Modeling and Forecasting Aurora

Modeling the physical processes needed for forecasting space-weather events requires multiscale modeling. This article discusses several models researchers use to treat the various auroral processes that influence space weather.

Over the past decade, space weather has become a growing concern in our technology-dependent society; some of its negative effects include malfunctioning satellites, human radiation exposure in airplanes or manned space flights, interruption of telecommunication and navigational systems, power transmission failures, and pipeline corrosion. The high latitudes that interest researchers involved with the International Polar Year (IPY) are particularly sensitive to space weather.

The most impressive geophysical space-weather effect is aurora, the light that appears in the upper atmosphere of Earth’s polar regions. During quiet times, aurora is typically confined to the region at 65° to 70° latitude, but an extreme geomagnetic storm in October 2003 caused it to be seen as far south as Athens, Greece, and the US–Mexico border (30° to 35° latitude). Researchers reported similar observations during the first International Geophysical Year (which was also the third IPY, 1957–1958), when people as far south as Mexico City spotted it.

To forecast the entire chain of events that lead to space-weather effects on Earth requires modeling the physical processes from the surface of the sun, the interplanetary medium, the near geospace environment, the upper atmosphere, and the interaction and coupling of these various regions. For this article, I’ve selected two topics: the modeling of the solar wind to predict the timing of space-weather events, and the modeling of the aurora.

The Auroral Process

Aurora is caused by energetic particles that penetrate into an increasingly dense gas and then lose their energy and momentum through collisions, which in turn leave upper atmospheric constituents in excited states. These energetic particles are protons and electrons, but only the electrons cause visible aurora. Cascading to lower energy states and eventually the ground state causes the excited atmospheric atoms and molecules to emit photons and thus the aurora’s light.

As the electrons penetrate into the atmosphere, they produce secondary electrons in ionizing collisions that then undergo elastic and inelastic collisions of their own, but all these electrons remain near the flux tube defined by the magnetic field line from which they originated. Proton collisions, on the other hand, cause
excitation, ionization, and electron capture; an energetic proton can strip an electron off a neutral atmospheric atom or molecule and thus become neutral hydrogen. Most of the momentum and kinetic energy remains with this newly created hydrogen atom, which isn’t bound to the magnetic field and can travel a large horizontal distance before making another collision, which might turn it back into a proton. Proton aurora is thus spread out horizontally, whereas electron aurora is confined to vertical structures aligned with the magnetic field.

**Aurora Forecasts**

As Figure 1 illustrates, the sun is the ultimate energy source of space weather and aurora. Its upper atmosphere is heated and accelerated to supersonic speed and then expands as solar wind out into space. Events on the sun, such as flares or coronal mass ejections (CMEs), send perturbations and shock waves through the solar wind; this fully ionized plasma also carries with it the sun’s magnetic field (which ultimately becomes the interplanetary magnetic field [IMF]). The superposition of the rotating sun and the radially expanding solar wind creates a spiral magnetic field structure analogous to water expelled from a lawn sprinkler.

Interaction of the solar wind with the Earth’s geomagnetic field and the plasma contained within it (the magnetosphere) pushes the solar wind around the Earth. Because the solar wind flows at supersonic speed, a shock wave precedes its magnetopause, the interface between the solar wind and the magnetosphere. We can estimate the magnetopause’s standoff distance on the sun-facing side from the balance of the interacting plasmas’ dynamics and magnetic pressure. On average, this...
distance is roughly 10 Earth radii (RE) at the sub-
solar point. On the downstream side, the magneto-
sphere forms a long magnetotail, which extends
past the lunar orbit.

Models
The energy transferred from the solar wind to
the magnetosphere is dissipated by the magneto-
sphere’s interaction with the upper atmosphere,
remains in the magnetosphere for extended peri-
ods, or returns to the solar wind by the ejection of
plasmoids from the magnetotail. A side effect of
this energy dissipation process is the aurora. An
overall aurora forecast model must therefore con-
sist of several other models:

- a solar wind model driven by solar observations
to predict the IMF and solar wind conditions
near Earth;
- a model describing how the solar wind and the
magnetosphere interact;
- a model of the magnetosphere itself, to predict
the processes that lead to auroral particle accel-
eration; and
- a model that describes the upper atmosphere’s
response to auroral particles.

