Toward Understanding Radiation Belt Dynamics, Nuclear Explosion-Produced Artificial Belts, and Active Radiation Belt Remediation: Producing a Radiation Belt Data Assimilation Model

Geoffrey D. Reeves, Reiner H. W. Friedel, Sebastien Bourdarie, Michelle F. Thomsen, Sorin Zaharia, Michael G. Henderson, Yue Chen, Vania K. Jordanova, Brian J. Albright, and Dan Winske

The space radiation environment presents serious challenges to spacecraft design and operations: adding costs or compromising capability. Our understanding of radiation belt dynamics has changed dramatically as a result of new observations. Relativistic electron fluxes change rapidly, on time scales less than a day, in response to geomagnetic activity. However, the magnitude, and even the sign, of the change appears uncorrelated with common geomagnetic indices. Additionally, observations of peaks in radial phase space density are not readily explained by diffusion processes. These observations lead to a complex picture of acceleration and loss processes all acting on top of adiabatic changes in the storm-time magnetic field. Of even greater practical concern for national security applications is the threat posed by artificial radiation belts produced by high altitude nuclear explosions (HANE). The HANE-produced environment, like the natural environment, is subject to global transport, acceleration, and losses. Radiation belt remediation programs aim to exploit our knowledge of natural loss processes to artificially enhance the removal of particles from the radiation belts. The need to open up new orbits and new capabilities has raised questions about the space environment that, up to this time, we have been unable to fully answer. Here we describe the development of a next-generation model for specifying natural and HANE-produced radiation belts using data-assimilation based modeling. We exploit the convergence of inexpensive high-performance parallel computing, new physical understanding, and an unprecedented set of satellite measurements to improve national capability to model, predict, and control the space environment.

NATURAL VARIABILITY OF THE EARTH’S ELECTRON RADIATION BELTS

The discovery of the Earth’s radiation belts was one of the first of the space age. Since that time many measurements of the radiation belts have been made and, as recently as ten years ago, the radiation belts and the processes affecting them were considered to be relatively well-understood. Text books still teach that radiation belt dynamics are primarily controlled by radial and pitch angle diffusion as described
by Schulz and Lanzerotti, [1974]. However, observations from a variety of satellite programs (such as CRRES, geosynchronous, GPS, HEO, SAMPEX, POLAR, Akebono, and others) have revealed fundamental holes in our understanding of radiation belt structure and dynamics (figure 1). The shift of radiation belt physics from a sleepy backwater of space physics to a cutting edge research topic [Friedel et al., 2002; Kintner et al., 2002] and a national science priority [NRC Space Studies Board, 2002; National Security Space Architect, 1997] can be traced to the March 1991 CRRES satellite observation that an entirely new belt of >13 MeV electrons was produced in a matter of minutes through the interaction of an interplanetary shock with the Earth’s magnetosphere [Blake et al., 1992].

To date most studies have focused on the radiation belt electron flux increases seen at geosynchronous orbit. Those studies showed that the peak fluxes are typically observed one to three days after the storm main phase, in the middle of the ring current recovery phase [Baker et al., 1990]. The delayed response was originally explained by the “recirculation” model of Fujimoto and Nishida [1990]. More recently, multi-spacecraft observations revealed that this delay is primarily a characteristic of the outer edges of the radiation belts near geosynchronous orbit, while in the heart of the radiation belts the enhancement can occur in a matter of hours, too fast for classical radial diffusion or recirculation [Reeves et al., 1998; Li et al., 1999]. New theories are being developed that account for enhanced diffusion through, for example, enhanced ULF drift resonance [e.g. Elkington et al., 1999; 2003] but other observations are even more of a challenge to the “diffusion-only” scenario. Green and Kivelson [2004] have published observations of peaks in the radial profile of phase space density that provide strong evidence that local stochastic acceleration and/or radially-localized pitch angle scattering from wave particle interactions may dominate over diffusive processes. Those proposed stochastic processes have led to new theoretical studies of relativistic wave-particle interactions [e.g. Horne and Thorne, 2003]. Another challenge for theory and models came from the discovery that enhanced geomagnetic activity could produce either large increases or large decreases in relativistic electron flux suggesting a delicate balance between enhanced electron acceleration and enhanced storm-time losses [Reeves et al. 2003]. Summers and Ma [2000] proposed a framework which included the combined processes of radial diffusion and interactions between electrons and both whistler and EMIC waves. The combined effect of the different interactions could produce both enhanced energization and enhanced precipitation operating simultaneously.

All of these proposed processes (and others not discussed here—See Friedel et al., [2002] for a review) are still somewhat speculative or poorly quantified because the observational evidence has not been combined with global physics-based models in a way that can definitively prove or rule out competing scenarios. One reason for this is that understanding acceleration, transport, and losses requires simultaneous multi-point measurements of phase space densities at fixed values of the three invariants of the particle motion, which in turn requires knowledge of the global, storm-time magnetic field—knowledge that can only come from global models. In order to enable a future space weather capability these models will need to be time-dependent and data-driven. But, they must also apply over sufficiently long time scales to enable reliable and cost-effective spacecraft design.

