

CHAPTER 8

SPECIAL PERTURBATION TECHNIQUES

- 8.1 Historical Background
- 8.2 Introduction to Perturbations
- 8.3 Encke's Formulation
- 8.4 Cowell's Formulation
- 8.5 Numerical Integration Methods
- 8.6 Disturbing Forces
- 8.7 Forming Numerical Solutions
- 8.8 Practical Considerations

8.1 Historical Background

The fundamental problem in perturbation analysis is orbit propagation. Unlike the two-body analysis of Algorithm 7, the most accurate way to analyze perturbations is numerically. We often discuss the sheer complexity of numerical integration, but to the early pioneers in celestial mechanics, the complexity often signaled a dead end. Of particular importance to the early scientists was their limited ability to calculate large numbers. Indeed, calculations weren't reasonable to perform until John Napier (1550–1617) invented logarithms in 1614. Incidentally, Napier (and later Jorst Borgi) worked on logarithms to help solve problems in astronomy.

As techniques for manipulating and dealing with larger numbers grew, so did the complexity of the theories. It quickly became apparent that the complexity of these theories could easily outpace the advances in computing technology. The problem then became how to efficiently evaluate complex theories for computational applications.

We'll see shortly that solution techniques for the perturbation problem fall into three broad categories: analytical, numerical, and semianalytical. The analytical approaches were generally developed first because the other methods required computational horsepower that didn't exist at the time. (We'll discuss analytical techniques in Chap. 9.) But today the modern computer has erased this limitation and opened a whole new era in analyzing perturbations.

The end of the 18th century brought significant developments in perturbation theory particularly in the modeling of the Earth's gravitational field. Laplace described a very useful analysis tool—the potential function. However, Laplace's most enduring legacy may be the phrase “it is easy to see that...” (Bell, 1986:177). He didn't care to do detailed derivations when the results were obvious (after the fact). Four volumes of his monumental work in celestial dynamics, *Méchanique céleste* (Celestial Mechanics) were published in 1799 and 1802–1805. The fifth and final volume was published between 1823 and 1825. It contained scant derivations but included his significant

contribution of the potential function. We'll explore several potential functions in this chapter.

Adrian Marie Legendre (1752–1833), a French mathematician and professor, did his main work on elliptic functions and number theory. Legendre functions are the solutions of the differential equations arising from his studies of the attraction of spheroids (Boyce and DiPrima, 1977:83). He published his discovery in 1783 in the first of four memoirs on spheroids. They're an integral part of solving the gravitational-potential problem.

In 1849, Sir George Gabriel Stokes (1819–1903), a professor at the University of Cambridge, published an extremely important formula which allowed accurate modeling of the Earth's shape (Caputo, 1967:62). Essentially, Stokes found that we can determine the shape of a geoid if we know the local gravity anomalies. This discovery permitted practical evaluation of the Earth's gravity field.

Johann Franz Encke (1791–1865) presented a formulation for computing orbits whenever the perturbations were small with respect to the two-body motion. This method was very popular for decades. He numerically integrated the differences of the osculating orbit from a reference orbit. By integrating the small differences between the orbits, rather than the complete attraction and specific forces involved, he attained enough precision with limited computing abilities. I'll only briefly introduce this method because computing technology has overcome the need to use such techniques.

Most numerical techniques didn't receive much attention until the 20th century because computing ability was limited. The *concept* of numerical integration was well understood, but it didn't receive much attention in astrodynamics until Phillip Herbert Cowell (1870–1949) used a numerical technique to determine the orbit of Jupiter's eighth satellite. Cowell and Crommelin also used a numerical procedure to predict two of three return visits of Halley's comet between 1759 and 1910. Their results were published in the "Appendix" to the *Greenwich Observations* 1909 as an "Investigation of the Motion of Halley's Comet." Cowell's method has been "rediscovered" many times and continues to gain favor in technical circles as computers become faster and more powerful. In modern times, it's common in astrodynamics to use *Cowell's formulation* to set up the equations of motion for numerical integration.

8.2 Introduction to Perturbations

Studying and modeling perturbations are key disciplines in astrodynamics. Although most of the solution techniques have been known for a long time, the literature only recently has begun to present (with increasing exactness) the precise methods needed to satisfactorily solve these problems. Leon Blitzer (1970) developed the *Handbook of Orbital Perturbations* and has graciously allowed me to use it as a basis for this chapter and Chap. 9. Blitzer was one of the early scientists to begin detailed analysis of the perturbative effects on near-Earth satellites, publishing several papers in the mid-1950s before satellites had become routine. I've also incorporated personal discussions with (listed alphabetically) Cefola, Fonte, Gabor, Junkins, and McClain to support this section. A complete list of their works would take many pages! Szebehely (1989:13–22)

lists other major contributors to the field of perturbations and astrodynamics. Lambeck (1988:654–709) includes extensive references on perturbations. Consult these sources for more detail.

Perturbations are deviations from a normal, idealized, or undisturbed motion. We tend to view the universe as highly regular and predictable. Yet accurate observational data often reveals unexplained irregularities of motion superimposed on the more regular or mean motions of the celestial bodies. Actual problems don't conform exactly to the assumptions made for the ideal two-body and restricted three-body problems. Therefore, these idealized solutions only approximate the observed motion. The actual motion will vary from the theoretical two-body path due to perturbations caused by other bodies (such as the Sun and Moon) and additional forces not considered in Keplerian motion (such as a nonspherical central body and drag). Notice we use the term *central body* to show these perturbation techniques apply about *any* central body. For missions to other planets, the Earth may not be the central body. Indeed, the formulas work equally well for the Earth, Mars, Saturn, and so on—assuming, of course, that we have accurately determined the data constants and coefficients for each central body.

Bate, Mueller, and White (1971:385–386) introduce perturbations as follows. We're familiar with random perturbations in almost every area of life. Seldom does anything go exactly as planned; rather, it is perturbed by unpredictable circumstances. An excellent example is an aircraft encountering a wind gust. All directional controls and power are kept constant; yet, the path changes abruptly from the predicted course that was based on previous measurements. In this example, the wind gust is a random perturbing force. Fortunately, most of the perturbing forces considered in orbital mechanics are more predictable and analytically tractable than a wind gust. The usual perturbing forces we consider for the orbit problem are both random and predictable. They include the asphericity of the central body, atmospheric drag and lift, third body effects, solar-radiation pressure, thrust, magnetic fields, solid-Earth and ocean tides, Earth re-radiation (albedo), relativistic effects, and others. I'll show how to analyze most of these later.

Don't get the idea that all deviations are small—they can be comparable to the primary attracting force (two-body gravitation). Whenever they're this large, they cease to be perturbations because the fundamental nature of the problem is different. Examples are atmospheric drag during reentry and the third body in the restricted three-body problem. In fact, many interplanetary missions would miss their target entirely if the perturbing effect of other attracting bodies weren't taken into account! Ignoring the effects of the central body's oblateness on any satellite keeps us from accurately predicting its position over a long time. Without analyzing perturbations, we couldn't explain or accurately predict the orbit of the Moon about the Earth.

We'll look at the three main approaches to examine the effects of perturbations: *special perturbation techniques* (using numerical methods—Sec. 8.5), *general perturbation techniques* (using analytical methods—Chap. 9), and *semianalytical techniques* (a combination of analytical and numerical techniques—Chap. 9).

Special perturbation techniques numerically integrate the equations of motion including all necessary perturbing accelerations. Because numerical integration is involved, we can think of numerical formulations as producing a *specific*, or *special*,

answer that is valid only for the given data (initial conditions and force-model parameters). Although numerical methods can give very accurate results and often establish the “truth” in analyses, they suffer from their specificity, which keeps us from using them in a different problem. Thus, new data means new integration, which can add lengthy computing times. NASA began the first complex numerical integrations during the late 1960s and early 1970s. Personal computers now compute sufficiently fast enough to perform complex perturbation analyses using numerical techniques. However, numerical methods suffer from errors that build up with truncation and round-off due to fixed computer wordlength. These errors can cause numerical solutions to degrade as the propagation interval lengthens.

Let's look at an example to illustrate the strengths and weaknesses of a numerical method. Start with the equation of motion in Example 8-1, which is simple, first order, and defined with a given initial condition. Find the value of this system when time, t , equals 0.5^s, assuming $\omega = 1 \text{ rad/s}$. Obviously, this example permits a direct solution if we recognize the integral as $(\sin(\omega t))/\omega$.

▼ Example 8-1: Using Direct Integration.

GIVEN: $\dot{y}(t) = \cos(\omega t)$

$y(t = 0) = y_o = 0$

$\omega = 1 \text{ rad/s}$

FIND: $y(t = 0.5)$

Remembering to evaluate both limits of the integration:

$$y(t = 0.5) = \int_0^{0.5} \cos(t) dt = \sin(0.5) - 0$$

▲ Be sure to use radians so that $y(0.5) = 0.479\,425\,538\,604\,203$.

Unfortunately, most systems don't have such simple equations of motion, and often the integrals are very complex, as with the orbit problem.

Now let's use a numerical technique to solve the problem in Example 8-1. A Taylor series is the underlying principle. For simplicity, use an Euler integrator (first-order derivative of the Taylor series) to propagate the system from $t = 0$ to $t = 0.5$ seconds. The general form of a Taylor series expansion is

$$y(t) = y(t_o) + \dot{y}(t_o)(t - t_o) + \frac{\ddot{y}(t_o)(t - t_o)^2}{2!} + \frac{\ddot{\ddot{y}}(t_o)(t - t_o)^3}{3!} + \dots$$

where the step size ($\Delta t = t - t_o$) is the difference in time. If we retain only the first-order term, the expansion becomes an Euler integrator and allows us to solve the problem. Remember that Δt must be small to permit the removal of higher-order terms.

▼ Example 8-2. Using a Numerical Technique.

GIVEN: $\dot{y}(t) = \cos(\omega t)$

$$y(t=0) = y_0 = 0, \omega = 1 \text{ rad/s}$$

FIND: $y(t = 0.5)$

For each trial, $y(t) = y(t_o) + \dot{y}(t_o) \Delta t$, where Δt is the step size ($t - t_o$) and $\dot{y} = \cos(t)$. Increment the time as $t = t_o + \Delta t$. Begin by assuming a step size of $\Delta t = 0.5$.

Let $\Delta t = 0.5$ and solve (to first order):

$$y(t = 0.5) = y(t_o) + \cos(t_o) (t - t_o) = 0.0 + \cos(0.0) (0.5 - 0) = \underline{\underline{0.5}}$$

Notice how far off from the correct answer this result is. Now try a step size of 0.1 (Let $\Delta t = 0.1$).

$$t_k = 0$$

$$y(t_{k+1} = 0.1) = y(t_k) + \cos(t_k) (t_{k+1} - t_k) = 0.0 + \cos(0.0) (0.1 - 0) = 0.1$$

$$t_k = t_{k+1}, t_{k+1} = t_k + (t_{k+1} - t_k) = t_k + \Delta t$$

$$y(t_{k+1} = 0.2) = y(t_k) + \cos(t_k) (t_{k+1} - t_k) = 0.1 + \cos(0.1) (0.2 - 0.1) = 0.199\ 500\ 4$$

$$t_k = t_{k+1}, t_{k+1} = t_k + \Delta t$$

$$y(t_{k+1} = 0.3) = y(t_k) + \cos(t_k) (t_{k+1} - t_k) = 0.199\ 500\ 4 + \cos(0.2) (0.3 - 0.2) = 0.297\ 507\ 1$$

$$t_k = t_{k+1}, t_{k+1} = t_k + \Delta t$$

$$y(t_{k+1} = 0.4) = y(t_k) + \cos(t_k) (t_{k+1} - t_k) = 0.297\ 507\ 1 + \cos(0.3) (0.4 - 0.3) = 0.393\ 040\ 7$$

$$t_k = t_{k+1}, t_{k+1} = t_k + \Delta t$$

$$y(t_{k+1} = 0.5) = y(t_k) + \cos(t_k) (t_{k+1} - t_k) = 0.393\ 040\ 7 + \cos(0.4) (0.5 - 0.4) = \underline{\underline{0.485\ 146\ 8}}$$



Notice several features. First, each subsequent equation uses the *previous* value for the function (0.0, 0.1, 0.199 500 4, ...). This stepwise process is a basic characteristic of all numerical techniques and is the reason the results are valid for only one set of initial conditions. If we change the original initial condition, we must recompute the entire sequence. In fact, let's suppose we want to determine the value of the function at $t = 0.48$. We could use interpolation, which is sometimes as accurate as the numerical integrator, or we could start at $t = 0.0$ and proceed with a new step size to determine the final answer. Notice how the step size is constant throughout this example. Step size changes could easily be input, and the example of changing to determine the result at $t = 0.48$ wouldn't be difficult. Some integrators used for solving the orbit problem employ this fixed step-size approach. Note the improvement with the smaller step size. The obvious conclusion from these two cases is that we should use the smallest possible step size of, say, 0.000 000 01, to obtain the most accurate answer of 0.479 425 539 216 302, right? Well, maybe. This choice *is* more accurate than the analytical method (see Sec. 9.2), but it uses too much computational time. At some point, the benefit of using a smaller and smaller step size is overwhelmed by the roundoff and truncation errors. The problem of determining an appropriate step size can be significant and is a challenge in any numerical process.

Besides selecting an integration technique, we must also consider the forces acting on the satellite. The forces causing the perturbative effects on the satellite are either *conservative* or *nonconservative*. The total energy (kinetic and potential) for conservative-force systems is constant, whereas nonconservative systems may lose or gain energy, mainly through heat exchange (friction) or external sources (thrust). Central-body and third-body gravitational effects are examples of conservative forces; solar-radiation

pressure, thrust, and drag are examples of nonconservative forces. Next, we'll see that we can determine accelerations for conservative forces as the gradient of a potential function.

You should know about *gradients*, *accelerations (specific forces)*, and *functions*. A **gradient** is really a directional derivative which gives the rate of change of a *scalar function* in a particular direction (Kreyszig, 1983:387). It's a vector quantity and the **del operator**, ∇ , designates the gradient process. The gradient gives an acceleration if the scalar function is a potential function related to a specific potential energy, such as the potential function of a central body's gravity field. I distinguish a potential function as the negative of the potential energy (see Sec. 1.3.3). Two conventions are "standard" in this area because many schools of thought have evolved over the last few decades. Brouwer and Clemence (1961), Battin (1987), Long et al. (1989), and others express one of the two main approaches, in which the acceleration is the negative gradient of the potential function. This implies that positive work is done as the potential decreases. The other approach, used mainly by the geophysical community, holds an acceleration to be the positive gradient of the potential function [Lambeck (1988), Kaula (1966), Moritz and Mueller (1987), Kaplan (1976), Roy (1988), and others]. Of course, both methods use potential functions that differ only by a minus sign; therefore, the results are identical! We'll follow the second method and place the sign change between the potential energy and the potential function. I'll also refer to the potential function instead of simply the *potential*, to avoid confusion with *potential energy*.

The distinction between a specific force (used interchangeably with acceleration in this text) and a potential is important because analysis of perturbations typically uses both concepts. It's common to analyze perturbations using a *disturbing function* and a *disturbing force*. The **disturbing force** simply expresses (in some coordinate system) the specific force (acceleration) that is perturbing the satellite's orbit. Nonconservative forces, such as the perturbing effects of drag and solar-radiation pressure, are usually modeled as a specific force. **Disturbing functions** are simply the difference between perturbed and unperturbed potential functions. They model conservative forces that perturb the orbit, such as the central body's nonsphericity and third-body attractions.

A **potential function** is one way to mathematically characterize a conservative force, such as the gravitational potential of a *spherical* central body ($U_{2\text{-body}} = \mu/r$). Some people distinguish a disturbing function from a disturbing potential by a minus sign. As mentioned earlier, considering the two to be equal is just as correct, as long as we maintain the correct sign convention. We'll soon develop a potential function for an *aspherical* central body, U (sometimes referred to as the *anomalous potential*). As you may have guessed, the form is much more complicated than the spherical one. It includes the spherical potential ($U_{2\text{-body}}$) as the first term. The term *geopotential* is often used for this aspherical potential when the central body is the Earth. Using the prefixes helps eliminate any possibility of confusion.

The existence of the spherical and aspherical potentials enables us to form a **disturbing-potential function**, R , defined as the difference between the two potentials ($R = U - U_{2\text{-body}}$). The disturbing-potential function represents *only* the perturbing effect of the nonspherical Earth (in this case), so we must carefully distinguish it when working

problems. The acceleration from the spherical central body's potential, $\nabla U_{2\text{-body}}$, yields the *two-body* acceleration; the gradient of the aspherical central body's potential, ∇U , yields the *complete* acceleration. The acceleration from the disturbing function, ∇R , will yield *only* the acceleration of the perturbing effect, which will equal the difference between the two-body and complete accelerations. We follow this order in defining R because it leaves the disturbing part of the potential as a positive quantity. The accelerations from the spherical potentials and the disturbing potential may differ by orders of magnitude, so be careful!

An example will help us understand the gradient vector operator as it applies to a scalar potential function. Examine the potential of the spherical central body identified previously. The del operator acts by taking the partial derivatives in each of the respective axes. In cartesian coordinates,

$$U_{2\text{-body}} = \frac{\mu}{r}$$

$$\nabla U_{2\text{-body}} = \frac{\partial U_{2\text{-body}}}{\partial x} \hat{I} + \frac{\partial U_{2\text{-body}}}{\partial y} \hat{J} + \frac{\partial U_{2\text{-body}}}{\partial z} \hat{K}$$

where IJK denote unit vectors along the axes of the coordinate system and xyz are the components of the position vector. With a position vector $\vec{r} = x\hat{I} + y\hat{J} + z\hat{K}$ having a magnitude of $r = \sqrt{x^2 + y^2 + z^2}$, the acceleration is

$$\nabla U_{2\text{-body}} = -\frac{1}{2} \frac{\mu}{(x^2 + y^2 + z^2)^{3/2}} [2x\hat{I} + 2y\hat{J} + 2z\hat{K}] = -\frac{\mu}{r^3} \vec{r}$$

Notice the result, as expected, is the two-body acceleration for a satellite. Thus we can write the two-body equation [Eq. (1-14)] very compactly:

$$\ddot{\vec{r}}_{2\text{-body}} = \nabla U_{2\text{-body}} \quad (8-1)$$

If we add conservative perturbative forces to the simple Newtonian attraction, the equation of motion becomes

$$\ddot{\vec{r}} - \nabla U_{2\text{-body}} = \nabla R \quad \text{or} \quad \ddot{\vec{r}} = \nabla U \quad (8-2)$$

The form of the disturbing-potential function, R , will depend on the particular type of perturbing source (aspherical central body, third-body attractions). If there are multiple perturbing forces, the disturbing functions due to each are added to give the total disturbing function, R .

8.3 Encke's Formulation

As mentioned in the introduction, Encke's formulation* isn't very popular today because computers can easily perform a numerical solution to many problems with enough precision to retain the accuracy of the theories. Still, it is historically relevant and it illustrates a computational technique that we can apply to other problems. Refer

to Plummer (1918:222–224), as well as Brouwer and Clemence (1961:176–185), for classical presentations and details on the formulation. Kaplan (1976:345–348) presents a concise algorithm.

Encke's formulation begins with an osculating orbit. Instead of integrating all the forces on the satellite, he integrates just the difference between the two-body acceleration and the perturbed acceleration. The perturbations to the orbit are integrated in cartesian elements. Because this technique integrates only the perturbations, the magnitudes are much smaller and the computational precision can actually be greater. The process continues until a *rectification point* where the osculating orbit is re-initialized. In Algorithm 62, this point is governed by a tolerance. Kaplan (1976:348) suggests about 1%. Battin (1987:450) notes that the rectification is advantageous in controlling the errors. This is especially useful for interplanetary trajectories. Figure 8-1 shows the geometry.

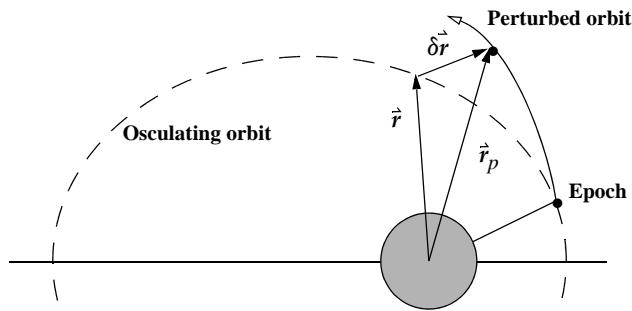


Figure 8-1. Encke's Formulation. The foundation for Encke's technique relies on the difference between the osculating (two-body) orbit and the actual perturbed motion, shown as $\delta\mathbf{r}$. By integrating this difference, it can save many computations.

We know the two-body and perturbed accelerations:

$$\ddot{\mathbf{r}} = -\frac{\mu}{r^3}\mathbf{r} \quad \ddot{\mathbf{r}}_p = -\frac{\mu}{r_p^3}\mathbf{r}_p + \hat{\mathbf{a}}_p$$

This means the difference between the two accelerations is

$$\ddot{\delta\mathbf{r}} = \hat{\mathbf{a}}_p + \frac{\mu}{r^3} \left\{ \left(1 - \frac{r^3}{r_p^3} \right) \mathbf{r}_p - \delta\mathbf{r} \right\}$$

Notice the appearance of the difference of two nearly equal quantities. Kaplan (1976: 347–348) shows how we can treat this parameter and form a solution using Algorithm 62, below (see also Sec. 8.6.3). We can use two-body relations to move the osculating orbit forward in time.

* I use the formulation terminology to be consistent with the discussion of Cowell's formulation. Encke did, however, use the phrase *neue methode* in Berliner Jahrbuch, c1857.

ALGORITHM 62: ENCKE ($\dot{\vec{r}}_o, \dot{\vec{v}}_o, t_{start}, t_{stop}, \Delta t \Rightarrow \dot{\vec{r}}, \dot{\vec{v}}$)

$$\delta\dot{\vec{r}} = \dot{\vec{0}} \quad \epsilon = 0 \quad f = 0 \quad \dot{\vec{r}}_p = \dot{\vec{r}}_o \quad \dot{\vec{v}}_p = \dot{\vec{v}}_o$$

$$t = t_{start}$$

LOOP

KEPLER ($\dot{\vec{r}}_o, \dot{\vec{v}}_o, t - t_{start} \Rightarrow \dot{\vec{r}}, \dot{\vec{v}}$)

$$\epsilon = \frac{\dot{\vec{r}} \cdot \delta\dot{\vec{r}}}{r^2} \quad f = \frac{1}{\epsilon} \left\{ 1 - \frac{1}{(1 - 2\epsilon)^{3/2}} \right\}$$

$$\ddot{\delta\dot{r}} = \ddot{\vec{a}}_p + \frac{\mu}{r^3} \{ f\epsilon\dot{\vec{r}}_p - \delta\dot{\vec{r}} \}$$

$$\begin{cases} \text{IF } \frac{\delta r}{r_p} > \text{tolerance } (\sim 1\%) \\ \quad \dot{\vec{r}} = \dot{\vec{r}}_p \quad \dot{\vec{v}} = \dot{\vec{v}}_p \\ \text{ELSE} \\ \quad \dot{\vec{r}}_p = \dot{\vec{r}} + \delta\dot{\vec{r}} \quad \dot{\vec{v}}_p = \dot{\vec{v}} + \delta\dot{\vec{v}} \end{cases}$$

$$t = t + \Delta t$$

UNTIL $t > t_{stop}$

8.4 Cowell's Formulation

As mentioned earlier, numerical methods are enjoying ever-increasing popularity over some analytical methods as computers become faster. One of the most attractive characteristics for numerical techniques is that we can incorporate any arbitrary disturbing acceleration. Although the two-body assumption neglects the effects of real-world perturbations, one way to account for these effects is to add the perturbing accelerations to the two-body equation to produce a more accurate equation of motion,

$$\ddot{\vec{r}} = -\frac{\mu}{r^3} \dot{\vec{r}} + \ddot{\vec{a}}_{perturbed} \quad (8-3)$$

Here, $\ddot{\vec{a}}_{perturbed}$ is the total acceleration caused by other forces acting on the satellite, shown as an additive effect to the original equation. The specific form of $\ddot{\vec{a}}_{perturbed}$ will depend on the number and type of perturbing sources. We'll examine perturbing effects of a nonspherical central body, atmospheric drag, third-body effects (Sun, Moon, and planets), solar-radiation pressure, thrust, and tides. Fortunately, the form of Eq. (8-3) permits us to add each effect linearly. Eq. (8-3) models the complex interactions among the effects but leaves us with several difficulties. For example, the particular expressions

for a perturbing acceleration can be very complex and computationally intensive. Obtaining the necessary data to use with each model may also limit us, such as in determining the gravitational coefficients for Pluto, a planet whose equatorial radius we don't know precisely. Calculating the Earth's atmospheric density is extremely difficult mainly because we don't fully understand the Sun's internal mechanics and the precise interaction of solar radiation with the molecules in the Earth's upper atmosphere. The Earth's geomagnetic field is also complex.

Equation (8-3) has come to be known as *Cowell's formulation*. I distinguish between *Cowell's method* and *Cowell's formulation*. The preceding discussion is based on Cowell's formulation, in which we formulate second-order differential equations of motion and numerically integrate them. *Cowell's method* is a technique using the *calculus of finite differences* to accomplish the integration. It's preserved in many places, including Brower and Clemence (1961:167–175).

Many techniques for numerical integration have been developed to operate on various first- and second-order differential equations. It's common to re-form the three second-order differential equations of motion into six first-order differential equations because we can then use a broader class of integration methods. This first-order system is known as a variation of Cowell's formulation.

$$\bar{\mathbf{X}} = \begin{bmatrix} \dot{\mathbf{r}} \\ \dot{\mathbf{v}} \end{bmatrix} \quad \dot{\bar{\mathbf{X}}} = \begin{bmatrix} \hat{\mathbf{v}} \\ -\frac{\mu \hat{\mathbf{r}}}{r^3} + \hat{\mathbf{a}}_p \end{bmatrix}$$

8.5 Numerical Integration Methods

The equations of motion used in astrodynamics aren't trivial to solve because they're usually coupled systems of equations—first-order or second-order, nonlinear, differential equations which have resisted direct solution over the last 300 years. But modern computers allow us to use numerical techniques. In this section, we'll discuss both *single-step* and *multi-step* methods for numerical integration problems. We could spend considerable time discussing the numerous types of integrators and which is appropriate for a particular problem, so I've introduced only some of the more popular techniques. Sadler (1974) presents fundamental information about numerical integration and notation used in the equations.

Single-step methods combine the state at one time with rates at several other times, based on the single-state value at time, t_o . The rates are readily obtained from the equations of motion and allow us to determine the state at succeeding times, $t_o + h$. For many applications, a fourth-order *Runge-Kutta* (RK4, single-step) method is sufficient. Runge-Kutta techniques are often preferable because they don't require a sequence of back values to start the integrator, which can sometimes increase the computer run-time. However, Runge-Kutta methods evaluate the function at several intermediate points, and these values are used only once.

Multi-step methods do an initial estimate (predictor) using previous estimates of the function's rate of change, and a second series of calculations (corrector) use the estimated answer to further refine the result. They are called *predictor-corrector* methods and they can sometimes improve the accuracy of the integration, but require us to maintain a series of back values for use in the algorithm. Consequently, they aren't self-starting, despite their efficiency and accuracy. The predictor-corrector methods are usually more accurate, but this improvement is achieved at the expense of increased complexity.

Step size techniques are employed with most numerical integration methods to limit error. **Fixed step-size** methods are most common for circular orbits because the step size is kept constant throughout the orbit. **Variable step-size** methods are better for eccentric orbits because they permit additional evaluations when the satellite is moving fastest, and fewer when the satellite is moving slowest. Variable step-size methods are more difficult to implement with multi-step techniques due to the use of back values, but it can be done in some situations.

Numerical integration methods are also classified by the equations they solve. **Single integration** techniques are designed for first-order differential equations. **Double integration** techniques solve second-order differential equations and produce direct answers. This is useful in astrodynamics [recall Eq. (8-3)], but when forces like atmospheric drag are included, the velocity is needed, requiring a single integration method.

It's also important to examine notation. For satellite calculations, we need $\dot{\vec{r}}$ when \vec{r} is known. If we model non-conservative forces like drag, we also need $\ddot{\vec{r}}$. Therefore, I'll change the notation here from y to show the use of position vectors and their derivatives. Maury and Segal (1969) state that summed forms are a way to limit the accumulated error from the integrated variable. **Summed forms** sum the *back differences*, $\nabla_{i-1}(\ddot{\vec{r}}_n)$, to form *summed back differences*, \hat{S}_n^{II} , and \hat{S}_n^{I} and they're only applicable for multi-step methods. The difference values come from summation tables that are usually presented with the derivation of numerical integration routines. The notation typically used with these tables can be very confusing. Sadler (1974:5) presents a concise explanation. I'll use parentheses to denote that the difference is a function of the state. Note the use of the binomial coefficient [Eq. (C-24)].

$$\begin{aligned}\nabla(\ddot{\vec{r}}_n) &= \ddot{\vec{r}}_n - \ddot{\vec{r}}_{n-1} & \nabla_2(\ddot{\vec{r}}_n) &= \ddot{\vec{r}}_n - 2\ddot{\vec{r}}_{n-1} + \ddot{\vec{r}}_{n-2} \\ \nabla_n(\ddot{\vec{r}}_n) &= \sum_{i=0}^j (-1)^i \binom{j}{i} \ddot{\vec{r}}_{n-i}\end{aligned}$$

Non-summed forms simply use the step-to-step back differences.

There are a few integrators that have gained popularity over the years and are included in existing orbit determination programs (See Table D-10 for a listing of programs). The Adams-Basforth-Moulton integrator is widely used in these programs. The Gauss-Jackson integrator also has a long history of operational use. Berry and Healy (2001) present a detailed discussion of this method and the required equations. Herrick comments (1972) that for near-circular, near-Earth orbits, Gauss-Jackson is about one order of magnitude more efficient than a fourth-order Runge-Kutta. However,

for higher eccentricities, thrusting, and high-drag orbits, the Runge-Kutta methods are usually competitive and less complex. Burlisch-Stoer represents an alternative that has fewer calls to the force-model equations (Pohlen and Titus, 1995). It reduces computer runtime, but it appears to be unreliable as a general-purpose integrator. It works best in quadruple precision when FORTRAN is used. Shampine-Gordon (1975, and Berry and Healy, 2004) has received recent interest and could become a plausible alternative.

Most numerical integrators actually derive from the simplest form of numerical integration—the *Taylor series* integrator.* Recall the form of a Taylor series:

$$y(t) = y(t_o) + \dot{y}(t_o)(t - t_o) + \frac{\ddot{y}(t_o)(t - t_o)^2}{2!} + \frac{\dddot{y}(t_o)(t - t_o)^3}{3!} + \dots \quad (8-4)$$

The difficulty here is two-fold: how do you include an infinite number of terms (or, if you don't, where do you truncate the series?), and how do you calculate all the derivatives, especially with extremely complex functions? A simple answer is the basis for an **Euler integrator** (after Leonhard Euler), which is simply a first-order Taylor series:

$$y(t) \approx y(t_o) + \dot{y}(t_o)(t - t_o) \quad (8-5)$$

Notice you need only the first derivative. The Euler integrator actually works by attempting to determine the slope (first derivative) of the function at the starting point. The subtle part of this scheme is that you must carefully choose the *step size*, defined here as ($h = t - t_o$). You can probably picture a situation in which a function changes drastically just after the point you've chosen to determine the slope. By examining terms only through the first order, you've implicitly *assumed* your step size is small enough to handle variations caused by the neglected higher-order derivatives. You've also assumed that you're operating in the near linear region.

Single-step: Runge-Kutta Methods

Perhaps the most well-known numerical integrators are the Runge-Kutta methods originally presented by Carl Runge (1856–1927) in 1895, and Wilhelm Kutta (1867–1944) in 1901, which also derive from a Taylor series. They differ from traditional Taylor series integrators because, instead of having to derive application-specific formulas for the higher derivative terms, we can form the approximation by simply using the slope at different points within the interval over which we'll integrate. The general equations for the classical fourth-order Runge-Kutta method are

* Brook Taylor (1685–1731) published his famous theorem in 1715 in *Methodus incrementorum directa et inversa*.

$$\begin{aligned}\dot{y}_1 &= f(t_o, y_o) & \dot{y}_2 &= f\left(t_o + \frac{h}{2}, y_o + \frac{h}{2}\dot{y}_1\right) \\ \dot{y}_3 &= f\left(t_o + \frac{h}{2}, y_o + \frac{h}{2}\dot{y}_2\right) & \dot{y}_4 &= f(t_o + h, y_o + h\dot{y}_3) \\ y(t) &= y(t_o) + \frac{h}{6}(\dot{y}_1 + 2\dot{y}_2 + 2\dot{y}_3 + \dot{y}_4) + O(h^5)\end{aligned}\quad (8-6)$$

Notice that we need only the first derivative, but we require four evaluations of the equations of motion to move forward one step in time—a characteristic of single-step methods.

