Physical, Numerical, and Computational Challenges in Modeling the Geospace Environment

J. Raeder, J. Dorelli, D. Larson, and B. Loring

Space Science Center, University of New Hampshire, 39 College Road, Durham, NH 03824

Abstract. We present a high level overview of Earth’s immediate space environment, called geospace, and contemporary efforts to model it. As with the modeling of many complex, multi-scale systems one faces the challenges of limited physical understanding of the system and its processes, less-than-ideal numerical procedures, and computer limitations. In spite of these difficulties, advanced modeling of geospace is important, because geospace is the only large-scale plasma physical system that is accessible to detailed in situ measurements, and because its variability, “Space Weather”, has a substantial impact on human activities in space and on the ground. As an example of state-of-the-art geospace modeling we discuss the global modeling of magnetotail turbulence.

1. Introduction

Geospace is the immediate environment of Earth that is dominated by the interaction of the Earth with the solar wind (SW) and the interplanetary magnetic field (IMF). Earth’s magnetosphere is the outermost part of geospace and forms as the result of the interaction of Earth’s intrinsic magnetic field with the oncoming solar wind and IMF (Kivelson and Russell 1995; Lyon 2000). Because the solar wind is supermagnetosonic a bow shock forms in front of the Earth with a typical standoff distance of \(14 R_E\) (geospace distance is usually measured in “Earth radii”, \(1 R_E = 6372 \text{ km}\)) from the center of the Earth. The bow shock is moderately strong, with a typical shock Mach number in the range 2-10.

Downstream of the bow shock the Earth’s intrinsic magnetic field is compressed on the day side to a “bubble” of the size of about 10 \(R_E\). In the night side the field stretches out to form the geomagnetic tail. The diameter of the tail is \(\sim 60 R_E\) and its length is virtually unlimited. The tail has been observed up to \(3,500 R_E\) from Earth (Walker et al. 1975).

On the earthward side of the magnetosphere lies the ionosphere (Kelley 1989; Schunk and Nagy 2000). The ionosphere forms mainly from the ionization of the upper atmosphere due to a variety of processes, such as solar electromagnetic radiation (UV, EUV, x-rays) and particle precipitation from the magnetosphere. The ionosphere is organized into layers, ranging in altitude from \(\sim 80 \text{ km}\) (D and E region) to several 100 km (F region). The ionosphere overlaps spatially with the upper layers of the neutral atmosphere, specifically the thermosphere (up to \(\sim 130 \text{ km}\), still collisional) and the exosphere (\(\sim 130 \text{ km}\) and up, collisionless.) The ionosphere is strongly coupled with the magnetosphere via the exchange of mass, momentum, and energy, and the flow of electric current between these two regions. Ionospheric plasma flows out into the magnetosphere. At low- and mid-latitudes the ionospheric plasma fills closed flux tubes, which
then form the plasmasphere. At high latitudes the plasma outflow occurs on open field lines and the plasma either is re-captured in the tail or leaves through the tail. On the other hand, magnetospheric particles precipitate into the ionosphere, at times (depending on energy and flux) visible as auroral emissions. There is also a strong electromagnetic coupling via currents and electric fields between the magnetosphere and the ionosphere.

Figure 1. Schematic view of geospace showing the main regions of the magnetosphere. (courtesy of C. T. Russell, UCLA).

Figure 1 shows schematically the regions and some of the coupling processes. Of fundamental importance is the surface that separates the solar wind and IMF from the Earth’s intrinsic magnetic field and plasma, which is called the magnetopause. To a first approximation the magnetopause is a MHD discontinuity. If the shear angle between the Earth’s field and the IMF is small the magnetopause is generally a tangential discontinuity. However, if the shear angle is large magnetic reconnection occurs. It is still a matter of debate as to how large the shear angle must be for reconnection to occur; the two principal possibilities are “anti-parallel reconnection” which only occurs for shear angles close to 180° and “component reconnection” which can occur at 90° or even smaller shear angles (Park et al. 2006; Birn and Priest 2006). In any case, in the presence of reconnection the magnetopause becomes a rotational discontinuity that allows for the entry of mass, momentum, and energy from the solar wind and IMF into the magnetosphere. If there were no reconnection the magnetosphere would be fairly calm (only wave energy would penetrate the magnetopause), whereas reconnection allows for substantial entry of energy. While it was not clear in the 80’s whether or not reconnection plays an important role for magnetospheric dynamics, it is now well established that reconnection occurs under virtually all IMF conditions (Frey et al. 2003). During southward IMF periods
reconnection at the magnetopause primarily occurs near the equator, while during northward IMF periods reconnection occurs tailward of the cusps, where the IMF and Earth’s field form a large shear region. Such a reconnection geometry was originally proposed by Dungey (1961). However, there is now evidence that this picture is not entirely correct and that component reconnection may occur near the subsolar magnetopause concurrently with cusp reconnection (Fuselier et al. 2000; Dorelli et al. 2006).