The timescales in these models differ widely—
the solar wind takes two to three days to reach
Earth, for example, but the magnetospheric
response that produces aurora happens on the
scale of hours or minutes. For forecast purposes,
it’s thus most desirable to have an operational
model of both the solar wind and the IMF. Sev-
eral approaches for solar wind models exist. An
alternative to physics-based models is statisti-
cal prediction, which works best for short-term
forecasts of geomagnetic and auroral activ-
ity from observed solar wind parameters. Such
models require satellites in stable orbits around
the first Lagrangian point (L1) just upstream
from Earth (the solar wind takes roughly an
hour to reach the magnetosphere from L1). In
situ observations near L1 provide the input for
short-term forecasts.

Researchers have developed several first-prin-
ciple models of the solar wind and its interaction
with the magnetosphere–ionosphere system. One
solar wind model, the Hakamada-Akasofu-
Fry (HAFv.2) model, achieved operational sta-
tus in October 2006. It uses a kinematic approach
to solve for solar wind conditions everywhere in
the solar system; its inner boundary is set at 2.5
solar radii. The model derives background cond-
tions from daily solar surface observations,
which provide synoptic magnetic field maps for
extrapolation to describe the magnetic field at
2.5 solar radii. The solar wind’s speed comes
from the divergence of the magnetic field’s flux
tubes and provides the HAFv.2 model’s inner
boundary condition.

Significant events on the sun—flares and
CMEs, for example—can launch a shock that
propagates though the solar wind. Forecast-
ers characterize these disturbances both by
their location on the sun and the expanding
shock wave’s speed and duration. They can
obtain such parameters from X-ray observa-
tions of flares from space (Geostationary Op-
erational Environmental Satellites [GOES]
X-ray detectors and X-ray imagers) and radio
burst detections or the CME imaging from the
Solar and Heliospheric Observatory/Large An-
gle and Spectrometric Coronagraph Experiment
(SOHO/LASCO). Figure 2 shows an example of
a CME that LASCO observed on 5 December
2003; an occultation disk blocks direct sunlight,
and density variations are made visible by sunlight
that scatters of electrons in the solar wind. The
recently launched Solar Terrestrial Relations Ob-
servatory (STEREO) spacecraft pair will greatly
improve CME imaging with higher resolution
and the ability to triangulate on CMEs.
Shock Propagation
As previously mentioned, disturbances propagate through the solar wind as shocks; depending on their intensity and the magnetic field’s direction, these shocks can decay into magneto-hydrodynamics (MHD) waves. The HAFv.2 model predicts the arrival time along with solar wind parameters on Earth, providing roughly two to three days’ lead time for forecasting large auroral events and space weather disturbances. Figure 3 shows an example from HAFv.2 on 4 November 2003; the figure shows magnetic field lines in the ecliptic plane—blue toward the sun and red away from it. The positions of Venus, Earth, and Mars are indicated, and a shock propagating to the right is clearly apparent by the disturbance in the magnetic field’s direction and strength. This is one of many shocks that caused the very intense auroral displays at the end of October 2003. The same active region on the sun rotated around and caused significant aurora again in the second half of November.

Input parameters at the sun cause the greatest uncertainty in predicting shock arrival times. The interpretation of radio burst data or CME images to obtain the initial speed is only partially supported by modeling and requires subjective interpretation—for example, the record-breaking storm in October 2003 had unusually high solar wind speeds, but they seemed so unrealistic that initial shock arrival forecasts were off by a large margin (the solar wind took only 18 hours to reach Earth instead of the usual two to three days). Overall, though, forecast success has been satisfactory—in a study of 173 events, researchers concluded that the accuracy of the shock arrival forecasts was within 11 hours, with a detection probability of 85 percent.

The Geophysical Institute at the University of Alaska has published an aurora forecast (www.gedds.alaska.edu/AuroraForecast/) based on the HAFv.2 model’s predictions for several years. This forecast predicts the time when aurora is expected to intensify from a predicted shock arrival time and the latitude to which it should be visible from the solar wind’s predicted speed and density.

Auroral Processes and Structure
The aurora is caused by electrons and protons accelerated to kinetic energies of a few hundred eV to tens of keV above the thermosphere, the uppermost part of the neutral atmosphere. Several mechanisms attempt to explain how the acceleration happens, but this research subject is still unresolved. The aurora has a characteristic spatial structure: thin curtains of light, sometimes 100 m or less, extend over lengths in excess of 1,000 km. These curtains can be homogeneous or can themselves contain structures with a horizontal scale length of 100 m or less. This poses the largest difficulty for models attempting to capture the acceleration mechanism. In a 1993 paper, for example, Joe Borovsky analyzed 21 models for auroral electron acceleration and dismissed them all for failing to generate the observed structure, although they were all plausible acceleration mechanisms and could generate the observed energies.