For the satellite problem, we often formulate the equations of motion as a first-order system with initial conditions. Using position and velocity vectors,

$$f(t, y) = \begin{bmatrix} \dot{\vec{r}} \\ \vec{v} \\ \ddot{\vec{r}} \end{bmatrix} = \begin{bmatrix} \dot{\vec{v}} \\ \vec{a} \end{bmatrix} \quad y(t_o) = \begin{bmatrix} \vec{r}_o \\ \vec{v}_o \end{bmatrix} \quad \therefore \dot{y}_1 = \begin{bmatrix} \dot{\vec{v}} \\ \vec{a} \end{bmatrix}_{t=t_o} \quad (8-7)$$

The fourth-order terminology comes from the fact that our integrator is matched to a fourth-order Taylor series expansion about the initial value $y(t_o)$. We can relate the higher-order time derivatives to the first derivative at other times. The way the other times are chosen forms the particular integration algorithm. The fourth-order Runge-Kutta is really an Euler-type extrapolation using a weighted average of four rates (\dot{y}_i) in the interval $[t_o, t_o + h]$. Figure 8-2 illustrates the different evaluations of the slope. It can be shown that the truncation error for a fourth-order Runge-Kutta is $O(h^5)$, which is the same as a fourth-order Taylor series. In general, we may distinguish several **orders** of numerical-integration routines: the order to which the method is locally or globally correct, or the order of its local or global error. Conventionally, a method is termed fourth-order if it's locally accurate to fourth-order, globally correct to third-order, has fifth-order local error and fourth-order global error (Berry, 2004).

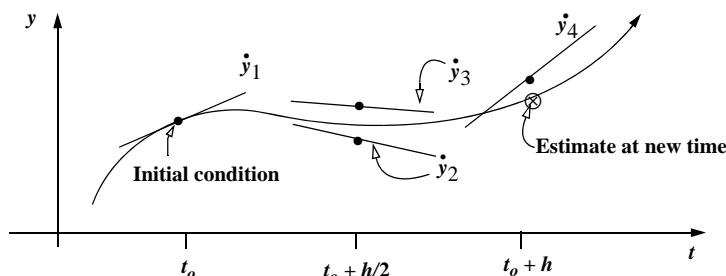


Figure 8-2. The Runge-Kutta Method. This algorithm works by evaluating the derivative (slope) at several different points (four for a fourth-order method) along the estimated trajectory. It's limited because all calculations depend on estimates of the function value. The intermediate estimates aid in finding the final estimate at the new time.

Several other forms derive from this basic method. There are eighth, twelfth, and other order Runge-Kutta methods, which operate as above, but with more terms. Consult Fehlberg (1968 and 1969) and Der (1995) for extensive treatments. They show the general iterative Runge-Kutta formulas, as well as an example six-stage, fifth-order, RK56, evaluation (i -order, j -evaluations per stage):

$$\begin{aligned} y(t) &= y_o + h \sum_{i=0}^{j-1} b_i \dot{y}_i + O(h^{n+1}) \quad n = 0, 1, 2, \dots, n-1 \\ \dot{y}_o &= f(t_n, y_n) \quad c_{io} = p_i - \sum_{j=1}^{i-1} c_{ij} \\ \dot{y}_i &= f\left(t_n + p_i h, y_n + h \sum_{j=0}^{i-1} c_{ij} \dot{y}_j\right) \quad i = 1, 2, \dots, j-1 \end{aligned}$$

The constants b_i , c_{ij} , and p_i must be determined separately. The actual process is lengthy and includes some user-chosen variables: $p_1, p_2, p_5 = 1$, $b_1 = 0$.

The *Runge-Kutta-Fehlberg* method, sometimes called an *embedded Runge Kutta*, uses a variable step size [Fehlberg (1968 and 1969)]. It greatly enhances the algorithm by adjusting h to keep local truncation errors within some tolerances. The method performs two numerical integrations, of different order, of the equations of motion at each step and compares the answers. If the answers are reasonably close, they're accepted. The step size is changed to maintain an even difference in the evaluations throughout the orbit. The final answer uses the initial values with a fourth- or fifth-order Runge-Kutta. This feature is especially nice on highly eccentric orbits like the Molniya. These orbits are best computed using variable step sizes to maintain some given level of accuracy. Without this variable step size, we waste a lot of time near apoapsis, when the integration is taking too small a step. Likewise, the integrator may not be using a small enough step size at periapsis, where the satellite is traveling very fast. Another approach uses an integrator with a fixed step-size and a variable time parameter, s , which varies as a function of the satellite's position (for example time-regularized Cowell on page 537). We use Eq. (8-7) as the first-order equations of motion for satellite problems. The Runge-Kutta-Fehlberg algorithm employs the following values:

$$k_1 = hf(t_o, y_o)$$

$$k_2 = hf\left(t_o + \frac{1}{4}h, y_o + \frac{1}{4}k_1\right)$$

$$k_3 = hf\left(t_o + \frac{3}{8}h, y_o + \frac{3}{32}k_1 + \frac{9}{32}k_2\right)$$

$$k_4 = hf\left(t_o + \frac{12}{13}h, y_o + \frac{1932}{2197}k_1 - \frac{7200}{2197}k_2 + \frac{7296}{2197}k_3\right)$$

$$k_5 = hf \left(t_o + h, y_o + \frac{439}{216}k_1 - 8k_2 + \frac{3680}{513}k_3 - \frac{845}{4104}k_4 \right)$$

$$k_6 = hf \left(t_o + \frac{1}{2}h, y_o + \frac{8}{27}k_1 + 2k_2 - \frac{3544}{2565}k_3 + \frac{1859}{4104}k_4 - \frac{11}{40}k_5 \right)$$

The two approximations use these values with a fourth-order and fifth-order Runge-Kutta algorithm. Their difference is

$$\Delta_{45} = \frac{1}{360}k_1 - \frac{128}{4275}k_3 - \frac{2197}{75240}k_4 + \frac{1}{50}k_5 + \frac{2}{55}k_6$$

If the absolute value of the delta is greater than some tolerance, say 1.0×10^{-5} , we determine whether to double or halve the step size based on the parameter, s .

$$s \cong 0.8408 \left[\frac{1 \times 10^{-8}h}{\Delta_{45}} \right]^{1/4}$$

Using predetermined minimum, $h_{min} = h/64$, and maximum, $h_{max} = h/64$, values for the step size, we determine any changes to the step size.

IF $s < 0.75$ and $h > 2h_{min}$ THEN $h = h/2$

IF $s > 1.5$ and $2h < h_{max}$ THEN $h = 2h$

Otherwise,

$$y = y_o + \frac{25}{216}k_1 + \frac{1408}{2565}k_3 + \frac{2197}{4104}k_4 - \frac{1}{5}k_5$$

Multi-step, Single Integration Techniques

The most popular methods here are the **Adams-Bashforth-Moulton** and **Shampine-Gordon**. Berry (2004) discusses these methods in detail, particularly Shampine-Gordon, which uses a divided difference (a similar operation to the back differences) to achieve a variable step-size. Adams-Bashforth-Moulton is well established in existing programs. It's a multi-step, fixed step-size method that estimates the state over time using previously determined back values of the solution. Although multi-step methods perform only one evaluation for each step forward (compared to four for the fourth-order Runge-Kutta), they usually have a predictor and a corrector formula, so they often require two evaluations per step. Because these methods require back values, many aren't self-starting. Berry and Healy (2001) show that you can start multi-step methods with iterative procedures that use the same method with the formulae shifted to correct the backpoints. You can also use a Runge-Kutta technique to supply the initial starting conditions, but you should match the order of both routines (e.g., eighth-order Runge-Kutta and an eighth-order Adams-Bashforth-Moulton). Consult Maury and Segal (1969), Long et al. (1989:6-2 to 6-20), Matthews (1987:434–437), and Press et al. (1992:740–744) for more information.

Most presentations start with sum and difference tables to evaluate each step. Herrick (1972:3-17 and 245-302) and Sadler (1974) describe this traditional approach. Montenbruck and Pfleger (1994:90–91) show how to derive multi-step methods, specifically Adams-Bashforth's fourth-order method proposed in 1883. Integrating both sides of the first-order equation, [$\dot{y}(t_{n+1}) = f(t, y(t))$], with initial condition [$y(0) = y(t_n)$], we get

$$y(t_{n+1}) = y(t_n) + \int_{t_n}^{t_{n+1}} f(t, y(t)) dt$$

Because we don't know the state at the future time, $y(t)$, we can use an interpolating polynomial. Choosing a cubic polynomial, we get the four points by evaluating the function at the current time, plus the last three times. Thus,

$$\begin{aligned} p(t) &= \alpha_0 + \alpha_1 \tau + \alpha_2 \tau^2 + \alpha_3 \tau^3 & \tau &= \frac{\Delta t}{h} \\ \alpha_0 &= f(y(t_n)) = f_n & \alpha_1 &= \frac{1}{6}(-2f_{n-3} + 9f_{n-2} - 18f_{n-1} + 11f_n) \\ \alpha_2 &= \frac{1}{6}(-3f_{n-3} + 12f_{n-2} - 15f_{n-1} + 6f_n) & \alpha_3 &= \frac{1}{6}(-f_{n-3} + 3f_{n-2} - 3f_{n-1} + f_n) \end{aligned}$$

Substituting the polynomial results in

$$y_{n+1} = y_n + \int_{t_n}^{t_{n+1}} p(t) dt$$

and the *predictor* formula for the fourth-order (**Adams-Bashforth ordinate** form) is

$$y_{n+1}^p = y_n + \frac{h}{24}\{55\dot{y}_n - 59\dot{y}_{n-1} + 37\dot{y}_{n-2} - 9\dot{y}_{n-3}\}$$

We'll discuss *ordinate* notations shortly but note that while back differences are also possible, ordinate forms are more popular. We find the *corrector* part similarly but use the estimated point from the predictor to find \dot{y}_{n+1}^p and discard the $n-3$ value. The fourth-order result (**Adams-Moulton ordinate** form) is

$$y_{n+1} = y_n + \frac{h}{24}\{9\dot{y}_{n+1}^p + 19\dot{y}_n - 5\dot{y}_{n-1} + \dot{y}_{n-2}\}$$

Having predicted and corrected values lets us compare the values and repeat the correction process if necessary. As we increase the degree of the polynomial, accuracy usually rises. However, complexity also rises, and certain orders can produce unstable results. Eighth- and twelfth-degree polynomials are common for many highly-accurate formulations. Maury and Segal (1969) show the derivation and give coefficients for orders 1 to 15 of Adams-Bashforth-Moulton integrators.

Maury and Segal (1969) show how to determine the initial and recurring values of the summed back differences at each step.

$$\hat{S}_n^I = \hat{S}_{n-1}^I + \ddot{\hat{r}}_n \quad \hat{S}_n^{II} = \hat{S}_{n-1}^{II} + \hat{S}_{n-1}^I + \ddot{\hat{r}}_n = \hat{S}_{n-1}^{II} + \hat{S}_n^I$$

The initial values are found from the corrector relations (Adams-Moulton and Cowell), and the previous function evaluations.

$$\hat{S}_{n-1}^I = \frac{1}{h} \ddot{\hat{r}}_{n-1} - \left\{ \frac{1}{2} \ddot{\hat{r}}_{n-1} + \alpha_{AM2} \nabla(\ddot{\hat{r}}_{n-1}) + \alpha_{AM3} \nabla_2(\ddot{\hat{r}}_{n-1}) + \alpha_{AM4} \nabla_3(\ddot{\hat{r}}_{n-1}) + \dots \right\}$$

$$\hat{S}_{n-1}^{II} = \frac{1}{h} \ddot{\hat{r}}_{n-1} - \left\{ \frac{1}{12} \ddot{\hat{r}}_{n-1} + \alpha_{C2} \nabla(\ddot{\hat{r}}_{n-1}) + \alpha_{C3} \nabla_2(\ddot{\hat{r}}_{n-1}) + \alpha_{C4} \nabla_3(\ddot{\hat{r}}_{n-1}) + \dots \right\}$$

Notice that this process will yield large numbers of values for storage, unfortunately, extensive tables aren't efficient for computer implementations, so I'll proceed to alternate forms. Maury and Segal (1969) show that some forms allow us to use only the back values. The resulting formulas are called ***ordinate*** forms and they avoid our maintaining extensive tables of back differences for the function.* I won't list these formulas because the logical extension is to find a ***summed-ordinate*** form that has improved truncation properties and uses only back values. These are the forms that are most useful for astrodynamics. For the Adams-Bashforth-Moulton numerical integrator,

PREDICT - ADAMS-BASHFORTH SUMMED-ORDINATE FORM (order j)

$$\dot{\hat{r}}_{n+1}^P = h \left\{ \alpha_{ABo} \hat{S}_n^I + \sum_{i=0}^j \beta_{ABi} \ddot{\hat{r}}_{n-i} \right\}$$

CORRECT - ADAMS-MOULTON SUMMED-ORDINATE FORM (order j)

$$\dot{\hat{r}}_{n+1} = h \left\{ \alpha_{ABo} \hat{S}_{n+1}^I + \sum_{i=0}^j \beta_{AMi} \ddot{\hat{r}}_{n+1-i} \right\}$$

Notice that this technique evaluates the derivative of the predicted state before doing the correction. For orbit determination with non-conservative systems, we must use the equations once to find velocity, and again to find position. We then form the predicted state evaluation $\dot{\hat{r}}_{n+1}^P$ using those two results, \hat{r}_{n+1}^P and $\dot{\hat{r}}_{n+1}^P$. This predicted value is used when $j = 0$ in the corrector phase. Varying the step-size is more difficult than with single-step methods because equal spacing is implicit in the formulation. This difficulty

* Berry and Healy (2001) show how to derive the ordinate forms by expressing the back difference operator in terms of binomial coefficients of the function points. This lets us write the back difference operator "in terms of the values of the functions."

can be overcome with simple interpolation routines and isn't difficult whenever we have to double the step size (Matthews, 1987:346–437).

Multi-step, Double Integration Techniques

I'll introduce three methods to implement this form of integration—a General Formulation, Störmer-Cowell, and Gauss-Jackson. These methods are sometimes called by different names so you should clearly identify which method is under consideration. I use the predictor-corrector nomenclature presented by Maury and Segal (1969), thus Störmer-Cowell implies a *Störmer predictor*, and a *Cowell corrector*. These techniques differ from single integration methods because they are designed especially for second-order systems. Maury and Segal (1969) present a detailed summary including tables of coefficients. Berry and Healy (2001) and Berry (2004) also give an excellent and expanded discussion. Herrick (1972:12–17, 260–275), Long et al. (1989:6–2 to 6–20), and Baker (1967:452–453) also introduce the formulations. Remember that the formulation is for second order systems—if we include drag, we must compute the first integral, or use a single integration method because the equations of motion depend on position and velocity vectors.

The same forms exist for these routines as before—non-summed and summed, and ordinate, back difference, and divided difference. Krogh's integrator (Krogh, 1974) uses the **General Formulation** in divided difference form. It includes a velocity term so it may not be a “true” double integration technique—also possibly why it is not as well known. It's a non-summed, variable step-size method that's self starting, although fixed step-size versions in back difference or ordinate forms are also possible. **Störmer-Cowell** is another non-summed method, but it has no velocity term. It can be written as a fixed step-size in back difference or ordinate form, and as a variable step-size in divided difference (Berry, 2004). **Gauss-Jackson** is simply the fixed step-size, summed form of Störmer-Cowell. There is no variable step-size divided difference form.

We usually prefer the summed-ordinate formulation for improved round-off error and ease of programming. The Gauss-Jackson method uses first and second, summed back-differences shown in the Adams-Bashforth-Moulton discussion. We use the predicted state in the corrector for $j = 0$.

PREDICT - GAUSS SUMMED-ORDINATE FORM (order j)

$$\hat{\vec{r}}_{n+1}^p = h \left\{ \alpha_{So} \hat{\vec{S}}_n^{II} + \alpha_{S1} \hat{\vec{S}}_n^{I} + \sum_{i=0}^j \beta_{Si} \ddot{\vec{r}}_{n-i} \right\}$$

CORRECT - JACKSON SUMMED-ORDINATE FORM (order j)

$$\hat{\vec{r}}_{n+1} = h \left\{ \alpha_{So} \hat{\vec{S}}_{n+1}^{II} + \{ \alpha_{So} + \alpha_{S1} \} \hat{\vec{S}}_{n+1}^{I} + \sum_{i=0}^j \beta_{Ci} \ddot{\vec{r}}_{n+1-i} \right\}$$

Table 8-1 lists coefficients needed to form different orders of numerical integrators. Consult Maury and Segal (1969) and Berry and Healy (2001) for complete tables.

TABLE 8-1. Numerical Integration Coefficients. I've listed sample coefficients for the Adams-Bashforth-Moulton (AB and AM), Störmer-Cowell (S and C), and Gauss-Jackson (G and J) integrators. The common denominators, D, assist computer implementations for the summed ordinate forms (β_i). The non-summed forms (α_i) may be truncated while the summed ordinate forms are designed for a finite number of terms. (Maury and Segal, 1969)

	Non-summed Forms				Summed Ordinate Forms			
	α_{ABi}	α_{AMi}	α_{Si}	α_{Ci}	β_{ABi}	β_{AMi}	β_{Gi}	β_{Ji}
D					3,628,880	3,628,880	3,628,880	3,628,880
0	1	1	1	1	10,468,447	1,070,017	2,137,435	229,124
1	1/2	-1/2	0	-1	-32,656,759	1,904,811	-6,996,434	304,443
2	5/12	-1/12	1/12	1/12	59,220,027	-2,696,283	13,508,829	-580,962
3	3/8	-1/24	1/12	0	-77,035,235	2,899,075	-6,555,520	677,885
4	251/720	-19/720	19/240	-1/240	60,933,245	-2,134,045	13,080,805	-516,840
5	95/288	-3/160	3/40	-1/240	-30,239,397	1,012,293	-6,485,130	249,861
6	19,087/ 60,480	-863/ 60,480	863/ 12,096	-221/ 60,480	8,594,089	-278,921	1,841,539	-68,658
7	5,257/ 17,280	-275/ 24,192	275/ 4,032	-19/ 6,048	-1,070,017	33,953	-229,124	8,547
8	1,070,017/ 3,628,800	-33,953/ 3,628,800	33,953/ 518,400	-9,829/ 3,628,800				

8.5.1 Implementing an Integrator and Determining Step Size

The advantages of a numerical approach may seem to exclude other solution techniques, but numerical analysis doesn't reveal as much of the problem's qualitative behavior. Also remember the inherent difficulties of computer runtime, and perhaps less discussed, the choice of integrator and step-size. The *type* of orbit integrator can drastically affect performance and precision.

The first-order equations in Eq. (8-7) allow us to easily implement a single-step method. Multi-step techniques require additional explanation. Suppose we want to numerically integrate a complex perturbation model like the one described in Sec. 8.7.2. We'll need both the single and double integration routines for a complete solution. First, we use the Adams-Bashforth and Störmer integrators to find the velocity and position, respectively, from the initial back values of the acceleration. Next, we use the predicted position and velocity vectors to determine the predicted acceleration at the new time. We then use the Adams-Moulton and Cowell integrators to find the velocity and position, respectively, using the predicted acceleration, and the remaining back values.

Finally, if the predicted and corrected values are close enough, the answer is accepted, otherwise, we repeat the corrector step using the new corrected values as inputs.

Finding initial conditions for multi-step methods can be difficult. However, we can often minimize errors by matching the order of all integrators. For instance, if we use an eighth-order, Gauss-Jackson multi-step integrator, we should use an eighth-order Runge-Kutta to determine the back values required by the algorithm. Another approach is to use a lower order integrator, but to also use a smaller step size (Maury and Segal, 1969). The important feature is for the error of the starting method and the multi-step method to be the same.

Once you've chosen an integrator, you'll need to determine the correct step size for methods using fixed step-sizes (shortly, we'll consider variable step sizes). Recall the example in Sec. 8.2 demonstrating special perturbation techniques. Although the numerical approach was eventually able to arrive at a correct answer, the computing process was long, and adjusting the step size was tedious.

Experience has shown that about 100 steps per revolution are satisfactory for standard applications. For eccentric orbits, however, this rule of thumb doesn't hold for fixed step-size integrators. Consider a circular orbit with a period of about 100 minutes. The simple solution suggests using a *Shannon-Sampling Theorem* (Phillips and Nagle, 1984:78) to determine step size. Shannon's theory states that a function with a given frequency, f , is uniquely determined by sampling points which are $1/(2f)$ apart. This is also the *Nyquist critical frequency*, $2f$. Numerical-integration techniques often calculate at five times this rate, so they often encounter a one-tenth factor. Using $1/10$ times the smallest period of oscillation seems reasonable. Yet, for an orbit with a 100-minute period using a 50×50 gravity field, the cyclical oscillations occur every two minutes ($100/50$). A step size of 12 seconds may be far too small! Even a typical step size of one minute will yield acceptable, moderate precision but will take 100 steps to propagate one revolution. Data for Landsat 4 and Landsat 6 (McClain, 1990), which are in basically circular orbits (~800 km), indicates that a one-minute step size yields about a 47 m error. A three-minute step size produces about a 936 m error (average). That's unacceptable for one revolution.

We can determine a fixed step-size for each class of satellites. For example, a satellite in LEO requires a much smaller step size than one in GEO. The particular choice of step size depends on the most rapidly varying component in the function we're integrating and how accurately we want to track it. For most orbital problems, this component is the highest frequency in the disturbing function we want to track.

Now consider an elliptical orbit with an eccentricity of about 0.7 and a period of one day. The same one- to three-minute step size is much too small near apogee because the satellite moves very slowly through gradually changing perturbational forces. Near perigee, however, this same step size may be too large to track the transition between drag and the central-body accelerations.

Variable step sizes are often used for highly elliptical or "difficult" orbits. Although they're accurate, they're not usually very efficient for most orbital problems. That's because many orbits are circular, or nearly circular, and a method using fixed step-sizes is fast and accurate. With variable step sizes, we can spend a lot of time finding a correct

value, usually through trial and error. The resulting values are typically very close to those from approaches using a fixed step-size, so the additional calculations actually hinder the performance for nearly circular orbits. Variable step sizes are available for most popular integrators from Runge-Kutta to the Adams-Bashforth methods. Unfortunately, one integrator isn't the best for *all* orbit types. The complexity of the integrating routine may be an additional hindrance because multi-step methods usually require equally spaced values to form the polynomials. Adjusting the step size can require us to determine additional past values if the new step size doesn't match existing derivatives. For a discussion of other issues, see Sec. 8.8.1.

Another well-known technique is ***regularization***—any approach that eliminates $1/r$ singularities of the motion and replaces time as the independent variable in the integration. Szebehely (1967:93–116) investigated potential applications for the restricted three-body problem. Stiefel and Scheifele (1971:1–16, 19–20, 77–87) and others cover a variety of applications. One approach uses an auxiliary variable, s , to replace time, so $s = \Delta t / r$. This particular scheme adjusts the new “time” variable(s) as the orbit's radius oscillates, which improves its performance. It's possible to show that s is proportional to the eccentric anomaly for an unperturbed orbit.

A very popular technique is known as ***time-regularized Cowell***, also called ***s-integration***,^{*} which causes a controlled variation of the time step by integrating with fixed step-sizes in another variable, such as the eccentric anomaly or true anomaly. From Eq. (2-17) and Eq. (9-17), we form

$$ndt = \frac{r}{a} dE = \frac{r^2}{a\sqrt{1-e^2}} d\nu \quad (8-8)$$

Selecting the best alternate variable to use depends on the size, nature, and frequency of the perturbation. One approach adopts the following time transformation [actually a ***generalized Sundman transformation*** because the index n is not 1 as in Eq. (2-40)].

$$dt = cr^n ds \quad (8-9)$$

Berry and Healy (2002 and 2003b) point out that if the constant $n = 1$ and $c = \sqrt{a/\mu}$ (as Sundman used), s is the eccentric anomaly, and that if $n = 2$ and $c = \sqrt{\mu a(1-e^2)}$, s is the true anomaly. Merson (1975) says that to evenly distribute the integration error throughout the orbit, $n = 1.5$ and $c = 1/\sqrt{\mu}$.

The constant s , the new step size, corresponds to a changing Δt , which produces the smallest time step at perigee and the largest time step at apogee. The price for this automatic time step variation is that we must solve a seventh-order differential equation [Eq. (8-9)]. Because we want the satellite state to be a function of time and because we no longer have equally time-spaced data, this can cause some extra work to obtain interpolated values of the state. Even so, the accuracy and efficiency of a time-regularized Cowell make it very attractive for eccentric orbits.

* Some sources distinguish ***t-integration*** to refer to the [common] situation where time is the independent variable.

8.6 Disturbing Forces

To numerically integrate Cowell's formulation, we must have mathematical models for each perturbing force. These can be analytical formulas, or even tabular representations (e.g., tables of atmospheric density). This section derives the more common disturbing forces, including the accelerations resulting from the central body, drag, third body, solar-radiation pressure, and other smaller forces.

8.6.1 Gravity Field of a Central Body

Given that the gradient of the potential for a spherical central body will yield the acceleration, we must examine how to form a potential function that includes the perturbing accelerations due to a nonspherical central body. The perturbing forces are mathematically described by an *aspherical-potential function*, U , which we'll develop following the methods of Roy (1988:171–177), Kaplan (1976:273–282), and Fitzpatrick (1970:288–308).

Deriving the Aspherical-Potential Function

Because the aspherical-potential function provides the means to determine the gravitational attraction on a satellite, it's logical to assume (from Newton) that a beginning point for the derivation would include the gravitational constant, G and the masses involved. If we examine an infinite number of masses, m_Q , at points Q , the potential per unit mass at point P is the summation of the potential due to each of these points, and the acceleration is ∇U . Therefore,

$$U = G \sum_{Q=1}^{\infty} \frac{m_Q}{\rho_Q}$$

Figure 8-3 shows the geometry.

I'll use a subscript with the mass, m_{\oplus} , to ensure it's not confused with the index m used later in this derivation. The change in potential due to an infinitesimal element of mass, dm_{\oplus} , is

$$dU = G \frac{dm_{\oplus}}{\rho_Q}$$

In reality, a large but finite number of very small masses are within a body of fixed size; thus, the summation of each mass approaches an integral, and integrating over the entire body gives us the total potential.

$$U = G \int_{body} \frac{1}{\rho_Q} dm_{\oplus}$$

The geocentric distance (r) of the point (P) where we're evaluating the potential, and the distance from an infinitesimal mass, dm_{\oplus} , to the center of the Earth, r_Q , are

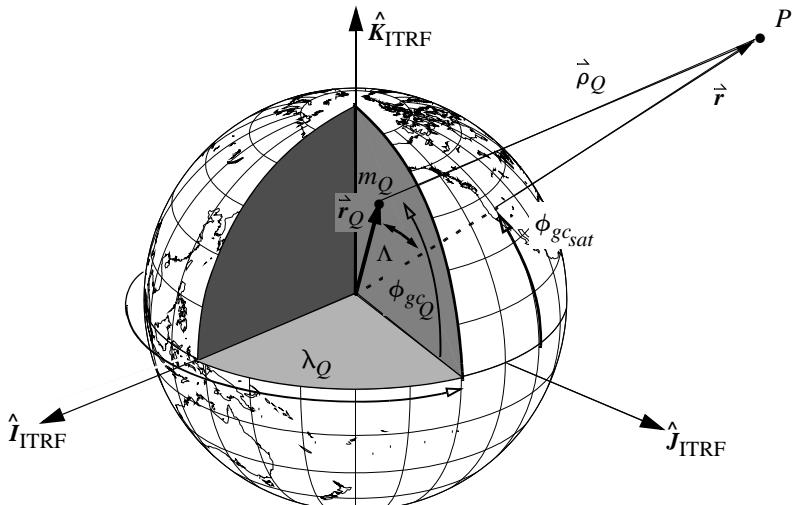


Figure 8-3. Deriving the Gravitational Potential. We consider each point in the Earth, m_Q , when determining the gravitational potential at P . The latitude is a geocentric value. Λ is the angle between the vectors r_Q and r .

$$r = \sqrt{x^2 + y^2 + z^2} \quad r_Q = \sqrt{\xi^2 + \eta^2 + \zeta^2}$$

Using the law of cosines [Eq. (C-21)], the slant range, ρ_Q , from an infinitesimal mass to the external point is

$$\rho_Q^2 = r^2 + r_Q^2 - 2rr_Q \cos(\Lambda)$$

We determine the angle between the two position vectors, Λ , by using the dot product:

$$\cos(\Lambda) = \frac{x\xi + y\eta + z\zeta}{rr_Q} = \frac{\hat{r} \cdot \hat{r}_Q}{rr_Q} \quad (8-10)$$

The slant range is now

$$\begin{aligned} \rho_Q &= r \sqrt{1 - 2 \frac{r_Q}{r} \cos(\Lambda) + \left(\frac{r_Q}{r} \right)^2} = r \sqrt{1 - 2\alpha \cos(\Lambda) + \alpha^2} \\ \alpha &= \frac{r_Q}{r} \end{aligned}$$

The potential (of the body) then becomes

$$U = G \int_{body} \frac{dm_{\oplus}}{r \sqrt{1 - 2\alpha \cos(\Lambda) + \alpha^2}}$$

The parameter (α) will always be less than 1.0 for a point, P , outside the central body ($r > R_{\oplus}$), and the absolute value of the quantity, $\cos(\Lambda)$, will always be less than or equal to 1.0. Thus, we can use the binomial theorem to expand the denominator of the potential in a series:

$$\frac{1}{\sqrt{1 - 2\alpha \cos(\Lambda) + \alpha^2}} = \frac{1}{\sqrt{1+x}} = \sum_{l=0} \alpha^l P_l[\cos(\Lambda)]$$

This expression is a series of *Legendre polynomials*, P_l . I'll use brackets to distinguish the argument of the Legendre polynomials, $\cos(\Lambda)$. Setting $\gamma = \cos(\Lambda)$, a form of Rodriguez's formula gives the *conventional* Legendre polynomials (Lambeck, 1988:10),

$$P_l[\gamma] = \frac{1}{2^l l!} \frac{d^l(\gamma^2 - 1)^l}{d\gamma^l}$$

$$P_l[\gamma] = \frac{1}{2^l} \sum_{j=0}^{\frac{l}{2}} \frac{(-1)^j (2l-2j)!}{j!(l-j)!(l-2j)!} \gamma^{l-2j} \quad (8-11)$$

where $d^l() / d\gamma^l$ is the l^{th} order derivative. The first six Legendre polynomials are

$$P_0[\gamma] = 1 \qquad P_3[\gamma] = \frac{1}{2}(5\gamma^3 - 3\gamma)$$

$$P_1[\gamma] = \gamma \qquad P_4[\gamma] = \frac{1}{8}(35\gamma^4 - 30\gamma^2 + 3)$$

$$P_2[\gamma] = \frac{1}{2}(3\gamma^2 - 1) \qquad P_5[\gamma] = \frac{1}{8}(63\gamma^5 - 70\gamma^3 + 15\gamma)$$

Notice I've factored the leading coefficients of the Legendre polynomials from those usually in the literature to make them computationally efficient. We'll explore recurrence relations to find additional polynomials in Eq. (8-57).

For now, let's write the potential as

$$U = \frac{G}{r} \int_{\text{body}} \sum_{l=0} \alpha^l P_l[\cos(\Lambda)] dm_{\oplus} \quad (8-12)$$

To continue the analysis we can use two methods: MacCullagh's (1855) geometric method and a technique for spherical-harmonic expansion. I'll introduce MacCullagh's technique because it eventually leads to a formula for moments of inertia. It's useful for certain perturbation analyses and in forming spherical-harmonic coefficients of other planets and moons. It also gives a more intuitive feel for gravity and inertia. The spherical-harmonic method may be easier to follow for an introduction to this topic.

With MacCullagh's approach, we express the potential in Eq. (8-12) as

$$U = U_0 + U_1 + U_2 + \dots$$

This permits us to evaluate each component. Let's find the first term using the first Legendre polynomial, P_0 , so that

$$U_0 = \frac{G}{r} \int dm_{\oplus} = \frac{Gm_{\oplus}}{r} = \frac{\mu}{r}$$

This is the two-body potential. Indeed, if we assume the central body is spherically symmetric and homogeneous, the potential reduces to the above expression.

The second term is slightly more difficult. If we use the Legendre polynomial terms and let $\ell = 1$ in Eq. (8-12), we can use Eq. (8-10) to find

$$\begin{aligned} U_1 &= \frac{G}{r} \int \cos(\Lambda) \alpha dm_{\oplus} = \frac{G}{r} \int \frac{x\xi + y\eta + z\zeta}{r^2} dm_{\oplus} \\ &= \frac{G}{r^3} \left(x \int \xi dm_{\oplus} + y \int \eta dm_{\oplus} + z \int \zeta dm_{\oplus} \right) \end{aligned}$$

Each integral (for ξ , η , and ζ) represents the component of the potential to account for a displacement of the center of mass from the origin of the coordinate frame. Even if the origin coincides with the geometric center of the body, it may not coincide with the center of mass. However, if the origin coincides with the center of mass,

$$\int \xi dm_{\oplus} = \int \eta dm_{\oplus} = \int \zeta dm_{\oplus} = 0$$

$U_1 = 0$ because we chose the origin at the center of mass. In fact, the expressions above are a definition of the center of mass.

