On the Development of the SAMI2 Ionosphere Model

J.D.Huba

Syntek Technologies, Fairfax, VA

October, 2022

Abstract

The development of the ionosphere model SAMI2 at the Naval Research Laboratory (NRL) is described. The genesis of the code and the adversities we faced in developing the model are described. The evolution of the numerical algorithms used is discussed as well as the decision to open-source the code. An example of a new discovery made with the code, the formation of an electron 'hole' in the nighttime, high-altitude, low-latitude ionosphere, is given.

1 The Beginning

In the late 1990s I initiated a major research program at the Naval Research Laboratory (NRL) in ionospheric physics. It was an internally funded Accelerated Research Initiative (ARI) that was cross-divisional and involved the Plasma Physics Division and the Space Science Division at NRL. I was in the Plasma Physics Division and my Section focused on theory and modeling of the ionosphere, while the Space Science Division provided experimental and observational data. The Principal Investigator of the ARI was Dr. Robert Meier of the Space Science Division. The overarching theme of the program was to advance our understanding of the ionosphere through a collaborative theory/modeling/data program.

A preliminary research proposal was first developed and submitted to the Director of Research of NRL, Dr. Timothy Coffey, for his thoughts; specifically, would he support such a large effort (roughly $1M/yr for 5 years). One of the research tasks in the proposal was to develop a new ionosphere model in my Section. Dr. Coffey's assessment of the proposal was extremely favorable and believed it would be supported by laboratory management with one caveat - there would be no ionosphere model development in the program. If an ionosphere model was needed then an existing model could be used. The proposed ARI was subsequently approved by NRL management *sans* model development.

We began the ARI in my Section focusing on the theoretical analysis of ionospheric instabilities but quickly realized that an ionosphere model was needed to make progress, especially in connecting to observational data. We contacted Dr. Phil Richards, the developer of the Field Line Interhemispheric Plasma (FLIP) model [*Richards and Torr*, 1996], and requested the source code which he kindly provided. I asked my colleague Dr. Glenn Joyce to run the FLIP model and assess how we could use it for our research. After a week or so I went to Glenn's office and asked him the status of FLIP. He said he could compile the code, run it, and get results. He then added that he didn't think we would be able to modify the code because of its complexity. At that point I said, then let's build our own ionosphere model, despite being explicitly 'prohibited' from doing this, and, quite frankly, having no idea on how to build a model of the ionosphere. Glenn agreed and our more than decade-long collaboration on the SAMI2/3 project began.

2 Off to Building SAMI2

In building a new ionosphere model we had two important assets to draw upon. First, an ionosphere model had been developed at NRL in the mid-seventies [*Oran et al.*, 1974] which we had access to. Albeit a very simplistic model, it was a 1D model that considered plasma dynamics along a single, straight magnetic field line at mid-latitude in the altitude range 90 - 1000 km, it was clearly written and well-documented. As such, it provided an architecture upon which to build a more general and realistic ionosphere model. Second, there were two articles published in the Step Handbook [*Schunk*, 1996]: Bailey and Balan (1996) and Millward et al. (1996). The Bailey and Balan (1996) article described the Sheffield University Plasmasphere Ionosphere Model (SUPIM) while the Millward et al. (1996) article described the Coupled Thermospheric-Ionospheric-Plasmasphere Model (CTIP). Both papers provided very detailed information regarding the models: the relevant physics equations, photoionization, chemical reactions and rates, heating rates, collision frequencies, geomagnetic field, and finally, numerical methods used in the models.

Armed with these assets, Glenn and I began work on a new ionosphere model in the summer of 1998. The early collaboration split the work into distinct components that we worked on independently. For example, Glenn focused on the implicit solvers for the continuity, velocity, and temperature equations, while I focused on photoionization, chemistry, and cross-field transport. As each component matured we worked together on unifying the various subroutines into a workable model. This initial phase of the project took about one year and we were obtaining 'reasonable' results, i.e., electron density profiles that resembled observations.

