2002

Coordinates

Clynch, James R.

Monterey, California, Naval Postgraduate School

Coordinates

James R. Clynch
Naval Postgraduate School, 2002

I. Coordinate Types

There are two generic types of coordinates:

Cartesian, and
Curvilinear of Angular.

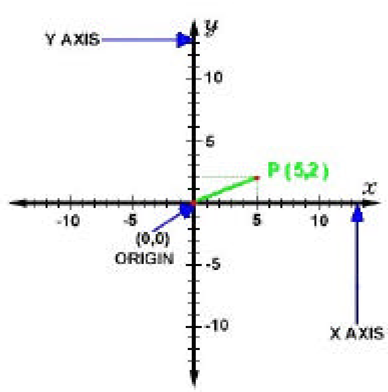
Those that provide x-y-z type values in meters, kilometers or other distance units are called **Cartesian**. Those that provide latitude, longitude, and height are called **curvilinear or angular**.

The Cartesian and angular coordinates are equivalent, but only after supplying some extra information. For the spherical earth model only the earth radius is needed. For the ellipsoidal earth, two parameters of the ellipsoid are needed. (These can be any of several sets. The most common is the semi-major axis, called "a", and the flattening, called "f".)

II. Cartesian Coordinates

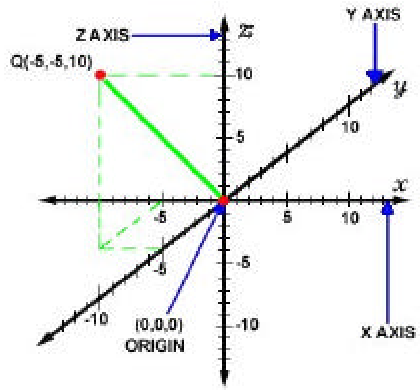
A. Generic Cartesian Coordinates

These are the coordinates that are used in algebra to plot functions. For a two dimensional system there are two axes, which are perpendicular to each other. The value of a point is represented by the values of the point projected onto the axes. In the figure below the point (5,2) and the standard orientation for the X and Y axes are shown.



Cartesian
2 Dimensional System

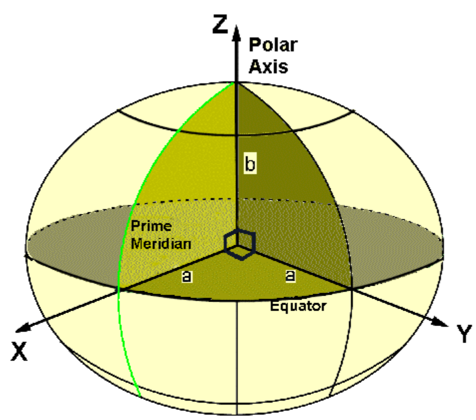
In three dimensions the same process is used. In this case there are three axis. There is some ambiguity to the orientation of the Z axis once the X and Y axes have been drawn. There are two choices, leading to right and left handed systems. The standard choice, a right hand system is shown below. Rotating a standard (right hand) screw from X into Y advances along the positive Z axis. The point Q at (-5, -5, 10) is shown.



Cartesian
3 Dimensional System

B. Earth Centered, Earth Fixed (ECEF) Coordinates

For the earth, the convention is to place the origin of the coordinates at the center of the earth. Then the positive Z axis goes out the north pole. The X-Y plane will be the equatorial plane. The X-Axis is along the prime meridian (zero point of longitude). The Y axis is then set to make the system right handed. This places it going out in the Indian Ocean.



Earth Fixed Cartesian Coordinates
X-Y Plane is Equatorial Plane
X On Prime Meridian
Z Polar Axis

These **Earth Centered, Earth Fixed ECEF** coordinates are the ones used by most satellites systems to designate an earth position. This is done because it gives precise values without having to choose a specific ellipsoid. Only the center of the earth and the orientation of the axis is needed. To convert to angular coordinates, more information is needed. Some high precision applications remain in ECEF to avoid additional error. High precision geodetic bench marks have both angular and ECEF coordinates recorded in the data bases.

C. Terrestrial Reference Frames

At the highest accuracy, geodesy is done in a ECEF coordinate system. Different organizations have defined a series of these over the last few decades. They are defined by specifying the locations of a small set of reference bench marks. These definitions are in ECEF.