**MAN-MADE BELTS FROM HIGH ALTITUDE NUCLEAR EXPLOSIONS (HANE)**

Like the March 1991 interplanetary shock event, high altitude nuclear explosions (HANE) are known to produce sudden, intense, and long-lived radiation belts (figure 2). Therefore the acceleration, transport, and loss processes that apply to the natural radiation environment are equally important for the HANE environment. However, understanding and mitigating the threat to space-based systems from man-made belts presents additional challenges to modeling. One of the biggest challenges to modeling is that we must apply those models to conditions that have never been observed or have been poorly observed. This requires a high degree of confidence in the physical understanding of key processes that are encoded in the models’ algorithms, a high degree of validation based on the variability of the natural environment, and event-specific scenarios that incorporate the full set of space observations.

The impact of a single high altitude nuclear explosion can be severe. It is estimated that the STARFISH explosion (1.4 Mt at 400 km altitude) set off in July 1962 produced about 10\(^{26}\) fission electrons with MeV energies [Brown et al., 1963; Van Allen et al., 1963]. No measurable high energy protons were produced. The belt was relatively narrow, being centered at L = 1.2 with a peak flux of \(10^9\) electrons/cm\(^2\)-s, with the electron density reduced to 10% at L = 1.8 and 1% at L = 2.2 [Hess, 1963]. However, some fission electrons were detected as far out as L \(\sim\) 5 – 6, implying some outward radial transport. (Similar cases have been proposed for the natural environment by Reeves et al. [1998] and Green and Kivelson [2004].) Most of the initial electrons were in the range 1–4 MeV, consistent with fission (where electrons occur up to about 10 MeV), and the artificial enhancement of the radiation belts was observed to have a lifetime of years. However, at low L, the calculated lifetimes (based on pitch angle diffusion and atmospheric precipitation) were
**Figure 1.** This schematic illustration shows some of the resources available for global data-assimilation based models of the Earth’s radiation belts. Illustrated are the color-coded MeV electron fluxes as viewed from above the equatorial plane. Also illustrated are the relative positions of satellites expected to be operational during the coming years: geosynchronous orbit (in which currently 2 GOES and 6 LANL satellites are currently operational), the GPS orbits (24 satellites with 4 in each of 6 orbits), Molniya orbits (with 2 polar highly elliptically orbiting satellites), and the 2 Radiation Belt Storm Probes (RBSP) that are one component of NASA’s Living With A Star program in an equatorial elliptical orbit shown in black.

**Figure 2.** A schematic illustration of the fluxes of natural radiation belts and those produced by a hypothetical high altitude nuclear explosion (HANE) of 1 Mt over Korea. The artificial, HANE belt is expected to be relatively narrow but have peak fluxes 100–1000 times higher than the average at the peak of the outer belt. Predicting the evolution and the potential effects of an artificial belt requires the same physical understanding and modeling that is currently being applied to the natural belts. This is particularly true when one considers that the observations and models must be extrapolated to conditions that have never been observed or have been only poorly observed.
longer than observed by an order of magnitude, implying an unknown mechanism source of strong diffusion over short periods of time. Similar results have been seen for interplanetary shock-produced radiation belts by the SAMPEX satellite, but possible mechanisms for enhanced wave-particle scattering have not yet been modeled or compared with observations.

Since the enactment of the Limited Nuclear Test Ban Treaty in 1963, the likelihood that a country would detonate a weapon in space has been a matter of policy debate and dependent on the current global political conditions. Currently, several scenarios have emerged as having the greatest risk to systems that are space-based or depend on critical space-based components (e.g., communications and navigation) [Murch, 2001]. The scenario currently considered the most likely is collateral damage from regional nuclear conflict. A high altitude nuclear explosion could be used as a nuclear warning shot in an escalating regional conflict or a deliberate effort to damage adversary forces and infrastructure through a nuclear-generated electromagnetic pulse (EMP). A second scenario is detonation of a salvage-fused warhead during an attempted exoatmospheric intercept. A third scenario has been described as a ‘Space Pearl Harbor’—a deliberate effort to cause economic damage and decreased military capability through asymmetric attack. Such a strategy could be used by a rogue state facing economic strangulation or imminent military defeat and could occur over their own sovereign territory. Increasingly, risk scenarios must include terrorist actions that seek to pose large-scale economic and cultural impact with lower risk of nuclear retaliation.

While there is ample room for debate over the probability that any of these scenarios would actually occur, the potential consequences are sufficiently severe that it would be unwise not to develop modern models that could better predict the creation and evolution of man-made radiation belts. Studies by the Defense Threat Reduction Agency (DTRA) [Murch, 2001] concluded that “One low-yield (10–20 kt), high-altitude (125–300 km) nuclear explosion could disable (in weeks to months) all LEO satellites not specifically hardened to withstand radiation generated by that explosion.” Satellites at risk include communications, imaging-mapping, and manned spaceflight. Replacement cost alone is estimated at over $50 billion. Reduced lifetimes of even a few critical Department of Defense (DoD) satellite systems would have a significant detrimental effect in conducting military campaigns [Metz and Babcock, 2004].