3 To Continue or Not Continue

As noted above, this research project was an internally funded at NRL. Internally funded research programs at NRL are subject to external reviews every 3 years. A panel of scientists from universities and other research laboratories would hear presentations from NRL scientists about their research and write a report for lab management on their assessment of the ongoing research addressing various questions: Is the research worthwhile? Good? Competitive? A worthwhile investment by the lab?

The review was set for October, 1999. I had about 4 - 5 research areas to present. One of them was on the ionosphere model Glenn and I were working on. I was actually excited about this because I thought we made an impressive amount of progress on the code and our preliminary results were encouraging.

On the day of the review I started my presentations. Eventually I placed a vugraph on the projector with the title 'Ionosphere Modeling' at which point a member of the Lab's upper management who was in attendance said in a loud voice, 'Hey, you guys aren't supposed to be building an ionosphere model. What's going on?' I was stunned that (1) he said this in front of the external review panel members, and (2) he actually remembered that we were told not to build an ionospheric model since the program had been approved several years earlier. Undaunted I proceeded with my presentation and thought that when everyone saw the progress Glenn and I had made all would be forgiven.

All in all the review went well as far as I could tell; the panel members seemed impressed overall with the research being done in space physics. The next step for the panel was to write a report of their findings and submit it to the director of the laboratory.

About a month later I received a copy of the report from the lab director with a note attached 'Joe, please respond.' The first page of the report was boilerplate stuff - how much the panel enjoyed visiting NRL, what a great place it is, what an impressive research staff, ... On page two they had specific recommendations. The number one recommendation was that 'NRL should cease and desist numerical modeling of the ionosphere immediately.' The primary reasons were that there were insufficient funds for a project of this magnitude and we didn't have the right personnel for the job.

I went over the report with Glenn and then proceeded to write a lengthy defense of our program to the director. I responded to each criticism in detail and explained why the panel was completely wrong and we should continue our modeling effort. Days passed I had not heard back from the director regarding the external review and my response. After about two weeks Glenn came in my office and asked if I had heard anything. I told him no - he asked then what are we going to do? I said, well we haven't been told to stop working on the model so let's keep going. And we did. My suspicion was that the director was not remiss in not responding, rather, he was leaving the 'ball in our court.'

4 But All Was Not Well

However, there were numerical issues solving the ion temperature equation: the implicit solver often failed and the code became unstable. This problem could be 'fixed' by simply ignoring one of the terms in the ion temperature equation that was small relative to the other terms, but we didn't feel this was an acceptable solution. To put into perspective the numerical problem and our solution, a brief digression on the equations and numerical methods used to solve them is warranted. A more complete discussion of this topic is given in Huba and Joyce (2014).

Only the dynamics of the density and velocity along the geomagnetic field are considered for illustrative purposes:

$$\frac{\partial n_i}{\partial t} + b_s^2 \frac{\partial}{\partial s} \frac{n_i V_{is}}{b_s} = \mathcal{P}_i - \mathcal{L}_i n_i \tag{1}$$

$$\frac{\partial V_{is}}{\partial t} + V_{is} \frac{\partial V_{is}}{\partial s} = -\frac{1}{n_i m_i} b_s \frac{\partial (P_i + P_e)}{\partial s} + g_s - \nu_{in}(V_{is} - V_{ns}) - \sum_{j(\text{ions})} \nu_{ij}(V_{is} - V_{js}) \tag{2}$$

The subscript s on vector quantities indicates the component of the vector in the s-direction. In the above, $\mathcal{P}_i$ is the production term associated with photoionization and chemistry (e.g., charge exchange), $\mathcal{L}_i$ is the loss term associated with chemistry (e.g., charge exchange, recombination), ν_{in} and ν_{ij} are the ion-neutral and ion-ion collision frequencies.

4.1 Implicit Method

The key assumption of an implicit algorithm to solve Eqs. (1) and (2) is to neglect ion inertia in Eq. (2)

$$0 = -\frac{1}{n_i m_i} b_s \frac{\partial (n_i T_i + n_e T_e)}{\partial s} + g_s - \nu_{in}(V_{is} - V_{ns}) - \sum_{j(\text{ions})} \nu_{ij}(V_{is} - V_{js}) \tag{3}$$

where we have used the definition of pressure $P = nT$. This is a very good assumption for the ionosphere because it is collision dominated and ion inertia is not important.