The US DoD system is called **World Geodetic System 84 (WGS84)**. This World Geodetic System has several components. One of these is the reference frame. WGS84 was used as the basis of the GPS solutions. This was done by finding the WGS84 ECEF locations of the stations that supply data for the Broad Cast Ephemeris (BCE) computation. WGS84 was a major successor to the previous system WGS72. There was a significant shift between the two systems in some parts of the world.

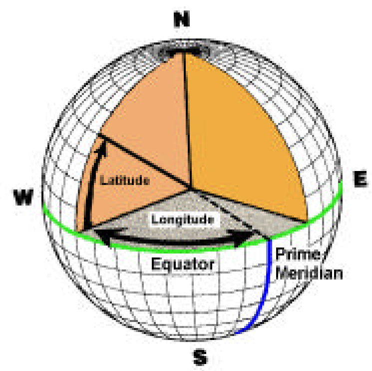
The science community has been working on a series of world reference systems that are called **International Terrestrial Reference Systems or ITRF's**. The earliest ones were ITRF92 and ITRF94, which was quite good. Modest improvements followed with ITRF97 and ITRF2000. The later two models were so accurate that models of the motion of the **crustal plates** of the earth had to be included.

III. Angular or Curvilinear Coordinates

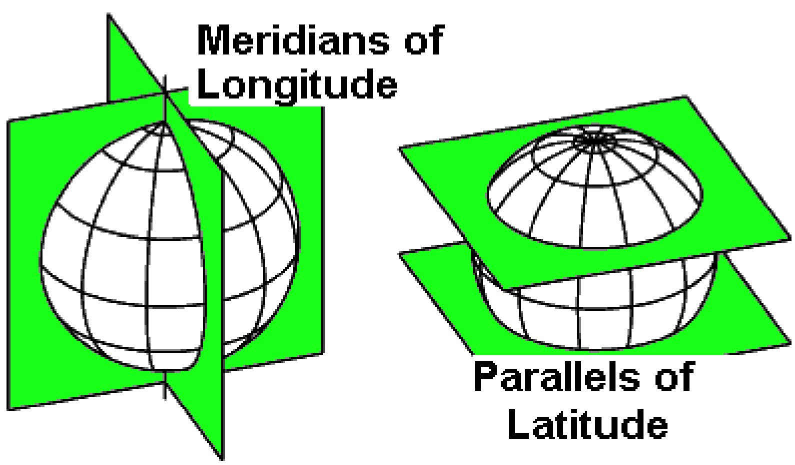
Angular coordinates or curvilinear coordinates are the **latitude, longitude and height** that are common on maps and in everyday use. The conversion is simple for the spherical earth model. For the ellipsoidal model, which is needed for real world applications, the issue of latitude is more complex. The height is even more complicated.

A. Latitude and Longitude on Spherical Earth

For a spherical earth, latitude and longitude are the grid lines you see on globes. These are angles seen from the center of the earth. The angle up from the equator is latitude. In the southern hemisphere is it negative. The angle in the equatorial plane is the longitude. The <u>reference for latitude is set by the equator - effectively set by the spin axis of the earth</u>. <u>There is no natural reference for longitude</u>. The zero line, called the **prime meridian**, is taken as the line through Greenwich England. (This was set by treaty in 1878. Before that each major nation had its own zero of longitude.)



The longitude used in technical work is positive going east from the prime meridian. The values go from 0 to 360 degrees. A value in the middle United States is therefore about 260 degrees east longitude. This is also the same as -100 degrees east. In order to make longitudes more convenient, often values in the western hemisphere are quoted in terms of angles west from the prime meridian. Thus the meridian of -100 E (E for East) is also 100 W (W for West). In a similar manor latitudes south of the equator are often given as "S" (for south) values to avoid negative numbers.



The lines of latitude and longitude form a pattern called the **graticule**. The lines of constant longitude are called **meridians**. These all meet at the poles. They are circles of the same radius which is the radius of the earth, R_e. They are examples of **great circles**. The lines of constant latitude are called **Parallels** of Latitude. They are circles with a radius dependent on the latitude. The radius is $\cos(\phi) R_e$. Parallels of latitude are called **small circles**.

