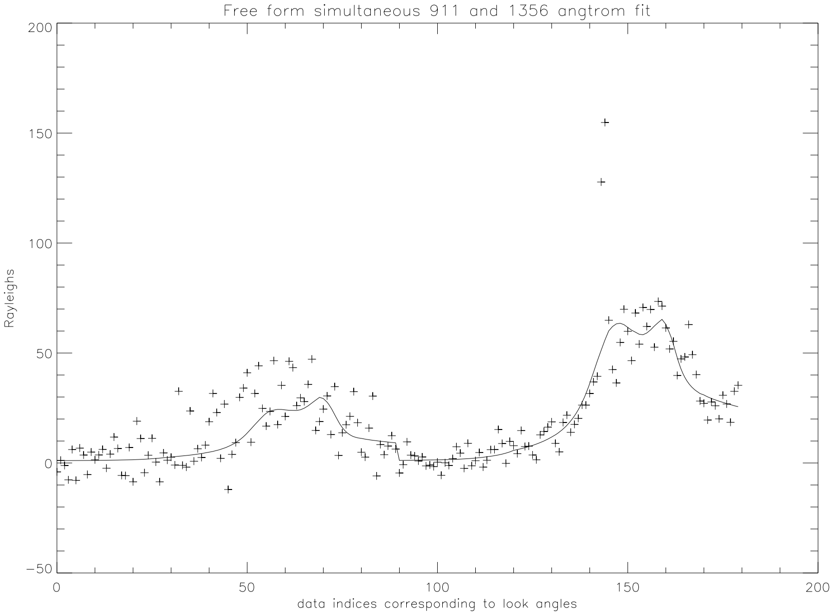


Figure 2.15: Free form simultaneous 911Å and 1356Å fit.

cubic spline interpolation plotted over the actual data point intensities. At the peak, there is a fluctuation in the data set that was not picked up by the Chapman profile. This may be a more realistic model. The objective function (square of the L_2 norm) was improved from 159 to 148. The reduced χ^2 value of 1.0131 reflects a worse fit but that is because the degrees of freedom went from 152 to 147. The corresponding p-value at 147 degrees of freedom is 0.4401 indicating a good fit. The ratio R of reduced χ^2 values for this data set is $.9723/1.0131 = .9597$. The upper 5% critical value $F_{.95}(152, 147) = 1.3104$. This indicates the extra parameters do not have a statistically significant effect on the fit at the 5% level of significance. Figures 2.16 and 2.17 show the residuals of the free form model applied to the first data set to be random $N(0, 1)$.

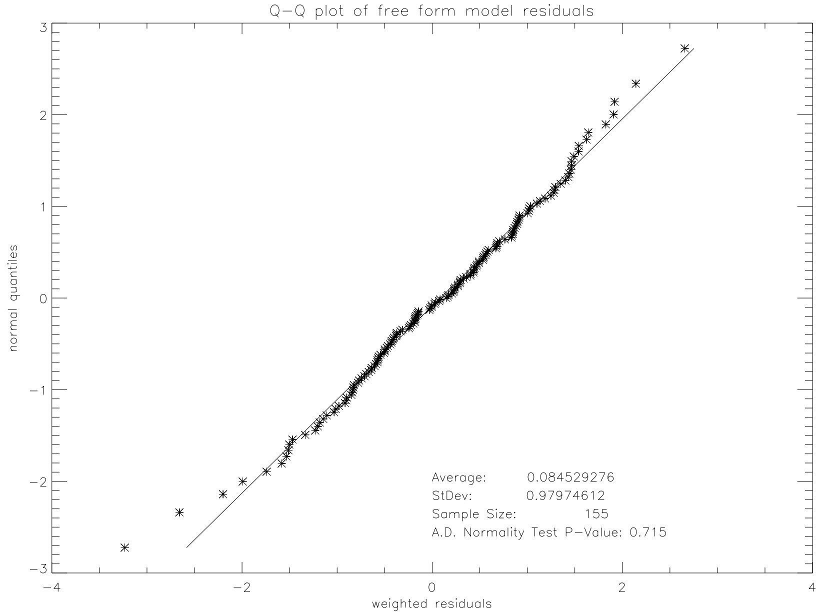


Figure 2.16: Q-Q plot showing free form model residuals are N(0,1).

One problem with using the free form method is that there is no information given about the scale height of the Ionosphere. Also, confidence limits are difficult to obtain because the variance is very large at parameter altitudes that are not closely spaced together. The free form model does pose an interesting way to fit the data, however. Figure 2.18 shows an electron density profile using the free form model on the Millstone Hill data. The asterisk on the plot is where the Ionosonde measured the peak electron density at 7:30 UT and the diamond is where the Ionosonde measured the peak electron density at 7:45 UT.The Millstone Hill Ionosonde measured a peak electron density of $5.0 \times 10^5 \mathrm{cm}^{-3}$ and $3.7 \times 10^5 \mathrm{cm}^{-3}$ at 7:30 UT and 7:45 UT respectively. The height of the peak electron density measured by the Ionosonde was 333 km and 276 km at 7:30 UT and 7:45 UT, respectively. It is noted that the

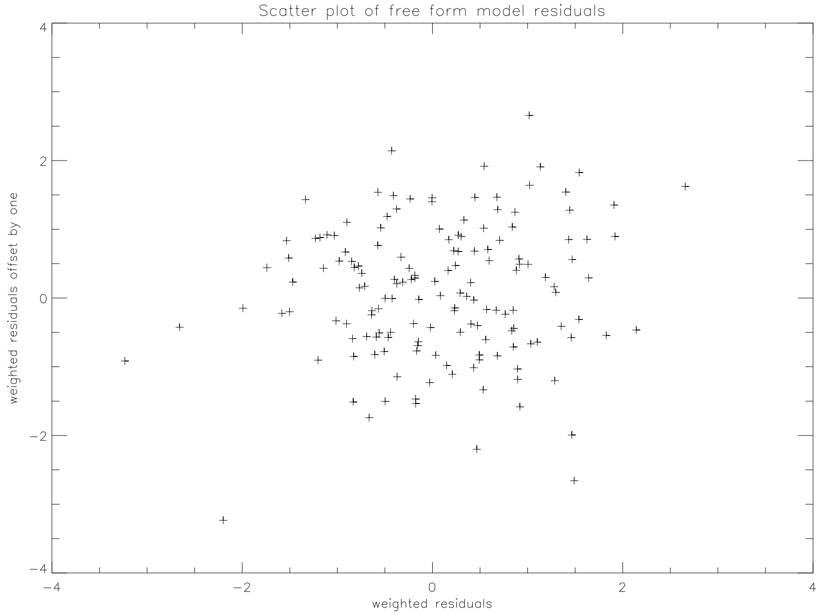


Figure 2.17: Scatter plot free form model residuals showing no correlation.

free form model has a higher peak density than the three and four parameter Chapman profiles. The results seem reasonable in comparison to the Ionosonde measurements.