The basic procedure is to solve Eq. (3) for V_{is} as a function of n_i (and the other variables) and substitute it into Eq. (1). The time discretization of Eq. (1) is then written as

$$(\frac{1}{\Delta t} + \mathcal{L}_i) n_i^{t+\Delta t} + b_s^2 \frac{\partial}{\partial s} \frac{n_i^{t+\Delta t} f(n_i^{t+\Delta t}, ...)}{b_s} = \frac{n_i^t}{\Delta t} + \mathcal{P}_i \tag{4}$$

where $f(n_i^{t+\Delta t}, ...)$ denotes the solution to V_{is}. Except for the right-hand-side of Eq. (4), the ion density n_i is defined at the upper time level $t + \Delta t$; this is the crux of the fully implicit scheme. Defining the spatial discretization as $\partial g/\partial s = (g_{j+1} - g_{j-1})/\Delta s$, one can write Eq. (4) in tridiagonal form

$$An_{i,j-1}^{t+\Delta t} + Bn_{i,j}^{t+\Delta t} + Cn_{i,j-1}^{t+\Delta t} = D \qquad (5)$$

which can be solved for $n_i^{t+\Delta t}$ using standard numerical algorithms (*Press*, 2003).

The above is a broad overview of the fully implicit differencing technique used in ionospheric modeling. One technique to solve these equations is to solve Eq. 3 for V_{is} and substitute it into Eq. 1. However, calculating V_{is} involves all ion species because $n_e = \sum_i n_i$; it's solution can be an algebraic nightmare because it involves a number of coupled equations. Another technique to solve Eqs. (1) and (3) is through iteration. The idea is to solve Eq. (3) for V_{is} without expanding the electron density into the ion densities. This allows V_{is} to be written in terms of n_i directly. The system of ion equations can then be solved where $n_e = \sum_i n_i^t$ is incorporated into D in Eq. (5). At each time step the equations are iterated until n_e no longer changes. This method was used in the 1D ionosphere model described in Oran et al. (1974). A shortcoming of this method is that there is no guarantee *a priori* that the solution will converge.

The primary benefit of solving Eqs. (1) and (3) fully implicitly is that a relatively large time step can be used. For example, time steps of 5 - 15 min are commonly used in ionospheric simulations using fully implicit schemes. On the other hand, the scheme can be unstable (as we discovered) and was problematic in the collisionless plasmasphere. To lowest order, the ion velocity given by Eq. (3) is proportional to ν_{in}^{-1}. At high altitudes ν_{in} becomes very small and the ion velocity becomes unphysically large.

4.2 Semi-Implicit Method

Recognizing our problems with the fully implicit scheme, Glenn recommended that we include ion inertia in ion velocity equation and simply time advance the velocity (as well as time advancing the ion temperature equations). The difference equation for continuity is written as

$$\frac{n_{i,j}^{t+\Delta t} - n_{i,j}^{t}}{\Delta t} + \frac{(n_i^{t+\Delta t} V_{is}^t)_{j+1/2} - (n_i^{t+\Delta t} V_{is}^t)_{j-1/2}}{\Delta s_j} = \mathcal{P} - n_{i,j}^{t+\Delta t}\mathcal{L} \tag{6}$$

where $\Delta s_j = (s_{j+1} - s_{j-1})/2$ and $\mathcal{P}$, and $\mathcal{L}$ are evaluated at time t. The density is evaluated at the upper time level $t + \Delta t$ so that the difference scheme is implicit (i.e., backward biased). However, the velocity V_{is} is evaluated the current time level t so the scheme is only 'semi-implicit.' This method allows the Courant condition ($\Delta t < \Delta l/V$) to be based upon the advection velocity $V = V_{is}$ and not the sum of the advection velocity and the sound speed $V = V_{is} + C_s$.

