Retrieval of global magnetospheric ion distributions from high energy neutral atom (ENA)measurements made by the IMAGE/HENA instrument

R. DeMajistre, E.C. Roelof, P. C:son Brandt, D. G. Mitchell

The Johns Hopkins University Applied Physics Laboratory, Laurel MD

Abstract. In this work we present a method for retrieving global magnetospheric ion distributions from Energetic Neutral Atom (ENA) measurements made by the IMAGE/HENA instrument. The technique itself is based on the well established method of constrained linear inversion. In addition to presenting the technique itself, we present a self consistent method for its optimization. The combination of the technique itself and the tuning process provide a direct measure of the fidelity of the resulting retrievals. We apply these techniques to a representative HENA image and show that the retrieval is indeed a useful tool in quantifying global magnetospheric ion distributions.

1. Introduction

Energetic Neutral Atom (ENA) imaging is currently establishing itself as a powerful method for global characterization of the Earth's magnetosphere. The IMAGE spacecraft is now carrying a payload of dedicated ENA instruments (LENA, MENA, and HENA) that have the capability of measuring ENA imagery over a very broad range of energies. Each image contains information about the entire magnetosphere, in contrast to charged particle measurements which are essentially *in situ* measurements. The power of ENA imaging has been recognized for several years (for example see *Roelof* [1987]), but with the launch of IMAGE, its utility is now being more fully demonstrated.

Each ENA image contains a great deal of information about the magnetosphere, specifically about the ion distributions that serve as a source of ENA flux via charge exchange in the geocorona and near the Earth's exobase. The information in the images, however, can be difficult to interpret directly without substantial experience. Clearly, quantitative techniques for extracting and presenting this information is required for ENA images to be truly useful.

Over the past few years, several authors have made significant progress in extracting this information from ENA images. *Roelof and Skinner* [2000] presented a non-linear parametric retrieval algorithm for the extraction of ion distributions. This method makes use of global functions representing the shape of the ion distribution in the magnetosphere. The free parameters of these functions are used in a non-linear least squares fitting algorithm to provide an optimal fit to the data. The basic weakness of this method is the global nature of the basis functions. That is, with a reasonable number of free parameters, the parametric model is not able to reproduce all the shapes that the ion distributions can take. In short, only solutions that can be represented by the basis functions will result from the retrieval, but the basis functions cannot sufficiently represent the complexity of the ion distributions being sought.

Another approach, which is similar in character to the one use in this work, has been described in *Perez et al.* [2000] and *Perez et al.* [2001]. This method uses cubic B-splines as local basis functions in a constrained linear retrieval. Using local basis functions, i.e., functions that are nonzero in a very restricted spatial area, allows a much broader range of spatial structure than the global functions of *Roelof and Skinner* [2000]. This technique has resulted in reasonable retrievals of both simulations and measured imagery (*Perez et al.* [2000] and *Perez et al.* [2001], respectively), though a systematic analysis of the accuracy of the results have not been published. Our work differs from that of *Perez et al.* in three important ways. First, we incorporate the effects (such as the angular response function) of the HENA instrument directly into the retrievals. This eliminates much of the systematic error in the retrievals that arise from instrument effects. Further, any evaluation of the accuracy of retrieval techniques must account for instrumental biases. Second, our model of ENA emission includes a component from charge exchange with atomic oxygen near the exobase. Third, we have adopted a lower order (linear) set of basis functions. In adopting a linear scheme, we have allowed for a wider range of spatial structure and have simplified the construction of different types of constraints (or so called penalty functions).

In section (2) we give a brief description of the HENA measurements. We develop the basic equations governing these measurements in section (3). Images simulated using the measurement equations are shown in (4). Our retrieval method is developed and evaluated in (5).

2. HENA Measurements

The HENA instrument is aboard the IMAGE spacecraft, which is in an eccentric near polar orbit (1000 km x 7 R_E altitude). IMAGE is a spin stabilized platform, whose spin axis is normal to the plane of the orbit with a nominal spin period of 2 minutes (see *Gibson et al.* [2000] for a more detailed description of the IMAGE spacecraft). The HENA instrument collects Energetic Neutral Atoms (ENAs) from

Paper number .
0148-0227/02/ $9.00

roughly 20-200 Kev in a rectangular area of 120° by 90°. The instrument is mounted such that the center of the 120° dimension lies in the orbital plane (i.e., perpendicular to the spin axis). Thus in the course of a single spin, the instrument views a 120° by 360° segment of the sky (see *Mitchell et al.* [2000] for a more detailed description of the HENA instrument). ENAs measured during a spin period are separated into energy channels and placed in spatial bins that are regular in the ϵ and β angles described in Figure 1.

For energies below 60 Kev, the images have a pixel size of roughly 6°x6°. At higher energies the pixels are 3°x3°. The HENA instrument counts individual ENAs, and under most circumstances has a negligible background count rate. For this reason, we assume that the measurements in each pixel obey Poisson counting statistics.

3. Equations of Measurement

The counts in each pixel of an ENA image, C_i can be represented by the measurement equation

$$C_i = \int_0^{\infty}\int_0^{\infty}\int_0^{2\pi}\int_0^{\pi} \sin\epsilon A_i(\epsilon,\beta,E,t) j_{ena} d\epsilon d\beta dE dt \qquad (1)$$

Where A_i is the response of the pixel i to and ENA intensity j_{ena} at time t, energy E and angular position ϵ and β. The ENA intensity from charge exchange can be written (see *Roelof* [2002]) as

$$j_{ena} = \int_0^{s_e} n^H(s)\sigma_H^{10}(E) j_{ion}(s,E) ds + j^e_{ena}(s_e) \qquad (2)$$

where s is the distance along the line of sight determined by ϵ and β, n^H is the number density of hydrogen, σ_{10}^H is the charge exchange cross section for protons on hydrogen and j_{ion} is the ion intensity. The limit of integration, s_e, is either the point where the line of sight first intersects the exobase, or $+\infty$ for lines of sight with no such intersection. In writing equation (2) we have separated the ENA intensity into two distinct components, the optically thin geocorona and the optically thick emission from the region near the exobase, $j^e_{ena}(s_e)$, which is described in subsection (3.3). In the geocorona we assume charge exchange with hydrogen atoms only. The emission near the exobase is modelled as a surface emission and only charge exchange with atomic oxygen is considered (see *Roelof* [2002]). For convenience, we can further separate the counts from the geocorona and the exobase

$$C_i = C_i^{high} + C_i^{low} \qquad (3)$$

where C_i^{high} are the counts from geocoronal emissions

$$C_i^{high} = \int_0^{2\pi}\int_0^{\pi}\int_0^{s_e} n^H \sigma_H^{10} \sin\epsilon < A_i(\epsilon,\beta) > j_{ion} ds d\epsilon d\beta \qquad (4)$$

and C_i^{low} are the counts from near the exobase

$$C_i^{low} = \int_0^{2\pi}\int_0^{\pi} \sin\epsilon < A_i(\epsilon,\beta) > j^e_{ena} d\epsilon d\beta \qquad (5)$$

Each of these contributions will be examined in the subsections 3.2 and 3.3.

In equations (4) and (5) we have made the substitution

$$< A_i(\epsilon,\beta) >= \int_0^{\infty}\int_0^{\infty} A_i(\epsilon,\beta,E,t) dE dt \qquad (6)$$

In making this substitution we have assumed that a particular pixel, i, has been collected for a brief interval of time $\triangle t$ and a narrow energy range $\triangle E$, and that these intervals are suitably narrow that j_{ena} can be considered constant over them. The instrument function $< A_i(\epsilon,\beta) >$ appropriate for HENA has been studied and presented in *Roelof* [2002].

3.1. Description of the Ion Intensity

In all that follows, we assume that over the collection time $\triangle t$, the ion intensity, j_{ion}, can be uniquely represented in a dipole coordinate system, i.e.,

$$j_{ion} = j_{ion}(L,\phi,\mu_{eq}) \qquad (7)$$

This system is closely related to the Solar Magnetic coordinate system (*Russell* [1971]) represented in the spherical coordinates (r,θ,ϕ). In the dipole representation $r = L\sin^2\theta$ (measuring distance in earth radii R_e) and μ_{eq} is the cosine of the angle of the velocity vector to the field line (the so-called pitch angle cosine) at the equator. In writing equation (7) we have assumed that the ions obey the adiabatic invariants, that is, the ions are constrained to follow field lines and that the pitch angle cosine at any point in an ion's trajectory can be determined from its equatorial pitch cosine by

$$1-\mu^2 = \frac{B(\theta)}{B(\pi/2)}(1-\mu_{eq}^2) \qquad (8)$$

or exploiting the dipole shape of the magnetic field

$$\mu^2 = 1-(1-\mu_{eq}^2)\frac{\sqrt{4-3\sin^2\theta}}{sin^6\theta} \qquad (9)$$

where $B(\theta)$ is the magnitude of the magnetic field (at constant L) for a given θ.