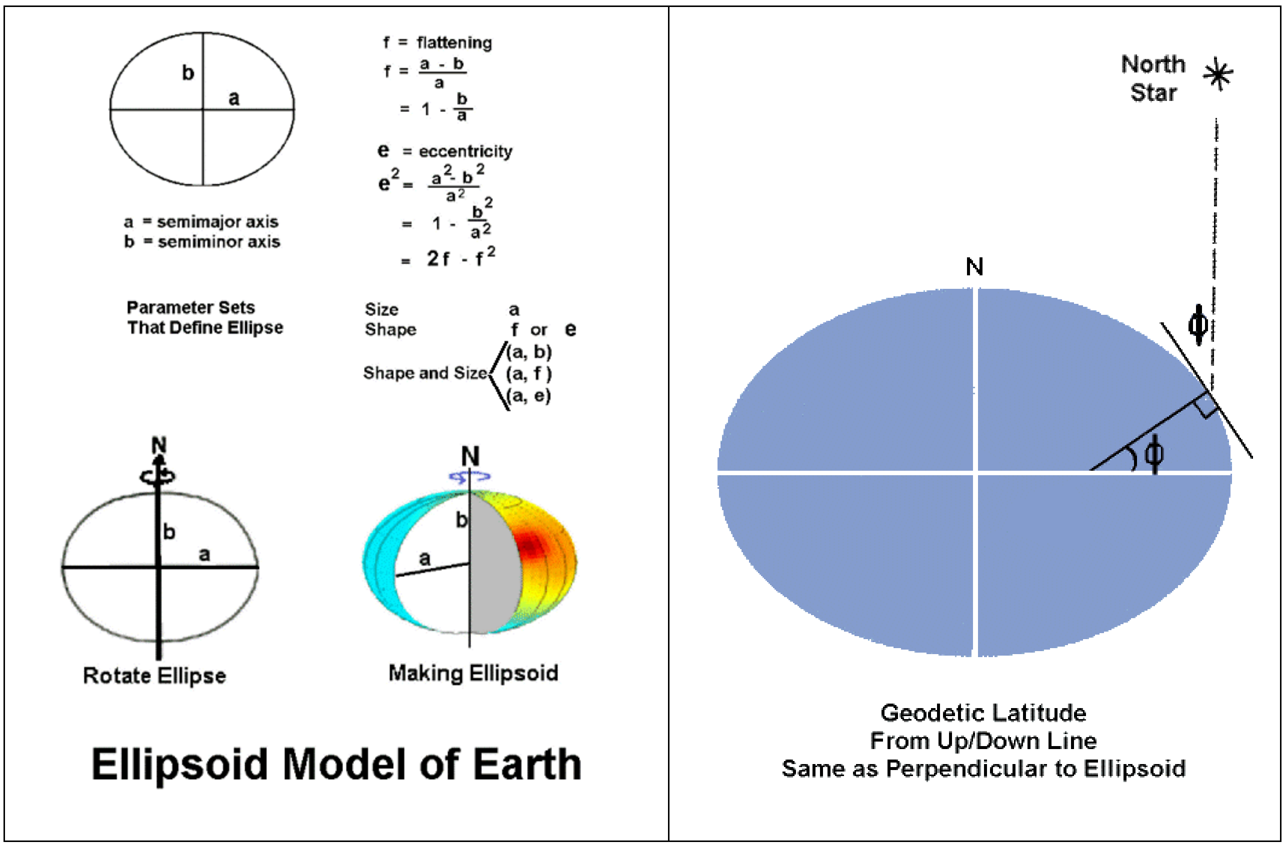
B. Latitude and Longitude on Ellipsoidal Earth

The earth is flattened by rotational effects. The cross-section of a meridian is no loner a circle, but an ellipse.[†] The ellipse that best fits the earth is only slightly different from a circle. The flattening, defined in the figure below, is about 1/298.25 for the earth.