Magnetic reconnection between the IMF and Earth’s field transports momentum into the magnetosphere. During southward IMF, Earth field lines that had been closed, i.e., anchored in both hemispheres, become open, i.e., with one end in the ionosphere and the other end in the solar wind. Since the solar wind end is frozen to the plasma field lines are dragged across the high latitude ionosphere from the day side to the night side, setting up a convection pattern across the ionosphere. Eventually the open field lines close again by reconnection in the tail plasma sheet and convect back into the day side. Thus a convection pattern is set up which comprises two convection cells in the high latitude ionosphere. A similar, but reverse pattern, exists for northward IMF, where often 4 ionospheric convection cells are observed, such that the high-latitude cells have a reversed sense and the low latitude cells have the same sense as during southward IMF. Convection is essential for magnetospheric dynamics. Often convection is unsteady because tail reconnection is slower than day side reconnection. In that case magnetic energy gets stored in the tail lobes. That energy is eventually released in an explosive fashion, which is called a “magnetospheric substorm” (Baker et al. 1997). Such substorms have also significant ionospheric manifestations, such as bright aurorae and strong ionospheric currents, called electrojets. Periods of very strong and prolonged southward IMF cause magnetic storms. During storms convection is more than an order of magnitude stronger compared to quiet times, and plasma is driven in large quantities into the inner magnetosphere where it enhances the ring current. The ring current can be measured from low latitude ground magnetometers, and an index formed from such measurements (the “Disturbance Storm Time”, or “Dst” index) and commonly used to characterize magnetic storms (McPherron 1991).

Unlike the plasma in astrophysical or solar settings the geospace plasma is accessible to in situ measurements, and thus an important place to study plasma-physical processes without the constraints of laboratory experiments, such as small scale and complicated boundary conditions. While such measurements (by at least 50 space probes since the beginning of the space age) have advanced our knowledge tremendously, the measurements are still limited, mostly by the sparse spatial coverage. Simulations thus play a vital role to put these measurements into context and to test our understanding of the processes. On the other hand, the data put a powerful constraint on the models.

The investigation of geospace has also significant societal implications. Over the past 50 years many man-made assets have been put into space and humans have ventured into geospace (even a bit beyond) themselves. Such presence in space requires that the medium be well understood to avoid unnecessary risks. Indeed, geospace is as variable as terrestrial weather, and thus the term ”space weather” arose (Wright 1995). Like atmospheric science, space science now has an application branch, that fulfills a societal role. And like in terrestrial weather
forecasting, numerical models are a critical part of any forecasting system. In the following we shall thus explore the challenges that modeling of geospace faces at various levels.

2. Physical Challenges

The most daunting physical challenge is the fact that most of the plasma in geospace is collisionless. This implies that, at least in principle, a kinetic treatment is required. A fully kinetic approach, without violating parameters too badly, is not computationally feasible. A hybrid kinetic approach, where the ions are treated as particles but the electrons are treated as a fluid (Winske and Omidi 1996) has now reached a development stage where at least small bodies (moon, mercury) can be modeled globally; however, global hybrid simulations of the Earth with realistic parameters are probably still a few years away.

The most basic fluid treatment that allows the construction of global models is magnetohydrodynamics (MHD). Careful consideration of the conditions under which MHD is derived shows that many of these conditions are violated in geospace. However, the MHD equations are the basic conservation equations for mass, momentum, energy, and magnetic flux, and thus describe at least the convective transport correctly. One of the primary simplifications of MHD is the truncation of the moment equations at the second moment (pressure). However, if the distributions of the particles are Maxwellian the higher moments vanish. Fortunately, in many regions of geospace the distribution functions stay close to Maxwellian because of waves and turbulence and thus the MHD closure of the moment equations is not a bad approximation.