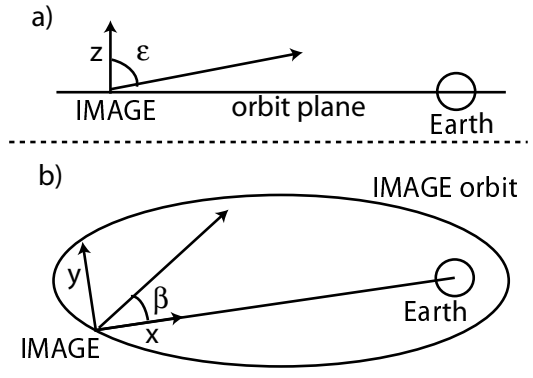

Figure 1. Geometry of HENA observations from a) within the plane of the orbit and b) above the plane of the orbit. The x axis points along the vector from the IMAGE spacecraft to the center of the Earth and the z axis points in the direction of the spacecraft spin axis. The angles ϵ and β are the polar and azimuthal angles in this system.

A consequence of equation (7) is that the pitch angle distribution is symmetric about $\mu_{eq} = 0$. That is, if α_{eq} is the equatorial pitch angle, then $\mu_{eq} = \cos(\alpha_{eq}) = \cos(-\alpha_{eq})$. Further, in equation (8) both μ_{eq} and μ enter as μ_{eq}^2 and μ^2; this implies the symmetry $j_{ion}(L, \phi, \mu_{eq}) = j_{ion}(L, \phi, -\mu_{eq})$. In view of these symmetries, the pitch angle distribution of j_{ion} is determined for all values of θ if $j_{ion}(\mu_{eq})$ is known for $0 \leq \mu \leq 1$.

Another important feature of $j_{ion}$is the possible presence of a depleted loss cone. Ions that mirror at or below the exobase are quickly removed from the distribution by charge exchange. The mirror altitude at a fixed L can be regarded as a function of the equatorial pitch angle μ_{eq}. The point at which the mirror radius is equal to the exobase radius, R_{atm}, defines the equatorial loss cone pitch angle cosine

$$\mu_{eqlc}^2 = 1 - \frac{R_{atm}^3}{L_3} \frac{1}{\sqrt{4 - 3R_{atm}/L}} \tag{10}$$

If a loss cone is present, the ion intensity, $j_{ion}(L, \theta, \mu_{eq})$, will be depleted where $\mu_{eq} \geq \mu_{eqlc}$. The existence and depth of the loss cone depends on the source of the ion intensity can cannot be predicted *a priori*.

In view of the above discussion, we have found it convenient to parameterize the pitch angle distribution as follows

$$j_{ion}(L, \theta, \mu_{eq}) = U(\mu_{eq}, \mu_{eqlc}(L), m) \sum_{p=0}^{\lambda} j_{ion}^{(k)} P_p^*(\mu_{eq}) \tag{11}$$

where the functions $P_p^*(\mu_{eq})$ are the Legendre polynomials scaled to the interval $(0, 1)$, the parameter m is set to 1 if the loss cone is present or 0 if is not. The function $U(\mu_{eq}, \mu_{eqlc}(L), m)$ is a binary function that is zero if $\mu_{eq} > \mu_{eqlc}(L)$ and $m = 1$. The U function therefore governs the behavior of the loss cone, which is considered either completely filled ($m = 0$) or completely empty ($m = 1$). Clearly, the number of pitch angle moments, n, determines the amount of detail in the pitch angle structure that can be represented by equation (11).

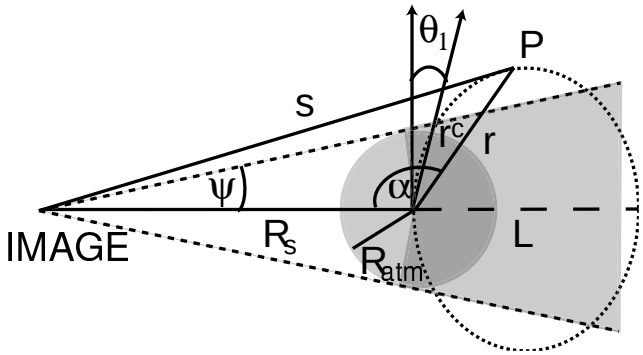

Figure 2. Obscuration of volume elements. For an observer at **IMAGE** viewing a volume element at **P** along line of sight **s**. The line of sight to **P** is obscured if it lies in the grey area. The dotted curve represents a contour of constant L. The angle $\theta_1 = \arcsin\left(\sqrt{R_{atm}/L}\right)$ is the polar angle where the northward L contour intersects with the exobase. The dashed rays originating from **IMAGE** represent the tangent cone. $\mathbf{r}$ is the line between **P** and the center of the Earth and $\mathbf{r^c}$ is the line along $\mathbf{r}$ from the center of the Earth to the tangent cone.

In this subsection, we have constrained the behavior of the ion intensity function j_{ion} by choosing an explicit representation and dependencies for it. We have assumed that over the period of observation, that the motion of the ions is dominated by a dipole magnetic field. All other forces, e.g., electric fields, drifts, etc. are neglected. Further, we have neglected all perturbations of the dipole field.

3.2. Geocoronal Emission

The geocoronal component of the measurement, equation (4), can be viewed as a volume integral over all space above the exobase that is not obscured by the Earth. This integral can be expressed in terms of (L, θ, ϕ) through the introduction of the Jacobean

$$J_0(L, \theta, \phi) = \sin^3\theta \frac{r(L, \theta)^2}{s(L, \theta, \phi)^2} \tag{12}$$

Before rewriting equation (4) in terms of these new variables, the limits of the integration must be determined as well. Equation (4) can be recast as a integral over all space through the introduction of an obscuration function O which has the value of 1 or 0 if the specified volume element is outside of or inside of the obscuration region, respectively.The geometry of O is shown in figure 2. Volume elements that lie below the exobase or are blocked from view by the Earth must be excluded from the integral, i.e., O must be set to 0. The function O can be written as (see *Roelof* [2002])

$$O(L, \theta, \phi) = \begin{cases} 0 & \text{if} \quad \alpha + \psi > \frac{\pi}{2} \quad \text{and} \\ & \sin^2\theta \sin(\alpha + \psi) < \sin^2\theta_1 \\ 0 & \text{if} \quad \sin\theta < \sin\theta_1 \\ 1 & \text{otherwise} \end{cases} \tag{13}$$

The condition $\sin\theta < \sin\theta_1$ can be incorporated directly in the limits of integration over both L and θ, as contributions to the integral can only be made where $L > R_{atm}$ and $\theta_1 < \theta < \pi - \theta_1$.

Using equations (11) and (12), equation (4) can be expressed in terms of the natural coordinates of j_{ion} as

$$C_i^{high} = \sum_p \int_0^{2\pi} \int_{R_{atm}}^{\infty} f_i^p(L, \phi, \mu_{eq}) j_{ion}^p(L, \phi) dL d\phi \tag{14}$$

where

$$f_i^p(L, \phi, \mu_{eq}) \equiv \sigma_H^{10} \int_{\theta_1}^{\pi - \theta_1} n^H < A_i > J_0 O P_p^*(\mu_{eq}) d\theta \tag{15}$$

The integral in equation (15) can be calculated numerically via Romberg integration (*Press et al.* [1988]).

Using the techniques developed in appendix A, equation (14) can be approximated by numerical quadrature as

$$C_i^{high} = \sum_p \sum_k \sum_l b_{ikl}^p j_{ion}^p(L_l, \phi_k) \tag{16}$$

where the quadrature weights b_{ikl}^p are determined by integrals of the form

$$\zeta_{ikl}^{pnm} = \int_{L_l}^{L_{l+1}} \int_{\phi_k}^{\phi_{k+1}} g_{kl}^{nm} f_i^p dL d\phi \tag{17}$$

with

$$g_{kl}^{nm} = \frac{(-1)^{n+m}(L_{l+m} - L)(\phi_{k+n} - \phi)}{(L_{l+1} - L_l)(\phi_{k+1} - \phi_k)} \tag{18}$$

The integrals in equation (17) are performed by means of the four point integration scheme described by *Abramowitz and Stegun* [1972]. Equation (16) is thus a well defined, linearized forward model for HENA geocoronal measurements. Equation (16)can be written in matrix form as

$$\boldsymbol{C}^{high} = \left|\boldsymbol{K}_{high}^0 \quad \boldsymbol{K}_{high}^1\right| \left|\begin{matrix}\boldsymbol{J}_{ion}^0 \\ \boldsymbol{J}_{ion}^1\end{matrix}\right| \tag{19}$$

where the superscripts on $\boldsymbol{K}_{high}^p$ and $\boldsymbol{J}_{ion}^p$ signify the dependence on the pitch angle moments.