The major drawback of this technique is that relatively small time steps are required: several seconds as opposed to several minutes for the fully implicit scheme. However, the advantages of this technique are (1) it's relatively simple to code, (2) it's stable (for sufficiently small time steps), it's flexible (i.e., additional ions are easily added), and (3) it provides a better description of the collisionless plasmasphere.

5 Something New?

In the summer of 2000 I hired a high school student as a summer intern (a classmate of one of my daughters). She worked on the graphics for SAMI2 using IDL. One afternoon I went to her office to see how things were going. She was making contour plots of the electron density as a function of latitude and altitude. The key issue we were looking into was the development of the Appleton anomaly, ionization crests that maximize roughly $\pm 15°$ off the magnetic equator and are caused by the so-called 'fountain effect' - a combination of the 'vertical' E $\times$ B drift and plasma motion along the magnetic field. The ionization peaks usually occur at altitudes 300 - 400 km so typical contour plots of the electron density usually had a maximum altitude in the range 800 - 1200 km. However, on this day she had the maximum altitude of the contour plot set at 4000 km. What caught my eye was not the Appleton ionization crests but an 'electron hole' at an altitude $\sim$ 2000 km. This seemed to be an unusual feature since I thought the electron density should monotonically decrease with altitude above the F peak. And I had never seen this type of feature reported in the literature.

This raised concerns that there was a numerical artifact in the model that gave this result. Glenn and I spent the next month investigating this result to determine if it was indeed a numerical problem but concluded it wasn't and that the result was physical. The 'electron hole' (roughly a factor of two lower in density than the surrounding plasma) was produced by transhemispheric O^+ flows that collisionally couple to H^+ and transport it to lower altitudes, thereby reducing the electron density at high altitudes. The transhemispheric O^+ flows are caused by an interhemispheric pressure anisotropy that can be generated by the neutral wind, primarily during solstice conditions.

Although we were confident that our finding was real it seemed the development of a high altitude electron density 'hole' should also be observed in other ionosphere models. At a meeting in Arecibo subsequent to the publication of our paper I met a post-doc named Brian MacPherson who used the Sheffield University Plasmasphere Ionosphere Mode (SUPIM) in his doctoral thesis. I discussed this new phenomenon with him and asked him if he could do a SUPIM run for the conditions used in our paper. He agreed and sent me several contour plots of the electron density several weeks later. And I was pleasantly surprised to see an 'electron hole' at high altitude similar to the SAMI2 results.

Having two ionosphere models give the same result was comforting but having observational data to support this finding would be the 'icing on the cake.' However, the altitude range 1500 - 2500 km is not commonly covered since low earth orbit (LEO) satellites are at altitudes *lesssim* 800 km. I spoke to some colleagues at Goddard Space Flight Center (GSFC) (Joe Grebowsky, Walt Hoegy, and Larry Brace) about this issue and found out there were two electrostatic cylindrical probes, operating as Langmuir probes, on the International Satellites for Ionospheric Studies (ISIS-1) satellite that operated in this altitude range. [Larry Brace was the PI of the instrument.] The data was not digitized but presented as line graphs in an internal GSFC research report. Interestingly, a dip in the electron density, by roughly a factor of two, was observed in the altitude range 1500 - 2500 km at local midnight during June of 1969, consistent with our results. Recently the ISIS-1 data has been digitized (https://omniweb.gsfc.nasa.gov/ftpbrowser/atmoweb.html) and an interesting project would be to do a more detailed data/model comparison.

6 What About Cross-field Transport

Section 4 describes plasma dynamics along the magnetic field. However, there is also plasma motion across the magnetic field associated with the E × B drift. In the original version of the model we used a Lagrangian scheme for cross-field transport. In the this method the motion of 'flux tubes' is calculated based on the $E \times B$ drift velocity. The ion density is updated based on conservation of particles and magnetic flux [*Huba et al.*, 2000a]. But this method is problematic, especially at high latitudes where the motion of flux tubes caused by the high latitude convective potential can lead to regions devoid of flux tubes and regions of closely-packed flux tubes. Alternatively, one can perform a Lagrangian 'push' of the plasma and then interpolate to a fixed grid to avoid the aforementioned problem. A shortcoming of this method is that it is diffusive.