The third term in the expansion is

$$U_2 = \frac{1}{2r^3} \int (3\gamma^2 - 1) r_Q^2 dm_{\oplus}$$

Remembering that $\gamma^2 = \cos^2(\Lambda)$, we can substitute using Eq. (C-10).

$$U_2 = \frac{G}{2r^3} \int 2r_Q^2 dm_{\oplus} - \frac{G}{2r^3} \int 3r_Q^2 \sin^2(\Lambda) dm_{\oplus}$$

The left-hand term represents the *moments of inertia* about the three coordinate axes:

$$\left. \begin{aligned} A &\equiv \int (\eta^2 + \zeta^2) dm_{\oplus} \\ B &\equiv \int (\xi^2 + \eta^2) dm_{\oplus} \\ C &\equiv \int (\xi^2 + \zeta^2) dm_{\oplus} \end{aligned} \right\} \quad A + B + C = \int 2r_Q^2 dm_{\oplus}$$

The remaining term is the *polar moment of inertia*, I , that refers to any point r_Q :

$$I \equiv \int r_Q^2 \sin^2(\Lambda) dm_{\oplus}$$

These definitions allow us to write $U_2 = \frac{G}{2r^3}(A + B + C - 3I)$. Substituting results of the three cases discussed so far, and assuming the coordinate frame's origin is at the center of mass, gives us the aspherical potential

$$U = \frac{Gm_{\oplus}}{r} + \frac{G}{2r^3}(A + B + C - 3I) + \dots \quad (8-13)$$

This is *MacCullagh's formula* (1855). Evaluating the integrals defining A , B , C , and I presents a practical problem because it requires detailed knowledge of the Earth's mass distribution. Making assumptions about the Earth's size and density distribution allows some progress. Some of the simplest assumptions are the ellipsoidal Earth (oblate spheroid) with uniform density or the prolate spheroid and triaxial ellipsoid. These assumptions aren't very realistic for representing the Earth's potential, but they're a first step if other knowledge is absent.

The geometric approach examines Eq. (8-12). The potential in the form of Eq. (8-12) isn't terribly useful because we can't directly find the angle, Λ .^{*} Spherical trigonometry allows us to develop an equation for this angle. Remember that the satellite's latitude [from Eq. (3-36)], will always be a *geocentric* value (ϕ_{gcsat}). The cosine law of spherical trigonometry yields

$$\begin{aligned} \cos(\Lambda) &= \cos(90^\circ - \phi_{gc_Q}) \cos(90^\circ - \phi_{gcsat}) \\ &\quad + \sin(90^\circ - \phi_{gc_Q}) \sin(90^\circ - \phi_{gcsat}) \cos(\lambda_Q - \lambda_{sat}) \end{aligned}$$

Reduction gives us

$$\cos(\Lambda) = \sin(\phi_{gc_Q}) \sin(\phi_{gcsat}) + \cos(\phi_{gc_Q}) \cos(\phi_{gcsat}) \cos(\lambda_Q - \lambda_{sat}) \quad (8-14)$$

The *addition theorem* (also called the *decomposition formula*) of spherical harmonics provides a way to expand the expressions for the Λ into Eq. (8-12) (Hobson, [1931] 1965:143):

$$\begin{aligned} P_l[\cos(\Lambda)] &= P_l[\sin(\phi_{gc_Q})]P_l[\sin(\phi_{gcsat})] \\ &\quad + 2 \sum_{m=1}^l \frac{(\ell-m)!}{(\ell+m)!} \{ A_{\ell,m} A'_{\ell,m} + B_{\ell,m} B'_{\ell,m} \} \end{aligned} \quad (8-15)$$

$$\begin{aligned} A_{\ell,m} &= P_{\ell,m}[\sin(\phi_{gc_Q})] \cos(m\lambda_Q) & A'_{\ell,m} &= P_{\ell,m}[\sin(\phi_{gcsat})] \cos(m\lambda_{sat}) \\ B_{\ell,m} &= P_{\ell,m}[\sin(\phi_{gc_Q})] \sin(m\lambda_Q) & B'_{\ell,m} &= P_{\ell,m}[\sin(\phi_{gcsat})] \sin(m\lambda_{sat}) \end{aligned}$$

* This angle is often used in Chap. 11, so we'll call it the ground range, or the total range angle, Λ .

where the summation introduces “ l ” and “ m ” indices as *degree* and *order*, respectively.

Now, let's introduce the *associated* Legendre functions, $P_{l,m}$. They have the general form given by Lambeck (1988:10):

$$P_{l,m}[\gamma] = \frac{1}{2^l l!} (1 - \gamma^2)^{m/2} \frac{d^{l+m}}{d\gamma^{l+m}} (\gamma^2 - 1)^l$$

or

$$P_{l,m}[\gamma] = (1 - \gamma^2)^{m/2} \frac{d^m}{d\gamma^m} P_l[\gamma] \quad (8-16)$$

Table 8-2 shows a few examples. Notice that for zero order ($m = 0$), the associated Legendre functions are simply the conventional Legendre polynomials. Practical computation of these functions is handled by recursion, which I'll discuss in Sec. 8.7.2.

TABLE 8-2. Associated Legendre Functions. This table gives a few sample expansions for the associated Legendre function, $P_{l,m}[\sin(\phi_{gc_{sat}})]$. For astrodynamics, we use the geocentric latitude. (Lambeck, 1988)

$P_{0,0}$	1	$P_{3,2}$	$15\cos^2(\phi_{gc_{sat}})\sin(\phi_{gc_{sat}})$
$P_{1,0}$	$\sin(\phi_{gc_{sat}})$	$P_{3,3}$	$15\cos^3(\phi_{gc_{sat}})$
$P_{1,1}$	$\cos(\phi_{gc_{sat}})$	$P_{4,0}$	$\frac{1}{8}\{35\sin^4(\phi_{gc_{sat}}) - 30\sin^2(\phi_{gc_{sat}}) + 3\}$
$P_{2,0}$	$\frac{1}{2}\{3\sin^2(\phi_{gc_{sat}}) - 1\}$	$P_{4,1}$	$\frac{5}{2}\cos(\phi_{gc_{sat}})\{7\sin^3(\phi_{gc_{sat}}) - 3\sin(\phi_{gc_{sat}})\}$
$P_{2,1}$	$3\sin(\phi_{gc_{sat}})\cos(\phi_{gc_{sat}})$	$P_{4,2}$	$\frac{15}{2}\cos^2(\phi_{gc_{sat}})\{7\sin^2(\phi_{gc_{sat}}) - 1\}$
$P_{2,2}$	$3\cos^2(\phi_{gc_{sat}})$	$P_{4,3}$	$105\cos^3(\phi_{gc_{sat}})\sin(\phi_{gc_{sat}})$
$P_{3,0}$	$\frac{1}{2}\{5\sin^3(\phi_{gc_{sat}}) - 3\sin(\phi_{gc_{sat}})\}$	$P_{4,4}$	$105\cos^4(\phi_{gc_{sat}})$
$P_{3,1}$	$\frac{1}{2}\cos(\phi_{gc_{sat}})\{15\sin^2(\phi_{gc_{sat}}) - 3\}$		

Now we make an important substitution to Eq. (8-15). By separating all the terms that are independent of the satellite's location in Eq. (8-12), we can arrive at a solution that isolates terms which depend only on the central body and those which relate the satellite's position. We define variables in Eq. (8-17).

$$\begin{aligned} C'_{l,m} &= \int_{\text{body}} r_Q^l \frac{(l-m)!}{(l+m)!} P_{l,m}[\sin(\phi_{gc_Q})] \cos(m\lambda_Q) dm \oplus \\ S'_{l,m} &= \int_{\text{body}} r_Q^l \frac{(l-m)!}{(l+m)!} P_{l,m}[\sin(\phi_{gc_Q})] \sin(m\lambda_Q) dm \oplus \end{aligned} \quad (8-17)$$

The coefficients in Eq. (8-17) represent the mathematical modeling for the Earth's shape using spherical harmonics. The special case for the *zonal harmonics* is

$$C'_{l,0} = \int_{\text{body}} r_Q^l P_l[\sin(\phi_{gc_Q})] dm \oplus$$

which uses the conventional Legendre polynomials. Note that $S'_{l,0}$ is zero. When we substitute these values into the potential in Eq. (8-12),

$$\begin{aligned} U &= \frac{G}{r} \sum_{l=0}^{\infty} \frac{P_l[\sin(\phi_{gc_{sat}})]}{r^l} C'_{l,0} \\ &+ \frac{G}{r} \sum_{l=1}^{\infty} \sum_{m=1}^l \frac{P_{l,m}[\sin(\phi_{gc_{sat}})]}{r^l} \left\{ C'_{l,m} \cos(m\lambda_{sat}) + S'_{l,m} \sin(m\lambda_{sat}) \right\} \end{aligned}$$

Finally, a unit analysis suggests we remove units from the result. This leads to the C and S coefficients. We can non-dimensionalize the gravitational coefficients (Lambeck, 1988:11):

$$C'_{l,m} = C_{l,m} R_{\oplus}^l m \oplus \quad S'_{l,m} = S_{l,m} R_{\oplus}^l m \oplus$$

R_{\oplus} and m_{\oplus} are the Earth's mean equatorial radius and mass, respectively.

Using the definition of the gravitational parameter, find the gravitational potential as

$$\begin{aligned} U &= \frac{\mu}{r} \sum_{l=0}^{\infty} P_l[\sin(\phi_{gc_{sat}})] \left(\frac{R_{\oplus}}{r} \right)^l C_{l0} \\ &+ \frac{\mu}{r} \sum_{l=1}^{\infty} \sum_{m=1}^l P_{l,m}[\sin(\phi_{gc_{sat}})] \left(\frac{R_{\oplus}}{r} \right)^l \left\{ C_{l,m} \cos(m\lambda_{sat}) + S_{l,m} \sin(m\lambda_{sat}) \right\} \end{aligned} \quad (8-18)$$

I'll present several of the common notations for the aspherical potential. The International Astronomical Union (1961) has adopted these forms, which you can find in Lambeck (1988:11), Kaplan (1976:281), and others. First, a double summation encompasses the associated Legendre polynomials, $P_{l,m}$, and the harmonic coefficients, and $S_{l,0} = 0$.

$$U = \frac{\mu}{r} \sum_{l=0}^{\infty} \sum_{m=0}^l \left(\frac{R_{\oplus}}{r}\right)^l P_{l,m} [\sin(\phi_{gc_{sat}})] \left\{ C_{l,m} \cos(m\lambda_{sat}) + S_{l,m} \sin(m\lambda_{sat}) \right\}$$

Notice that both summation indices begin at zero. Recall, as part of the derivation, if the center of the coordinate system coincides with the attracting body's center of mass, the coefficients $C_{1,0}$, $C_{1,1}$, and $S_{1,1}$ are all zero ($S_{1,0}$ is also zero by definition). This result corresponds to $l=1$ in MacCullagh's approach. It leads to another very common form of this relation, which separates the spherical potential and therefore requires us to adjust the summation limits. The 0th term is within the μ/r term, and the first-degree terms are zero.

$$U = \frac{\mu}{r} \left[1 + \sum_{l=2}^{\infty} \sum_{m=0}^l \left(\frac{R_{\oplus}}{r}\right)^l P_{l,m} [\sin(\phi_{gc_{sat}})] \left\{ C_{l,m} \cos(m\lambda_{sat}) + S_{l,m} \sin(m\lambda_{sat}) \right\} \right] \quad (8-19)$$

This expression describes the gravitational attraction resulting from the irregular distribution of the Earth's mass using a potential function.

Some authors use a "J" notation for the zonal harmonics. The correct use of the "J" notation is that it's the negative of the C notation for zonal harmonics.

$$-C_{l,0} = J_l \quad (8-20)$$

Though this is merely a sign change, it can be very significant in forming the perturbation equations discussed in Sec. 9.6.

We can also separate the zonal and tesseral terms from Eq. (8-19) by recalling that $J_l = -C_{l,0}$ [Eq. (8-20)]. Thus,

$$U = \frac{\mu}{r} \left[1 - \sum_{l=2}^{\infty} J_l \left(\frac{R_{\oplus}}{r}\right)^l P_l [\sin(\phi_{gc_{sat}})] + \sum_{l=2}^{\infty} \sum_{m=1}^l \left(\frac{R_{\oplus}}{r}\right)^l P_{l,m} [\sin(\phi_{gc_{sat}})] \left\{ C_{l,m} \cos(m\lambda_{sat}) + S_{l,m} \sin(m\lambda_{sat}) \right\} \right] \quad (8-21)$$

This form has several important features. The *conventional* Legendre polynomials are merely special cases of the associated Legendre functions. Eq. (8-21) also depends on the C and S coefficients. Fortunately, the analysis of satellite motion allows us to empirically determine the C and S coefficients from observations. This is a common method in developing gravitational fields today.

Older computers sometimes used normalized coefficients because the conventional gravitational coefficients become very small as the degree and order (l, m) get large. This is a consequence of the factorial ratios appearing in Eq. (8-17). Computer truncation

tions may introduce errors so it's common to normalize in order to correct this deficiency. The following formula permits these transformations. The overall transformation, $\Pi_{l,m}$, is often used to abbreviate the relation. The overbar denotes normalized variables. Due to the size of the coefficients, $\Pi_{l,m}$ is usually determined with the maximum precision available on the computer. Don't confuse this process with non-dimensionalizing described earlier.

$$\begin{aligned}\Pi_{l,m} &= \sqrt{\frac{(l+m)!}{(l-m)!k(2l+1)}} \\ \bar{S}_{l,m} &= \Pi_{l,m} S_{l,m} \quad \bar{C}_{l,m} = \Pi_{l,m} C_{l,m} \quad \bar{P}_{l,m} = \frac{P_{l,m}}{\Pi_{l,m}} \\ k &= 1 \text{ if } m = 0 \\ k &= 2 \text{ if } m \neq 0\end{aligned}\tag{8-22}$$

When using normalized coefficients, you must also normalize the Legendre polynomials (conventional and associated) because the product must remain the same. Thus,

$$\bar{C}_{l,m} \bar{P}_{l,m} = C_{l,m} P_{l,m} \quad \bar{S}_{l,m} \bar{P}_{l,m} = S_{l,m} P_{l,m}$$

Note that most modern floating point computers don't require normalized values and can accurately process unnormalized coefficients. Normalizing requires an extra step in each calculation that *can* be eliminated. Also, the summation should always begin with the smallest coefficient, proceeding to the largest (as in the reduction calculations in Ch. 3) to preserve accuracy. An example will show how normalizing works.

▼ Example 8-3. Normalizing Gravitational Coefficients.

GIVEN: Gravitational coefficients in Table D-1 and Table D-2.

FIND: Normalized values for (4,0), (41,1), (70,65).

For $l = 4$ and $m = 0$,

$$\begin{aligned}\bar{C}_{4,0} &= \sqrt{\frac{(l+m)!}{(l-m)!k(2l+1)}} C_{4,0} = \sqrt{\frac{(4+0)!}{(4-0)!1(2(4)+1)}} 1.619\ 621\ 591 \times 10^{-6} \\ &= 5.398\ 738\ 638 \times 10^{-7}\end{aligned}$$

Notice that $k = 1$ and S don't exist for zonal harmonics ($m = 0$). For $l = 41$ and $m = 1$,

$$\begin{aligned}\bar{C}_{41,1} &= \sqrt{\frac{(l+m)!}{(l-m)!k(2l+1)}} C_{41,1} = \sqrt{\frac{(41+1)!}{(41-1)!2(2(41)+1)}} (-1.522\ 218\ 288 \times 10^{-9}) \\ &= -4.902\ 746\ 800 \times 10^{-9} \\ \bar{S}_{41,1} &= \sqrt{\frac{(l+m)!}{(l-m)!k(2l+1)}} S_{41,1} = \sqrt{\frac{(41+1)!}{(41-1)!2(2(41)+1)}} (-1.283\ 602\ 318 \times 10^{-9}) \\ &= -4.134\ 214\ 655 \times 10^{-9}\end{aligned}$$

Notice now that $k = 2$ because m isn't zero. For $l = 70$ and $m = 65$,

$$\begin{aligned}\bar{C}_{l,m} &= \sqrt{\frac{(l+m)!}{(l-m)!k(2l+1)}} C_{l,m} = \sqrt{\frac{(70+65)!}{(70-65)!2(2(70)+1)}} 1.431\ 238\ 529 \times 10^{-122} \\ &= 1.019\ 277\ 424 \times 10^{-9} \\ \bar{S}_{l,m} &= \sqrt{\frac{(l+m)!}{(l-m)!k(2l+1)}} S_{l,m} = \sqrt{\frac{(70+65)!}{(70-65)!2(2(70)+1)}} (1.951\ 368\ 590 \times 10^{-122}) \\ &= 1.775\ 623\ 037 \times 10^{-9}\end{aligned}$$

▲ Notice the coefficients' similar orders of magnitude. The normalization process also benefits the terms of higher degree and order.

We can identify the term which gives us an idea of the Earth's overall nonsphericity. Specifically, the ***normalized, second-degree, zonal gravitational coefficient***, $\bar{C}_{2,0}$, represents the Earth's equatorial bulge. Eq. (8-22) discusses normalization. The remaining parameters help represent the geopotential due to the central body's specific shape.

$$\bar{C}_{2,0} = -484.165\ 466\ 3 \times 10^{-6}$$

This value doesn't include the permanent tidal deformation (Sec. 8.6.5 and Sec. 9.5.3). The conventional (un-normalized) form of this equation uses the following relation:

$$C_{l,0} = \bar{C}_{l,0} \sqrt{2l+1} \quad (8-23)$$

This is simply a special case of general formulas. Thus the first zonal coefficient is (with a second-order zonal harmonic of $J_2 = 0.001\ 082\ 626\ 9$)

$$C_{2,0} = -0.001\ 082\ 626\ 9$$

Spherical Harmonics

The trigonometric argument of the Legendre polynomials in Eq. (8-19) constitutes surface **spherical harmonics**, for they are periodic on the surface of a unit sphere. When the surface's spherical harmonics are divided by r^{l+1} , they're usually called solid spherical harmonics. The Sturm-Liouville theorem states that the solid spherical harmonics are eigenfunctions that constitute an independent basis for the gravitational model. In essence, they are a Fourier series. The indices l and m determine lines on the sphere along which the functions vanish. These spherical harmonics further break down into three types of terms—zonal, sectoral, and tesseral harmonics—which I'll introduce in this section. Later in this chapter, we'll see how they affect orbital elements.

Zonal harmonics are defined by zeroth order ($m = 0$), where the dependence of the potential on longitude vanishes and the field is symmetrical about the polar axis. These are simply bands of latitude. For any $P_l[\sin(\phi_{gcsat})]$, there are l circles of latitude along which P_l is zero, and hence $(l+1)$ zones in which the function is alternately increasing (shaded) and decreasing (white). J_2 is by far the strongest perturbation due to the Earth's shape. As the table on the inside back cover shows, J_2 is almost 1000 times larger than the next largest coefficient (J_3). Each of the boundaries of the zonal harmonics represents a root of the Legendre polynomial. They are a transition between positive and negative values. Figure 8-4 shows several depictions of the zonal harmonics.

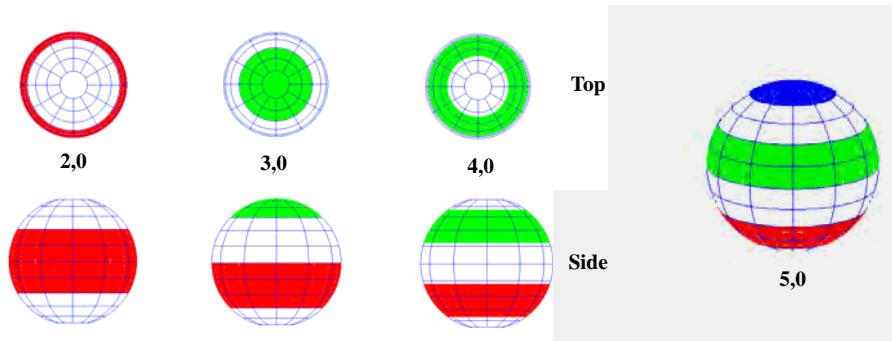


Figure 8-4. **Zonal Harmonics.** J_2 accounts for most of the Earth's gravitational departure from a perfect sphere. This band (and others) reflects the Earth's oblateness. The shading indicates regions of additional mass and the numbers link regions between the views. The third harmonic appears similar to the J_2 from the top but is reversed for the bottom view. J_5 is shown in a perspective view.

Sectoral harmonics occur when $l = m$ and represent bands of longitude. Figure 8-5 shows some sectoral harmonics. The polynomials $P_{l,l}[\sin(\phi_{gcsat})]$ are zero only at the poles ($\phi_{gcsat} = \pm 90^\circ$). On the other hand, the terms [$\cos(l\lambda)$ and $\sin(l\lambda)$] are zero for $2l$ different values of λ . Hence the lines along which the functions [$\cos(l\lambda)$ and $\sin(l\lambda)$] times $P_{l,l}[\sin(\phi_{gcsat})]$ vanish are meridians of longitude, which divide the sphere into $2l$ “orange-slice” sectors. Each sector represents l positive (shaded) and l negative (white) mass concentrations.

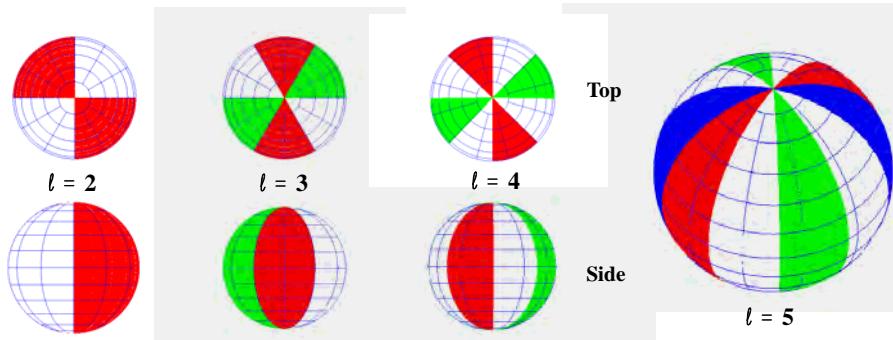


Figure 8-5. **Sectoral Harmonics.** Sectoral harmonics take into account the extra mass distribution in longitudinal regions.

For cases in which $l \neq m \neq 0$, the physical depiction is a specific region on the Earth. These functions [$\cos(m\lambda)$ and $\sin(m\lambda)$] times $P_{l,m}[\sin(\phi_{gcsat})]$ are referred to as **telescopical harmonics**.

serial harmonics (because they look like tiles), for the sphere is divided up into a checkeredboard array. The number of circles of latitude along which $P_{l,m}[\sin(\phi_{gcsat})]$ is zero is equal to $(l - m)$, whereas the terms [$\cos(m\lambda)$ and $\sin(m\lambda)$] vanish along $2m$ meridians of longitude. These zero lines represent the center of the latitude/longitude bands. Figure 8-6 shows the tesseral harmonics. Notice the similarities to, and differences from, the sectoral harmonics. Some texts refer to the sectorals as a subset of the tesserales. That's correct, but we can also keep them separate to distinguish their physical differences, as I do in this book.

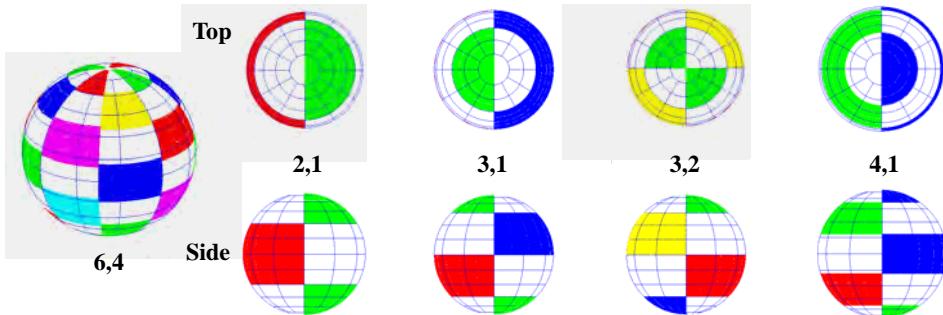


Figure 8-6. Tesseral Harmonics. Tesseral harmonics attempt to model specific regions on the Earth which depart from a perfect sphere. Depicting these and tying to the gravitational coefficients can be difficult.

Partial Derivatives

Now that we've developed a potential function for the central-body effects, we must take the gradient (partial derivatives) to determine the acceleration we can use in numerical propagations. These partial derivatives are also necessary for the variational equations we'll use in Chap. 10. To determine the acceleration, we use the gradient operation. We'll change the notation slightly to reinforce the fact that we're using the satellite's geocentric latitude, ϕ_{gcsat} , and longitude, $\lambda = \lambda_{sat}$. From Long et al. (1989:4–11 to 4–16), using matrix differentiation with $\dot{\vec{r}} = r_I \hat{I} + r_J \hat{J} + r_K \hat{K}$ in the ITRF frame,* we can directly determine the partial derivatives of the aspherical-potential function (including the two-body attraction). We often differentiate the Legendre functions in spherical coordinates $(r, \phi_{gcsat}, \lambda)$, but find the acceleration in cartesian coordinates, so we have to apply the chain rule. The acceleration is

* Remember that we usually evaluate the accelerations in the body-fixed frame (ITRF), but we numerically integrate in an inertial frame (GCRF).

$$\ddot{a} = \frac{\partial U}{\partial r} \left(\frac{\partial r}{\partial \dot{r}} \right)^T + \frac{\partial U}{\partial \phi_{gc_{sat}}} \left(\frac{\partial \phi_{gc_{sat}}}{\partial \dot{r}} \right)^T + \frac{\partial U}{\partial \lambda_{sat}} \left(\frac{\partial \lambda_{sat}}{\partial \dot{r}} \right)^T \quad (8-24)$$

Now, using the nonspherical portion of Eq. (8-19), (notice there is no discontinuity at $\phi_{gc_{sat}} = 90^\circ$)

$$\begin{aligned} \frac{\partial U}{\partial r} &= -\frac{\mu}{r^2} \sum_{\ell=2m=0}^{\infty} \sum_{\ell} \left(\frac{R_\oplus}{r} \right)^\ell (\ell+1) P_{\ell,m} [\sin(\phi_{gc_{sat}})] \\ &\quad \left\{ C_{\ell,m} \cos(m\lambda_{sat}) + S_{\ell,m} \sin(m\lambda_{sat}) \right\} \\ \frac{\partial U}{\partial \phi_{gc_{sat}}} &= \frac{\mu}{r} \sum_{\ell=2m=0}^{\infty} \sum_{\ell} \left(\frac{R_\oplus}{r} \right)^\ell \{ P_{\ell,m+1} [\sin(\phi_{gc_{sat}})] - m \tan(\phi_{gc_{sat}}) P_{\ell,m} [\sin(\phi_{gc_{sat}})] \} \\ &\quad \times \{ C_{\ell,m} \cos(m\lambda_{sat}) + S_{\ell,m} \sin(m\lambda_{sat}) \} \\ \frac{\partial U}{\partial \lambda_{sat}} &= \frac{\mu}{r} \sum_{\ell=2m=0}^{\infty} \sum_{\ell} \left(\frac{R_\oplus}{r} \right)^\ell m P_{\ell,m} [\sin(\phi_{gc_{sat}})] \left\{ S_{\ell,m} \cos(m\lambda_{sat}) - C_{\ell,m} \sin(m\lambda_{sat}) \right\} \end{aligned} \quad (8-25)$$

The derivatives of the position vector also proceed directly as

$$\begin{aligned} \frac{\partial r}{\partial \dot{r}} &= \frac{\dot{r}^T}{r} \\ \frac{\partial \phi_{gc_{sat}}}{\partial \dot{r}} &= \frac{1}{\sqrt{r_I^2 + r_J^2}} \left(-\frac{\dot{r}^T r_K}{r^2} + \frac{\partial r_K}{\partial \dot{r}} \right) \\ \frac{\partial \lambda_{sat}}{\partial \dot{r}} &= \frac{1}{r_I^2 + r_J^2} \left(r_I \frac{\partial r_J}{\partial \dot{r}} - r_J \frac{\partial r_I}{\partial \dot{r}} \right) \end{aligned} \quad (8-26)$$

Notice the three partial derivatives in Eq. (8-26) are unit vectors. Putting the acceleration terms together and breaking out individual components make Eq. (8-24)

$$\begin{aligned} a_I &= \left\{ \frac{1}{r} \frac{\partial U}{\partial r} - \frac{r_K}{r^2 \sqrt{r_I^2 + r_J^2}} \frac{\partial U}{\partial \phi_{gc_{sat}}} \right\} r_I - \left\{ \frac{1}{r_I^2 + r_J^2} \frac{\partial U}{\partial \lambda_{sat}} \right\} r_J - \frac{\mu r}{r^3} \\ a_J &= \left\{ \frac{1}{r} \frac{\partial U}{\partial r} - \frac{r_K}{r^2 \sqrt{r_I^2 + r_J^2}} \frac{\partial U}{\partial \phi_{gc_{sat}}} \right\} r_J + \left\{ \frac{1}{r_I^2 + r_J^2} \frac{\partial U}{\partial \lambda_{sat}} \right\} r_I - \frac{\mu r}{r^3} \\ a_K &= \frac{1}{r} \frac{\partial U}{\partial r} r_K + \frac{\sqrt{r_I^2 + r_J^2}}{r^2} \frac{\partial U}{\partial \phi_{gc_{sat}}} - \frac{\mu r}{r^3} \end{aligned} \quad (8-27)$$

8.6.2 Atmospheric Drag

Next to the oblateness of the Earth, atmospheric drag most strongly influences the motion of a satellite *near Earth*; in fact, during the last few revolutions of the satellite's life, drag effects can be more dominant than those from the Earth's oblateness. For more distant satellites, third-body effects and solar-radiation pressure dominate more than oblateness and drag. Investigations of aerobraking and satellite tethers require accurate atmospheric models. We must also understand the troposphere and the ionosphere to predict the behavior of tracking and communications systems. The idea is to have a good enough model to account for the atmospheric density while accurately modeling the effects of drag on an orbit. The Sun's interaction with the upper atmosphere and the Earth's magnetic field influence drag variations and uncertainty the most. Many references describe the effects and various analytical procedures to model the drag effect [see King-Hele (1987) for example], but models are still incomplete. The study of drag usually divides into three areas: (1) determining orbits under the influence of drag, (2) estimating a satellite's lifetime, and (3) determining physical properties of the upper atmosphere. We'll introduce each in this section.

The cause of drag is the atmospheric particles, which retard the satellite's motion. Calculating density is extremely complex for real-world problems. Drag is a nonconservative perturbation because total energy isn't conserved (loss of energy due to friction). In fact, all velocity-dependent perturbations are nonconservative because they change the total energy. Drag changes mainly the semimajor axis and eccentricity of the orbit. There are periodic effects in the other orbital elements and some coupling aspects with the aspherical potential. We'll examine these ideas shortly.

Studying astrodynamics is most difficult in the atmosphere. To rigorously model the effects of atmospheric perturbation, we must know molecular chemistry, thermodynamics, aerodynamics, hypersonics, meteorology, electromagnetics, planetary sciences, as well as orbital mechanics. Accurately determining atmospheric properties is crucial to determining drag on satellites.

This book considers various atmospheric models which allow you to use varying levels of sophistication in simulation programs. Some of the basic information in this section is from Jablonski (1991, 1992), who has allowed me to reproduce sections of her papers concerning general atmospheric concepts and the Russian GOST model.

Acceleration Due to Drag

The basic equation for aerodynamic drag combines several factors; I'm showing it here as a specific force or acceleration:

$$\hat{\vec{a}}_{drag} = -\frac{1}{2} \frac{c_D A}{m} \rho v_{rel}^2 \frac{\hat{\vec{v}}_{rel}}{|\hat{\vec{v}}_{rel}|} \quad (8-28)$$

The **coefficient of drag**, c_D , is a dimensionless quantity which reflects the satellite's susceptibility to drag forces. The drag coefficient for satellites in the upper atmosphere is often approximately 2.2 (using a flat plate model). Spheres have $c_D \sim 2.0$ to 2.1. The drag coefficient is satellite configuration-specific and is seldom approximated to more

than three significant digits. The **atmospheric density**, ρ , indicates how dense the atmosphere is at the satellite altitude and is perhaps the most difficult parameter to determine. Another difficult parameter to estimate is the exposed **cross-sectional area**, A , defined to be the area which is normal to the satellite's velocity vector. For high-precision studies, we must use attitude determination to help calculate A . For a tumbling reentry vehicle, such as Skylab in 1980, it's nearly impossible to know the attitude accurately, so A is inherently uncertain. We also need the satellite's mass, m , along with the relative-velocity vector. For some satellites, we can't assume mass is constant.