RADIATION BELT REMEDICATION

In space as on the ground the old joke still applies: “Everybody talks about the weather but nobody does anything about it.” Increasingly though, the civilian and military space community is moving from a strategy that has been described as “cope and avoid” to one that is characterized as “predict and control”. Improved physical understanding of the natural loss processes in the radiation belts has the potential to enable systems that can exploit those processes and reduce or remEDIATE the threat from natural or man-made radiation belt electron fluxes.

Relativistic electron fluxes are depleted either by loss through the magnetopause, precipitation into the atmosphere, or possibly through de-energization—removing electrons from the system or reversing the initial acceleration processes. Comparisons between satellites such as SAMPEX and POLAR show that the fluxes in the drift loss cone (which are precipitated in one drift period or less) track the fluxes at high altitude closely [Kanekal et al., 1999]. This shows that relativistic electron precipitation occurs nearly continuously. However, in contrast to the quasi-steady “drizzle” of electrons from the radiation belts, strong geomagnetic activity during storms can produce very rapid rates of precipitation leading to permanent and dramatic reductions in the trapped electron fluxes [Onsager et al., 2002; Reeves et al., 2003]. Green et al., [2004] have reviewed the possible causes of this loss, evaluated those mechanisms against observations, and concluded the most probable mechanism for loss is electron precipitation through enhanced pitch angle scattering through wave-particle interactions [e.g. Horne and Thorne, 2003]. Low-altitude satellite measurements of electron precipitation “bands” and “microbursts” yield electron loss rates that could completely remove all relativistic electrons from the radiation belts in a matter of days. (See e.g. Nakamura et al., [1995, 2000]; Lorentzen et al., [2000, 2001]; Millan et al., [2002]; Blake et al., [1995] for further discussion.)

While some radiation belt remediation schemes are, to say the least, impractical (for example, wrapping the Earth’s equator with a solenoid to cancel out the geomagnetic dipole field) others aim to exploit our knowledge of natural processes to artificially enhance the rates of electron precipitation at specific times, energies, or altitudes to mitigate radiation hazards from natural or man-made events. One promising method of radiation belt remediation under current investigation involves enhancing the electron pitch-angle scattering rate via cyclotron-resonant wave-particle interactions. VLF radio waves can be injected either from space or from ground-based sources (such as ionospheric heaters that modulate ionospheric conductivity). Properly coupled, the wave-induced scattering will reduce the magnetic mirror altitude of trapped electrons, increase atmospheric collisions, and dramatically increase precipitation losses [e.g. Inan et al., 2003]. Other methods to increase...
the pitch-angle scattering rate being investigated include electrostatic and magnetostatic processes implemented by space based tethers or DC magnets.

In order to evaluate any of the proposed radiation belt remediation techniques, or to optimize the effectiveness of any particular technique, one must be able to accurately and quantitatively predict the effects on a global scale. To achieve this level of understanding and predictive capability requires improved understanding of natural processes, targeted active experiments in space, and the global, data-driven physical models that we describe in further detail below.

DATA ASSIMILATION FOR RADIATION BELT MODELING

To address the needs and solve the questions posed in the preceding sections requires a focused international effort with three components: (1) a targeted, multi-satellite observational campaign such as the NASA Living With A Star (LWS) Radiation Belt Storm Probes [Kintner et al., 2002] to fill in holes in our knowledge of radiation belt dynamics; (2) a strong program to develop improved theoretical descriptions of key processes such as wave-particle interactions and multi-dimensional relativistic electron diffusion; and (3) development of global, time-dependent, data-driven but physics-based models of the radiation belts. Here we address the elements that would enable the successful execution of the third component—a next-generation radiation belt model.

To be useful for the applications described above, a next-generation radiation belt model must have several features. It must have high fidelity to the known physical equations governing the particles and fields in the inner magnetosphere—not just for physical understanding, but also to be useful in extrapolating to conditions or scenarios that have not yet been observed. (However, this does not necessarily require a first-principles model, such as global MHD models, that start with conditions in the solar wind or at the solar surface.) At the same time, the model must use all available observations in order to accurately represent the dynamic changes that occur during any specific individual event. The model must also be able to accurately represent the changes in the global magnetic field during geomagnetic storms. This is particularly critical for transforming spacecraft observations from a spatial coordinate system to a magnetic coordinate system where data from multiple satellites can be properly compared and physical equations can be solved consistently. While purely empirical and purely first-principle physics models are most appropriate for some applications, the requirements discussed here lead us toward data assimilation models that use physical equations together with all relevant observations to produce a “best fit” description of the dynamics of the radiation belts.

Data assimilation techniques are ideally suited for combining the data and models in such a way that the limitations of one component are balanced by the strengths of another component. Data assimilation models have been used extensively in other fields such as meteorology and climate modeling [e.g. Ghil et al., 1997] but, except in the area of ionospheric physics models, they have not been extensively applied to space. Radiation belt dynamics are well-suited to the methods of data assimilation—more so than other problems in space plasma physics. Compared to other regions of the magnetosphere, the inner magnetosphere is relatively well-ordered by the geomagnetic field, the physical equations governing the majority of particle dynamics relatively well-known, and there is a relatively large number of satellites (tens) covering the volume of the system.