Next in importance are diffusion terms and the truncation of Ohm’s law, which is essentially the electron momentum equation (Strangeway and Raeder 2001). Although the geospace plasmas are mostly collisionless, in critical regions diffusion must be present, for example in the bow shock or in boundary layers. Similarly, the “ideal” MHD approximation, namely that the electric field is entirely the motional electric field (and thus magnetic field lines are frozen to the plasma) is certainly violated in current sheets where magnetic reconnection takes place. The detailed physical processes that cause such deviations from ideal MHD are currently not well understood and the topic of intense research (Birn et al. 2001; Birn and Priest 2006). For the purpose of global geospace modeling these “anomalous” processes are either parameterized in an ad-hoc fashion or even produced by numerical effects (Raeder 2003).

Some key processes in geospace are impossible to describe with a fluid model and must be treated with a separate analytic or empirical model. For example, the field-aligned currents (FACs) that flow between the magnetosphere and the ionosphere may create a condition where the thermal plasma has not enough charge carriers to support the current which is imposed by the magnetosphere. This situation arises particularly on field lines with upward FAC where the downward going electrons that carry the current would mirror in the converging magnetic field above the ionosphere. In such a situation a parallel electric field develops that forces just enough electrons through the converging field to satisfy the current requirement (Knight 1972). These electrons are then accelerated up to tens of keV and cause discrete auroral arcs. Such a process cannot
be treated self-consistently with MHD and must thus be parameterized using analytic models (e.g., Raeder 2003).

Some plasma populations of the inner magnetosphere are very energetic, with temperatures of tens to 100s keV (ring current plasma), or in the MeV range (radiation belts). These populations coexist with colder plasma, but their transport is not well described by MHD. Traditionally these plasma populations have been modeled with different approaches that take the gradient and curvature drifts into account while neglecting inertia (Wolf 1983; Fok et al. 1993; Jordanova et al. 1996; Chen et al. 1997). Such models also limit the simulation region to the inner magnetosphere, apply proper plasma boundary conditions, and prescribe the electric and magnetic fields from simplified models. Although the energy density of the radiation belts is too small to affect the magnetosphere dynamics, the ring current couples strongly with the magnetosphere and the ionosphere. The challenge is now to embed drift models into the MHD based global geospace models. Several groups are pursuing this and the main challenge is to find the proper coupling between these models, which for the most part occupy the same regions.

![Figure 2. Cut through a numerical grid used by the OpenGGCM. In reality, the grid is larger and also has a higher resolution as shown here. The cutout contains the Earth, and in this region the MHD equations are not solved. Instead, the magnetosphere MHD model couples to the CTIM ionosphere-thermosphere model via FACs and the electric potential. See Raeder (2003) for details.](image)

3. Numerical Challenges

Geospace encompasses a large volume, yet also contains many critical features, such as discontinuities and boundary layers, that have spatial scales many orders of magnitude smaller than the system size. The numerical grids are either adapted to the small scale features in the system, or a brute force approach is used with as high numerical resolution as possible while pushing the limits of available computational power. For a typical system size of $300 \times 80 \times 80 \, R_E^3$ at a resolution of $0.1 \, R_E$ (comparable to the thickness of current sheets) one
would need \(\sim 2 \times 10^9\) grid cells, which is at least an order of magnitude beyond current computer capabilities. A stretched cartesian grid, such as the one depicted in Figure 2 and used in the OpenGGCM (http://openggcm.sr.unh.edu, http://ccmc.gsfc.nasa.gov/) can still achieve \(0.1 R_E^3\) resolution in the critical regions while requiring a factor of \(10^2\) - \(10^3\) fewer cells. Adaptive mesh refinement (AMR) promises to resolve boundary layers with even less cells (DeZeeuw et al. 2000) but also incurs a substantial computational overhead.

Besides sufficient resolution geospace models also need robustness to handle the large range of plasma parameters that is encountered. For example, the plasma density, the Alfvén velocity, and the plasma \(\beta\) vary over many orders of magnitude. This also implies a large variation of the CFL (Courant-Fridrichs-Levy, see Sod (1985)) stability limit, which needs to be overcome, for example, by using the Boris correction (Boris 1970; Brecht 1985) of including the displacement current in the equations and artificially limiting the light speed in the simulation.