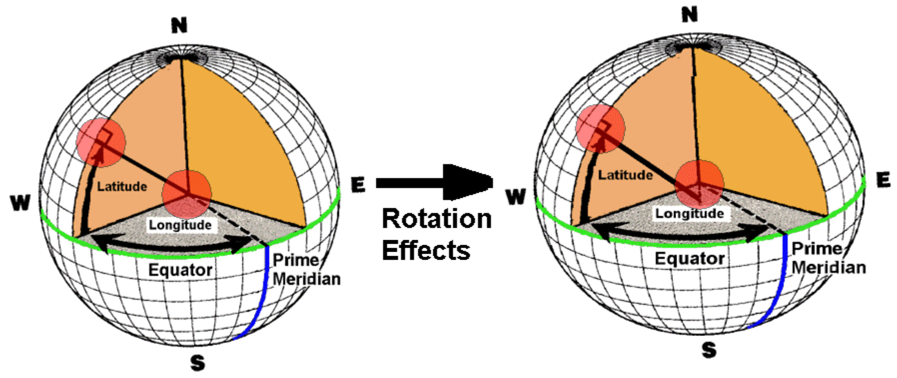
[†] The precise parameters of this ellipse, as well as the origin of it, have slightly different values in different parts of the world. This will be covered elsewhere in the Datums document.

The latitude and longitude are defined to be "intuitively the same as for a spherical earth". The longitude is the precisely the same. The way latitude is handled was defined by the French in the 17th century after Newton deduced that the world had an elliptical cross-section.

The latitude was measured before satellites by observing the stars. In particular observing the angle between the horizon and stars. In defining the horizon a plumb bob or spirit level was used. The "vertical line" of the plumb bob was thought to be perpendicular to the sphere that formed the earth. The extension to an ellipsoidal earth is to use the line perpendicular to the ellipsoid. This is essentially the same as the plumb bob. The small difference are discussed below under astrodetic coordinates.



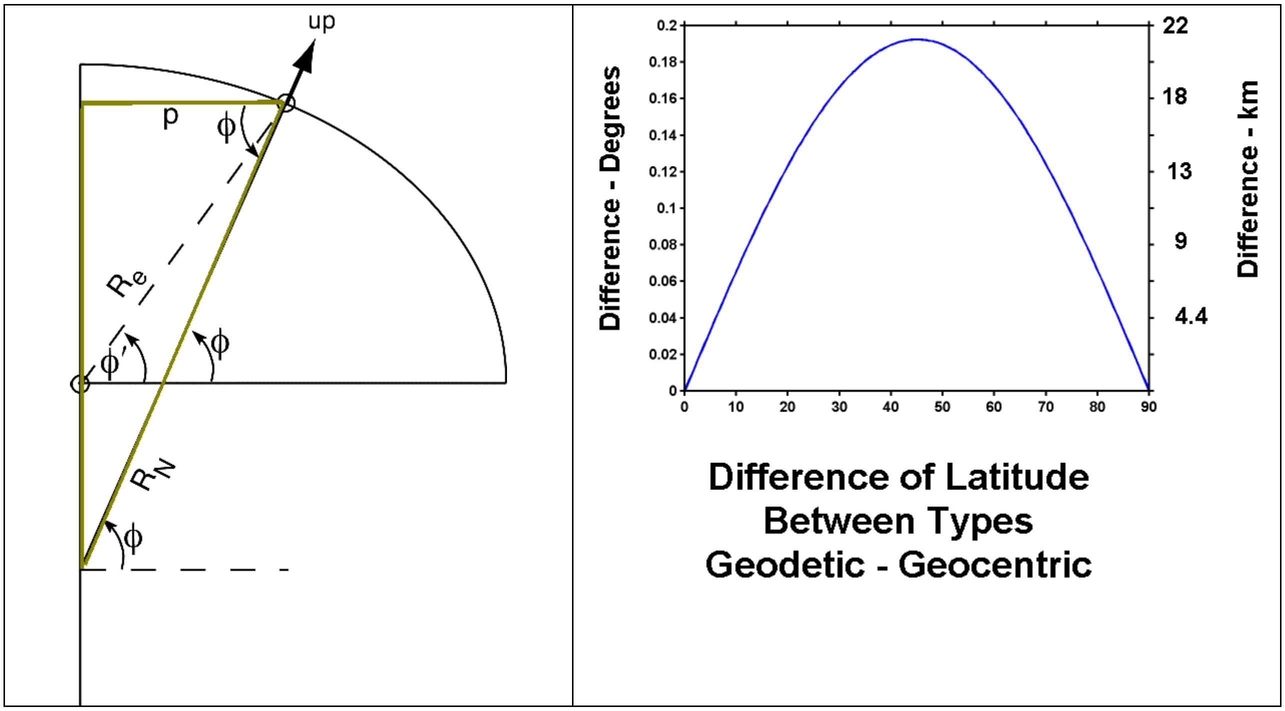
The figure below shows the key effects of rotation on the earth and coordinates. The latitude is defined in both the spherical and ellipsoidal cases from the line perpendicular to the world model. In the case of the spherical earth, this line hits the origin of the sphere - the center of the earth. For the ellipsoidal model the up-down line does not hit the center of the earth. It does hit the polar axis though.