There are a number of well-established techniques for data assimilation but among the most powerful and widely-used is Kalman filtering [Kalman, 1960] (and here we use the term to included extended Kalman filtering). Kalman filtering is a technique to simultaneously incorporate data (with specified errors) and adjust physical parameters within the model using a recursive solution to the discrete-data linear filtering problem. The Kalman filter is a set of mathematical equations that provides an efficient computational (recursive) means to estimate the state of a process, in a way that minimizes the mean of the squared error. The filter is very powerful in several aspects: it supports estimations of past, present, and even future states, and it can do so even when the precise nature of the modeled system is unknown [Welsh and Bishop, 1995]. Since the technique was first described in the 1960’s the method has been the subject of extensive research and application due in large part to advances in digital computing that allow for the solution of highly-coupled systems like the storm-time inner magnetosphere.

Irrespective of the specific data assimilation technique used, there are several components that need to be combined to solve the coupled system of ring current, radiation belts, electric potentials, magnetic fields and waves in the inner magnetosphere. A realistic model of radiation belt dynamics that is valid for geomagnetic storm times must also include a self-consistent calculation of the storm-time ring current (carried by keV protons), a sophisticated description of diffusion in energy, pitch angle and L-shell (including off-diagonal matrix elements), a specification of the spatial and temporal distribution of whistler and EMIC wave fields, and a calculation of the stochastic effects of wave particle interactions. All the necessary components of such a model now exist.
To be realistic, this model must also be consistent with all the available data sources—measurements of radiation belt particles, ring current particles, wave fields, local magnetic fields, and solar wind inputs. An unprecedented set of all these measurements has now been collected and critical new measurements will be added by the LWS Radiation Belt Storm Probes (and possibly, by other proposed missions such as ORBiTAlS and COMPASS).

A comprehensive program is needed to coordinate all the aspects of theory, modeling, data validation, and application of data assimilation techniques. Such a comprehensive program is a significant, but highly valuable, endeavor. We now outline the components of such a program and the initial steps that have been taken to bring these pieces together.

### DIRECT DATA INSERTION USING THE SALAMMBO CODE

One example of the value of even very simple data assimilation is provided by a three-satellite study of one month of storms using the Salammbô code [Boscher et al., 1996; Bourdarie et al., 1996]. The Salammbô model is a diffusion model that solves the Fokker Planck equation in three dimensions: L-shell, Energy, and pitch angle. Salammbô currently uses the simplifying assumption that the magnetic field is a pure dipole and therefore not time-dependent and not azimuthally asymmetric. The model uses statistical relationships between solar wind parameters and indices of geomagnetic activity (such as Kp) to parameterize processes such as diffusion rates or wave-particle interactions where the radiation belts overlap plasmasphere.

Figure 3 shows the period of September 2–October 20, 1998: one of the intervals selected for study by the NSG Geospace Environment Modeling (GEM) working group on radiation belt dynamics. Figure 3-a shows the results of a run in which only a single geosynchronous spacecraft was used as input into the simulation. The top panel in figure 3-a shows the fluxes measured by the Aerospace Corporation’s instruments on a highly elliptical orbit (HEO) satellite. Electron flux above ≈4 MeV is plotted as a function of L-shell and time. The bottom panel shows the Dst index. There are three clear enhancements of fluxes (of diminishing intensity) in the range L ≈ 3–8 in response to three intervals of enhanced geomagnetic activity. The second panel of figure 3-a shows the output of the Salammbô model using the geosynchronous data as input and the next panel shows the ratio of the measured and model fluxes on a log scale and in the same format of L vs. time. In the plot of ratios, black represents a value of 1 or perfect agreement; red represents measured fluxes that are ten times higher than the model predicts and dark blue represents measured fluxes that are up to ten times lower.

One, known, problem with the location of plasmapause in the Salammbô model produces a pair of bright red bands near L=3. More troubling is the relative lack of dynamic changes in the Salammbô model compared to the observations. This is particularly true in regions above L=5.5. One might expect that region to be more well-specified since the input from geosynchronous is at L=6.6. However, the HEO satellite, which is the basis for the observational comparison, crosses L=6.6 at high magnetic latitudes and is therefore quite sensitive to the assumed (isotropic) pitch angle distribution at geosynchronous orbit. Therefore, even at L=6.6 the agreement between model and measurement is poor.