We usually call $m/(c_D A)$ the **ballistic coefficient**, BC; it's another measure of a satellite's susceptibility to drag effects. With this definition, a low BC means drag will affect the satellite a lot—and vice versa. Note that *many* organizations use the reciprocal $(c_D A)/m$, so it's very important to understand which one is in use.

The velocity is *not* the velocity vector typically found in the state vector. This velocity vector is relative to the atmosphere, hence the subscript. In actuality, the Earth's atmosphere has a mean motion due to the Earth's rotation, and the winds are superimposed on this mean motion. Notice also that the force of drag opposes the velocity vector at all times. This is a primary use for the NTW coordinate system. For a nonspherical satellite, we must also consider companion aerodynamic forces such as lift and side forces. Remember that although the atmosphere is rotating, it does so with a "profile" that follows a little behind the Earth. Due to friction with the Earth, the atmosphere closest to the Earth rotates a little faster than higher altitudes. The velocity vector relative to the rotating atmosphere is

$$\vec{v}_{rel} = \frac{d\vec{r}}{dt} - \vec{\omega}_{\oplus} \times \vec{r} = \begin{bmatrix} \frac{dx}{dt} + \omega_{\oplus} y & \frac{dy}{dt} - \omega_{\oplus} x & \frac{dz}{dt} \end{bmatrix}^T$$

Escobal ([1965] 1985:60) gives a more general expression including wind variations:^{*}

$$\vec{v}_{rel} = \begin{bmatrix} \frac{dx}{dt} + \omega_{\oplus} y + v_w \{-\cos(\alpha)\sin(\delta)\cos(\beta_w) - \sin(\alpha)\sin(\beta_w)\} \\ \frac{dy}{dt} - \omega_{\oplus} x + v_w \{-\sin(\alpha)\sin(\delta)\cos(\beta_w) + \cos(\alpha)\sin(\beta_w)\} \\ \frac{dz}{dt} + v_w \{ \cos(\delta)\cos(\beta_w) \} \end{bmatrix}$$

which requires the wind's speed, v_w , and azimuth, β_w , and the satellite's right ascension and declination. But many applications don't use this expression because the additional information usually isn't available. Rather, the satellite's specific orientation and shape are determined to help determine the satellite's effective cross-sectional area.

* The signs differ from Escobal because this equation finds the contribution of the wind using v_w and β_w in the SEZ system, and then rotates to the geocentric system $\{\text{ROT3}(-\alpha) \text{ROT2}(-(90-\delta))\}$.

Determining Density

The density of the upper atmosphere changes because of a complex interaction between three basic parameters: the nature of the atmosphere's molecular structure, the incident solar flux, and geomagnetic (auroral) interactions. Molecular composition strongly affects the density, so we'll look at it shortly. **Solar flux**, or the incident radiation arriving from the Sun, affects the atmospheric density through nearly instantaneous heating from *Extreme Ultraviolet Radiation* (EUV or F_{EUV}). Geomagnetic activity affects the atmosphere through delayed heating of atmospheric particles from collisions with charged energetic particles from the Sun. Together, these effects increase atmospheric density at higher altitudes by increasing particle collisions. The level of solar flux and geomagnetic activity are difficult to predict, but very important for precise models.

The gravitational attraction of molecules in the atmosphere mainly determines its pressure and density. The development of both the static and time-varying models relies on a few basic hydrostatic principles which model atmospheric effects. The ***ideal-gas law*** relates the *absolute pressure*, p_o , the *mean molecular mass of all atmospheric constituents*, M , the *acceleration due to gravity*, g_o , the *universal gas constant*, R , and the *absolute temperature*, T (Kelvin):

$$\rho = \frac{p_o M}{g_o R T} \quad (8-29)$$

The linkage with temperature is important because it causes much of the difficulty in determining an exact model for the density. The Earth's rotation exposes the atmosphere to the Sun, and the resulting solar heating affects density.

The second relation is the ***hydrostatic equation***, which relates the change in pressure, Δp , to the density, gravity, and change in altitude, Δh :

$$\Delta p = -\rho g \Delta h \quad (8-30)$$

I won't develop the density equations and associated relations here; see Baker and Makemson (1967:210–213) and Escobal ([1968] 1979:18–25) for further information.

To gain an idea of the importance of modeling atmospheric parameters for astrodynamics, consider Fig. 8-7. The complexity of the atmosphere is apparent from the sheer number of regimes. In addition, although values are shown for temperature and altitude, they all change over time and are very difficult to predict.

Overall, atmospheric models are either *static* or *time-varying*. Of course, the ***static*** models are the simplest to use because we assume all the atmospheric parameters remain constant. Yet, some factors affect even static models:

1. **Latitudinal variations:** These effects are easiest to visualize with a satellite in a circular, inclined orbit. Passing over the Earth's equatorial bulge effectively changes the actual altitude and, therefore, the density. This changes the drag at virtually every point on the orbit.
2. **Longitudinal variations:** Although we consider these effects mainly with time-varying models [specifically the diurnal (daily) effect], they can also influence static models. Consider the difficulty in determining a

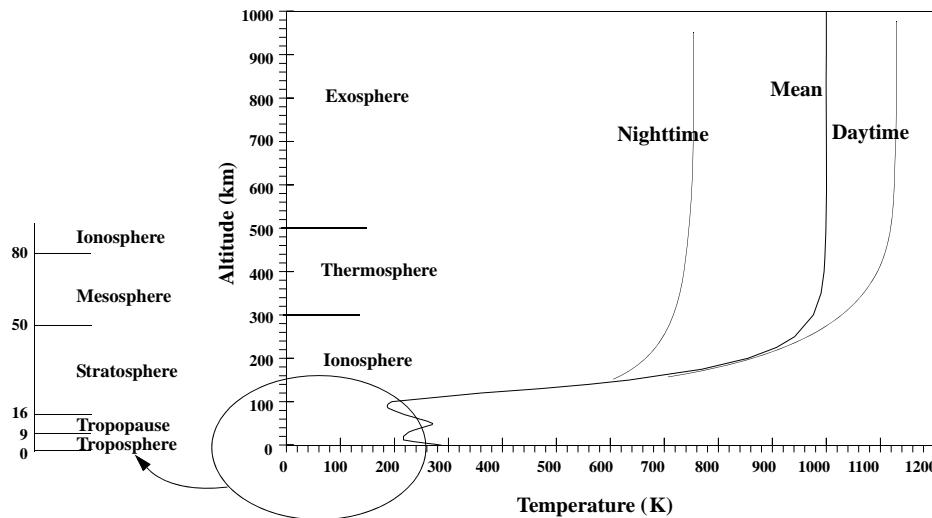


Figure 8-7. Earth’s Atmosphere. The Earth’s atmosphere is very complex. Notice how rapidly the temperature changes. The bands shown above are general areas where the atmosphere has similar properties. Use this figure carefully—values can change drastically over time.

completely symmetrical atmosphere to conform to every region on the Earth. Mountain ranges in the Himalayas cause significant differences in the wind, density, and temperature from the vast expanses of the open ocean at the same latitude but different longitudes.

Additional complexities appear in the *time-varying* models as we consider more real-world effects. For example, we must model

1. **Diurnal variations:** These variations occur every day as the Earth rotates. A pronounced atmospheric bulge (and therefore a density maximum) lags the general direction of the Sun, where the atmosphere is warmest. The bulge centers on meridians where the local time is 2:00–2:30 P.M. It centers on the equator at the equinoxes but moves to higher latitudes depending on the Sun’s declination, which varies throughout the year. Because of the diurnal bulge, the atmospheric density depends on latitude, local time, and time of year (de Lafontaine, 1986). A minimum value also occurs approximately opposite the bulge at 4:00 A.M. or so each day.
2. **27-day solar-rotation cycle:** This effect comes from the Sun’s rotational period and causes a fluctuation which is correlated with the solar decimetric-wavelength radio flux. Irregular changes to the solar flux are related to the growth and decay of active solar regions. These active regions not only have many different patterns of growth, stability, and decay but also cause

an uncertain cyclical pattern because they rotate with the Sun over its 27-day period. These patterns are difficult to predict, and it's also difficult to determine how they heat the Earth's atmosphere. This uncertainty affects the accuracy of solar-flux predictions (Nostrand, 1984).

3. **11-year cycle of Sun spots:** The Sun-spot cycle, or *Solar Cycle** strongly varies the amount of incoming solar radiation that reaches Earth. At the height of the cycle, this effect can create a larger disturbance through solar-radiation pressure than through drag, depending on the satellite's altitude. Recent cycles have experienced solar minima in March 1976, July 1986, August 1996, and November 2008 (Hathaway, 2010).
4. **Semi-annual / Seasonal variations:** These variations last about six months and are related to the varying distance of the Earth from the Sun and the Sun's declination during the year. They're usually small.
5. **Cyclical variations:** An 11-year cycle roughly parallels but lags a few years behind the Sun-spot cycle. The minimum of this cycle isn't halfway between the two maxima because the recovery from a maximum is about 6–7 years. The magnitude of the actual peak of the 11-year cycle varies from cycle to cycle. The exact cause is unknown but likely is related to sunspot activity.
6. **Rotating atmosphere:** The atmosphere rotates with the Earth to some extent, causing additional time-varying changes in the density. The velocity is usually larger closer to the Earth's surface due to increased friction.
7. **Winds:** Accounting for atmospheric weather patterns and the effect of atmospheric winds on the density is very difficult and extremely complex. They cause temperature variations and therefore changes in density. The effect of actual weather systems is less well known. Much work remains if we are to fully understand the dynamics of the upper atmosphere.
8. **Magnetic-storm variations:** Usually, fluctuations in the Earth's magnetic field slightly affect the atmosphere, but the effect can be large when geomagnetic activity increases.
9. **Irregular short-periodic variations:** These effects are also small and are often associated with transient geomagnetic disturbances. Included in this section are random solar flares and other very small effects, such as variation in hydrogen currents within the atmosphere.
10. **Tides:** Ocean tides (and even atmospheric tides) also cause very small variations in the atmospheric density.

* The solar cycle is based primarily on the number of sunspots and there are Boulder and International Sunspot numbers (and some others). The Wolf 1848 formula is used by both groups, but the observational data is somewhat different. As a result, the dates of the minimums do not always agree.

Magnetic Field Models

The effect of the magnetic variations of the Earth and Sun is useful with calculations of atmospheric density because it's believed that magnetic variations are related to fluctuations in atmospheric density. The Earth's magnetic field also has a measurable (but small) influence on satellite operations. The effect appears in four ways: (1) the charged particles from any magnetic disturbances cause ionization in the upper atmosphere, thereby affecting the density and, subsequently, the drag; (2) the charges on the particles can actually alter the attractive forces experienced by the satellite; (3) the ionization interferes with satellite tracking and communication; and (4) variations in the magnetic field of the charged particles can interfere with onboard electromagnets that impose torques and carry out slow attitude maneuvers. The attractive force on the satellite is very small and is almost always ignored. But the effect of ionization on atmospheric density is noticeable, so we'll discuss it further.

The magnetic field's strength varies with the environment around the Earth, and the mean magnetic field is usually modeled with a low degree-and-order spherical-harmonic expansion (exactly analogous to the gravitational models). Direct collisions of the solar wind and air particles interacting with the Earth's geomagnetic field heat the atmosphere. Geomagnetic activity must be measured to determine the heat generated. The commonly used geomagnetic ***planetary index***, K_p , is a quasi-logarithmic, worldwide (hence the title) average of geomagnetic activity below the auroral zones. Twelve stations record values of K measured every three hours. Once latitude corrections are applied, the average K_p is formed. The quasi-logarithmic identifier is used because values range from low activity ($K_p = 0.0$) to extreme geomagnetic activity ($K_p = 9.0$) and are quoted to the nearest third of an integer. Some atmospheric models, like the Jacchia-Roberts model, use K_p as an input. The geomagnetic ***planetary amplitude***, a_p , is a linear equivalent of the K_p index, designed to minimize differences at 50° latitude. It's also known as the ***3-hourly index*** because eight values of a_p are averaged to create the ***daily planetary amplitude***, A_p , shown in Fig. 8-8. The effects of drag resulting from magnetic disturbances are noticeable for satellites at altitudes between 300 km and 1000 km.

Planetary geomagnetic indices (K_p and a_p) are compiled using measurements from twelve observatories which lie between 48° N and 63° S latitudes; three of these are in the United Kingdom, two in Canada, three in the USA, and the remaining four in New Zealand, Australia, Sweden, and Denmark. The most accepted compilation of the measurements from these observatories is from the Institut für Geophysik at Göttingen University, Germany. The daily planetary amplitude is in gamma units, where

$$\text{one gamma} = 10^{-9} \text{ Tesla} = 10^{-9} \text{ kg}\cdot\text{s} / \text{m}$$

The range of values for A_p is from 0 to 400; values greater than 100 are rare, and values of 10–20 are average. The daily planetary amplitude tends to follow the 11-year cycle of sunspots, although consistently large maxima of A_p usually occur in the declining phase of each 11-year cycle of $F_{10.7}$. There is also a secondary semi-annual cycle due to the variable position of the solar wind with respect to the Earth's magnetosphere. This cycle is just as variable and hard to predict as the sunspot cycle. Variations of A_p from the sun-

spot and semi-annual cycles are mainly due to solar flares, coronal holes, disappearing solar filaments, and the solar-wind environment near the Earth (Fraser-Smith, 1972; Nostrand, 1984). Intense geomagnetic activity at the auroral zones affects the shape of the atmosphere and makes atmospheric density depend on latitude (de Lafontaine, 1986). Sample values of A_p are shown in Fig. 8-8.

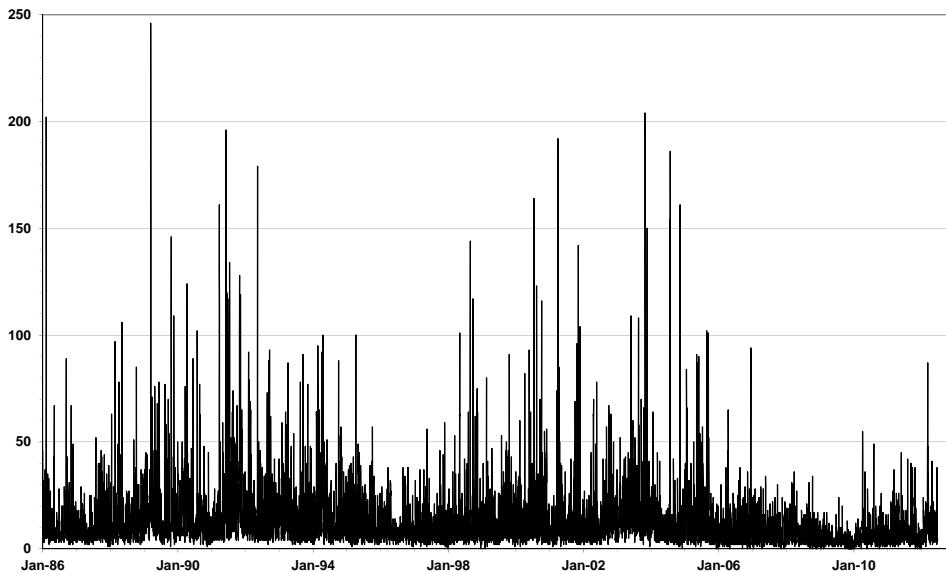


Figure 8-8. Daily Geomagnetic Planetary Index. Notice the seemingly chaotic behavior at almost all times for the average daily A_p . There may be some correlation to the solar cycle, but accurate prediction is very challenging! All units are gammas.

The 3-hourly index, a_p , is measured every three hours. It can be quite different from the daily values, and including it in calculations often results in substantially different (hundreds of meters) results. Some values are shown in Fig. 8-9.

Converting between K_p and a_p indices over their entire range, other than those actually measured is important. In over 50 years of recorded space weather data, there are only discrete recorded values of K_p and a_p quantities. Chapman and Bartels (1940) originally defined these discrete values (Table 8-3).

Examining a plot of the quantities, it's readily apparent that the relation is not linear.

Two practical situations arise. For the last month, only the 3-hourly K_p values are given, but any atmospheric model using a_p will need to convert the values, and when the values are finalized, they will be from the set of discrete values mentioned above. The second situation arises when assembling the predicted data, for which only a_p data is available. Specifically, values of a_p occur that do not have an exact corresponding K_p value per Chapman and Bartels (1940). To consistently use the data, a standard technique should be employed to convert the data to and from the a_p and K_p values, for both of these situations.

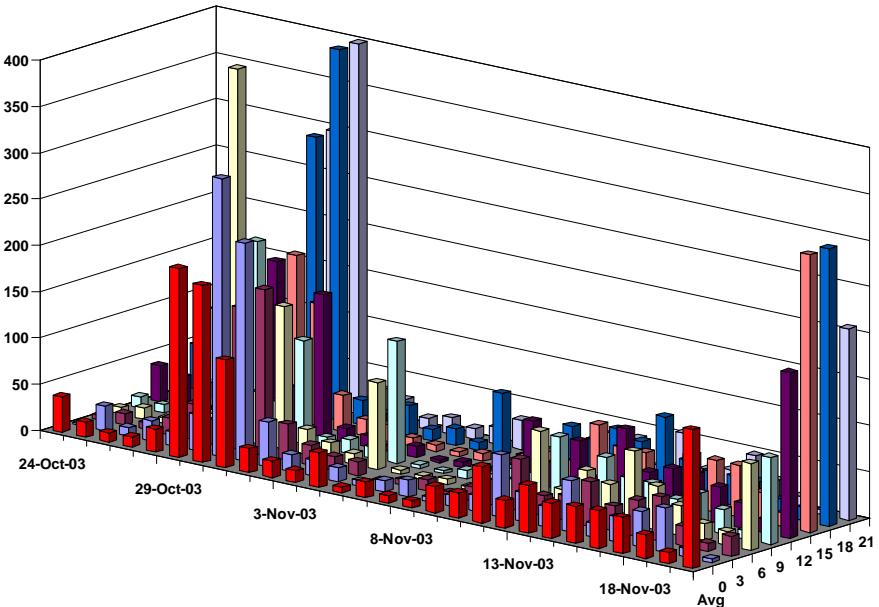


Figure 8-9. 3-hourly Geomagnetic Index. Notice the variability within a 24-hour day. The average daily values are shown in the “front” with each 3-hourly value behind it. All units are gammas.

TABLE 8-3. K_p and a_p Values. This table gives the discrete values for the geomagnetic indices. Note the plus and minus indicators represent 1/3 values and should not be rounded. Only two decimal places are shown for brevity.

K_p	00	0+	1-	10	1+	2-	20	2+	3-	30	3+	4-	40	4+
K_p	0.0	0.33	0.67	1.0	1.33	1.67	2.0	2.33	2.67	3.0	3.33	3.67	4.0	4.33
a_p	0	2	3	4	5	6	7	9	12	15	18	22	27	32
K_p	5-	50	5+	6-	60	6+	7-	70	7+	8-	80	8+	9-	90
K_p	4.67	5.0	5.33	5.67	6.0	6.33	6.67	7.0	7.33	7.67	8.0	8.33	8.67	9.0
a_p	39	48	56	67	80	94	111	132	154	179	207	236	300	400

A cubic splining technique seems to best replicate the observed data values, while simultaneously maintaining closure properties. The approach for the cubic spline is shown below (note this is modified to match all four points instead of 2 points and slope).

ALGORITHM 63: AP2Kp CONVERSION ($K_p \Rightarrow a_p$)

Setup defining arrays of a_p and K_p from Table 8-3.

Locate input a_{pin} or K_{pin} within the appropriate defining array input a_p or K_p (respectively)

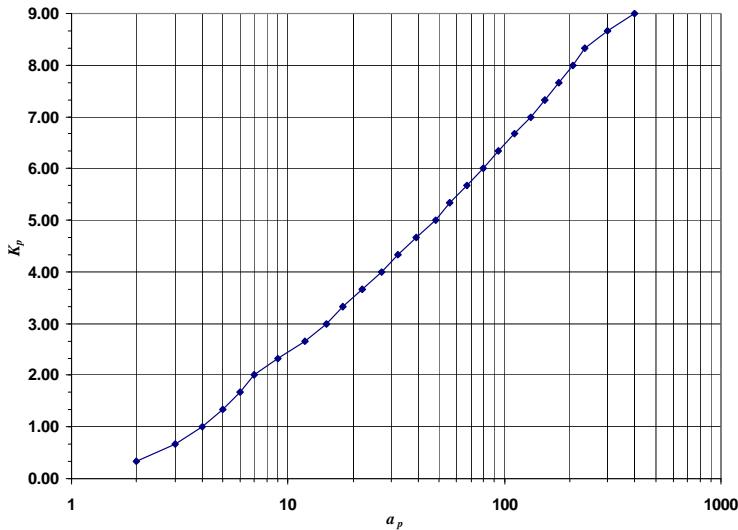


Figure 8-10. Relationship between a_p and K_p . The semi-logarithmic correlation between K_p and a_p is shown. Note that there are only distinct values that correspond between each scale and the K_p values are divided by 10 from published space weather values.

Find 4 adjacent points in the defining array for K_p and a_p (bracket the input K_{pin} and a_{pin} between points 2 and 3 ($K_{p1}, K_{p2}, K_{p3}, K_{p4}, a_{p1}, a_{p2}, a_{p3}, a_{p4}$)

Form coefficients for the cubic polynomials (one for a_p and one for K_p)

$$\alpha_{k0} = K_{p2}$$

$$\alpha_{a0} = a_{p2}$$

$$\alpha_{k1} = -\frac{K_{p1}}{3} - \frac{K_{p2}}{2} + K_{p3} - \frac{K_{p4}}{6}$$

$$\alpha_{a1} = -\frac{a_{p1}}{3} - \frac{a_{p2}}{2} + a_{p3} - \frac{a_{p4}}{6}$$

$$\alpha_{k2} = \frac{K_{p1}}{2} - K_{p2} + \frac{K_{p3}}{2}$$

$$\alpha_{a2} = \frac{a_{p1}}{2} - a_{p2} + \frac{a_{p3}}{2}$$

$$\alpha_{k3} = -\frac{K_{p1}}{6} + \frac{K_{p2}}{2} - \frac{K_{p3}}{2} + K_{p4}$$

$$\alpha_{a3} = -\frac{a_{p1}}{6} + \frac{a_{p2}}{2} - \frac{a_{p3}}{2} + a_{p4}$$

Solve the cubic polynomial for the real root (x) between 0.0 and 1.0 for either K_p or a_p (both are shown for completeness) using the input K_{pin} or a_{pin} value.

$$\alpha_{a3}x^3 + \alpha_{a2}x^2 + \alpha_{a1}x + \alpha_{a0} - a_{pin} = 0$$

$$\alpha_{k3}x^3 + \alpha_{k2}x^2 + \alpha_{k1}x + \alpha_{k0} - k_{pin} = 0$$

Solve for the remaining a_p or K_p value (both are shown for completeness)

$$\begin{aligned} a_p &= \alpha_{a3}x^3 + \alpha_{a2}x^2 + \alpha_{a1}x + \alpha_{a0} - K_{pin} \\ K_p &= \alpha_{k3}x^3 + \alpha_{k2}x^2 + \alpha_{k1}x + \alpha_{k0} - a_{pin} \end{aligned}$$

Solar Flux Data

Because the Sun's overall activity influences changes in the atmosphere, the terms solar maxima and solar minima are often used to refer to its total output. During periods of **solar minima**, the incoming radiation is less than normal and only slightly influences most satellite orbits. But periods of **solar maxima** can produce large unpredictable perturbations in many satellites.

The contribution of solar flux to atmospheric density is mainly from incoming solar radiation. **Solar flux** (or **Extreme Ultra-Violet**, EUV, radiation that heats the upper atmosphere, F_{EUV}) is impossible to measure at the Earth's surface because the atmosphere doesn't allow transmission of EUV radiation. Today, there are a few space-borne systems to measure EUV flux, but no atmospheric-density models designed specifically to use these space-based measurements. However, scientists have determined that both EUV and incoming solar radiation with a wavelength of 10.7 cm, $F_{10.7}$ ($f = 2800$ MHz), originate in the same layers of the Sun's chromosphere and corona. Because the Earth's atmosphere is transparent to $F_{10.7}$ radiation, we can infer the relative strength of F_{EUV} from Earth-based measurements of 10.7 cm-length radio waves. Regular measurements of $F_{10.7}$ exist from about 1940. It's measured in *Solar Flux Units*, SFU, where

$$1 \text{ SFU} = 1 \times 10^{-22} \frac{\text{watt}}{\text{m}^2 \text{Hz}} \quad (8-31)$$

Typical values range from less than 70 to more than 300 SFU (Nostrand, 1984). The most commonly accepted measurement of $F_{10.7}$ is distributed daily by the National Oceanic and Atmospheric Administration (NOAA) at the National Geophysical Data Center in Boulder, Colorado. Measurements were routinely made at the Algonquin Radio Observatory in Ottawa, Ontario, Canada from 1947 until May 31, 1991 at 1700 UT. Since then, the measurements have been made at the Dominion Radio Astrophysical Observatory (DRAO) Penticton, British Columbia, Canada, at 2000 UT. Daily values are averaged to produce 81-day average values (3 solar rotations) denoted with a bar, $\bar{F}_{10.7}$. Figure 8-11 shows the values of $F_{10.7}$ for about one solar cycle.

The solar flux data may be either observed (at the true Sun-Earth distance), or adjusted to 1.0 AU (see Fig. 8-12). The raw observed data (and adjusted in DAILY-PLT.ADJ) may be found at NGDC (2013) and the conversion uses the current Earth-Sun distance ($r_{\oplus-Sun}$) and the AU distance (149,597,870.66 km).

$$F_{10.7}(obs) = \frac{F_{10.7}(adj)AU^2}{r_{\oplus-Sun}^2}$$

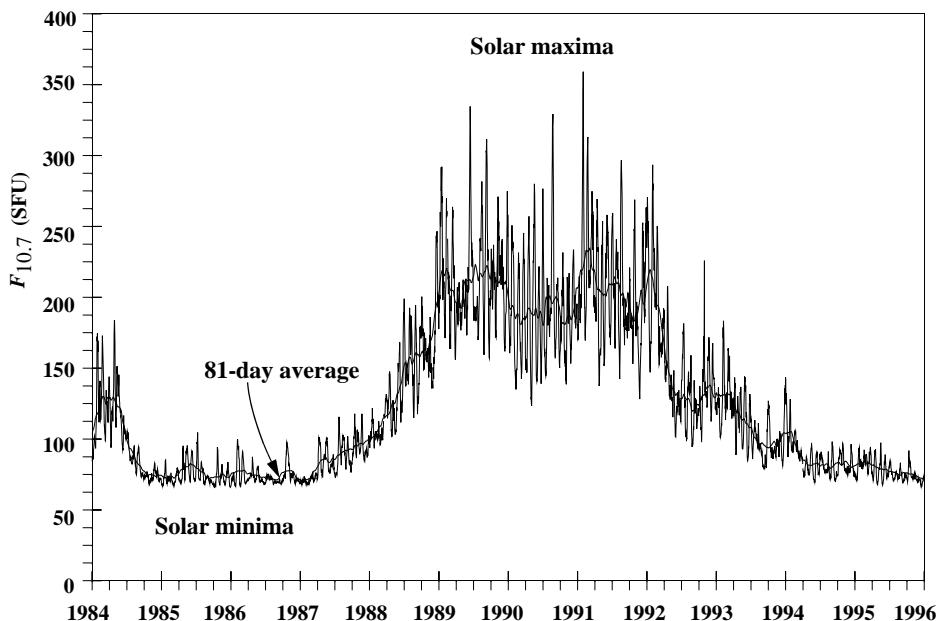


Figure 8-11. Values of Solar Flux. I've shown several years of $F_{10.7}$ values. Remember, $F_{10.7}$ is a radiation that originates at the same place as EUV which can't be measured on the Earth's surface. Notice the erratic behavior but a prominent rise during solar max. I've drawn the 81-day average, which we often use in density calculations. $F_{10.7}$ is also very difficult to predict.

Although the differences are small, some atmospheric models specifically require the solar flux at the true Earth-Sun distance. The difference is large enough to require separate values for the 81-day average, and because many programs use either the centered or trailing 81-day averages (about the date of interest), a total of six values are available for the solar flux. Also note that the values found in the GEOMAGNETIC and the SOLAR FLUX directories on NGDC do not agree completely because they are actually different data. The GEOMAGNETIC files contain Lenhart adjusted data while the SOLAR FLUX files contain observed and adjusted DRAO data (Knapp, 2005). Observed – adjusted values should show the cyclical variations of each solar cycle. Differences of the adjusted DRAO and Lenhart data should be zero. The data spikes in Fig. 8-12 appear to be random in that neither data source appears to be “correct” for all times.

Predicting Solar Flux

Solar flux receives a lot of attention because it is an important parameter in determining atmospheric density. The primary quantity, $F_{10.7}$, has been used as a proxy for the EUV radiation for many years. Several new indices have been under consideration for the past few years—Mg II (from NOAA and other satellites at about 280 nm), EUV (from the

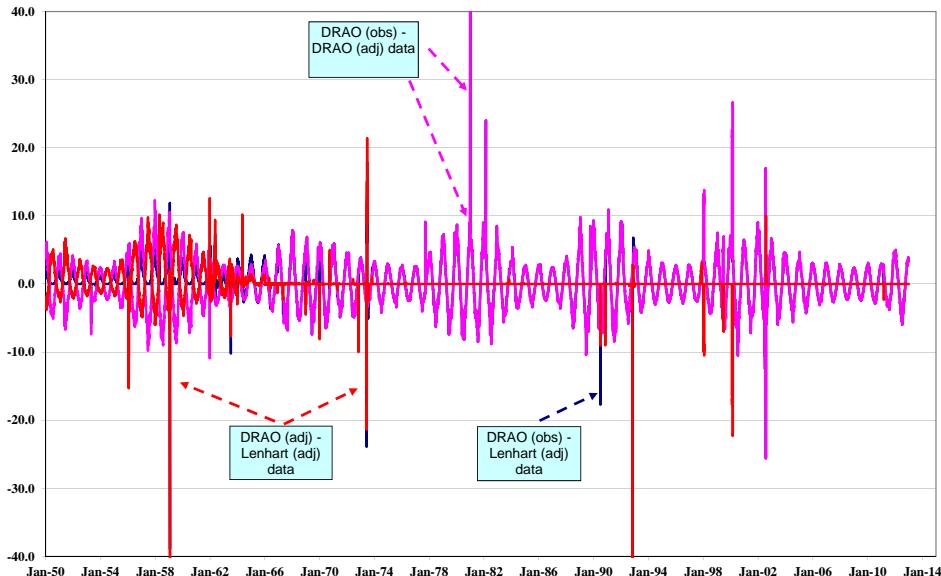


Figure 8-12. Difference of Observed and Adjusted Solar Flux Values. The solar cycles are evident in the observed minus adjusted solar flux differences (in SFU), along with the seasonal variations that drive the original variations over time. The data spikes occur in both observed and adjusted values, and in the DRAO and Lenhart data for a variety of reasons including multiple Sun-Earth distance corrections.

NASA/ESA SOHO satellite at the Lagrange point at 26–34 nm), etc.—but none has been unequivocally proven better than another and established as a new leader that will dramatically improve the accuracy of satellite operations. Moreover, predicting the $F_{10.7}$ values into the future poses significant challenges, and introduces significant uncertainty into operations requiring accurate forecasts of atmospheric behavior. Schatten and Sofia (1987) and Schatten (1988) have developed a monthly estimate of $F_{10.7}$ and a_p . Figure 8-13 shows some of those predictions and current data. There are early, mid and late timing, as well as min, mid, and maximum values for each prediction.

The difficulty of the prediction process becomes apparent for any cycle that departs significantly (i.e., cycle 24) from previous trends. Consider Fig. 8-14 for the plethora of predictions.

Even the short term predictions (3, 27 and 45 day) exhibit significant variations. Figure 8-15 shows the solar flux short term predictions.