Examples of Observational Difficulties
The structure and temporal evolution of visible aurora isn’t just an important model requirement—it also poses a significant observational problem. Two common methods for observing aurora are imaging the aurora from the ground or flying a satellite or rocket through the aurora—for example, for in situ measurements of the energetic electron flux. Low light levels require either long exposure times or light amplification in intensified detectors (Figure 4 shows an example of an intensified image with short exposure time). The more structure an aurora has, the faster its movement;
therefore, long exposure times tend to show diffuse images of aurora.

*In situ* observations of the auroral particles from rockets or satellites suffer from the same problem. A fast-moving satellite or rocket spends a very short time inside the auroral structure and might not resolve typical spatial and temporal variations. An example of how the auroral structure can taint observation comes from an energy spectrum of precipitating auroral electrons measured on the Freja satellite at a 1,650-km altitude over the aurora. The electron detector worked by sweeping through the energy range from 25 keV to roughly 20 eV in 32 steps within 31.25 ms, alternating with a 31.25-ms observation period at a fixed energy and with the same 1-ms sample rate. Assuming a spatially homogenous precipitation over the distance traveled in 31.25 ms (with a satellite speed of 7 km/s, this would be roughly 200 m), scientists can obtain the precipitation's energy spectrum during one of these sweeps. However, during one of the sampling periods at constant energy, researchers detected a five-fold increase of the electron flux within 2 ms, which means that within the space of 10 m the auroral electron flux increased to its full strength.

This example not only demonstrates that an energy spectrum that takes 32 ms to measure might not be representative of actual auroral particles, but it also raises theoretical difficulties in explaining how such a structure can exist in the plasma at a 1,650-km altitude in the first place. A 100-eV electron’s gyroradius (the size of the circular motion around a magnetic field line) is 1 m, and its skin depth (the scale length of electromagnetic waves interacting with the plasma) is 100 m. It’s difficult to maintain any propagating structure shorter than the skin depth, so the only plausible explanation for the anomalous structure’s presence is that the Freja spacecraft happened to be in the exact region in which the electron acceleration occurred and where the steep gradient was generated.

**Auroral Electron Acceleration**

This and other observations (one was from two conjugate spacecraft at different altitudes) show that auroral electron acceleration occurs not far above the upper atmosphere. This is also the most likely location in which the aurora’s small-scale structure is generated. The energy source for electron acceleration, however, is located much further away—in the magnetosphere’s tail—which in turn receives energy through its interaction with the solar wind. Several mecha-
Figure 5. Model simulation. This horizontal slice at a 400-km altitude is through the 3D three-fluid simulation of a current sheet, 4.5 seconds into the simulation. The originally broad and unstructured current sheet has evolved into structures that are similar to observed aurora. The color scale on the left shows the field-aligned current, and the arrows show the horizontal plasma flow. The right panel shows the quasi-potential derived from the simulation’s electric field. The horizontal scale is in km.

nisms can transport this energy to the auroral acceleration region: field-aligned currents (electrons and ions flowing in opposing directions along the magnetic field) can connect the tail with the inner magnetosphere, high-energy plasma beams (electrons and ions flowing in the same direction), or large-amplitude Alfvén waves (the plasma moves perpendicular to the magnetic field and transmits energy along the magnetic field). Depending on the energy transport mechanism in the auroral acceleration region, different conditions are set up that eventually produce the structure and evolution of auroral arcs. Field-aligned currents can become unstable to generate local parallel electric fields; plasma flow can lead to double layers (structures with an embedded parallel electric field); and kinetic (or inertial) Alfvén waves carry a quasi-static electric field to accelerate electrons.

**Acceleration by Waves**

Measurements of the Poynting flux in kinetic Alfvén waves at very high altitude have occasionally shown that these waves carry as much energy as the aurora typically dissipates. The large-scale waves generate small-scale waves with very large wave numbers in a perpendicular direction, which means that although these waves extend thousands of kilometers, they’re highly structured perpendicular to the magnetic field (the structure that we expect from the aurora’s observed form and shape). Particle plasma simulations in two dimensions of space and three dimensions in velocity show the development of structure from large-scale kinetic Alfvén waves. These simulations show strong up- and downward electric fields at altitudes of a few thousand km above the atmosphere with a horizontal north-south structure of the same scale size as observed in aurora. Long-term observational studies of satellite data have also shown that the kinetic Alfvén wave auroral acceleration is likely to account for roughly one-third of aurora during geomagnetic active periods.