2.7 Conclusions

After examining the results of the three different model parameterizations it seems apparent that they all fit the model well enough to pass the χ^2 test easily. Also, they all seem to agree closely to their corresponding Ionosonde measurements. The most efficient model is the three parameter Chapman function. It was the most efficient in the sense that all the F tests performed showed no significantly better fit when more parameters were used. This does not mean that only the three parameter Chapman model should be used. It is easy to

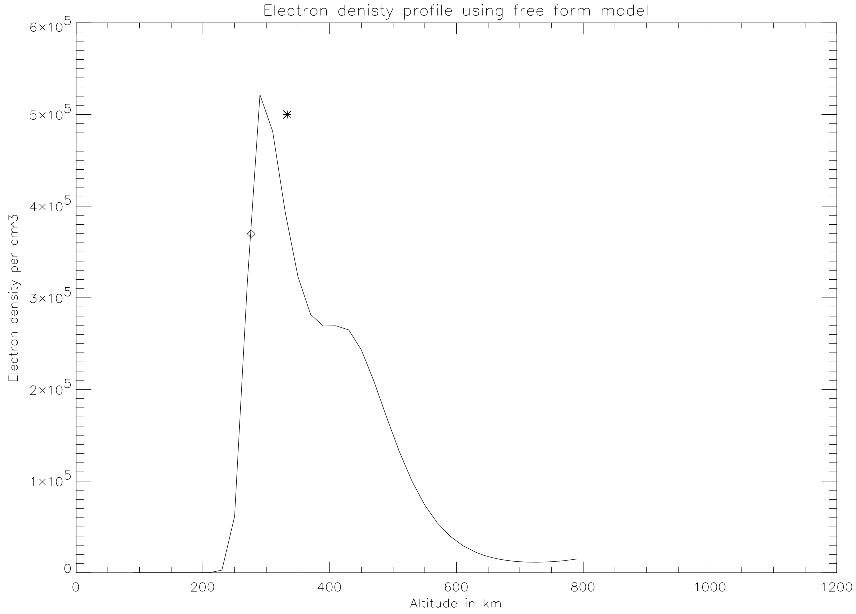


Figure 2.18: Free form electron density profile with Millstone Hill results.

switch to more parameters, if the programs are written properly, in order to test results on other data sets.

The free form model does have its advantages in that fluctuations in the density can be found that could never be found with the Chapman function. It is important to remember here that the Ionosphere is assumed to be homogeneous along latitudes and longitudes. For this reason, we only attempted to fit profiles that were in the mid-latitude regions. This does not mean that profiles cannot be fit at or near the equator. It is unclear at this point if the results would be accurate. The next Chapter attempts to fit a two dimensional model where by we do not make these assumptions. Then, results of the one-dimensional model fit at the equator will be compared to the two-dimensional model.

Chapter 3

The Two Dimensional Model

3.1 Overview

In exploring the one-dimensional model, data sets from LORAAS were used to accurately obtain a vertical profile of the Ionosphere as measured against ground truth. The line of sight of the LORAAS instrument cuts across many degrees of latitude and longitude. For this reason, obtaining a vertical profile using the Chapman function relies on the assumption that the latitude-longitude gradient of the electron densities is small. There was concern that fitting profiles at or near the equator might not be very accurate since gradients in the tropics are generally higher. Our interest lies in finding out if we can obtain a two-dimensional picture of the Ionosphere that spans across the mid-latitude and equatorial regions of the nighttime sky. Densities found in the two-dimensional model will be compared to those obtained by the one-dimensional model to see if they agree.

The problem is really a three-dimensional one (latitude, longitude, and altitude), but since the limb scans are in the orbit plane of ARGOS, it reduces to a two-dimensional problem. The algorithm used in the two-dimensional model is the same as in the one-dimensional case except that a two-dimensional grid of the ionosphere is maintained that contain electron densities. The area the grid covers is determined by the size of the data set used.

The data sets used are a series of consecutive limb scan profiles (unaffected by auroral precipitation) that span the mid-latitude and equatorial regions coinciding with the ARGOS orbit path. The grid is evenly spaced in angle along the orbit plane which roughly corresponds to a similar spacing in latitude. The number of elements in the grid array depends on the angle spacing specified by the user and the number of altitudes specified by the program. The number of altitudes is set at fifty ranging from 90 km to 1070 km in 20 km increments.

The elements in the grid have corresponding values of altitude, latitude, and longitude calculated from the orbit path of ARGOS. These corresponding values are saved in other arrays. If a grid spacing of ten degrees along the orbit plane is used, then each column of the grid roughly corresponds to one limb scan made by LORAAS. This is because ARGOS moves 10 deg in its orbit during each complete limb scan. A more or less dense grid can be created by changing the grid angle spacing value. The trade off is that the denser the grid the more parameters we have to fit. If we make the grid twice as dense in orbit angle spacing then we double the number of parameters. The iterative algorithm has to take the derivative with respect to each parameter. This means evaluating the forward function twice as many times. It was found that a grid density of ten degrees is sufficient for proper resolution of the problem.

The electron densities at each grid angle, an array with a density at each altitude, is parameterized as in the 1-D case with a Chapman function. Thus, there is also a parameter grid containing parameter values for the Chapman function that correspond to the columns (or angle spacing) of the electron density grid. These parameter values are adjusted so that the corresponding electron densities they calculate for the electron density grid are optimal in fitting the data. Since the data are measurements of intensities, the electron densities in the grid are used to calculate the intensities as would be observed by LORAAS. The formulas used for the intensities are the same as are used for the 1-D problem except that the two-dimensional grid of electron densities is used. This grid has varying electron densities across the orbit angles instead of the constant densities, used in the 1-D problem. This gives a more accurate description of the ionosphere especially where the electron density is changing rapidly along the orbit plane.

Another two-dimensional array is needed of the nitrogen and oxygen densities, and temperatures corresponding to the geometry of the electron density grid. The MSIS-86 model described in the one-dimensional problem returns an array that contains these values at each point in the grid. Once the grids are created a good set of initial values are needed. The initial values used for the one-dimensional problem are not sufficient here since the Chapman parameters vary considerably across the mid-latitude to equatorial regions.

The initial values used for the two-dimensional problem are obtained from a function called the Parameterized Ionospheric Model (PIM) [Dan 95]. The PIM is a global model of the theoretical ionospheric climatology based on diurnally reproducible runs of four physics based numerical models of the ionosphere. The four numerical models, taken together, cover the E and F layers for all latitudes, longitudes and local times. PIM consists of a semi-analytic representation of diurnally reproducible runs of these models for low, moderate, and high levels of both solar and geomagnetic activity, and conditions during June and December solstice, as well as the March equinox. The output obtained from the PIM used in our algorithm is of the form of a latitude and longitude grid in the region of the flight path of ARGOS. It returns a grid of electron densities in the F layer of that region. For more information on the PIM function see [Dan 95].