The variability of the predictions is apparent. Many predictions rely on the solar sunspot activity. Routinely monitored since the 1700's, several relations exist between the sunspot number, R , averaged over a month or longer, and $F_{10.7}$ (Svalgaard, 2009).

$$F_{10.7} = 63.7 + 0.728 R + 0.00089 R^2$$

$$F_{10.7} = 67 + 0.97 R + 17.6 (e^{-0.035R} - 1.0)$$

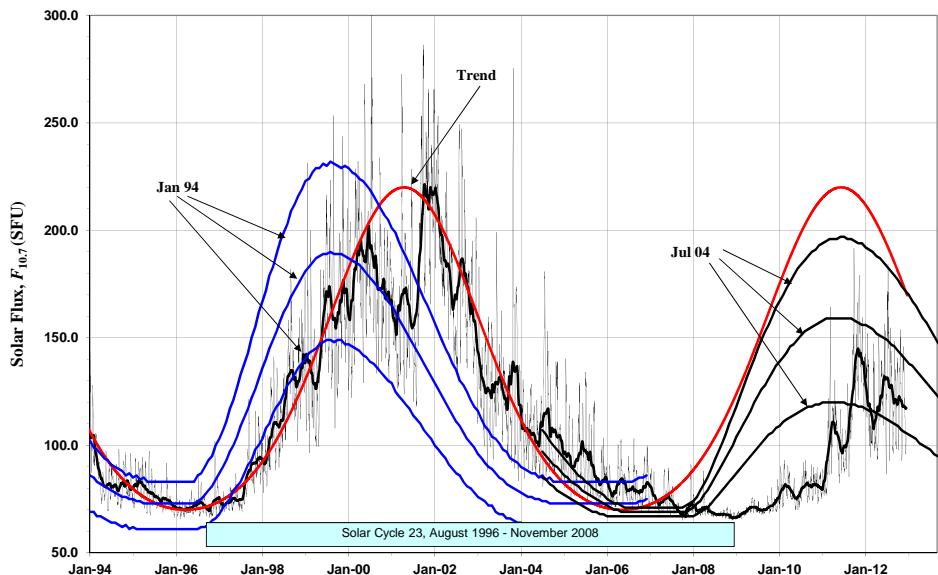


Figure 8-13. Schatten Predicted $F_{10.7}$. Actual data is shown for two Schatten predictions of $F_{10.7}$. Notice there are min, mid, and maximum predictions of the solar maxima (Schatten, 1988). Each prediction also includes early, mid, and late timing for the solar cycle (not shown). Comparing the predictions to actual data, the Jan 94 prediction was reasonably accurate while the Jul 04 prediction was significantly off. The 81-day average is close to the nominal prediction, although there is a lag. The polynomial is a simple heuristic expression that matches the last four solar cycles.

A polynomial intended to match the last few solar cycles, and assuming the next will not change dramatically from the previous, could also be used. In Eq. (8-32), t is the number of days from January 1, 1981. While matching cycle 23, it misses cycle 24.

$$F_{10.7} = 145 + 75 \cos(0.001\,696t + 0.35 \sin(0.000\,016\,95t)) \quad (8-32)$$

Another interesting approach is given by Oltrogge and Chao (2007). Their method for extrapolating and interpolating atmospheric density is based on adjusting or modifying actual density values, not the proxies. The rationale is that the dependence of density on the proxies is nonlinear; hence averaging or interpolating proxies does not yield correct average or interpolated density. They were examining the effect of solar flux prediction estimates on satellite lifetimes and noted that most orbital lifetime prediction errors were caused by (1) the poor performance of current mean solar activity predictions as compared to actual activity variations; (2) non-availability of predictions more than one solar cycle away; and (3) nonlinear density as a function of solar and geomagnetic activity. To circumvent these obstacles, they used the complete set of solar and geomagnetic data available (from Feb 1947). Thus, five solar cycles were used and combined into a single cycle of 10.82546 years (3954 days). Rather than attempting to produce a single

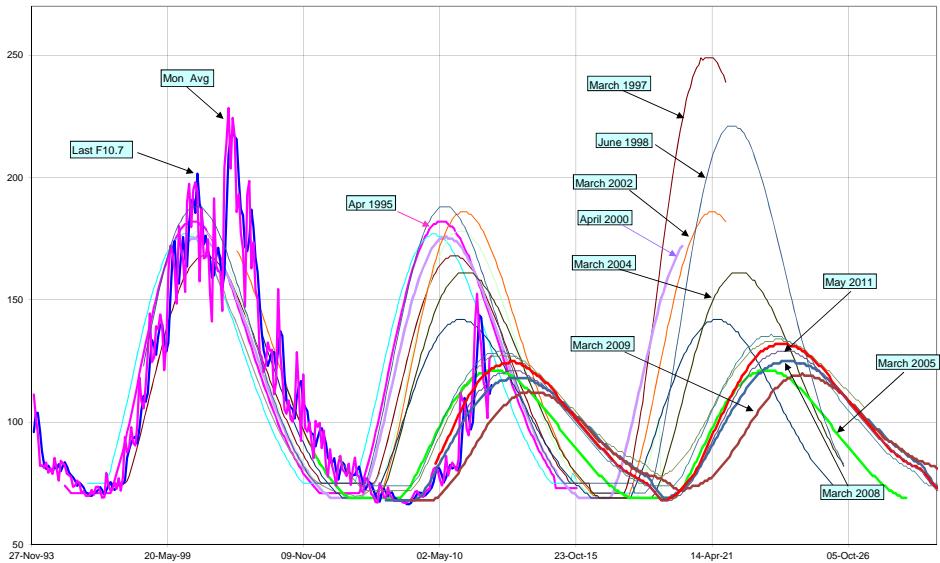


Figure 8-14. Recent Solar Flux Predictions. Several Schatten predictions of solar flux are shown. Notice that each Schatten prediction covers about two solar cycles and the newer predictions suggest a lower solar max for cycles 24 and 25. Units are SFU.

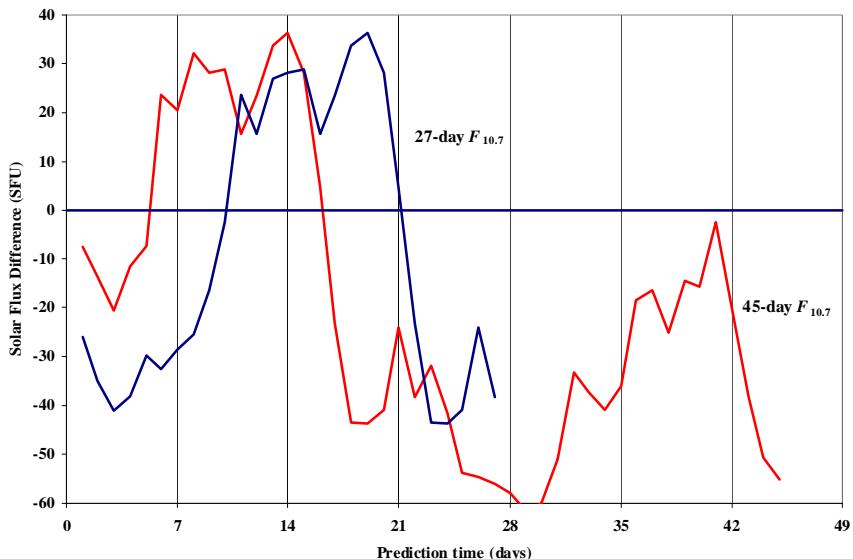


Figure 8-15. Short term $F_{10.7}$ Predictions. The 27 and 45 day predictions are shown when compared to the actual data. Notice the significant departures. The geomagnetic indices exhibit the same behavior.

mean solar cycle, they chose to directly use the solar flux and geomagnetic values from the previous cycles. Essentially, the technique creates a day of solar and geomagnetic data (a combined triad of values including the $\bar{F}_{10.7}$ 81-day average) from one of the five cycles at random, and does so for each day of the simulation. Thus, the existing data is used to construct a future solar cycle, while retaining all the variability and interrelations of the existing data.* This technique is certainly worthy of additional analysis and use as it avoids the vagaries of other prediction techniques that often vary widely for estimates just a few years apart.

Model Atmospheres

Numerous density models have been developed over the past few decades from two main approaches: 1) combining conservation laws and atmospheric-constituent models into a physical model; 2) using simplified physical concepts developed from in-situ measurements and satellite-tracking data. This book presents both static and time-varying models to find density. Marcos et al. (1993) discuss the development history of many of the representative atmospheric models, as shown in Fig. 8-16.

There are various models to satisfy differing accuracy models. We could always use a time-varying model, like the Jacchia-Roberts atmosphere, but the computational requirements might be too great. Models of this fidelity are the most complete and need accurate data, but they also demand the highest computational power. On the other hand, the simple exponential model may be too inaccurate for some applications. Gaposchkin and Coster (1988) discuss various atmospheric models in detail. They conclude that no model is best for *all* applications. This is because there are physical reasons that the models have errors—winds, gravity waves, etc. They note that Jacchia-71 performs quite well, and is the fastest among the models tested (J71, DTM, J77, and three variations of J77). However, they prefer the J77 with certain updates for the best combination of speed, accuracy, and applicability.

One feature of the analytical density models is that their relatively simple structure allows analytical computation of the integrals required for analytical and semianalytical satellite propagation theories using these densities. Numerical implementations, and some models in semianalytical theories like the DSST require numerical quadrature for the same integrals.

Exponential Model (0–1000 km)

This simple, static model assumes the density of the atmosphere decays exponentially with increasing altitude. It also assumes a spherically symmetrical distribution of particles, in which the density, ρ , varies exponentially according to

$$\rho = \rho_o \text{EXP}\left[-\frac{h_{ellp} - h_o}{H}\right] \quad (8-33)$$

* Their resulting solar and geomagnetic activity computer code tdt2f10ap_60yr is available upon request from the author (Oltrogge@1EarthResearch.com) for users wishing to explore the model in greater detail.

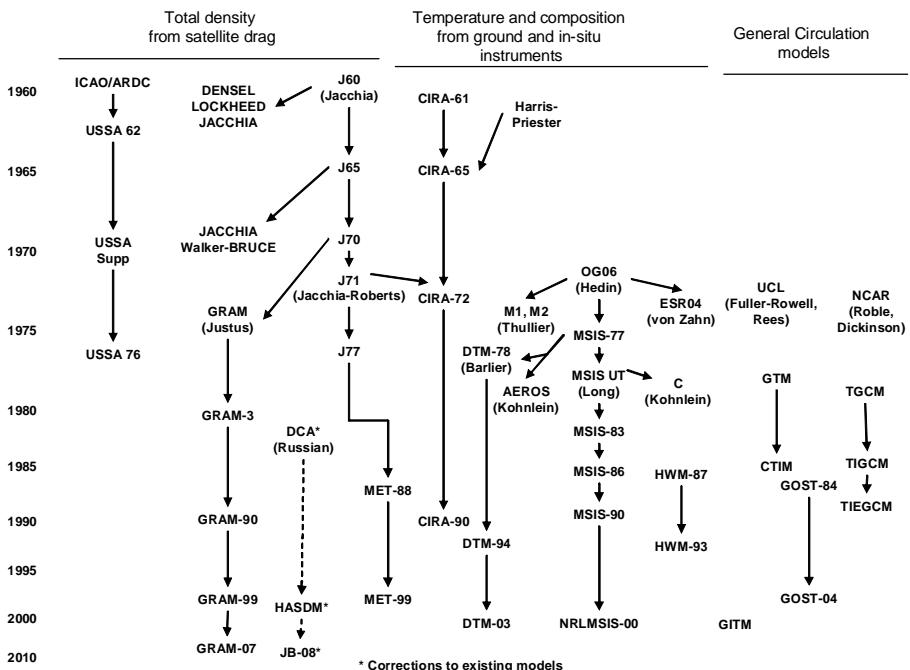


Figure 8-16. Development of Atmosphere Models. Notice the variety of models. Flow of information among the three overall categories is limited (Marcos, et al. 1993, 20). The main models in use today are the Standard Atmosphere, USSA76; variations of the Jacchia-Roberts, J71, J77, and GRAM99; COSPAR International Reference Atmosphere, CIRA90; Mass Spectrometer Incoherent Scatter, NRLMSIS-00; Drag Temperature Model (DTM), Marshall Engineering Thermosphere (MET), the Russian GOST, Global Ionosphere Thermosphere Model (GITM), and general circulation models. Dynamic Calibration of the atmosphere (DCA), High Accuracy Satellite Drag Model (HASDM), and Jacchia-Bowman (JB-08) are corrections to other models, usually J70.

where a *reference density*, ρ_o , is used with the *reference altitude*, h_o , the *actual altitude*, above the ellipsoid, h_{ellp} , and the *scale height*, H .^{*} Table 8-4 gives values and descriptions. Although this model approximates much of the atmosphere, it lacks the accuracy and sophistication needed for highly accurate studies. To fix this deficiency, we'd first try to segregate the data into closely related bands. One's choice of bands determines the accuracy. Within limits, this approach works for simulating drag effects for design studies. Extensive tabulated data on atmospheric parameters (density, pressure, temperature,

* Scale height is the fractional change in density with height. It can be useful in determining numerical partial derivatives. Wertz (1978:108) shows that scale height is equal to $k \times \text{Temperature} / \text{molecular weight} \times \text{gravity}$, where k is the Boltzmann constant.

scale height) are in the *U.S. Standard Atmosphere* (1976), together with procedures for calculating values at specified hours of local time, date, and location.

TABLE 8-4. Exponential Atmospheric Model. Although a very simple approach, this method yields moderate results for general studies. Source: Wertz, 1978:820, which uses the *U.S. Standard Atmosphere* (1976) for 0 km, CIRA-72 for 25–500 km, and CIRA-72 with exospheric temperature, $T_\infty = 1000$ K for 500–1000 km. The scale heights have been adjusted to maintain a piecewise-continuous formulation of the density.

Altitude h_{ellp} (km)	Base Altitude h_o (km)	Nominal Density ρ_o (kg/m ³)	Scale Height H (km)	Altitude h_{ellp} (km)	Base Altitude h_o (km)	Nominal Density ρ_o (kg/m ³)	Scale Height H (km)
0–25	0	1.225	7.249	150–180	150	2.070×10^{-9}	22.523
25–30	25	3.899×10^{-2}	6.349	180–200	180	5.464×10^{-10}	29.740
30–40	30	1.774×10^{-2}	6.682	200–250	200	2.789×10^{-10}	37.105
40–50	40	3.972×10^{-3}	7.554	250–300	250	7.248×10^{-11}	45.546
50–60	50	1.057×10^{-3}	8.382	300–350	300	2.418×10^{-11}	53.628
60–70	60	3.206×10^{-4}	7.714	350–400	350	9.518×10^{-12}	53.298
70–80	70	8.770×10^{-5}	6.549	400–450	400	3.725×10^{-12}	58.515
80–90	80	1.905×10^{-5}	5.799	450–500	450	1.585×10^{-12}	60.828
90–100	90	3.396×10^{-6}	5.382	500–600	500	6.967×10^{-13}	63.822
100–110	100	5.297×10^{-7}	5.877	600–700	600	1.454×10^{-13}	71.835
110–120	110	9.661×10^{-8}	7.263	700–800	700	3.614×10^{-14}	88.667
120–130	120	2.438×10^{-8}	9.473	800–900	800	1.170×10^{-14}	124.64
130–140	130	8.484×10^{-9}	12.636	900–1000	900	5.245×10^{-15}	181.05
140–150	140	3.845×10^{-9}	16.149	1000–	1000	3.019×10^{-15}	268.00

▼ Example 8-4. Calculating Atmospheric Density.

GIVEN: $r = 7125.3489$ km

FIND: ρ

Eq. (8-33) requires knowledge of the altitude, found by subtracting the Earth's radius (6378.137 km) from the satellite's radius r ($h_{\text{ellp}} = 747.2119$ km). Now, using values from Table 8-4, we use Eq. (8-33)

$$\rho = 3.614 \times 10^{-14} \exp \left[-\frac{747.2119 - 700}{88.667} \right] = 2.1219854 \times 10^{-14} \frac{\text{kg}}{\text{m}^3}$$

Note that the units in the exponential cancel (all are km), and the result is less than the base value at 700 km, as we would expect from being at a higher altitude. In practice, the height is found above the ellipsoid and does not assume a spherical Earth.

We can also determine pressure through a similar formula that approximates pressure at any altitude. The formula is

$$p = p_{SL} \text{EXP} \left[-\frac{h_{ellp}}{H} \right]$$

where the variables are similar to the density formula. The *scale height*, H , is 7.0104 km, and the *height above the ellipsoid* is h_{ellp} . The sea-level pressure, p_{SL} , which is 1.0133 N/m², lets us find the pressure, p , above sea level, usually in N/m².

Standard Atmosphere (0–1000 km)

The Standard Atmosphere has been published twice (1962 and 1976) in an attempt to provide a standard reference model. It is an ideal, steady-state model of the Earth's atmosphere at a latitude of 45°N during moderate solar activity. The U.S. Committee on Extension to the Standard Atmosphere (COESA) has adopted two definitions that show its intent:

A hypothetical vertical distribution of atmospheric temperature, pressure and density which, by international agreement, is roughly representative of year-round, midlatitude conditions. Typical usages are as a basis for pressure altimeter calibrations, aircraft performance calculations, aircraft and rocket design, ballistic tables, and meteorological diagrams. The air is assumed to obey the perfect gas law and the hydrostatic equation which, taken together, relate temperature, pressure and density with geopotential. Only one standard atmosphere should be specified at a particular time and this standard atmosphere must not be subjected to amendment except at intervals of many years. (US Standard, 1976:xiv)

Satellites operating at much higher altitudes required the Committee to adopt another goal for the Standard Atmosphere:

This atmosphere shall also be considered to rotate with the Earth, and be an average over the diurnal cycle, semi-annual variation, and the range of conditions from active to quiet geomagnetic, and of active to quiet sunspot conditions. Above the tropopause (about 110 km) generalized forms of the hydrostatic equations apply. (US Standard, 1976:xiv).

CIRA (25–2500 km)

The Committee on Space Research (COSPAR) of the International Council of Scientific Unions periodically determines an atmospheric model. The current version is the COSPAR International Reference Atmosphere (CIRA-90) model. The first model was produced in 1965 (CIRA-65), which was basically a new model for altitudes from 30–300 km, based on Champion (1963) and the Harris-Priester (1962) model from 120–800 km. CIRA-72 included mean values from 25–500 km, and Jacchia (1971) prepared models from 110–2000 km. The model is a semi-theoretical technique, but it does have some free variables. Data comes mainly from measurements of satellite drag and ground-based measurements.

Harris-Priester (120–2000 km)

This model is basically static, but several tables account for the densities we may observe within the solar cycle. Interpolation then determines the density at a particular time. This model is recommended for comparing propagation algorithms because it will

give fairly accurate results and is computationally efficient. The model was part of CIRA-65, but it continues to be useful separately for simulations requiring an estimate of density during different regions of the solar cycle.

The Harris-Priester model determines the physical properties of the upper atmosphere by averaging several of the predominant variations. They include approximations for fluxes from the extreme ultraviolet but average the semiannual and seasonal latitudinal variations. Harris-Priester doesn't model the effects of the 27-day solar rotation. Long et al. (1989:4-57) lists "discrete values of density for the maximum- and minimum-density profiles [shown in Table 8-5]... [These values] correspond to the mean solar activity and are stored in tabular form as $\rho_{min}(h_i)$ and $\rho_{max}(h_i)$, respectively. Different maximum and minimum profiles correspond to different levels of solar activity."

Jacchia-Roberts (70–2500 km)

Jacchia's models (J70, 1970; J71, 1971; J77, 1977) are popular in military operations (J70) and some analyses for astrodynamics. The Jacchia-Roberts atmosphere contains analytical expressions for determining exospheric temperature as a function of position, time, solar activity, and geomagnetic activity. With a computed temperature, we get density from empirically determined temperature profiles or from the diffusion equation.

One advantage of the Jacchia model over CIRA-65 is that it applies over all latitudes. NASA used the Jacchia model in preference to CIRA-65 but employed the diffusion equation instead of Jacchia's numerical approximation to eliminate extensive tables. Jacchia developed this approach in his 1970 models of the upper atmosphere (Jacchia, 1970). These empirically defined models represent atmospheric density as a function of exospheric temperature and altitude. Jacchia divided the upper atmosphere into several regions: 90–125 km and above 125 km. He determined these regions by assuming that mixing dominates between 90 and 100 km, with a fixed-boundary atmospheric condition at 90 km. In addition, Jacchia assumed diffusive equilibrium above 100 km. This led to using a low-altitude temperature profile in the diffusion differential equation between 100 and 125 km and a high-altitude temperature profile above 125 km. Jacchia solved these differential equations by numerically integrating them over various constant values of exospheric temperature. As a point of interest, the procedure routinely uses an 81-day average of solar parameters. At first glance, this may seem arbitrary, but it's necessary to average out differences caused by the solar-rotation cycle (27 days), so we have a three-period cycle. Jacchia tabulated these results for use in atmospheric-drag simulations through interpolation (Jacchia, 1970). If we analyze historical data, the 81-day average is easy to find, but predictions need some estimate such as the reasonably accurate Schatten approximation.

In 1971, Jacchia re-formulated his model using newer, and in many cases, more complete data. The changes are dramatic in comparison to the 1970 model. Jacchia (1971) remarks:

While overhauling the basic models, we have also tried to reanalyze these variations. In so doing, we have found that for some of them—the geomagnetic effect, the semiannual variation, and the Helium variation—the analytical formulation we had used was inadequate and had to be altered, or even drastically changed. In particular, the dissociation of the

TABLE 8-5. Tables of Density by Altitude. The Harris-Priester model uses data tables of this type. Long, et al. (1989:4-58)

Height (km)	Minimum Density (kg/m ³)	Maximum Density (kg/m ³)	Height (km)	Minimum Density (kg/m ³)	Maximum Density (kg/m ³)
100	4.974×10^{-7}	4.974×10^{-7}	420	1.558×10^{-12}	5.684×10^{-12}
120	2.490×10^{-8}	2.490×10^{-8}	440	1.091×10^{-12}	4.355×10^{-12}
130	8.377×10^{-9}	8.710×10^{-9}	460	7.701×10^{-13}	3.362×10^{-12}
140	3.899×10^{-9}	4.059×10^{-9}	480	5.474×10^{-13}	2.612×10^{-12}
150	2.122×10^{-9}	2.215×10^{-9}	500	3.916×10^{-13}	2.042×10^{-12}
160	1.263×10^{-9}	1.344×10^{-9}	520	2.819×10^{-13}	1.605×10^{-12}
170	8.008×10^{-10}	8.758×10^{-10}	540	2.042×10^{-13}	1.267×10^{-12}
180	5.283×10^{-10}	6.010×10^{-10}	560	1.488×10^{-13}	1.005×10^{-12}
190	3.617×10^{-10}	4.297×10^{-10}	580	1.092×10^{-13}	7.997×10^{-13}
200	2.557×10^{-10}	3.162×10^{-10}	600	8.070×10^{-14}	6.390×10^{-13}
210	1.839×10^{-10}	2.396×10^{-10}	620	6.012×10^{-14}	5.123×10^{-13}
220	1.341×10^{-10}	1.853×10^{-10}	640	4.519×10^{-14}	4.121×10^{-13}
230	9.949×10^{-11}	1.455×10^{-10}	660	3.430×10^{-14}	3.325×10^{-13}
240	7.488×10^{-11}	1.157×10^{-10}	680	2.620×10^{-14}	2.691×10^{-13}
250	5.709×10^{-11}	9.308×10^{-11}	700	2.043×10^{-14}	2.185×10^{-13}
260	4.403×10^{-11}	7.555×10^{-11}	720	1.607×10^{-14}	1.779×10^{-13}
280	2.697×10^{-11}	5.095×10^{-11}	760	1.036×10^{-14}	1.190×10^{-13}
290	2.139×10^{-11}	4.226×10^{-11}	780	8.496×10^{-15}	9.776×10^{-14}
300	1.708×10^{-11}	3.526×10^{-11}	800	7.069×10^{-15}	8.059×10^{-14}
320	1.099×10^{-11}	2.511×10^{-11}	840	4.680×10^{-15}	5.741×10^{-14}
340	7.214×10^{-12}	1.819×10^{-11}	880	3.200×10^{-15}	4.210×10^{-14}
360	4.824×10^{-12}	1.337×10^{-11}	920	2.210×10^{-15}	3.130×10^{-14}
380	3.274×10^{-12}	9.955×10^{-12}	960	1.560×10^{-15}	2.360×10^{-14}
400	2.249×10^{-12}	7.492×10^{-12}	1,000	1.150×10^{-15}	1.810×10^{-14}

semiannual variation from temperature variations has cleared up many puzzling results from the Helium-hydrogen region and eliminated the necessity of introducing ad hoc variations for these constituents.

The Jacchia 1977 revision incorporates satellite mass spectrometer data. The model again revised equations from the previous models, and there are several variations of this technique.

Roberts (1971) recognized that tabular determination of atmospheric density and numerical integration to calculate partial derivatives for drag is computationally intensive. So, in 1971, he analytically evaluated the 1970 Jacchia models. Roberts uses partial fractions to integrate values between 90 km and 125 km. For altitudes above 125 km, he introduces a different asymptotic function than the one Jacchia introduced to achieve an integrable form. With this substitution, Roberts closely approximates Jacchia's results above 125 km. Although he based his analytical equations on the 1970 Jacchia model (before Jacchia published his 1971 models), we can still use his corrections with Jacchia's later model. Draper Laboratory uses this approach and has corrected an error in the $W(v)$ function in Equation 12 of Roberts' 1971 paper (Long et al. 1989:4-51). According to Long et al. (1989:4-36), the Jacchia-Roberts density deviates no more than 6.7% from Jacchia values. Sec. B.1 of the Appendixes details the formulas for the Jacchia-Roberts atmosphere. Oza and Freierstag (1995) suggest that the Jacchia-Roberts model is very responsive to changes in geomagnetic activity. Partial derivatives are available for the Jacchia-Roberts model in Long et al. (1989:4-54 to 4-57).

DTM (200–1200 km) and NRLMSIS-00 (0–2000 km)

These two additional models are based on air-glow temperatures (DTM—Barlier, et al. 1978) and incoherent radar scatter (MSIS—Hedin, et al. 1977), respectively. Barlier uses spherical harmonics to incorporate data on satellite drag from over two complete solar cycles, and significant observational data based on the Thuillier et al. (1977) model of global exospheric temperature. An advantage is the smaller amount of code required for execution (compared to Jacchia-Roberts, for instance). Although both models were developed after the Jacchia-Roberts models and incorporate data which was unavailable, Oza and Freierstag (1995) note that a modified form of Jacchia-Roberts still can out-perform these models in some applications. They suggest that the DTM model is more sensitive to changes in the solar flux activity. They also have found that the 81-day $F_{10.7}$ average should be centered on the day of interest, rather than at the end of the cycle. Unfortunately, some operational programs routinely use the latest [trailing] 81-day average because the predicted values are not well known (Jablonski, 1992).

The MSIS models derive from the DTM models. They are quite accurate and have been successfully applied to many stressing problems. For instance, Hujšák (1999) shows results of using the MSIS model in determining reentry of a decaying Molniya satellite. Because the satellite is traveling so fast at perigee, it can cross several bands of the atmosphere during a single integration step. MSIS does not suffer these difficulties. A newer release is the NRLMSISE-00 model which is extremely popular for all applications. Complete source code is available for all the 1986, 1990, and 2000 models. The models run a little slower than the Jacchia models.

Russian GOST Model (120–1500 km)

The Russian GOST density model (Yurasov, 1999, Voiskovskii, 1973, Volkov, 1984, Jablonski [Boelitz], 1992, and Amelina et al. 1996) is an analytical method to obtain atmospheric density in an aspherical upper atmosphere from observations of Russian Cosmos satellites. It's been part of the Russian operational space surveillance system for

more than 30 years, was used with the Apollo-Soyuz Test Project in 1975, and it continues to incorporate updates from new satellite data. The elegant part of this algorithm is that it can turn factors affecting atmospheric density on or off by simply omitting certain k factors—for example, if diurnal variations aren't needed, set k_2 to 1. Consequently, the model runs very fast. The model also includes long and short period variations in solar activity. The variability in the Russian model gives a 1σ prediction error similar to the position error of actual data calculated with a Harris-Priester density model (Carter et al. 1987). The GOST model differed only 5–10% (Volkov, 1982) from the 1971 Jacchia density model.

The model emerges empirically from observations of the Cosmos satellites' orbital motion. It includes the dependence of the density on solar flux and geomagnetic activity as well as the diurnal and semi-annual density variations. This model is valid for satellites at altitudes of 120–1500 km. The current version is GOST 25645.166-2004. Appendix B further discusses this model.

Dynamic Calibration of the Atmosphere (DCA)

Unlike the previous models, this approach isn't really a model, but rather a technique for improving or correcting atmospheric density. It gives fundamental scientific information about the variations in the density and the statistics of these variations. The work was pioneered by Nazarenko in the early 1980's (Gorochov and Nazarenko, 1982) and researched by Draper Laboratory (many documents including Cefola and Nazarenko, 1999, Granholm, 2000, Bergstrom, 2002, Yurasov et al. 2005, and Wilkins et al. 2006). Granholm (2000) and Wilkins et al. 2006 provide an excellent description of the technique, and Bergstrom (2002) showed results of implementing and testing the theory. The concept relies on changing the density directly. Figure 8-17 shows the general process.

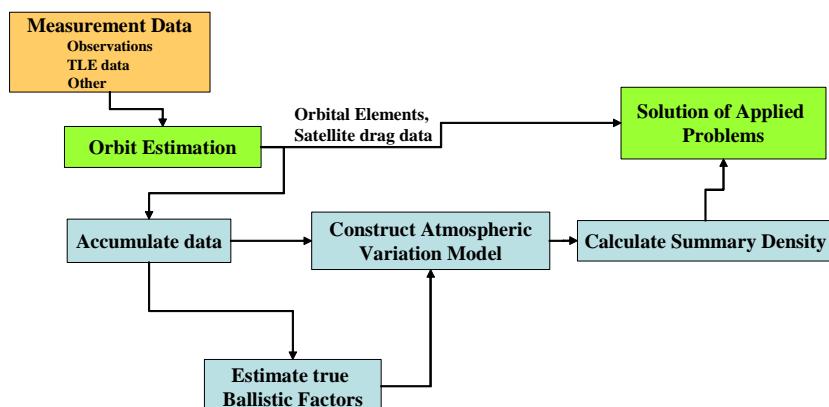


Figure 8-17. Process Flow for DCA models. This figure shows the interrelations for orbital estimation and DCA corrections. Note that the construction of the density model follows the estimation in parallel. Also, the determination of the true ballistic factors should be performed in parallel.

The model determines the density corrections about every 3 hours from a set of “calibration” satellites.* The three hour rate for density corrections is not unique. It came from a Russian consideration of determining the three hourly terms in the density from empirical inputs rather than observed geomagnetic data which they considered as unreliable in the early to mid 1980s. Note that in the recent work determining the density corrections from NORAD TLE information, the density corrections are determined once per day.

The “true” ballistic coefficient is used as an input, with processing about once every 20 days or so. This information becomes the basis of changing the atmospheric densities from an existing model—often Jacchia-Roberts-71 or an MSIS model.

The dynamic calibration work was later examined by the AFSPC (Storz, 1999). They investigated changing the temperature calculations and using spherical harmonics to extrapolate the results for global coverage. At the present time, there are a variety of AFSPC atmospheric model approaches implementing variations on the original DCA technique. Some are based on J70, while others use J71 with modifications. They use a variety of indices, some of which limit the time applicability of the model, and the performance is many times slower than the J70 model. See Marcos et al. (2006) and Bowman et al. (2006, 2008) for additional information.

Wright and Woodburn (2004) have independently shown that using a Kalman filter, you can separate the atmospheric density errors from the modeling errors and bypass the additional computations with the density, spherical harmonics, fit spans, etc. This work was limited to the simultaneous estimation of the ballistic coefficient and atmospheric density correction for each satellite independently without extension to additional satellites. The separation of estimates for two linearly dependent states was enabled by the filter formulation which allows for both variables to be unknown functions of time, but restricting their variability with time to have greatly different time scales. Thus, all the variation on short time scales becomes attributable to density corrections while a long term bias is moved into the ballistic coefficient. The filter formulation could be extended to simultaneously estimate corrections to parameters within the global density model to provide an analog to the dynamic calibration work mentioned above.

Most studies have focused on a perceived variability of the ballistic coefficient (independent from the attitude induced changes through the frontal area) because batch least squares techniques can not track the time-varying atmospheric model errors, nor the rapid and dynamic changes induced by atmospheric drag. The result is that the ballistic coefficient supposedly exhibits large variations caused by atmospheric density variations. In reality, the attitude contributes the largest portion of the error and the atmospheric density models simply add to that.

* The calibration satellites are an ad-hoc collection of LEO satellites that receive additional observational data, providing for slightly better than average orbit determination results. Selecting the calibration satellites for this technique is important because as the solar cycle raises and lowers the altitude at which atmospheric drag effects are prominent, the accuracy of the data from these calibration satellites will change.

The distinguishing feature of all the dynamic calibration approaches are that they permit the first breakthrough of the generally accepted 10-15% error in atmospheric models. All these approaches have significant applications to operational orbit determination.

Partial Derivatives

In general, we find the partial derivatives for drag using the chain rule and the density, ρ .

$$\frac{\partial \dot{\vec{r}}_{drag}}{\partial \vec{r}} = \frac{\partial \dot{\vec{r}}_{drag}}{\partial \rho} \frac{\partial \rho}{\partial \vec{r}}$$

The variety of atmospheric models would require significant space to list all the possibilities. Thus, I've only begun the general technique here. Consult Long et al. (1989:4-32) for additional information.