Rotation Effects Change Latitude Defination

The length of the line to the center of the earth for a spherical model is the radius of the sphere. For the ellipsoidal model the length from the surface to the polar axis is one of three radii needed to work with angles and distance on the earth. (It is called the **radius of curvature in the prime vertical**, and denoted R_N here. See the document on radii of the earth for details.)

There are not two types of latitude that can easily be defined. (A third will be discussed under astrodetic coordinates.) The angle that the line makes from the center of the earth is called the geocentric latitude. **Geocentric latitude** is usually denoted as ϕ', or ϕ_c. It does not strike the surface of the ellipsoid at a right angle. The line perpendicular to the ellipsoid makes an angle with the equatorial plane that is called the **geodetic latitude**. ("Geodetic" is usually implies something taken with respect to the ellipsoid.) The latitude on maps is geodetic latitude. It is usually denoted as ϕ, or ϕ_g



Difference of Latitude Between Types Geodetic - Geocentric

Above on the left is a diagram showing the two types of latitude. The physical radius of the earth, R_e, and the radius of curvature in the prime vertical, R_N, are also shown. The highlighted triangle makes it clear that the radius of the parallel of latitude circle, called p, is just $R_N \cos(\phi)$, where the geodetic latitude is used.

From the figure it is also clear that the geodetic latitude is always greater or equal in magnitude to the geocentric. The two are equal at the poles and equator. The difference for the real world, with a flattening of about 1/300, is a small angle. However this can be a significant distance on the earth surface. The difference in degrees and kilometers on the earth surface are shown on the right above.

C. Heights

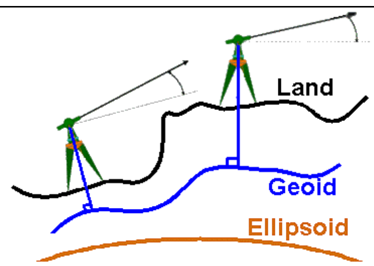
Rather than define the radial distance from the center of the earth for a point, it is more convenient to define the height. The question is what is the reference surface. For the spherical earth model, the answer is simple, the sphere with the radius of the mean earth is the reference.

For the ellipsoidal model of the earth, the ellipsoid can be used as the reference surface. This is done and the resulting height is called **ellipsoidal height or geodetic height**. But this is not the height found on maps.

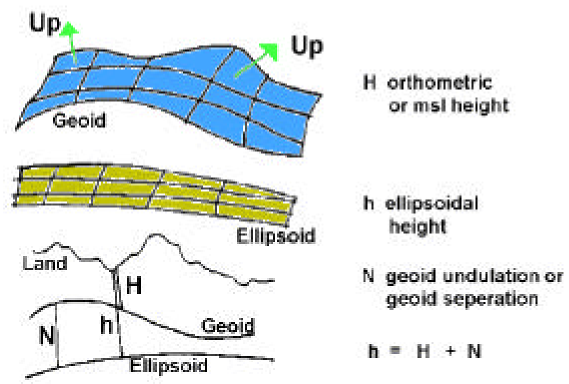
For practical purposes the height used on maps is **mean sea level (MSL) heights**. It has the official name of **orthometric height**. It is measured by going from point to point inland from points near the sea where mean sea level is determined by tide gauges. This is the height on maps.

One might assume that geodetic (ellipsoidal) height and MSL height would just be different by some constant that is the error in the tide gauge calibration. But this is not the case. The sea is not "flat", that is it is not an ellipsoid. The true gravity field of the earth is lumpy. Just as the up direction is perpendicular to the ellipsoid in the homogeneous earth case, for the real earth up is perpendicular to the geoid. This causes small variations in the "up" defined by the perpendicular to the ellipsoid and that defined operationally by a plumb bob or spirit level. This difference is called the **deflection of the vertical.**

On the left below is a diagram of how height are surveyed from point to point. At each station the local vertical is used via a spirit level. This causes the survey to follow a surface of constant gravity potential, not the smooth ellipsoid. This surface is called the geoid. It is the extension of sea level.