The grid of electron densities obtained from the PIM is inverted to obtain optimal Chapman parameters for each grid angle along the orbit path. These are used as the initial parameter values. The initial values fit include all four Chapman parameters. When we run the three parameter Chapman model the fourth parameter (the slope of the top side scale height) is fixed at this initial value.

The next major difference between the one and two dimensional models is the calculation of the intensities. In the one dimensional model the electron, nitrogen, and oxygen densities are considered constant along latitudes and longitudes. In the two-dimensional model the differing densities are found in the grids, so when the intensities are calculated they reflect the variable electron densities at differing latitudes and longitudes. Also, there is a certain amount of interpolation going on so intensities calculated along a line a sight come from a bilinear interpolation between points in the electron density grid.

Since the forward function as a whole is very flat near the optimal solution the problem is ill-conditioned. This leads to a very rough and physically unrealistic solution. This roughness will be seen when results of three and four parameter runs are presented.

Once the best two-dimensional model has been determined regularization will be introduced to smooth out the results. The regularization used is described in section 3.3. At this point, any statistical inference about the fit is lost including confidence intervals, since the bias introduced into the solution will be too great.

As different data sets were looked at, it was observed that the method used to subtract out the background in the one-dimensional model did not always work in the two-dimensional case. As mentioned, it was observed that the given ratios of 1356Å and 911Å to 1216Å and 1304Å contamination subtracts off too much of the 1216Å and 1304Å counts. The way to correct this is to adjust these ratios so that (when the data set is plotted) the section of data which is expected to be nearly zero should have just as many data elements randomly distributed above and below zero. Before the algorithm is run the data set under consideration is examined to make sure that all contamination by stars are excluded and that the subtracted background noise is correct.

In the two-dimensional model, where a series of limb scan profiles make up the data set, the correct ratio was not constant across the entire data set. The baseline of the data from each profile seemed to fluctuate up and down time wise through the data set. It was discovered, upon inquiry to the Naval Research Lab, that there was a “Hot Spot” on the detector that could artificially increase the apparent background. The way around this is to use the counts obtained from nearby bins (of the 911Å and 1356Å emissions) as the background. There are no reactions taking place in the nighttime sky that contribute anything to the nearby bins on the detector other than background noise. Making the necessary changes to the procedure and plotting the data, the background was much more reasonable for the previously troubled data sets. We can confirm this fact by adding in to the forward function, temporarily, two extra parameters that were used to find the multiplier of background noise that would give the best fit to the data. These multiples stayed very close to one, indicating that the sum of the counts of the nearby bins to the 911Å and 1356Å frequencies served well as the entire background noise.

Another variation to the algorithm inspired by trouble with the background, was the idea to leave the background in. Here, the idea is to take the total counts in the bins corresponding to 911Å and 1356Å emissions (including the background) and use these to find the optimal Chapman parameters that, when the background is added to the retrieved data, best fit the total counts. There is another advantage here, other than avoiding trying to fit negative data, and that is the resulting distribution. Before, Poisson total counts were subtracted from Poisson background noise. The resulting distribution is not Poisson. If the background is left in the original counts, the result is a purely Poisson distribution with standard deviations estimated as the square root of the counts. It is necessary to smooth the background noise added in after the retrieved data is calculated from the Chapman function to obtain an acceptable fit. The smoothing is done by fitting a polynomial curve through the background data points using the IDL function polyfit. From these arguments it is obvious this method makes much more sense and the one-dimensional algorithm has since been modified to work this way. The retrieved parameters were close to being the same using either technique for the one-dimensional model.

Convergence of the two-dimensional model was heavily tested and the printouts are extensive[1]. What they reveal are that the first three parameters can be randomly adjusted up and down by 35% of the their initial values and still converge to the same minimum objective function value, as well as the same parameter values. The fourth parameter (the slope of the top side scale height) is much more sensitive and could only be adjusted up and down by 15% of its initial value before the algorithm would start converging to different objective function values. The good news is that the other objective function values the algorithm converged to (when the parameters where adjusted outside the limits just described) were always larger than the minimized objective function value obtained from the PIM initial values.

3.2 The Three and Four Parameter Chapman Model

The Chapman function, described in Section 2.2, is given by equation 2.7. In this section, we will explore results from the optimization of a grid of three of the the four parameters in equation 2.7 using the algorithm described in section 2.2 and in the overview of this chapter. The fourth parameter is the slope of the scale height and it will remain fixed at the value obtained from inverting the PIM. To help find a more precise parameterization of the data, 911Å and 1356Å emissions will be fit simultaneously. The objective function we will be minimizing for two dimensional problem is

$$\| \text{ weighted residuals } \|_2^2 = \sum_{i=1}^{n} \text{residuals}_i^2/\sigma_i^2 + \text{penalty}^2 \tag{3.1}$$

---

[1] Listings of the convergence results are available from the author or Dr. Borchers

where the residuals are the differences between actual data and the retrieved data, sigma is the standard deviation of each data point taken as the square root of the full count and n is the number of data points. The penalty data point is always zero unless the algorithm tries to take a descent step that puts a parameter value outside imposed limits. The penalty data point is assigned such a high value that the step is never taken. The step is completed when a proper direction of descent is found, then the penalty point is put back to zero.

The first data set investigated is an arbitrarily chosen series of scan profiles selected from a data set obtained from LORAAS on November 24, 1999. The section of data we will analyzed corresponds to tangent point latitudes and longitudes along the ARGOS flight path ranging from 70.53 deg N - 145.73 deg E to 47.33 deg S - 109.46 deg E . The Universal times corresponding to these measurements start at 17:35 UT and end at 18:09 UT. A total of 52 data points were excluded from this set of 2156 because they were found to be contaminated by star light. This works out to be about two points per limb scan per spectra, the number that generally had to be removed when fitting the one-dimensional model. There is no need to be concerned here with small count values since the total counts are used, including the noise. The smallest values found in the data sets start at around 10 counts per bin and go up to 150 counts per bin. These values from a Poisson distribution are generally considered to approximate a normal distribution well enough, especially when extended to the residuals.