We can simplify the interpretation of the pitch angle moments, $\boldsymbol{J}_{ion}^p$, if only the first two terms are taken ($p = 0, 1$). Considering only the first two moments

$$\begin{aligned}\sum_{p=0}^{2} j_{ion}^{(k)} P_p^*(0) &= j_{ion}^{(0)} - j_{ion}^{(1)} \\ \sum_{p=0}^{2} j_{ion}^{(k)} P_p^*(1) &= j_{ion}^{(0)} + j_{ion}^{(1)}\end{aligned} \tag{20}$$

we can immediately identify the quantities

$$\begin{aligned}j_{ion}^{\perp} &\equiv j_{ion}^{(0)} - j_{ion}^{(1)} \\ j_{ion}^{\parallel} &\equiv j_{ion}^{(0)} + j_{ion}^{(1)}\end{aligned} \tag{21}$$

which are the equatorial ion intensities perpendicular and parallel to the field lines, respectively. Equation (19) can now be written in terms of these more intuitive quantities as

$$\boldsymbol{C}^{high} = \left|\boldsymbol{K}_{high}^{\parallel} \quad \boldsymbol{K}_{high}^{\perp}\right| \left|\begin{matrix}\boldsymbol{J}_{ion}^{\parallel} \\ \boldsymbol{J}_{ion}^{\perp}\end{matrix}\right| \tag{22}$$

where we have defined

$$\begin{aligned}\boldsymbol{K}_{high}^{\parallel} &= \tfrac{1}{2}\left(\boldsymbol{K}_{high}^0 + \boldsymbol{K}_{high}^1\right) \\ \boldsymbol{K}_{high}^{\perp} &= \tfrac{1}{2}\left(\boldsymbol{K}_{high}^0 - \boldsymbol{K}_{high}^1\right)\end{aligned} \tag{23}$$

Experience with the HENA measurements and retrievals have shown that no more than two pitch angle moments can be resolved. Thus the representation of equation (22)is slightly preferable to that of equation (19). This representation is not only more intuitive, but provides a more direct method of calculating the exospheric emission (as will be detailed in section 3.3), and allows for the possibility of implementing non-negativity constraints on the retrievals. This last advantage is due to the fact that in order to be physically acceptable, $\boldsymbol{J}_{ion}^{\parallel}$ and $\boldsymbol{J}_{ion}^{\parallel}$ must be positive, though physical acceptable values of $\boldsymbol{J}_{ion}^1$ can have either sign.

3.3. Emission Near the Exobase

Emission from near the exobase results from charge exchange of nearly mirroring ions with atmospheric neutrals, primarily atomic oxygen. This emission occurs in a thin layer due to the small scale height (relative to that of the geocorona) of atomic oxygen and other gasses in this region. The emission is also optically thick, i.e., there is a finite probability that an ENA created by charge exchange will undergo stripping, i.e., re-ionization, before exiting this layer. We will model this as a surface emission in view of the fact that the layer is too thin (a few hundred kilometers or so) to be resolved by the HENA instrument. Further,it should be noted that the emission comes from a rather narrow range of pitch angles that correspond to ions mirroring within this narrow layer. Ions that mirror above this layer (those with smaller pitch cosines) never encounter atmospheric densities suitable for large ENA production. Those ions with larger pitch cosines penetrate to altitudes below the emission layer and the ENAs they produce do not escape. For this reason, the contribution of low altitude emission to the HENA images is strongest (and is, in fact, only significant) where the line of sight is nearly perpendicular to the field lines as the enter the atmosphere. When this range of pitch angles is mapped back to the equator, we find that in order for the ions to mirror below the exobase, the equatorial pitch cosine must be close, to one, i.e., $\mu_{eq} > .95$. For this reason, we consider only the $\boldsymbol{J}_{ion}^{\parallel}$ component of the ion distribution as a source of emission near the exobase.

Roelof has shown that the ENA intensity due to emission near the exobase can be written as (*Roelof* [2002])

$$j_{ena}^{e}(L, \phi, \mu) = M(\mu) j_{ion}^{\parallel}(L, \phi) \tag{24}$$

where $M(\mu)$ is defined by

$$M(\mu) \equiv \frac{\sigma_O^{10}}{\sigma_O^c}\left(1 - e^{\sigma_O^c \xi(\mu)}\right) e^{-\kappa \sigma_O \xi(\mu)} \tag{25}$$

where σ_O^{10} is the charge exchange cross section for protons on atomic oxygen, σ_O^c is the combined cross charge exchanged-stripping cross section $\sigma_O^{10} + \sigma_O^{01}$ and σ_O is the total inelastic interaction cross section for protons on oxygen

$$\sigma_O = \sigma_O^{10} + \sigma_O^1 \cong \sigma_O^{01} + \sigma_O^2 \tag{26}$$

where σ_O^1 and σ_O^2 are the sums of the ionization and excitation cross sections for ions and ENAs, respectively. The constant κ represented the ratio of the average energy loss per interaction, W ($\cong$ .028 kev), to the size of the HENA energy bin, E^* ($\cong$ 5 kev). The optical path function, ξ is given by *Roelof* as

$$\xi = \sigma_O n^O(r_m)\sqrt{\frac{\pi H_o r_m}{3}}\operatorname{erfc}\left[-\frac{\sqrt{R_{atm} - r_m}}{H_o}\right] \tag{27}$$

where the n^O is the atomic oxygen density, H_o is the oxygen scale height near the exobase and r_m is the mirror radius

$$r_m = \left(1 - \mu^2\right)^{1/3} \tag{28}$$

Thus the ENA intensity, j_{ena} emitted from the surface near the exobase is very closely related to the ion intensity with small pitch angles, $j_{ion}^{\parallel}$.

We now combine equations (24) and (5) and write the result as an integral over all space,

$$C_i^{low} = \int_0^{2\pi}\int_0^{\pi}\int_0^{\infty} \sin\epsilon < A_i > M\delta(s - s_e) j_{ion}^{\parallel} ds d\epsilon d\phi \tag{29}$$

where we have used the Dirac delta function, δ to reinterpret the equation as a volume integral. We now introduce the Jacobean connecting the system (s, ϵ, β) with the (r, θ, ϕ) system

$$J_1(r, \theta, \phi) = \frac{r^2 \sin\theta}{s^2 \sin\epsilon} \tag{30}$$

and make use the identity

$$\sin\theta d\theta = -\frac{dL}{2L^2\sqrt{1-R_{atm}/L}} \tag{31}$$

to write

$$C_i^{low} = \int_0^{2\pi}\int_r^{\infty}\int_0^{\infty} \frac{r^2 < A_i > M}{2s^2L^2\sqrt{1-r/L}}\delta(s-s_e)j_{ion}^{\parallel}drdLd\phi \tag{32}$$

For a given field line, determined by L and ϕ, there are two points at which the argument of the delta function in equation (32) attain a zero value. These points occur where the field line intersects the exobase in the northern and southern hemispheres. Clearly, at these two points the value of r becomes R_{atm}; we subscript the the other quantities in the integral with N or S to indicate which intersection point is required. We now rewrite equation (32) as

$$C_i^{low} = \int_0^{2\pi}\int_{R_atm}^{\infty} \rho_i(L,\phi)j_{ion}^{\parallel}(L,\phi)dLd\phi \tag{33}$$

with

$$\rho_i(L,\phi) = \sum_H \frac{R_{atm}^2 < A_i(\epsilon_H,\beta_H) > M(\mu_H)}{2s_H^2L^2\sqrt{1-R_{atm}/L}} \tag{34}$$

where the summation is carried out over the two hemispheres.

Again, we use the method described in appendix A, to approximate the integral in equation (33) as a numerical quadrature, writing

$$C_i^{low} = \sum_k\sum_l c_{ikl}j_{ion}^{\parallel}(L_l,\phi_k) \tag{35}$$

where the weights c_{ikl} are calculated via integral of the form

$$\vartheta_{ikl}^{nm} = \int_{L_l}^{L_{l+1}}\int_{\phi_k}^{\phi_{k+1}} g_{kl}^{nm}\rho_i dLd\phi \tag{36}$$

with g_{kl}^{nm} given by equation (18).

Just as for the high altitude emission, we write write equation (35) in matrix form

$$\boldsymbol{C}^{low} = \boldsymbol{K}_{low}^{\parallel}\boldsymbol{J}_{low}^{\parallel} \tag{37}$$

Now combining equations (3), (22) and (37) we can now write

$$\boldsymbol{C} = \boldsymbol{K}\begin{vmatrix}\boldsymbol{J}_{ion}^{\parallel}\\ \boldsymbol{J}_{ion}^{\perp}\end{vmatrix} = \boldsymbol{K}\boldsymbol{J}_{ion} \tag{38}$$

where $\boldsymbol{C}$ is the vector representation of the measured counts from both the geocoronal source and the source near the exobase, and the so call kernel matrix $\boldsymbol{K}$ is defined as

$$\boldsymbol{K} \equiv \begin{vmatrix}\left(\boldsymbol{K}_{high}^{\parallel} + \boldsymbol{K}_{low}^{\parallel}\right) & \boldsymbol{K}_{high}^{\perp}\end{vmatrix} \tag{39}$$

Equation (38) is the linearized equation of measurement for the HENA instrument. This equation can be used to simulate HENA observations given an ion density on a specified grid. It also contains all the information (at least within the approximations made) about the relationship between the measured quantities, $\boldsymbol{C}$ and the object of the remote sensing measurement $\boldsymbol{J}_{ion}$. This information will be exploited in the following sections where we seek to determined the inverse relationship, i.e., $\boldsymbol{J}_{ion}$ as a function of $\boldsymbol{C}$.