Figure 3-b is identical to figure 3-a except that in this run two satellites were used as input into the model LANL-GEO 1994-084 and GPS NS-33. The data were incorporated through simple ‘direct data insertion’ which involves adjusting the values at certain grid points to exactly fit the observations when the satellite was at that location and then allowing the solution to propagate from those points throughout the system as defined by the equations of motion. Once again the HEO fluxes were used as an independent comparison and not as an input to the model so the top panel of figure 3-b is the same as previously. In the second and third panels of figure 3-b we can see that the inclusion of GPS makes a significant difference in the output from the model. While the GPS data (with coverage only above L>4.2) do not change the problems at very low L-shells, the addition of those data make the three intervals of enhanced electron flux much more apparent in the model. The ratio plot shows significantly better quantitative agreement with more regions of black throughout the plot. The substantially better agreement in the vicinity of geosynchronous orbit is due to the fact that GPS crosses L=6.6 at high magnetic latitudes (similar to HEO) and the two point measurements on the same L-shell allow a direct (but limited) determination of the pitch angle distribution.

This relatively simple illustration shows the value of incorporating multiple data sets but does not yet show the full potential of data assimilation. With even as few as two satellites it is possible to allow adjustable, time-dependent diffusion rates in the model—a technique known as adaptive data assimilation. For example, Koller and Friedel, [2005] have recently demonstrated that we can use this technique to not only determine radial diffusion rates that best reproduce the observations, but actually to determine the time dependence of those diffusion rates. The result is not just a better match to the HEO data (after all we could just use the HEO observations for that) but rather an improved understanding of the time-dependence of the diffusion rates and their rela-
Figure 3 shows observations and models of a storm on September 21–October 20, 1998. Column A shows the results of a run in which only a single geosynchronous spacecraft was used as input into the simulation. The top panel in figure 3-a shows the fluxes measured by the Aerospace Corporation’s instruments on a highly elliptical orbit (HEO) satellite. Electron flux above ≈4 MeV is plotted as a function of L-shell and time. The bottom panel shows the Dst index. The second panel of figure 3-a shows the output of the Salammbó model using the geosynchronous data as input and the next panel shows the ratio of the measured and model fluxes on a log scale and in the same format of L vs. time. In the plot of ratios, black represents a value of 1 or perfect agreement; red represents measured fluxes that are ten times higher than the model predicts and dark blue represents measured fluxes that are up to ten times lower. Column B is identical to Column A except that in this run two satellites were used as input into the model LANL-GEO 1994-084 and GPS NS-33. Again the HEO fluxes were used as an independent test and not as an input to the model. In the second and third panels of figure 3-b we can see that the inclusion of GPS makes a significant difference in the output from the model.
tion to geomagnetic activity. Improved understanding provides improved confidence in our ability to extrapolate other energies, locations, and conditions.

As noted above, Salammbo uses a relatively simple dipole description of the magnetic field. Recently, though, the model has been extended to better incorporate non-dipole fields. This is done by transforming the spacecraft observations from spatial coordinates (latitude, longitude, radius) into magnetic coordinates (most importantly pitch angle and L*). All calculations and spacecraft intercomparisons are done in magnetic coordinates. To compare against an independent set of observations such as HEO, the transformation is reversed. This improvement allows much more accurate and consistent treatment of the data which is especially important as additional data from Geo, GPS, POLAR, SAMPEX, HEO, and Akebono are added.

**STORM-TIME RING CURRENT MODELING**

While diffusion is certainly one of the most important processes affecting the radiation belts, full understanding of the structure and dynamics of the radiation belts requires a more physically realistic description that self-consistently represents a variety of interacting populations and processes. Those include, time-dependent convection electric fields, the development of the storm-time plasma sheet and ring current, the interaction of the ring current with the plasmasphere to produce EMIC waves, the interaction of EMIC and storm-time whistler waves with the radiation belt electrons, the inflation of the geomagnetic field due to the ring current, and the adiabatic response of the radiation belt electrons to the changing geomagnetic field.

Several models of this type have been developed, originating with work done at the University of Michigan [e.g. Jordanova et al., 1994; Liemohn 2001; Fok, 2001]. We have been using the UNH version of the RAM code which incorporates all the above-mentioned processes and has already been used with considerable success to reproduce realistically the storm-time evolution of the near-Earth plasma sheet and ring current [e.g., Jordanova et al., 2003a,b].

The model numerically solves the bounce-averaged kinetic equation for the distribution function of charged particles in specified global electric and magnetic fields. The model treats ions (H+, O+, and He+) with kinetic energies from 15 eV to 400 keV and has recently been extended up to relativistic energies for electrons [Jordanova et al., 2005]. Like Salammbo, the present version of UNH-RAM represents the magnetic field of the Earth as a dipole. However, the code is capable of solving the equations of motion in an arbitrary magnetic field. The model has been updated to use any electric field specification including new models with high temporal and spatial resolution like the AMIE model [Richmond and Kamide, 1988], the Weimer [2001] model, or real-time data-driven descriptions that are under development.

All major loss processes of magnetospheric particles are included in the UNH-RAM model, including charge exchange, Coulomb collisions, wave-particle interactions, and loss due to atmospheric precipitation (see Jordanova et al. [1996; 1997] for more details). For example, the convective growth rates are obtained from the dispersion relation, which is coupled and solved simultaneously with drift transport in order to treat the process of wave-particle interactions self-consistently.