Figure 3.1 shows the two dimensional picture of the electron densities retrieved from the 3-parameter Chapman model. This solution is very rough as can be seen by the very sharp drop offs from the peaks of about 9×10^5

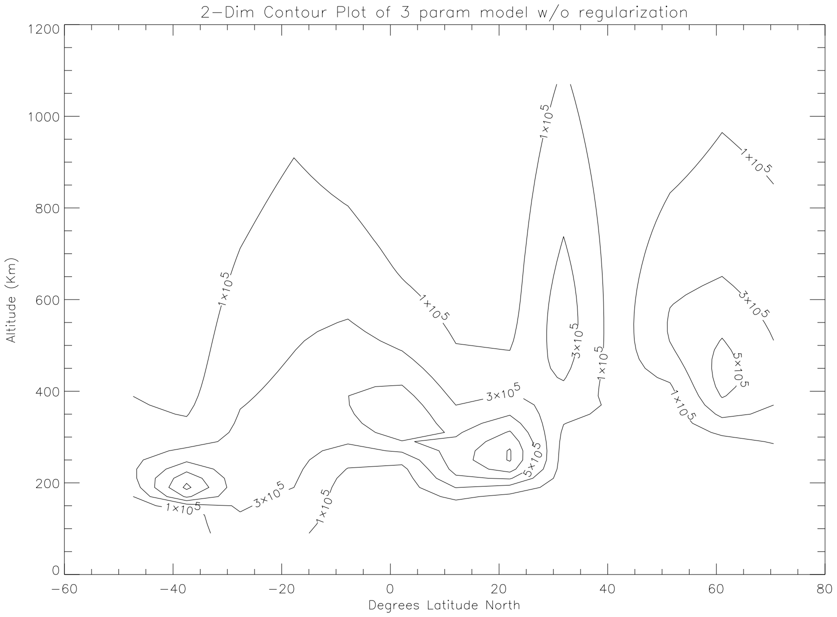


Figure 3.1: Two dimensional contour plot showing the retrieved electron densities from the 3-parameter Chapman function without any smoothing of the solution. The peaks represent electron densities of 9e5 .

hmF2(Km)	nmF2 (per cm^3)	H_{top} (Km)	H_{con}
428	339948	95	1.06
438	580791	79	1.06
541	316860	76	1.02
359	77016	10	1.00
545	410956	129	1.02
259	931514	41	1.00
251	598850	52	1.02
339	663877	56	1.03
383	499563	88	1.02
311	373789	154	1.02
230	361167	130	1.01
190	949267	29	1.02
227	276658	54	1.07

Table 3.1: Optimal parameter values for the two-dimensional Chapman model. The H_{con} parameter remains fixed.

electrons per cm^3 down to 7.7×10^4 per cm^3 in only 20 deg of latitude. Also, some of the parameters are unrealistic, for instance, the top side scale heights that fluctuate wildly. The results will be smoothed out using regularization later in this chapter.

It is interesting to note when looking at Figure 3.1 that there are two peaks on both sides of the 10 deg latitude mark. It is known that electron densities between the mid-latitude regions peak just above and below the magnetic equator. It can be verified at the NOAA national data centers, NGDC web page that the magnetic equator during that time and at that position longitude (120 degE) was at 10 degN latitude. This is further evidence that the model represents (albeit rough) a realistic picture of the ionosphere. The peak at -40 degN latitude is really at -50 degN latitude in relation to the magnetic equator. This is still not that far south and makes this peak very suspect. It probably is an effect of the unregularized two dimensional fit.

The χ^2 value from this fit was 1.0049 with 2065 degrees of freedom. The corresponding p-value is 0.4335 indicating a good fit. We have already discussed the bias that is introduced into the solution merely by the fact that interpolation is used to maintain the grids and that the lines of sight being fit overlap. This overlap has a tendency to pull individual profiles away from there optimal fit if doing so causes neighboring profiles to obtain a better fit which improves the objective function value to a greater degree. For this reason, we do not expect our residuals to show the degree of normality that they do in the one-dimensional case. Figures 3.2 and 3.3 show that the residuals are randomly distributed but they do not pass the A.D. normality test anymore. Figure 3.4 shows a plot of the histogram of the residuals with a normal curve overlay. It can be seen that the residuals are approximately normal but look slightly skewed to the right due to bias.

Using the algorithm described in section 2.2 and in the overview of this chapter the results from the optimization of a grid of all four parameters in equation 2.7 will be looked at. The fourth parameter is the slope of the scale height and it will be allowed to vary from the initial value obtained from inverting the PIM. For comparison the same data set will be used as was used for the three parameter two-dimensional model. Every thing else will remain the same in the algorithm.

The χ^2 value from this fit was 1.0053 with 2052 degrees of freedom. The corresponding p-value is 0.4286 indicating a good fit. The residuals looked

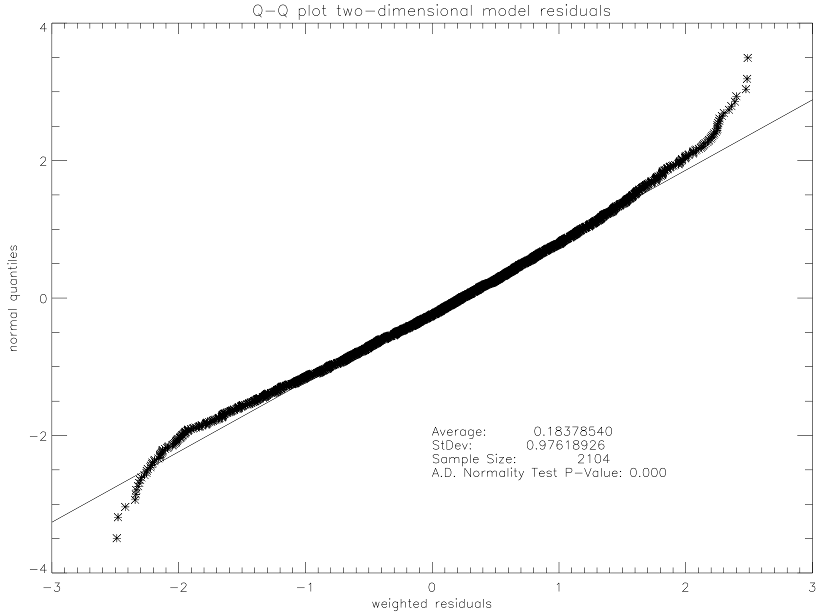


Figure 3.2: Q-Q plot of 2-D model residuals.

the same as the three parameter model. The nmF2 values where not as erratic as the three-dimensional model but the $\mathrm{H}_{\mathrm{con}}$ parameter must have absorbed the effects of the bias since their values where very erratic and unrealistic. The question is, is this an improvement in the model? The answer is no. The ratio R of reduced χ^2 values for this data set is $1.0049/1.0053 = 0.9996$. The upper 5% critical value $F_{.95}(2065, 2052) = 1.0752$. This indicates the extra parameters do not have a statistically significant effect on the fit at the 5% level. This result is consistent with the one-dimensional model.