4. Simulated HENA images

Before proceeding to inversions, we will use the equations developed in the previous section to simulate HENA measurements. This will provide a groundwork for much of the discussion in the sections relating to constraint tuning and error analysis.

Simulations are based on equation (38), which in turn depends on the construction of the kernel, $\boldsymbol{K}$. In order to calculate the kernel, several quantities must be specified. The geocoronal hydrogen density is described in *C:son Brandt et al.* [2002] and is based on the work by *Rairden et al.* [1986]. The charge exchange cross section for hydrogen, σ_H^{10}, has been taken from *Barnett* [1990]. The cross sections for atomic oxygen, σ_O^{10}, σ_O^{10}, σ_O^{1} and σ_O^{2} have been taken from *Basu et al.* [1987]. For the calculation of low altitude emission, we assume an exobase height, R_{atm}, of 500km, an atomic oxygen scale height, H_O, of 50km and an atomic oxygen density at the exobase, $n^O(R_{atm})$, of 4.4×10^6. We can now calculate the kernel using these quantities, the HENA instrument function and the ion distribution.

Figure 3 shows the results of a simulation of an actual HENA observation that occurred on the fourth of October, 2000. The test ion distribution has a maximum at $L=4$ in the midnight-dawn sector. This is reflected in the simulated image which also shows a midnight-dawn enhancement. The viewing perspective, the geocoronal distribution and the instrument spread function make visual inference of the test ion distribution from the simulated image somewhat difficult. The ENA intensity is a product of the ion distribution and the neutral density, thus for a given field line the maximum ENA intensity will occur where the density is highest, i.e., where the distance to the earth is smallest. Thus the maximum intensity generally occurs where the ions mirror (this is made even more significant by the long path lengths the ions traverse near the mirror point). For this reason, the peak of the ion density will appear to occur closer to the Earth in the image. In addition, since the intensity is brightest near the mirror point, even when the equatorial ion distribution is isotropic, the ENA intensity will exhibit significant anisotropy where the larger intensities are being generated. This is particularly true of the near exobase emission, where nearly all the ENA emission is perpendicular to the field lines. Finally, the HENA instrument has a finite spread function; sharp bright areas of the intensity are broadened by the instrument. This is evident in the simulated image where there appears to be significant emission on the sunward side of the Earth. This apparent sunward side emission is, however, due to post-midnight emission spread by the instrument.

From the figure it is clear that the ENA image contains a good deal of information about the ion distribution. The interpretation of this information directly from the image, even qualitatively, can be difficult. It is for this reason that we seek to develop a recovery technique that can re-produce, as far as possible, the ion distribution from the image.

5. Constrained Linear Inversion

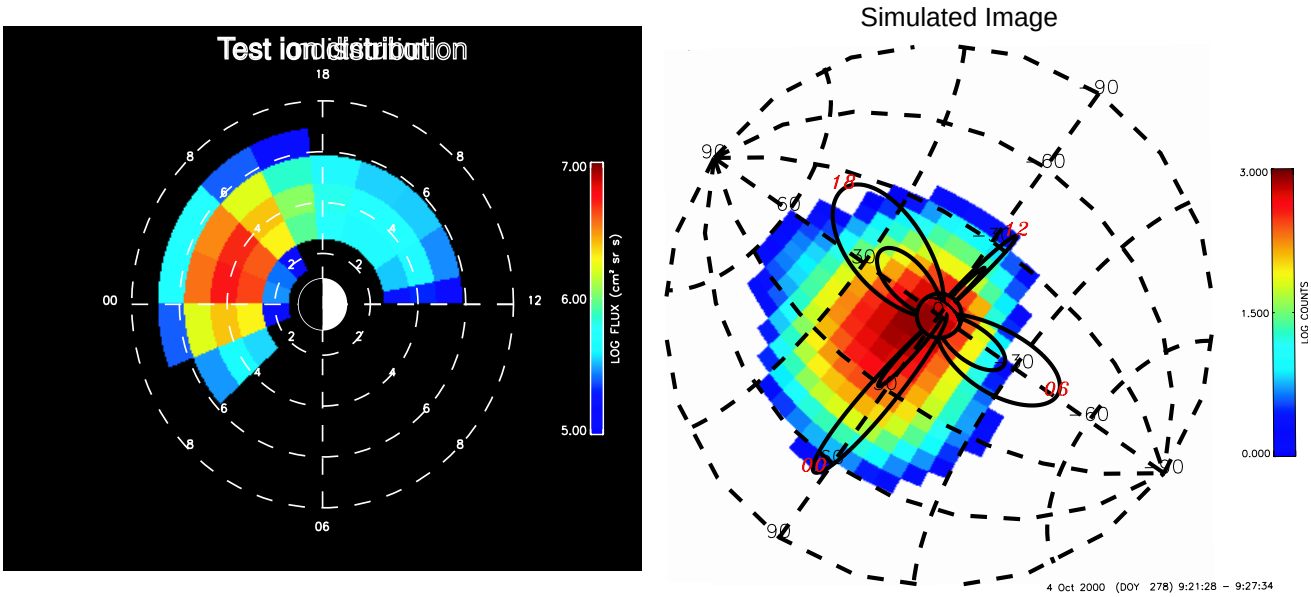

Figure 3. Simulated HENA image (right) using the test equatorial ion distribution (left) based on the actual viewing geometry that occurred on the fourth of October,2000. On the ion distribution panel, the closed dashed curves represent the scale of L (2-12) in Earth radii and the local time scale is given outside of the perimeter of the outermost L boundary. The simulated image is presented in a projection described in detail by *C:son Brandt et al.* [2002]. The dashed black lines on the image represent the scale in ϵ and β, the solid black curves show curves of constant L ($L = 4, 8$) at four local times (which are labelled in red).

5.1. Inversion Equations

In a strict sense, equation (38) can only be solved uniquely if the matrix $\boldsymbol{K}$ is both square and non-singular. In practice, meeting these two criteria simultaneously is not possible. Instead of seeking a direct solution, we could pursue a least squares solution, i.e., determine $\boldsymbol{J}_{ion}$ through minimizing the quantity

$$(\boldsymbol{C} - \boldsymbol{K}\boldsymbol{J}_{ion})^T \boldsymbol{\sigma}_C^{-2} (\boldsymbol{C} - \boldsymbol{K}\boldsymbol{J}_{ion}) \tag{40}$$

where $\boldsymbol{\sigma}_C^{-2}$ is the inverse of the measurement covariance matrix. For the HENA data, $\boldsymbol{\sigma}_C^{-2}$ is a diagonal matrix whose elements are $1/\sigma_i^2$ with σ_i as the uncertainty corresponding to each pixel. Minimizing (40) yields

$$\boldsymbol{J}_{ion} = \left(\boldsymbol{K}^T \boldsymbol{\sigma}_C^{-2} \boldsymbol{K}\right)^{-1} \boldsymbol{K}^T \boldsymbol{\sigma}_C^{-2} \boldsymbol{C} \tag{41}$$

Unfortunately, in the case of the HENA measurements, the matrix $\boldsymbol{K}^T \boldsymbol{\sigma}_C^{-2} \boldsymbol{K}$ is very poorly conditioned. For this reason, application of equation (41) results in a highly oscillatory solution that is dominated by numerical noise. In spite of the fact that we have more measurements than unknowns, the measurements do not contain enough information to uniquely determine $\boldsymbol{J}_{ion}$. We must, therefore, add information about $\boldsymbol{J}_{ion}$ in the form of *a priori* constraints. Instead of minimizing the expression (40) we instead minimize

$$(\boldsymbol{C} - \boldsymbol{K}\boldsymbol{J}_{ion}) \boldsymbol{\sigma}_C^{-2} (\boldsymbol{C} - \boldsymbol{K}\boldsymbol{J}_{ion}) + \gamma \boldsymbol{J}_{ion}^T \boldsymbol{H} \boldsymbol{J}_{ion} \tag{42}$$

where γ is a constant and $\boldsymbol{H}$ is a constraint matrix. Both γ and $\boldsymbol{H}$ will be described in detail below. The result of the minimization is

$$\boldsymbol{J}_{ion} = \left(\boldsymbol{K}^T \boldsymbol{\sigma}_C^{-2} \boldsymbol{K} + \gamma \boldsymbol{H}\right)^{-1} \boldsymbol{K}^T \boldsymbol{\sigma}_C^{-2} \boldsymbol{C} \tag{43}$$

which for appropriate values of γ and $\boldsymbol{H}$ has a unique solution. Solutions of this form have been treated extensively by other authors (see for example *Twomey* [1977], *Rodgers* [2000] and *Menke* [1989]); only a brief summary has been described here.