8.6.3 Third-Body Perturbations

In Chap. 1, we examined multiple bodies and their effects on satellites from the standpoint of determining orbital motion. We now turn to the *perturbing* effects that other bodies have on satellites. Third bodies, such as the Sun or Moon, have a greater affect on satellites in higher altitude orbits. Their effects become noticeable about when the effects of drag begin to diminish. Because the cause of perturbations from a third body (Sun and Moon) is the gravitational attraction, which is conservative, it's reasonable to use a disturbing-function solution. We can also develop the forces. I'll present both forms in this section so you can use them in different applications.

Accelerations Due to Third Body

I developed the equations of motion for a three-body system in Eq. (1-37). Let the third body be 3 and assume the mass of the satellite is negligible.

$$\ddot{\vec{r}}_{\oplus sat} = -\frac{\mu_{\oplus} \dot{\vec{r}}_{\oplus sat}}{r_{\oplus sat}^3} + \mu_3 \left(\frac{\dot{\vec{r}}_{sat3}}{r_{sat3}^3} - \frac{\dot{\vec{r}}_{\oplus 3}}{r_{\oplus 3}^3} \right) \quad (8-34)$$

At this point, we can numerically integrate to determine the third-body's effect on the satellite; however, careful examination of Eq. (8-34) reveals a potential difficulty. Suppose the disturbing third body is the Sun. The distance from the satellite to the Sun (in km) and the distance from the Earth to the Sun (also in km) are very similar, and in the final term of the acceleration, the cube of each of these distances is in the denominator. Each value will be very small; their numerically computed difference is even smaller and may introduce errors during a simulation. Long et al. (1989:4-8) state this numerical difficulty is not a problem for Earth satellites, but it can be a problem around other central bodies and may produce significant errors for the Moon. Roy (1978:202) shows how to find a solution directly.

$$Q = \frac{(r_{\oplus sat}^2 + 2(\dot{\vec{r}}_{\oplus sat} \cdot \dot{\vec{r}}_{sat3})))(r_{\oplus 3}^2 + r_{\oplus 3} r_{sat3} + r_{sat3}^2)}{r_{\oplus 3}^3 r_{sat3}^3 (\dot{\vec{r}}_{\oplus 3} + \dot{\vec{r}}_{sat3})} \quad (8-35)$$

$$\frac{\dot{\vec{r}}_{sat3}}{r_{sat3}^3} - \frac{\dot{\vec{r}}_{\oplus 3}}{r_{\oplus 3}^3} = \dot{\vec{r}}_{sat3} Q - \frac{\dot{\vec{r}}_{\oplus sat}}{r_{\oplus 3}^3}$$

Geyling and Westerman (1971:113) give us another solution using a Taylor series expansion and neglect small terms because $r_{\oplus sat} \ll r_{\oplus 3}$. Including the last term in the parenthesis [Neta (1996)] the final result is

$$\ddot{\vec{r}}_{\oplus sat} \cong -\frac{\mu_{\oplus} \dot{\vec{r}}_{\oplus sat}}{r_{\oplus sat}^3} - \frac{\mu_3}{r_{\oplus 3}^3} \left(\dot{\vec{r}}_{\oplus sat} - 3\dot{\vec{r}}_{\oplus 3} \frac{\dot{\vec{r}}_{\oplus sat} \cdot \dot{\vec{r}}_{\oplus 3}}{r_{\oplus 3}^2} - \frac{15}{2} \left(\frac{\dot{\vec{r}}_{\oplus sat} \cdot \dot{\vec{r}}_{\oplus 3}}{r_{\oplus 3}^2} \right)^2 \dot{\vec{r}}_{\oplus 3} \right) \quad (8-36)$$

Although this form is numerically stable, it's approximate (the last term in the parenthesis is often omitted) and therefore less accurate. When the third body is the Moon, the distances from the Earth to the satellite and from the Earth to the Moon are much closer than that from the Earth to the Sun, we may need more terms from the Taylor series to preserve the accuracy.

We can also express the modified form of the equations of motion using Legendre functions [Long et al. (1989:4-8 to 4-9)]. If we operate on the direct term [$\dot{\vec{r}}_{sat3} / r_{sat3}^3$ in Eq. (8-34)], we can find the magnitude of $\dot{\vec{r}}_{sat3}$ using the cosine law plus the sides of the triangle formed by the Sun, satellite, and center of the inertial coordinate system.

Designating ζ as the angle between the third body and the satellite as seen from the Earth, the cosine law gives

$$r_{sat3}^2 = r_{\oplus sat}^2 + r_{\oplus 3}^2 - 2r_{\oplus sat}r_{\oplus 3} \cos(\zeta)$$

and letting

$$h = \frac{r_{\oplus sat}}{r_{\oplus 3}} \quad B = \sum_{j=1}^{\infty} P_j [\cos(\zeta)] h^j$$

we expand $1/r_{sat3}$ in a series that is the generating function for Legendre polynomials

$$\frac{1}{r_{sat3}} = \frac{1}{r_{\oplus 3}} \{ P_0[\cos(\zeta)] + P_1[\cos(\zeta)]h + \dots \} = \frac{1+B}{r_{\oplus 3}}$$

We can substitute this into Eq. (8-34) and obtain an alternate form for the acceleration.

$$\ddot{\vec{r}}_{\oplus sat} = -\frac{\mu_{\oplus} \dot{\vec{r}}_{\oplus sat}}{r_{\oplus sat}^3} + \mu_3 \left\{ \dot{\vec{r}}_{sat3} \frac{(1+B)^3}{r_{\oplus 3}^3} - \frac{\dot{\vec{r}}_{\oplus 3}}{r_{\oplus 3}^3} \right\}$$

By expanding $(1+B)^3$

$$\ddot{\vec{r}}_{\oplus sat} = -\frac{\mu_{\oplus} \dot{\vec{r}}_{\oplus sat}}{r_{\oplus sat}^3} + \mu_3 \left\{ \dot{\vec{r}}_{sat3} \frac{(1 + 3B + 3B^2 + B^3)}{r_{\oplus 3}^3} - \frac{\dot{\vec{r}}_{\oplus 3}}{r_{\oplus 3}^3} \right\}$$

Recognizing that $\dot{\vec{r}}_{sat3} - \dot{\vec{r}}_{\oplus 3} = -\dot{\vec{r}}_{\oplus sat}$ and distributing signs, we have

$$\ddot{\vec{r}}_{\oplus sat} = -\frac{\mu_{\oplus} \dot{\vec{r}}_{\oplus sat}}{r_{\oplus sat}^3} - \mu_3 \left\{ \frac{-\dot{\vec{r}}_{sat3}(3B + 3B^2 + B^3) + \dot{\vec{r}}_{\oplus sat}}{r_{\oplus 3}^3} \right\} \quad (8-37)$$

Now, let's generalize this expression to multiple disturbing third bodies, $k = 1, 2, \dots, n$. Thus,

$$\ddot{\vec{r}}_{\oplus sat} = -\frac{\mu_{\oplus}}{r_{\oplus sat}^3} \dot{\vec{r}}_{\oplus sat} - \sum_{k=1}^n \frac{\mu_k}{r_{\oplus k}^3} (\dot{\vec{r}}_{\oplus sat} - \beta_k \dot{\vec{r}}_{sat k}) \quad (8-38)$$

where $\beta_k = 3B_k + 3B_k^2 + B_k^3$ because $B_k = B(\zeta_k)$. This form eliminates the numerical difficulty in Eq. (8-34) when $r_{\oplus k} \approx r_{sat k}$, but only at the expense of introducing an infinite series in the formulation. Fortunately, this series converges rapidly for those satellites where $r_{\oplus sat} \ll r_{\oplus 3}$, which is the case for most satellite missions. Thus a few terms suffice for most practical applications.

We find the disturbing potential function by expressing the third-body potential in terms of the Legendre polynomials.

$$R_{3-body} = \frac{\mu_3}{r_3} \sum_{l=2}^{\infty} \left(\frac{r}{r_3} \right)^l P_l[\cos(\zeta)]$$

Here, r is the geocentric distance, and μ_3 and r_3 are the third-body gravitational parameters and geocentric distance, respectively. We've already introduced the angle between the position vectors to the satellite and the Sun, ζ .

We use the addition theorem with which we developed the gravitational potential in Eq. (8-15), to obtain

$$P_l[\cos(\zeta)] = \sum_{m=0}^l k_m \frac{(l-m)!}{(l+m)!} P_{l,m}[\sin(\delta)] P_{l,m}[\sin(\delta_3)] \cos(m(\alpha - \alpha_3))$$

The index, k_m , assumes values of 1 for $m = 0$, and 2 for $m \neq 0$, and α, δ and α_3, δ_3 are the right ascension and declination of the satellite and the third body respectively. Using Eq. (C-16) for $\cos(m(\alpha - \alpha_3))$, define

$$A_{t,m} = \frac{\mu_3}{r_3^{t+1}} k_m \frac{(t-m)!}{(t+m)!} P_{t,m} [\sin(\delta_3)] \cos(m\alpha_3)$$

$$B_{t,m} = \frac{\mu_3}{r_3^{t+1}} k_m \frac{(t-m)!}{(t+m)!} P_{t,m} [\sin(\delta_3)] \sin(m\alpha_3)$$

Notice that we group the parameters relating to the third body into $A_{t,m}$ and $B_{t,m}$ as we did in the development of the Earth's gravitational potential. These coefficients are analogous to the $C_{t,m}$ and $S_{t,m}$ coefficients for the geopotential, but they depend on the physical location of the third body. Thus, they reflect the distribution of mass over the orbital motion of the third body, not the shape of the third body itself. We now have

$$R_{3-body} = \sum_{t=2}^{\infty} \sum_{m=0}^t r^t P_{tm} [\sin(\delta)] \{ A_{t,m} \cos(m\alpha) + B_{t,m} \sin(m\alpha) \} \quad (8-39)$$

Partial Derivatives

We require partial derivatives of the third-body acceleration for orbit-determination and estimation techniques in Chap. 10. Begin with Eq. (8-34) and assume the Sun is the third body. Consider first the partial derivative with respect to position $\dot{\vec{r}}_{\oplus sat}$:

$$\begin{aligned} \frac{\partial \ddot{\vec{r}}}{\partial \dot{\vec{r}}_{\oplus sat}} &= -\mu_{\oplus} \left[\frac{\partial (1/r_{\oplus sat}^3)}{\partial \dot{\vec{r}}_{\oplus sat}} \dot{\vec{r}}_{\oplus sat} + \frac{1}{r_{\oplus sat}^3} \frac{\partial \dot{\vec{r}}_{\oplus sat}}{\partial \dot{\vec{r}}_{\oplus sat}} \right] \\ &\quad + \mu_{\odot} \left[\frac{\partial (1/r_{sat\odot}^3)}{\partial \dot{\vec{r}}_{\oplus sat}} \dot{\vec{r}}_{sat\odot} + \frac{1}{r_{sat\odot}^3} \frac{\partial \dot{\vec{r}}_{sat\odot}}{\partial \dot{\vec{r}}_{\oplus sat}} \right] \end{aligned} \quad (8-40)$$

Note that the indirect term $(\dot{\vec{r}}_{\oplus\odot}/r_{\oplus\odot}^3)$ in Eq. (8-34) is independent of the satellite's position and velocity; thus, the partial derivative vanishes. Clearly, we'll need to use the identities $\partial r / \partial \dot{\vec{r}} = \dot{\vec{r}} / r$ and $\partial \dot{\vec{r}} / \partial \dot{\vec{r}} = I$, where I is a 3×3 identity matrix. Also note that because the acceleration and position vectors both have three components, we expect a 3×3 matrix as the result. Furthermore, in view of the identity, the partial derivative in the first term of Eq. (8-40) will produce a result proportional to $\dot{\vec{r}}_{\oplus sat}$ resulting in the apparent product of two vectors $\dot{\vec{r}}_{\oplus sat} \dot{\vec{r}}_{\oplus sat}^T$. This isn't a dot product. Rather, it's the exterior product. If we assume the usual convention of a column representation for a vector, then the exterior product of two vectors is defined as the product of an $n \times 1$ column vector with a $1 \times m$ row vector, which results in an $n \times m$ matrix. We use this notation to represent our result. Evaluation of the partial derivatives in Eq. (8-40) yields

$$\frac{\partial (1/r_{\oplus sat}^3)}{\partial \dot{\vec{r}}_{\oplus sat}} = \frac{-3}{r_{\oplus sat}^4} \frac{\partial r_{\oplus sat}}{\partial \dot{\vec{r}}_{\oplus sat}} = \frac{-3}{r_{\oplus sat}^5} \dot{\vec{r}}_{\oplus sat}$$

$$\frac{\partial(1/r_{sat\odot}^3)}{\partial \dot{\vec{r}}_{\oplus sat}} = \frac{-3}{r_{sat\odot}^4} \frac{\partial r_{sat\odot}}{\partial \dot{\vec{r}}_{\oplus sat}}$$

Further evaluation and simplification yield the final result. Notice that, because the acceleration is independent of the velocity, the partial derivatives are zero. I've shown the general form using "k" disturbing bodies. A single third body uses k and $i = 1$.

$$\begin{aligned} \frac{\ddot{\vec{r}}_{\oplus sat}}{\partial \dot{\vec{r}}_{\oplus sat}} &= \left\{ -\frac{\mu_{\oplus}}{r_{\oplus sat}^3} - \sum_{i=1}^k \mu_i \frac{1}{r_{sat_i}^3} \right\} \mathbf{I} + 3 \left(\frac{\mu_{\oplus} \dot{\vec{r}}_{\oplus sat} \dot{\vec{r}}_{\oplus sat}^T}{r_{\oplus sat}^5} + \sum_{i=1}^k \mu_i \frac{\dot{\vec{r}}_{sat_i} \dot{\vec{r}}_{sat_i}^T}{r_{sat_i}^5} \right) \\ \frac{\ddot{\vec{r}}_{\oplus sat}}{\partial \dot{\vec{r}}_{\oplus sat}} &= 0 \end{aligned}$$

8.6.4 Solar-Radiation Pressure

Like drag, solar-radiation pressure is a nonconservative perturbation that becomes more pronounced at higher altitudes. One of the more difficult aspects of analyzing solar radiation is accurately modeling and predicting the solar cycles and variations. During periods of intense solar storms, this effect may be much larger than all the other perturbations (depending on the altitude); at times of low activity, the effect may be negligible. The same difficulties arise for cross-sectional area as for drag; however, solar-radiation pressure also requires us to determine the shadowing effect on the spacecraft.

To examine the overall effect of solar radiation, we must first develop expressions for the specific force (acceleration) and how it is measured. Because the incoming radiation from the Sun exerts a force on the satellite, the apparent size of the satellite that faces the Sun is crucial in accurately determining the amount of acceleration. The pressure is simply the force divided by the incident area exposed to the Sun. This means that the pressure distribution is very critical, and this depends on the satellite's shape and composition (a balloon versus a cannonball, for instance). Incorporating the mass then permits us to determine the acceleration. The entire process involves determining the Sun's precise location; the correct satellite orbital attitude; the exact value of the solar-radiation pressure; the effective, time-varying, cross-sectional area exposed to the incoming radiation; and the correct and usually time-varying coefficients that model the satellite's reflectivity. I've discussed the first two items in calculating drag and third-body effects.

To arrive at the solar-radiation pressure, begin with the intensity of the energy of the incoming radiation from the Sun. Baker (1967:186) states that 8×10^{17} photons / cm^2 with $\lambda_{avg} = 5560 \text{ \AA}$ reach the Earth with a frequency, f , at the mean Earth-Sun distance. Each photon has a certain amount of energy, hf (Planck's constant is $h = 6.6256 \times 10^{-34} \text{ J.s}$). A *solar-radiation constant*—often called the *intensity, irradiance*, or *solar flux* is

$$SF = 1367 \text{ W/m}^2$$

Many programs use this value (or another similar value) for the solar constant because determining the actual frequencies and energy is very difficult and varies over time. Sources of current values are listed in Appendix D, and Fig. 8-18 shows historical values. Values are sometimes given in Watts (e.g., 3.823×10^{26}) which we convert to Watts/m^2 by dividing by 4π times the Earth distance from the Sun squared— $(4\pi)(149,597,870,000)^2$.

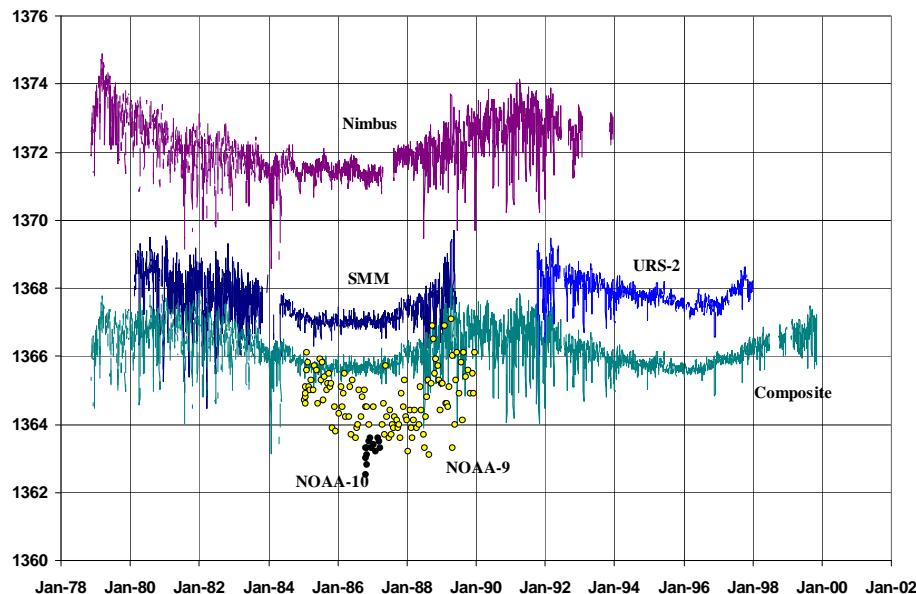


Figure 8-18. Solar Irradiance Values. Data from several satellites is shown for solar irradiance measurements. Notice the variability of values and lack of definitive trends. Solar Maximum Mission (SMM) and Upper Atmospheric Research Satellite (UARS) satellite data are also included. Units are Watts/m^2 .

Many of the same factors affecting atmospheric-density calculations (27-day, diurnal, etc.) also affect the magnitude of solar-radiation pressure. Wertz (1978:130) gives a time-varying formula instead of the constant value, accounting for the variation over a year (notice the different solar irradiance value):

$$SF = \frac{1358}{1.004 + 0.0334 \cos(D_{aphelion})} \frac{\text{W}}{\text{m}^2} \quad (8-42)$$

where $D_{aphelion}$ is 2π times the days from when the Earth is at aphelion, as a fraction of the whole year. This aphelion point usually occurs about July 4, although it does vary. For instance, in 1996, it was July 6.

We must still determine the effect of the pressure, or the amount of momentum, imparted. Using Einstein's law relating energy with mass, $E = mc^2$, and with c being the speed of light, we can solve for the momentum, mc , and find

$$mc = \frac{E}{c}$$

which allows us to find the force of *solar pressure*, p_{srp} , per unit area, or the change in momentum:

$$p_{srp} = \frac{1367 \text{ W/m}^2}{3 \times 10^8 \text{ m/s}} = 4.57 \times 10^{-6} \frac{\text{W} \cdot \text{s}}{\text{m}^3} = 4.57 \times 10^{-6} \frac{\text{N}}{\text{m}^2}$$

Remember that this value will change if we use more precise values of the solar-radiation intensity, as is possible with Eq. (8-42).

Now we can develop expressions for the force. Using the *reflectivity*, c_R , the *solar-radiation pressure*, p_{srp} , and the *exposed area to the Sun*, A_\odot , the force is

$$\vec{F}_{srp} = -p_{srp} c_R A_\odot \frac{\vec{r}_\oplus \odot}{|\vec{r}_\oplus|} \quad (8-43)$$

Notice the symbol for the area facing the Sun has a subscript because it's usually *not* the same as the projected frontal area used in calculations of drag. The *reflectivity*, c_R , is a value between 0.0 and 2.0, which indicates how the satellite reflects incoming radiation. A value of 0.0 means the object is translucent to incoming radiation. No momentum is transmitted (and therefore no force), but there may be some refraction. A value of 1.0 means that all the radiation is absorbed, and all the momentum is transmitted (i.e., it's a black body) resulting in a perturbing force. Finally, 2.0 indicates that all the radiation is reflected and twice the momentum is transmitted to the satellite (i.e., a flat mirror perpendicular to the light source). In this case, the resulting force is twice the absorption force due to Newton's first law. It turns out that determining c_R properly is extremely difficult. It changes over time and is virtually impossible to predict. This is especially true for complex satellites made of various materials, that enter and exit eclipse regions, and have a constantly changing orientation. For this reason, it's almost always a solution parameter, determined in differential correction (just like the drag coefficient).

It's common to assume the surface maintains a constant attitude perpendicular to the Sun, thereby setting $\phi_{mc} = 0^\circ$ (See Fig. 8-19). Though not very realistic, it does give us rough estimates for the problem of solar radiation. Burns et al. (2000) describe an attitude guidance law which yields a time-varying orientation. Newton's second law permits us to determine the acceleration with the given force:

$$a_{srp} = \frac{F_{srp}}{m} = \frac{p_{srp} c_R A_\odot}{m}$$

Because the direction is always away from the Sun, a unit vector from the satellite to the Sun yields the correct direction (the sign is positive if $\vec{r}_{\odot sat}$ is used instead). The acceleration commonly used for numerical simulations is

$$\ddot{\vec{a}}_{srp} = - \frac{p_{srp} c_R A_\odot}{m} \frac{\dot{\vec{r}}_{sat\odot}}{|\dot{\vec{r}}_{sat\odot}|} \quad (8-44)$$

In forming Eq. (8-44), we made several assumptions concerning satellite attitude and other parameters. Satellites have complex geometries. Some surfaces reflect diffusely; others reflect specularly and with changing aspect toward the Sun. Hence, at any moment, the satellite will experience a net force, *not* along the Sun-satellite vector, plus a net torque. This implies the possibility of a *solar sail*, in which solar-radiation pressure becomes the “wind” to move the satellite in a “low-thrust-like” maneuver. Because we’re interested in orbital motion, we often ignore the torque and make the simplifying assumption that the disturbing acceleration is given by Eq. (8-44), where A_\odot is an average effective cross-section that implicitly incorporates c_R .

In reality though, the reflection process is two-fold: absorption followed by reflection. Figure 8-19 shows the forces imparted by absorption and reflection. Remember that the absorbed and reflected forces aren’t orthogonal, so they don’t combine the same as normal and tangential expressions. Burns et al. (2000) shows additional formulations for calculating different surfaces, and how to analyze these forms through VOP analysis (Sec. 9.3.2).

There are also two kinds of reflection—*specular* and *diffuse*. We won’t discuss absorbed reflection because this requires us to have complex three-dimensional models of the satellite. Now consider solar radiation incident on a surface element of area A_\odot , whose normal makes a *solar-incidence angle*, ϕ_{inc} , with the Sun-satellite line. The individual forces are (absorbed and reflected: specular and diffuse), and they occur in the normal \hat{n} and source incident \hat{s} directions.

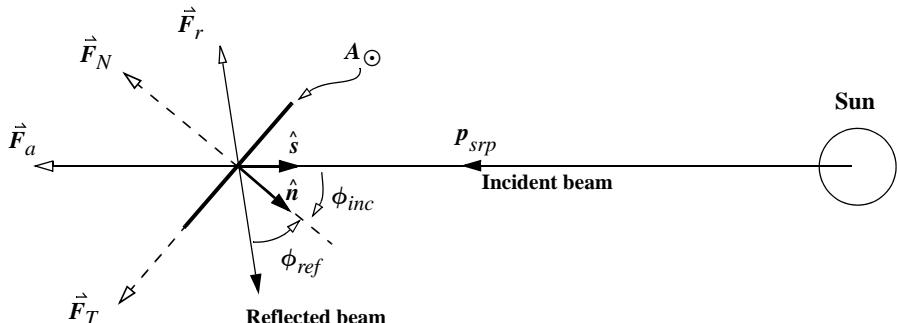


Figure 8-19. Incident Solar Radiation. The incident radiation produces one reflected beam resulting in reflected, F_r , and absorbed, F_a , forces. The reflected, ϕ_{ref} , and incident, ϕ_{inc} , angles are the same. We can also examine the interaction of the incoming radiation by looking at the normal, F_N , and tangential, F_T , force components. The satellite area that’s exposed to the Sun is depicted as a flat plate.

$$\begin{aligned}\vec{F}_{rs} &= -2p_{srp}c_{Rs}A_{\odot} \cos^2(\phi_{inc})\hat{n} \\ \vec{F}_a &= -p_{srp}c_{Ra}A_{\odot} \cos(\phi_{inc})\hat{s} \\ \vec{F}_{rd} &= -p_{srp}c_{Rd}A_{\odot} \cos(\phi_{inc})\left\{\frac{2}{3}\hat{n} + \hat{s}\right\}\end{aligned}$$

We can model diffuse and specular radiation forces by assuming a Lambertian diffusion (Luthcke et al. 1997). Summing the components of a macro-model

$$\vec{a}_{srp} = -\sum_{i=1} \frac{p_{srp}A_{\odot i} \cos(\phi_{inc_i})}{m} \left\{ 2\left(\frac{c_{Rd_i}}{3} + c_{Rs_i} \cos(\phi_{inc_i})\right)\hat{n} + (1 - c_{Rs_i})\hat{s} \right\} \quad (8-45)$$

In this relation, we use the incoming solar pressure, p_{srp} , the angle between the surface normal and the incoming radiation, ϕ_{inc} , and the satellite mass, m . We also require the diffuse, c_{Rd} , and specular, c_{Rs} , reflectivities (Note that $c_{Ra} + c_{Rd} + c_{Rs} = 1.0$). The surface normal vector and the source (Sun) incidence vectors are needed for orienting the satellite. Macro-models approximate the shape and orientation of the satellite to more precisely find the acceleration. The summation adds each flat plate surface of the satellite model. The i -subscripts indicate that the area, angles, and coefficients will vary for each macro-model plate.

ROCK Solar Radiation Pressure Models

Well-known examples of the solar radiation pressure modeling discussed previously are the ROCK4 and ROCK42 models for GPS satellites (Fliegel et al. 1992). These models were originally developed by Rockwell (hence the name) and included methods to account for various surfaces on the satellites, and precise attitude. Because the programs were complex, Fourier series representations were developed to capture the essential behavior. Figure 8-20 shows the general orientation and nomenclature for the GPS body coordinate system. The angle formed by the Sun-satellite-Z axis is important with these models.

There are perturbing accelerations in both the X and Z directions. A y -bias also exists, which yields a smaller perturbing force. Several equations are relevant. Each block of GPS satellites have a specific set of equations. Standard (S_{10} and S_{20}) models match the AF implementation, and thermal models (T_{10} and T_{20}) also exist. The accelerations are as follows. The angle α is measured from the Sun to the Z -axis.

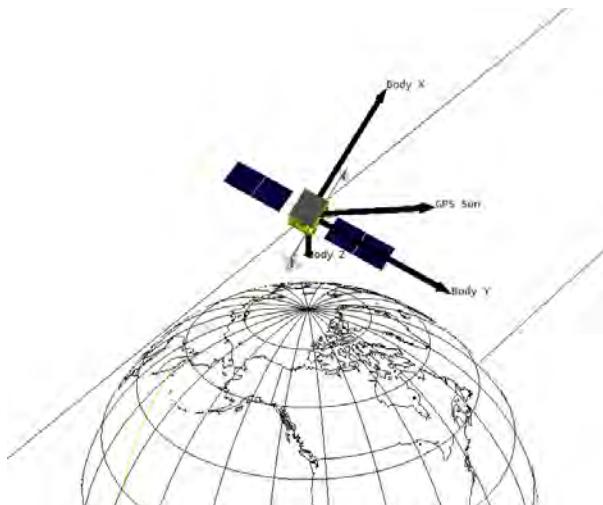


Figure 8-20. GPS body coordinate system. The orientation of a GPS satellite is shown with the satellite body coordinate system. The Z-axis points towards the Earth and the Y-axis lies along a solar panel arm, normal to the Sun-satellite vector. The X-axis completes a right handed coordinate system. (STK images courtesy of AGI).

$$S_{10}$$

$$X = -4.34 \sin(\alpha) + 0.1 \sin(2\alpha + 1.1) - 0.05 \cos(4\alpha) + 0.06$$

$$Z = -4.34 \cos(\alpha) + 0.17 \sin(2\alpha - 0.4) - 0.05 \sin(4\alpha) - 0.06$$

$$T_{10}$$

$$X = -4.55 \sin(\alpha) + 0.08 \sin(2\alpha + 0.9) - 0.06 \cos(4\alpha + 0.08) + 0.08$$

$$Z = -4.54 \cos(\alpha) + 0.20 \sin(2\alpha - 0.3) - 0.03 \sin(4\alpha)$$

(8-46)

$$S_{20}$$

$$X = -8.10 \sin(\alpha) + 0.05 \cos(2\alpha) - 0.056 \sin(4\alpha + 1.4) + 0.07$$

$$Z = -7.80 \cos(\alpha) + 0.024 \sin(2\alpha - 0.8) - 0.047 \sin(4\alpha + 0.9) + 0.02$$

$$T_{20}$$

$$X = -8.96 \sin(\alpha) + 0.16 \sin(3\alpha) + 0.10 \sin(5\alpha) - 0.07 \sin(7\alpha)$$

$$Z = -8.42 \cos(\alpha)$$

Earth Albedo and Infrared Radiation

Solar radiation that immediately reflects off the Earth back onto a satellite is called **albedo**. The amount of reflected radiation is about 30% of the incoming solar radiation. The remaining solar radiation is absorbed and re-emitted at a later time as infrared radiation, or **emissivity**. The infrared energy emanating at the Earth's surface (*IR*) is about 237 W/m^2 . Both effects can be measurable on some satellites. Unlike solar radiation

pressure, they are usually separated into specific wavelengths—each of which transfer through the atmosphere differently. We find the acceleration as before [Eq. (8-45)], but now we must first calculate the ground albedo and emissivity, and also consider each region of the Earth that is visible to the satellite, or a satellite plate model component.

The albedo and emissivity at the ground are shown by Knocke et al. (1988). They use a spherical harmonic expansion (limited to zeroth, first, and second degree, Table 8-2) to represent the functions worldwide. The generic expression is

$$Pradiation = a_0 + [c_0 + c_1 \cos(\omega(JD - t_o)) + c_2 \sin(\omega(JD - t_o))]P(\cos(\phi)) + a_2 P(\sin(\phi))$$

where ϕ is the central latitude of the grid, and ω establishes the periodicity, $2\pi/365.25$.* The base epoch (t_o) is December 22, 1981 ($JD = 2,444,960.5$). They give expressions for albedo and emissivity.

$$p_{alb} = 0.34 + [0.10 \cos(\omega(JD - t_o))] \cos(\phi) + 0.29 \sin(\phi) \quad (8-47)$$

$$p_{emiss} = 0.68 + [-0.07 \cos(\omega(JD - t_o))] \cos(\phi) - 0.18 \sin(\phi)$$

To find the acceleration on a satellite, you must sum the accelerations induced on each component of a macro model, and from each grid location. Note the similarity with Eq. (8-45). For a single wavelength with albedo pressure (p_{alb})

$$albedo = -\sum_i \sum_j \frac{p_{alb} A_{ij} \cos(\phi_{inc_{ij}})}{mc} \left\{ 2 \left(\frac{c_{Rd_i}}{3} + c_{Rs_i} \cos(\phi_{inc_{ij}}) \right) \hat{n} + (1 - c_{Rs_i}) \hat{s} \right\}$$

The i components are the macro plate models, while the j indices represent grid points on the Earth. STK/HPOP documentation suggests that for calculations within about 2% of the actual magnitude on satellites above 200 km altitude, you can use a grid spacing on the Earth of about 10° , decreasing to about 1° below 300 km altitude. Note that the area, incident angle, and incoming radiation will vary for each Earth location (j) considered. Denser grids produce more accurate results, but at significantly higher processing times. You can account for the infrared radiation as an additional perturbing force using the same equation, and replacing p_{alb} with p_{emiss} .

A simple expression may be found from sample data in Wertz (1978:571). The albedo flux is $-197.65 \ln(h_{ellp}) + 1833.3 \text{ W/m}^2$, and the emissivity flux is $-39.451 \ln(h_{ellp}) + 391.44 \text{ W/m}^2$.

8.6.5 Other Perturbations

Many other forces affect a satellite's orbit, but most are very small and are often neglected. Tides are being modeled more often as computational resources expand and

* Note that longitude is not included as the variations for longitude are generally significantly less than those for latitude.

allow us to consider them. Thrust is a large perturbation, but unlike the other forms we've discussed, it's planned. Upper atmosphere winds are under investigation as accuracies have improved over the years (Sutton et al. 2006).