3.3 Regularization of the Best 2-D Model

The best two dimensional model is the grid formed when three of the four Chapman parameters are allowed to vary. The fourth parameter (slope of

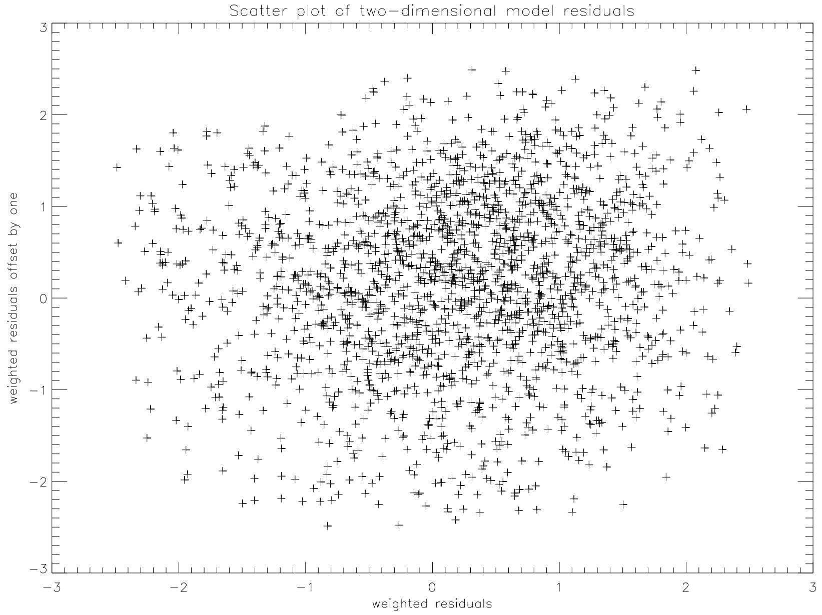


Figure 3.3: Scatter plot free form model residuals showing no correlation.

the top side scale height) remains fixed at its initial value obtained by inverting the PIM.

Now that the appropriate model has been chosen, it is necessary to smooth out the unregularized solution. In this case, what is meant by regularize is to stabilize the solution so that the transition between columns in the grid of ion densities is more gradual than the rough solution. It is also necessary to keep the parameter values somewhat close to there initial values and thereby keeping the parameter values within physically realistic limits. Without these two types of regularization, the penalty function is invoked on nearly every iteration, and as seen in table 3.1 the parameter values turn out to be erratic. The smoothness and stability of this type of regularization is obviously reflected

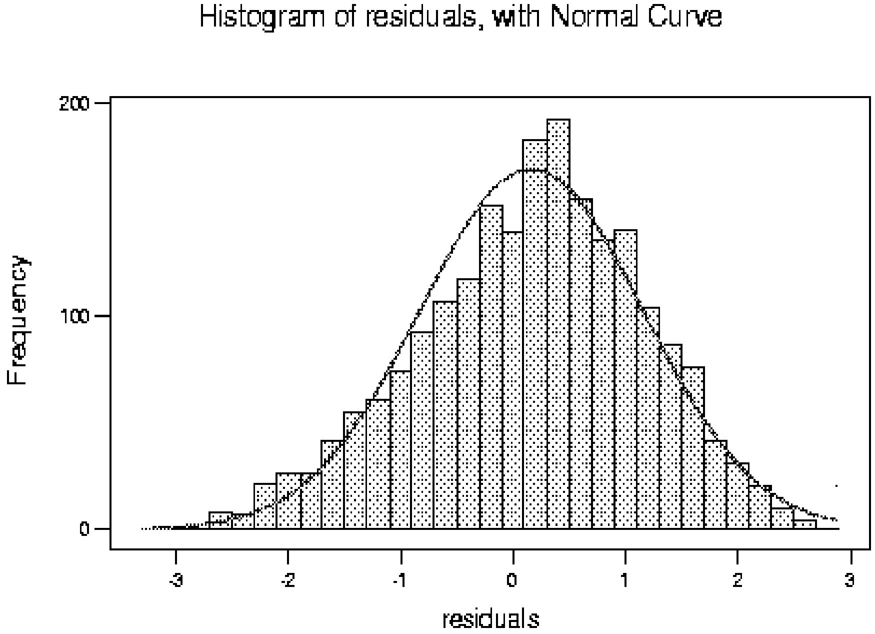


Figure 3.4: Histogram of residuals with normal curve overlay.

in the smoothness of the grid of parameter values.

The way regularization is imposed on our solution is to add two new terms to the objective function. The first new term added is a measure of the distance the parameters are getting from their initial values. The second new term is a measure of the roughness between columns in the grid of ion densities. The way the roughness in the grid of ion densities is calculated is an estimate of the second derivative between three consecutive points (at the same altitude in the grid) is taken and then normalized by the total O^+ value obtained by the PIM at that altitude. This is done at different altitudes staggered from top to bottom in the ion density grid. This array of derivative estimates is reshaped into a vector. Minimizing the norm of this vector has the effect of finding optimal values for our parameters that keep the grid of ion densities smooth. The new objective function is now

$$\text{minimize } \parallel \text{residuals}_w \parallel_2^2 + \alpha^2 (\parallel \text{m} - \text{m}_{\text{PIM}} \parallel_2^2 + L^2 \parallel \text{s} \parallel_2^2) + \text{penalty}^2 \tag{3.2}$$

where residuals_w is the vector of misfit measurements divided by their respective data point standard deviations, m represents the optimal parameter values, m_{PIM} are the initial parameter values obtained from the PIM profile, s is the vector of roughness calculations, L is the weight or importance assigned to the roughness vector versus the distance our optimal parameter values get from their initial values, α is the weight or importance assigned to the sum of the two regularization terms versus the misfit, and penalty is the penalty data point that is invoked if the algorithm tries to assign parameter values outside the physically real boundaries. This last term is always zero at the end of any iteration.

After some trial and error it was found that letting $L = 3.5$ distributed the regularization norms so that the magnitude of each of the two terms were close to being equal. The program was then run for values of alpha ranging from 0.0 to 3.8. The square of the norm of the weighted misfit was plotted against the sum of the squares of the norms of the two regularization terms. The optimal value for alpha was chosen to be the sixth data point or $\alpha = 1.0$. This value of alpha is where the model is smoothed to an optimal point and further smoothing only adversely affects the misfit. The algorithm was very stable and converged in only six iterations.