**STORM-TIME MAGNETIC FIELD SPECIFICATION**

As discussed above, a major challenge in understanding important radiation belt processes such as diffusion, stochastic acceleration, or particle precipitation is to accurately calculate the phase space density at fixed values of the adiabatic invariants. While the first invariant (defined by electron gyromotion) can be calculated based on local measurements of particle pitch angle distributions and magnetic field strength, the second (bounce) invariant requires an integral along a magnetic field line and the third (drift) invariant requires an integral around the entire drift orbit of the electron. Therefore, highly accurate, time-dependent, and event-specific magnetic field models are required.

The requirements on the global magnetic field model are most stringent during the main phase of geomagnetic storms when acceleration and loss processes may both be most intense but when the perturbation to the magnetic field is also the most dramatic. The diamagnetic effects of the ring current decrease the field, moving particles outward and decreasing their energy in order to conserve the third invariant [Kim and Chan, 1997]. This can result in flux changes of 2–3 orders of magnitude during a storm main phase—which is exactly the time when the other processes have their largest effects. In order to calculate radiation belt dynamics correctly we need to separate the nonadiabatic acceleration and loss processes from the adiabatic effects of the ring current (the “Dst effect”). Equally important is including the full local time asymmetries of the inflated magnetic field (the asymmetric ring current) in order to correctly map spacecraft at different local times to the same magnetic coordinate system.

Two classes of storm-time magnetic field models currently exist. One provides a statistical representation of the average storm time field [Tsyganenko et al., 2002] but is not event specific and is only parameterized by solar wind and geomagnetic conditions. The other class is represented...
by the UNH-RAM code which calculates the perturbation field produced by the ring current but solves the electron and ion motion in a dipole field. We have investigated two approaches to solving this problem which can ultimately be combined together.

Self-Consistent Magnetic Field-Ring Current Calculations

No ring current models currently calculate ring current dynamics in a self-consistent magnetic field. Particle dynamics are calculated in a dipole field and then used to determine the perturbation field produced by those particle distributions. In reality, though any perturbation to the field perturbs the trajectory of the ring current particles themselves and hence, to be realistic, the particle trajectories and global magnetic field must be calculated self-consistently.

This adds several layers of complexity. First, the codes must be generalized to be able to solve the equations of motion in a non-dipole field, which requires more numerical integration (and computation time) but is not conceptually complex. In an arbitrary field, bounce-averaging of the general gradient-curvature drift is necessary [Shukhtina, 1992]. In order to study the process of wave-particle interactions and address questions related to acceleration and loss of energetic particles by plasma waves, the full pitch angle dependence of the distribution function in the equatorial plane must be retained. Incorporating a self-consistent field specification therefore has impacts on pressure distributions, wave growth, wave particle interactions, and particle trajectories as well as the large adiabatic effects discussed above. We have recently implemented a parallel computing version of the UNH-RAM code using the Message Passing Interface standard to implement domain decomposition, making the added computations easily feasible.

The second problem is that the complex 3-dimensional magnetic field that would be in force balance with the particle (plasma) population needs to be found and can no longer be represented analytically. Additionally, in the inner magnetosphere large temperature anisotropies are common and equilibrium solutions must accommodate those anisotropies. There are several ways to solve the problem of the magnetic field/plasma equilibrium in 3 dimensions. Among the most promising, and the method we have chosen, is an iterative solution [Zaharia et al., 2004] that uses an Euler Potential specification of the field and finds the magnetic configuration in force balance with a prescribed pressure distribution (figure 4). The pressure only needs to be prescribed at one location along each field line (e.g. on the equatorial plane), as mapping along the field line provides it everywhere else.

These two changes to the numerical calculations will enable a fully self-consistent calculation of the ring current and the perturbed global, storm-time magnetic field. To do so, first we calculate the plasma pressure distributions as already specified by UNH-RAM code. We then calculate a 3-dimensional magnetic configuration in force balance with those pressure distributions, using our Euler potential equilibrium code. Once the field and the electric currents have been found, we need to replace the dipole field in UNH-RAM with the new, more realistic field. This cycle will be repeated iteratively until the solution converges. The first step has been successfully demonstrated by Zaharia et al. [2005]. The resulting field can be used to calculate the dynamics of radiation belt electrons directly since they do not carry sufficient currents to further perturb the magnetic field.

Empirical Magnetic Field Specification

Another, complementary, technique to specify the global magnetic field is to exploit the magnetic drift motion of energetic particles and Liouville’s theorem, which states that particle phase space density is conserved along a dynamic trajectory. The technique is illustrated in figure 5 [Reeves et al., 1997]. For simplicity, imagine particles with 90 degree equatorial pitch angles that drift along contours of constant equatorial magnetic field strength. Two satellites that cross the same contour of constant equatorial field strength at nearly the same time should, according to Liouville’s theorem, measure the same phase space density provided that there has been no appreciable acceleration or loss of particles as they drift from one satellite location to the other. Furthermore, this must be true for all energies and all pitch angles (which follow slightly different trajectories in a non-dipole field). Thus since all particles move in the same magnetic field, each energy and each pitch angle measured can essentially provide an independent constraint on the configuration of the large-scale magnetic field. Those constraints can be used either to verify the accuracy of the magnetic field calculated by other means (e.g. through self-consistent ring current modeling) or can be used to specify where the phase space densities should match and hence how the magnetic field model should be modified to ensure that match.