A principal problem of many simulations is the introduction of numerical errors due to the limited temporal and spatial resolution. Instead of solving the equations of the system the model produces the solution of a modified set of equations that includes the error terms, in particular numerical dispersion and numerical diffusion. Numerical dispersion leads to oscillations at discontinuities, and if no measures are taken to counteract its effects, it leads to a rapid breakdown of the solution. Numerical diffusion, on the other hand, tends to smear out the solutions. Modern numerical schemes try to achieve monotonicity by blending a diffusive, low order scheme with a more accurate and less diffusive, but dispersive scheme, such that just enough diffusion is introduced at discontinuities to avoid a non-monotone solution without degrading the high-order solution in regions of smooth field and flow. There are numerous schemes of this kind, for example Harten’s hybrid method (Harten and Zwas 1972), Flux Corrected Transport (FCT) (Boris and Book 1973; Zalesak 1979, 1981; DeVore 1991), Total Variance Diminishing (TVD) schemes (Harten 1984; Sweby 1984; Yee 1985, 1987), Essentially Non-oscillatory (ENO) schemes (Liu et al. 1994; Jiang and Wu 1999), and the Van Leer flux limited schemes (Van Leer 1973, 1974, 1977). The OpenGGCM uses a variant of Harten’s hybrid scheme for the gas dynamic variables and a van Leer flux limited scheme for the magnetic field integration. It should be noted, however, that all of these schemes are roughly equivalent, as long as they are of second order accuracy in time and space and consistent with the MHD equations.

A very specific problem of all electrodynamic codes, including MHD codes, is the conservation of magnetic flux. Faraday’s law has the property that it conserves \(\nabla \cdot \mathbf{B}\), which can be seen from:

\[
\nabla \cdot \frac{\partial \mathbf{B}}{\partial t} = \frac{\partial (\nabla \cdot \mathbf{B})}{\partial t} = -\nabla \cdot \nabla \times \mathbf{E} = 0
\]

(1)

Most numerical schemes do not \textit{a priori} preserve \(\nabla \cdot \mathbf{B}\). For such schemes the accumulation of \(\nabla \cdot \mathbf{B}\) can lead to serious errors, in particular spurious parallel acceleration, wrong magnetic topology (field lines that are not closed), and significant errors in the shock jumps (Brackbill and Barnes 1980; Toth 2000). To address that problem one can “divergence clean” the field by using a projection method (Toth 2000), or as in Powell et al. (1999) modify the equations. The
latter approach, however, violates the Lax convergence theorem (namely, that stability and consistency is necessary and sufficient for convergence). Because the Powell scheme is not consistent with the MHD equations convergence is not guaranteed, and indeed, Toth (2000) has shown that this scheme may lead to wrong shock solutions. Evans and Hawley (1988) have outlined a staggered grid approach for Faraday’s equation that is equivalent to the conservative differencing of gasdynamic equations and preserves $\nabla \cdot \mathbf{B}$ to roundoff error. Many global geospace models, including the OpenGGCM (Raeder 2003) and the Lyon-Fedder-Mobarry (LFM) model (Lyon et al. 2004) use this approach now.

4. Computational Challenges

Global geospace codes are extremely computationally expensive. They can thus only be run at reasonable speed on parallel machines, which makes load balancing a critical issue. For a numerical grid like the one depicted in Figure 2 a simple domain decomposition can yield very good load balancing since the processor workload and the communication load is very similar for each processor. However, that is not necessarily true for other grids, in particular AMR. Also note, that the grid shown in Figure 2 solves only part of the problem. The ionosphere-thermosphere model (CTIM, see Fuller-Rowell et al. (1996) and Raeder (2003)) that provides the earthward boundary usually runs on a separate processor. Fortunately, the time scale of the ionosphere and thermosphere is of the order of 1 min, whereas the MHD time step is typically 0.1 - 0.3 sec as dictated by the CFL criterion. Thus, the ionosphere-thermosphere model has not much impact on load balancing. However, when other models are coupled to the MHD part, for example ring current and radiation belt models, it may become much more difficult to achieve efficient load balancing.

Even without model coupling data input and output (I/O) becomes increasingly difficult as the size of the model increases with increasing CPU capabilities. First, the speed of hard disk drives has increased much slower that the CPU speed over the past decade or so. Second, I/O does not scale very well because several (or sometimes all) compute nodes try to write to the same file system at the same time. Increasing the number of nodes makes the problem often worse, because more CPUs try to write to the same file system. For simulation runs that require a large amount of I/O, for example for the creation of animations, the wall clock time spent on I/O operations can easily exceed the time needed for the calculations. The resulting amounts of data, often several 100 GB, are also difficult to handle. Although hard disc capacity increases have more or less kept pace with increases in CPU speed, there are no cost effective backup media available. Tape drives are expensive and notoriously unreliable, whereas optical media (DVD±R) have too little capacity to be practical. A new generation of writable optical discs (blue-ray, etc.) promises a 10-fold capacity increase which would at least temporarily be useful for backup.