Now, the results of our regularized two dimensional model can be observed. The reduced χ^2 value for this run was 1.09145, which fails the χ^2

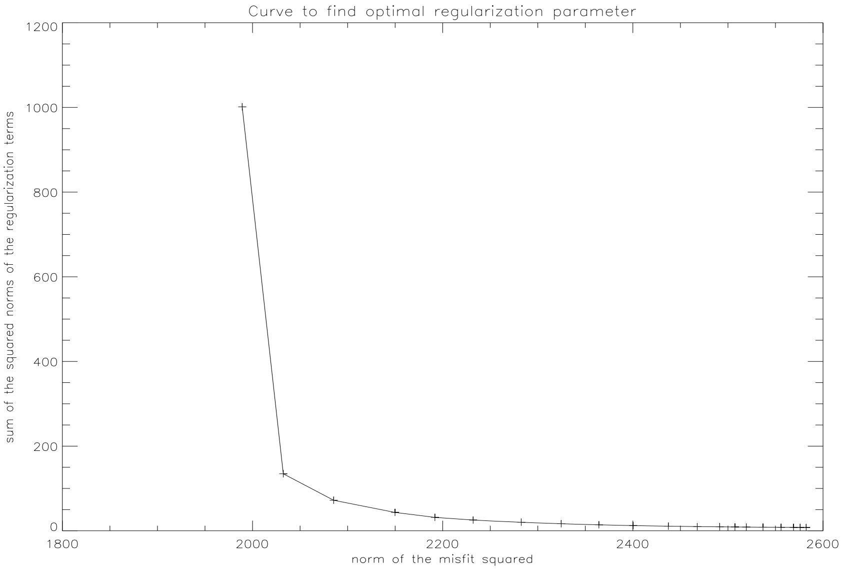


Figure 3.5: The regularization curve found by trying a range of values for alpha and plotting the norm of the misfit squared against the sum of the squared norms of the regularization terms .

test at the 5% level of significance with a p-value of 0.0098. This is to be expected when using regularization. The bias introduced into the solution when smoothing it out usually renders any accurate statistical analysis useless. However, the residuals still look very similar to the unregularized solution. Figure 3.6 shows the two dimensional contour plot of the smoothed solution. This solution looks very different from Figure 3.1. The suspicious peak at $-40\,$degN latitude is gone and there is a new peak forming at $70\,$degN latitude as the solution heads into the auroral region at the pole. One effect of the smoothing is that the two distinct peaks on either side of the magnetic equator have been lost. This has been blurred into one peak. By and large, this mapping of the ionosphere is far more realistic than the unregularized solution. Table 3.2 shows the final optimal parameter values that are used to construct the 2-D grid of ion densities. The observed values are also much smoother in their transition from limb scan to limb scan. Remember the chosen grid density corresponds to one set of parameters per limb scan.

A one-dimensional fit of the model at $21\,$degN latitude is compared by Looking at Figure 3.6 and row six of Table 3.2. It can see that the peak electron density is at 6.28×10^5 per cm^3 located at a height of 274 km. Fitting the one-dimensional limb scan data at this position yields almost the same results. The peak electron density is at 6.64×10^5 per cm^3 located at a height of 271 km. One can also fit the 1-D profile at $10\,$deg to the other side of the magnetic equator located at at $2\,$degN. These parameters are in row eight of Table 3.2 for the 2-D model. The optimal parameter values obtained from the 1-D model have the peak electron density at 5.15×10^5 per cm^3 located at a height of 375 km. These values are close to their corresponding parameter

hmF2(Km)	nmF2 (per cm^3)	H_{top} (Km)	H_{con}
493	455505	72	1.06
489	377412	68	1.06
492	263303	85	1.02
487	166294	95	1.00
282	216893	84	1.02
274	628492	42	1.00
284	638197	56	1.02
326	569919	73	1.03
365	484296	87	1.02
370	338856	120	1.02
327	304570	149	1.01
258	331307	155	1.02
249	390875	125	1.07

Table 3.2: Optimal parameter values for the regularized two-dimensional model using a three parameter Chapman function..

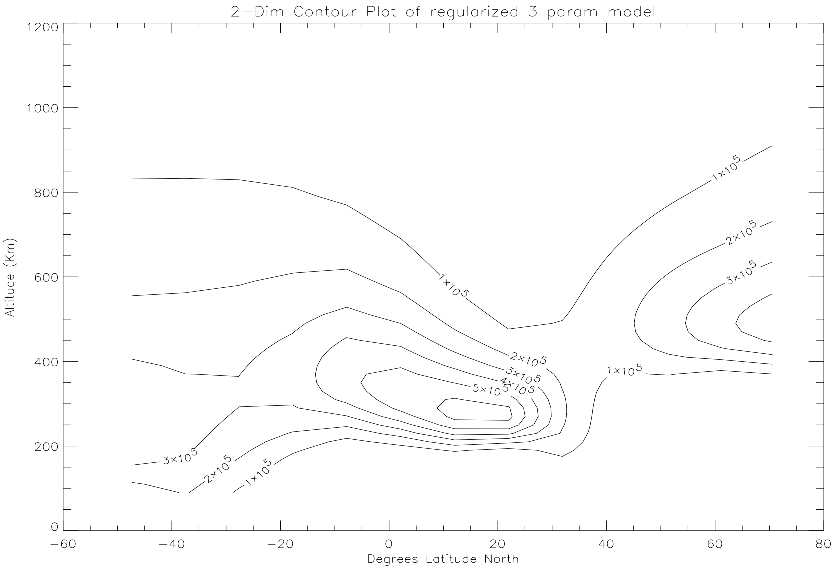


Figure 3.6: Two dimensional contour plot showing the retrieved electron densities from the 3-parameter Chapman function with optimal smoothing of the solution.

values from the 2-D model.

It is important to point out that there is a disagreement between the third and fourth rows of Table 3.2 to the corresponding 1-D results. The corresponding 1-D results (which were presented in detail in Chapter Two) show the peak electron density at 5.7×10^5 per cm^3 located at a height of 257 km. One reason for this may be that, according to the two-dimensional model, the gradient in the 30 degN to 50 degN latitudes is high for this data set. This could explain why the ion density is so much higher in the 1-D model. The 1-D model assumes the atmosphere is homogeneous and therefore any measurements that might be influenced by the line of sight stretching into regions where the ion density is higher (far away from the tangent point altitude) will be used as if they where from readings at the tangent point latitude. This can lead to peak ion densities that are much higher than they should be. One way to check this is to take the part of the solution from the 2-D model corresponding to the fourth row of Table 3.2 and use it as the data set to fit in the 1-D model. If measurements have been effected by high ion densities that are very close or very far from the detector, with respect to the tangent point altitude in the line of sight of LORAAS, then the peak ion density found by the 1-D model should be much higher than the 2-D model.

After substituting the solution from the latitude of interest obtained from the 2-D model into the 1-D model the result is the peak electron density is $5.36 \times 10^5/\mathrm{cm}^3$. This result indicates that the 1-D model is being affected by the high ion density gradient in that region and the results are not to be trusted. The 2-D model is able to detect the valley in the electron density by way of measurements taken before and after the region of interest.

The results of the 2-D model are still not perfect. They are obtained from regularizing the raw solution in an inexact way. The electron densities are calculated by interpolation between points in the ion density grid and corresponding intensities are derived using oxygen levels obtained from the MSIS model. These electron densities are therefore a sort of averaging of the real ionosphere with the peaks and valleys slightly exaggerated.