To overcome the deficiencies using the Lagrangian method I thought it best to use a fixed grid. I developed a non-uniform, orthogonal dipole grid and used the donor cell method for cross-field transport. This technique was used in a Hall MHD code I was developing at the time [*Huba*, 2003]. The new code performed well and reproduced the results of the Lagrangian code. However, it could not be easily extended to high latitudes and retain complete dipole field lines because of the orthogonality condition. One could impose an altitude limit on the grid but this would lead to the implementation of boundary conditions at the top boundary which in itself could be problematic.

At a Fall AGU meeting (maybe 2000) I complained to a colleague John Lyon about this problem. He suggested that I lay, say, 200 points along each field line, with the base of each field below the E region ($\sim$ 85 km), and then simply 'connect the dots.' This would solve the problem but I noted that the grid would now be non-orthogonal. If my memory serves correct, his response was along the lines of 'deal with it.'

I spent much of my Christmas break that year at my parents house with my laptop developing the geometric factors (cell volumes, normals, face areas, etc.) needed for cross-field transport on a non-orthogonal grid. Again, the donor cell method was used as in the orthogonal grid code. The contrast between the orthogonal Eulerian and non-orthogonal Eulerian grids is shown in Fig. 1. A comparison of the electron density using the orthogonal and non-orthogonal grids is show in Fig. 2 which shows electron density contour plots as a function of latitude and altitude. The results are consistent between the two grids. However, for the orthogonal grid (top) there is a 'rattiness' in the E-region and a minor discontinuity near the outer field line associated with the boundary conditions. These 'features' do not occur for the non-orthogonal grid (bottom).

7 To Open Source or Not To Open Source

One of the first meetings in which I presented results from the new SAMI2 model was at the International Symposium on Equatorial Aeronomy (ISEA-10) in Antalya Turkey in 2000. Overall my presentation was well-received but there were several comments expressing skepticism given the model was 'brand new' and used new numerical algorithms.

While at the airport leaving the meeting I discussed the model with a colleague

Vince Eccles. He was favorable about its development and thought it would be a good addition to the aeronomy community. I pointed out that given the effort Glenn and I put into developing the code it would be nice if SAMI2 could be used by other researchers in the ionosphere community. Vince agreed and I said that the way to do this would be to 'open source' the code. Note, this was very early in the days of open sourcing and not quite in the mainstream for geophysical models.

When I returned to NRL I told Glenn we should open source SAMI2. He was not enthusiastic about this. He viewed it as a 'lose-lose' proposition. If the code was used and didn't work we would get bad publicity, and if it did work perhaps we wouldn't receive credit. He was sensitive to this because it happened to him several years earlier with an electron beam code he had written. [His code was correct but was run incorrectly by another scientist who made critical comments about the code at a workshop.]

Despite Glenn's concerns I discussed the matter with the Plasma Physics Division superintendent Sid Ossakow. He was also opposed to open sourcing SAMI2. But his concern was that we would be giving up our competitive edge in future funding opportunities if other research groups had access to the code. My position was that the current version of SAMI2 was simply a stepping stone to a more comprehensive ionosphere model and we wouldn't be losing any advantage in future proposals. Finally he said if I wanted to continue down this path I should discuss it with the Director of Research Timothy Coffey.

I made an appointment to see Dr. Coffey and met with him to discuss open sourcing SAMI2. When I asked him if NRL would allow me to open source SAMI2 his response was basically 'it's your code - you can do with you want with it.' He said there was no generic proscription against open sourcing codes at NRL and noted that several Navy funded codes were already available to the public. After raising a number of issues he advised I contact the NRL legal department and have them develop a disclaimer to be added to the beginning of the source code if I decided to open source the code.