Along with the difficulty of storing the data comes the difficulty to visualize the simulation output and to extract useful information from the raw simulation results. Some methods are relatively straight-forward. For example, code verification is often done by comparing simulation results with in situ satellite measurements (Frank et al. 1995; Raeder et al. 1997, 2000; Le et al. 2001; Raeder et al. 2001a; Wang et al. 2003; Wiltberger et al. 2003; Oieroset et al. 2005) which
only requires the extraction of virtual time series from the results. Sometimes a comparison with two-dimensional patterns is called for (Raeder et al. 1998; Wiltberger et al. 2003), which is also easy to accomplish. However, in order to truly understand the physical processes one often needs to create intricate time dependent three-dimensional visualizations. One such example is the formation of Flux Transfer Events (FTEs) which are small magnetic flux ropes that form sporadically at the Earth’s magnetopause by magnetic reconnection (Raeder 2006). These are truly three-dimensional structures and their formation mechanism could only be elucidated by using interactive visualization tools such as Data Explorer (http://www.opendx.org) and Visit (http://www.llnl.gov/visit), and by using three-dimensional animations.

5. Magnetotail Turbulence

While the OpenGGCM has been used for a wide variety of studies, ranging from theoretical reconnection studies to space weather effects, and by many authors (Raeder et al. 2001b,c; Raeder and Lu 2005; Le et al. 2001; Wang et al. 2003, 2004; Dorelli et al. 2004; Moretto et al. 2005; Vennerstrom et al. 2005; Rastatter et al. 2005; Li et al. 2005; Raeder 2006), we present here for the lack of space only one example of a demanding application.

The plasma sheet in the geomagnetic tail, which separates the northern lobe from the southern lobe, has long been known to be a turbulent medium (Borovsky et al. 1997; Borovsky and Funsten 2003; Weygand et al. 2005). Until fairly recently global models could not resolve the turbulent motion in the plasma sheet. More specifically, the models could not reproduce “Bursty Bulk Flows” (BBFs), which are short duration (a few min), predominantly earthward flow bursts with velocity of 400 km/s or more (Baumjohann et al. 1990; Angelopoulos et al. 1992, 1994). BBFs are considered both a part of the turbulence itself and a driver of smaller scale turbulence through cascading (Borovsky and Funsten 2003). It was originally not possible to determine the lateral scale size of BBFs since only single spacecraft observations were available. Recently, Nakamura et al. (2004, 2005) have shown, using the 4 spacecraft Cluster constellation, that the lateral dimensions are of the order of 2 - 4 $R_E$.

Figure 3 shows a cut in the equatorial plane from a recent high-resolution OpenGGCM simulation. In this simulation the IMF is southward while all other solar wind quantities are at average values. The flows in the tail plasma sheet are by no means steady as depicted in many cartoons (McPherron 1991; Kivelson and Russell 1995) but are confined to narrow flow channels. These flow channels are either tailward or earthward flow. The flows also emanate from small x-line segments. The characteristics of these flows are thus very close to the observed ones.

An important aspect of this simulation is the lack of any specific trigger for these reconnection events. Apparently the magnetotail is in a meta-stable state where each of the reconnection events relieves some of the stress in a certain region, and thereby increasing the stress in an adjacent region, which subsequently becomes unstable. This mechanism appears to be similar to some proposed earthquake mechanisms where an earthquake relieves stress on one part of a fault line while increasing stress on a different part. Further details of
this process, such as dependencies on numerical resolution, resistivity models, and SW/IMF parameters still need to be worked out and will be subject of subsequent publications.

Acknowledgments. This work was supported by the National Aeronautics and Space Administration (NASA) Living With a Star (LWS) program under grant NNG04GP7 and by the National Science Foundation under grant ATM-0427754. Computations were performed at the San Diego Supercomputer Center (SDSC), the National Center for Supercomputer Applications (NCSA), and on the UNH “Zaphod” Beowulf Cluster, which was in part funded by the Major Research Instrumentation (MRI) program of the National Science Foundation under grant ATM-0420905.

References