Classical Surveying Follows Geoid



Geoid - Ellisoid Diagram
Two and Three Dimensions

Thus there are two heights corresponding to the two reference surfaces. The operationally defined heights are msl or orthometric heights. They use the geoid as the reference. These will be called H here. The heights from the ellipsoid, geodetic or ellipsoidal heights, will be called h. The difference between then, the height of the geoid measured from the ellipsoid is called the undulation of the geoid or geoid separation. It is denoted by N. N must be measured. This is hard to do.

If you determine an ECEF Cartesian coordinate of a point, you can easily find the geodetic height h. Satellite based positions work this way. GPS positions are inherently ECEF Cartesian locations. Any other format is the result of a transformation. The error is on the order of a meter for absolute positions. If you determine a height by classical surveying you get msl or orthometric heights. Good surveys are accurate at the few cm level. You can measure N several different ways. Today the difference in N at near points has little error. But the error over the US might be 10's of cm. This is improving however.

D. SUMMARY of MODELS AND ANGULAR COORDINATE TYPES

MODEL	Surface	Latitude Longitude	Height
Spherical/Globe Used in elementary descriptions	Sphere	Geocentric	spherical
Ellipsoid/Ellipsoidal Used in mapmaking	Ellipse of Revolution	Geodetic **Used on maps**	ellipsoidal Produced by Satellite Systems (GPS etc.)
Real World	Geoid A Level[‡] Surface	Astrodetic or Astronomic	Orthometric (Mean Sea Level) **Used on maps**

[‡] Note: **Level surfaces** are surfaces of constant gravity potential. This is the potential energy per unit mass of something rotating with the earth. The sea is one level surface. (If we ignore some small effects due to ocean currents.) The up-down line as measured by a plumb bob or spirit level is always perpendicular to a level surface. The geoid is the level surface that represents mean sea level, but is also extended over the entire earth. It is usually under the land.

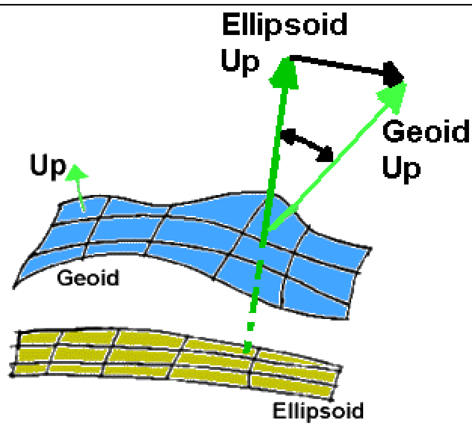
F. Angular to/from Cartesian

There is a detailed separate document on the transformation between these two types of coordinates. It will not be repeated here. The key issue is what information is needed.

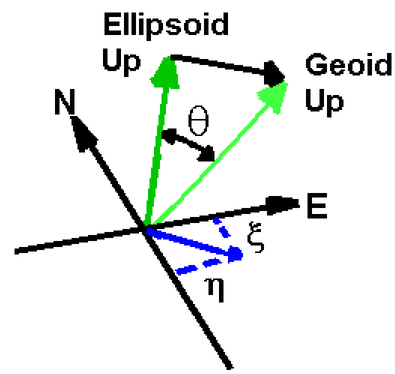
Spherical Model:	
ECEF to Longitude	No additional information needed
ECEF to Latitude	No additional information needed
ECEF to height	Radius of Earth needed
Latitude, longitude, height to ECEF	Radius of Earth
Ellipsoidal Model:	
ECEF to Longitude	No additional information needed
ECEF to Latitude	Ellipsoid parameter e (or f)
ECEF to ellipsoidal height	Ellipsoid parameters (a, f)
Ellipsoid height to MSL height	Geoid separation (undulation) N
Latitude, Longitude, Ellipsoidal height to ECEF	Ellipsoid parameters (a, f)

IV. Deflection of Vertical and Astrodetic Coordinates

The real world **"up" vector, or vertical**, is along the line of the gradient of the true gravity potential. It is perpendicular to the geoid. The gradient of the theoretical potential will give the perpendicular to the ellipsoid - the official definition of up. There will be a difference. This small difference in theoretical and measured verticals is called the **deflection of the vertical**. It is usually under an arc minute, often only a few arc seconds. It has components in the north-south and east-west.