In the expression (42) the constant γ serves a tuning parameter, determining the relative contribution of the *a priori* constraints to the solution. As γ becomes larger, the solution is more tightly bound to the constraint. The constraint matrix $\boldsymbol{H}$ can take several forms, and is usually written in the form

$$\boldsymbol{H} = \boldsymbol{D}^T \boldsymbol{D} \tag{44}$$

where $\boldsymbol{D}$ is a linear operator that when applied to $\boldsymbol{J}_{ion}$ is expected to be small. For example, defining $\boldsymbol{D}$ as $\boldsymbol{D}_0$, the identity operator, the constraint in equation (43) would favor solutions with smaller $\boldsymbol{J}_{ion}^T \boldsymbol{J}_{ion}$, i.e., solutions where the absolute values of the elements of $\boldsymbol{J}_{ion}$ are small. If,in turn, we define $\boldsymbol{D}$ as $\boldsymbol{D}_1$, the first derivative operator, solutions that tend toward constant values of $\boldsymbol{J}_{ion}$ would be favored (small first derivative).

Note that we can also calculate the expected covariance of the solution, $\boldsymbol{\sigma}_J^2$ through the relationship (*Menke* [1989])

$$\boldsymbol{\sigma}_J^2 = \boldsymbol{G} \boldsymbol{\sigma}_C^2 \boldsymbol{G}^T \tag{45}$$

defining the generalized inverse $\boldsymbol{G}$ as

$$\boldsymbol{G} \equiv \left(\boldsymbol{K}^T \boldsymbol{\sigma}_C^{-2} \boldsymbol{K} + \gamma \boldsymbol{H}\right)^{-1} \boldsymbol{K}^T \boldsymbol{\sigma}_C^{-2} \tag{46}$$

We must now determine the best γ and $\boldsymbol{H}$ to use with the HENA data.

5.2. Constraint Matrices

As we have pointed out earlier the HENA measurements do not contain enough information to unambiguously determine the ion distribution, J_{ion}. The added information

takes the form of an *a priori* assumption that along with minimizing the expression (40) the best solutions will also minimize the quantity $\boldsymbol{J}_{ion}^T \boldsymbol{H} \boldsymbol{J}_{ion}$. In this section we will introduce specific forms of the constraint matrix $\boldsymbol{H}$ that will subsequently be evaluated for applicability to the HENA data.

It should be pointed out that in the absence of constraints, the estimated $\boldsymbol{J}_{ion}$ oscillates rapidly between unreasonable large positive values and large negative (and therefore completely unphysical) values. In a certain sense, the constraints serve to damp this oscillation. The simplest constraint would favor smaller values of J_{ion} and reject larger ones. As we have pointed out before, the identity matrix $\boldsymbol{H}_0 = \boldsymbol{D}_0^T \boldsymbol{D}_0 = \boldsymbol{I}$ serves this purpose well. The drawback, of course, is that this constraint will tend to force the solution to be smaller than it should be.

We could also damp the oscillations in $\boldsymbol{J}_{ion}$ by requiring the solution to be smooth rather than small. For this purpose we choose a slightly modified version of the second derivative operator, i.e., the Laplacian, in the construction of $\boldsymbol{J}$ via equation 44. The (L, ϕ) is a polar coordinate system so the modified second derivative operator is written

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \alpha \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2} \tag{47}$$

where we have introduced the asymmetry parameter α, which can be used to adjust the smoothness more strongly in either the L or ϕ direction. This free parameter α is introduced because we have no *a priori* reason to expect an equal amount of smoothness in L and ϕ. We implement equation (47) using a second order finite difference stencil. Two other modifications are required before this constraint can be used. First, the boundary in ϕ is cyclical, thus the finite difference stencil must be modified to calculate the proper derivatives across the boundary at each L. Second, we require that the ion distribution be small at the boundaries in L. We implement this 'clamping' by replacing the smoothness constraint for the minimum and maximum L with the identity constraint. We can thus express the smoothness constraint as

$$\boldsymbol{H}_2(\alpha, \lambda) = \boldsymbol{D}_2^{*T}(\alpha) \boldsymbol{D}_2^*(\alpha) + \lambda \boldsymbol{H}_0^b \tag{48}$$

where $\boldsymbol{D}_2^*(\alpha)$ is the modified second derivative operator, $\boldsymbol{H}_0^b$ is the $\boldsymbol{H}_0$ operator that acts only on the L boundaries, and λ is used to adjust the relative strength of the clamping constraint. The $\boldsymbol{H}_2$ constraint favors solutions with small spatial second derivatives, that is, solutions that increase or decrease linearly.

Finally, instead of the more mathematically constructed constraints above, we could take a more physically based approach, making the straightforward assumption that elements of $\boldsymbol{J}_{ion}$ that lie close to one another are more closely correlated to one another than elements that are widely spatially separated. This type of constraint has been described by *Rodgers* [2000] who relates its construction to a Markhov type process. The two dimensional Markhov constraint matrix has elements

$$\boldsymbol{H}_m^{ij} = e^{-d_{ij}/d_m} \tag{49}$$

where d_{ij} is the Euclidian distance between element (L_i, ϕ_i) and (L_j, ϕ_j) and d_m is an adjustable parameter that describes the characteristic length of the covariance. In the construction of d_{ij} we implement the cyclical boundary in ϕ. We also add the clamping constraint at the boundaries, so that the Markhov constraint is most properly expressed as $\boldsymbol{H}_m(d_m, \lambda)$. Note that in the limit of small d_m, the Markhov constraint becomes identical to the identity constraint. As d_m increases, the correlation between neighboring points increases, and the constraint tends to favor solutions where the covariance weighted mean around any given point is smaller. Like the identity constraint, the Markhov constraint will tend to force the solution to be smaller than it should be.

We now proceed to describe methods for selecting constraints and optimizing their free parameters.

5.3. Method for Tuning the Constraints

Clearly, the optimal choice of γ, $\boldsymbol{H}$ and the free parameters of $\boldsymbol{H}$ should yield a solution that is both consistent with the measurement and reflective of the underlying ion distribution. Consistency with the measurement can be easily quantified via the expression (40). For this purpose we define the goodness of fit, g_f as

$$g_f(\boldsymbol{p}_H) = \frac{1}{n_C} (\boldsymbol{C} - \boldsymbol{K}\boldsymbol{J}_{ion}) \boldsymbol{\sigma}_C^{-2} (\boldsymbol{C} - \boldsymbol{K}\boldsymbol{J}_{ion}) \tag{50}$$

where n_C is the number of measurements (i.e., the number of pixels in the HENA image) and $\boldsymbol{p}_H$ is the appropriate array of constraint parameters (γ for $\boldsymbol{H}_0$, $(\gamma, \alpha, \lambda)$ for $\boldsymbol{H}_2$ and (γ, d_m, λ) for $\boldsymbol{H}_m$) . We expect g_f to be approximately 1 for a retrieved ion distribution that is consistent with the data.

For real measurements, it is not possible to gauge how well the solution reflects the underlying ion distribution. If we had access to the underlying ion distribution, we wouldn't have bothered to make the ENA measurement. We can, however, simulate images from a given ion distribution, use equation (42) to estimate ion distribution from this image and then compare the given ion distribution to

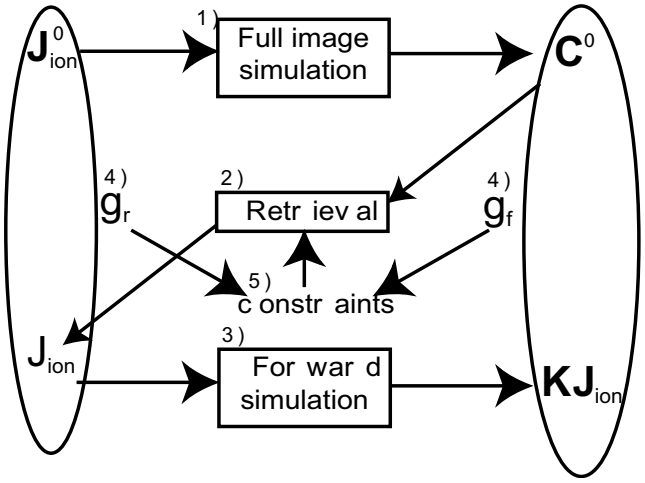

Figure 4. Method of constraint tuning. 1) an assumed ion distribution $\boldsymbol{J}_{ion}^0$ is used to create a simulated image, $\boldsymbol{C}^0$ via the full integration method of *Roelof and Skinner* [2000]. 2) An ion distribution, $\boldsymbol{J}_{ion}$ is retrieved from the simulated image using an initial guess for the constraints. 3) An image, $\boldsymbol{K}\boldsymbol{J}_{ion}$ is reconstructed from the retrieved ion density. 4) The goodness of fit and retrieval, g_f and g_r are calculated from $\boldsymbol{K}\boldsymbol{J}_{ion}$ and $\boldsymbol{J}_{ion}$. 5) The constraints are modified based on g_f and g_r. The process resumes at step 2 using the new constraints and continues until the quantity q is minimized.

the estimated ion distribution. Following this process, we can define the goodness of the retrieval, g_r as

$$g_r(\boldsymbol{p}_H) = \frac{1}{n_J}\left(\boldsymbol{J}^0_{ion} - \boldsymbol{J}_{ion}\right)\boldsymbol{\sigma}_J^{-2}\left(\boldsymbol{J}^0_{ion} - \boldsymbol{J}_{ion}\right) \qquad (51)$$

where n_J is the number of elements of $\boldsymbol{J}_{ion}$ and $\boldsymbol{J}^0_{ion}$ is the ion distribution used for the image simulation. We could, in principle, use equation (38) to create the simulated image. We have chosen, however, to simulate the images using the the parameterization of *Roelof and Skinner* [2000]. In doing so, the quantity g_r reflects the systematic errors incurred in linearizing the problem (since *Roelof and Skinner* [2000] use a non-linear, direct integration method). After the simulated image is constructed, gaussian noise that is consistent with the actually measured image is added. In order to do so, the simulated image is scaled so that the maximum simulated pixel is equal to the maximum measured pixel, and then uncertainties consistent with counting statistics are assumed for the scaled simulated image.