So I asked the legal department at NRL to write a disclaimer for the open sourced SAMI2 code and released it. I spent a couple of weeks learning HTML and developing a web site for the initial release. All in all the process went smoothly. Over the next decade or so the code was downloaded several hundred times and used in a variety of ionosphere research projects and Ph.D. theses. Eventually NRL stopped supporting in-house web sites for codes and the open sourced version is now on GitHub [https://github.com/NRL-Plasma-Physics-Division/SAMI2].

8 Some After Thoughts

After the basic SAMI2 model was finished by Glenn and myself, a number of other scientists have made substantial and valuable contributions to the model - Marc Swisdak, Jon Krall, Paul Berhnardt, Joel Fedder, and Roger Varney to name a few. Additionally, users reported problems with the code which we did not foresee. One example, in the same week, two users reported the code simply did not work: 'NaN' printed in the screen output. They were considering very low solar activity conditions, i.e., F10.7 = 70. We had never tested the code for these conditions. We were able to find the errant code in the photoionization subroutine and fix it. Thus, one benefit of open sourcing the code was to have it be tested by many users under a wide range of conditions to uncover bugs and problems that could be fixed.

As noted in the beginning, development of what came to be SAMI2 was not initially supported by NRL management. The reason for this was understandable: the funding and personnel level of the program was considered insufficient to support a large-scale model development effort. The original, simplistic ionosphere model developed at NRL [*Oran et al.*, 1974] involved 7 scientists with distinct capabilities to contribute to the model (e.g., numerics, chemistry, photoionization) and we did not have this type of personnel in our group. Additionally, an external panel of experts reviewing our program strongly recommended that the modeling effort Glenn and I had embarked on be stopped immediately. However, the director of research, Dr. Timothy Coffey, did not order us to discontinue our model development despite our 'unauthorized' effort and the recommendation of the panel. I can only presume he had confidence in Glenn and me, and that we were worth the investment. Ultimately we were extremely fortunate that NRL management allowed us to continue our research and trust our judgment.

9 Final Word

Following the development of SAMI2, Glenn and I worked side by side for the next 10 years developing SAMI3, a three-dimensional model of the ionosphere [e.g., *Huba and Joyce*, 2010]. During the last several years of our collaboration Glenn battled cancer. He scheduled his chemotherapy to maximize the amount of time he could come to NRL and work on SAMI3 with me. Sadly, he passed away in December 2011. His contributions to both SAMI2 and SAMI3 cannot be overstated.

Acknowledgments: This work has been supported by the National Science Foundation (AGS1931415).

References

Bailey, G.J. and N. Balan, A low-latitude ionosphere-plasmasphere model, in *STEP: Handbook of Ionospheric Models*, ed. R.W. Schunk, p. 173, Utah State Univ., Logan, Utah, 1996.

Huba, J.D., G. Joyce, and J.A. Fedder, SAMI2 (Sami2 is Another Model of the Ionosphere): A New Low-Latitude Ionosphere Model *J. Geophys. Res., 105*, 23,035, 2000a.

Huba, J.D., G. Joyce, and J.A. Fedder, The formation of an electron hole in the topside equatorial ionosphere, *Geophys. Res. Lett. 27*, 181, 2000b.

Huba, J.D., A Tutorial on Hall Magnetohydrodynamics, in *Space Simulations*, eds. M. Scholer, C.T. Dum, and J. Büchner (Springer, New York) p. 170, 2003.

Huba, J.D. and G. Joyce, Numerical methods in modeling the ionosphere, in *Modeling the Ionosphere/Thermosphere System*, eds. J.D. Huba, R.W. Schunk, and G.V. Khazanov, (Geophysical Monograph, ISSN 0065-8448), p. 49, 2014.

Huba, J.D. and G. Joyce, Global modeling of equatorial plasma bubbles, *Geophys. Res. Lett. 37*, L17104, doi:10.1029/2010GL044281, 2010.

Milward, G.H., R.J. Moffett, W. Quegan, and T.J. Fuller-Rowell, A Coupled Thermospheric- Ionospheric-Plasmasphere Model (CTIP), in *STEP: Handbook of Ionospheric Models*, ed. R.W. Schunk, p. 207, Utah State Univ., Logan, Utah, 1996.