Deflection of Vertical



Deflection of Vertical
North and East Components

The up vector responds to minor bumps in the geoid. The undulation can vary only a small amount, while the normal can vary quite a bit. The defection is more sensitive to near by density variations (mountains, etc.) than the undulation.

The deflection of the vertical will normally be in some general direction, not north-south or east-west. Therefore it has two components. The north south (latitude) component is usually called ξ (Greek Xi) and the east west η (Greek Eta). Therefore the deflection of the vertical has two components. These define the difference between geodetic and **astrodetic (or astronomic) coordinates.**

The **astrodetic latitude** and **longitude** are measured using the local vertical. They are related to the geodetic latitude and longitude (which use the perpendicular to the ellipsoid) by,

$$\sin(\phi) = \cos(\eta)\sin(\Phi - \xi)$$
$$\sin(\eta) = \cos(\phi)\sin(\Lambda - \lambda)$$

Where the capital Phi and Lambda are astrodetic values and the small letters are geodetic. Because the deflection is small, $\cos(\eta)$ can be taken as 1 and $\sin(\xi)$ as ξ. Then

$$\xi = (\Phi - \phi),$$
$$\eta = (\Lambda - \lambda)\cos(\phi)$$

V. Earth Reference Systems and Satellites

A. Earth and Sky Connections - Irregular Earth Motion

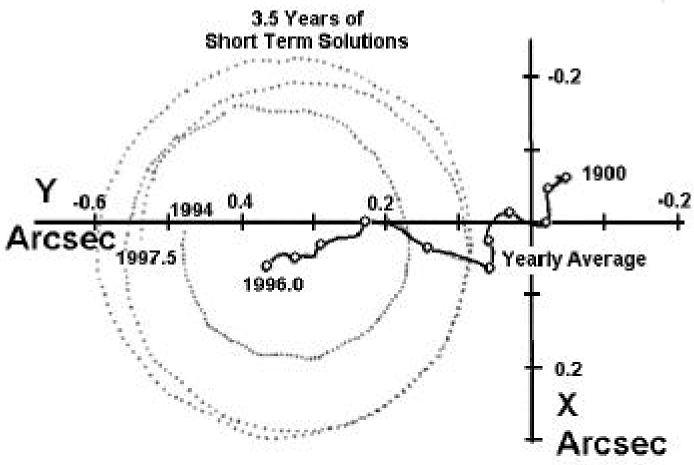
In dealing with earth based surveying and coordinates you can just drive a stake in the earth, call it the primary reference point and go from there. But if you want to connect to the sky, to the stars, to an **inertial reference system** you need more. Any time you deal with satellites, you are dealing with this **inertial space**, a non-rotating system. The facts may be hidden from you, as is done in GPS, but they are used somewhere.

In inertial space satellites go in nice ellipses (almost), but the ground tracks are complicated. For geostationary satellites, the track would be a point in theory. A polar orbiting satellite at 1200 km altitude makes a revolution in about 100 minutes. To this satellite it makes regular repeatable orbits and the earth rotates underneath it. The ground track is a set of curved lines.

The big complication comes in connecting inertial space to earth coordinates. The problem is the non-uniform motion of the earth. Not only does the earth go around the sun at a varying speed, faster in January when the earth is a little closer to the sun and slower in July, but there an many smaller motions. The polar axis is not fixed in inertial space. It moves in a circle with period of 26,000 years. The rotation axis points almost at the star Polaris now, but in 14,000 years will point at Vega. This is called the **precision or precision of the equinox**. In addition there is a smaller oscillation with a period of 18.6 years due to the moon. This is called **nutation**. These are smooth and predictable. But there are smaller, unpredictable motions. These are called the **polar motion**.

The polar motion has an irregular circular motion with a period of about 1.3 years. If one takes the average of these revolutions, then it is clear that even that center location is walking. This is evident in the following graph adapted from the **International Earth Rotation Service (IERS)**. Before satellite techniques were available, only long period averages were measurable with any accuracy. Today daily positions are computed by several groups and assembled by the IERS.