When currents flow along the magnetic field’s direction, they encounter almost no resistance—that is, the field-aligned conductivity is nearly infinite in the magnetosphere. Collisions in the upper atmosphere, however, do impose a resistance, so a field-aligned electric field is needed to maintain the current. In ideal plasma, a field-aligned current sheet means that the magnetic field has a shear across it—turning on a current thus implies that an initially parallel magnetic field is sheared. As the current propagates along the magnetic field, this shear propagates with the current’s head in what is called a shear Alfvén wave. As this wave approaches the upper atmosphere, the decreasing conductance causes the wave to be reflected, and a small amount of wave energy is dissipated. The ionospheric plasma is then set in motion, horizontal and field-aligned currents in the ionosphere are induced, and electric fields are generated. To understand the coupling between the ionosphere and the magnetosphere, we can model this transition of a current (or an Alfvén wave) propagating from the magnetosphere with infinite conductivity into the ionosphere with smaller conductivity and the reflection back into the magnetosphere in a fluid simulation.

**Acceleration by Instability**

Although the processes that cause auroral emissions, the auroral electrons’ energy loss, and the production of ions and secondary electrons through auroral ionization are captured in a fluid description only by parameterizations,
many important aspects—especially the currents’ structure and dynamics, as well as the electric and magnetic fields—evolve self-consistently and can be studied with a fluid assumption. To describe the interaction and coupling of the magnetosphere and ionosphere, we need a three-fluid simulation (electrons, ions, and neutrals). We can do it in two or three dimensions, but some processes are inherently 3D and require a 3D simulation.

Antonius Otto and colleagues developed such a simulation model to produce an aurora with realistic shape and structure. The model is driven by an unstructured current sheet at the upper boundary. The magnetic field’s shear propagates into the simulation domain, setting the plasma in motion and distorting the current sheet. At some altitude, the resistivity becomes large enough that a parallel electric field develops (with realistic model parameters, this altitude is compatible with the observed auroral acceleration altitude). This not only provides the downward acceleration of auroral electrons but also allows a decoupling of the dynamics of the magnetic field and the plasma. This process is called a tearing instability (or component reconnection) because it seems to tear the magnetic field apart and recon- figure the flux tubes. The electrons that would accelerate in this region are still bound to follow the magnetic field direction downward, but the current sheet becomes very thin and develops folds. The energy for this process is taken from the magnetic field’s reconfiguration; the original sheared field contains stress, which the reconfiguration reduces.

Figure 5 shows a slice through this model at a 400-km altitude, which is the ionosphere’s upper region. An unstructured current sheet as a boundary condition at a 10,000-km altitude provides the energy source and magnetic shear. The left panel shows the distorted current sheet as a color scale, with blue indicating upward current (downward-moving electrons) and red indicating the downward return current. The arrows show the plasma flow’s horizontal components. Across the current sheet, the horizontal flow exhibits strong shear. We can integrate the calculated electric field along the magnetic field lines to produce a quasi-potential drop that would accelerate electrons (the right panel). A very thin structure develops within 4.5 seconds of the simulation and must be interpreted as the area of accelerated electrons and the aurora. The dynamics, shape, and energy (6.4 keV) is consistent with observed aurora. To illustrate this better, we can use the calculated 3D structure to generate a picture of the aurora by selecting a point of view on the ground below this simulation domain and integrating along all lines of sight. Figure 6 shows the result—the green color is artificially added, but the structure and brightness are derived from the value of the calculated quasi-potential.

### Aurora in the Ionosphere

Auroral electron acceleration occurs in the collisionless plasma above the thermosphere and ionosphere, but the collision-dominated atmosphere’s proximity is an important ingredient for the acceleration mechanism. Modeling the auroral ionosphere must therefore be combined with modeling the magnetosphere.

### Challenges

Some of the challenges in auroral modeling come from the large differences in the densities in the upper atmosphere and magnetosphere, the energy scale, and the complexity of the upper atmosphere’s chemistry. We can describe the transport of energetic electrons into the atmosphere with a Boltzmann equation that equates
the change in the electron-distribution function in a given phase-space volume to the changes in moving to a different altitude, changing the direction in elastic scattering, changing the energy in inelastic scattering, and the production of secondary electrons in ionizing collisions. Proton aurora has two different types of energetic particles—the protons themselves and hydrogen atoms—that switch back and forth in charge exchange and ionization collisions. The hydrogen isn’t bound by the magnetic field, so we have two coupled transport equations to solve in at least two dimensions. We can treat auroral electrons in a 1D model, and because transport through the thermosphere’s entire altitude range is usually fast compared to changes in the input of auroral electrons at the top of the atmosphere, we can treat the problem in most cases as stationary. Time dependence is important for flickering aurora, which can vary with a frequency of 10 Hz or more. For the stationary case, we’re left with an integral equation that’s formally similar to a radiative transfer equation and can be solved with Monte Carlo simulations, finite differencing, or eigenfunction expansion.