Another strength of this technique is that it can be applied to observations for which only particle measurements, and not magnetic field measurements, are available. In that case, phase space density can be calculated based on the model magnetic field. Again phase space densities at two (or more) satellite locations can be compared and adjustments made to the field as needed. Of course this modifies the phase space density based on the model field calculation so the process must be repeated iteratively. For spacecraft with magnetometer measurements the field is
Figure 4 illustrates the methods we have applied to the problem of determining the global storm-time magnetic field using representations that (a) are self consistent with the ring current particle distributions and (b) can be efficiently used in numerical calculations. The top panel illustrates the use of Euler potential specification of the geomagnetic field. In the Euler specification two scalars are used the magnetic field vector at any point is given by the cross product of the gradient of those scalars. Added benefits are that the field is, by design, divergence free, and that magnetic field lines are simply the intersection of those two iso-potential contours. In the bottom panel we show a first step in determining a self consistent calculation of the ring current and storm-time magnetic field using the UNH RAM code (after Zaharia et al., [2005]). Zaharia et al. calculated the plasma pressure distributions in the UNH-RAM code (using a dipole field), then calculated the 3-dimensional magnetic configuration in force balance with those pressure distributions, using an Euler potential equilibrium code. The bottom panel shows the difference in the field calculated using this procedure and the original dipole field. Values of B-B_dip over 100 nT during the main phase illustrate the need to calculate particle motions in a self-consistent magnetic field model. The middle panel shows Dst for this storm (beginning October 21, 2001) with four selected times marked.
known and those known values of the field need to be used in the process of optimizing the magnetic field model so that any adjustments to the field do not produce contradictory values of the field. Practically this is done through the use of an error function that one minimizes in order to get the best agreement between phase space densities and measured vs model field values.

At low energies electric field drifts dominate over magnetic gradient-curvature drifts and at intermediate energies both electric and magnetic drifts contribute. Thus, in principle, with a sufficiently dense set of measurements over the right range of energies the electric field can be constrained in the same manner as the magnetic field (although validation of the electric field is more straightforward than specification of the field).

Whether magnetic and electric fields or just magnetic fields are being specified, the more satellite measurements that are available the more powerful this technique can be. Currently there are six geosynchronous satellites carrying LANL particle instruments, two GOES satellites carrying SEC particle and field instruments, POLAR, HEO, and Akebono in inclined orbits, SAMPEX and numerous other low-Earth-orbit satellites, and the GPS constellation. Within the next six years we expect the full constellation of 24 GPS satellites to be equipped with energetic particle detectors. In about the same time frame NASA also plans to launch the two Radiation Belt Storm Probes into CRRES-like equatorial elliptical orbits. This represents a sufficiently dense set of satellites that many and frequent conjunctions occur over a broad range of L-shells.

This technique has already been demonstrated for the field in the vicinity of geosynchronous orbit by Onsager et al., [2004] and Chen et al., [2005a, b]. Onsager et al., [2004] used the fact that two geosynchronous satellites at different longitudes are also at different latitudes and therefore trace out different trajectories through L* (which is a common representation of the drift invariant [Roederer, 1970]). At two points on the orbit the satellites cross the same L* where the phase space densities should match. (See figure 6.). At other locations the two satellites measure the instantaneous local radial gradient in phase space density. Chen et al., [2005a, b] extended this technique using both the GOES and LANL measurements. They calculated phase space density for a variety of equatorial pitch angles which all follow slightly different contours of L* and cross in slightly different locations. They found that, at times the standard empirical magnetic field models (e.g. Olsen Pfitzer [1974]; Tsyganenko [2002]; etc.) gave excellent agreement over a very broad range of adiabatic invariants. Chen et al., [2005b] then went beyond verification to use the particle measurements to specify the magnetic field. They implemented 7 different magnetic field models and constructed an error function based on matching both the magnetic field (vector or unit vector) and particle phase space densities. By calculating the error function in each field they could determine which empirical field best matched the global magnetic conditions near geosynchronous orbit at any given time and then switch, dynamically, between models to determine the evolution of the magnetic field throughout a storm (figure 7).

So far this technique has been limited by the choice of spacecraft and field models. The next logical steps are to extend it in L-shell by using other spacecraft and to use a continuously deformable magnetic field to remove the current temporal discontinuities caused by instantaneously switching magnetic field models. Ultimately, the two techniques described in this section—empirical and self-consistent calculations of the magnetic field—can be combined using the standard, but powerful, techniques of data assimilation that we have discussed briefly here.