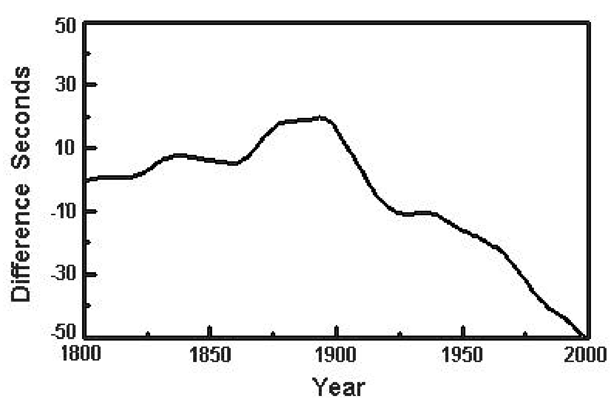
For reference, an arc-second measured from the center of the earth represents a distance of about 30 m. Therefore the diameter of these recent 1.3 year tracks is about 15 m.



Motion of Earth's Pole

In addition to the motion of the spin axis, the rate of spin of the earth is variable. This has been known for two centuries. Today the IERS also determines this from measurements and publishes the offset of a clock based on the earth rotation (called **Universal Time 1 or UT1**) and a "perfect" atomic clock (called **UT or UTA**). **Universal Time Coordinated (UTC)** is UTA with occasional jumps of 1 second (leap seconds) to keep it within 0.9 sec. of UT1.

The UT1 minus UT difference represents the integrated effect of small variations in the earth spin. A plot of the difference is shown below. Notice that several seconds of change have occurred, that the changes are increasing and decreasing over time, and that lately the rate has been running consistently lower.



Accumulated Earth Clock Error From 1800
IERS Measurements and Analysis

Unfortunately, in order to find spacecraft positions at high accuracy, computations must be done in an **inertial frame**. Thus measurement are made, sometimes from earth and adjusted to an inertial frame. The orbit is computed in this frame. The results must then be adjusted to earth fixed coordinates in some specific datum. This must include all the smooth (precession and nutation) effects as well as irregular effects of the earth's motion.

For the Global Positioning System this is all done at the central processing center. A model of the satellite motion in earth fixed coordinates is produced and broadcast to the user. The difficult work must be done, but it is done once for all the users of the broadcast ephemeris (BCE).

B. Celestial Reference Frames

Prior to the 1960's stars were the only "points of reference" that could be used as bench marks for a celestial reference system. There were a series of **catalogues of star positions** used just like bench marks for realizing the celestial coordinate system These were called the Fundamental Katalog (in German) and the systems were know as FK1, FK2 etc. The last such "celestial" datum was **FK5 J2000.5** . The date implies that the precision and nutation were used to adjust the coordinates to how they would be seen in the middle of year 2000. The locations were coupled to the motions of the earth.

In addition stars move. The apparent position of stars has small, but measurable changes. These are called proper motions. They are only important for near by stars, but those are generally the brightest stars and the most important in establishing the celestial reference frames. This meant that another factor would have to be included in the definitions.

Beginning in the 1960's radio astronomy began to use antennas separated by distance up to the diameter of the earth to form an effective single large antenna. This is called **Very Long Baseline Interferometry (VLBI)**. The angular resolution was extremely small. It was much smaller than could be achieved with optical telescopes on a practical basis. Thus there were separate celestial reference system for optical and radio astronomy.

In 1998 a new system was accepted by the International Astronomic Union that was based on point like radio sources outside the galaxy. This united the two systems and decoupled the reference system for inertial space measurements from models of the earth's motion.

C. ICRS and ICRF

The **International Celestial Reference System (ICRS)** is similar to the mathematical definition of a datum. It is the idealized reference frame. In this case it has some substance in the extragalactic radio sources. One particular source, 3C 273B has adopted, fixed coordinates that define the system. The IAU adopted this to replace FK5 in 1998. There are 23 extragalactic sources that are used to form the primary reference system. These have all been connected with VLBI measurements.

The **International Celestial Reference Frame (ICRF)** is analogous to the realization of a datum. It has measured coordinates of 212 extragalactic sources that form the first order network. (Some workers use 608 radio sources.) Each of these is measured with respect to several of the 23 primary sources.

In addition this system has been connected to a new, very accurate star catalogue. Between 1989 and 1993 the **Hippocras satellite** measured the very precise location of 118,218 stars. These star positions and proper motions are in the ICRF frame. There is an extended catalogue from this mission of over 2.5 million positions (called the **Tycho catalogue**) that are known at less accuracy. Between these, optical astronomers have a set of "bench marks in the sky" that can be used to locate objects.