We expect that for an acceptable solution, g_r should be approximately 1. If g_r is significantly less greater that 1, the retrieved ion density,$\boldsymbol{J}_{ion}$ differs too greatly from the input ion density $\boldsymbol{J}^0_{ion}$. If in turn, g_r is significantly less than 1, the covariance of the retrieved values, $\boldsymbol{\sigma}^2_J$ is too large, i.e., the uncertainty of $\boldsymbol{J}_{ion}$ can be made smaller through tightening the constraints.

If we regard both g_r and g_f as functions of the constraints, we can determine a set of acceptable constraints for a particular observation geometry by minimizing the quantity

$$q(\boldsymbol{p}_H) \equiv (1 - g_r(\boldsymbol{p}_H))^2 + (1 - g_f(\boldsymbol{p}_H))^2 \qquad (52)$$

for a simulated image similar in character to the actual measurements. When this quantity is made small, we are assured that for the simulated image, the retrieval is both reflective of the assumed underlying ion distribution and consistent with the input image. The constraints so determined then can be used with the actual measurements and

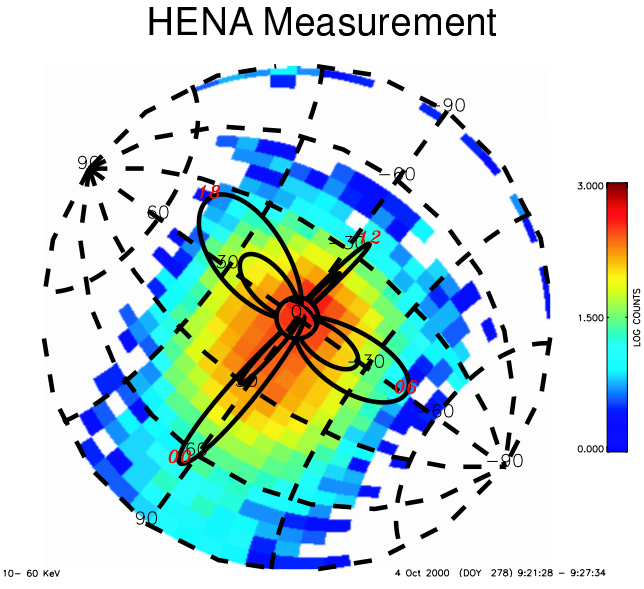

Figure 5. HENA image from the fourth of October, 2000 displayed in the same format as Figure (3)

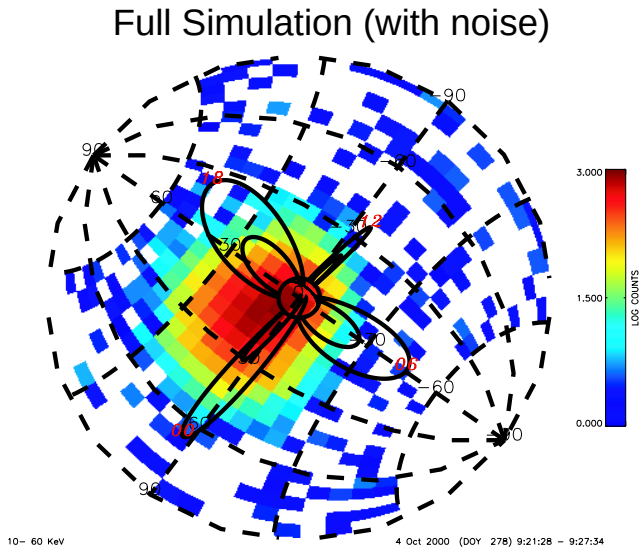

Figure 6. Full non-linear simulation using the ion distribution in Figure (3) and the image geometry of Figure (5) with noise added.

expected to similarly retrieve the actual underlying ion distribution.

Our method for minimizing q and therefore optimizing the constraints are shown in Figure 4.The constraints are adjusted to minimize the quantity q by means of a non-linear simplex minimization given by *Press et al.* [1988].

Thus for a given form of $\boldsymbol{H}$, we can systematically select the constraint parameters, $\boldsymbol{p}_H$, that optimally retrieve a test image. We then postulate that if we are given an image that is similar in character to the test image, this set of constraint parameters will be suitable. In this case, similar in character implies a similar viewing geometry, similar signal to noise ratio and a qualitatively similar ion distribution.

5.4. Example Tuning and Error Analysis

The HENA instrument acquires images with very disparate viewing geometries and ENA intensities; for this reason it is difficult to provide a comprehensive assessment of the quality of our inversion algorithm. The tuning process described above, however, provides information on how well the inversion process should be expected to work for a particular measurement. In this section, we examine the application of the tuning process and the quality of the retrieval for a particular image that is reasonably representative of much of the HENA data.

Our analysis focuses on the image shown in Figure (5), on which the simulation shown in Figure (3) was based. Note that the image in Figure (5) exhibits a slightly more diffuse region of emission that appears to extend to higher values of L on the night-side. Also the measured emission is slightly less intense at the peak than in the simulation. Because the signal to noise ratio is higher for brighter emission, the uncertainties in the simulated image must be scaled accordingly to be representative of the measured data. This process of scaling is described above.

The full non-linear simulation (using the model from *Roelof and Skinner* [2000] as described above) with the appropriate noise added is shown in Figure (6). This image employs the same ion distribution as the one used in Figure