SUMMARY

The Earth’s radiation belts provide a rich field of study for basic physical processes as well as an important topic for the design and operation of space-based technology systems. Recent observations and newly-emerging theories have made significant advances in our understanding but still leave many important questions unanswered. Better understanding and quantitative prediction of changes in Earth’s natural radiation belts have high value for spacecraft design, systems operations, and manned exploration programs. These issues are generally well-known within the space physics community. Less well-known are the threats posed by high altitude nuclear explosions (HANE) or the possible steps that could be taken to mitigate those threats through active radiation belt remediation (RBR). Evaluating the production and dynamics of artificial radiation belts, understanding their potential impact on national space-based infrastructure, and quantitatively evaluating the effectiveness of mitigation strategies requires the same ability to understand and model physical processes as does study of the natural variation of the Earth’s belts.

Developing physical understanding of the key transport, acceleration, and loss processes requires a three-pronged approach: (1) a targeted, multi-satellite observational campaign to fill in holes in our knowledge of radiation belt dynamics; (2) a strong program to develop improved theoretical descriptions of key processes; and (3) development of global, time-dependent, data-driven but physics-based models. We have described one promising approach to provide the third of these critical components, a next-generation data assimilation model of the radiation belts. Our approach combines data-driven but physics-based modeling of the storm-
Figure 5 shows how particle measurements can be used to further improve the accuracy of global magnetic field models. Shown, schematically, are the drift paths of 100 keV and 30 keV protons along with a proposed configuration for a multi-satellite inner magnetosphere mission [Reeves et al., 1997]. Particle measurements can be used along with field models to calculate phase space density at each observation point. Liouville’s theorem specifies that (in the absence of non-adiabatic effects) the phase space density must be constant where satellites measure particles on the same dynamic trajectory. This must be true at all energies and all pitch angles independently (but for the same global field). Thus Liouville’s theorem can be used with multi-point particle measurements to adjust the global field (and recalculate the adiabatic invariants) until optimal matching of phase space densities is obtained.

Figure 6 shows the application of phase space density matching to determination of the optimal global magnetic field configuration in the vicinity of geosynchronous orbit. Panel A illustrates the technique. Because of the Earth’s dipole tilt geosynchronous satellites at different longitudes are at slightly different magnetic latitudes and therefore trace out different paths in $L^*$ as they complete their 24-hour orbit around the Earth. Panel B shows the phase space density color coded as a function of location (in $L^*$) and time for two geosynchronous satellites. Each satellite traces a different path in $L^*$. Where those paths cross the phase space densities should match (as shown by the same color dots). Where satellites sample different $L^*$ the phase space density gradient over that range of $L^*$ can be measured [Chen et al., 2005a, b]. (This plot shows phase space density calculated by optimally fitting 7 different magnetic field models to measurements of the geosynchronous magnetic field made by GOES. Some discontinuities in $L^*$ appear where the optimal magnetic field model changes.)
time ring current, a self-consistent specification of the global magnetic field, and powerful data assimilation techniques that are well-developed in other fields but only recently applied to the radiation belt problem. One of the promises of this proposed approach is that many of the limitations of the current (extensive) set of spacecraft observations can be overcome through assimilation with physics-based models. For example the lack of magnetometers, pitch angle measurements, or the specification of adiabatic invariants can be compensated for if one applies a magnetic field model that is physically realistic. At the same time the observations can be used not only as input or as a consistency check on the models but can actually modify the models to have the accuracy and realistic dynamics required.

The space physics community has begun to take critical steps to implement the program outlined in this paper. We have verified that the key techniques do work, that critical observations are available and can be assimilated with physical models, that the needed improvements in numerical methods and implementation are possible: in short, the problem can be solved if the right resources are applied. Clearly the program outlined here is not trivial and certainly requires effort significantly larger than commonly provided by individual research grants. However, those challenges do not seem insurmountable—especially when compared to the potential reward in scientific understanding, in real cost savings to satellite designers and operations, and to improved reliability and risk assessment for commercial and military programs that depend on systems in space.

Acknowledgements. We gratefully acknowledge R. A. Christensen, E. Noveroske, T. E. Cayton, J. B. Blake, J. F. Fennell, T. G. Onsager, and many colleagues from the NSF Geospace Environment Modeling (GEM) workshops for providing data used in this study and for many helpful and inspiring discussions. We thank the US Department of Energy, the NSF GEM program, and NASA’s Living With A Star (LWS) program for financial support of this study.

REFERENCES


Summers, D., and C. Ma, A model for generating relativistic electrons in the earth’s inner magnetosphere based on gyroresonant


Geoffrey D. Reeves, Reiner H. W. Friedel, Michelle F. Thomsen, Sorin Zaharia, Michael G. Henderson, and Yue Chen, Space Sciences and Applications, Los Alamos National Laboratory, Mail Stop D-644, Los Alamos NM, 87545

Sebastien Bourdarie, CERT-ONERA/DERTS, 2 Avenue Edouard Belin, PO Box 4025, 31055 Toulouse, France

Vania K. Jordanova, Institute for the Study of Earth, Oceans, and Space Science Center, University of New Hampshire, Durham NH, 03824

Brian J. Albright, and Dan Winske, Plasma Physics, Los Alamos National Laboratory, Mail Stop B-259, Los Alamos NM, 87545