Oran, E.S., T.R. Young, D.V. Anderson, T.P. Coffey, P.C. Kepple, A.W. Ali, and D.F. Strobel, A Numerical Model of the Mid-Latitude Ionosphere, NRL Memo Report 2839, 1974.

Richards, P.G. and D.G. Torr, The field line interhemispheric plasma model, in *STEP: Handbook of Ionospheric Models*, ed. R.W. Schunk, p. 281, Utah State Univ., Logan, Utah, 1996.

Schunk, R.W. (Ed.), *STEP: Handbook of Ionospheric Models*, Utah State Univ., Logan, Utah, 1996.

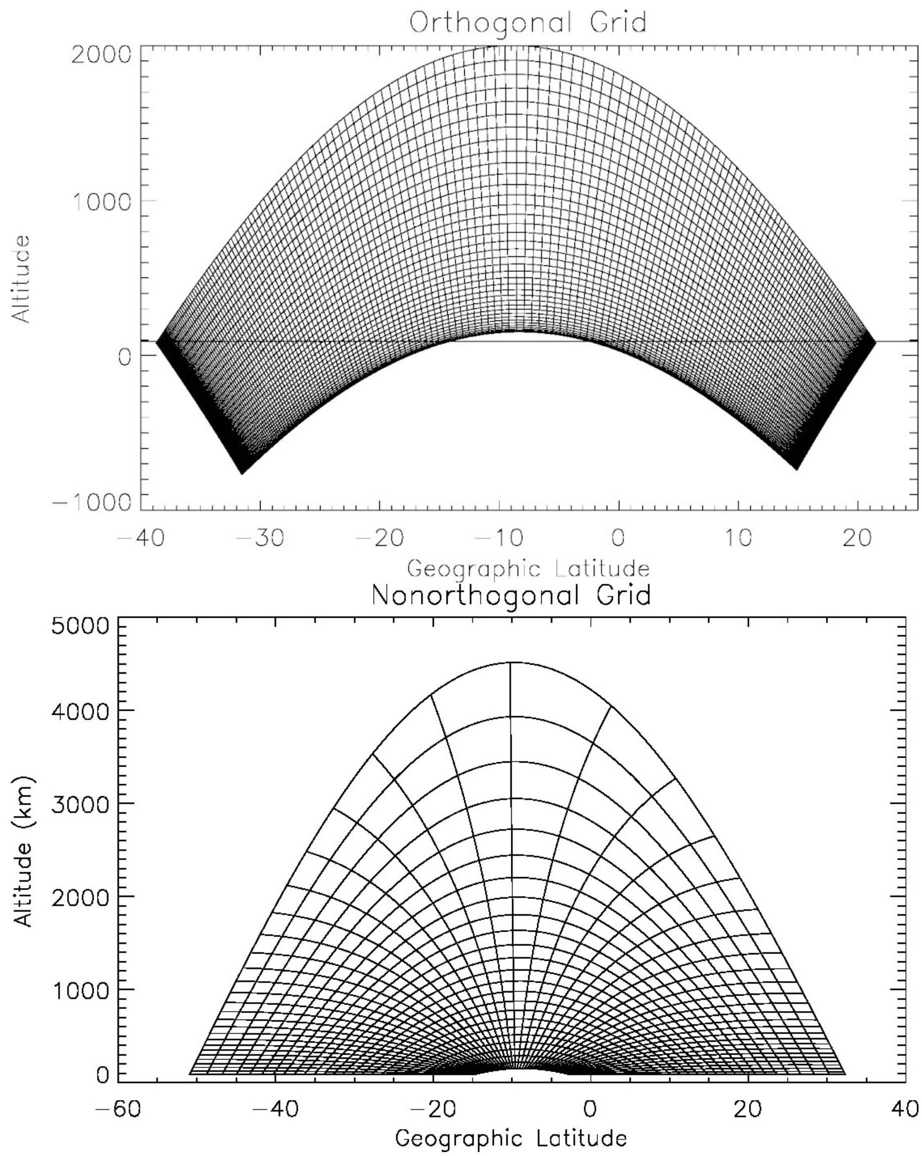

Figure 1: The SAMI2 (a) orthogonal Eulerian grid and (b) non-orthogonal Eulerian grid.