Numerical difficulties come from the required resolution in energy, in which electrons with several tens of keV energy suffer energy losses (the energy to excite, dissociate, or ionize and the energy imparted on the secondary electron in ionization collisions) of 2 to 100 eV in a single collision and from the altitude range, which varies by several orders of magnitude when expressed in terms of scattering depth. Unfortunately, the production of additional secondary electrons hampers Monte Carlo simulations. Every incident energetic electron suffers, on average, an energy loss of 35 eV for each ion produced, which means we have 100 to 1,000 times as many secondary electrons as primary (incident) auroral electrons. Although this is a problem for Monte Carlo simulations, it offers a very simple method of approximating the transport of electrons into the upper atmosphere. Because of this almost constant average energy loss per ionization collision, we can scale the penetration of an electron beam into a laboratory gas of constant pressure to the real atmosphere by counting the number of molecules encountered and adding up the energy loss.

**Chemistry in the Upper Atmosphere**

As the electrons penetrate the thermosphere, they deposit energy, excite neutrals, dissociate molecules, and create ions. This provides the energy for a neutral and ion chemistry process involving a large number of constituents. The brightest auroral emission in the visible wavelength band—the atomic oxygen green line—is the result of the chemical interactions; direct excitation plays a minor role. The dominant neutral at auroral altitude is molecular nitrogen (just like in the air at ground level), and most of the excitation, dissociation, and ionization starts with electron impact on N₂. Yet, the dominant ions at auroral altitude are NO⁺ and O₂⁺, which are the end product of a chemical reaction scheme. Fred Rees lists 80 chemical reactions in his textbook on the upper atmosphere involving 17 constituents that play a role in the auroral ionosphere. Some of these reactions depend on the temperature of neutrals, ions, or electrons, which in turn is coupled to the energy input from auroral electrons, the energy released in the chemical reactions, and additional heating from Joule and Ohmic heating by currents. Some constituents are long-lived (minutes to tens of hours), and we must take into account transport and diffusion.

Space weather and aurora forecasts depend mostly on solar observations and models that propagate solar disturbances though the solar wind. This is true not just for aurora on Earth but also for space weather on other planets and the interplanetary medium. The HAFv.2 model has successfully identified solar disturbances from the October 2003 storm as far out as the heliopause by comparing measurements made by Voyager at 100 astronomical units (1 AU is the distance of Earth from the sun) with shock-propagation predictions.

As of today, no model exists that combines all these different scales and processes: the solar surface, the solar wind, the magnetosphere, and the chemistry of the ionosphere. Usually, we extract from one model an essential feature in the form of a parameterization and include this into another model.

For example, from a model that combines auroral electron transport and ionospheric chemistry and energetics, we can obtain auroral emissions and ionospheric conductivity. We derive parameterizations by relating the relative brightness of high- and low-altitude emissions to the mean energy of the incident electrons or the ionospheric conductivity that would result from auroral electrons with that mean energy. We include these parameterizations of the ionospheric response to aurora in global models of the magnetosphere to globally model space-weather events.
The scales involved in these models vary over many orders of magnitude in time and space—for example, the solar wind has densities on the order of 1 particle/cm$^3$, the ionosphere has plasma densities of $10^3$ cm$^{-3}$ embedded in a neutral gas with densities of $10^{13}$ cm$^{-3}$, and the magnetospheric densities are in between. Magnetic reconnection depends on physical processes that occur on scales smaller than a gyroradius but affects a region that is 100,000 times as large. Auroral structures have scales of 100 m but are generated by a magnetosphere that is 10$^8$ m. Some processes require physics that can be captured only through particle interaction (for example, reconnection), but they have large-scale consequences accessible only to fluid models. Thus, the requirements for future model development go beyond multiscale modeling—a hybrid approach of particle and fluid simulations is also necessary. Several existing models treat the auroral processes separately, but the coupling of these processes is still an active research area: the requirements for spatial and temporal resolution make holistic models challenging.

Acknowledgments

Support for this work came from NASA Living with a Star (LWS) grant number NNG05GM42G.

References


Dirk Lummerzheim is a research professor in the Geophysical Institute at the University of Alaska Fairbanks. His technical interests include auroral electron and proton transport modeling, local and global ionospheric modeling, and interpretation of auroral observations from space- and ground-based instrumentation. Lummerzheim has a PhD in space physics from the University of Alaska. Contact him at lumm@gi.alaska.edu; http://odin.gi.alaska.edu.