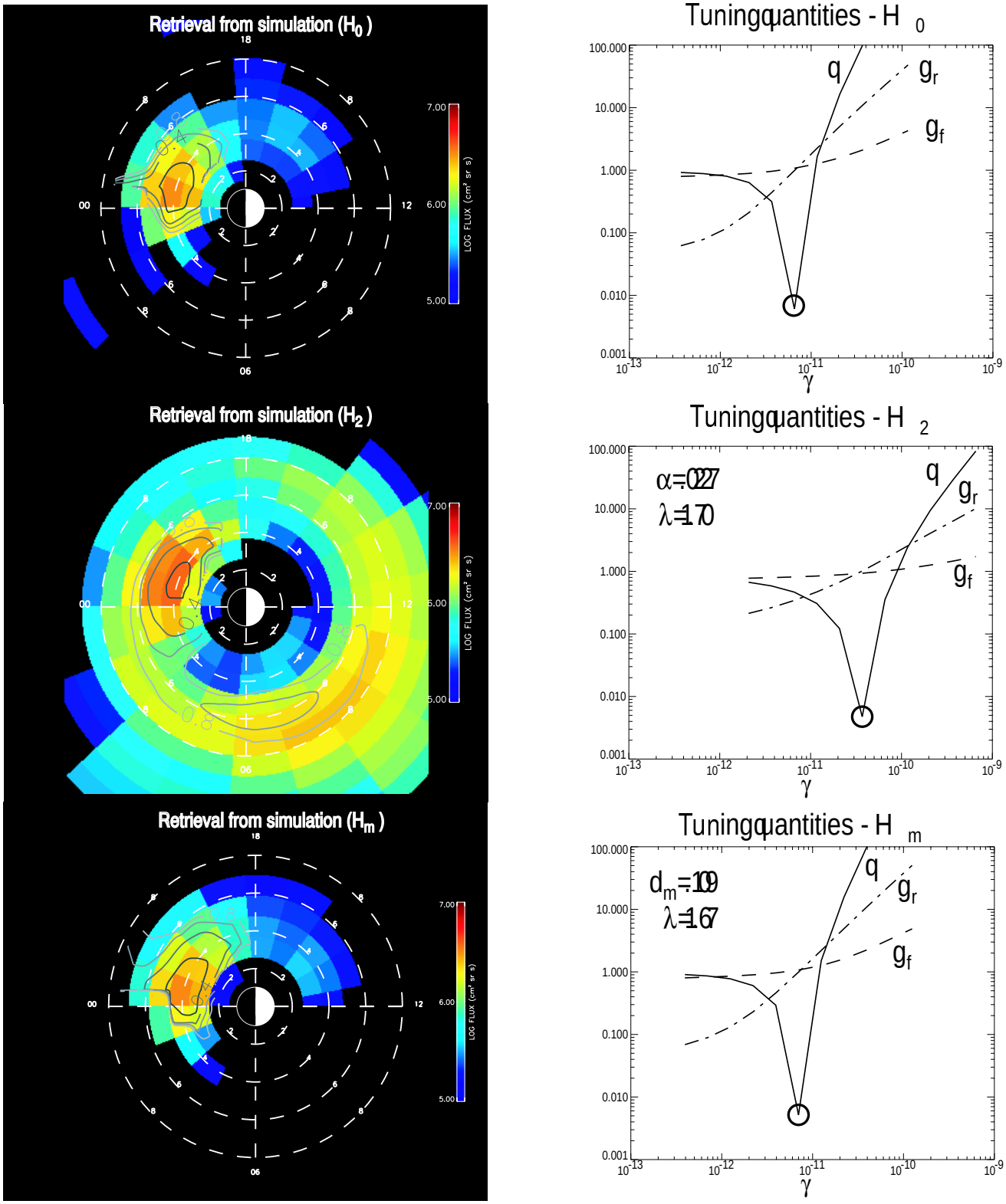

Figure 7. Optimal retrievals and tuning quantities for various constraint types. Optimized retrievals of the simulated image in Figure (6) for the three constraint types, H_0, H_2 and H_m are shown on the left. The plots on the right show the various tuning measurements (g_f, g_r and q) as a function of the constraint strength γ. The images on the left are rendered on the same scale as the ion distribution in Figure (3) and can be compared to this figure directly. Contours of relative uncertainties are also shown on the ion distribution images. The plots on the right were made for the optimal values of the additional tuning parameters (α, λ and d_m); The value of these parameters are also shown on the plot. The minimum value of q at the optimal γ are circled in each plot.

(3). As is to be expected, the principle difference between Figure (6) and Figure (3) is addition of noise.

The results of the tuning process and a summary of the dependence of the tuning measurements, g_f, g_r and q, on the constraint strength, γ, are shown in Figure (6). Be-

fore examining the details of the retrievals themselves, we will first discuss the general behavior of the tuning process shown in the right column of the figure.

The qualitative behavior of the various tuning measurements, g_f, g_r and q, as a function of the constraint strength, γ, is similar for all of the constraint types. At small values of γ, the quality of the fit, g_f, is close to one, as the constraints merely push the retrieval away from the mathematically optimal retrieval (which oscillates wildly) towards a more physically realistic one. As γ increases, g_f climbs away from one, and the retrieval becomes more and more biased towards the constraint. The quality of the retrieval, g_r, also monotonically increases as a function of γ, though at small values of γ it is significantly less than one. This is not because the retrieved ion distribution faithfully reproduces the test ion distribution, rather, it is because the retrieved covariance is very large when γ is small. As γ grows, the covariance shrinks and g_r increases. For this reason it is important to regard g_r as a measure of the fidelity of the retrieval *with respect to the estimated uncertainty of the retrieval*. Thus an optimized retrieval will always be reasonably accurate in a formal sense (taking the retrieval and the uncertainty both into account), however the uncertainties may be so large that the retrieval has very limited utility.

We observe that though each of the constraints tested exhibited a deep minimum in q, the resulting ion distributions vary significantly. We will consider each constraint type in turn.

The identity constraint, H_0 is by far the simplest constraint to implement and optimize. In this case q can be regarded as a simple function of the constraint strength, γ, and a one dimensional optimization can be carried out. In practice, as can be seen in Figure (7), this constraint also provides a reasonable retrieval. Clearly, the peak intensity is underestimated, however the morphology of the ion distribution is accurately reproduced. The underestimation of the peak is not unexpected, as the identity constraint tends to force the retrieval to smaller values. It is important to take into account the contours of constant relative uncertainty in the figure. From these contours it is clear that although the morphology of the ion distribution is produced quite well away from the peak value, it is quantitatively unreliable. Further, even near the peak value there are only a few pixels whose relative uncertainty is less than 50%.

The second derivative constraint, H_2, yields what appears at first glance to be a qualitatively different retrieval.The retrieval near the peak is still underestimated, but to a lesser extent than the H_0 constrained retrieval. The shape of the peak is also better reproduced, in that it is appropriately broader than the H_0 retrieval. Further, the number of pixels with less than 40% relative error has increased. This favorable behavior near the peak is offset by rather poor behavior away from the peak. A spurious second peak in the dawn-noon sector has appeared. This is mitigated somewhat by the relative uncertainties, which are rather large in this region. This secondary peak arises from a combination of the nature of the constraint (i.e., it favors solutions that have a constant slope) and the geometry of the observation. From Figure (6) we see that the field lines in the area of the spurious peak are blocked from view by the Earth. This is an area in which the retrieval is essentially blind, i.e., it is an area that cannot be well determined by the data and is controlled by the constraint. Since the constraint favors equally positive and negative slopes, it is not unusual that the retrieval would grow spuriously in this region. If the uncertainties are taken into account, the H_2 constraint yields a reasonable retrieval, particularly in the region of the peak intensity, however the behavior in the less well determined areas is undesirable.

The Markhov constraint, H_m, yields a retrieval that is similar in character to the identity constraint, H_0. This is not surprising, as the optimal characteristic length, d_m is fairly small (.109), and thus has only small off-diagonal elements and is therefore H_m is very similar to H_0. The retrieval with the H_m constraint is, however, slightly preferable to the retrieval with the H_0 constraint. The range where the uncertainties are less than 40% is slightly larger, and the morphology off-peak is slightly more representative. The behavior of the retrieval near the outer L boundary has somewhat less noise, no doubt because of the introduction of the boundary clamping parameter λ. We therefore consider H_m to be the preferable form for the constraint.

Clearly, the choosing the the constraints is a somewhat subjective enterprize. Different constraints will enhance some aspects of the retrieval at the expense of other aspects. This is clear from the discussion above in regard to the differences between the H_0 and H_m constraints. If we wish to be as careful as possible in the region of the emission peak, we should choose to use the H_2 constraint. If, however, we wish to better reproduce global morphology, the H_m constraint is preferable. In some cases the tradeoffs can be quantified (indeed the method of tuning the constraint parameters represents a quantitative tradeoff). In other cases selections must be made on more qualitative criteria. In these situations, the retrieval metrics, g_r, g_f, and q can be relied upon to signal the danger of misleading results. In general if the value of q is small, the results of corresponding retrievals, taken together with their uncertainties, will be reliable. Whether these results are useful or not is another matter. In some cases a retrieval with a small q may have error bars that are so large that the results are basically useless.

Before concluding, we present the results of the retrieval using the data (Figure (5)) on which the above optimization was based. The results of the retrieval using the H_m constraint (using the optimal parameters) are shown in Figure

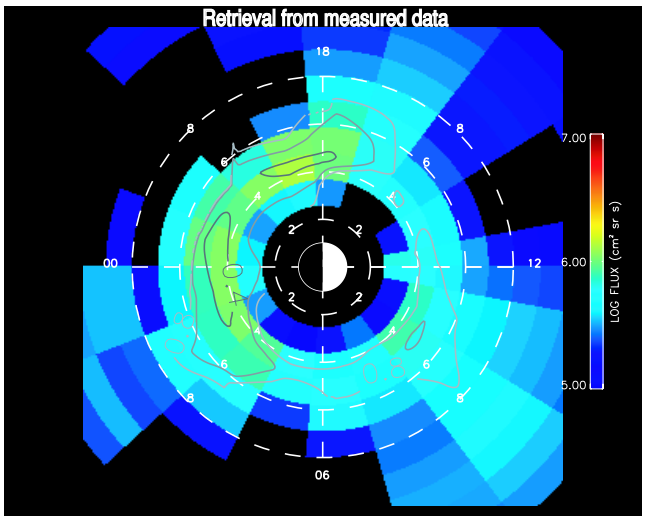

Figure 8. Retrieval results for the image in Figure (5) using the optimal H_m constraints shown in Figure (7). The color scale and legend are the same as in Figure (7).

(8). The retrieval shows quantitatively some of the features that we surmised when comparing the simulated and actual emissions (Figures (6) and (5), respectively). Namely that the peak of the ion intensity is somewhat smaller than in the test image and the peak intensity is much more diffuse.This figure shows a ring current that extends over nearly the entire night-side rather than the sharply peaked spatial distribution shown in the test distribution.

6. Summary and Conclusions

In this work we have developed a technique for the retrieval of global magnetospheric ion densities from high energy neutral atom measurements from the HENA instrument. The basic equations governing the HENA measurements were summarized. These include both the geocoronal source and the source near the exobase. The HENA response function is also explicitly included in our treatment. These equations are treated in more detail by *Roelof* [2002]. After presenting the equations of measurement, simulations based on these equations were shown and their important features described. We then turned to developing the retrieval technique. It is noted that the retrieval of ion distributions is an ill-posed problem, i.e., it has not unique solution. For this reason, we described the introduction of constraints and a quantitative method of determining what they should be. Finally, we calculated optimal constraints for a representative HENA image and performed retrievals on simulated data in order to explore the reliability of the retrievals.