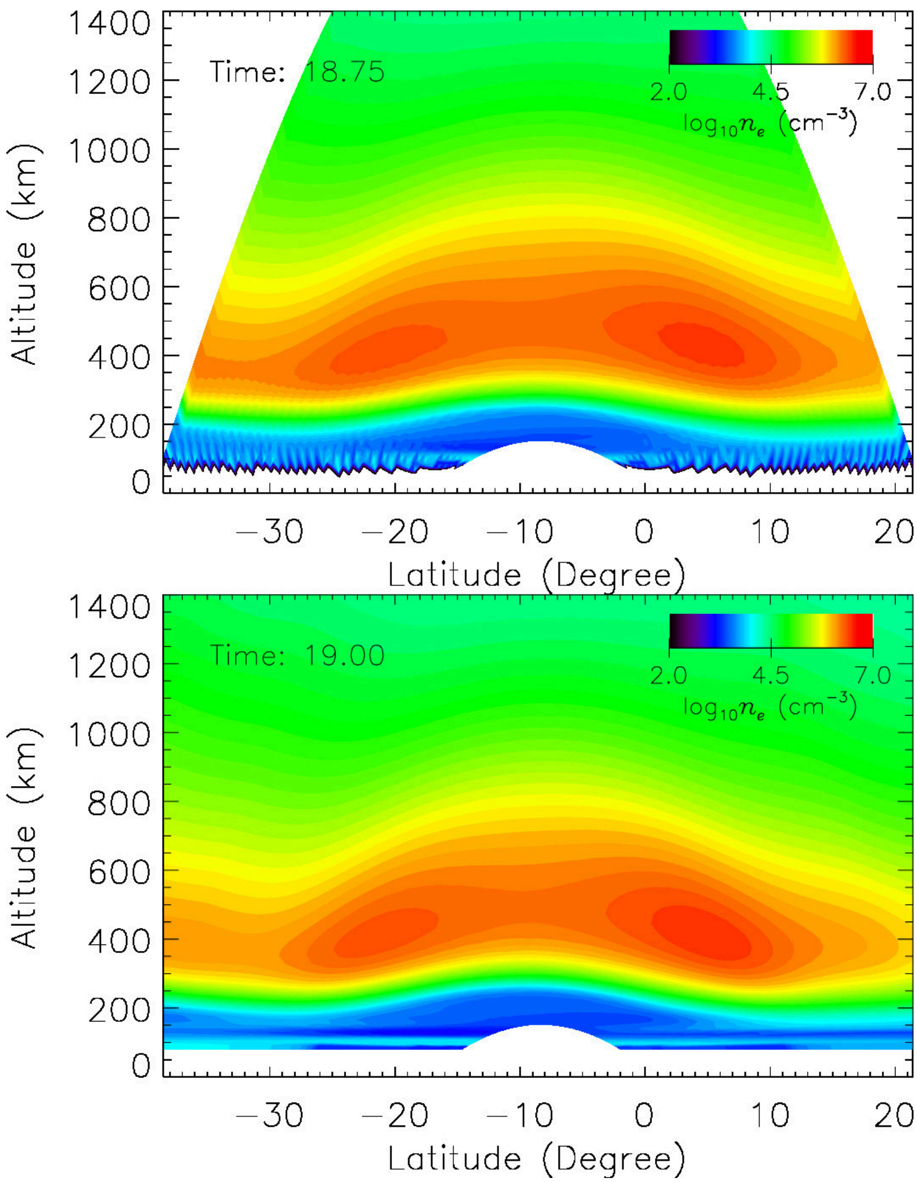

Figure 2: Comparison of electron density for the SAMI2 (a) orthogonal Eulerian grid and (b) non-orthogonal Eulerian grid.