Tides

Although the first mention of tides is generally associated with ocean tides, the term also includes solid Earth tides, and to a lesser extent, pole tides, and atmospheric tides from air mass movement. Tides are complex and have only recently been studied in detail as computational and observational processes have become readily accessible. In general, tides are a result of a gravitational distortion caused by an external body—for the Earth, the Moon and Sun are the primary sources. The Earth rotation introduces periodicity into these effects.

We've seen that the orbits for the Sun and Moon are complex, and this requires us to consider many different periods in any analysis. The periods are related to the Earth's diurnal motion, the Sun's rotation, the Moon's motions, and so on. Each period has a repetitive frequency, so it's common to model the tidal potential as a harmonic expression, similar to the Earth's aspherical potential. The number of harmonics required to accurately model each of the forces varies from about 13 for six months, to more than 500 for periods of several years and longer.

Many sources define the tide generating potential (TGP) similar to the gravity potential we found in Eq. (8-21). The tidal potential is not directly observable, but we can derive it from measurements of the vertical and horizontal components of gravity. It's common to analyze the potential as a function of time, which involves determining the equilibrium behavior of a solid, spherical, oceanless Earth. The three parameters used here are the Love numbers, h , k , and l . The Love numbers let us examine the Earth's response to perturbations from a frequency perspective. Although specific forms of equations differ, they stem from many of the same sources. See Jursa (1985:23-24 to 23-29), Kaula (1969), and Casotto (1989) for more information.

Solid-Earth tides are deformations of the Earth due to perturbing forces of external gravitational attraction, particularly the Moon. They are generally the largest perturbing effects of tides discussed in this section. Internal forces also result from the Earth's interior structure and involve complex models of the motions for the liquid and solid properties of the matter within the Earth. These forces are beginning to be used routinely with numerical propagation. McCarthy and Petit (2003:59) show the basic formulae for implementation using corrections to the unnormalized gravitational coefficients. Note the use of the Love numbers ($k_{l,m}$) and summations for the Moon ($j = 2$) and Sun ($j = 3$).

$$\Delta C_{l,m} = \frac{k_{l,m}}{2l+1} \sum_{j=2}^3 \frac{\mu_j}{\mu_{\oplus}} \left(\frac{R_{\oplus}}{r_j} \right)^{l+1} P_{l,m} [\sin(\phi_{gc_j})] \cos(m\lambda_j)$$

$$\Delta S_{l,m} = \frac{k_{l,m}}{2l+1} \sum_{j=2}^3 \frac{\mu_j}{\mu_{\oplus}} \left(\frac{R_{\oplus}}{r_j} \right)^{l+1} P_{l,m} [\sin(\phi_{gc_j})] \sin(m\lambda_j)$$
(8-48)

These corrections are determined for all values of m (0 to ℓ) for $\ell = 2$ and 3, and m (0 to 2) for $\ell = 4$. McCarthy and Petit (2003:60) list the nominal values. Note that for $\ell = 4$, alternate Love coefficients are used, and the Legendre polynomial has $\ell = 2$.

The centrifugal effect of the Earth's rotation also causes a solid Earth **pole tide**. A potential could be found, but the result is ultimately a change to the C_{21} and S_{21} coefficients. The parameters m_1 and m_2 are usually in arcseconds.

$$\Delta C_{21} = -1.333 \times 10^{-9} (m_1 - 0.0115 m_2)$$

$$\Delta S_{21} = -1.333 \times 10^{-9} (m_2 + 0.0115 m_1)$$

The tide generating potential has a zonal component ($\ell = 2$) that is non-zero. Thus, when using gravitational models such as JGM-3, you must first restore the permanent portion of the tidal effect in the model. See McCarthy and Petit (2003:67) for the procedure.

Ocean tides cause a large change in mass distribution as the water reacts to the various gravitational attractions. They are also the main reason for the differences between tidal theory and observational measurements. Casotto (1989) suggests that ocean tides are only about 10-15% of the magnitude of solid Earth tides. The ocean tides affect: (1) the direct attraction of the water, (2) the physical bending of the Earth's crust due to the mass of the water, and (3) the distortion of the geoid surface resulting from the previous effects. The perturbative effect of the oceans has been understood conceptually for quite some time; however, only recently have modern computers permitted realistic analysis of the perturbation. Developing correct tidal models is complicated because the resonant frequencies of many of the tidal basins, ocean basins, and so forth, are nearly the same as the overall tidal motion. In addition, swells and troughs associated with large currents (like the Gulf Stream) introduce systematic departures from ideal predictions of tidal models.

From McCarthy and Petit (2003:67), we find the corrections to the gravitational coefficients. Note the use of the density of seawater ($\rho_w = 1.025 \text{ gm/m}^3$), and the load deformation coefficients $k'_{\ell, m}$.

$$\Delta C_{\ell,m} = \frac{4\pi R_\oplus^2}{m_\oplus} \rho_w \left(\frac{1 + k'_{\ell,m}}{2\ell + 1} \right) \sum_{j=1}^2 \frac{\mu_j}{\mu_\oplus} \left(\frac{R_\oplus}{r_j} \right)^{\ell+1} P_{\ell,m} \sin(\phi_{gc_j}) \cos(m\lambda_j)$$

$$\Delta S_{\ell,m} = \frac{k'_{\ell,m}}{2\ell + 1} \sum_{j=1}^2 \frac{\mu_j}{\mu_\oplus} \left(\frac{R_\oplus}{r_j} \right)^{\ell+1} P_{\ell,m} \sin(\phi_{gc_j}) \sin(m\lambda_j)$$

A primary problem remaining today is the accurate global measurement of tides using tide gauges. Correctly placing these gauges is crucial, but experience has shown that almost any finite set of locations requires undesirable trade-offs. Today, most measurements for tide models come from satellite altimeter data. A second major problem is our limited understanding of the world's oceans, especially the current effects and interactions at the coasts.

There is a coupling between solid Earth and ocean tides. The ocean tides cause mass redistribution, and a resultant loading on the crust. This causes a regional deformation to the crust. The Coriolis force of Earth rotation also induces a tidal deformation due to centrifugal acceleration. The exact dynamics of the Earth's core, mantle, and crust are not fully understood, but there are numerous wobbles that occur in each. Resonances occur in these wobbles, introducing additional phases to account in any response. Some of these periods range from hours to years. There is also a correlation with the nutation calculations discussed in Chap. 3.

Mathematically modeling tides usually involves a three-step process: (1) Calculate the tidal variations resulting from the Sun and Moon, (2) Update the corrections for smaller effects, and (3) Adjust for zero-tide solution if necessary.

Magnetic Field Effects

We get the field strength of the magnetic field, \vec{B} , if we want to calculate the perturbative effects of the charged particles on the satellite (such as magnetic torquing). It isn't needed for analyzing atmospheric density. Notice the use of a spherical harmonic representation of the magnetic-potential function, V . To be consistent with the geopotential function, I show the magnetic field as the positive gradient of the potential function.

$$\vec{B} = \nabla V$$

$$V = -R_{\oplus} \sum_{l=1}^k \left(\frac{R_{\oplus}}{r}\right)^{l+1} \sum_{m=0}^l \bar{P}_{l,m} [\cos(90^\circ - \phi_{gc})] \{ \bar{g}_{l,m} \cos(m\lambda) + \bar{h}_{l,m} \sin(m\lambda) \} \quad (8-49)$$

where the variables are as defined with the gravitational potential. The bars represent normalized parameters [as in Eq. (8-22)], which we'll discuss shortly. The *Schmidt functions*, $\bar{P}_{l,m}[\cos(90^\circ - \phi_{gc})]$, combine the normalization with the Legendre functions. Note that we need more coefficients if we consider magnetic currents above the Earth as well as those arising only from the Earth. The values in Table 8-6 approximate the *Schmidt coefficients*, $\bar{g}_{l,m}$ and $\bar{h}_{l,m}$. See the *International Geomagnetic Reference Field* (IGRF) for current data. We use these coefficients for about five years. Derivatives of the coefficients permit accurate extrapolations.

It's common to normalize as with the gravitational coefficients, but the magnetic-field calculation only partly normalizes, and we refer to the resulting parameters as *Gaussian coefficients*, in honor of Karl Friedrich Gauss.

$$\begin{aligned} \Pi_{l,m}^{Schmidt} &= \sqrt{\frac{(2 - \delta_{0,m})(l-m)!}{(l+m)!} \frac{(2l-1)!!}{(l-m)!}} \\ \bar{g}_{l,m} &= \Pi_{l,m}^{Schmidt} g_{l,m} & \bar{h}_{l,m} &= \Pi_{l,m}^{Schmidt} h_{l,m} \\ \delta_{l,m} &= 1 \text{ if } m = l \\ \delta_{l,m} &= 0 \text{ if } m \neq l \end{aligned}$$

TABLE 8-6. Gaussian Geomagnetic Coefficients. Example Schmidt coefficients are listed for a 13×13 magnetic field. Consider these values valid for 2010–2015. Appendix D lists Internet sites for current values.

l	m	$\bar{g}_{t,m}$	$\bar{h}_{t,m}$												
1	0	-29496.5	0.0	6	6	-77.9	54.9	9	8	-8.4	-6.1	12	1	-0.2	-0.8
1	1	-1585.9	4945.1	7	0	80.4	0.0	9	9	-10.1	7.0	12	2	0.3	0.3
2	0	-2396.6	0.0	7	1	-75.0	-57.8	10	0	-2.0	0.0	12	3	1.0	2.2
2	1	3026.0	-2707.7	7	2	-4.7	-21.2	10	1	-6.3	2.8	12	4	-0.7	-2.5
2	2	1668.6	-575.4	7	3	45.3	6.6	10	2	0.9	-0.1	12	5	0.9	0.5
3	0	1339.7	0.0	7	4	14.0	24.9	10	3	-1.1	4.7	12	6	-0.1	0.6
3	1	-2326.3	-160.5	7	5	10.4	7.0	10	4	-0.2	4.4	12	7	0.5	0.0
3	2	1231.7	251.7	7	6	1.6	-27.7	10	5	2.5	-7.2	12	8	-0.4	0.1
3	3	634.2	-536.8	7	7	4.9	-3.4	10	6	-0.3	-1.0	12	9	-0.4	0.3
4	0	912.6	0.0	8	0	24.3	0.0	10	7	2.2	-4.0	12	10	0.2	-0.9
4	1	809.0	286.4	8	1	8.2	10.9	10	8	3.1	-2.0	12	11	-0.8	-0.2
4	2	166.6	-211.2	8	2	-14.5	-20.0	10	9	-1.0	-2.0	12	12	0.0	0.8
4	3	-357.1	164.4	8	3	-5.7	11.9	10	10	-2.8	-8.3	13	0	-0.2	0.0
4	4	89.7	-309.2	8	4	-19.3	-17.4	11	0	3.0	0.0	13	1	-0.9	-0.8
5	0	-231.1	0.0	8	5	11.6	16.7	11	1	-1.5	0.1	13	2	0.3	0.3
5	1	357.2	44.7	8	6	10.9	7.1	11	2	-2.1	1.7	13	3	0.4	1.7
5	2	200.3	188.9	8	7	-14.1	-10.8	11	3	1.6	-0.6	13	4	-0.4	-0.6
5	3	-141.2	-118.1	8	8	-3.7	1.7	11	4	-0.5	-1.8	13	5	1.1	-1.2
5	4	-163.1	0.1	9	0	5.4	0.0	11	5	0.5	0.9	13	6	-0.3	-0.1
5	5	-7.7	100.9	9	1	9.4	-20.5	11	6	-0.8	-0.4	13	7	0.8	0.5
6	0	72.8	0.0	9	2	3.4	11.6	11	7	0.4	-2.5	13	8	-0.2	0.1
6	1	68.6	-20.8	9	3	-5.3	12.8	11	8	1.8	-1.3	13	9	0.4	0.5
6	2	76.0	44.2	9	4	3.1	-7.2	11	9	0.2	-2.1	13	10	0.0	0.4
6	3	-141.4	61.5	9	5	-12.4	-7.4	11	10	0.8	-1.9	13	11	0.4	-0.2
6	4	-22.9	-66.3	9	6	-0.8	8.0	11	11	3.8	-1.8	13	12	-0.3	-0.5
6	5	13.1	3.1	9	7	8.4	2.2	12	0	-2.1	0.0	13	13	-0.3	-0.8
l	m	dg/dt	dh/dt												
1	0	11.4	0.0	4	2	-8.9	3.2	6	2	-0.3	-2.1	7	6	-0.8	-0.3
1	1	16.7	-28.8	4	3	4.4	3.6	6	3	1.9	-0.4	7	7	0.4	0.2
2	0	-11.3	0.0	4	4	-2.3	-0.8	6	4	-1.6	-0.5	8	0	-0.1	0.0
2	1	-3.9	-23.0	5	0	-0.5	0.0	6	5	-0.2	0.8	8	1	0.1	0.0
2	2	2.7	-12.9	5	1	0.5	0.5	6	6	1.8	0.5	8	2	-0.5	0.2
3	0	1.3	0.0	5	2	-1.5	1.5	7	0	0.2	0.0	8	3	0.3	0.5
3	1	-3.9	8.6	5	3	-0.7	0.9	7	1	-0.1	0.6	8	4	-0.3	0.4
3	2	-2.9	-2.9	5	4	1.3	3.7	7	2	-0.6	0.3	8	5	0.3	0.1
3	3	-8.1	-2.1	5	5	1.4	-0.6	7	3	1.4	-0.2	8	6	0.2	-0.1
4	0	-1.4	0.0	6	0	-0.3	0.0	7	4	0.3	-0.1	8	7	-0.5	0.4
4	1	2.0	0.4	6	1	-0.3	-0.1	7	5	0.1	-0.8	8	8	0.2	0.4

Notice the Kronecker delta, $\delta_{l,m}$, and the double product in the formulation. Don't confuse this process with normalizing the gravitational coefficients as shown in Eq. (8-22). Also, the difference in normalizing schemes is arbitrary and results from different traditions for geophysics and magnetism. Remember to normalize the Legendre functions if you use normalized coefficients, so the operation ultimately cancels out. You need $(l -$

m)! in both the numerator and denominator to scale the numbers and avoid potential numerical difficulties. Finally, ℓ is greater than 1, and m is greater than 0. As with the gravitational coefficients, a recursion scheme is possible to compute the geomagnetic potential. Wertz (1978:781) shows the process:

$$\Pi_{0,0}^{Schmidt} = 1$$

$$\Pi_{\ell,0}^{Schmidt} = \Pi_{\ell-1,0}^{Schmidt} \frac{2\ell-1}{\ell} \quad \ell = 1, 2, \dots$$

$$\Pi_{\ell,m}^{Schmidt} = \Pi_{\ell,m-1}^{Schmidt} \sqrt{\frac{(\ell-m+1)(\delta_{1,m}+1)}{\ell+m}} \quad m = 1, 2, \dots$$

Notice these values are independent of satellite parameters, so we can calculate them once and place them in an external location.

South Atlantic Anomaly

The radiation arriving at the Earth is deflected by the Van Allen radiation belts which were discovered in 1958 (Van Allen et al. 1958) by James Van Allen (1914-2006). The belts are formed by the interaction of the particles with the magnetic field of the Earth. There are two primary belts—an inner one extending from about 1.1 to 3.3 ER, and an outer belt from about 3.3 to 4.0 ER. Figure 8-21 shows an approximate shape. A third belt was discovered in about 1990 from the Combined Release and Radiation Effects satellite (CRRES). Some sources suggest belts to 10.0 ER.

The South Atlantic Anomaly (SAA) is a dip in the Van Allen belts centered roughly over South America and the South Atlantic. Figure 8-22 shows the approximate location. Its altitude is about 250 km and varies over time.

Thrust

We can consider thrust as an acceleration that induces a perturbation to the orbit. We saw this in the continuous-thrust discussion in Chap. 6. Thrust may quickly produce significant changes in any element of the satellite's orbit. Literature on this subject is extensive, and I've included some discussion in Chap. 6. The simplest formulation uses the acceleration caused by the motor's action.

Simply speaking, thrust is a function of the motor's mass-flow rate and the specific impulse of the fuel. Both thrust and mass-flow rate change over time and must be modeled, including the time dependencies. The acceleration is simply the time-varying thrust, divided by the vehicle's time-varying mass:

$$F_{thrust} = I_{sp} \frac{dm}{dt} \quad \ddot{a}_{thrust} = \frac{\vec{F}_{thrust}(t)}{m_c}$$

$$\Delta v = g I_{sp} \text{LN} \frac{m_{init}}{m_{init} - m_{prop}} \quad (8-50)$$

Now, find the vehicle's current mass, m_c , using

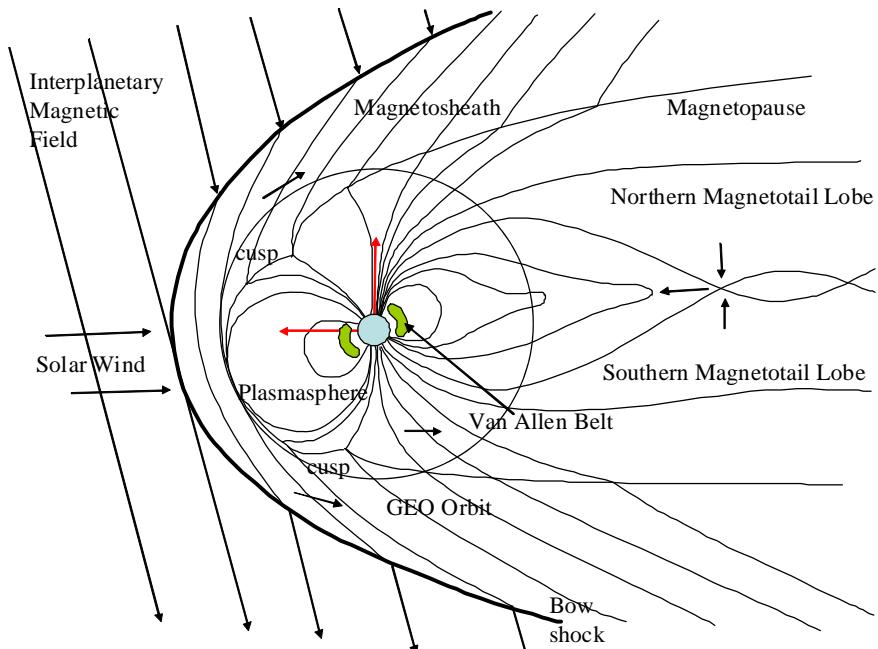


Figure 8-21. Van Allen Radiation Belts. The Earth's magnetic field distorts the incoming solar wind (radiation) from the Sun. The Van Allen radiation belts trap high energy particles and are located in a doughnut shape around the earth from about 1.1 to 2.0 ER, and 3.3 to 4.0 ER. The effect on satellites is primarily radiation. The figure is approximately to scale, and the GEO orbit is included for reference only. The basic figure is adapted from http://www.astro.phys.ethz.ch/edu/theses/berthoud/main.htm#title_the_solar_wind, Dec 20, 2006.

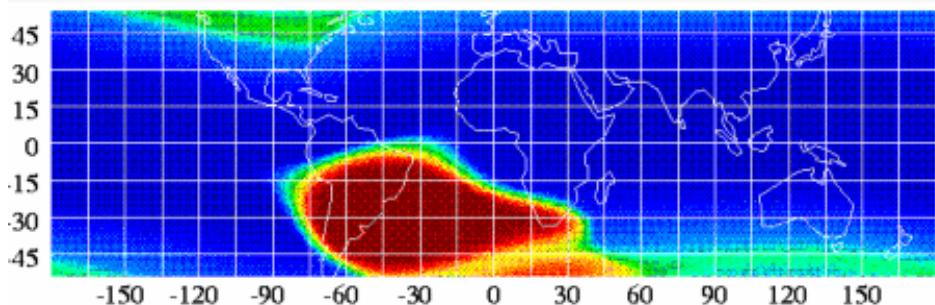


Figure 8-22. South Atlantic Anomaly. The approximate location of the SAA is over South America and the South Atlantic. The shape is variable over time, and is located at an average altitude of approximately 250 km. Source http://image.gsfc.nasa.gov/docs/ask_astro/answers/961004.html, Dec 20, 2006.

$$m_c = m_o - \int_{t_o}^{t_c} \frac{dm}{dt}(t) dt \quad (8-51)$$

Next, consider the real-world effects of thrust. Actual motor firings are finite in length and don't occur instantaneously. In addition, the thrust profile is never ideal in that ignition and tail-off characteristics diverge from a commanded behavior, and variations in the mass flow rate or specific impulse introduce time-varying thrust anomalies. Typically, throttling controlled by real-time feedback can compensate for these nonideal effects. McClain (1993) describes ways to account for ramped thrust.

Impulsive thrust is sometimes addressed in numerical operations. For short runs—say about five minutes—this approach works and introduces little error in the final solution. The technique uses the numerical-propagation routine to advance the satellite to the time of the maneuver, and the change in velocity is added to the velocity vector. The numerical propagation continues using the new velocity vector.

8.7 Forming Numerical Solutions

Special perturbations (a term used interchangeably with numerical methods and Cowell's formulation) are the simplest and most straightforward of all the perturbation methods. An overwhelming advantage is the fact that the solution contains all secular and periodic variations introduced by the original model of the perturbing forces. Although the modern computer has eliminated many of the computational constraints, obtaining the result usually takes longer than analytical methods, but produces more accurate results. The solution rests with solving the perturbed equation of motion [Algorithm 64] for several perturbing forces. To get the total result, add each of the accelerations and numerically integrate. Remember, the acceleration labeled nonspherical refers only to the perturbing acceleration caused by the nonspherical portion of the mass distribution of the central body, so don't forget to include the two-body component. The vectors may be in any coordinate system, but typically, integration is done in ECI and force model evaluations are done in ECEF. The algorithm contains some shorthand—e.g., τ represents all the time parameters required for reduction operations, and the position and velocity vectors used for all acceleration calculations are in ECEF.

ALGORITHM 64: NUMERICAL INTEGRATION ($\hat{\vec{r}}_o, \hat{\vec{v}}_o, \tau_{start}, \tau_{stop}, \Delta\tau \Rightarrow \hat{\vec{r}}, \hat{\vec{v}}$)

$$t = \tau_{start}$$

LOOP

$$\begin{aligned} \text{FK5}(\hat{\vec{r}}_{ECI}, \hat{\vec{v}}_{ECI}, \hat{\vec{a}}_{ECI}, \text{yr, mo, day, UTC, } \Delta\text{UT1, } \Delta\text{AT}, x_p, y_p \Rightarrow \\ \hat{\vec{r}}_{ECEF}, \hat{\vec{v}}_{ECEF}, \hat{\vec{a}}_{ECEF}) \end{aligned}$$

$$\begin{aligned}
U &= \frac{\mu}{r} \left[1 + \sum_{l=2m=0}^{\infty} \sum_{l=2m=0}^{\ell} \left(\frac{R_{\oplus}}{r} \right)^l P_{l,m} [\sin(\phi_{gc_{sat}})] \times \right. \\
&\quad \left. \{ C_{l,m} \cos(m\lambda_{sat}) + S_{l,m} \sin(m\lambda_{sat}) \} \right] \\
\hat{\vec{a}}_{nonspherical} &= \nabla(U - U_{2-body}) = \nabla \left(U - \frac{\mu}{r} \right) = \nabla R \\
\hat{\vec{a}}_{drag} &= -\frac{1}{2} \rho \frac{c_D A}{m} v_{rel}^2 \frac{\hat{\vec{v}}_{rel}}{|\hat{\vec{v}}_{rel}|} \\
\hat{\vec{a}}_{3-body} &= \mu_3 \left(\frac{\hat{\vec{r}}_{sat3}}{r_{sat3}^3} - \frac{\hat{\vec{r}}_{\oplus3}}{r_{\oplus3}^3} \right) \\
\hat{\vec{a}}_{srp} &= -\frac{P_{srp} c_R A_{\odot}}{m} \frac{\hat{\vec{r}}_{sat\odot}}{|\hat{\vec{r}}_{sat\odot}|} \\
\hat{\vec{a}} &= -\frac{\mu}{r^3} \hat{\vec{r}}_{ECEF} + \hat{\vec{a}}_{nonspherical} + \hat{\vec{a}}_{drag} + \hat{\vec{a}}_{3-body} + \hat{\vec{a}}_{srp} + \hat{\vec{a}}_{tides} + \hat{\vec{a}}_{other}
\end{aligned}$$

$$\begin{aligned}
\text{FK5}(\hat{\vec{r}}_{ECEF}, \hat{\vec{v}}_{ECEF}, \hat{\vec{a}}_{ECEF}, yr, mo, day, UTC, \Delta UT1, \Delta AT, x_p, y_p \Rightarrow \\
\hat{\vec{r}}_{ECP}, \hat{\vec{v}}_{ECP}, \hat{\vec{a}}_{ECI})
\end{aligned}$$

$$t = t + \Delta t$$

Numerically integrate (Runge-Kutta, Gauss Jackson, etc.)

UNTIL $t > \tau_{stop}$

Although this algorithm looks innocuous, its solution continues to challenge the greatest intellects. To account for all the major perturbations (central body, atmospheric drag, third body, solar radiation, thrust, other), we use expressions for the accelerations developed earlier in this chapter.

We can also numerically integrate the variation of parameter equations I'll introduce in Chap. 9 [Eq. (9-12), Eq. (9-14), and Eq. (9-24)]. This process is really a special perturbation solution, but I've kept it with the discussion in Chap. 9 because it focuses on orbital elements, rather than position and velocity vectors. Depending on your application, the immediate availability of orbital elements throughout a program simulation may be very beneficial.

I'll examine two approaches for numerical solutions of Algorithm 64. The simplified approach is useful for academic applications and it illustrates the general process. The technique is quite easy to adapt to the Runge-Kutta methods discussed in Sec. 8.5. The more complex method is a brief sketch of the steps required to fully implement Algorithm 64. Actually forming a complex numerical integration routine can take on several forms depending on the number of desired force models.

8.7.1 Application: Simplified Acceleration Model

Quite often a simplified version of each acceleration is useful for applications needing moderate accuracy. In these cases we model each perturbing force but without rigorous accuracy. We'll look at each force and the possible approximations.

It's possible to simplify the coordinate system transformations for certain applications, although ignoring the coordinate system altogether is not recommended. A common simplification is changing between just the PEF and TOD frames. This introduces additional error into the solution because the accelerations are applied at slightly different locations of the orbit, but for low-precision applications, this may be acceptable.

The central-body acceleration is simplified by excluding all tesseral terms (and therefore sectoral terms) in the aspherical potential and examining only the first few zonal harmonics. Although **PKEPLER** (Algorithm 65) does this analytically (for first-order J_2 secular effects), numerical integration provides the means to obtain the secular *and* periodic contributions of these zonal harmonics.

From the aspherical potential [Eq. (8-18)], we use the gradient to determine the accelerations resulting from the central body. The gradient operation produces acceleration components along each axis, so we'll show each component separately. This is actually a special case (for J_2) of the general partial derivative section beginning on page 549. Begin with the disturbing function (subscripts indicate the zonal harmonic under consideration):

$$R_2 = -\frac{3J_2\mu}{2r}\left(\frac{R_{\oplus}}{r}\right)^2\left(\sin^2(\phi_{gc}) - \frac{1}{3}\right)$$

Taking the position vector in spherical coordinates, vector r in the ITRF frame, and ϕ_{gc} , determine the accelerations for each axis. To illustrate the process, consider just the I component of the acceleration vector. Distribute the common term and recognize ($\sin(\phi_{gc}) = r_K/r$) to obtain

$$R_2 = -\frac{3J_2\mu R_{\oplus}^2}{2r^3}\left(\frac{r_K^2}{r^2}\right) + \frac{J_2\mu R_{\oplus}^2}{2r^3}$$

Differentiate to get

$$\frac{\partial R_2}{\partial r_I} = -\frac{3J_2\mu R_{\oplus}^2 r_K^2}{2}\left(-\frac{5(2r_I)}{2r^7}\right) + \frac{J_2\mu R_{\oplus}^2}{2}\left(-\frac{3(2r_I)}{2r^5}\right)$$

Simplify to produce

$$\frac{\partial R_2}{\partial r_I} = -\frac{3J_2\mu R_{\oplus}^2 r_I}{2r^5}\left(-\frac{5r_K^2}{r^2} + 1\right)$$

You can handle the remaining axes similarly; however, the K component is slightly different because r_K appears explicitly in the disturbing function. The results are

$$\begin{aligned}
 a_I &= \frac{\partial R_2}{\partial r_I} = -\frac{3J_2\mu R_{\oplus}^2 r_I}{2r^5} \left(1 - \frac{5r_K^2}{r^2}\right) \\
 a_J &= \frac{\partial R_2}{\partial r_J} = -\frac{3J_2\mu R_{\oplus}^2 r_J}{2r^5} \left(1 - \frac{5r_K^2}{r^2}\right) \\
 a_K &= \frac{\partial R_2}{\partial r_K} = -\frac{3J_2\mu R_{\oplus}^2 r_K}{2r^5} \left(3 - \frac{5r_K^2}{r^2}\right)
 \end{aligned} \tag{8-52}$$

The perturbations resulting from the third harmonic are developed from the disturbing potential resulting when $\ell = 3$, so

$$R_3 = -\frac{J_3\mu}{2r} \left(\frac{R_{\oplus}}{r}\right)^3 \{5\sin^3(\phi_{gc}) - 3\sin(\phi_{gc})\}$$

Similar to J_2 , the accelerations due to J_3 are

$$\begin{aligned}
 a_I &= \frac{\partial R_3}{\partial r_I} = -\frac{5J_3\mu R_{\oplus}^3 r_I}{2r^7} \left(3r_K - \frac{7r_K^3}{r^2}\right) \\
 a_J &= \frac{\partial R_3}{\partial r_J} = -\frac{5J_3\mu R_{\oplus}^3 r_J}{2r^7} \left(3r_K - \frac{7r_K^3}{r^2}\right) \\
 a_K &= \frac{\partial R_3}{\partial r_K} = -\frac{5J_3\mu R_{\oplus}^3}{2r^7} \left(6r_K^2 - \frac{7r_K^4}{r^2} - \frac{3}{5}r^2\right)
 \end{aligned} \tag{8-53}$$

Express the perturbations resulting from the fourth harmonic, $\ell = 4$, as

$$R_4 = -\frac{\mu J_4}{8r} \left(\frac{R_{\oplus}}{r}\right)^4 \{35\sin^4(\phi_{gc}) - 30\sin^2(\phi_{gc}) + 3\}$$

and the accelerations as

$$\begin{aligned}
 a_I &= \frac{\partial R_4}{\partial r_I} = \frac{15J_4\mu R_{\oplus}^4 r_I}{8r^7} \left(1 - \frac{14r_K^2}{r^2} + \frac{21r_K^4}{r^4}\right) \\
 a_J &= \frac{\partial R_4}{\partial r_J} = \frac{15J_4\mu R_{\oplus}^4 r_J}{8r^7} \left(1 - \frac{14r_K^2}{r^2} + \frac{21r_K^4}{r^4}\right) \\
 a_K &= \frac{\partial R_4}{\partial r_K} = \frac{15J_4\mu R_{\oplus}^4 r_K}{8r^7} \left(5 - \frac{70r_K^2}{3r^2} + \frac{21r_K^4}{r^4}\right)
 \end{aligned} \tag{8-54}$$

Escobal ([1965] 1985:50–51) shows the partials for the J_5 and J_6 zonal harmonics:

$$\begin{aligned} a_I &= \frac{\partial R_5}{\partial r_I} = \frac{3J_5\mu R_{\oplus}^5 r_I r_k}{8r^9} \left(35 - 210 \frac{r_k^2}{r^2} + 231 \frac{r_k^4}{r^4} \right) \\ a_J &= \frac{\partial R_5}{\partial r_J} = \frac{3J_5\mu R_{\oplus}^5 r_J r_k}{8r^9} \left(35 - 210 \frac{r_k^2}{r^2} + 231 \frac{r_k^4}{r^4} \right) \\ a_K &= \frac{\partial R_5}{\partial r_K} = \frac{3J_5\mu R_{\oplus}^5 r_k^2}{8r^9} \left(105 - 315 \frac{r_k^2}{r^2} + 231 \frac{r_k^4}{r^4} \right) - \frac{15J_5\mu R_{\oplus}^5}{8r^7} \end{aligned} \quad (8-55)$$

$$\begin{aligned} a_I &= \frac{\partial R_6}{\partial r_I} = -\frac{J_6\mu R_{\oplus}^6 r_I}{16r^9} \left(35 - 945 \frac{r_k^2}{r^2} + 3465 \frac{r_k^4}{r^4} - 3003 \frac{r_k^6}{r^6} \right) \\ a_J &= \frac{\partial R_6}{\partial r_J} = -\frac{J_6\mu R_{\oplus}^6 r_J}{16r^9} \left(35 - 945 \frac{r_k^2}{r^2} + 3465 \frac{r_k^4}{r^4} - 3003 \frac{r_k^6}{r^6} \right) \\ a_K &= \frac{\partial R_6}{\partial r_K} = -\frac{J_6\mu R_{\oplus}^6 r_K}{16r^9} \left(245 - 2205 \frac{r_k^2}{r^2} + 4851 \frac{r_k^4}{r^4} - 3003 \frac{r_k^6}{r^6} \right) \end{aligned} \quad (8-56)$$

We can simplify drag with three major assumptions: using an exponential atmosphere; ignoring the complex interactions with the winds in the relative-velocity calculations (and hence, the rotating atmosphere); and assuming \vec{v}_{rel} is the same as the velocity of the satellite. The result is an acceleration which computes quickly but can sacrifice considerable accuracy.

$$\ddot{\vec{a}}_{drag} = -\frac{1}{2}\rho \frac{c_D A}{m} v_{rel}^2 \frac{\vec{v}_{rel}}{|\vec{v}_{rel}|}$$

Third-body perturbations depend foremost on the locations of the other bodies (typically the Sun or Moon). Planetary-ephemeris files provide accurate (numerically generated) positions of each body's location, which programs can interpolate during operation. Unfortunately, some of these files can be very large. Several alternatives exist. First, we can calculate the position vectors of each body from analytical techniques presented in Sec. 5.1.1 and Sec. 5.2.1. A second simplification, for runs of only a few hours of simulation time, could be to store the position vector at the middle time. That's possible because the Sun and Moon will change only a small amount in "short" simulation times.