In short, we have developed a retrieval technique and provided a means of quantifying its fidelity. We have also applied these to a representative image and shown that these technique are indeed useful for determination of ion densities. Without this or a similar method of retrieval, only qualitative conclusions can be drawn from the ENA images about the source ion distribution.

There are several ways in which these techniques can be improved. First, much work has yet to be done on refining the constraints. We have presented some very simple examples of how these retrievals can be constrained. A more sophisticated treatment, e.g., one that has a strong physical basis, may yield a significant increase in the quality of the retrievals. Second, in the absence of extremely high count rates (i.e., unless the signal to noise ratio is very high) we can only poorly quantify the pitch angle distribution of the ions. We intend to approach this problem by combining the information in images taken at multiple viewpoints. Early attempts at this have been encouraging, but no detailed study has yet been conducted. Thirdly, the retrieval equations are based on a simple dipole magnetic field. This may lead to errors particularly at higher values of L. We are currently seeking a method of incorporating a more realistic magnetic field without prohibitively increasing the computational cost. Finally, the retrievals are only as good as our knowledge of the instrument. Information about the instrument must be included in the retrievals as it becomes available.

Appendix A: Linear Quadratures

The approximation of double integrals by numerical quadratures is an important element in the techniques developed above. In short, we wish to find an approximation of the form

$$\sum_{k=1}^{N}\sum_{l=1}^{M} b_{kl} f(x_k, y_l) \cong \int_{a_x}^{b_x}\int_{a_y}^{b_y} K(x,y) f(x,y) dx dy \tag{A1}$$

where $K(x,y)$ is some known function, $f(x,y)$ is a continuous function that can be sampled at $N \times M$ points $f(x_k, y_l)$ and b_{kl} are weights which will now be developed.

In order to formulate the approximation (A1), some assumption about the behavior of $f(x,y)$ between the quadrature points, $f(x_k, y_l)$, must be assumed. In the simplest case we assume that $f(x,y)$ remains constant around each grid point and $f(x_k, y_l)$ is given on a regular grid with separations $\triangle x$, $\triangle y$. In this case we obviously have $b_{kl} = \triangle x \triangle y$. A more robust technique would assume some functional dependence of $f(x,y)$ between the quadrature points. The success of the quadrature scheme depends on the accuracy of such an assumption.

In this work, we adapt a one dimensional linear quadrature technique presented by *Twomey* [1977] and briefly summarized here. In the one dimensional version of the problem, we seek an approximation to the integral of the form

$$\sum_{j=1}^{N} b_j f_j \cong \int_{a}^{b} K_1(x) f_1(x) dx \tag{A2}$$

The integral may be broken up into intervals

$$\int_{a}^{b} K_1(x) f_1(x) dx = \sum_{j=1}^{N} \int_{x_j}^{x_{j+1}} K_1(x) f_1(x) dx \tag{A3}$$

If we assume that the function $f(x)$ is linear in each subinterval, we can write

$$\int_{a}^{b} K_1(x) f_1(x) dx = \sum_{j=1}^{N-1} \int_{x_j}^{x_{j+1}} K_1(x)(A_j + B_j x) dx \tag{A4}$$

where A_i and B_i are constants determined by the $2 \times (N-1)$ equations

$$\begin{aligned} f(x_j) &= A_j + B_j x_j \\ f(x_{j+1}) &= A_j + B_j x_{j+1} \end{aligned} \tag{A5}$$

If equations are solved in terms of the $f(x_j)$, the result substituted into equation (A4) and placed in the form of equation (A2) the final expression for the quadrature weights, b_j becomes

$$b_j = \frac{\varphi^1_{j-1} - x_{j-1}\varphi^0_{j-1}}{x_j - x_{j-1}} + \frac{x_{j+1}\varphi^0_j - \varphi^1_j}{x_{j+1} - x_j}, \quad 2 \le j \le N-1$$

$$b_1 = \frac{x_2\varphi^0_1 - \varphi^1_1}{x_2 - x_1} \tag{A6}$$

$$b_N = \frac{\varphi^1_{N-1} - x_{N-1}\varphi^0_{N-1}}{x_N - x_{N-1}}$$

where the moments φ^m_j are

$$\varphi^m_j = \int_{x_j}^{x_{j+1}} x^m K_1(x) dx \tag{A7}$$

Thus the one dimensional quadrature weights in equation (A2) can be expressed through the moments of the function $K_1(x)$ over the subintervals (x_j, x_{j+1}).

The two dimensional analog of the one dimensional quadrature can be derived through repeated application of equation (A6). This process results in the two dimensional quadrature weights b_{kl} from equation (A1)

$$b_{kl} = \sum_{m=0}^{1} \sum_{n=0}^{1} \zeta^{m,n}_{(k+m-1)(l+n-1)} \tag{A8}$$

where

$$\zeta^{nm}_{kl} = \int_{x_l}^{x_{l+1}} \int_{y_k}^{y_{k+1}} g^{nm}_{kl} f dx dy \quad \text{if} \quad \begin{array}{l} 1 \le k \le N-1 \\ 1 \le l \le M-1 \end{array} \tag{A9}$$

$$\zeta^{nm}_{kl} = 0 \quad \text{otherwise}$$

and

$$g^{nm}_{kl} = \frac{(-1)^{n+m}(x_{l+m} - x)(y_{k+n} - y)}{(x_{l+1} - x_l)(y_{k+1} - y_k)} \tag{A10}$$

Thus we have derived a numerical approximation for equations of the form (A1). This approximation is accurate when the function $f(x, y)$ in equation (A1) can be approximated by a bilinear function.

Acknowledgments. Help me here

References

Abramowitz, M., and I. Stegun, *Handbook of Mathematical Functions*, National Bureau of Standards; reprinted by Dover Publications, 1972.

Barnett, C. F., *Collisions of H, H_2, He, and Li atoms and ions with atoms and molecules, in Atomic Data for Fusion*, vol. 1 of *ORNL-6086*, Controlled Fusion Atomic Data Center, Oak Ridge National Laboratory,, Oak Ridge, Tennessee, 1990.

Basu, B., J. Jasperse, R. Robinson, R. Vondrak, and D. Evens, Linear transport-theory of auroral precipitation - a comparison with observations, *J. Geophys. Res.*, *92*, 1987.

C:son Brandt, P., R. Demajistre, E. C. Roelof, D. G. Mitchell, and S. Mende, IMAGE/HENA: Global ENA imaging of the plasmasheet and ring current during substorms, *J. Geophys. Res.,* , 2002, submitted February.

Gibson, W. C., J. L. Burch, J. R. Scherrer, M. B. Tapley, R. L. Killough, F. A. Volpe, W. D. Davis, D. C. Vaccarello, G. Grismore, D. Sakkas, and S. J. Housten, The image observatory, *Space Sci. Rev.*, *91*, 15–50, 2000.

Menke, W., *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press Inc., 1989.

Mitchell, D. G., S. E. Jaskulek, C. E. Schlemm, E. P. Keath, R. E. Thompson, B. E. Tossman, J. D. Boldt, J. R. H. abd G. B. Andrews, N. Paschalidis, D. C. Hamilton, R. A. Lundgren, E. O. Tums, P. Wilson IV, H. D. Voss, D. Prentice, K. C. Hsieh, C. C. Curtis, and F. R. Powell, High energy neutral atom (HENA) imager for the IMAGE mission, *Space Sci. Rev.*, *91*, 67–112, 2000.

Perez, J. D., M.-C. Fok, and T. E. Moore, Deconvolution of energetic neutral atom images of the earth's magnetosphere, *Space Sci. Rev.*, *91*, 421–436, 2000.

Perez, J. D., G. Kozlowski, P. C. Brandt, D. G. MItchell, J. M. Jahn, C. J. Pollock, and X. Zhang, Initial ion equatorial pitch angle distributions from energetic netural atom images obtained by IMAGE, *Geophys. Res. Lett.*, *28*, 1155–1158, 2001.

Press, W., B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes in C*, Cambridge University Press, 1988.

Rairden, R. L., L. A. Frank, and J. D. Craven, Geocoronal imaging with dynamics explorer, *J. Geophys. Res.*, *91*, 13,613–13,630, 1986.

Rodgers, C., *Inverse Methods for Atmospheric Sounding*, World Scientific, 2000.

Roelof, E. C., Energetic neutral atom image of a storm-time ring current, *Geophys. Res. Lett.*, *14*, 652–655, 1987.

Roelof, E. C., xxx, *J. Geophys. Res.*, *xx*, xx, 2002.

Roelof, E. C., and A. J. Skinner, Extraction of ion distributions from magnetospheric ENA and EUV images, *Space Sci. Rev.*, *91*, 437–459, 2000.

Russell, C. T., Geophysical coordinate transformations, in *Cosmic electrodynamics*, vol. 2, pp. 184–196, D. Reidel Publishing Company, 1971.

Twomey, S., *Introduction to the mathematics in remote sensing and indirect measurements*, Developments in geomathematics 3, 1st ed., Elsevier scientific publishing company, 1977.

R. DeMajistre, E.C. Roelof, P. C:son Brandt, D. Mitchel, The Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel MD, 20723. (robert.demajistre@jhuapl.edu)

(Received .)