Chapter 4

Conclusions

We can conclude from the two dimensional results that the best model to use in calculating the ion density grid is the three parameter Chapman model. The unregularized solution obtained is too rough to be physically realistic and thus must be smoothed out using regularization.

One reason we explored the two-dimensional model was that we were concerned that the one dimensional model might not give us accurate parameter values at the regions in the ionosphere where the gradient of the ion density is high. Upon comparison we can see that the one and two dimensional models give different results in these regions. Due to the geometry of the problem, the best method is one that reflects changes in ion density in the orbit plane of LORAAS.

An experiment was run that constructed a two dimensional mapping of the ionosphere by fitting consecutive limb scans using the 1-D model. Figure 4.1 shows the results of the experiment. It shows a peak density of 5×10^5 consistently across latitudes ranging from -20 N to 70 N. There is a slight peak at 20 N corresponding to the 10 deg mark north of the magnetic equator. At latitudes of -50 N to -30 N the algorithm failed to converge and gave results in excess of 12×10^5. The result is a blurred picture of the ionosphere with constant peak densities at altitudes that are unrealistic.

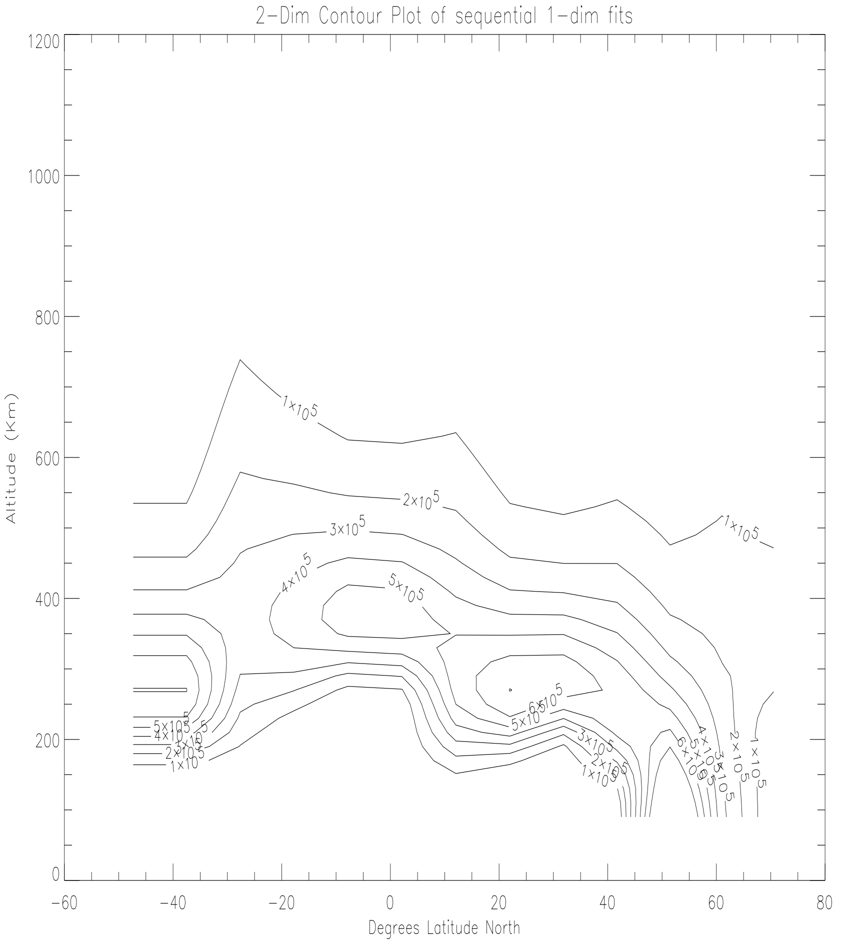


Figure 4.1: Two dimensional contour plot constructed by fitting consecutive 1-D profiles.

We have found that in some cases the one dimensional model works well and has advantages over the two-dimensional model in terms of faster run times and the ability to calculate accurate confidence intervals. However, we conclude here that the regularized two-dimensional model gives a much more accurate continuous picture of the state of the ionosphere.

References

[And 54] Anderson, T.W., D.A. Darling, A Test of Goodness of Fit, *J. Am. Statist. Assoc., 49:765-769*, 1954.

[Bar 70] Barth, C.A., and Schaffner, OGO 4 spectrometer measurements of the tropical ultraviolet airglow, *J. Geophysical Res., 75*, 4299, 1970.

[Bil 90] Bilitza, D., International reference ionosphere 1990, *NSSDW/WDC-A-R&S 90-20*, Natl. Space Sci. Data Cent., Greenbelt, Md., 1990.

[Bru 78] Brune, W.H., P.D. Feldman, R.C. Anderson, W.G. Fastie, and R.C. Henry, Mid-latitude oxygen ultraviolet nightglow, *Geophys. Res. Lett., 5*, 383, 1978.

[Cha 84] Chakrabarti, S., R. Kimble, and S. Bowyer, Spectroscopy of the EUV (350-1400 Å) nightglow, *J. Geophys. Res.,89*, 5660,1984

[Cha 87] Chamberlain, J.W., D.M. Hunten, *Theory of planetary Atmospheres: An Introduction to Their Physics and Chemistry*, Academic Press, Inc., Orlando, FL., 1987.

[Chn 75] Chandra, S., E.I. Reed, R.R. Meier, C.B. Opal, and G.T. Hicks, Remote sensing of the ionosphere F layer by use of O I 6300-Å and O I 1356-Å observations, *J. Geophys. Res., 80*, 2327, 1975

[Dan 95] Daniel R.E., L.D. Brown, D.N. Anderson, M.W. Fox, P.H. Doherty, D.T. Decker, J.J. Sojka, and R.W. Schunk, Parameterized Ionospheric Model: A Global Ionospheric Parameterization Based on First Principles Models, *Radio Science, Vol 30, Number 5, pages 1499-1510, Sep-Oct 1995*

[Dym 93] Dymond K.F., and R.P. McCoy, Ultraviolet Spectrograph for Thermospheric and Ionospheric Remote Sensing, *Proc. SPIE, 1940*, 117-127, 1993.