Locking the position of the Sun or Moon—or any other time varying quantity—can introduce effects in the resulting calculations. For example, consider a one-day simulation for both the Sun and the Moon. The geocentric position vector for the Sun changes about 1° (averaged throughout the year), whereas the Moon moves about 12.85° per day. Given these types of results, we may need to redetermine the acceleration at each point along a one-day simulation. On the other hand, a one-month simulation involving only the Sun may require us to rectify far less often. This gives rise to a phenomenon known as ***weak time dependence***, in that the calculations will reflect slight variations as the program updates positions of the Sun and Moon. Fonte (1995) shows that, for sys-

tems assuming an orbital period on the order of one day (GEO), the weak time dependence of the influence of the lunar motion on the satellite can contribute up to 200 m of error.

Solar radiation presents additional challenges to a simplified result for two main reasons. First, the available data for reliably forecasting future solar activity is extremely limited. Second, the method for determining the regular eclipse interval is complex to implement and may ultimately require us to know the Sun's precise location (see the previous discussions in Sec. 5.1.1 and Sec. 8.6.4). This knowledge presents a chance to share data between perturbation sections (third-body and solar-radiation) but doesn't reduce the computational load of the overall solution. A common approach assumes a spherical shape with no diffuse reflection. Simplified approaches also don't address shadowing often—a fact that slightly degrades accuracy.

8.7.2 Application: Complex Acceleration Model

Forming a complete, highly accurate, special perturbation method is a complex activity. With care, patience, and discipline, you can develop an extremely flexible and accurate tool. The equations are in Algorithm 64 just as for the simplified approach. The difference is that we rigorously determine each acceleration.

The first task is to determine the type of integration to perform. The variation of parameter solution is useful when analysis of the orbital elements and computational speed are important. Conventional integration processes the state vectors directly and is common in many flight dynamics programs. However, due to the additional processing of each complete force model, the computational requirements can be longer than the VOP approach.

The coordinate frame is very important in highly accurate programs and it requires correct EOP data (see Sec. 8.8.2). You usually integrate the equations of motion in an inertial frame, like the GCRF, which would allow you to avoid the complexity of having to account for rotating coordinates as discussed in Chap. 3 but other possibilities exist as in older programs that use the IAU-76/FK5, or TOD frames. These may not be desirable though for longer simulations because the true-of-date system continually moves. When applying the accelerations, remember that the satellite sees the “true” orientation of the Earth over time. The natural choice for coordinate systems is the Earth fixed (ITRF) system when you evaluate the accelerations. Therefore, you must use a coordinate frame that's consistent with where the satellite is at each instant of time. The pseudo-Earth fixed frame (PEF) is sometimes used for computational speed.

To start, use the aspherical potential in calculating high-precision ephemerides. Recall the form of the aspherical potential from Eq. (8-19). Implementing a model based on the infinite series of the nonspherical gravitational model requires certain decisions. First, coefficients for a given model are available to a given degree and order (41×41 , 70×70). Computer runtime may keep you from incorporating an entire field. If you use a subset of the field, say a 50×50 subset of a 70×70 field, you will introduce some small differences into the solution. Vallado (2005) showed that if truncation of the gravitational potential is necessary, additional zonal harmonic terms should be retained to improve the accuracy over simple ‘square’ truncations that are sometimes discussed.

Now consider the calculations for the associated Legendre functions. The equation [Eq. (8-16)] to determine these functions is sufficient for a few hand calculations but is terribly inefficient for computers. Recursion is an efficient way to compute orthogonal polynomials and other functions. Recursions replace intensive calculations with simpler ones. For instance, Eq. (8-57) uses simple addition and multiplications instead of more costly trigonometric terms. You should see from Eq. (8-16) that subsequent values depend on previously computed values. Long et al. (1989:4–13) give specific relations:

$$\begin{aligned} P_{t,0}[\gamma] &= \frac{(2t-1)\gamma P_{t-1,0}[\gamma] - (t-1)P_{t-2,0}[\gamma]}{t} & t \geq 2 \\ P_{t,m}[\gamma] &= P_{t-2,m}[\gamma] + (2t-1)\cos(\phi_{gc})P_{t-1,m-1}[\gamma] & m \neq 0, \quad m < t \\ P_{t,t}[\gamma] &= (2t-1)\cos(\phi_{gc})P_{t-1,t-1}[\gamma] & t \neq 0 \end{aligned} \quad (8-57)$$

with the starting values of $[\cos(\phi_{gc})]$ can be replaced by $P_{1,1}[\gamma]$ in Eq. (8-57)]

$$\begin{aligned} P_{0,0}[\gamma] &= 1 \\ P_{1,0}[\gamma] &= \gamma = \sin(\phi_{gc}) \\ P_{1,1}[\gamma] &= \cos(\phi_{gc}) \end{aligned}$$

For computer implementation, remember that m represents the order of the derivative. If m is larger than t , the derivative value is zero because the t^{th} derivative is always one. This affects only $P_{t-2,m}[\gamma]$. We sum t from 2 to the order desired, and then m goes from 0 to t . Remember to initialize each calculation with $P_{0,t-1} = 0$. We'll see later that resonance (when the Earth's rotation and the satellite's motion are commensurate) can cause large deviations in a satellite's orbit. Thus, we often want to determine this effect. Unfortunately, only a few gravitational terms usually contribute to this effect. Finding all the Legendre polynomials is very inefficient. In these cases, we can use the recursion relation [Eq. (8-57)] to find a smaller subset of the gravity field (say 12×12) and then use Eq. (8-16) to find the individual terms responsible for the resonance.

Notice that the matrix of associated Legendre functions can be calculated *a priori* to use with Eq. (8-12), but you can't perform this operation directly if you use normalized coefficients. In this case, you must store two matrices because $P_{t,m}$ and $P_{t,m+1}$ both multiply by the same gravitational coefficients.

We always address the topic of *stability* whenever we use a recursion algorithm. **Stability** is determined by the particular structure of the recurrence relation and the precision of the computation. For example, suppose the recurrence is given by the difference of two nearly equal quantities that are the same to three significant digits. If our precision is ten digits, the first application of the recursion loses five digits of accuracy. Repeated application of this process quickly produces useless results. Now suppose the recurrence is given by a summation rather than a difference. We retain the precision with only a possibility of losing the final digit due to round-off error. The process would corrupt the final results, but only after thousands of applications of the recurrence relation. These examples don't indicate the relations for the associated Legendre functions

but illustrate the concept of stability. A clear indication of an unstable recurrence relation is the appearance of small divisors. Bessel functions appear as coefficients in the Fourier series representation of Kepler's equation (see Sec. 2.2.5). The recurrence relation has eccentricity in the denominator of one term. When e is small, this single term dominates all others, effectively eliminating their information content.

It's useful to obtain recursion expressions for the latitude and longitude terms:

$$\begin{aligned}\sin(m\lambda) &= 2\cos(\lambda)\sin\{(m-1)\lambda\} - \sin\{(m-2)\lambda\} \\ \cos(m\lambda) &= 2\cos(\lambda)\cos\{(m-1)\lambda\} - \cos\{(m-2)\lambda\} \\ m\tan(\phi_{gc}) &= (m-1)\tan(\phi_{gc}) + \tan(\phi_{gc})\end{aligned}\quad (8.58)$$

Treating drag in highly accurate programs remains difficult, despite the deceptively simple form of the equation of motion. This is because accurate density calculations require a complex atmospheric model. Typically, the more complete atmospheric models require extensive knowledge about the incoming solar radiation and the atmosphere's molecular composition. Although we know these inputs with good precision from observational measurements, forecasting the data through solar cycles and other variations is much less accurate. Wright and Tanygin (2004) show the importance of interpolating the solar weather parameters and Vallado and Kelso (2005) show how to interpolate the geomagnetic and solar flux parameters. In fact, interpolation differences can sometimes exceed model differences. Nevertheless, real studies for mission analysis and planning typically warrant the more sophisticated model, despite the associated computational burden. For example, an 81-day running average of $F_{10.7}$ is reasonable for fitting past data; however, it becomes a limitation and potential source of error when considering mission *planning*. Indeed, the 81-day running average becomes less realistic the greater the extrapolation into the future. In some of these cases, it might be better to use a model such as Harris-Priester (page 568), which attempts to model certain tendencies during discrete intervals of the solar cycle. Schatten's (1988, 1993, etc.) prediction estimates for atmospheric parameters may also be of use in reasonably estimating the future. We can still miss the solar activity when selecting a table value but can greatly reduce variations in the real world.

The relative-velocity calculation is:

$$\vec{v}_{rel} = \vec{v}_{GCRF} - \vec{\omega} \times \vec{r}_{GCRF} + \vec{v}_{wind}$$

For ease of implementation, all the vectors should be in an inertial coordinate system (I show GCRF here and recommend it). Accurate drag analysis (and solar radiation analysis) requires us to know the mass and shape of the spacecraft and its time-varying orientation, or *attitude*. Values for the coefficient of drag for various shapes are available in books describing moments of inertia. We often use a simple sphere to represent the spacecraft's shape. The sphere works reasonably well in many cases, except when the satellite is very large, small, or complex. But highly accurate studies show that even slight changes in c_D (1.98 vs. 2.00, for instance) can strongly change the satellite's orbit (up to hundreds of kilometers); thus, we almost always solve for the drag parameter (c_D) or mean-motion rates (\dot{n}, \ddot{n}) during orbit determination. The estimated c_D helps

account for the difference in shape from the assumed model. This is only an issue for extremely accurate studies involving large, complex spacecraft with time-varying attitudes. ***Macro models*** help in that they try to piece together well-known shapes to approximate the satellite's true shape. They're especially useful for atmospheric drag and solar-radiation calculations. Luthcke et al. (1997) used a macro model for TDRS to improve the orbit determination accuracy to a few meters. Knowing the satellite's precise orientation is equally difficult and, in many cases, may actually involve commanding and controlling the satellite to receive health and status information from it.

Treatment of third-body accelerations is virtually the same as for the simplified approach, but we usually use the JPL recovery technique to recreate the planetary ephemerides as closely as possible. Of course, we must use planetary ephemerides (see Sec. 5.3.1) consistent with the inertial frame we use for integrating the equations of motion.

Lastly, the equations in Sec. 8.6.4 allow us to model solar-radiation pressure. Difficulties arise that are similar to those for the drag accelerations. Remember, now the effective cross section is the projection of the satellite's area (A_{\odot}) that is exposed to the Sun and *not* the component normal to the velocity vector. The solar-ephemeris data required to calculate solar-radiation pressure is the same as that for some atmospheric density models and for the third body. The acceleration due to solar radiation is also very sensitive to the time in which the Earth eclipses the satellite. A cylindrical shadow model is common (see page 299). But more precise determination requires us to analyze the times when the satellite is in the umbra and penumbra regions, uses an elliptical-Earth model, and updates the location of the Sun and Moon to determine precise entrance and exit times.

8.8 Practical Considerations

Overall, propagating orbits with high accuracy is complicated. There are many techniques, and people often select one method over another for the wrong reason. Computer source code is closely guarded and usually evolves as a conglomeration of several techniques. Thus, even if software were available, making it work still takes a lot of effort. Development from scratch is another option; yet it also has many drawbacks. Its main disadvantages are trying to prepare for all input types, coupling with a differential-correction program, and the availability of suitable test cases for validation. Because we discuss differential-correction in Chap. 10, we'll concentrate on propagator-related issues in this section.

8.8.1 Verifying and Validating the Propagator

Perhaps the largest misconception in evaluating methods of propagation is what represents *truth*. We ***verify*** to existing data or results (ensure it's coded properly), while we ***validate*** that the model accurately reflects the truth (consistent and reliable results). Determining the “true” trajectory is ultimately tied to differential correction and involves the discussions of Sec. 10.4.2 and Sec. 10.9. It usually uses either existing or actual observational data. Although each type of data has advantages, both also have important drawbacks we must address. We must consider *both* types to completely ver-

ify and validate a propagation routine. Also important are precision and accuracy. **Precision** is the fineness (or reproducibility) with which a measurement is made while **accuracy** is how close the measurement is with respect to truth. These terms are sometimes confused.

First, let's introduce some terms that allow us to discuss how to verify and validate propagation algorithms. Because any propagator can coincidentally produce the location of the satellite at certain instants of time, we usually use an **ephemeris**, or a set of position and velocity vectors over time. This ensures we get the average behavior of the propagation routine. We refer to tabular values of position and velocity over time for several satellites as **ephemerides**. Highly accurate ephemerides that come from high-fidelity numerical techniques are called **Precision Orbit Ephemerides** (POE). POE's are available for a few satellites such as TOPEX, LAGEOS, ETALON 1 and 2, GPS, TDRSS, and a few others. You should contact the appropriate organization to determine their availability. For instance, the Center for Space Research at the University of Texas at Austin produces POE's for TOPEX. See Appendix D for additional information.

When we use data to test a propagator, there are two options. If we use the same program, for instance during a system upgrade, we can use existing test cases and benchmarks to check our results (verification). Although this allows us to check the new implementation, any errors in the existing test cases could migrate to the new system, so this should only be used as an initial check. A necessary extension seeks an independent program to more fully verify the results. This technique overcomes the "familiarity" of the original program, and allows an independent evaluation of the propagation routine. Excellent examples of independent benchmarks are the sample orbits documented in Metzinger (1993) and Chao et al. (2000). The first document gives various orbits and their expected output from several different computing platforms. The propagation algorithm used by this study is the Draper version of the **Goddard Trajectory Determination System** (GTDS) with complete force models. GTDS is comparable with other legacy programs that I list in Appendix D (Vallado, 2005) and it is capable of producing results to well under a meter of the absolute truth. Because this document has been updated over many years, it contains a wealth of information for any validation effort. Be careful to use the same data (coordinate systems, gravitational constants, atmospheric parameters, etc.) when using different programs. It's especially difficult to ensure atmospheric data have identical actual and predicted values.

Simulated data can solve some of the problems encountered using actual observations, but it still doesn't model all perturbational forces. Using simulated data allows us to develop consistency for orbit determination because we can know exactly how the underlying orbits correspond to the measurements. We can also develop continuous data streams from which we can estimate the orbit. Remember, this is usually *continuous* data for the entire period; it doesn't represent the real world, but it's enough for initial analyses.

The second primary means of testing involves tests with real-world data and satellites (verification). The advent of modern laser-tracking systems permits very accurate analysis of how well a propagation routine models the real world. On first thought, actual observations should be the most accurate means to determine the *truth*, but as

shown in the estimation section (Sec. 10.9), this process isn't perfect. In fact, a review of Chap. 10 could lead us to think the truth is unachievable. Fortunately, this isn't the case and many satellite orbits (TOPEX, for instance) are known to centimeter accuracy (*post-processing*).

8.8.2 Physical Data and Sources

Collecting physical data for use with any type of perturbation analysis ensures consistent results. Data is what we determine and measure empirically or what we determine experimentally based on known physical models. Of course, the most accurate data exists for past events because we can retrieve actual measurements (*post-processing*). However, we often need predicted values of unavailable current or future data to plan satellite missions (prediction). The accuracy of the predicted data strongly affects the final results. See Appendix D.5, "Data Sources," for additional information.

Gravitational Parameters

We get gravitational data from three main sources: measurements taken at the Earth's surface (terrestrial), altimetry (height above the Earth), and the observed motion of satellites. Historically, only terrestrial measurements were available. The *Ohio State University* model (OSU-91A) is complete through degree and order 360. Models based on satellite observations are usually smaller. Other models also include the *Joint Gravity Model* (JGM-3, 70×70), developed through collaboration between NASA/Goddard and the University of Texas at Austin (Nerem et al. 1994). The JGM models updated NASA's *Goddard Earth Model* (GEM-T3) and have improvements using *International Earth Rotation Service* (IERS) constants, analysis of long-wavelength data, and TOPEX altimetry data. The Defense Mapping Agency's *World Geodetic Survey* (WGS-84, 41×41) has several different dates of release (1960, 1966, 1972, 1984). Each revision is referenced as WGS-##, with the appropriate year inserted. There are separate models from the University of Texas (UT/CSR TEG3), the NASA *Goddard Earth Model* (GEM), and others. A larger and combined model (EGM-96, 360×360) was published in late 1996.

Typically, accuracy requirements drive the maximum degree and order. For example, 4×4 fields are often adequate for deep-space orbits, whereas some low-Earth satellites need at least a 50×50 field. Increasing the field size isn't always enough. For highly accurate missions, we usually need to consider other force models. Table D-1 and Table D-2 in Appendix D show representative coefficient values from the EGM-96 model.

The development of gravitational models over the last few decades shows a number of different models (Fig. 8-23). That's because it's hard to determine absolute gravitational coefficients, and having to resolve multiple data sources and satellite classes makes the process even more difficult (Table 8-7). Vetter (1994) concisely recounts the development of most commonly used gravitational models, whose use rests on two important facts:

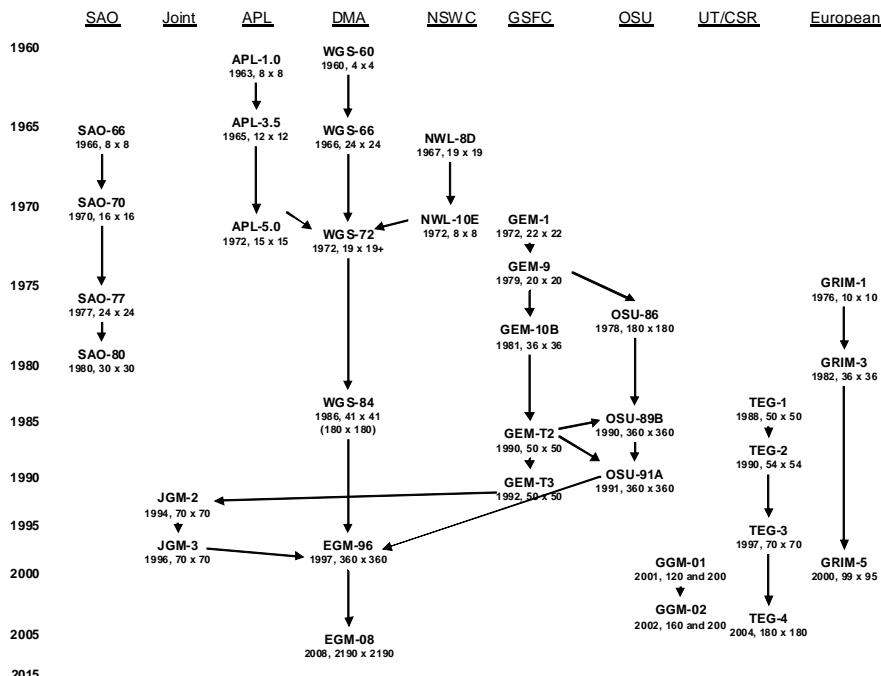


Figure 8-23. Development of Earth Gravity Models. The *Joint Gravity Models* (JGM) come from Goddard Space Flight Center (GSFC), Ohio State University (OSU), University of Texas at Austin (UT), and the European communities. The *Earth Gravity Model* (EGM) combines the JGM work with Defense Mapping Agencies efforts. The *Goddard Earth Models* (GEM) were produced annually beginning with GEM-1 in 1972. Even numbered models contain satellite and surface gravity data. Odd numbered models contain only satellite data. The Grace Gravity models (GGM-02) are derived solely from satellite data. *Standard Earth* (SAO) and *Applied Physics Laboratory* (APL) models were among the first models. The basic information is from Vetter (1994) and (2001).

1. We must use the complete theory and associated constants to obtain documented accuracy because mixing constants from different theories creates errors. This complete set includes the four fundamental defining parameters in the table on the inside back cover, for each model, the speed of light, masses of the planets, a sea-surface model, etc. We *must use a complete and consistent set of physical constants (as in that table) for any high-accuracy study.*
2. The gravitational coefficients are entirely accurate only *for the particular orbit classes used in their development*. This point is extremely important because constants derived from a single class of satellite orbits will work

TABLE 8-7. Current Gravitational Model Data Sources. This table lists the major sources of data to create each gravity field (Vetter, 1994, 2001). The data types consist of Optical, Laser, Doppler, Range Rate (RR), Satellite to satellite tracking (S2), GPS, and DORIS information. Some models include satellite only, or combined results.

Model	Data Types	Perigee Alt km	Inclination °	Surface Gravity	Altimeter	# Sats
JGM-3	O, L, D, RR, S2	600-2000	1-144	1 × 1°MGA	Seasat, Geosat	30
	SLR, DORIS	5900, 35,000			GEOS-3, SPOT-2	
WGS-84	L, D, GPS	800-1100	50-90	1 × 1°MGA	Seasat	10
		5700, 20,500			GEOS-3	
EGM-96	All types	600-2000	1-144	1 × 1°MGA	ERS-1, Geosat,	30+
		5900, 35,000		30 × 30'MGA	TOPEX	
GEM T3	O, L, D, RR, S2	600-2000	1-144	1 × 1°MGA	Seasat, Geosat	30
		5900, 35,000			GEOS-3	
OSU 91A	O, L, D, RR, S2	600-2000	1-144	1 × 1°MGA	TOPEX	30
TEG-3	L, D, RR	780-1600	15-108	1 × 1°MGA	Seasat, Geosat	30
		5900, 19200			GEOS-3	
GRIM 5	O, L, D	19,200	15-115	1 × 1°MGA	SPOT-2	32

very well for that same class, but they may diverge considerably for different classes. Thus, we need to use a model derived from many satellite orbits.

In this light, Vetter (1994) recommended the JGM-2 (Nerem et al. 1994) and the newer JGM-3 models because they're derived from many orbit classes. The EGM-96 effort is developed among the University of Texas at Austin, the Defense Mapping Agency, Ohio State University, and the Goddard Spaceflight Center. The model uses data from 30 satellites and from surface gravity measurements; it is a complete 360×360 field. Any of these gravity fields may be *tuned* for a particular orbit, although tuning is uncommon because it's complex.

Earth Orientation Parameters

Earth orientation parameters let us precisely model the orientation of the Earth at a particular instant of time. First we need the precise location of the North Pole. Recall from page 223 that the pole precesses and nutates slightly over time. Highly accurate calculations must include corrections for polar motion. The data consists of x_p and y_p displacements from the **Celestial Intermediate Pole** (CIP), defined as the axis of rotation which is normal to the true equator of the Earth. The x_p and y_p angles represent displacements from the 0° and 270° longitude meridians, respectively. These coordinates are available from several sources.

The effect of using the EOP parameters is small but noticeable. Precise operations always use the proper EOP values. The EOP data should be interpolated to recover the

sub-diurnal values in UT and polar motion. These values are removed during processing, but should be re-added during operations. The values are valid at 0 hours, except the average values which are valid at noon. It can be particularly challenging to assemble a complete and consistent set of values for use in operations because the recent data is updated at irregular intervals.

Next, we must convert time. To convert between time systems (Algorithm 16), we need to know the difference between atomic time and coordinated universal time ($\Delta AT = TAI - UTC$), as well as the difference between universal time and coordinated time ($\Delta UT1 = UT1 - UTC$). Other users may have a particular epoch, such as January 1, 1970, and would therefore require timing coefficients to enable recovery of positions at that particular time. We must be sure to have access to the correct time associated with *all* observational equipment. The broadcast of universal time by the United States National Institute for Standards and Technologies (WWV in Fort Collins, Colorado, and WWVH in Hawaii) is integral to this process. Special receivers take the WWV and WWVH signals and place them in computers. See Sec. 3.5.4 for additional information.

NGA produces coefficients that will reproduce about one week of data for the polar motion coefficients, and the timing corrections. They are reasonably accurate. NGA reintroduced inclusion of zonal tide corrections into the data beginning with week 2 of 2005. This may change the accuracy slightly.

The EOP parameters, historical data, and analysis were discussed in detail in Vallado and Kelso (2005). That information is summarized in App. D.5, “Data Sources”.

Atmospheric and Solar Data

The most accepted compilation of the measurements of planetary geomagnetic indices (K_p and a_p) is created by the Institute für Geophysik, at Göttingen University, Germany (Geomagnetic). The solar flux measurements are taken at the Dominion Radio Astrophysical Observatory in Penticton, British Columbia. The most commonly accepted measurement of $F_{10.7}$ is distributed daily by the National Oceanographic and Atmospheric Administration (NOAA) at the National Geophysical Data Center (NGDC) in Boulder, Colorado. It's worth noting that NOAA publishes values for (a_p , K_p , and $F_{10.7}$) at weekly intervals and includes 45-day predictions for each value. High-fidelity programs will find NOAA to be a unique source for accurate measurements and predictions. The solar flux data is generally adjusted to an Earth distance of 1.0 AU, however observed values are also required by some atmospheric models. Vallado and Kelso (2005) also studied the various sources of data and show how to produce a continuous file of data for use in operations. Their data is also available on the internet as shown in Appendix D.5, “Data Sources”. For studies discussing the accuracy of predictions, Eller et al. (1994), Jablonski (1991), and others.

Planetary Ephemerides

The main source for all planetary and lunar ephemerides is the Jet Propulsion Laboratory (JPL). Data is available for the planets (e.g., Development Ephemeris, DE-245) and the Moon (e.g., Lunar Ephemeris, LE-245). Specific details on forming these ephemerides are in Standish (1990:252–271). JPL creates these ephemerides by numerical inte-

grations accounting for the influence of the planets. The numerical integration uses a variable step-size that averages about 0.3 days. Once the ephemerides are created, the results are fit with Chebyshev polynomials and collected into 32-day blocks. For DE-245 and DE-405, each polynomial set covers 16 days for Venus, the Earth, and the Sun, 4 days for the Moon, 8 days for Mercury and the librations, and 32 days for the remaining objects. The general accuracy of the ephemerides is about $0.01''$. The Moon ephemerides are accurate to about 2 m ($0.001''$) and the Sun is accurate to about 200 m ($0.0003''$). Long et al. (1989:3-84 to 3-88) describes how to retrieve the Chebyshev polynomials.

TABLE 8-8. Available Planetary and Lunar Ephemerides. The table shows available ephemerides and when they apply. Lunar ephemerides (LE-###) are produced with each development ephemerides. The nutations are the $\Delta\Psi$ and $\Delta\epsilon$ values for the IAU 1980 theory of nutation, and the librations are the 3 Euler velocities obtained by interpolating the Moon's position vector.

Designation	Basis	Year	Start	Stop	Comments
DE-102	FK4	1977	Apr 16, 1410	Dec 22, 3002	Long span
DE-118	FK4	1980	Dec 9, 1599	Mar 31, 2169	Previous standard
DE-200	FK5	1982	Dec 9, 1599	Mar 31, 2169	Includes nutations
DE-245	FK5	1993	Jun 16, 1750	Aug 1, 2051	Previous standard
DE-405	ICRF	1997	Dec 9, 1599	Dec 31, 2199	Includes nutations & librations
DE-406	ICRF	1997	Feb 23, -3000	May 6, 3000	Long span
DE-418	ICRF	2007	1899	2051	for New Horizons
DE-422	ICRF	2009	Feb 23, -3000	May 6, 3000	Incorporates new satellite data

There are also ephemerides available from NASA's Navigation and Ancillary Information Facility (NAIF). The so-called SPICE files (spacecraft, planet, instrument, c-matrix, events) are a collection of data and programs to aid navigation and trajectory solutions. The ephemerides come in variable and uniform time steps, as well as spice kernel (SPK), binary (BSP), and planetary constant kernel (PCK) formats. See <http://naif.jpl.nasa.gov/naif/about.html> for additional information.

IAU Resolution #3 (July 2006) recommends that TDB may be used for T_{Eph} as the independent time argument for the planetary ephemerides (DE-405).

Problems

1. I've taken the convention of Kaula and used the positive gradient for this book, whereas other authors adopt the acceleration as a *negative* of the potential. Am I justified when examining the choice of c in the development of ξ and the discussion of gradients in Sec. 8.2? How are potential energy, potential functions, and accelerations related?
2. Program the Jacchia-Roberts model discussed in Appendix B, and compare the results to the exponential model for daily values in the months of January, 1986, and May, 1996. Explain differences and similarities. Hint: you'll need to obtain data using information in Appendix D.5.
3. Develop coefficients for an eighth-order Runge-Kutta, Adams-Basforth, and Gauss-Jackson numerical integrator. Test the resulting propagators on a Molniya orbit.
4. Derive the form of third-body acceleration in Eq. (8-36). Also show that Eq. (8-36) is a simplified form of Eq. (8-37).
5. Compare computing times for Encke's method and the simplified numerical technique discussed previously using a GPS satellite and a one-week simulation in May, 1996. Use the same force models with Encke's method as in the simplified approach. Do your answers differ for TOPEX? (Hint, it's in a roughly circular, 1300 km altitude, 66.0° inclined orbit) Additional credit: obtain the precision orbit ephemerides for TOPEX and analyze which method is more accurate. With Encke's method, was the rule of thumb for rectification points accurate?
6. Calculate entry and exit times for a satellite for one week (January 17 to January 25, 1997) given the following information.

$$\hat{r} = -26,175.1034 \hat{i} + 12,757.0706 \hat{j} + 14,626.6556 \hat{k} \text{ km}$$

$$\hat{v} = 2.376\,641 \hat{i} + 0.139\,677 \hat{j} + 2.078\,097 \hat{k} \text{ km/s}$$

What's the percentage increase for the solar-radiation pressure if we ignore the periods when the satellite is in the Earth's shadow?

7. A satellite is about to launch for an Earth Resources Technology mission. It has a radar that can map out crop resources, accurate to about 15 m. Someone forgot to examine the astrodynamical needs for orbit determination, so you've been asked to help. If the planned orbital elements are $a = 6800$ km, $e = 0.02$, $i = 28.5^\circ$, what forces should you include in your analysis?
8. Calculate a one week ephemeris for the TOPEX satellite using the simplified and complex numerical propagation techniques discussed earlier. The data references the ECI (IAU-76/FK5) frame. The orbital data is as follows:

$$\hat{\vec{r}} = -2703.635\,550 \hat{i} + 1965.968\,884 \hat{j} - 6954.830\,225 \hat{k} \text{ km}$$

$$\hat{\vec{v}} = -5.779\,958\,666 \hat{i} - 4.125\,847\,273 \hat{j} + 1.079\,893\,238 \hat{k} \text{ km/s}$$

Epoch = December 8, 1996, 00:00:00.00 UTC

Research and locate the POE information for TOPEX after the week of interest and compare your answers.

9. a. In Sec. 3.3, we introduced the Solar System, and Earth barycenters. Draw a figure to show the magnitude and orientation of the barycenter locations on August 21, 2004 at 00:00 TDB.
 b. Use the following data from DE200 for August 21, 2004 at 00:00 TDB ($JD_{TDB} = 2453238.5$) to find the vector locations for the Sun and Earth barycenters from the center of each object. All units are in km.

Earth Center — Jupiter		
$\hat{\vec{r}}_I$	$\hat{\vec{r}}_J$	$\hat{\vec{r}}_K$
-941,461,060.3967	107,142,732.9047	66,072,182.1543
Sun Center — Jupiter		
-812,813,500.2447	34,019,092.0896	34,370,010.4158
Solar System Barycenter — Jupiter		
-812,196,708.2857	33,853,712.3193	34,283,473.3035
Earth-Moon Barycenter — Jupiter		
-941,456,961.8638	107,144,697.0723	66,072,980.4147
Sun Center — Earth		
128,647,560.1520	-73,123,640.8150	-31,702,171.7385
Sun Center — Earth-Moon Barycenter		
128,643,461.6191	-73,125,604.9826	-31,702,969.9989
Solar System Barycenter — Earth		
129,264,352.1110	-73,289,020.5854	-31,7887,08.8509
Solar System Barycenter — Earth-Moon Barycenter		
129,260,253.5782	-73,290,984.7530	-31,789,507.1112
Earth Center — Moon		
337,311.6606	161,652.1478	65,697.2955

- c. Does your answer in part *a* make sense in light of your answers to part *b*?