[Dym 97] Dymond K.F., S.E. Thonnard, R.P. McCoy, R.J. Thomas, An optical remote sensing technique for determining nighttime F region electron density, *Radio Science, Vol 32, Number 5, pages 1985-1996, Sep-Oct 1997*

[Dym 01] Dymond K.F., S.A. Budzien, S.E. Thonnard, R.P. McCoy, R.J. Thomas, Electron densities determined by inversion of ultraviolet limb profiles, *Geophys. Res. Letters, (submitted)*

[Dym 02] Dymond K.F., S.A. Budzien, S.E. Thonnard, R.P. McCoy, R.J. Thomas, Electron densities determined by the HIRAAS experiment and comparisons with Ionosonde measurements, *Geophys. Res. Letters, (submitted)*

[Dym 03] Dymond K.F., S.A. Budzien, S.E. Thonnard, R.P. McCoy, R.J. Thomas, Nighttime F-region electron densities determined by tomographic inversion of ultraviolet limb scans, *Journal of Geophys. Res., (submitted)*

[Dym 04] Dymond K.F., S.A. Budzien, S.E. Thonnard, R.P. McCoy, R.J. Thomas, Electron densities determined by inversion of ultraviolet limb scans, *Journal of Geophys. Res., (submitted)*

[Fel 92] Feldman P.D., A.F. Davidson, W.P. Blair, C.W. Bowers, S.T. Durrance, G.A. Kriss, H.C. Ferguson, R.A. Kimble, and K.S. Long, The spectrum of the tropical oxygen nightglow observed at 3-Å resolution with the Hopkins Ultraviolet Telescope, *Geophys. Res. Letters, 19*, 453, 1992.

[Fel 01] Feldman, P.D., D.J. Sahnow, J.W. Kruk, E.M. Murphy, and H.W. Moos, High resolution FUV spectroscopy of the terrestrial day airglow with FUSE, *Jrn. Geophys. Res.*, (Accepted 2001).

[Gen 79] Gentieu, E.P., P.D. Feldman, and R.R. Meier, Spectroscopy of the extreme ultraviolet dayglow at 6.5-Å resolution: Atomic and ionic emissions between 530 and 1240 Å *Geophys. Res. Letters,6*, 325, 1979.

[Gen 81] Gentieu, E.P., P.D. Feldman, R.W. Eastes, and A.B. Christensen, Spectroscopy of the extreme ultraviolet dayglow during active solar conditions, *Geophys. Res. Letters, 8*, 1242, 1981.

[Gen 84] Gentieu, E.P., P.D. Feldman, R.W. Eastes, and A.B. Christensen, EUV airglow during active solar conditions, 2, Emission between 530 and 930 Å *Jrn. Geophys. Res., 89*, 11053, 1984.

[Han 69] Hanson, W.B., Radiative recombination of atomic oxygen ions in the nighttime F region, *Jrn. Geophys. Res., 74*, 3720, 1969.

[Hed 87] Hedin, A.E., MSIS-86 Thermospheric model, *J. Geophys. Res. 92, 4649*, 1987

[Hic 70] Hicks, G.T., and T.A. Chubb, Equatorial aurora/airglow in the far ultraviolet, *Jrn. Geophys. Res., 75*, 6233, 1970.

[IDL 95] Research Systems, Inc., *IDL Reference Guide*, 1995.

[Knu 70] Knudsen, W.C., Tropical ultraviolet nightglow from oxygen ion-ion neutralization, *Jrn. Geophys. Res., 75*, 3862, 1970.

[Law 00] Law, Averill, W.D. Kelton, *Simulation Modeling and Analysis*, McGraw Hill, 2000.

[Mad 99] Madsen, K., H.B. Nielsen, O. Tingleff, Methods for Nonlinear Least Squares Problems, *IMM Department of Mathematical Modeling, No. H38 7.7.1999*

[McC 92] McCoy, R.P., K.F. Dymond, G.G. Fritz, S.E. Thonnard, R.R. Meier, and P.A. Regeon, Far and extreme ultraviolet limb imaging spectrograph for the DMSP satellites, *Proc. SPIE, 1745*, 310-321, 1992.

[McC 94] McCoy, R.P., K.F. Dymond, G.G. Fritz, S.E. Thonnard, R.R. Meier, and P.A. Regeon, Special sensor ultraviolet limb imager: An ionospheric and neutral density profiler for the Defense Meteorological Satellite Program satellites, *Opt. Eng., 33*, 423, 1994.

[McC 95] McCoy, R.P., K.F. Dymond, J.M. Picone, G.R. Caruthers, O.A. Kelley, and D.D. Cleary, Hyper-spectral imaging of the global ionosphere from the ARGOS satellite, paper presented at the International Symposium on Spectral Sensing Research, Int. Soc. for Photogramm. and Remote Sens., Melbourne, Australia, Nov. 26 to Dec. 1, 1995.

[Mel 99] Melendez-Alvira, D.J., R.R. Meier, J.M. Picone, P.D. Feldman, and M. Mclaughlin, Analysis of the Oxygen nightglow Measured by the Hopkins Telescope: Implications for the Ionospheric Partial Radiative Recombination Coefficients, *Journal Geophysical Research, 104, 14901*, 1999.

[Men 89] Menke, W., *Geophysical Data Analysis: Discrete Inverse Theory, International Geophysics Series, No. 45*, Academic Press, San Diego, 1989.

[Mei 91] Meier, R.R., Ultraviolet spectroscopy and remote sensing of the upper atmosphere, *Space Sci. Rev., 58*, 1991.

[Mye 90] Myers, R.H., *Classical and Modern Regression with applications, 2nd edition*, Duxbury Press, 1990.

[Nas 96] Nash, S.G., A. Sofer, *Linear and Nonlinear programming*, McGraw-Hill series in industrial and management science, 1996.

[Pre 92] Press, W.H., B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, New York, NY, 1992.

[Ree 89] Rees, M.H., *Physics and Chemistry of the Upper Atmosphere*, Cambridge University Press, 1989.

[Ste 74] Stephens, M.A., EDF Statistics for Goodness of Fit and some Comparisons, *J. Am. Statist. Assoc., 69: 730-737*, 1974.

[Tin 72] Tinsley, B.A., O I and N I allowed transitions in the airglow and aurora, *Ann. Geophys., 28*, 155, 1972.

[Tin 75] Tinsley, B.A., and J.A. Bittencourt, Determination of F region height and peak electron density at night using using airglow emissions from atomic oxygen, *Jrn. Geophys. Res., 80*, 2333, 1975.

[Tin 73] Tinsley, B.A., A.B. Christensen, J. Bittencourt, H. Gouveia, P.D. Angreji, H. Takahashi, Excitation of oxygen permitted line emissions in the tropical nightglow, *Jrn. Geophys. Res., 78*, 1174, 1973.

[Tho 99] Thonnard, S.E., S.A. Budzien, A.C. Nicholas, and K.F. Dymond, An update on the calibration and performance of the Special Sensor Ultraviolet Limb Imager (SSULI's), *SPIE Vol. 3818, Ultraviolet Atmospheric and Space Remote Sensing: Methods and Instrumentation II*, 90, 1999.

This thesis is accepted on behalf of the faculty of the Institute by the following committee:

Advisor

Date